



A pilot study of genetic structure of *Porphyra umbilicalis* Kützing in the Gulf of Maine using SNP markers from RNA-Seq

Yuanyu Cao¹ · Lindsay A. Green-Gavrielidis² · Renée L. Eriksen³ · Anita S. Klein³

Received: 31 October 2017 / Revised and accepted: 13 August 2018 / Published online: 1 September 2018
© The Author(s) 2018

Abstract

The red alga *Porphyra umbilicalis* Kützing has a broad distribution within the North Atlantic. In the Northeast Atlantic, *P. umbilicalis* is dioecious and reproduces both sexually and asexually, while in the Northwest Atlantic, only asexual reproduction has been observed. In this study, transcriptomes were mined to identify putative single nucleotide polymorphisms (SNP) markers. A computational pipeline was developed that accounts for the specific characteristics of transcriptome dataset, filtered against the available red alga *Chondrus* genome and *P. umbilicalis* EST library to eliminate microbial contamination. Five hundred forty-nine putative SNPs were detected within a single population (Schoodic Point, ME, USA). Five of the validated SNP markers were applied in a pilot study of genetic diversity and population structure of seven *P. umbilicalis* populations within the Gulf of Maine. Results of this study revealed the genetic diversity and structure of *P. umbilicalis* populations in the Gulf of Maine. Novel genotypes were found in the open coastal populations at Reid State Park, Schoodic Point, and the estuarine tidal rapid population at Wiscasset. Our study represents the first attempt to develop suitable bioinformatic pipeline for RNA-seq to detect SNP markers for red alga *Porphyra umbilicalis* and successfully used these SNP markers for population study.

Keywords *Porphyra* · RNA-seq · SNPs · Asexual population · Population genetics · Population structure

Introduction

The identification of genomic variation helps to clarify the relationship between genotype and phenotype. Single nucleotide polymorphisms (SNPs) are single base differences between DNA sequences of individuals or strains. SNP markers represent the most common type of variation across a genome (Kwok 2001). Studies of SNP markers facilitate marker-aided selection (MAS), detect alleles associated with disease, analyze population history, and are used to produce genetic maps (Cardon and Bell 2001; Flint-Garcia et al. 2003; Goya et al. 2010). The SNP markers were identified in silico from expressed sequence tag (EST) or genomic sequences, or by Sanger Sequencing of

candidate genes or PCR products (Gupta et al. 2008). The application for commercial SNP array has expanded from genotyping DNA to the detection and characterization of copy number variation and the loss-of heterozygosity (LaFramboise 2009). However, the rapid development and the reduced expense of next-generation sequencing technologies have provided unprecedented opportunities for researchers to find SNP markers, especially for non-model organisms, with greater resolutions and accuracy (LaFramboise 2009; Fonseca et al. 2016).

Transcriptome sequencing techniques (i.e., RNA-seq) can detect variants in the coding regions of the genome. RNA-seq provides large amount of sequencing data with reduced costs (Cloonan et al. 2008; Wilhelm et al. 2008). SNPs in noncoding regions may alter transcript levels by disrupting functional *cis*-regulatory elements, while those in coding regions of the genome may be silent or responsible for altered forms of proteins (Chepelev et al. 2009). RNA-seq is commonly employed to quantify gene expression levels under different conditions, and to detect alternative splicing, allele-specific expression, gene fusions and RNA editing (Wang et al. 2009; Eswaran et al. 2013; Piskol et al. 2013; Rapaport et al. 2013; Crowley et al. 2015; Conesa et al. 2016). Intrinsic complexities in transcriptomes (e.g., alternative splicing) lead to some challenges when using RNA-

✉ Yuanyu Cao
yov2@wildcats.unh.edu

¹ Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham 03824, USA

² Department of Natural Resources Science, University of Rhode Island, South Kingstown, USA

³ Department of Biological Sciences, University of New Hampshire, Durham, USA

seq data to identify genomic variants at computational analysis steps. Several methods have been developed to call variants from RNA-seq data with the availability of complete genome information (Piskol et al. 2013; Quinn et al. 2013; Deelen et al. 2015). Studies suggested imposing strong variant filtering criteria, having sufficient coverage, using relevant tissue, imposing suitable quality control screens and having additional whole-exome sequencing (Cirulli et al. 2010; Seo et al. 2012) can increase the accuracy of variant identification using transcriptome data. Recently, RNA-seq has been employed to systematically identify variants in transcribed regions in different species, mostly in humans (Chepelev et al. 2009; Cirulli et al. 2010; Quinn et al. 2013; Kim et al. 2014) and in some plants (Paritosh et al. 2013; Shearman et al. 2015).

The marine red alga *Porphyra umbilicalis* Kützinger is found from the Northeast Atlantic to the Northwest Atlantic, while its global distribution requires further review (Brodie et al. 2008). Sutherland et al. (2011) showed evidence from molecular data that the name of *P. umbilicalis* has been applied to more than one species. *Porphyra umbilicalis* is an important food source with high level of proteins, free amino acids, and a high ratio of $\omega 3:\omega 6$ fatty acids (Mouritsen 2013). It was demonstrated that *P. umbilicalis* can be used as a biofilter in integrated multi-trophic aquaculture (IMTA) and as a partial replacement for fishmeal (Day et al. 2009; Walker et al. 2009). *Porphyra umbilicalis*, which is native to northwest Atlantic, is an important component of intertidal communities and can be a good target to develop its domesticated strains for aquaculture and for components of IMTA (Blouin et al. 2007). The use of native strains of *P. umbilicalis* in the northwest Atlantic can potentially reduce the time, infrastructure, and costs associated with sexual spores (Blouin et al. 2007) and limit the potential ecological impact of introducing *Pyropia* species from Asia (Neefus et al. 2008).

Amplified fragment length polymorphism (AFLP; Blouin and Brawley 2012) and simple sequence repeat (SSR or microsatellite) markers (Eriksen et al. 2016) have been applied to analyze the genetic diversity and variation of the Northwest Atlantic populations of *P. umbilicalis*. However, the limited number of SSR markers used hinder the direct comparison of the microsatellite study with the AFLP assessment of *P. umbilicalis* genetic variation. Furthermore, there are large microbial communities associated with *P. umbilicalis*, living on the surface and inside the cell wall of the alga (Miranda et al. 2013). Bacterial DNA contamination of the *Porphyra* DNA pool may have generated more variable AFLP profiles. It is important to have more genetic markers that can be firmly tied to the *Porphyra* genome to analyze the genetic diversity and structure of Northwest Atlantic populations, and to aid the strain selection for integrated multi-trophic aquaculture.

