J Appl Phycol (2016) 28:2479–2484
DOI 10.1007/s10811-015-0747-6

Development of genic SSR markers from an assembled *Saccharina japonica* genome

Jie Peng[1] · Linan Zhang[2] · Xiaojie Li[1] · Cuiju Cui[1] · Ruina Wu[1] · Pingping Tian[1] · Yan Li[1] · Yanling Liu[1]

Received: 29 June 2015 / Revised and accepted: 28 October 2015 / Published online: 9 November 2015

Abstract *Saccharina japonica*, a marine brown alga, is an economically important species that has been cultivated in China for approximately a century. We identified 11,973 simple sequence repeats (SSRs) in 58.3 Mb of predicted coding sequence from an assembled *S. japonica* genome using Illumina paired-end sequencing data. Trinucleotide SSRs were the most abundant motif. Twenty-six loci amplified by the 57 unique primer pairs we designed from trinucleotide genic SSRs (repeat number >10) were polymorphic, among the 55 *S. japonica* individuals tested. The number of alleles per locus ranged from 2 to 7 (average 3.46). The observed and expected heterozygosity per locus ranged from 0.128 to 0.652 and from 0.130 to 0.676, respectively. Two loci deviated from the Hardy–Weinberg equilibrium and two others were in linkage disequilibrium. We demonstrate that genic SSRs can be efficiently identified from assembled and annotated genomes using Illumina sequencing data. The novel polymorphism markers that we have identified, with informative flanking sequences and unique positional relationships to the genome, should facilitate genetic diversity analysis and further genetic studies in this species.

Keywords Genic SSRs · Illumina sequencing · Microsatellite makers · *Saccharina japonica*

✉ Linan Zhang
zhanglinan1206@163.com

[1] National Engineering and Technique Research and Development Center of Algae and Sea Cucumber of China, Provincial Key Laboratory of Genetic Improvement and Efficient Culture of Marine Algae of Shandong, Shandong Oriental Ocean Sci-tech Co., Ltd, Yantai, Shandong 264003, People's Republic of China

[2] Marine Science and Engineering College, Qingdao Agriculture University, Qingdao 266109, People's Republic of China

Introduction

Saccharina japonica, which is indigenous to Japan but was accidentally introduced to China in the early twentieth Century, has been extensively cultivated in China due to its high biomass productivity and its economic value in the food, fertilizer, chemical, drug, and phycocolloid industries (Tseng 2001). Various classical breeding methods for terrestrial crops, including continuous selection, hybridization, and the development of interspecies hybrids, have been used as a part of its cultivation since the 1950s for the genetic improvement and development of a set of elite varieties and hybrids. Representatives of these varieties include Haiqin no.1, 90-1, and interspecies hybrids Dongfang no. 2 and Dongfang no. 6, which have all contributed significantly to its aquaculture performance, particularly its yield (Fang et al. 1962; Li et al. 2007, 2015; Zhang et al. 2007). Despite these achievements, the genetic study of *S. japonica* is far behind that of most model plants.

Microsatellites, also known as simple sequence repeats (SSRs), are hypervariable, co-dominant, and abundant in all organisms; therefore, they are frequently the target of genetic research. SSRs are divided into genomic SSRs (random genomic sequences) versus genic SSRs (coding sequences/expressed sequence tags (ESTs)) according to location in the genome. SSRs in coding sequences are more useful for genetic study compared with those in non-coding sequences, because they are potentially linked with particular regions that may contribute to agronomic phenotypes and have high transferability. Various genomic SSR markers have been developed for *S. japonica* and applied to genetic studies in previous studies, including diversity assessment (Shi et al. 2007; Liu et al. 2012; Zhang et al. 2015), linkage map construction, and quantitative trait loci mapping (Yang et al. 2009; Liu et al. 2010a). However, no genic SSR markers are available in this species

Table 1 Characteristics of 26 polymorphic trinucleotide genic SSRs loci for *Saccharina japonica*

