CrossMark

Development of EST-SSR markers derived from transcriptome of *Saccharina japonica* and their application in genetic diversity analysis

Jing Zhang[1] · Tao Liu[2] · Fengping Rui[2]

Abstract

Expressed sequence tag (EST)-derived simple sequence repeat (SSR) markers were developed and characterized for *Saccharina japonica* using transcriptome data. A total of 98,627 ESTs were screened for SSRs using microsatellite search module (MISA) tool. From those, 9688 SSRs in 8039 ESTs were identified and trinucleotides accounted for 57.69% were the predominant types. Among these EST-SSRs, 1120 SSR primer pairs were ultimately designed and 146 could be used for genetic study by polymorphism tests using four *Saccharina* DNAs as templates. In the present study, 52 EST-SSR markers with good representativeness were selected to conduct genetic analyses of ten Chinese strains. A total of 139 alleles were detected and alleles per locus ranged from 2 to 5. The percentage of polymorphic loci (P) per strain ranged from 28.85 to 73.08%, Nei's genetic diversity (H) ranged from 0.1207 to 0.3376 and Shannon's information index (I) from 0.1819 to 0.5267. Genetic identity and cluster analysis of *Saccharina* strains were also performed based on the obtained EST-SSR data in our work. As shown by genetic structure analysis, ten strains could be classified into three groups which associated with their parental origin. These obtained EST-SSR markers will be helpful for genetic analysis and molecular marker assisted breeding in *Saccharina*.

Keywords *Saccharina* · Phaeophyta · Expressed sequence tag (EST) · Simple sequence repeat (SSR) · Transcriptome · Genetic diversity

Introduction

Saccharina (Laminariales, Phaeophyceae), one of the most important economic algae, can be used as food and raw biochemical industry materials due to its algin, mannitol, and iodine extraction (Jensen 1993; Porse and Rudolph 2017). It also plays a key role in maintaining offshore ecological balance (An et al. 2010). Additionally, *Saccharina* has been considered as a potential model organism for genetic study of seaweeds with respect to the heterogenesis, parthenogenesis, and apogamy in the life history (Waaland et al. 2004). Because

of its important economic, ecological, and genetic value, *Saccharina* is cultured in large scale in the region of Western Pacific, especially in China (Liu et al. 2012). So far, more than 20 elite varieties have been bred in China and its mariculture industry ranks first among the world (Liu et al. 2010b). However, some problems for *Saccharina* breeding, including genetic confusion, inbreeding depression, and limited improved varieties have not been resolved at present. Applying molecular markers for genetic study systematically will improve the breeding work.

Recently, several molecular markers for *Saccharina* such as randomly amplified polymorphism of DNA (RAPD) (Billot et al. 1999a; Yotsukura et al. 2001; Wang et al. 2004; Bi et al. 2011), inter-simple sequence repeat (ISSR) (Wang et al. 2005), and amplified fragment length polymorphism (AFLP) (Li et al. 2007b; Yang et al. 2009; Liu et al. 2010a; Shan et al. 2011) have been used in germplasm identification, genetic diversity, population genetics, genetic mapping, and QTL analysis. SSRs which also known as microsatellites are widely utilized in many species including tobacco, rice, maize, wheat, poplar, pine, apple, and peach (Gupta et al. 1996; Rungis et al. 2004; Chen et al. 2005).

✉ Tao Liu
 myscience@yeah.net

[1] Qilu University of Technology, 3501 Daxue Road,
 Jinan 250353, Shandong Province, People's Republic of China

[2] College of Marine Life Sciences, Ocean University of China, 5
 Yushan Road, Qingdao 266003, Shandong Province, People's
 Republic of China

Compared with other molecular markers, SSR markers with Mendelian inheritance and co-dominance are more simple, stable, abundant, highly reproducible, and polymorphic (Varshney et al. 2002; Chen et al. 2005). For *Saccharina* and *Laminaria*, studies associated with SSRs have made a lot of progress in recent years. There were ten polymorphic genomic SSRs (gSSRs) isolated from *Laminaria digitata* genomic libraries of small inserts (Billot et al. 1999b), 18 gSSR markers developed from *S. japonica* using FIASCO method (Shi et al. 2007), 27 polymorphic trinucleotide gSSR markers developed from *S. japonica* through paired-end Illumina sequencing data (Zhang et al. 2014b), and 23 gSSRs from the SSR-enriched genomic library of *S. japonica* (Zhang et al. 2015). In addition, nine EST-SSR markers with polymorphism were generated from *S. japonica* EST sequences (Liu et al. 2010b), 23 EST-SSR polymorphic markers were developed from *L. digitata* ESTs and *S. japonica* ESTs (Wang et al. 2011), 13 EST-SSR markers for wild *S. japonica* population using *L. digitata* ESTs (Liu et al. 2012). Even so, SSR works of this important seaweed is lagged far behind that of the high plants and more SSR markers are indispensable to be developed effectively in order to satisfy the need of molecular genetics.

In contrast with gSSRs developed from genome by enrichment and sequencing of genomic libraries (Edwards et al. 1996), EST-SSRs are more efficient and less expensive (Bouck and Vision 2007). Moreover, they associated with functional genes that can be more useful in construction of physics map and marker-assistant selection. However, only 984 ESTs from *Saccharina* and 4102 ESTs from *Laminaria* are available in the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) for the moment which had hampered the advancement of EST-SSR markers. Transcriptome sequencing is a useful method to produce EST datasets for expression pattern construction, molecular marker development, and novel gene identification (Li et al. 2012). Fortunately, transcriptome of *S. japonica* has been de novo sequenced by our team using the next generation sequencing (NGS) technologies (Liang et al. 2014) and yielded a large amount of ESTs. Here, we developed a batch of EST-SSR markers based on transcriptome data and conducted their practicability test. In addition, genetic diversity and population structure of ten *Saccharina* strains which represented almost all the most widely used commercial varieties in China were evaluated using the developed polymorphic EST-SSR markers in our study. This work will enhance the genetic research and promote the molecular marker-assistant selection of *Saccharina*.