In the Northeast Atlantic, the red alga *P. umbilicalis* can reproduce both sexually and asexually (Brodie and Irvine 2003); however, the Northwest Atlantic populations

reproduce only asexually (Bird and McLachlan 1992; Blouin et al. 2007; Gantt et al. 2010). The Northwest Atlantic *P. umbilicalis* populations are found over a wide latitudinal range and occupy both rocky, open coastal habitats and estuarine tidal rapids (Eriksen et al. 2016). How these asexual populations adapt to different habitats is still unclear. According to Ingolfsson (1992), the Northwest Atlantic rocky coast species are extirpated by the previous glacial period, and the current populations are assumed to be the descendants of the Northeast Atlantic species that were introduced from glacial refugia via post-glacial trans-Atlantic currents. Elucidating genetic diversity and variation of *P. umbilicalis* in the NW Atlantic could assist in the further understanding of how *P. umbilicalis* colonized the Gulf of Maine after the last glacial maximum, and how asexual *P. umbilicalis* populations survive in different habitat environments, such as open coastal and estuarine tidal rapid habitats.

In this study, a computational pipeline was developed to identify SNP markers from RNA-seq data in *P. umbilicalis*. This pipeline accounts for the specific characteristics of RNA-seq experiments, the biological characteristics of *P. umbilicalis* (microbial contamination), and the absence of a genome assembly of *P. umbilicalis* at the time this study was carried out. Twenty-five of these SNP markers were validated and five of them were then used in a pilot study to examine genetic diversity and population structure of *P. umbilicalis* within the Gulf of Maine.

Material and methods

Species identification: gametophyte culture, RNA isolation, and sequencing

Four “individuals” (each defined as thalli connected to a single holdfast) were collected randomly along the rocky shore during low tide at Schoodic Point, ME (44° 20' 11.3" N 68° 03' 23.3" W). Because morphological identification of *P. umbilicalis* is error prone (Klein et al. 2003), the *rbcL-rbcS* intergenic spacer gene was sequenced for all individuals used in this study to confirm species identity (Teasdale et al. 2002). The individuals that were confirmed to be *P. umbilicalis* were put into culture.

Neutral spores (asexual spores) release was induced from the four wild *P. umbilicalis* blades from Schoodic Point, ME, USA, which were isolated, and cultured in the lab-controlled environment according to Redmond et al. (2014). Blade materials from each culture were pooled, and then each pool was subjected to one of the five different treatment conditions (Table 1). The total RNA was extracted and assessed for quality and quantity using a NanoDrop 2000c spectrophotometer (ThermoFisher Scientific) and an Agilent 2100 Bioanalyzer. The cDNA libraries were prepared by Illumina Truseq RNA

Table 1 Treatments applied to each pooled culture

Library	Treatment conditions
A	Air dried for 4 h, frozen at $-20\text{ }^{\circ}\text{C}$ for 3 months, then cultured for 3 h
B	Cultured for 3 months
C	Air dried for 4 h, frozen at $-20\text{ }^{\circ}\text{C}$ for 3 months, then cultured for 2 weeks
D	Fully hydrated fresh tissue
E	Air dried for 4 h

Preparation Kit (Illumina, USA) and then sequenced on Illumina HiSeq 2000 at the Hubbard Center for Genome Studies in University of New Hampshire.

Reference transcriptome construction

Five libraries of transcriptome sequences from *P. umbilicalis* were generated under five environmental treatments; these transcriptomes were used to build the reference transcriptome assembly. Reads were first error corrected by BLESS (Heo et al. 2014) with kmerlength 25 and then trimmed with Trimmomatic (Bolger et al. 2014) at Phred score of 2 (Macmanes 2014) to get rid of low-quality reads and adapters. Subsequently, Trinity (Grabherr et al. 2011) was applied for de novo assembly with digital normalization. Due to known bacterial contamination on the surface and inside the cell walls of *P. umbilicalis* (Miranda et al. 2013), the raw de novo assembly was screened by BLAST analysis against the genome of the red alga *Chondrus crispus* Stackhouse (Collén et al. 2013), as well as the *P. umbilicalis* EST reference (Chan et al. 2012) to eliminate potential bacterial contamination. Only contigs that were similar to either the *C. crispus* genome or the *P. umbilicalis* EST reference with *E* values higher than 1e^{-10} were retained, resulting in a contamination-free, partial transcriptome reference for *P. umbilicalis*. The quality of the contamination-free reference was checked by the number of “genes,” the number of “transcripts,” the GC content and the N50 value. The completeness of the contamination-free reference was checked by Core Eukaryotic Genes Mapping Approach (CEGMA, Parra et al. 2007).

SNP detection

Reads from all five libraries were first trimmed at a Phred score of 20 to eliminate low-quality reads and then mapped to the reference transcriptome assembly respectively by BWA (Li and Durbin 2010). Samtools (Li et al. 2009a) was used to filter alignments with mapping quality lower than 30. Picard (The Broad Institute; <http://picard.sourceforge.net>) was used to mark duplicates and sort. SNP calling was performed using the Genome Analysis Toolkit (GATK, McKenna et al. 2010)

with Split’N’Trim. Indels and large structural variants were not analyzed. The raw putative SNPs were further filtered with sliding window of 35 and cluster of 3 based on the following criteria: (1) sequence depth at the SNP position ≥ 20 ; (2) FisherStrand (FS) ≤ 60 ; (3) RMSMappingQuality (MQ) ≥ 40 ; (4) MappingQualityRankSumTest (MQRankSum) ≥ -12.5 ; (5) ReadPosRankSumTest (ReadPosRankSum) ≥ -8.0 . After filtering, SNPs that were common to all five libraries were considered *P. umbilicalis* putative population SNPs and used for further validation. The genes containing SNPs that were unique to an individual treatment were annotated using Blast2GO (Conesa et al. 2005).

SNP validation

Twenty-five putative SNPs identified by computational pipeline were randomly chosen for validation using five individuals collected from five sites: Fort Stark, NH; Dover Point, NH; Lubec, ME; Reid State Park, ME; Nubble Light, ME (Fig. 4). Among these putative 25 SNPs, “true SNPs” were SNPs that were confirmed by Sanger Sequencing to be polymorphic among five individuals. “True SNP rate” is used as a measurement of the success of using current pipeline to detect SNP markers from transcriptome data in *P. umbilicalis*. The adjacent sequence around target SNP was retrieved from RNA-seq data to design primers. Primer pairs targeting each putative SNP were designed with Primer 3 software (available at: <http://primer3.ut.ee/>). Polymerase chain reaction (PCR) conditions were optimized for each primer pair and then used to amplify various DNA templates for the targeted SNP regions. Amplification of each target region was performed in 25 μL reaction volumes containing about 25–125 ng genomic DNA, $1\times$ Q5 One Taq Standard Reaction Buffer (New England BioLabs, USA), 200 μM dNTPs, 0.2 μM of each forward and reverse primer (Table 2), and 0.75 U One Taq Hot Start DNA Polymerase (New England BioLabs). PCR conditions started with an initial denaturation step of 94 $^{\circ}\text{C}$ for 3 min and was followed by 35 cycles of 30 s at 95 $^{\circ}\text{C}$, 1 min at a primer-specific annealing temperature (Table 2), and 1 min extension at 72 $^{\circ}\text{C}$. The amplification ended with a final extension at 72 $^{\circ}\text{C}$ for 5 min. The amplicon sizes were the same as predicted from the transcriptome sequence. The amplicons were purified using both QIAquick Gel extraction kit and ExoSAP-IT (Affymetrix.com) and then sent for Sanger Sequencing at GENEWIZ Company, USA. The Sanger sequencing results were aligned using MAFFT (<http://mafft.cbrc.jp/alignment/server/>) to detect SNPs. Trace files were utilized to further confirm SNPs found by MAFFT.

