

RESEARCH ARTICLE

Open Access

Development of microsatellite markers by transcriptome sequencing in two species of *Amorphophallus* (Araceae)

Xingfei Zheng^{1†}, Cheng Pan^{1†}, Ying Diao^{1,2*}, Yongning You¹, Chaozhu Yang³ and Zhongli Hu^{1*}

Abstract

Background: *Amorphophallus* is a genus of perennial plants widely distributed in the tropics or subtropics of West Africa and South Asia. Its corms contain a high level of water-soluble glucomannan; therefore, it has long been used as a medicinal herb and food source. Genetic studies of *Amorphophallus* have been hindered by a lack of genetic markers. A large number of molecular markers are required for genetic diversity study and improving disease resistance in *Amorphophallus*. Here, we report large scale of transcriptome sequencing of two species: *Amorphophallus konjac* and *Amorphophallus bulbifer* using deep sequencing technology, and microsatellite (SSR) markers were identified based on these transcriptome sequences.

Results: cDNAs of *A. konjac* and *A. bulbifer* were sequenced using Illumina HiSeq™ 2000 sequencing technology. A total of 135,822 non-redundant unigenes were assembled from about 9.66 gigabases, and 19,596 SSRs were identified in 16,027 non-redundant unigenes. Di-nucleotide SSRs were the most abundant motif (61.6%), followed by tri- (30.3%), tetra- (5.6%), penta- (1.5%), and hexa-nucleotides (1%) repeats. The top di- and tri-nucleotide repeat motifs included AG/CT (45.2%) and AGG/CCT (7.1%), respectively. A total of 10,754 primer pairs were designed for marker development. Of these, 320 primers were synthesized and used for validation of amplification and assessment of polymorphisms in 25 individual plants. The total of 275 primer pairs yielded PCR amplification products, of which 205 were polymorphic. The number of alleles ranged from 2 to 14 and the polymorphism information content valued ranged from 0.10 to 0.90. Genetic diversity analysis was done using 177 highly polymorphic SSR markers. A phenogram based on Jaccard's similarity coefficients was constructed, which showed a distinct cluster of 25 *Amorphophallus* individuals.

Conclusion: A total of 10,754 SSR markers have been identified in *Amorphophallus* using transcriptome sequencing. One hundred and seventy-seven polymorphic markers were successfully validated in 25 individuals. The large number of genetic markers developed in the present study should contribute greatly to research into genetic diversity and germplasm characterization in *Amorphophallus*.

Keywords: *Amorphophallus*, Microsatellite marker, Transcriptome, Genetic diversity

* Correspondence: diaoy78@163.com; huzhongli@whu.edu.cn

†Equal contributors

¹State Key Laboratory of Hybrid Rice, College of Life Sciences, Wuhan University, 430072 Wuhan, Hubei, PR China

²College of Forestry and Life Sciences, Chongqing University of Arts and Sciences, 402160 Yongchuan, Chongqing, PR China

Full list of author information is available at the end of the article

Background

Amorphophallus is a genus of perennial plants with an underground stem in the form of a corm and an umbrella-shaped leaf blade, which belongs to the family of Araceae, order of Alismatales [1,2]. The genus includes more than 170 species, which are mainly distributed in the tropics or subtropics of West Africa and South Asia [3]. *Amorphophallus* has been cultivated and utilized in China as a medicinal herb and food source for 2000 years. *Amorphophallus* contains abundant water-soluble dietary fiber known as glucomannan, which is beneficial to the human health, possibly being responsible for lowering lipids, systolic blood pressure and glycemia [4,5]. Thus *Amorphophallus* has great potential for commercial application. The diploid species *A. konjac* [6] is widely planted in China, its corm contained high-level glucomannan and possessed wide adaptability, but poor disease resistance and low coefficient of propagation [7]; in the contrast, *A. bulbifer* is a triploid ($2n = 39$) [6] with high disease resistance, characteristics of multi-seeding relay growth and high coefficient of propagation [8]. However, molecular genetic research of *Amorphophallus* is limited. Many studies have focused on extracting bioactive components [9], and some studies have attempted to identify *Amorphophallus* species taxonomically and to assess the genetic diversity of *Amorphophallus* [10-14]. However, few genomic molecular makers have been developed until now: only 32 microsatellite (SSR) markers were exploited by Santosa *et al.* and Pan *et al.* [15,16]. SSR markers have been shown to be an effective tool to carry out germplasm characterization and genetic diversity studies [17-21]. Thus, more markers are needed for an in-depth understanding of the natural diversity of *Amorphophallus* and to develop strategies for their sustainable utilization. Finally, as far as we know, gene discovery and molecular breeding of new *Amorphophallus* economic cultivars have not been initiated because of the lack of genetic and genomic information.

SSR makers have been widely used in population genetics and crop breeding [22]. According to the original sequences used for development of SSRs, SSRs were divided in two categories: genomic-SSRs, identified from random genomic sequences, and expressed sequence tag (EST)-SSRs, identified from transcribed RNA sequences. Compared with EST-SSRs, the development of genomic-SSRs is very expensive, labor intensive and timing consuming. Genomic-SSRs have uncertain linkage to the transcribed regions of the genome, while EST-SSRs are potentially linked with particular transcriptional regions that contribute to agronomic phenotypes [23,24]. In addition, EST-SSRs have high transferability, because sequences containing EST-SSRs are more conserved than sequences containing genomic-SSRs and can be utilized in other related species [25]. Large numbers of EST sequences produced by deep sequencing technology

have been deposited in the NCBI database and are being used for DNA fingerprinting, comparative mapping, and evolutionary analysis [26,27]. However, only a limited number of *Amorphophallus* EST sequence are available in the NCBI database.

Here, we report the generation of a large expressed sequence dataset based on Illumina HiSeq™ 2000 sequencing data from the young leaves of two *Amorphophallus* species, *A. konjac* and *A. bulbifer*. In addition, we developed EST-SSR markers and described their transferability in 14 species and 5 wild unknown “species”. The relationship among cultivars and wild individuals was investigated using the developed EST-SSRs.

Results

Assembly of *Amorphophallus* transcriptome data from Illumina sequencing

After stringent quality assessment and data filtering, Illumina HiSeq™ 2000 sequencing produced 54,986,020 reads for *A. konjac* and 52,334,098 reads for *A. bulbifer* with 92.27% Q20 bases and 93.48% Q20 bases (those with a base quality greater than 20), respectively (Table 1). The total length of the reads was about 9.66 gigabases (Gb). The raw data files are available in the Sequence Read Archive at the National Center for Biotechnology information (NCBI), with accession number SRA057020. Using the Trinity assembler software [28], short-read sequences from *A. konjac* and *A. bulbifer* were assembled into 187,459 contigs and 199,259 contigs, respectively (Table 1). The size distribution of these contigs is shown in Figure 1. Little difference was found in the contig lengths from these two *Amorphophallus* species; the contigs produced from both *A. konjac* and *A. bulbifer* were, on average, 276 bp long, with 381 bp and 372 bp median contig lengths (N50), respectively.