Locus	Accession no.	Primer sequence (5′-3′)	Repeat motif	*Ta* (°C)	*Na*	Size range (bp)	*Ho*	*He*
ZpsjE-1	KR869825	F: TTGCCCTACGAGTATTGCGC R: CCAAACGTGACGGAAAGCTG	$(GTG)_{17}$	59	3	229–244	0.287	0.303
ZpsjE-4	KR869826	F: CTCTCTCCTCTCAAACCGC R: CGGGCATACTCTGTGACA	$(GGA)_{15}$	56	5	272–293	0.430	0.435
ZpsjE-5	KR869827	F: AGCGAGTTCAGCGACAAA R: TCAACAGAGCTCTTTCCCAG	$(GGC)_{15}$	57	2	254–266	0.128	0.130
ZpsjE-10	KR869828	F: CTCCACCGCTTCTTCTTGTG R: CAACGGGGAAGAGTGTCAAC	$(CCA)_{13}$	55	6	251–275	0.371	0.350
ZpsjE-12	KR869829	F: CCATCCTCGCCAACACATC R: CTTCCCACCCTTCTGCTTTG	$(CTC)_{14}$	57	4	287–302	0.652	0.676
ZpsjE-13	KR869830	F: GTAGTATAGTGGGCCTGTTC R: AAAGCTGGGACACAAAGG	$(GGA)_{14}$	53	2	219–225	0.211	0.229
ZpsjE-15	KR869831	F: CTTCTTCGTTCCGCTTCTC R: GGTCTGGCCGTATTTACTTG	$(TCC)_{14}$	57	2	290–302	0.171	0.159
ZpsjE-16	KR869832	F: GCCGTAGCGTTGTTTCTC R: GTTCCCCTCTTAGTCGCTG	$(TCC)_{14}$	57	3	258–273	0.439	0.448
ZpsjE-17	KR869833	F: GACCTCCCGGCTTTACTACA R: CCCACCTCCCACTTGTCTTT	$(TGG)_{14}$	53	3	357–366	0.271	0.276
ZpsjE-20	KR869834	F: CAGTCCAGATTTGTACCTTC R: CGTATACCAATCCAAGGAGT	$(CAC)_{13}$	57	3	286–295	0.171*	0.303
ZpsjE-21	KR869835	F: CACCCCAAACGAGTCTATAG R: ACTTGAACGAACTGGCTC	$(CCA)_{13}$	57	2	287–299	0.425	0.437
ZpsjE-23	KR869836	F: TGCTACAACACGGTTCTG R: TGTACAGCAGAGCGATTC	$(CGC)_{13}$	55	6	218–245	0.441	0.610
ZpsjE-25	KR869837	F: CGCACTACTACGCACTGATA R: TCAGGTTTCCAATCCCATCC	$(GAG)_{13}$	55	2	215–227	0.150	0.156
ZpsjE-28	KR869839	F: CGATGATGCCGCACTAGA R: GAGATGCCAACACTGCACT	$(GGA)_{13}$	53	5	369–381	0.457	0.466
ZpsjE-35	KR869840	F: CACCCACGATCTAATCCCC R: GTCGTGGAAAGGCAGAGAC	$(AGG)_{12}$	55	4	280–295	0.247	0.242
ZpsjE-36	KR869841	F: AGCCCTCTGAAATAACGCAC R: GGTGTGGGAATCGATGAACA	$(AGG)_{12}$	56	3	236–251	0.229	0.316
ZpsjE-37	KR869842	F: GGCGCTCATTACGTTGTT R: CAGATTTCGGCACTTCCTC	$(AGG)_{12}$	57	5	208–223	0.606	0.592
ZpsjE-38	KR869843	F: CCACAGCCCTTTCTTTCT R: GTAGAGGATGGGATTGACG	$(CAC)_{12}$	53	2	282–285	0.132	0.135
ZpsjE-40	KR869845	F: TTTTGAGTCGCTGTCATCG R: CAAATCTGGACTGTTCTCGA	$(CAC)_{12}$	56	2	277–280	0.272	0.354
ZpsjE-41	KR869846	F: CCGAACAACAGACCCCATG R: GAATCTCGCTCAGCCCAAC	$(CAC)_{12}$	57	2	285–294	0.287	0.301
ZpsjE-44	KR869848	F: CAATGAACACCCCACCTTC R: CCCAAGTGTAGCTGATTCG	$(CCA)_{12}$	57	5	262–280	0.447	0.448
ZpsjE-45	KR869849	F: CCCCATTCTTTTCTCACCC R: GTTATCGAGTAGCCCGAC	$(CCA)_{12}$	56	4	270–285	0.171	0.187
ZpsjE-49	KR869850	F: CTTCGGGTGTTCTTCGACG R: CTCCTGCATCTCGGTCAGT	$(CCT)_{12}$	55	2	392–401	0.429	0.486
ZpsjE-53	KR869853	F: ACCACGGACCTCTACAAGTA R: GACTCTTTGCCCACTTTGCT	$(CGG)_{12}$	55	4	234–246	0.242*	0.414
ZpsjE-54	KR869854	F: GCTGCGAAGGTGATTTCA R: AGCTGTTACAAATGGGGC	$(CGG)_{12}$	57	5	272–290	0.286	0.311
ZpsjE-57	KR869856	F: GACCGTCCTCTACTGCAA R: CGGCCTCGTACTTTCTTG	$(CGG)_{12}$	53	4	310–325	0.314	0.403