Materials and methods

EST-SSRs identification and primer design

ESTs used in this study were generated from the transcriptome of *S. japonica* which has been sequenced using Illumina HiSeq 2000 by our team (Liang et al. 2014) as one part of 1KP Project (http://www.onekp.com/). Microsatellites were searched from these ESTs using microsatellite search module (MISA) of PrimerPro (http://webdocs.cs.ualberta.ca/~yifeng/primerpro/) with 2–6 bp repeats as described in Thiel et al. (2003). The parameters were set for detection of dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide motifs with a minimum of ten, six, five, four, and three repeats, respectively (Liu et al. 2012). Mononucleotide repeats such as A, T, G, and C were not considered in this work.

SSR Primers were designed in the conserved flanking regions with Primer Premier software (version 5.0). The principles for primer design were as follows: product length 150–500 bp, primer size 18–24 bp, primer melting temperature of 45–65 °C, the difference in annealing temperatures between the forward and the reverse primer did not exceed 1 °C, and none secondary structure. Oligonucleotides were synthesized at Invitrogen, Shanghai, China.

Practicability test of the designed EST-SSR primers

All designed primer pairs were initially tested for amplification using four DNA samples as templates from "Ailunwan" (*S. japonica* × *Saccharina latissima*), "Zaohoucheng" (*S. japonica*), "Dongfang No. 2" (*S. japonica* × *Saccharina logissima*), and "Rongfu" (*S. japonica* × *S. latissima*), respectively. These four *Saccharina* varieties were all widely cultured in China. Total DNA was isolated from sporophytes using improved CTAB methods (Guillemaut and Drouard 1992). PCR amplifications were carried out in 20 μL reaction mixture containing 1× buffer, 250 μM dNTPs, 200 μM of each primer, 30–50 ng template DNA, 1.5 mM MgCl$_2$, and 0.5 U Taq DNA polymerase (Zhang et al. 2015). The thermal program with gradient temperature control was as follows: 5 min at 94 °C, then 35 cycles of 50 s at 94 °C, a 45-s gradient from 45 to 65 °C and 45 s at 72 °C, finishing with 10 min at 72 °C.

Amplified products were separated by electrophoresis in 3% agarose gels and visualized by GoldView (5 μL GoldView was added into 100 mL cooled agarose gel solutions). The primer pairs which yielded clear bands were used to optimize the annealing temperature and to determine the PCR cycle number.

Genetic diversity analysis

A total of 128 individuals from ten strains including "Ailunwan," "Fujian," "Rongfu," "Pingbancai," "Sanhai," "Gaojia," "Zaohoucheng," "Dongfang No. 2," "Dongfang No. 3," and "Pengza No. 2" which covering almost the full widely used commercial *Saccharina* varieties in China were utilized to evaluate the genetic diversity and population structure using the developed SSR primer pairs from this work. We performed SSR PCR amplifications using the reaction

conditions described in this paper. The amplified products were separated through 6% denaturing polyacrylamide gel electrophoresis and then visualized using silver-staining method (Bassam et al. 1991). Molecular sizes of the amplifications were estimated using 20-bp DNA ladder (Takara).

SSR products which were amplified by each primer pair were scored manually as binary data using "1" (presence) and "0" (absence) based on the SSR pattern according to the previously reported method (Sun et al. 2006). The percentage of polymorphic loci (P) among strains, Nei's genetic diversity (H), Shannon's information index (I), effective number of alleles (Ne), genetic identity, and genetic distance were calculated from the EST-SSR data using POPGENE version 1.31 (Yeh et al. 1999). The cluster analysis was constructed on the similarity matrix by employing the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm using NTSYS-pc (version 2.1) software. Polymorphism information content (PIC) was calculated according to the following formula: $PIC = 1 - \sum_{i=1}^{n} P_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2P_i^2 P_j^2$ ("n" is the number of alleles at one locus, "P_i" and "P_j" are the frequencies of the ith and jth alleles at one locus) (Botstein et al. 1980).

To further investigate the genetic relationship of Chinese *Saccharina* strains, the possible population structure was analyzed with the STRUCTURE (version 2.2) program (Pritchard et al. 2000) based on the admixture model. To calculate an accurate number (K) of subpopulations inferred, six independent runs were performed at K levels, ranging from $K = 2$ to $K = 10$. Both the length of burn-in period and the number of iterations were set at 100,000. The model choice criterion to detect the most probable value of K was ΔK, an ad hoc quantity related to the second-order change in the log probability of data with respect to the number of clusters inferred by STRUCTURE (Evanno et al. 2005).

Results

Characterization of searched microsatellites

A set of 98,627 ESTs from the transcriptome of *S. japonica* with a total length of 44,362,190 bp was obtained in this work

Table 1 Searching results of EST-SSRs from the transcriptome of *S. japonica*

Parameter	Value
Total number of sequences examined	98,627
Total size of examined sequences (bp)	44,362,190
Total number of identified SSRs	9688
Number of SSR containing sequences	8039
Number of sequences containing more than 1 SSR	1237

Table 2 Distribution of EST-SSR repeat types found in *S. japonica*

Repeat type	Number	Frequency (%)	Number of motif type
Dinucleotide	285	2.94	3
Trinucleotide	5589	57.69	10
Tetranucleotide	953	9.84	33
Pentanucleotide	934	9.64	86
Hexanucleotide	1927	19.89	260
Total	9688	100.00	392

(Table 1). Here, 9688 SSRs in 8039 ESTs were isolated using the MISA program, which accounted for 9.82% (9688 sequences out of 98,627). It appears that there is one microsatellite sequence every 4.6 kb *Saccharina* transcriptome sequences. From those, most ESTs contained only one SSR while there were 1237 ESTs contained two or more SSRs. The identified microsatellites from this study contained five repeat types: dinucleotide (285, 2.94%); trinucleotide (5589, 57.69%); tetranucleotide (953, 9.84%); pentanucleotide (934, 9.64%); and hexanucleotide (1927, 19.89%). Trinucleotides which represented 57.69% of the isolated microsatellites were