With paired-end reads, contigs can be identified from the same transcript and the distance between these contigs can be estimated. Trinity can be used to map the reads back to the contigs, and to connect the contigs into unigenes that cannot be further extended either end. As a result, 108,651 and 119,678 unigenes were obtained, with N50 values of 534 and 524 bp, from *A. konjac* and *A. bulbifer*, respectively (Table 2). To expand the utility of Illumina sequencing data, the contigs from the two species were pooled and assembled into a non-redundant unigene set for further analysis. As shown in Table 2, 135,822 non-redundant unigenes with an average length of 523 bp and an N50 of 635 bp, ranging from 271 to 2,431 bp, were obtained. The length distribution of two species and combined non-redundant unigenes are shown in Figure 2. The results revealed that the contig and unigene size distributions for two species were consistent, which implied the Illumina sequencing solution was reliable and appropriate. Among combined non-redundant

Table 1 Transcriptome reads and assembled contig information for two *Amorphophallus* species

Species	Total reads	Total clean Nucleotides (Nt)	Q20 percentage	GC percentage	Total number of contigs	Total length of contigs (Nt)	N50 of contigs	Mean (Nt)
<i>A. konjac</i>	54,986,020	4,948,741,800	92.27%	56.93%	187,459	51,822,856	381	276
<i>A. bulbifer</i>	52,334,098	4,710,068,820	93.48%	54.29%	199,257	54,982,674	372	276

unigenes, the length of 88,155 (64.90%) ranged from 100 to 500 bp, 33,529 unigenes (24.69%) ranged from 500 to 1000 bp, and 14,138 unigenes (10.41%) were more than 1000 bp in length.

Frequency and distribution of different types of SSR markers

Using a perl script known as MISA, 19,596 SSR loci were detected from 16,027 non-redundant unigenes, representing 11.8% of 135,822 unigenes. A total of 2,840 unigene sequences contained more than one SSR loci. In this study, the mononucleotide repeats were not considered. The probability of SSRs from combined non-redundant unigenes was 14.4% and the distribution density was one SSR loci per 3.63kb. Dinucleotide repeats were the most common type, with a frequency of 61.6% (12,069), followed by trinucleotide (30.3%, 5,943), tetranucleotide (5.6%, 1,083), pentanucleotide (1.5%, 300) and hexanucleotide repeats (1%, 201) (Figure 3A). The combined number of SSR loci from both dinucleotide and trinucleotide repeats was 18,012 with a frequency of 91.9%. The frequencies of SSR distributed in different reiteration numbers are shown in Figure 3B. The number of SSR

repeats ranged from 4 to 19, and SSRs with six repeats were the most abundant, followed by those with five tandem repeats, seven tandem repeats and eight random repeats. Motifs that showed more than 12 reiterations were very rare, with a frequency of <0.3%. SSR loci with a length of 18 bp were the most frequent (19.7%, 3,865) followed by those with 15 bp (17.3%, 3,388), 12bp (16.8%, 3,293), 16bp (14.6%, 2,860) and 14 bp (12.6%, 2,467). The longest SSR locus was 57 bp (Figure 3C). From 19,596 SSR loci, 205 different repeat motifs, considering sequencing complementary, were identified. The 13 most abundant SSR repeat motifs with different levels of repeats are shown in Table 3. The most abundant motifs were dinucleotide repeat units AG/CT, AT/AT and AC/GT with frequencies of 45.2%, 8.0% and 7.8%, respectively. Among the trinucleotide repeat units, AGG/CCT, AAG/CTT and CCG/CGG were the most abundant with frequencies of 7.1%, 6.1% and 5.3%, respectively (Table 3).

Development, validation and transferability of SSR markers
 Based on combined non-redundant unigene sequences containing 19,596 SSR loci, 10,754 primer pairs (Additional file 1) were designed using Primer3.0 [29]. Primer pairs for

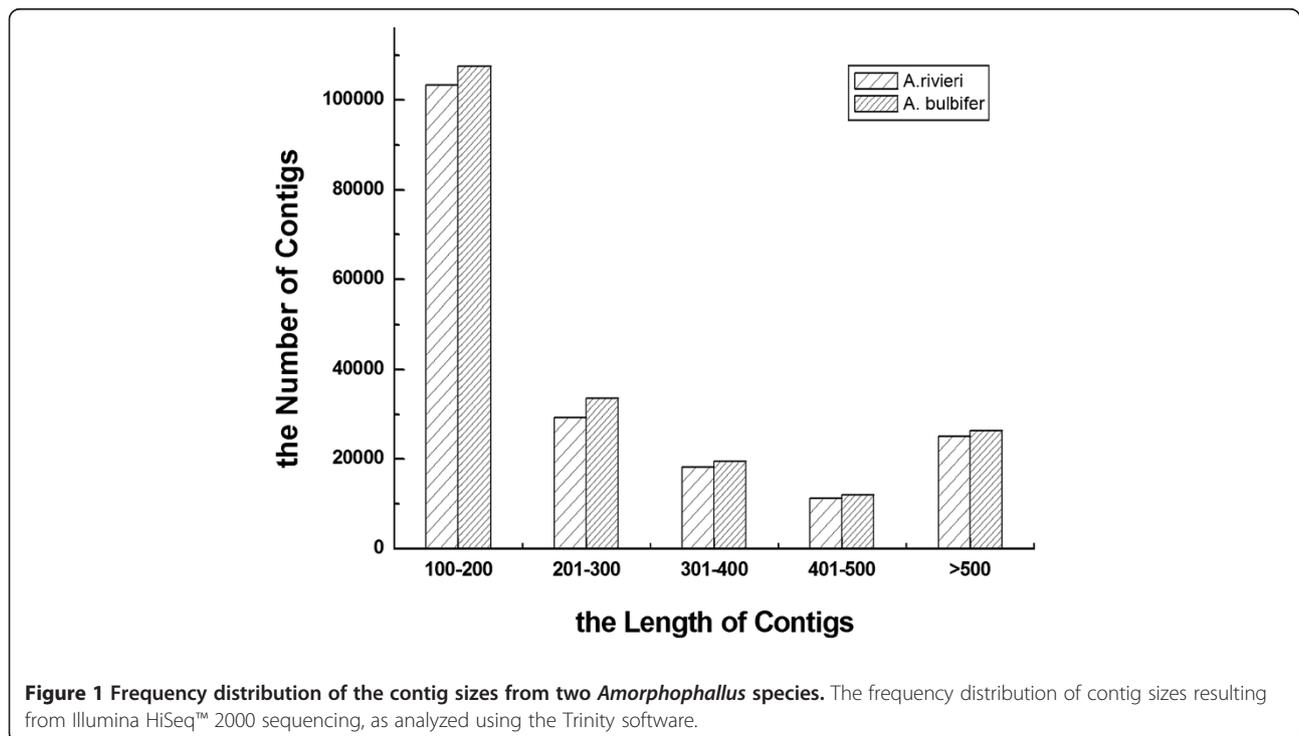


Figure 1 Frequency distribution of the contig sizes from two *Amorphophallus* species. The frequency distribution of contig sizes resulting from Illumina HiSeq™ 2000 sequencing, as analyzed using the Trinity software.