Genetic diversity and population structure

A total of five SNP markers (2, 8, 13, 18, and 22) that were validated by Sanger Sequencing were further used in the pilot

Table 2 Primers used to amplify regions containing SNP markers in *P. umbilicalis*

Primer	Major allele	Minor allele	Amplicon length	Annealing T _m (°C)	Sequence (the first is the forward primer and the second is the reverse primer)
2	C	T	599	62	5'-TGGGATCGTCACTCTCATCA-3' 5'-GAGGGTGTTC AAGGGAAAGG-3'
3	C	T	407	60	5'-TCGGGCGAATACTGTCTCTT-3' 5'-CGCCTCACCGATAATACGAG-3'
4	G	A	999	60	5'-CCTGTCGCTCTCTCTCTTG-3' 5'-TTGAGAGCCTTTTCCAGTCG-3'
8	G	T	561	62	5'-GTATGAGCGGACGGATAAGG-3' 5'-ATGAGCGAGCGTGGATAAAG-3'
10	G	A	398	60	5'-ATCGCAAAGGCGTTATCAAG-3' 5'-GTGTCGTCAACTGCAATCGT-3'
13	C	T	448	62	5'-GATGAAATCTCGGACCCTGA-3' 5'-TGAGGTTGAGCTTCGTGTTG-3'
14	T	C	752	60	5'-CTCTCCATCTGCTCCTTGCT-3' 5'-TTCTGTCACGCTCATTGGTC-3'
15	C	G	500	60	5'-CAACGTGATCCGCTACTTCC-3' 5'-CACCCGACTGCATCATCA-3'
16	G	T	382	60	5'-GGTCAATGCACAGGATGAGA-3' 5'-CACGAGCAGATTGTGTTGTC-3'
17	A	C	395	65	5'-GTGCTCGCCCTTCTACTAA-3' 5'-AGACAGAAGGTGGTGGGTTG-3'
18	C	G	450	60	5'-GACGTATGCTTTGTTCCGAGA-3' 5'-ACCTCCATTGCGTGGTAGAC-3'
19	C	T	447	60	5'-GTGGTGGATGTAGGCTCGTT-3' 5'-AGGTGCGTTGACAGGTGACT-3'
22	A	G	397	62	5'-GTTATTGCGCTCCTCATCGT-3' 5'-CGCACAGAGACCCTTCGT-3'

population study. Genes containing these SNPs were annotated for their functions using Blast2GO (Conesa et al. 2005). A total of 37 individuals from seven sampling sites (Fig. 4) were assayed for these five SNP markers. Among these seven populations, there are five open coastal and two estuarine tidal-rapid populations, Dover Point NH and Wiscasset ME that are geographically close to open coastal populations (Fort Stark NH and Reid State Park ME), respectively. PCR amplification conditions were the same as described above.

The major allele frequency, gene diversity, polymorphism information content (PIC), and Nei's genetic distances for each population were calculated using Power Marker 3.25 (Liu and Muse 2005). The SNP data was coded as follows: A = 1, C = 2, G = 3, T = 4 and missing data was coded as 0 as suggested in GenAlEx V6.5 user manual (Peakall and Smouse 2006, 2012). Analysis of molecular variance (AMOVA) and principal coordinate analysis (PCoA) were performed in GenAlEx v. 6.41 (Peakall and Smouse 2006, 2012). In addition, the Mantel test, with 9999 permutations, was conducted using the program GenAlEx 6.5 for correlation between

Nei's genetic distance and geographic distance. Because neutral spores travel along currents, geographic distance was calculated according to Eriksen (2014).

Results

Reference transcriptome

Comparisons between the raw Trinity transcriptome and the transcriptome reference corrected for contamination are summarized in Table 3. There were 182,905 contigs in the raw reference directly from Trinity output, with GC content of 56.47% and N50 of 512 bp. There were many fewer contigs in the contamination-free transcriptome reference after clean-up step (42,802 reads), with higher N50 length (1090 bp) and higher GC contents (62.28%). However, according to CEGMA analysis, the transcriptome reference without contamination was 4% less complete than the raw Trinity transcriptome reference.

Table 3 Assembly result comparison between Trinity output and Trinity minus contamination output

	# of contigs	N50	GC%	248 Core conserved eukaryotic genes (CEGMA)	
				# of protein	% complete
Trinity output	182,905	512	56.47	216	87
Trinity minus contamination	42,802	1090	62.28	206	83

SNP calling and functional annotation of library-specific SNP-containing genes

With stringent filtration, about 90% of the raw SNPs were filtered out. There were 549 putative SNPs in common to all five libraries (Fig. 1). The functional annotation containing unique putative SNPs in each library is showed in Fig. 2. There were less SNPs unique to the stress stimulus in library C (air dried, frozen, and then cultured for 2 weeks) compared to other libraries.

SNP validation

Primer pairs were designed for 25 gene transcripts containing putative SNPs. Target DNAs were amplified successfully for 13 of the 25 primer pairs. These primer pairs were used to amplify DNA from five individuals, each from one of the five different populations. The resulting amplicons were sent for Sanger Sequencing at GENEWIZ (<http://www.genewiz.com/>). Products of primer pair 4 did not contain the designed SNP, so this gene was dropped from further analysis. Amplicons from primer pairs 3 and 17 exhibited double peaks in the trace file from some individuals, even when sequenced from both ends. Among the remaining 10 primers pairs, seven were polymorphic based on screening five individual *P. umbilicalis* samples (SNP 2, 8, 13, 14, 15, 18, and 22). Two of the primer pairs (SNP 10 and SNP 16) produced monomorphic amplicons for the five *P. umbilicalis* individuals tested. SNP 19 was monomorphic in the predicted SNP position but was polymorphic at another position. The true SNP detection rate was likely higher than 70% since only five individuals were used to validate each putative SNP.