Table 3 Primer pairs of polymorphic EST-SSR characterized by repeat-unit type

Repeat type	Motif	Number	Percent (%)
dinucleotide	AT/TA	28	19.18
	AG/GA	14	9.59
	CT/TC	13	8.91
	TG/GT	11	7.53
	AC/CA	11	7.53
	GC/CG	6	4.11
	Subtotal	*83*	*56.85*
trinucleotide	GAC/CAG/CGA/AGC	11	7.53
	ATG/GTA/AGT	9	6.17
	ATC/TAC/ACT	7	4.79
	CTG/GCT	6	4.11
	GGT/TGG	5	3.43
	others	7	4.79
	Subtotal	*45*	*30.82*
tetranucleotide	CGTA/CAGT/ACGT/GTAC	4	2.74
	TGAT/TATG	2	1.37
	others	7	4.80
	Subtotal	*13*	*8.91*
pentanucleotide	AGGAT	1	0.684
	CGAAC	1	0.684
	AAACG	1	0.684
	ACAAA	1	0.684
	TTGAT	1	0.684
	Subtotal	*5*	*3.42*
Total		146	100

Table 4 Characteristics of 52 representative microsatellite loci of *S. japonica*

Primer name	Product Size (bp)	Forward primer(5′–3′)	Reverse primer(5′–3′)	SSR motif	Ta (°C)	NA	*PIC*
HDS120	154	CGGTAGCGTCTTTTGTTTTC	ACCTGGTGTATGTGTTTGC	(TA)7	50.3	2	0.117
HDS125	157	ATAGTTGTCGGACATCTCAGG	CGGACGCTTCGTACTTGTT	(TA)6	50.3	2	0.204
HDS127	153	GCCTCTTGCCTTATTTGCTT	TGATGGTGGAATTGATGTGC	(ATC)7	50.3	2	0.374
HDS128	166	ATGTCGCTGGACGGATTG	CGGATATTTGCTGCCTCTG	(AT)5	50.3	4	0.619
HDS130	190	TATCCAGATCGCTTAGGTGC	GCCAATCAGAGCCAACA'	(TA)6	50.3	2	0.255
HDS187	178	TGGCTTTAGTTGGGTAGGC	CGTATCGTTTATTACAGGCTCA	(AG)5	52.5	2	0.141
HDS209	204	CCGCAGATCGAAAGGAAAC	CAACCAGCCATCACCCAT	(TAC)8	52.5	5	0.489
HDS210	167	ACTCACAAGGTGCGAGAATAAC	CCACCAACCATTGCGTTA'	(CAC)7	52	3	0.414
HDS212	170	GACAGTGGCGAAGAACAAAG	CAGAACACTCAAAACCCATACA	(TC)6	52.5	3	0.462
HDS237	178	GATCAAATCTGCCTACGATGC	TAACAGTTTGGGCGGGTG	(AC)5	53	3	0.121
HDS282	198	AGGCAGAGGCACGAGCAGAAC	CCCTAACGACCGAATGGAA	(ATG)6	53.5	2	0.371
HDS287	252	AGACCACTCGCTTTATTTGAC	TGCGTACCGACCACAGACA	(GT)6	53.5	3	0.123
HDS308	286	TTCGTGTCATTGTGGCTGTG	GCTGTAATAACGCCCTTCG	(GGT)5	54.5	4	0.398
HDS312	150	TGGAACCACAACCAACCG	CCTTCCCTTTCTGGCTACG	(ACT)6	54.5	4	0.444
HDS318	210	GTAGGATAGCCCCATTTGCC	TGCTGTTGCCACGTTCATT	(GCT)6	54.5	2	0.332
HDS321	183	GCAGTATCTCCCTGTTGTCG	GTGCTTTGAGTGCGTTGG	(CAA)5	54.5	2	0.339
HDS326	176	CACATTATGCGGAGGGAG	GCAGTCGGTGTCGGTTTT	(CGTA)5	54.5	2	0.233
HDS329	224	CCGATTGTCTGATTGAGGACT	TAGAGGATGTGCCCAGAGTG	(GCA)5	54.5	2	0.274
HDS343	189	AGAAAGCAGGAAGGCGTGTC	TTTGTCGGGGTTGGGTCA	(AGT)6	55	3	0.382
HDS386	182	CAGCAACAGCAGCCTCCTC	GCTTCGTCTGTGCCCTCT	(CAG)6	56	2	0.335
HDS404	281	ATTACAACTAGAATGGCAC	ATAGACAGGCGAGGAACT	(AGT)5	51	2	0.228
HDS416	317	GAACGGGGAGAAGAAGAATAAT	TGTAAAGAAAGAGGGGTGGAG	(ACT)6	52	3	0.125
HDS437	455	ATTTTCTCCCGTAATGAGCC	GCAGCGGAAGGAAGTGTT	(ATG)6	53.5	3	0.388
HDS446	225	TCCGAGAAGGAACAGTAGGAC	AGCGATAACCACCAACAGC	(AT)5	53.5	4	0.385
HDS450	231	GAAGGAAGGGGTGGAAGAC	ATCATGCAGGTTATGCGACT	(AAG)5	53.5	2	0.282
HDS475	266	CATCGTGTTGTGCTTTCGC	TCCATTCGGTCGTTAGGG	(ATC)6	53.5	3	0.303
HDS479	368	GGCGAGGCGTAATAATCAG	ATGGCAACATCACGAGGG	(TTGAT)5	53.5	2	0.117
HDS482	248	TGCGAGTCAGGCAAACG	GCTGTCCGAACAATACCACC	(ACGT)5	53.5	2	0.344
HDS484	181	TGTCATCAGTGGCATCAGTTT	GGTCACAGGGATAGTTGGAAG	(AG)5	53.5	2	0.366
HDS488	195	CCTCCCAGCGTTACAGAT	GTCCGACAACCAACAAGC	(GCA)5	52.9	3	0.478
HDS492	242	GCTTCAATGGCACTGTTTACTG	CGATTTTGTCGGAGGGGAT	(TGG)5	53	2	0.361
HDS593	157	TCACGGAAGGCAAGAAACT	CTCGCCTTTTAATTGTGGTT	(ACT)6	50	3	0.064
HDS596	227	AGTGGGAAGAAAGGATGGTC	GCTTACTCTGCCGCAATA	(TA)6	50.2	4	0.416
HDS597	195	GATACGGAATAAGCAGAGTG	ACGACGCAAGTCCAAAGT	(TGT)6	50.3	3	0.570
HDS600	213	TATCGGTCACAAACACCAACG	AAGCAGGGAGACGGGAGAAG	(ACT)6	50.5	2	0.355
HDS602	180	ATGACGAGGGCTCCTTTTG	GCCACTTGAGGGGCTTTA	(TC)5	50.7	4	0.562
HDS657	208	CAAGAAGACCATCCCACTC	GTTTTGCACAGCATACCATT	(GA)5	52.6	3	0.175
HDS676	447	CCGAATCGTACCGTTGCT	CGTGTTTGTTCTCCCACTTTT	(CG)5	53.1	3	0.247
HDS682	283	TAAGAGGGGCTTTCACGC	CAAACCAACCCAACCACTAC	(TA)5	53.2	2	0.339
HDS0714	193	AAATACAGCCTGCTCACG	CGCTTAACTGCACCTCTTG	(CGAAC)9	53.6	3	0.278
HDS842	162	CAAGTTTTGATGTTTGGGGAGA	AGGTGCCCGCAGTTTTCT	(TC)5	52.2	2	0.305
HDS911	175	GGCAGAAGACAGCTCACGA	CTCAATGTTCAAATGACTCCGT	(AT)5	52	2	0.371
HDS922	152	CCATCAGCGTCGTCAAAT	TGGATCGTCGTGGTGTCT	(CA)9	52.5	3	0.260
HDS925	184	CACTCTTGGCAATGTTAGCG	TGGCCTCTGCCGTGTACTGT	(TA)5	52.6	2	0.204
HDS927	190	GTTTCCCCTTAGGCTAATGAT	AGCGTGAGGCACTTGGT	(AGT)9	52.8	4	0.639
HDS941	244	CTGCTTCCAACGAGGCTAA	TGAAGGGCGAAAGGAGTT	(GT)6	53.5	2	0.124
HDS957	173	AGAACGGGGACATGACAGC	CGTTGGTTTGATAGTGTTGGA	(GCA)6	54.1	2	0.305
HDS958	158	CGCACAACATAGCACAGCA	CGATGAGTGAATCCACGAAC	(CAG)5	54.1	2	0.336
HDS964	200	CTTGACTGGACCAATGCCC	CCGACAACCACTGCGATAAT	(AC)7	54.4	4	0.657
HDS980	202	CTGGCGAATGCCTGTCTT	TCACTGTTGCTGGTGTTGG	(CAG)7	55.4	2	0.305
HDS984	170	TGCGTTTGGTTTTCCGTG	CGTCCGTAACAACAGTGCC	(TG)5	55.5	2	0.332
HDGS073	235	TACAGCAGTACATTTGGAGCC	ACTTTGCCTTTTGTGGGTG	(TGC)6	55	3	0.381