Table 2 Summary of the unigenes from two *Amorphophallus* species

Unigenes source	Total number of unigenes	Total length of unigenes (Nt)	Mean length of unigenes (Nt)	N50 of unigenes
<i>A. konjac</i>	108,651	47,721,327	439	534
<i>A. bulbifer</i>	119,678	51,480,410	430	524
All	135,822	71,092,160	523	635

the remaining 8,842 loci could not be designed successfully because the sequences flanking the SSR loci were either too short or were not appropriate for designing primers. From the 10,754 primer pairs, 320 were randomly selected for primer synthesis and validation (Additional file 2).

Among the 320 primer pairs, 270 were successful in PCR amplification using genomic DNA from *A. konjac* and *A. bulbifer*. The remaining 50-pair primers failed to generate PCR products, even when the annealing temperature was reduced by 8°C. Of the 270 working primer pairs, 198 yielded PCR products of the expected fragment size. The other 31 primer pairs generated multiple bands that did not include the target bands. In addition, 36 primer pairs amplified fragments larger than the expected size, probably because of the presence of an insertion mutation. Only five primer pair amplified a shorter than expected fragment, suggesting that a deletion mutation had occurred in the amplified region. To test the inter-species transferability, all 230 validated primer pairs, excluding the 31 pairs that generated multiple bands and nine pairs that generated more than 500 bp bands, were screened for polymorphism among 25 individual plants.

Thirty five pairs of primers could amplify PCR products from all *Amorphophallus* species (Additional file 3). Twelve pair of primers only produced PCR fragments in *A. bulbifer*; seven pairs (AK-EST-SSR55, AK-EST-SSR63, AK-EST-SSR72, AK-EST-SSR78, AK-EST-SSR83, AK-EST-SSR120 and AK-EST-SSR162) amplified successful from *A. bulbifer* and *A. bulbifer* cultivars; four pairs (AK-EST-SSR21, AK-EST-SSR55, AK-EST-SSR57 and AK-EST-SSR79) amplified successful from *A. bulbifer*, *A. yuloensis* and *A. bulbifer* cultivars; and twelve pairs amplified successful from *A. konjac*. The remaining 161 primer pairs successfully amplified fragments from 3 to 19 species. Therefore, the EST-SSRs developed from *A. konjac* and *A. bulbifer* could be successfully applied to other *Amorphophallus* species, with a transferability rate of 86.9%.

Of 230 primers, 25 primers were monomorphic; the other 205 primers were polymorphic; the proportion of polymorphic primers was 89.1%. From the 205 polymorphic loci, the number of alleles per locus ranged from 2 to 14 alleles. A total of 1,030 alleles were identified, with an average of 5.02 alleles per locus. Across 188 loci,

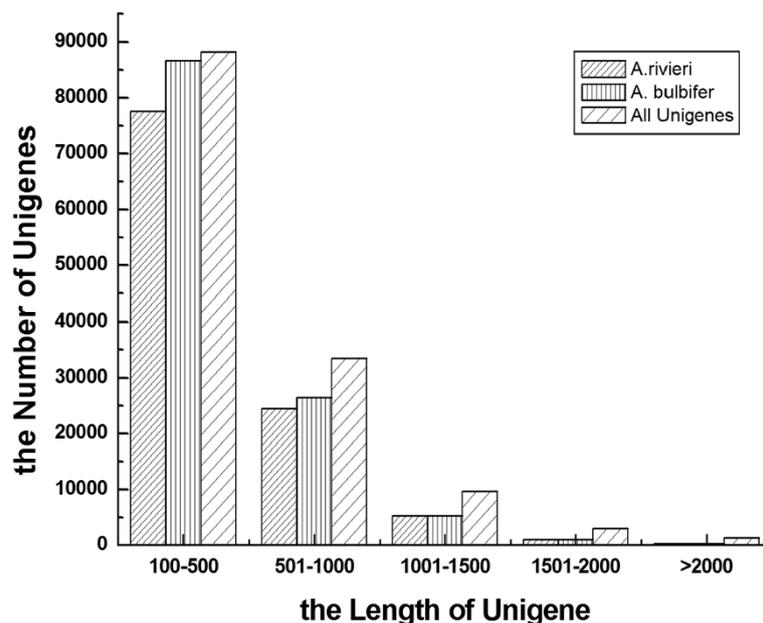
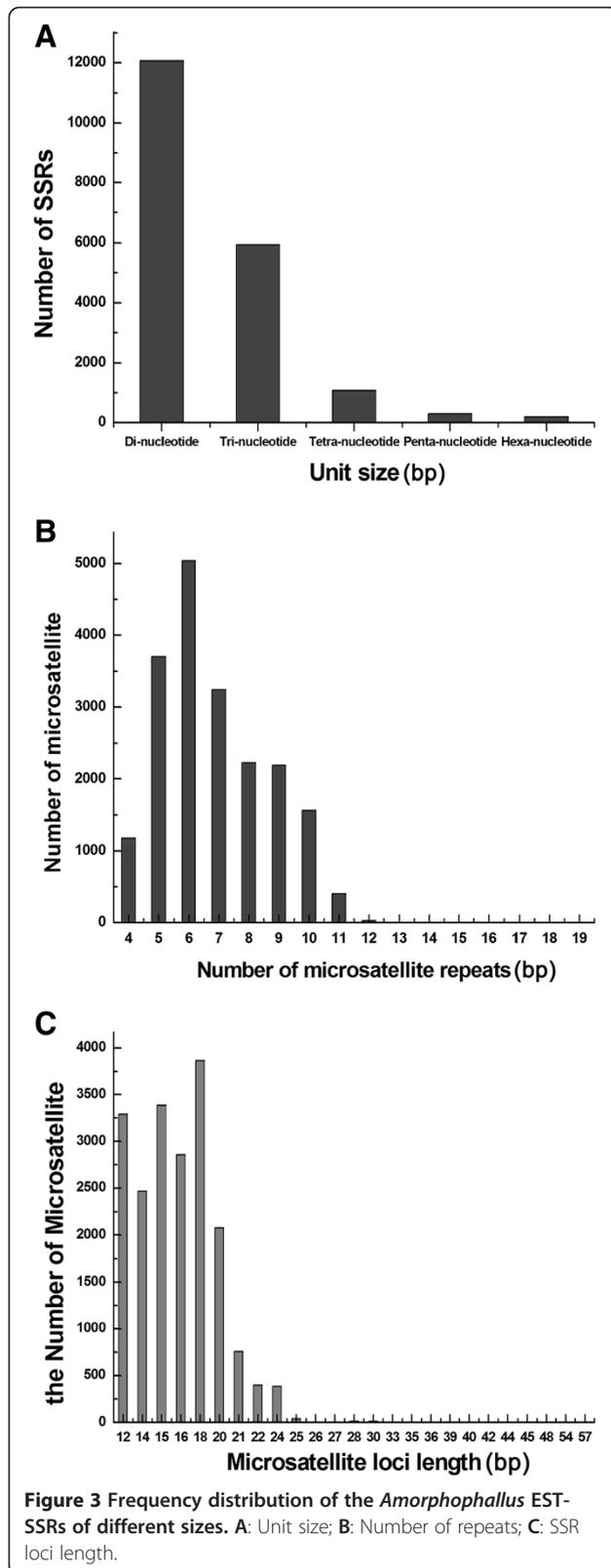


