Aquaculture 452 (2016) 351–356

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aquaculture

Developing a core collection of *Pyropia haitanensis* using simple sequence repeat markers

Yan Xu [a,b], Changsheng Chen [a,b], Dehua Ji [a,b], Kai Xu [a,b], Xiaoxue Xie [a,b], Chaotian Xie [a,b,*]

[a] College of Fisheries, Jimei University, Xiamen 361021, PR China
[b] Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of Agriculture, Xiamen 361021, PR China

ARTICLE INFO

Article history:
Received 3 August 2015
Received in revised form 4 November 2015
Accepted 10 November 2015
Available online 10 November 2015

Keywords:
Pyropia haitanensis
Core collection
Genetic diversity
SSR marker

ABSTRACT

Pyropia haitanensis is one of the most economically important mariculture crops of southern China. Core collection construction is important for the study, management and utilization of germplasm resources of *P. haitanensis*. In this study, methods for the construction of a core collection of *P. haitanensis* based on SSR marker data were developed and a core collection was created. Using 126 SSR alleles of 77 strains of *P. haitanensis*, 30 core subsets were constructed by stepwise clustering using two sampling strategies (random sampling and preferred sampling), three genetic distances (Nei & Li genetic distance (ND), simple matching genetic distance (SM), and Jaccard genetic distance (JD)) and five sample sizes (10%, 15%, 20%, 25%, 30%). Three parameters (Ne, H, and I) of genetic diversity were used to evaluate the genetic diversity of the 30 core subsets. The core subset that was constructed using preferred sampling at 15% of sample size with the SM genetic distance best represented the initial collection of the 30 core subsets and is recommended to be used as the core collection of *P. haitanensis*.

Statement of relevance

Pyropia haitanensis is one of the most economically important mariculture crops of southern China. Core collection construction is important for the study, management and utilization of germplasm resources of *P. haitanensis*.

1. Introduction

Pyropia/*Porphyra* is one of the most economically important mariculture crops with an annual harvest of more than 120,000 tons (dry weight) and a value of over $2 (USD) billion per year and can be used as a source of food, fertilizer, medicine, chemicals and other uses (Sahoo et al., 2002; Blouin et al., 2010). The genus *Pyropia*/*Porphyra* is distributed worldwide, and >130 species have been reported (Zhang et al., 2005; Sutherland et al., 2011). Several species, including *P. yezoensis*, *P. haitanensis*, *P. tenera* and *P. seriata*, have been cultivated for the aquaculture industry in East Asia. *Pyropia haitanensis*, a typical warm temperate zone species originally found in the Fujian Province of China, has been widely cultivated along the coasts of southern China for more than fifty years. In recent years, *P. haitanensis* has comprised 75% of the total production of cultivated *Pyropia* in China (Xie et al., 2013). Through years of genetic study and breeding, some improved varieties of *P. haitanensis* have been obtained and cultivated widely (Chen et al., 2008a; Chen et al., 2008b; Yan et al., 2010). To some degree, this enhanced the cultivation of *P.haitanensis* and promoted the industrial development of this economic seaweed. However, *P. haitanensis* cultivation still faces many problems. First, the cultivation of *P. haitanensis* to date in some areas still relies on natural populations, with very limited germplasm development and genetic improvement. Second, varieties of *P. haitanensis* with high yield or high quality are still lacking (Xu et al., 2012). Thus, it is highly desirable to carry out breeding studies and to cultivate improved species to increase the industrial economic efficiency and to expand the scale of *P. haitanensis* cultivation.

Plant breeding is a dynamic area of applied science, relying on genetic variation combined with selection to improve plant characteristics that are of interest to the grower and consumer. Creating a germplasm collection is an important and efficient strategy for preservation of available genetic variability before it is lost due to widespread use of improved cultivars and disruption of the environment (Tanksley and

Abbreviations: A, polymorphic alleles; CV, coefficient of variability; H, Nei's gene diversity; I, Shannon's information index; JD, Jaccard genetic similarity coefficient; Na, number of observed number alleles; ND, Nei & Li genetic similarity coefficient; Ne, number of effective alleles; P, percentage of polymorphic alleles; PES, Provasoli's enrichment solution; PCO, principal coordinates; PS, preferred sampling; RS, random sampling; SM, simple matching coefficient; STDV, standard deviation; UPGMA, Unweighted Pair Group Method of Arithmetic Average.