Ta the optimal annealing temperature, *NA* number of alleles

the predominant types. The distribution of EST-SSR repeat types is shown in Table 2.

We used the three-class categorization according to the report of Weber (1990) in order to conduct a more detailed SSR analysis. Five hundred seventy-five microsatellites (5.93%) were regarded as compound type, 5921 (61.12%) as perfect, and 3192 (32.95%) as imperfect.

Evaluation for EST-SSR markers

Some microsatellites were located too close to the end of the flanking region or their base composition of the flanking sequence was unsuitable to accommodate primer design (Wang et al. 2011). Therefore, not all the searched microsatellites were suitable for primer design. Finally, a total of 1120 PCR

Table 5 Genetic diversity statistics of ten *Saccharina* populations based on 52 EST-SSR markers

Population	P (%)	NA	Ne	H	I	He	Ho
Ailunwan	73.08	1.8654	1.6726	0.3312	0.4965	0.3974	0.4423
Fujian	57.69	1.6346	1.512	0.2618	0.3826	0.3141	0.4423
Pingbancai	48.08	1.500	1.39	0.2083	0.3027	0.2500	0.3269
Sanhai	69.23	1.7692	1.5814	0.2957	0.4407	0.3647	0.4006
Gaojia	65.38	1.7500	1.549	0.2853	0.4271	0.3436	0.3526
Zaohoucheng	28.85	1.3269	1.2267	0.1207	0.1819	0.1487	0.1795
Dongfang No. 3	73.08	1.9808	1.7331	0.3376	0.5267	0.4051	0.4615
Pengza No. 2	67.31	1.7500	1.5294	0.2767	0.4170	0.3321	0.4295
Rongfu	63.46	1.6346	1.4663	0.2601	0.3784	0.3205	0.3654
Dongfang No. 2	73.08	1.7885	1.5447	0.2925	0.4390	0.3532	0.4071
Total	100	2.6731	1.7690	0.3882	0.6360	0.3949	0.3799

P percentage of polymorphic loci, *NA* observed number of alleles, *Ne* effective number of alleles, *H* Nei's genetic diversity, *I* Shannon's information index, *He* expected heterozygote, *Ho* observed heterozygote

primer pairs were designed and synthesized according to the criteria of primer design in this work.

Among the 1120 PCR primer pairs, 631 accounting for 56.34% have the good amplification in four *Saccharina* templates at the target region. Other primers were excluded which showed no amplification or multiple bands. Subsequently, 146 of the screened 1120 (13.04%) SSR primer pairs revealed polymorphism in four templates. Of the 146 primer pairs, 83 (56.85%) pairs were dinucleotide repeats, 45 (30.82%) were trinucleotide repeats, 13 (8.91%) were tetranucleotide repeats, and 5 (3.42%) were pentanucleotide repeats (Table 3), which indicated primers of dinucleotide repeat type had the highest development efficiency. Moreover, the most abundant repeat motif was (AT/TA)n (19.18%), followed by (AG/GA)n (9.59%), and (CT/TC)n (8.91%).