Figure 2 Size distribution of unigenes from two *Amorphophallus* species. The frequency distribution of unigene sizes from *A. konjac* and *A. bulbifer* and the combined unigenes, as analyzed with the Trinity software.



the polymorphic information content (PIC) ranged from 0.10 for AK-EST-SSR207 to 0.90 for AK-EST-SSR4 (Additional file 3), suggesting that the developed EST-SSRs were highly polymorphic. To determine the likely function of these EST-SSRs, they were subjected to BLAST analysis. The results showed that 124 transcriptome sequences from the 205 polymorphic SSR makers shared significant homology to other plants functional loci. The functional annotations of the developed markers are listed in Additional file 3.

The genetic diversity and cluster analysis in genus *Amorphophallus*

The 177 pair primers that yielded clear, highly polymorphic bands were used to assess the genetic diversity in a set of 25 individual plants representing diverse wild unknown species and cultivated genotypes of *Amorphophallus* (Table 4). A phenogram tree based on Jaccard's similarity coefficients was constructed, which showed three distinct clusters at a cut-off similarity index of 0.55 (Figure 4). Cluster I contained five species and five unknown "species", and was divided into three sub-clusters: Ia, Ib and Ic, at a cut-off similarity index of 0.71. Sub-cluster Ia comprised two wild *A. konjac*, three *A. konjac* cultivar genotypes, three Yunnan species and three unknown "species". Sub-cluster Ib comprised two unknown "species" from Chongqing and Yunan, respectively. Sub-cluster Ic comprised one wild *A. konjac* and one *A. kiusianus*. The wild *A. konjac* species from Kunming was distant from the other five *A. konjac* species. Cluster II included seven species and it was divided into three sub-clusters IIa, IIb and IIc, at a cut-off similarity index of 0.69. Sub-cluster IIa comprised one *A. bulbifer* and one *A. bulbifer* cultivar; sub-cluster IIb comprised one Hunan species and two Yunnan species; sub-cluster IIc comprised three species from Yunnan. Cluster III contained *A. albus* and *A. paeoniifolius* from Yunan and Chongqing, respectively.

Discussion

SSR marker frequency and distribution in *Amorphophallus* transcriptome

Genetic markers are important for studying population genetic structure, diversity analysis and the genetic basis of adaptive traits [30,31]. Deep transcriptome sequencing provides a good resource for the development of SSRs because of the enormous amount of sequence of data that it generates. Markers based on transcriptome sequences are useful for the detection of functional variation and gene associated genetic analysis [32]. Based on the 135,822 non-redundant unigene sequences, a total of 19,596 SSRs were identified. About 11.8% of the transcriptomic sequences possess SSR loci. This rate is higher than wheat (7.41%),

Table 3 Frequency distribution of the 13 most frequent SSR repeat motifs in the *Amorphophallus* transcriptome with different levels of repeats

s. no	Repeats motif	Number of repeats of the motif										total	
		5	6	7	8	9	10	11	12	13			
1	AG/CT	-	2233	1797	1632	1808	1166	217	11	2	8866	45.2%	
2	AT/AT	-	518	320	252	206	195	72	10	1	1574	8.0%	
3	AC/GT	-	487	325	222	173	200	106	11		1524	7.8%	
4	AGG/CCT	800	421	161	10					1	1393	7.1%	
5	AAG/CTT	638	358	177	22			1		2	1198	6.1%	
6	CCG/CGG	674	261	102	6	2					1045	5.3%	
7	AGC/CTG	508	259	128	12		1			1	909	4.6%	
8	ACC/GGT	261	128	59	11					1	460	2.3%	
9	ATC/ATG	157	86	60	13						316	1.6%	
10	ACG/CGT	136	56	24	6					2	223	1.1%	
11	AAT/ATT	107	53	24	10			2		3	199	1.0%	
12	AAC/GTT	79	29	15	9			1		1	134	0.7%	
13	CG/CG	-	55	25	18	3	2	2			105	0.5%	

barley (2.8%) and *Epimedium* (3.67%), but lower than coffee (18.5%) [33-36]. The EST-SSRs frequency is dependent on the search parameters for exploring SSR markers, e.g. the repeat length threshold and the number of repeat motifs unit. In our study, the mononucleotide repeat motifs were excluded for identifying SSRs because they can result from sequencing errors. In addition, the software used to detect SSR loci can affect the SSR frequency. Some tools can detect imperfect SSRs, e.g. Sputnik, while SSRIT and MISA (used here) can only detect perfect SSRs. The distribution density in *Amorphophallus* is one SSR loci per 3.63 kb, compared with 3.4 kb in rice, 5.4 kb in wheat, 7.4 kb in soybean, 14 kb in Arabidopsis, and 20 kb in cotton [34,37]. The difference in SSR density could partially depend on the minimum length of the SSR repeat motif [24]. In addition, Varshney *et al.* assumed that the high frequency of SSR in rice EST sequences might be due to its small genome size [38].

Dinucleotide repeat motifs are the most frequent repeat type among the *Amorphophallus* unigenes analyzed, representing 61.6% of the SSR loci identified, followed by Trinucleotide (30.3%), Tetranucleotide (5.6%), Pentanucleotide (1.5%) and Hexanucleotide (1%) (Figure 3A). This is consistent with the EST-SSRs distributions reported in pigeonpea, peach, spruce, pumpkin, kiwifruit and coffee, where the dinucleotide motif is the most frequent [39-43]. As shown in Table 3, the most abundant dinucleotide repeat motif was AG/CT (45.2%) in our transcriptome sequences. The same results were also found in other plant species [44-48]. The most abundant trinucleotide repeat motif was AGG/CCT (7.1%), closely followed by AAG/CTT (6.1%). These results agree with the report

from Poncet *et al.* [47]. Unlike the results obtained in *Amorphophallus*, the most frequency trinucleotide repeat motifs were AAC/TTG in wheat, AAG/CTT in soybean and *Epimedium sagittatum*, CCG/GGC in maize, barley and sorghum [36,38,49-51]. The previous studies suggested that the trinucleotide AAG/CTT is a common motif and that CCG/CGG is very rare in dicotyledonous plants. Interestingly, the trinucleotide CCG/CGG motif is common in monocots [52]. In this study, the trinucleotide CCG/CGG motif (5.3%) is the third most common repeat type, closely followed by AGG/CCT (7.1%) and AAG/CTT (6.1%). These results strongly support the view that the abundance CCG/CGG motif is a specific feature of monocot genomes, which could have resulted from the high GC content and consequent codon usage bias [53,54].