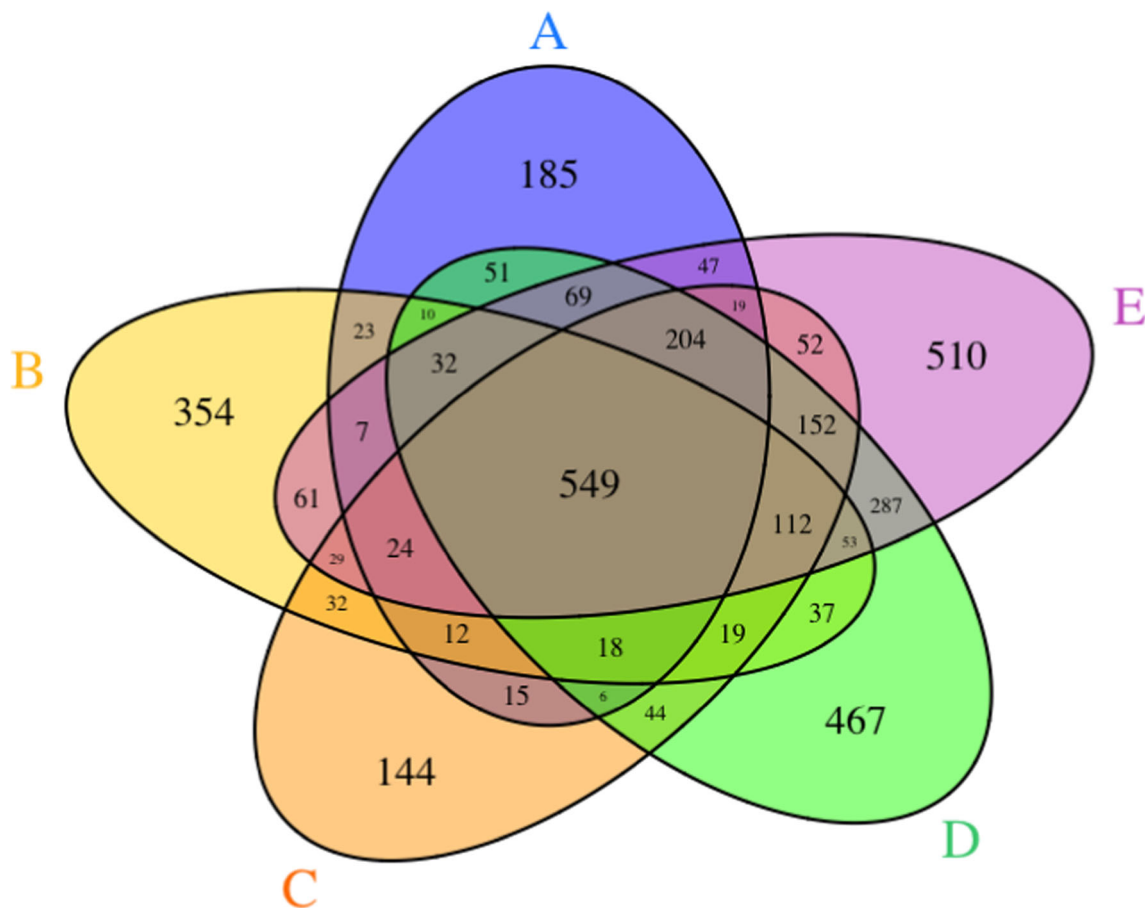
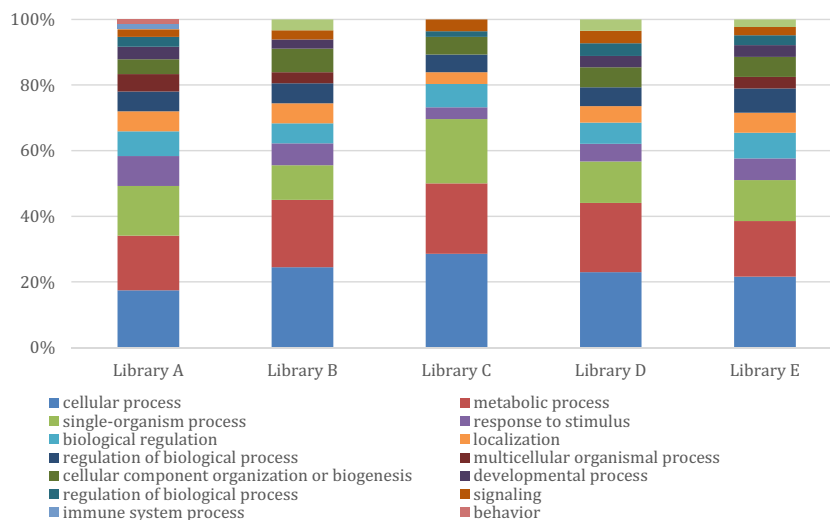


Fig. 1 Venn diagram of putative SNPs in each treatment library. Each oval represents each library. A total of 549 putative SNPs were found in common to all five libraries. Unique putative SNPs found: A-185; B-354; C-144; D-467; E-510

Fig. 2 Comparison of functional annotation of genes containing library-specific SNPs. Y-axis (%) means percentage of genes that have the corresponding functions illustrated



Pilot population study

Five random SNP markers from the seven validated polymorphic SNP loci were used for the population genetic study. The functions of these five amplicons were 6-phosphogluconate dehydrogenase decarboxylating (SNP 8), mRNA export factor/elongation factor (SNP 13), translocation protein 3Ec63 homolog (SNP 18), and unknown (SNP 2 and 22). There were more SNPs in the sequenced regions besides those inferred by bioinformatics of RNA-seq libraries and an additional eight SNPs were identified in comparing two or more amplicons from the same primers. Thus, 13 polymorphic SNPs were characterized in this study. The gene diversity and Polymorphic Information Content (PIC) value for each population are shown in Table 4. The gene diversity ranged from 0 to 0.16 and the PIC value ranged from 0 to 0.131. Dover Point, Nubble Light, Reid Park, and Wiscasset had the lowest genetic diversity and PIC values. Fort Stark and Lubec populations had intermediate levels of genetic diversity. The highest genetic diversity occurred from populations collected from Schoodic Point.

Table 4 Genetic diversity of seven *P. umbilicalis* populations

Sampling site	Abbr.	Sample size	Major allele freq.	Gene diversity	PIC
Dover Point, NH	D	5	1.000	0.000	0.000
Fort Stark, NH	F	5	0.969	0.037	0.028
Lubec, ME	L	5	0.892	0.160	0.131
Nubble Light, ME	N	5	1.000	0.000	0.000
Reid Park, ME	R	5	1.000	0.000	0.000
Schoodic Point, ME	S	7	0.890	0.144	0.115
Wiscasset, ME	W	5	1.000	0.000	0.000

Population structure

Genetic distance and genetic differentiation across the sampling sites are summarized in Table 5. The genetic distances were highest between Reid State Park and the rest of the populations, ranging from 0.171 to 0.308. Based on Mantel test, there was no evidence of isolation by distance between geographic distance and genetic distance ($R_{xy} = -0.344$, $p = 0.613$). Based on genetic distance, PCoA clustered three open coast populations (Fort Stark, Lubec, and Schoodic Point) into a central group, and Nubble Light and Dover Point into another group (Fig. 4). Reid State Park and Wiscasset were both on the right side of the PCoA but separated by PCoA2 (Fig. 3). AMOVA showed that there was more variation among populations (59%) than within populations (41%) indicating some level of colonality within populations.