Genetic diversity analysis

A total of 128 individuals belonged to ten *Saccharina* strains were analyzed with 52 representative polymorphic EST-SSR markers which obtained in this work. One hundred thirty-nine alleles were amplified in total and the average number of alleles per locus was 2.67 (ranging from 2 to 5). The *PIC* index was between 0.064 (HDS593) and 0.657 (HDS964) with an average of 0.327. The detailed information of the 52 EST-SSR markers was present in Table 4.

And then the genetic diversity was assessed using these 52 EST-SSR primer pairs. The percentage of polymorphic loci (*P*) per strain ranged from 28.85% ("Zaohoucheng") to 73.08% ("Ailunwan," "Dongfang No. 2," and "Dongfang No. 3"). Nei's genetic diversity (*H*) ranged from 0.1207 ("Zaohoucheng") to 0.3376 ("Dongfang No. 3") and Shannon's information index (*I*) from 0.1819 ("Zaohoucheng")

Table 6 Nei's genetic identity (above diagonal) and genetic distance (below diagonal) among ten *Saccharina* populations

Population	1	2	3	4	5	6	7	8	9	10
1	****	0.8317	0.8384	0.8577	0.9156	0.8010	0.6172	0.9109	0.8528	0.7568
2	0.1844	****	0.8102	0.7923	0.8317	0.8354	0.6409	0.8260	0.7645	0.7419
3	0.1763	0.2006	****	0.8635	0.8813	0.8052	0.5555	0.8740	0.8878	0.7429
4	0.1535	0.2188	0.1468	****	0.8628	0.7878	0.6261	0.8390	0.9035	0.7116
5	0.0882	0.1739	0.1264	0.1476	****	0.8263	0.6475	0.9214	0.8569	0.7327
6	0.2051	0.1828	0.2101	0.2386	0.1784	****	0.7760	0.8302	0.7989	0.7686
7	0.4826	0.4449	0.5879	0.4683	0.4347	0.3294	****	0.6267	0.6177	0.7054
8	0.0933	0.1810	0.1347	0.1756	0.081	0.1727	0.4673	****	0.8558	0.7750
9	0.1592	0.2490	0.1190	0.1014	0.1544	0.2226	0.4817	0.1557	****	0.7594
10	0.2538	0.2721	0.2709	0.3088	0.2745	0.2472	0.4054	0.2288	0.2509	****

Numbers 1–10 indicate the ten *Saccharina* populations: (1) "Ailunwan," (2) "Fujian," (3) "Pingbancai," (4) "Sanhai," (5) "Gaojia," (6) "Zaohoucheng," (7) "Dongfang No. 3," (8) "Pengza No. 2," (9) "Rongfu," and (10) "Dongfang No. 2"

Fig. 1 The UPGMA dendrogram of ten *Saccharina* populations based on 52 representative polymorphic SSR markers developed in this study
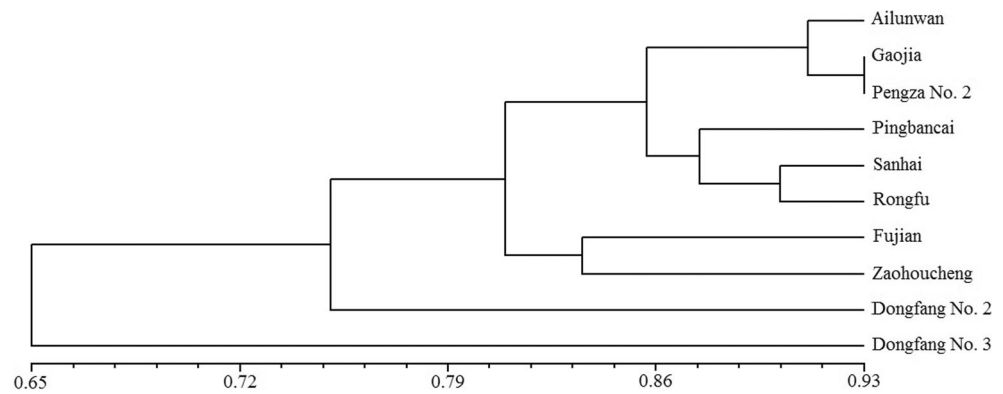
to 0.5267 ("Dongfang No. 3") (Table 5). The results showed that genetic diversity of "Zaohoucheng" was the lowest while "Dongfang No. 3" was the highest. As shown by genetic identity and genetic distance from EST-SSR markers (Table 6), the genetic identity coefficients ranged from 0.5555 to 0.9214. Among these ten strains, "Dongfang No. 3" and "Pingbancai" had the lowest similarity while "Gaojia" and "Pengza No. 2" showed the highest similarity value.

All the amplified alleles were used for cluster analysis of the ten *Saccharina* strains via the UPGMA method. The results indicated that different samples from the same strain were clustered together and ten *Saccharina* strains were divided into two main groups (Fig. 1). The first group contained nine *Saccharina* strains which were divided into two subgroups, while the second group contained only "Dongfang No. 3." For the clade of nine strains, "Dongfang No. 2" was the basal strain, "Fujian" and "Zaohoucheng" subsequently derived. The remaining six strains fell into two clusters: "Gaojia," "Pengza No. 2," and "Ailunwan" clustered as one clade while "Sanhai," "Rongfu," and "Pingbancai" clustered in another clade.

Population structures were determined using the STRUCTURE program in order to further investigate the relationship of Chinese *Saccharina* strains in this work. The
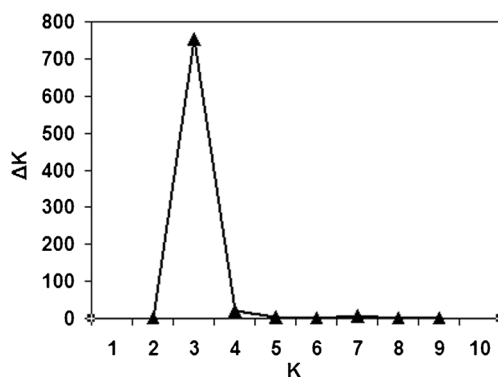
estimated likelihood values for a given *K* (from 2 to 10) in six independent runs were performed and the maximum log likelihood was attained at *K* = 3 (Fig. 2). At this level, all samples from ten strains could be classified into three groups with greatest probability. Clustering bar plots with *K* = 3 were shown in Fig. 3. The strains such as "Zaohoucheng" and "Fujian" originated from *S. japonica* had a similar trend and clustered into one group (red), "Dongfang No. 2" and "Dongfang No. 3" originated from hybrids of *S. japonica* and *S. logissima* had a trend of clustering into one group (green) and then the strains such as "Ailunwan," "Pingbancai," "Sanhai," "Rongfu," "Pengza No. 2," and "Gaojia" had a trend of clustering into another group (blue).