Polymorphism and genera transferability of SSR markers

EST-SSRs are very useful and attractive because they are located in coding regions of the genome. Therefore, they often show a high degree of transferability to related species [55-57]. In this study, 320 designed primer pairs were used for validation of the EST-SSR markers in *Amorphophallus*, 270 primer pairs (84.4%) yielded amplicons in *A. konjac* or *A. bulbifer*. This result was similar to the success rate of 60-90% amplification in previous studies [41,44,48,52,55,58,59]. From 230 primer pairs, we obtained 205 polymorphic EST-SSR markers within 25 samples, with the polymorphic proportion of 89.1%. Compared with other plants, the polymorphic ratio of EST-SSRs in *Amorphophallus* is very high [51,60]. Using these polymorphic markers, 1,030 alleles were identified, with an average of 5.02 alleles per locus in 25 individual plants. PIC value ranges from 0.10 to 0.90.

Table 4 The 25 individual plants (belonging to 14 species and 5 unknown “species”) used for validation and the genetic diversity study

No.	Species	Source	Characteristics
1	<i>A. konjac</i>	Chengdu, Sichuan	Cultivar
2	<i>A. konjac</i>	Fujian	Cultivar
3	<i>A. konjac</i>	Enshi, Hubei	Wild
4	<i>A. konjac</i>	Enshi, Hubei	Cultivar
5	<i>A. konjac</i>	Zhangjiajie, Hunan	Wild
6	<i>A. konjac</i>	Kunming, Yunnan	Wild
7	<i>A. yuloensis</i>	Kunming, Yunnan	Cultivar
8	<i>A. bulbifer</i>	Wuhan, Hubei	Cultivar
9	<i>A. bulbifer</i>	Wuhan, Hubei	Wild
10	<i>A. yunnanensisengl</i>	Kunming, Yunnan	Wild
11	<i>A. guripingensis</i>	Kunming, Yunnan	Wild
12	<i>A. sp*</i>	Kunming, Yunnan	Wild
13	<i>A. kiusianus</i>	Kunming, Yunnan	Wild
14	<i>A. maireiLevl.</i>	Kunming, Yunnan	Wild
15	<i>A. bannaensis</i>	Kunming, Yunnan	Cultivar
16	<i>A. tokinensis</i>	Kunming, Yunnan	Wild
17	<i>A. albus</i>	Kunming, Yunnan	Cultivar
18	<i>A. pingbianensis</i>	Kunming, Yunnan	Wild
19	<i>A. krauseiengler</i>	Kunming, Yunnan	Cultivar
20	<i>A. sp*</i>	Chongqing	Wild
21	<i>A. sp*</i>	Chongqing	Wild
22	<i>A. paeoniifolius</i>	Chongqing	Wild
23	<i>A. dunnii</i>	Changsha, Hunan	Wild
24	<i>A. sp*</i>	Enshi, Hubei	Wild
25	<i>A. sp*</i>	Myanmar	Wild

* *A. sp* means a wild unknown “species” (different variations or genotypes) with no clear classification.

Santosa *et al.* reported 19 polymorphic SSR markers with an average of 14.5 alleles per locus in *A. paeoniifolius* [15]. Santosa *et al.* also reported ten SSR markers with an average of 0.403 gene diversity (PIC) in 61 *A. paeoniifolius* individual plants [11]. The transferability of EST-SSRs among different related genera has been reported in many crop plants. Ellis *et al.* summarized the transferability of EST-SSRs and showed that the ratio of EST-SSRs cross-genera transferability ranged from 10% to 90% [61]. In this study, the *Amorphophallus* species had a high transferability rate (86.9%). This high success rate indicated that the different species in the genus *Amorphophallus* may evolutionarily closely related.

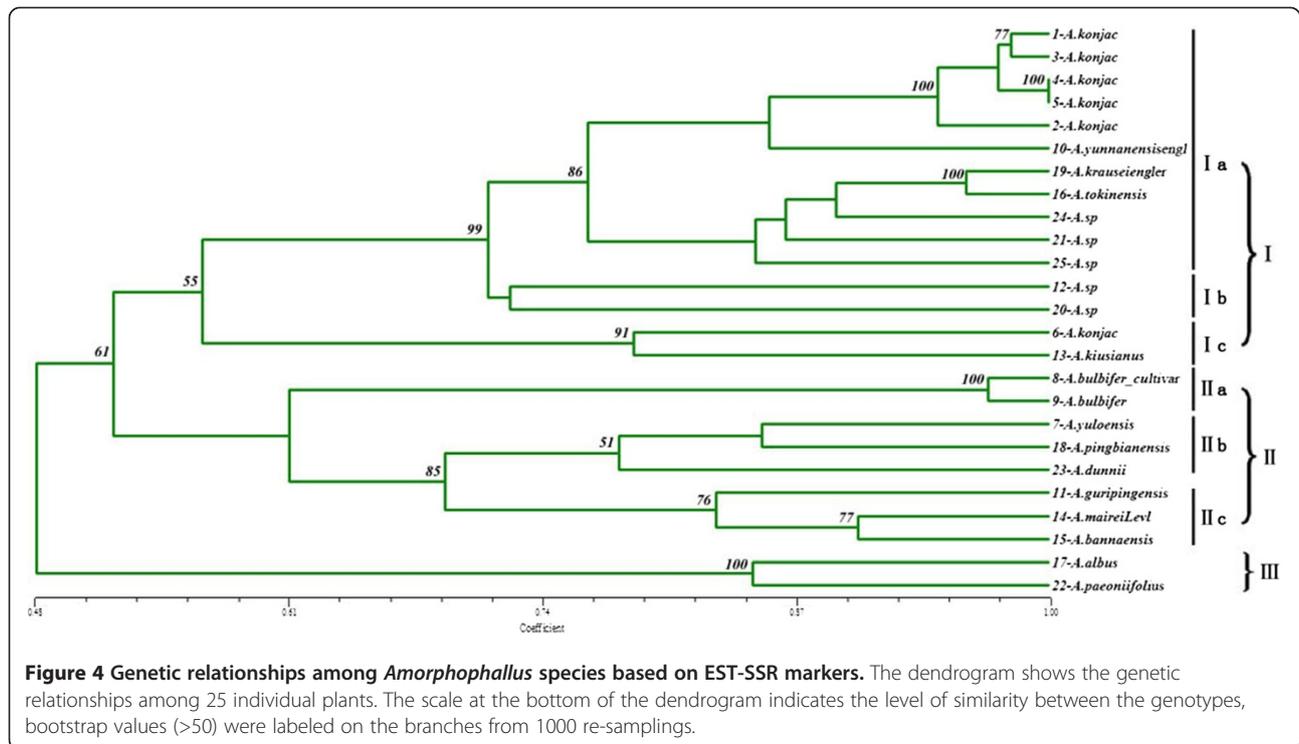
Evaluation of genetic relationships among different species of *Amorphophallus*