Figure 4 shows the frequency of genotypes in each population. Besides in Schoodic Point (genotype 4–genotype 9), two other unique genotypes were found in Wiscasset (genotype 3) and Reid State Park (genotype 1). A common genotype (G2) was found in all populations except Reid State Park NH and Wiscasset ME. Schoodic Point ME had the highest genotypical diversity, followed by Lubec ME and Fort Stark NH.

Table 5 Genetic distance (below diagonal) and genetic differentiation (above diagonal) among seven sample sites of *Porphyra* within the Gulf of Maine

	D	F	L	N	R	S	W
D		0.500	0.071	0.000*	1.000*	0.237	1.000*
F	0.028		0.000	0.500	0.875*	0.318*	0.813*
L	0.058	0.042		0.071	0.639*	0.200*	0.458*
N	0.000	0.028	0.058		1.000*	0.237	1.000*
R	0.231	0.171	0.223	0.231		0.704*	1.000*
S	0.069	0.097	0.127	0.069	0.265		0.517*
W	0.077	0.105	0.135	0.077	0.308	0.146	

*Means significant with *p* value smaller than 0.05

Discussion

SNP discover and validation

Several bioinformatic pipelines have been developed to detect SNPs using genome information; these methods were reviewed by Pabinger et al. (2014). However, most of the existing pipelines were designed with the availability of a good reference genome and thus were species-specific, such as SOAPSnp (Li et al. 2009b), SNPdetector (Zhang et al. 2005), and SNPiR (Piskol et al. 2013). Several other pipelines have been successfully applied to both model and non-model species using only transcriptome data without a genome reference (Li and Godzik 2006; Van Belleghem et al. 2012; Piskol et al. 2013; Romiguier et al. 2014; Lopez-Maestre et al. 2016). In green algae, SNP markers have been identified using transcriptome data only (Li et al. 2014). This study benchmarks the detection of SNP markers from *P. umbilicalis* using only transcriptome dataset.

The computational pipeline we used accounted for the lack of a reference genome for *P. umbilicalis* at the time the research was conducted, RNA-Seq experiment intrinsic characteristics, and the extensive microbial contamination of *P. umbilicalis* tissues (Miranda et al. 2013). Filtering the

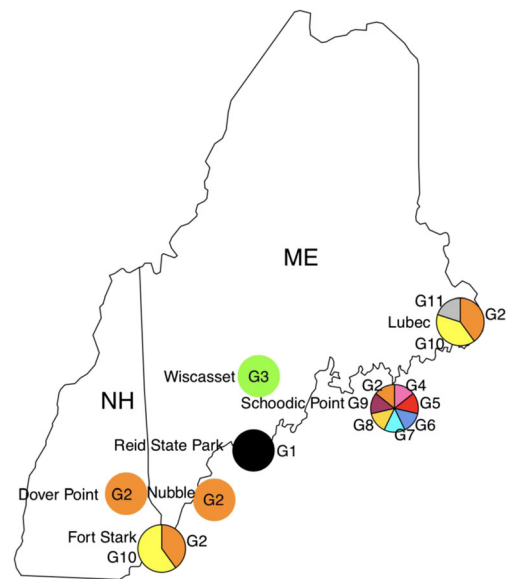
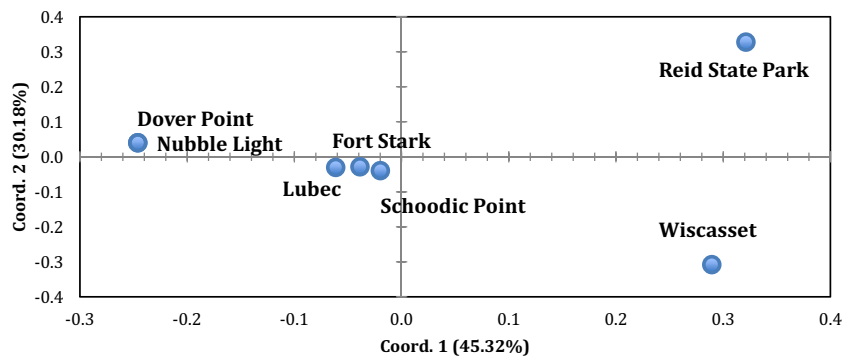


Fig. 4 Pie graph showing genotype frequency in each population. Each color represented a different genotype. The number represented the name of the genotype. NH stands for New Hampshire and ME stands for Maine

transcriptome assemblies against the *Porphyra* EST database (Chan et al. 2012) and *C. crispus* genome (Collén et al. 2013) eliminated microbial contamination and ensured the SNP markers found came from *Porphyra*. After filtering, the contamination-free transcriptome reference assembly was estimated to be slightly less complete (83 versus 87%) than the original Trinity reference transcriptome.

We found a total of 549 putative SNPs in common to five *Porphyra* RNA-Seq libraries. SNP validation result showed that at least 70% of the SNPs identified by the bioinformatic pipeline were true SNPs. The true SNP detection rate was likely higher than 70% since only five individuals from different algal populations were screened for polymorphism for each amplicon. Increasing the number of individuals or increasing the number of sampling sites will likely increase the true SNP detection rate. These 549 SNPs are good candidate markers for further genetic diversity and population structure analysis, especially to study the evolutionary history of

Fig. 3 Principal coordinate analysis (PCoA) based on genetic distance among seven populations. Dover Point and Nubble Light populations mapped to the same PCoA coordinate. Percentage of the total variation explained by each axes was shown



how mutations accumulate in asexual *Porphyra* populations in the Northwest Atlantic.

Library C went through air-drying and freezing stress conditions, which might be expected to cause more stress-related genes to be expressed. However, there were less unique SNPs detected responding to stimulus (stress related) in library C compared to other libraries. Based on annotation results, we suspect that some library-specific SNPs were identified due to differences in transcript coverage in each library rather than due to differential gene expression. We showed that RNA-seq data is suitable to find SNPs without bias towards specific treatments. We also identified two SNPs (3 and 17) that had high background noise in the sequencing files. This high background noise could be the result of polymorphisms from recent gene duplications; this is a drawback to not having a reference genome with information about underlying gene family structure.

Population genetics analysis

A total of 13 polymorphic SNP sites were used to investigate the genetic diversity and population structure of *P. umbilicalis* from seven populations within the Gulf of Maine. A total of 11 genotypes were found in our study. Genotype 2 was present in a wide range of environmental conditions from the northern border of Maine to New Hampshire, as well as the estuarine environment at Dover Point. A “general purpose genotype” (Baker 1965) that can confer broad environmental tolerance will rapidly increase its frequency in a population without the selection of a locally adapted variant. Genotype 2 in our study is likely a “general purpose genotype.” However, this “general purpose genotype” may not be able to deal with all changing environments (Selander and Hudson 1976) as genotype 2 was not detected in Wiscasset ME and Reid State Park ME, which have unique environmental characteristics.