Ta Annealing temperature, *Na* Number of alleles, *Ho* Observed heterozygosity, *He* Expected heterozygosity

*Signification deviation from Hardy–Weinberg equilibrium

except the EST-SSR markers from Liu et al. (2010b). Moreover, the relationship between these EST-SSR markers and other genes and markers is unclear due to limited flanking sequences and insufficient genome information. Thus, the EST-SSR markers' applicability is no different than genomic SSR markers. We believe that it is necessary to develop genic SSRs with sufficiently informative flanking sequences to increase application depth and width.

The development of next-generation sequencing technologies, such as the Roche/454 and Illumina Solexa platforms, has enabled the highly efficient identification of large numbers of SSRs, including genic SSRs, for an increasing number of species (Mikheyev et al. 2010; Dutta et al. 2011; Frenkel et al. 2012; Zheng et al. 2013). The Illumina platform operates with a much higher throughput and lower costs than the 454 sequencing platform. Although the Illumina platform generates shorter reads, those reads can be assembled into a draft genome and then used to discover SSRs with sufficient flanking sequences. Thus, the use of Illumina sequencing data to identify SSRs has evolved into a highly successful approach (Cai et al. 2013; Silva et al. 2013; Shi et al. 2014). Genomic SSRs have previously been identified based on a 500 bp insert Illumina paired-end shotgun survey for *S. japonica* (Zhang et al. 2014). In this study, we report the first identification of genic SSRs from the assembled and annotated *S. japonica* genome, from which we developed and characterized novel polymorphic trinucleotide genic SSR markers, providing a useful tool for studying breeding programs and germplasm conservation.

Materials and methods

Genomic DNA was extracted from inbred lineage using a previously described protocol (Cock et al. 2010). Three paired-end shotgun libraries (180, 500, and 800 bp) were constructed following the standard Illumina TrueSeqDNA sample preparation kit protocol, and then sequenced on an Illumina HiSeq 2000 (USA) following standard protocol. The raw reads were cleaned up by trimming the adapter and removing low-quality bases. The genome was assembled using SOAPdenovo (Li et al. 2010). Next, Augustus software was used to predict genes to divide the assembled genome into non-coding versus coding sequences (Stanke et al. 2008). Genic SSRs were then identified using the MISA perl script (Thiel et al. 2003), with the default minimum number of repeats set to 12, 6, 5, 5, 5, and 5 for the mono-through hexanucleotide simple sequence repeats, respectively. Primers were designed from the trinucleotide SSRs (repeat number more than 10) with sufficient flanking regions using the online software Primer3 (Koressaar and Remm 2007; Untergrasser et al. 2012), and synthesizedby Sangon Biotech (China).

Fifty-seven trinucleotide repeat microsatellite primer pairs were synthesized and evaluated by DNA amplification and agarose gel electrophoresis. Primer pairs that successfully amplified fragments were picked for subsequent polymorphism analysis. The polymorphisms of these loci were tested in 55 *Saccharina japonica* individuals, which were collected from an aquafarm located in Chudao Island in Shandong Peninsula, China (37°01′N, 122°33′E) in March, 2015. A 2.0-cm-diameter blade tissue disc was scrubbed with absorbent cotton,