Collection of germplasm has been carried out in China, India and Japan. However, until now, evaluation of the

genetic diversity of *Amorphophallus* germplasm has been limited. In this study, three major groups were identified at a cut-off similarity index of 0.55. The level of genetic similarity (0.48-1.00) was revealed in the 25 individual plants, which included 14 species and 5 wild unknown “species”, using 177 polymorphic primers. The genetic variation between *A. konjac* and *A. paeoniifolius* was the highest among species in *Amorphophallus*. The intra-specific variation in *A. konjac* species is higher than that of some inter-species, and the *A. konjac* wild genotypes were clustered with cultivated genotypes. These results implied that *Amorphophallus* has a rich genetic diversity; the variation between wild and cultivar genotypes in *A. konjac* species was very small. This may be because the cultivars have been directly derived from the wild species without systematic genetic breeding. In addition, five unknown *Amorphophallus* “species” were used to evaluate their taxonomic status. 21-*A. sp* from Chongqing, 24-*A. sp* from Hubei and 25-*A. sp* from Myanmar were clustered in one group with *A. krauseiengler* and *A. tokinensis*. 20-*A. sp* and 12-*A. sp* comprised one group: sub-cluster Ib. Evaluation of the genetic status of these unknown species is valuable for utilizing new resources and for breeding. Sedayu *et al.* reported that the molecular phylogeny of the genus *Amorphophallus* contained 69 taxa, based on a combined analysis of trnL, rbcL and LEAFY second intron sequences, which reflected the biogeographical distribution of *Amorphophallus* [13]. Grob *et al.* also reported the phylogeny of 46 *Amorphophallus* using the nuclear FLORICAULA/LEAFY second intron [12]. Our results complement and support these previous reports.

Conclusions

A large-scale transcriptome dataset with 135,822 unigenes from *A. konjac* and *A. bulbifer* is reported. A total of 10,754 primer pairs were successfully designed based on the transcriptome sequences. Three hundred and twenty primer pairs from the combined non-redundant unigenes of the two sequenced *Amorphophallus* species were validated. Using the validated primers, the diversity of 25 individual plants containing 14 species and 5 wild unknown “species” was analyzed. The developed SSR markers are a valuable resource for studying genetic diversity, linkage mapping, and germplasm characterization analysis in *Amorphophallus* (Araceae). The ESR-SSR markers were developed based on conserved expressed sequences; therefore, they may be applied to other species because of their high transferability. To the best of our knowledge, this is the first attempt to exploit a transcriptome database and develop a large set of SSR markers in *Amorphophallus*.



Methods

Plant materials

Tender leaves of two species in the genus *Amorphophallus* (*A. bulbifer* and *A. konjac*) were collected for RNA extraction and transcriptome sequencing. Twenty-five individual plants were used for DNA extraction, PCR amplification, SSR marker validation and diversity analysis. All materials in this study were planted in the Wuhan University nursery (Wuhan, Hubei province, China). Detailed information for the plant materials is listed in Table 4.

RNA extraction, reverse transcription and sequencing

Total RNA was isolated using TRIzol kit (Invitrogen) according to the manufacturer's instructions and then purified to exclude the rRNA or tRNA for enrichment of mRNA using magnetic oligo(dT) microspheres. Fragmentation buffer was added to break the mRNA into short fragments (200nt-700nt). The mRNA fragments were reverse transcribed using Power-script™II with random hexamers primers. cDNA were then synthesized using DNA polymerase I. Double-stranded cDNA (ds cDNA) were purified using QiaQuick PCR kit and eluted with EB buffer. After ds cDNA end repair, poly (A) tailing and the addition of sequencing linkers, the correctly sized fragments were purified by Agarose Gel Electrophoresis gel and amplified using sequencing primers. The PCR products were sent to Illumina for HiSeq™ 2000 sequencing.

Development of SSR markers and primer designing of two *Amorphophallus* species

Two sets of expressed sequences reads were generated by high-throughput transcriptome sequencing from cDNA libraries of the two *Amorphophallus* species. High quantity filtered transcriptome reads were obtained by Illumina HiSeq™ 2000, and contigs were generated for the two species separately by *de novo* assembly using Trinity tools [28]. A non-redundant dataset of unigene sequences was then created using paired-end reads, which ensures the distance between different contigs from the same transcriptome. This unigene dataset was used for detecting SSR loci by a Perl script known as MicroSATellite (MISA, <http://pgrc.ipk-gatersleben.de/misa>) and primers were designed using Primer 3.0 [29]. The SSR loci were only considered to contain two to six nucleotides motifs with a minimum of 6,5,4,4 and 4 repeats, respectively. Mononucleotide repeats were excluded.

The parameters for designing PCR primers were as follows: (1) primer length ranging from 18 to 28 bases; (2) PCR product size range of 100 to 280 bp; (3) melting temperature between 57°C and 63°C, with 60°C as the optimum annealing temperature; (4) a GC content of 40%-60%, with an optimum of 50%.

Plant DNA extraction, PCR conditions and separation of SSR markers

Plant DNA was isolated from leaf samples of 25 individuals (Table 4) according to CTAB method [62]. Electrophoresis

through a 1.5% agarose gel was used to check DNA integrity. The SSR markers were initially tested for amplification using DNA from the two species used to generate the transcriptome sequences to optimize the annealing temperature. PCR amplifications were carried out using MyCycler™ Thermal Cycler (Bio-RAD) in a final volume of 15 µl. Each reaction tube contains 1.5 µl of 10 × PCR buffer, 0.6 µl of dNTP (10 mM), 1.0 µl of MgCl₂ (20 mM), 1 µl of each primer (10 pmol), and 2 µl of genomic DNA (50 ng), 0.5 µl of Taq DNA polymerase (1 U, Biostar). The PCR reaction program was: DNA denaturation at 94°C for 4 min; 36 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 1 min.; and 72°C for 7 min as a final extension. The primers that were not successful for amplification or produced multiple bands were reanalyzed using the touchdown PCR method with 1°C increments. The optimized SSR primers were used to amplify DNA from 25 plant individuals for genetic diversity analysis. The PCR products were separated using 6% polyacrylamide gel, 0.5 × TBE buffer in an EC160 DNA Sequencing System (Thermo EC). Gels were stained with silver nitrate, following the protocol described in Han *et al.* [63]. A PBR322 DNA marker ladder (Tiangen Biotech, Beijing co., LTD.) was used for assessing the length of the DNA bands.

Genetic analysis and data scoring

The PCR products were scored manually in binary format, with the absence of a band being scored as 0 and its presence as 1, thus generating a binary matrix. The polymorphism information content (PIC) of each EST-SSR primer pair was calculated using the formula as following:

$$PIC = 1 - \sum_{i=0}^n P_i^2$$

where, P_i is the frequency of the i^{th} allele for a given SSR loci, and n is the total number of alleles detected for the SSR markers [64]. The genetic similarity between any two individuals was assessed based on Jaccard's similarity coefficient. All 25 individuals were clustered using UPGMA analysis and the SHAN clustering program by NTSYS-pc v2.10 t [65]. Bootstrapping analysis (1000 re-samplings) was carried out using software FREETREE V.0.9.1.50 [66]. Bootstrap values over 50 are considered significant and provided on the dendrogram.