Discussion

Due to the intrinsic advantages of EST-SSRs, exploring SSR markers from ESTs has attracted our more and more attentions. The major disadvantage of gSSR or EST-SSR markers is their strong species-specificity, and the markers need to be developed de novo for different species (Squirrell et al. 2003). Therefore, it requires the information of genome or transcriptome sequences in advance for primer designing. Traditionally, ESTs were obtained by constructing cDNA library and sequencing a number of clones. With the advent of high-throughput next generation sequencing methods such as Roche/454, ABI/, and Illumina/Hiseq2000, large-scale transcriptome sequences can be generated economically which enable researches to develop a batch of EST-SSR markers more easily. EST-SSRs have been identified using NGS technology in many species such as *Cyprinus carpio* (Ji et al. 2012), Alfalfa (Liu et al. 2013), *Carassius auratus* (Zheng et al. 2014), *Pelteobagrus fulvidraco* (Zhang et al. 2014a), *Neolitsea sericea* (Chen et al. 2015), and *Verasper variegatus* (Ge et al. 2017) . At the same time, gSSR markers for *Saccharina* using paired-end Illumina sequencing data have been reported (Zhang et al. 2014b). Here, we identified 9688 EST-SSRs and explored 146 novel EST-SSR markers based on 98,627 *S. japonica* ESTs produced by NGS technology.

Fig. 2 Assessing the population structure by ΔLn P (D). Log probability data [LnP(D)] as function of *K* (number of groups) from the STRUCTURE run. STRUCTURE simulation demonstrated that *K* value showed a modest peak at *K* = 3, suggesting that three groups could contain all individuals with greatest probability

Fig. 3 Estimated population structure of 100 individuals for ten *Saccharina* populations using the program STRUCTURE (*K* = 3). Each group is represented by a different color as listed: red: "Zaohoucheng," "Fujian,"; green: "Dongfang No. 2," "Dongfang No. 3,"; blue: "Ailunwan," "Pingbancai," "Sanhai," "Rongfu," "Pengza No. 2," and "Gaojia"

Previously, nine EST-SSR markers were generated from *S. japonica* EST sequences (Liu et al. 2010b) and 23 were developed from the *L. digitata* ESTs and *S. japonica* ESTs (Wang et al. 2011). Compared to these *Saccharina* EST-SSR reports, the obvious advantage of NGS methods used in this work is their ability to produce large numbers of ESTs from which to isolate and develop numerous gene-associated microsatellite markers at lower cost and effort (Zalapa et al. 2012).

Usually, most SSRs are present in genomic DNA while only a little is present in transcriptome (Sun et al. 2006). Some studies showed that 7–10% of EST sequences contains SSRs in land plants (Chen et al. 2005). A total of 9688 EST-SSRs accounting for 9.82% were identified in this work using the MISA program. This frequency was in coincident with plants but higher than those seaweeds which have informed EST-SSR markers such as *Ulva prolifera* (2.91%) (Zhang et al. 2014c) and *Porphyra haitanensis* (6.02%) (Xie et al. 2009). In earlier reports of *S. japonica*, SSRs exist in 3–6% of EST sequences (Wang et al. 2011; Liu et al. 2012). In general, the NGS method could provide comprehensive information about the genomic organization of repeat sequences in species (Zheng et al. 2014). However, other explanations should also be considered such as SSR search criteria and the size of dataset which usually leads to the varying frequency of EST-SSRs (Varshney et al. 2005; Liang et al. 2009). Although the criteria for screening for EST-SSRs vary in different species, trinucleotide repeats are the most common types, not only in many land plants but also in some seaweeds (Varshney et al. 2002; Kantety et al. 2002; Sun et al. 2006; Xie et al. 2009; Wang et al. 2011). Our results were also in accordance with this universal pattern. In addition, if four nucleotides (A, T, G, C) are present in random combinations, all the SSR motifs could be represented by 4 different dinucleotide motifs (AC, AG, AT, CG), 10 different trinucleotide motifs, 33 different tetranucleotide motifs, 102 different pentanucleotide motifs, and 350 different hexanucleotide motifs (Rota et al. 2005). In this work, all types of trinucleotide, tetranucleotide and most types of dinucleotide, pentanucleotide, and hexanucleotide occurred in *S. japonica* EST-SSRs (Table 2). The results proved that the distribution of different types of EST-SSRs in *S. japonica* was obviously extensive. Of course, if this distribution is the actual distribution of *S. japonica*, EST-SSRs still requires more evidence and further research.

The key step in an effective breeding or conservation program is to accurately evaluate the available genetic resources (Xie et al. 2009), and SSR analysis is a well-established tool for evaluating genetic diversity. The value of *P*, *H*, *Ne*, *He*, and *I* which are all parameters of genetic diversity varies with abundance (Nei 1972). On the basis of the present study, "Zaohoucheng" which in an inbred line of *S. japonica* (Tian and Yuan 1989) exhibited the lowest genetic diversity (*P* = 28.85%, *H* = 0.1207) while "Dongfang No. 3" (*P* = 73.08%, *H* = 0.3376) showed higher genetic diversity. "Dongfang No. 3" is the direct hybrid of the first filial generation from the female gametophytes of *S. japonica* and the male of *S. longissima* (Li et al. 2008). EST-SSR survey in our work indicated that continuous selfing and/or inbreeding had reduced the genetic diversity and might lead to inbreeding depression. Although high-genetic diversity could enhance the adaptability to environments, self-crossing and selection would stabilize the traits of the varieties (Zhao et al. 2013). Therefore, it needs to be considered comprehensively, and evaluating genetic diversity is indispensable in *Saccharina* breeding.