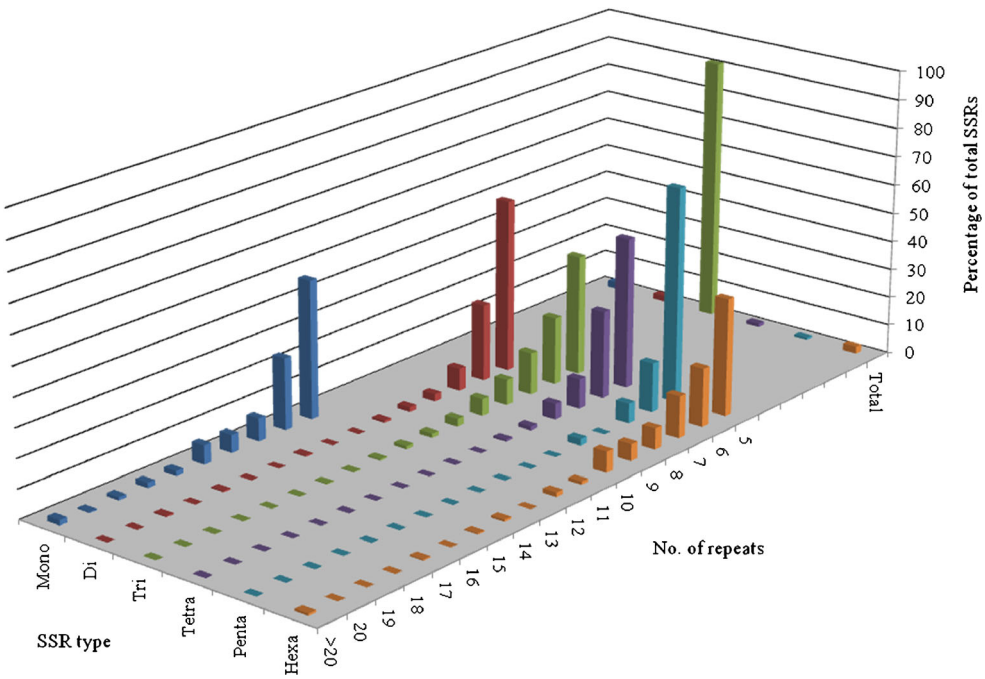

Fig. 1 Relative frequency (%) of SSR type in the *Saccharina japonica* assembly genome by number of repeats. The graph is based on a total of N=11,973 genic SSRs detected in 58 Mbp genomic coding sequence

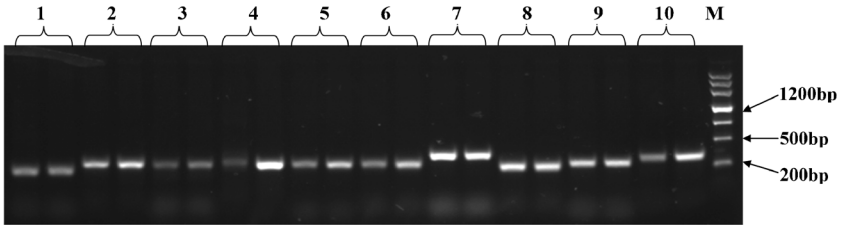

Fig. 2 Agarose gel electrophoresis of PCR products by 10 novel *Saccharina japonica* trinucleotide genic SSR markers. *Lanes1–10:ZspjE37, 38, 40, 41, 44, 45, 49, 53, 54, 57*; each primer pair was used to amplify DNA from two sporophyte individuals

washed with sterile seawater, and frozen at −70 °C for DNA extraction. Genomic DNA was extracted using a plant genomic DNA kit (Tiangen Biotech, China) according to the manufacturer's protocol. Polymerase chain reactions were performed in a total volume of 25 μL containing 1× buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs (each), 0.2 μM primers (each direction), 1U Taq polymerase (Tiangen Biotech), and about 100 ng DNA as the template. Thermocycling conditions are as follows: an initial step at 94 °C for 3 min; followed by 32 cycles of denaturation at 94 °C for 45 s, annealing at the appropriate temperature for every pair of primers (Table 1) for 45 s, and extension at 72 °C for 1 min; and a final extension at 72 °C for 10 min. PCR products were separated on a 6 % denaturing polyacrylamide gel, and visualized by silver staining. To assess genetic diversity, the number of alleles and the observed/expected heterozygosity were calculated using POPGENE 32 (Yeh et al. 2000). Linkage disequilibriumand Hardy–Weinberg equilibrium were tested using Arlequin version 3.5 (Excoffier and Lischer 2010).

Results and discussion

In total, 11,973 SSRs were identified in 58.3 Mb of assembled genomic coding sequence, a frequency of 205.5 genic SSRs per Mb of coding sequence. Trinucleotide SSRs were dominant (>90 %) among the nearly 12 thousand genic SSRs, followed in abundance by the mono-, hexa-, and dinucleotide SSRs; whereas tetra- and pentanucleotide SSRs were not commonly identified (<2 %) (Fig. 1). Therefore, we used trinucleotide SSRs to develop primer pairs for potential application as SSR markers. Microsatellite abundance generally decreased significantly with increasing motif repeat numbers (Fig. 1). Schlttöerer demonstrated that the mutation rate of SSRs increases with repeat number in previous mutational dynamics studies (2000). Although no study exists showing a robust relationship between microsatellite marker polymorphism level and the motif length, we randomly selected trinucleotide genic SSRs with repeat numbers of more than 10 to design primers.