Additional files

Additional file 1: Primer sequences for SSR loci. SSR primers were designed using Primer 3.0. The sequence ID, Unigene Source, SSR type, SSR motif type, motif size, location of the SSR, location of both primers, primer sequences, length of primers, melting temperature, and amplicon product size are provided for each locus.

Additional file 2: Characterization of 320 primer pairs amplification in 25 individual plants. Primer name, sequence ID, repeat motif, sequence of both primers, melting temperature and expected size (bp) are indicated.

Additional file 3: Details of 230 validated EST-SSR markers and predicted functions of their genes, based on BLASTX. Primer name, sequence ID, repeat motif, expected size (bp), objective size (bp), number of species from which it was successfully amplified, N. alleles, PIC value, Swissprot-annotation, and COG-Function-Description are indicated.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PC and YYN collected plant materials and performed the SSRs validation and diversity analysis. DY and HZL designed the study. The transcriptome data processing and writing of this manuscript were completed by ZXF. In addition, YCZ participated in data analysis and material collection. All the authors read and approved the final manuscript.

Acknowledgements

We would like thank Professor Yunchun Song for correcting the English. We also thank to the editor and reviewer for their constructive suggestions and comments. This study was supported by grants from the National Science and Technology Supporting Program (No. 2011BAD33B03), the National Natural Science Foundation of China (No. 30970195) and the Science and Technology Project of Chongqing Education Commission (No. KJ101202).

Author details

¹State Key Laboratory of Hybrid Rice, College of Life Sciences, Wuhan University, 430072 Wuhan, Hubei, PR China. ²College of Forestry and Life Sciences, Chongqing University of Arts and Sciences, 402160 Yongchuan, Chongqing, PR China. ³Agricultural Science Academy of Enshi Autonomous Prefecture, 445002 Enshi, Hubei, PR China.

Received: 20 September 2012 Accepted: 12 July 2013

Published: 19 July 2013

References

- Gille S, Cheng K, Skinner ME, Liepman AH, Wilkerson CG, Pauly M: Deep sequencing of voodoo lily (*Amorphophallus konjac*): an approach to identify relevant genes involved in the synthesis of the hemicellulose glucomannan. *Planta* 2011, **234**(3):515–526.
- Chua M, Baldwin TC, Hocking TJ, Chan K: Traditional uses and potential health benefits of *Amorphophallus konjac* K. Koch ex NE Br. *J Ethnopharmacol* 2010, **128**(2):268–278.
- Hettterscheid W, Ittenbach S: Everything you always wanted to know about *Amorphophallus*, but were afraid to stick your nose into. *Aroideana* 1996, **19**:7–131.
- Sood N, Baker WL, Coleman CI: Effect of glucomannan on plasma lipid and glucose concentrations, body weight, and blood pressure: systematic review and meta-analysis. *Am J Clin Nutr* 2008, **88**(4):1167–1175.
- Arvill A, Bodin L: Effect of short-term ingestion of konjac glucomannan on serum cholesterol in healthy men. *Am J Clin Nutr* 1995, **61**(3):585–589.
- Chauhan K, Brandham P: Chromosome and DNA variation in *Amorphophallus* (Araceae). *Kew Bulletin* 1985, **40**(4):745–758.
- X-q CUN, Y-c LU, QUAN H, Q-y ZHAO, S-q XIE: Yield and Quality Analysis of Different Varieties of *Amorphophallus* konjac in Yunnan Province. *J Yunnan Agric Univ* 2009, **5**:027.
- D-h ZHANG, Q-p WANG, Z-b DUAN, K-x MI: Mechanism of Relay Multi-seedling Release *Amorphophallus* bulbifer and Its Application in Southeast Asia. *Resour Dev Market* 2009, **8**:004.
- Das D, Mondal S, Roy SK, Maiti D, Bhunia B, Maiti TK, Islam SS: Isolation and characterization of a heteropolysaccharide from the corm of *Amorphophallus campanulatus*. *Carbohydr Res* 2009, **344**(18):2581–2585.
- Lekhak MM, Yadav SR: Cytotaxonomy of some species of *Amorphophallus* sect. *Rhaphiophallus* (Schott) Engl. (Araceae) of the Indian Subcontinent. *Nucleus* 2011, **54**(3):169–176.
- Santosa E, Mine Y, Nakata M, Lian C, Sugiyama N: Genetic diversity of cultivated elephant foot yam (*Amorphophallus paeoniifolius*) in