In our work, genetic identity and cluster analysis of ten *Saccharina* strains which are widely cultured in China were analyzed using all the amplified EST-SSR alleles. It is well known that the genetic identity and clustering order can reflect the relationships among strains. At the same time, the previous paper has reported that lower genetic parental identity resulted in more obvious heterosis to a certain extent (Melchinger et al. 1992). In other words, stronger heterosis could be achieved in breeding if parental combinations with lower genetic identity are used. Therefore, our research can help in the selection of parents for *Saccharina* breeding purposes. In addition, population structure analysis displayed three categories of ten strains: one was constituted of strains bred from *S. japonica*, one was made up of hybrids derived from *S. japonica* and *S. longissima*, while another was constituted of strains most from hybrids between *S. japonica* and *S. latissima* such as "Ailunwan," "Rongfu," and "Sanhai" (Zhang et al. 2011, 2016). Generally

speaking, the accessions with the same background usually have similar Q values which could be within a cluster (Hou et al. 2011). We infer that *Saccharina* strains cultured in China at present are divided into three groups based on genetic background. This is consistent with our traditional understanding of the parental origin of these strains (Tian and Yuan 1989; Li et al. 2007a, 2008; Zhang et al. 2011, 2016). Therefore, the genetic sources of *Saccharina* strains were revealed using EST-SSR markers developed from our work.

All the above research supported that the EST-SSR analytical system using the polymorphic markers obtained in this study has been successfully adapted for the genetic analysis of Chinese *Saccharina* strains. These developed EST-SSR markers will facilitate most fields of genetic analysis and promote the program of molecular marker-assisted breeding in *Saccharina*.

References

An XL, Li XM, Xu CX, Li ZX (2010) Ecological effects of seaweeds on the coastal environment. Fish Sci 29:115–119

Bassam BJ, Caetano-Anolles G, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80–83

Bi YH, Hu YJ, Zhou ZG (2011) Genetic variation of *Laminaria japonica* (Phaeophyta) populations in China as revealed by RAPD markers. Acta Oceanol Sinica 30:103–112

Billot C, Boury S, Benet H, Kloareg B (1999a) Development of RAPD markers for parentage analysis in *Laminaria digitata*. Bot Mar 42: 307–314

Billot C, Rousvoal S, Estoup A, Epplen JT, Saumitou-Laprade P, Valero M, Kloareg B (1999b) Isolation and characterization of microsatellite markers in nuclear genome of the brown alga *Laminaria digitata* (Phaeophyceae). Mol Ecol 7:1778–1780

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. Mol Ecol 16:907–924

Chen JF, Ren ZL, Gao LF, Jia JZ (2005) Developing new SSR markers from EST of wheat. Acta Agron Sinica 31:154–158 (In Chinese with English abstract)

Chen LY, Cao YN, Yuan N, Nakamura K, Wang GM, Qiu YX (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 Breed 35(9):187. https://doi.org/10.1007/s11032-015-0379-1

Edwards KJ, Barker JH, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plant. BioTechniques 20:758–760

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Ge J, Chen S, Liu C, Bian L, Sun H, Tan J (2017) Characterization of the global transcriptome and microsatellite marker information for spotted halibut *Verasper variegates*. Genes Genom 39:307–316

Guillemaut P, Drouard LM (1992) Isolation of plant DNA: a fast, inexpensive, and reliable method. Plant Mol Biol Rep 10:60–65

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci 45:45–54

Hou JF, Wang CL, Hong XJ, Zhao JM, Xue CC, Guo N, Gai J, Xing H (2011) Association analysis of vegetable soybean quality traits with SSR markers. Plant Breed 130:444–449

Jensen A (1993) Present and future needs for algae and algal products. Hydrobiologia 269:15–23

Ji P, Zhang Y, Li C, Zhao Z, Wang J, Li J, Xu P, Sun X (2012) High throughput mining and characterization of microsatellites from common carp genome. Int J Mol Sci 13:9798–9807

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Li XJ, Cong YZ, Yang GP, Shi YY, Qu SC, Li ZL, Wang G, Zhang Z, Luo S, Dai H, Xie J, Jiang G, Liu J, Wang T (2007a) Trait evaluation and trial cultivation of Dongfang no. 2, the hybrid of a male gametophyte clone of *Laminaria longissima* (Laminariales, Phaeophyta) and a female one of *L. japonica*. J Appl Phycol 19:139–151

Li YH, Yang YX, Liu JD, Wang XL, Gao TX, Duan D (2007b) Genetic mapping of *Laminaria japonica* and *L. longissima* using amplified fragment length polymorphism markers in a "two-way pseudotestcross" strategy. J Integr Plant Biol 49:392–400

Li XJ, Liu JL, Cong YZ, Qu SC, Zhang ZZ, Dai HL et al (2008) Breeding and trial cultivation of Dongfang no.3, a hybrid of *Laminaria* gametophyte clones with a more than intraspecific less than interspecific relationship. Aquaculture 280:76–80

Li DJ, Deng Z, Qin B, Liu XH, Men ZH (2012) De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.) BMC Genomics 13(1):192. https://doi.org/10.1186/1471-2164-13-192

Liang XQ, Chen XP, Hong YB, Liu HY, Zhou GY, Li SX et al (2009) Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and Arachis wild species. BMC Plant Biol 9:1–9

Liang XY, Wang XM, Chi S, Wu SX, Sun J, Liu T et al (2014) Analysis of *Saccharina japonica* transcriptome using the high-throughput DNA sequencing technique and its vanadium-dependent haloperoxidase gene. Acta Oceanol Sinica 33:27–36

Liu FL, Shao ZR, Zhang HN, Liu JD, Wang XL, Duan DL (2010a) QTL mapping for frond length and width in *Laminaria japonica* Aresch (Laminarales, Phaeophyta) using AFLP and SSR markers. Mar Biotechnol 12:386–394