Among the 57 primer pairs we designed, 32 easily amplified DNA, producing discrete, clear agarose gel electrophoresis bands (Fig. 2). Of these, 26 loci displayed polymorphism among the 55 *S. japonica* individuals tested (Fig. 3). The number of alleles per locus ranged from 2 to 7, with an average of 3.46, which is lower than the 5.7 average number of alleles per locus Liu et al. (2010b) reported for *S. japonica* EST-SSR loci. *Saccharina japonica* breeding strategies have included selecting only desirable individuals with excellent economic traits as seedlings, whereas others are discarded. Therefore, we propose that the lower allele count average that we found in this farmed population may be caused by the continuous artificial selection of desirable economic traits during the domestication process, further increasing allelic losses, reducing genetic diversity, and losing genetic resources (Liu et al. 2012). Alternatively, continuous inbreeding may also play a similar role after a few generations of repeated use (Shan et al. 2011). Either way, this highlights the need for protecting, managing, and utilizing the limited natural *S. japonica* resources still available during its continued cultivation. Additionally, the lower polymorphism level seen in this study may be due to the influence of different motif repeat lengths or repeat types. In *Brassica*, the polymorphism level of microsatellite markers for penta- and hexanucleotide repeats was significantly higher than that for mono- to tetranucleotide repeats (Shi et al. 2014). However, no study has yet shown a strong relationship between microsatellite marker polymorphism level and motif length. The 26 polymorphic loci we found had observed and expected heterozygosities varying between 0.128–0.652 and 0.130–0.676, respectively (Table 1). After adjustment by Bonferroni correction, significant deviation from the Hardy–Weinberg equilibrium was detected at two of our SSR loci (*ZpsjE-20* and

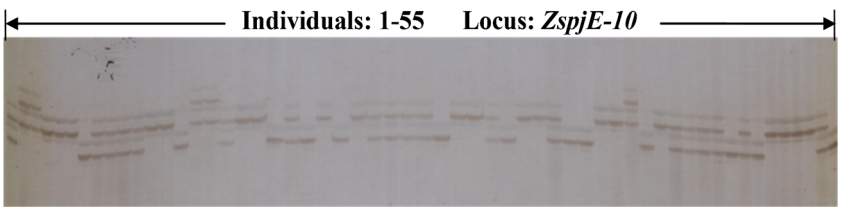

Fig. 3 SSR pattern of 55 *Saccharina japonica* individuals amplified by SSR primer *ZpsjE-10*

ZpsjE-53), probably also the result of the samples being from a cultivated population. There was no evidence of significant linkage disequilibrium except for markers *ZpsjE-1* and *ZpsjE-16*, which were strongly linked ($P<0.0002$).

A preliminarily assembled and annotated *S. japonica* genome enabled us to identify a large numbers of genic SSRs with sufficiently informative flanking sequences in a highly efficient manner. The novel genic SSR markers we developed in this study are sufficiently variable for population diversity analyses. These markers will be valuable tools for the conservation and management of *S. japonica* germplasm resources, and for further genetic studies, including parentage analysis, linkage mapping, and marker-assisted breeding. Furthermore, the *S. japonica* draft genome was reported recently (Ye et al. 2015). This will enable us to map our genic SSRs on the genome and determine positional relationships between the genic SSR markers and functional genes. The existence of genic SSR markers linked to quantitative traits that can be easily scanned and identified will greatly facilitate genetic breeding research in *Saccharina*.

Acknowledgments This study was funded by the National Supporting Program of Science and Technology of China (2012BAD55G01), National High Technology Research and Development Program of China (2012AA10A406), Municipal Science and Technology Research and Development Project of Yantai (2013LGS002), and Shandong Province Higher Educational Science and Technology Program (J15LE14).