- Kuningan, West Java as revealed by microsatellite markers. *Complimentary Copy* 2010, **12**(2):125–128.
12. Grob G, Gravendeel B, Eurlings M: Potential phylogenetic utility of the nuclear FLORICAULA/LEAFY second intron: comparison with three chloroplast DNA regions in *Amorphophallus* (Araceae). *Mol Phylogenet Evol* 2004, **30**(1):13–23.
 13. Sedayu A, Eurlings MCM, Gravendeel B, Hettterscheid WLA: Morphological character evolution of *Amorphophallus* (Araceae) based on a combined phylogenetic analysis of trnL, rbcL and LEAFY second intron sequences. *Bot Stud* 2010, **51**(4):473–490.
 14. Wenbing C, Chunlin L, Jianfu Z: Study on genetic diversity of RAPD markers in *Amorphophallus*. *J Agric Biotechnol* 2001, **4**:024.
 15. Santosa E, Lian C, Piskoosantivatana Y, Sugiyama N: Isolation and characterization of polymorphic microsatellite markers in *Amorphophallus paeoniifolius* (Dennst.) Nicolson, Araceae. *Mol Ecol Notes* 2007, **7**(5):814–817.
 16. Pan C, You Y, Diao Y, Hu Z, Chen J: Isolation and characterization of microsatellite loci for the herbaceous tuber crop, *Amorphophallus konjac* (Araceae). *Genet Mol Res* 2012, **11**(4):4617–4621.
 17. Hao C, Zhang X, Wang L, Dong Y, Shang X, Jia J: Genetic diversity and core collection evaluations in common wheat germplasm from the Northwestern Spring Wheat Region in China. *Mol breeding* 2006, **17**(1):69–77.
 18. Barkley NA, Roose ML, Krueger RR, Federici CT: Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *Theor Appl Genet* 2006, **112**(8):1519–1531.
 19. Folkertsma RT, Rattunde HFW, Chandra S, Raju GS, Hash CT: The pattern of genetic diversity of Guinea-race *Sorghum bicolor* (L.) Moench landraces as revealed with SSR markers. *Theor Appl Genet* 2005, **111**(3):399–409.
 20. Song Z, Xu X, Wang B, Chen J, Lu B-R: Genetic diversity in the northernmost *Oryza rufipogon* populations estimated by SSR markers. *Theor Appl Genet* 2003, **107**(8):1492–1499.
 21. Belaj A, Satovic Z, Cipriani G, Baldoni L, Testolin R, Rallo L, Trujillo I: Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. *Theor Appl Genet* 2003, **107**(4):736–744.
 22. Gupta P, Varshney R: The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 2000, **113**(3):163–185.
 23. Bozhko M, Riegel R, Schubert R, Müller-Starck G: A cyclophilin gene marker confirming geographical differentiation of Norway spruce populations and indicating viability response on excess soil-born salinity. *Mol Ecol* 2003, **12**(11):3147–3155.
 24. Varshney RK, Graner A, Sorrells ME: Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 2005, **23**(1):48–55.
 25. Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ: Analysis of SSRs derived from grape ESTs. *TAG Theor Appl Genet* 2000, **100**(5):723–726.
 26. Yu JK, La Rota M, Kantety R, Sorrells M: EST derived SSR markers for comparative mapping in wheat and rice. *Mol Genet Genomics* 2004, **271**(6):742–751.
 27. Luro F, Costantino G, Terol J, Argout X, Allario T, Wincker P, Talon M, Ollitrault P, Morillon R: Transferability of the EST-SSRs developed on *Nules clementine* (*Citrus clementina* Hort ex Tan) to other *Citrus* species and their effectiveness for genetic mapping. *BMC Genomics* 2008, **9**(1):287.
 28. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011, **29**(7):644–652.
 29. Rozen S, Skaletsky H: Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000, **132**(3):365–386.
 30. Beaumont MA: Adaptation and speciation: what can F_{ST} tell us? *Trends Ecol Evol* 2005, **20**(8):435–440.
 31. Plaschke J, Ganai M, Röder M: Detection of genetic diversity in closely related bread wheat using microsatellite markers. *TAG Theor Appl Genet* 1995, **91**(6):1001–1007.
 32. Vasemägi A, Nilsson J, Primmer CR: Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Mol Biol Evol* 2005, **22**(4):1067–1076.
 33. Varshney R, Grosse I, Hähnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmid L, Graner A: Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. *TAG Theor Appl Genet* 2006, **113**(2):239–250.
 34. Peng J, Lapitan NLV: Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. *Funct Integr Genomics* 2005, **5**(2):80–96.
 35. Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L: Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *TAG Theor Appl Genet* 2007, **114**(2):359–372.
 36. Zeng S, Xiao G, Guo J, Fei Z, Xu Y, Roe BA, Wang Y: Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. *BMC Genomics* 2010, **11**(1):94.
 37. Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R: Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 2000, **156**(2):847–854.
 38. Varshney RK, Thiel T, Stein N, Langridge P, Graner A: In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett* 2002, **7**(2A):537–546.
 39. Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Raje RS, Bandhopadhy TK: Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol* 2011, **11**(1):17.
 40. Xu Y, Ma R-C, Xie H, Liu J-T, Cao M-Q: Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. *Genome* 2004, **47**(6):1091–1104.
 41. Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K: Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *TAG Theor Appl Genet* 2004, **109**(6):1283–1294.
 42. Gong L, Stift G, Kofler R, Pachner M, Lelley T: Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L. *TAG Theor Appl Genet* 2008, **117**(1):37–48.
 43. Fraser L, Harvey C, Crowhurst R, Silva HND: EST-derived microsatellites from *Actinidia* species and their potential for mapping. *TAG Theor Appl Genet* 2004, **108**(6):1010–1016.
 44. Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang L, May GD: Tall fescue EST-SSR markers with transferability across several grass species. *TAG Theor Appl Genet* 2004, **109**(4):783–791.
 45. Gao L, Tang J, Li H, Jia J: Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol breeding* 2003, **12**(3):245–261.
 46. Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, Chen X, Li Y: De novo assembly and characterization of root transcriptome using illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genomics* 2010, **11**(1):726.
 47. Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, De Kochko A, Hamon P: SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Mol Genet Genomics* 2006, **276**(5):436–449.
 48. Liang X, Chen X, Hong Y, Liu H, Liu H, Zhou G, Li S, Guo B: Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. *BMC Plant Biol* 2009, **9**(1):35.
 49. Thiel T, Michalek W, Varshney R, Graner A: Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor appl genot* 2003, **106**(3):411–422.
 50. Kantety RV, La Rota M, Matthews DE, Sorrells ME: Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* 2002, **48**(5):501–510.
 51. Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A: Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. *Mol Biol Evol* 1998, **15**(10):1275–1287.
 52. Zhangying W, Jun L, Zhongxia L, Lifei H, Boping F, Yujun L, Jingyi C, Xiongjian Z: Characterization and development of EST-derived SSR markers in cultivated sweetpotato (*Ipomoea batatas*). *BMC Plant Biol* 2011, **11**(1):139.
 53. La Rota M, Kantety R, Yu JK, Sorrells M: Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* 2005, **6**(1):23.

54. Morgante M, Hanafey M, Powell W: **Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes.** *Nat Genet* 2002, **30**(2):194–200.
55. Gupta P, Rustgi S, Sharma S, Singh R, Kumar N, Balyan H: **Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat.** *Mol Genet Genomics* 2003, **270**(4):315–323.
56. Bory S, Da Silva D, Risterucci AM, Grisoni M, Besse P, Duval MF: **Development of microsatellite markers in cultivated vanilla: Polymorphism and transferability to other vanilla species.** *SciHortic-Amsterdam* 2008, **115**(4):420–425.
57. Vendramin E, Dettori M, Giovinazzi J, Micali S, Quarta R, Verde I: **A set of EST-SSRs isolated from peach fruit transcriptome and their transportability across Prunus species.** *Mol Ecol Notes* 2007, **7**(2):307–310.
58. Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ: **Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum.** *Plant Sci* 2001, **160**(6):1115–1123.
59. Yu JK, Dake TM, Singh S, Benscher D, Li W, Gill B, Sorrells ME: **Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat.** *Genome* 2004, **47**(5):805–818.
60. Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A: **Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice.** *Plant Sci* 2005, **168**(1):195–202.
61. Ellis J, Burke J: **EST-SSRs as a resource for population genetic analyses.** *Heredity* 2007, **99**(2):125–132.
62. Porebski S, Bailey LG, Baum BR: **Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components.** *Plant Mol Biol Rep* 1997, **15**(1):8–15.
63. Han YC, Teng CZ, Hu ZL, Song YC: **An optimal method of DNA silver staining in polyacrylamide gels.** *Electrophoresis* 2008, **29**(6):1355–1358.
64. Botstein D, White RL, Skolnick M, Davis RW: **Construction of a genetic linkage map in man using restriction fragment length polymorphisms.** *Am J Hum Genet* 1980, **32**(3):314.
65. Rolf J: *Numerical Taxonomy and Multivariate Analysis System, version 2.11 T Exeter Software.* Setauket, NY, USA; 2000.
66. Flegr J, Hrdá S, Pavlíček A: **Free-tree=free-ware program for construction of phylogenetic trees on the basis of distance data and bootstrap/jackknife analysis of the tree robustness. Application in the RAPD analysis of genus *Frenkelia*.** *Folia Biol* 1999, **45**(3):97.

doi:10.1186/1471-2164-14-490

Cite this article as: Zheng *et al.*: Development of microsatellite markers by transcriptome sequencing in two species of *Amorphophallus* (Araceae). *BMC Genomics* 2013 **14**:490.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