The genetic diversity of asexual populations depends on the number of possible mutations that occur over time in the populations, and the proportion of these mutations that persist through time (Good et al. 2012). In general, postglacial recolonization from refugia is related to the genetic diversity of the rocky shore fauna and marine macroalgae in the Northwest Atlantic (Ingolfsson 1992; Teasdale and Klein 2010). The population from Schoodic Point had the highest genotypic diversity, with each individual examined having a unique genotype; this population also had the highest genetic diversity. Schoodic Point is the north most population among the seven populations, which is very likely to be the somewhat older than more southerly populations in New England (Teasdale and Klein 2010). However, the high genetic diversity observed for the Schoodic Point could also suggested there may be population ascertainment bias as the SNPs were originally discovered using cultures established from multiple individuals from Schoodic Point. This would explain why the

Schoodic Point population had the highest genetic diversity and is most genotypically diverse. The populations with the low (or absence of) allelic diversity (Dover Point, Nubble Light, Wiscasset, and Reid State Park) are all marginal populations. Marginal populations were showed with decreased genetic and allelic diversity in brown alga *Laminaria digitata* (Oppliger et al. 2014). Also, Dover Point, Wiscasset, and Reid State Park populations inhabit somewhat atypical environments for *P. umbilicalis* (estuarine tidal rapids, or sandy substrata). Selection for these atypical environments may further reduce the genetic diversity in these populations. The low genetic diversity would limit the ability of populations to adapt to the changing environments and thus impact their long-term survival potential, especially under stressful conditions (Markert et al. 2010).

In addition to seven unique genotypes found within the Schoodic Point population, there were two other novel genotypes. One was from the estuarine population at Wiscasset ME and the other was from the nearby coastal population at Reid State Park ME. The unique genotype from Wiscasset (genotype 3) was only one locus different from genotype 2 and may be the result of genetic drift. Another possible explanation is that the low salinity in the estuarine environment induces the unique genotype as suggested by Ram and Hadany (2014); stress-induced mutagenesis can help generate a better adaptive genotype in an extreme environment. The other unique genotype (genotype 1) was only present in Reid State Park population. This genotype differs at three loci from genotype 2. These three loci were not in the same gene but all reach fixation based on our limited sampling in the population (five individuals). Besides the possibility of genetic drift, it is also likely that these three fixed mutations occurred in succession as three hard selective sweeps events or one clone under the effect of clonal interference. As Lang et al. (2011) suggested that clonal interference is far more likely to happen to any mutation compared to selective sweeps in asexual population. It is very likely that we missed other clones in the Reid State Park population. The genetic distance and differentiation also showed that significantly high genetic differentiation existed between Reid State Park and the rest of the populations. It is not clear why Reid State Park has such a high genetic differentiation to the rest populations. Bottlenecks or extirpation events in the past followed by subsequent recolonization could lead to high genetic differentiation. Also, it can be caused by the unique environment in Reid State Park, as the sampling site in Reid State Park has more sandy sediment nearby. However, larger sample sizes from Reid State Park are needed to draw any conclusion for the high genetic variation in Reid State Park.

The previous study of genetic diversity of these same populations (Eriksen et al. 2016) looked at a much larger number of individuals (221) using three polymorphic SSR markers. A total of six genotypes were identified by SSR, compared to the

11 genotypes found in this study that sampled 37 individuals. One explanation for the lower level of polymorphisms identified by SSR loci is that SSR markers used by Eriksen et al. (2016) were developed from protein coding regions. Expansion and contraction of SSR regions in the protein coding regions may be more functionally constrained than SNPs in the same region. Different genotype patterns were also observed by Eriksen et al. (2016); they reported two or more SSR genotypes in each population, with the largest number of genotypes (4) were observed for the Fort Stark NH population. By contrast, the present study found four populations had single genotypes (Reid State Park ME, Wiscasset ME, Nubble Light ME, and Dover Point NH) and Schoodic Point had the highest number of genotypes. Different patterns of genetic diversity reported by SNP markers microsatellites markers were also reported in red alga *Chondrus crispus* (Provan et al. 2013). However, we found much less amounts of genotypes in our study comparing to those in Blouin and Brawley (2012), which showed 41 clones in 51 individuals from 2 populations in ME. The difference between these two studies were probably due to that AFLP DNA fingerprinting was influenced by non-target DNA caused by cryptic contaminants.

With predominant asexual reproduction of *P. umbilicalis* in the Northwest Atlantic (Blouin and Brawley 2012), genetic differentiation among *Porphyra* populations is restricted by the dispersal ability of neutral spores. There is no report on how far *Porphyra* neutral spores can travel. In natural environment with water constantly stirred, spores tend to stay near the surface. The rate of spore sinking is low, irrespective of spore size in other algae species (Hoffmann and Camus 1989). While under lab conditions, *Porphyra* spores tend to settle down and attach to substrate within 12–24 h of release in still water (unpub. data). In general, the genetic differentiation among seven populations reported was much higher than that reported by EST-SSR markers (Eriksen et al. 2016), which agrees with the finding in red alga *Furcellaria lumbricalis* that showed the level of genetic differentiation reported by SNP markers was higher than that reported by EST-derived neutral microsatellites markers (Olsson and Korpelainen 2013). The genetic differentiation among the three geographically close southern populations (Fort Stark, Nubble Light, and Dover Point) was low and insignificant, possibly because these three populations are within the dispersal range of neutral spores (< 30 km apart). It is possible that the genetic differentiation among these populations is mainly driven by migration. The high genetic differentiation found among the rest populations (Fort Stark, Nubble Light, and Dover Point) was consistent with the finding in red alga *Corallina officinalis* by SNP markers.

We found that genetic variation was slightly higher among populations than within populations based on AMOVA results, suggesting that genetic differentiation may be caused by genetic drift or selection (Excoffier et

al. 1992). Genetic drift and selection are more likely to happen to small relatively isolated estuarine populations like Wiscasset ME and Dover Point NH or populations in atypical environments like Reid State Park ME. Although Reid State Park is geographically closest to the estuarine population Wiscasset, it was genetically most distant to Wiscasset. Genetic drift and selection for specific environments (estuarine environment for Wiscasset and more sandy sediment for Reid State Park) may play an important role in the high genetic differentiation between these two nearby populations (27 km apart).

It is worth mentioning that observed genetic structure and genotype variation for *P. umbilicalis* may also vary year-round and between years. Drenth et al. (1994) showed that the genetic structure of asexual fungal populations was different every year as less than 10% of the genotypes survived every year. Selection has the ability to change the genetic structure of small populations rapidly (Worrall 2012) and selection can vary considerably from year to year within a population (Price et al. 1984, Milner et al. 1999). In *P. umbilicalis* as population size drops in the summer due to thermal and UV stress; random genetic drift will increase as the power of random genetic drift is inversely proportional to the effective population size (Lynch et al. 2016). Environmental impacts on genetic diversity suggest that it is better to collect samples at the same season in order to accurately compare genetic diversity and population structure. The samples used in our study were originally collected by Eriksen et al. (2016), at different times of the year and thus our inferences about population structure should be viewed with caution. In future studies, inclusion of a larger number of markers and more samples per population should yield a better estimate of genetic diversity and population structure of *P. umbilicalis* within the Gulf of Maine.