References

Cai GH, Leadbetter CW, Muehlbauer MF, Molnar TJ, Hillman BI (2013) Genome-wide microsatellite identification in the fungus *Anisogramma anomala* using illumine sequencing and genome assembly. PLoS ONE 8:e82408

Cock JM, Sterck L, Rouzé P et al (2010) The *Ectocarpus* genomeand the independent evolution of multicellularity in brown algae. Nature 465:617–621

Dutta S, Kumawat G, Singh BP et al (2011) Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. BMC Plant Biol 11:17. doi:10.1186/1471-2229-11-17

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Fang TC, Wu CY, Jiang BY, Li JJ, Ren GZ (1962) The breeding of a new breed of Haidai (*Laminaria japonica* Aresch) and its preliminary genetic analysis. Acta Bot Sin 10:197–209 (in Chinese with English abstract)

Frenkel O, Portillo I, Brewer MT, Peros JP, Cadle-Davidson L, Milgroom MG (2012) Development of microsatellite markers from the transcriptome of *Erysiphe necator* for analysing population structure in North America and Europe. Plant Pathol 61:106–119

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–1291

Li XJ, Cong YZ, Yang GP, Shi YY, Qu SC, Li ZL, Wang GW, Zhang ZZ, Luo SJ, Dai HL, Xie JZ, Jiang GL, Liu JL, Wang TY (2007) 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 RQ, Zhu HM, Ruan J et al (2010) De novo assembly of human genomes with massively parallel short read sequencing. Genome Res 20:265–272

Li XJ, Zhang ZZ, Qu SC, Liang GJ et al (2015) Breeding of an intraspecific kelp hybrid Dongfang no. 6 *(Saccharina japonica*, Phaeophyceae, Laminariales) for suitable processing products and evaluation of its culture performance. J Appl Phycol. doi:10.1007/s10811-015-0562-0

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. 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, Yao JT, Wang XL, Repnikova A, Galanin DA, Duan DL (2012) Genetic diversity and structure within and between wild and cultivated *Saccharina japonica* (Laminariales, Phaeophyta) revealed by SSR markers. Aquaculture 358/359:139–145

Mikheyev AS, Vo T, Wee B, Singer MC, Parmesan C (2010) Rapid microsatellite isolation from a butterfly by de novo transcriptome sequencing: performance and a comparison with AFLP-Derived distances. PLoS ONE 5:e11212

Schlttöerer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109:365–371

Shan TF, Pag SJ, Zhang YR, Yakovleva IR, 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 et al (2007) Development of 18 polymorphic microsatellite DNA markers of *Laminaria japonica* (Phaeophyceae). Mol Ecol Resour 7:620–622

Shi JQ, Huang SM, Zhan JP, Yu JY, Wang XF, Hua W, Liu SY, Liu GH, Wang HZ (2014) Genome-wide microsatellite characterization and marker development in the sequenced *Brassica* crop species. DNA Res 21:53–68

Silva PIT, Martins AM, Gouvea EG, Pessoa-Filho M, Perreira ME (2013) Development and validation of microsatellite markers for *Brachiaria ruziziensis* obtained by partial genome assembly of illumina single-end reads. BMC Genomics 14:17–25

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24:637–644

Thiel T, Michalek W, 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

Tseng CK (2001) Algal biotechnology industries and research activitiesin China. J Appl Phycol 13:375–380

Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3–new capabilities and interfaces. Nucleic Acids Res 40:e115

Yang GP, Sun Y, Shi YY, Zhang LN, Guo SS, Li BJ, Li XJ, Li ZL, Cong YZ, Zhao YS, Wang WQ (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

Ye NH, Zhang XW, Miao M et al (2015) *Saccharina* genomes provide novel insight into kelp biology. Nat Commun 6:6986. doi:10.1038/ncomms7986

Yeh FC, Yang R, Boyle TJ, Ye Z (2000) POPGENE 32, Microsoft window-based freeware for population genetic analysis. Molecular

Biology and Biotechnology Centre, University of Alberta, Edmonton

Zhang QS, Tang XX, Cong YZ, Qu SC, Luo SJ, Yang GP (2007) Breeding of an elite *Laminaria* variety 90-1 through inter-specific gametophyte crossing. J Appl Phycol 19:303–311

Zhang LN, Peng J, Li XJ, Liu YL, Cui CJ, Wu H, Wu RN, Tian PP, Li Y (2014) Development of 27 trinucleotide microsatellite markers for *Saccharina japonica* using next generation sequencing technology. Conserv Genet Resour 6:341–344

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

Zheng XF, Pan C, Diao Y, You YN, Yang CZ, Hu ZL (2013) Development of microsatellite markers by transcriptome sequencing in two species of *Amorphophallus* (Araceae). BMC Genomics 14:490