Liu FL, Wang XL, Yao JT, Fu WD, Duan DL (2010b) Development of expressed sequence tag-derived microsatellite markers for *Saccharina* (*Laminaria*) *japonica*. J Appl Phycol 22:109–111

Liu FL, Wang FJ, Duan DL (2012) EST-SSR markers derived from *Laminaria digitata* and its transferable application in *Saccharina japonica*. J Appl Phycol 24:501–505

Liu Z, Chen T, Ma L, Zhao Z, Zhao PX, Nan Z, Wang Y (2013) Global transcriptome sequencing using the Illumina platform and the development of EST-SSR markers in autotetraploid alfalfa. PLoS One 8: e83549

Melchinger AE, Boppenmaie J, Dhillon BS, Pollmer WG, Herrmann RG (1992) Genetic diversity for RFLPs in European maize inbreds. II Relation to performance of hybrids within versus between heterotic groups for forage traits. Theor Appl Genet 84:672–681

Nei M (1972) Genetic distance between populations. Anim Nat 106:283–292

Porse H, Rudolph B (2017) The seaweed hydrocolloid industry: 2016 updates, requirements, and outlook. J Appl Phycol 29:2187–2200

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Rota LM, Kantety RV, Yu JK, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6:23–34

Rungis D, Bérubé Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K (2004) Robust simple sequence repeat markers for spruce (Picea spp.) from expressed sequence tags. Theor Appl Genet 109:1283–1294

Shan TF, Pang SJ, Zhang YR, Yakovleva IM, Skriptsova AV (2011) An AFLP-based survey of genetic diversity and relationships of major farmed cultivars and geographically isolated wild populations of Saccharina japonica (Phaeophyta) along the northwest coasts of the Pacific. J Appl Phycol 23:35–45

Shi YY, Yang GP, Liu YJ, Liao MJ, Li XJ, Cong YZ (2007) Development of 18 polymorphic microsatellite DNA markers of Laminaria japonica (Phaeophyceae). Mol Ecol Notes 7:620–622

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:1339–1348

Sun JW, Liu T, Guo BT, Jin DM, Weng ML, Feng YB, Xu P, Duan D, Wang B (2006) Development of SSR primers from EST sequences and their application in germplasm identification of Porphyra lines (Rhodophyta). Eur J Phycol 41:329–336

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

Tian Z, Yuan B (1989) Report of the breeding of a novel kelp variety, "Zaohoucheng No.1". Mariculture 1:7–17 (in Chinese)

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis of frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotech 23:48–55

Waaland JR, Stiller JW, Cheney DP (2004) Macroalgal candidates for genomics. J Phycol 40:26–33

Wang XL, Yang YX, Cong YZ, Duan DL (2004) DNA fingerprinting of selected Laminaria (Phaeophyta) gametophytes by RAPD markers. Aquaculture 238:143–153

Wang XL, Liu CL, Li XJ, Cong YZ, Duan DL (2005) Assessment of genetic diversities of selected Laminaria (Laminarales, Phaeophyta) gametophytes by inter-simple sequence repeat analysis. J Integr Plant Biol 47:753–758

Wang GL, Tan XL, Shen JL, Li J, Zhang L, Sun JW et al (2011) Development of EST-SSR primers and their practicability test for Laminaria. Acta Oceanol Sinica 30:112–117

Weber JL (1990) Informativeness of human (dC-dA)n. (dG.dT)n polymorphisms. Genomics 7:524–530

Xie CT, Cheng CS, Ji DH, Xu Y (2009) Characterization, development and exploitation of EST-derived microsatellites in Porphyra haitanensis Chang et Zheng (Bangiales, Rhodophyta). J Appl Phycol 21:367–374

Yang GP, Sun Y, Shin YY, Zhang LA, Guo SS, Li BJ, Li X, Cong Y, Zhao Y, Wang W (2009) Construction and characterization of a tentative amplified fragment length polymorphism-simple sequence repeat linkage map of Laminaria (Laminariales, Phaeophyta). J Phycol 45:873–878

Yeh FC, Yang RC, Boyle T (1999) POPGENE, version 1.31. Microsoft window-based freeware for population genetic analysis. Quick userguide. FrancisYeh, University of Alberta, Canada. www.ualberta.ca/-fyeh

Yotsukura N, Kawai T, Motomura T, Ichimura T (2001) Random amplified polymorphic DNA markers for three Japanese Laminaria species. Fish Sci 67:857–862

Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot 99:193–208

Zhang J, Liu Y, Yu D, Song HZ, Cui JJ, Liu T (2011) Study on high-temperature-resistant and high-yield Laminaria variety "Rongfu". J Appl Phycol 23:165–171

Zhang J, Ma W, Song XM, Lin QH, Gui JF, Mei J (2014a) Characterization and development of EST-SSR markers derived from transcriptome of yellow catfish. Molecules 19:16402–16415

Zhang L, Peng J, Li X, Liu Y, Cui C, Wu H, Wu R, Tian P, Li Y (2014b) Development of 27 trinucleotide microsatellite markers for Saccharina japonica using next generation sequencing technology. Conserv Genet Genet Resour 6:341–344

Zhang L, Wang GL, Liu C, Chi S, Liu T (2014c) Development and utility of EST-SSR markers in Ulva prolifera of the South Yellow Sea. Acta Oceanol Sinica 33:105–113

Zhang J, Li W, Qu JQ, Wang XM, Liu C, Liu T (2015) Development and characterization of microsatellite markers from an enriched genomic library of Saccharina japonica. J Appl Phycol 27:479–487

Zhang J, Liu T, Bian DP, Zhang L, Li XB, Liu D et al (2016) Breeding and genetic stability evaluation of the new Saccharina variety "Ailunwan" with high yield. J Appl Phycol 28:3413–3421

Zhao C, Liu C, Li W, Chi S, Feng RF, Liu T (2013) Application of restriction site amplified polymorphism (RSAP) to genetic diversity in Saccharina japonica. Chin J Oceanol Limnol 31:830–834

Zheng XH, Kuang YY, Lü WH, Cao DC, Sun XW (2014) Transcriptome-derived EST–SSR markers and their correlations with growth traits in crucian carp Carassius auratus. Fish Sci 80:977–984