Acknowledgements The authors thank Kelley Thomas of the Hubbard Genome Center and Matthew MacManes for the helpful discussions.

Funding information Partial funding was provided by the New Hampshire Agricultural Experiment Station (Scientific Contribution Number 2749). This work was supported by the USDA National Institute of Food and Agriculture Hatch Project 1004051. Additional support also was provided by the Leslie S. Hubbard Marine Program Endowment (to L. Green-Gavrielidis), NSF 0929558 (to Susan H. Brawley and Arthur R. Grossman), NOAA contract NA060AR4170108 (to SHB), and by the Joint Genome Institute (U.S. Dept. of Energy) under a Community Science Program award from the Office of Science of the U.S. Dept. of Energy under Contract No. DE-AC02-05CH11231 (to Susan H. Brawley, Elizabeth Gantt, Arthur R. Grossman and John Stiller).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Baker HG (1965) Characteristics and modes of origins of weeds. In: Baker HGSG (ed) The genetics of colonizing species. Academic Press, New York, pp 147–172
- Bird CJ, McLachlan JL (1992) Seaweed flora of the maritimes: 1. Rhodophyta - The red algae. Biopress Ltd, Bristol
- Blouin N, Brawley S (2012) An AFLP-based test of clonality in wide-spread, putatively asexual populations of *Porphyra umbilicalis* (Rhodophyta) in the Northwest Atlantic with an in silico analysis for bacterial contamination. *Mar Biol* 159:2723–2729
- Blouin N, Xiugeng F, Peng J, Yarish C, Brawley SH (2007) Seeding nets with neutral spores of the red alga *Porphyra umbilicalis* (L.) Kützinger for use in integrated multi-trophic aquaculture (IMTA). *Aquaculture* 270:77–91
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Brodie JA, Irvine LM (2003) Seaweeds of the British Isles Vol. 1 Rhodophyta Part 3B Bangiophycidae. The Natural History Museum, London
- Brodie J, Mortensen AM, Ramirez ME, Russell S, Rinkel B (2008) Making the links: towards a global taxonomy for the red algal genus *Porphyra* (Bangiales, Rhodophyta). *J Appl Phycol* 20:939–949
- Cardon LR, Bell JI (2001) Association study designs for complex diseases. *Nat Rev Genet* 2:91–99
- Chan CX, Blouin NA, Zhuang Y et al (2012) *Porphyra* (Bangiophyceae) transcriptomes provide insights into red algal development and metabolism. *J Phycol* 48:1328–1342
- Chepelev I, Wei G, Tang Q, Zhao K (2009) Detection of single nucleotide variations in expressed exons of the human genome using RNASeq. *Nucleic Acids Res* 37:106
- Cirulli ET, Singh A, Shianna KV et al (2010) Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. *Genome Biol* 11:R57
- Cloonan N, Forrest ARR, Kolle G et al (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5:613–619
- Collén J, Porcel B, Carré W et al (2013) Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc Natl Acad Sci U S A* 110:5247–5252
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676
- Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol* 17:13
- Crowley JJ, Zhabotynsky V, Sun W et al (2015) Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nature Genet* 47:353–360
- Day JP, Neefus CD, Yarish C (2009) Development of a modular integrated recirculating aquaculture system using *Porphyra* for bioremediation of marine finfish effluent. *J Phycol* 45:31–32
- Deelen P, Zhenakova DV, de Haan M, van der Sijde M, Bonder MJ, Karjalainen J, van der Velde KJ, Abbott KM, Fu J, Wijmenga C, Sinke RJ, Swertz MA, Franke L (2015) Calling genotypes from public RNA-sequencing data enables identification of genetic variants that affect gene-expression levels. *Genome Med* 7:30
- Drenth A, Tas ICQ, Govers F (1994) DNA fingerprinting uncovers a new sexually reproducing population of *Phytophthora infestans* in the Netherlands. *Eur J Plant Pathol* 100:97–107
- Eriksen RL (2014) Population genetics and organism-environment interactions of *Porphyra umbilicalis* Kützinger in the Gulf of Maine. Dissertation, University of New Hampshire
- Eriksen RL, Green LA, Klein AS (2016) Genetic variation within and among asexual populations of *Porphyra umbilicalis* Kützinger (Bangiales, Rhodophyta) in the Gulf of Maine, USA. *Bot Mar* 59: 1–12
- Eswaran J, Horvath A, Godbole S, Reddy SD, Mudvari P, Ohshiro K, Cyanam D, Nair S, Fuqua SAW, Polyak K, Florea LD, Kumar R (2013) RNA sequencing of cancer reveals novel splicing alterations. *Sci Rep* 3:1689
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Flint-Garcia SA, Thomsberry JM, Buckler ES (2003) Structure of linkage disequilibria in plants. *Annu Rev Plant Biol* 54:357–374
- Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madriral J, Sibbesen JA, Maretty L, Zepeda-Mendoza ML, Campos PF, Heller R, Pereira RJ (2016) Next-generation biology: sequencing and data analysis approaches for non-model organisms. *Mar Genomics* 30:3–13
- Gantt E, Berg GM, Bhattacharya D et al (2010) *Porphyra*: complex life histories in a harsh environment. *P. umbilicalis*, an intertidal red alga for genomic analysis. In: Seckbach J, Chapman D (eds) Red Algae in the Genomic Age. Cellular Origin, Life in Extreme Habitats and Astrobiology Springer. Springer, New York, pp 129–148
- Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM (2012) Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc Natl Acad Sci U S A* 109:4950–4955
- Goya R, Sun MGF, Morin RD, Leung G, Ha G, Wiegand KC, Senz J, Crisan A, Marra MA, Hirst M, Huntsman D, Murphy KP, Aparicio S, Shah SP (2010) SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics* 26:730–736
- Grabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652
- Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity* 101:5–18
- Heo Y, Wu X, Chen D, Ma J, Hwu WM (2014) BLESS: bloom filter-based error correction solution for high-throughput sequencing reads. *Bioinformatics* 30:1354–1362
- Hoffmann AJ, Camus P (1989) Sinking rates and viability of spores from benthic algae in central Chile. *J Exp Mar Biol Ecol* 126:281–291
- Ingólfsson A (1992) The origin of the rocky shore fauna of Iceland and the Canadian maritimes. *J Biogeogr* 19:705–712
- Kim S, Kim D, Cho SW, Kim J, Kim JS (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* 24:1012–1019
- Klein AS, Mathieson AC, Neefus CD (2003) Identifications of northwestern Atlantic *Porphyra* (Bangiaceae, Bangiales) based on sequence variation in nuclear SSU and rbcL genes. *Phycologia* 42:109–122
- Kwok P (2001) Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet* 2:235–258
- LaFramboise T (2009) Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucl Acids Res* 37:4181–4193
- Lang GI, Botstein D, Desai MM (2011) Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics* 188:647–661
- Li H, Durbin R (2010) Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* 26:589–595
- Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009a) The sequence alignment map format and SAMtools. *Bioinformatics* 25:2078–2079
- Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, Wang J (2009b) SNP detection for massively parallel whole-genome resequencing. *Genome Res* 19:1124–1132

- Li Q, Liu J, Zhang L, Liu Q (2014) *De novo* transcriptome analysis of an aerial microalga *Trentepohlia jolithus*: pathway description and gene discovery for carbon fixation and carotenoid biosynthesis. *PLoS One* 9:e108488
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129
- Lopez-Maestre H, Brinza L, Marchet C, Kielbassa J, Bastien S, Boutigny M, Monnin D, Filali AE, Carareto CM, Vieira C, Picard F, Kremer N, Vavre F, Sagot MF, Lacroix V (2016) SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. *Nucleic Acids Res* 44:e148
- Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL (2016) Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet* 17:704–714
- Macmanes MD (2014) On the optimal trimming of high-throughput mRNA sequence data. *Front Genet* 5:13
- Markert JA, Champlin DM, Gutjahr-Gobell R, Grear JS, Kuhn A, McGreevy TJ Jr, Roth A, Bagley MJ, Nacci DE (2010) Population genetic diversity and fitness in multiple environments. *BMC Evol Biol* 10:205
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
- Milner JM, Albon SD, Illius AW, Pemberton JM, Clutton-Brock TH (1999) Repeated selection of morphometric traits in the Soay sheep on St Kilda. *J Anim Ecol* 68:472–488
- Miranda LN, Hutchison K, Grossman AR, Brawley SH (2013) Diversity and abundance of the bacterial community of the red macroalga *Porphyra umbilicalis*: did bacterial farmers produce macroalgae? *PLoS One* 8:e58269
- Mouritsen OG (2013) Seaweeds: edible, available & sustainable. The University of Chicago Press, Chicago
- Neefus CD, Mathieson AC, Bray TL, Yarish C (2008) The distribution, morphology, and ecology of three introduced Asiatic species of *Porphyra* (Bangiales, Rhodophyta) in the northwestern Atlantic. *J Phycol* 44:1399–1414
- Olsson S, Korpelainen H (2013) Single nucleotide polymorphisms found in the red alga *Furcellaria lumbricalis* (Gigartinales): new markers for population and conservation genetic analyses. *Aquat Conserv* 23:460–467
- Oppliger LV, Von Dassow P, Bouchemousse S, Robuschon M, Valero M, Correa JA, Mauger S, Destombe C (2014) Alteration of sexual reproduction and genetic diversity in the kelp species *Laminaria digitata* at the southern limit of its range. *PLoS One* 9(7):e102518
- Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efreanova M, Krabichler B, Speicher MR, Zschocke J, Trajanoski Z (2014) A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform* 15:256–278
- Paritosh K, Yadava SK, Gupta V, Panjabi-Massand P, Sodhi YS, Prahan AK, Pental D (2013) RNA-seq based SNPs in some agronomically important oleiferous lines of *Brassica rapa* and their use for genome-wide linkage mapping and specific-region fine mapping. *BMC Genomics* 14:463
- Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- Piskol R, Ramaswami G, Li JB (2013) Reliable identification of genomic variants from RNA-seq data. *Am J Hum Genet* 93:641–651
- Price TD, Grant PR, Gibbs HL, Boag PT (1984) Recurrent patterns of natural selection in a population of Darwin's finches. *Nature* 309:787–789
- Provan J, Glendinning K, Kelly R, Maggs CA (2013) Levels and patterns of population genetic diversity in the red seaweed *Chondrus crispus* (Florideophyceae): a direct comparison of single nucleotide polymorphisms and microsatellites. *Biol J Linn Soc* 108:251–262
- Quinn EM, Cormican P, Kenny EM, Hill M, Anney R, Gill M, Corvin AP, Morris DW (2013) Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid cell lines and evaluation using 1000 genomes data. *PLoS One* 8:e58815
- Ram Y, Hadany L (2014) Stress-induced mutagenesis and complex adaptation. *Proc Biol Sci* 281
- Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND, Betel D (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* 14:R95
- Redmond S, Green L, Yarish C, Kim J, Neefus C (2014) New England seaweed culture handbook. Seaweed Cultivation
- Romiguier J, Gayral P, Ballenghien M et al (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261–263
- Selander RK, Hudson RO (1976) Animal population structure under close inbreeding: the land snail *Rumina* in southern France. *Am Nat* 110:695–718
- Seo J, Ju YS, Lee W, Shin JY, Lee JK, Bleazard T, Lee J, Jung YJ, Kim JO, Shin JY, Yu SB, Kim J, Lee ER, Kang CH, Park IK, Rhee H, Lee SH, Kim JI, Kang JH, Kim YT (2012) *Genome Res* 22:2109–2019
- Shearman JR, Sangsakru D, Jomchai N, Ruang-areerate P, Sonthirod C, Naktang C, Theerawattanasuk K, Tragoonrun S, Tangphatsornruang S (2015) SNP identification from RNA sequencing and linkage map construction of rubber tree for anchoring the draft genome. *PLoS One* 10:e0121961
- Sutherland JE, Lindstrom SC, Nelson WA, Brodie J, Lynch MD, Hwang MS, Choi HG, Miyata M, Kikuchi N, Oliveira MC, Farr T, Neefus C, Mols-Mortensen A, Milstein D, Müller KM (2011) A new look at an ancient order: generic revision of the Bangiales (Rhodophyta). *J Phycol* 47:1131–1151
- Teasdale BW, Klein AS (2010) Genetic variation and biogeographical boundaries within the red alga *Porphyra umbilicalis* (Bangiales). *Rhodophyta Bot Mar* 53:413–417
- Teasdale B, West A, Taylor H, Klein A (2002) A simple restriction fragment length polymorphism (RFLP) assay to discriminate common *Porphyra* (Bangioophyceae, Rhodophyta) taxa from the Northwest Atlantic. *J Appl Phycol* 14:293–298
- Van Belleghem SM, Roelofs D, Van Houdt J, Hendrickx F (2012) *De novo* transcriptome assembly and SNP discovery in the wing polymorphic salt marsh beetle *Pogonus chalceus* (Coleoptera, Carabidae). *PLoS One* 7:e42605
- Walker AB, Fournier HR, Neefus CD et al (2009) Partial replacement of fish meal with laver *Porphyra* spp. in diets for Atlantic cod. *North Amer J Aquacult* 71(1):39–45
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63
- Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453:1239–1243
- Worrall JJ (ed) (2012) Structure and dynamics of fungal populations. Springer, Dordrecht
- Zhang J, Wheeler DA, Yakub I, Wei S, Sood R, Rowe W, Liu PP, Gibbs RA, Buetow KH (2005) SNPdetector: a software tool for sensitive and accurate SNP detection. *PLoS Comput Biol* 1:e53