* Corresponding author at: College of Fisheries, Jimei University, Xiamen 361021, PR China.
E-mail address: ctxie@jmu.edu.cn (C. Xie).

http://dx.doi.org/10.1016/j.aquaculture.2015.11.016
0044-8486/

McCouch, 1997). Many countries and organizations have created hundreds of germplasm collections, and millions of crop resources have been preserved (Van Hintum et al., 2000). However, with continuous collection of germplasm resources, the sizes of collections have been becoming larger; thus, the management, evaluation and use of large germplasm collections is expensive and inefficient due to redundancies and/or duplications, as well as difficulty in detailed analysis of all of the conserved accessions (Brown, 1989). To solve this problem, Frankel (1984) initiated the concept of core collections, defined as a representative sample of an entire collection of germplasm resources with minimal repetitiveness and maximal genetic diversity of a given species and its relatives. The core collection serves as a working collection, which can be extensively examined, with accessions excluded from it being retained as a reserve collection (Brown, 1989). The core collection is more easily studied, more conducive to utilization of germplasm resources and aids in the management of the entire collection. Since Brown et al. (1989) developed the first core collection from Australian collections of perennial *Glycine* spp., dozens of core collections have been successfully developed, having very large to relatively small numbers of accessions, such as *Arabidopsis thaliana* (McKhann et al., 2004), wild apple (Richards et al., 2009; Volk et al., 2005), pear (Miranda et al., 2010) and rice (Li et al., 2011). To date, no core collections of mariculture crops have been constructed.

Since the concept of the core collection was developed, various data have been used to analyze genetic diversity in crops, including phenotypic traits, biochemical data and DNA molecular markers (Odong et al., 2013). Most phenotypic traits of crop varieties are polygenic quantitative traits, which are easily affected by environmental conditions and experimental errors. Moreover, interactions between genes and the environment affect phenotypic traits. Therefore, genetic sorting based on phenotypic data cannot correctly reflect the genetic diversity of the initial germplasm resources. The same phenotype can be found in different genotypes, and accessions with a similar phenotype may sometimes be evolutionarily unrelated (Odong et al., 2013). Biochemical markers are disadvantaged by their bulk and difficulty in accurate measurement; thus it is not realistic to obtain accurate intact data of an entire collection, or even a large fraction of it (Hu et al., 2000). Because DNA markers such as RFLPs and SSRs are easily obtained, have a higher polymorphism and can reflect direct changes at the DNA sequence level independent of environment and plant growth stage, core collections based on DNA markers data can be measured more accurately and have better representation of the initial collection than those based on phenotypic traits or biochemical data; therefore DNA markers are now becoming an ideal tool for the development of core collections (Van Treuren et al., 2006).

By definition, a core collection should avoid identical or near-identical accessions, and contain as much diversity as possible. Therefore, an appropriate and efficient sampling strategy is critical during the establishment of core collections (Odong et al., 2013). To validate the representativeness of core accessions, appropriate parameters are required to evaluate representative core accessions in order to ensure genetic diversity (Wang et al., 2007a). There are now a number of sampling strategies and evaluative parameters that have been introduced for developing core collections; however, the optimal sampling strategy and evaluative parameters can be different for each species (Van Hintum et al., 2000). Therefore, selecting an optimal sampling strategy and a series of appropriate evaluative parameters are important aspects of research into core collection construction.

The coastal area of Fujian province of China, where *P. haitanensis* originated, has the richest resources of *P. haitanensis* of both wild and cultivated varieties. Through the efforts of the past ten years, one germplasm collection of *P. haitanensis* has been created at the Jimei University of China and contained more than 500 accessions (Xie et al., 2010). To better evaluate and utilize the germplasm of *P. haitanensis*, the main goal of this work was to establish a procedure to develop a core collection of *P. haitanensis* based on SSR markers.

2. Materials and methods

2.1. Plant material

For this study, 77 germplasm resources from the germplasm collection of *P. haitanensis* were used as an initial collection. The 77 germplasm strains included cultivars collected from coastal areas of Fujian provinces of China and some strains selected by wild breeding, mutation breeding or cross breeding (Additional file 1). All of the germplasm strains were conserved as free-living conchocelis at the Jimei University of China. The free-living conchocelis were cultured in natural seawater with Provasoli's enrichment solution (PES) medium at 21 °C and were illuminated by 30 μmol photons m^{-2} s^{-1} (10:14, L:D). The culture medium was refreshed every 15 days.

2.2. Molecular characterization of the initial collection

2.2.1. DNA extraction

DNA was isolated from free-living conchocelis of each germplasm strain. The collected free-living conchocelis were ground into powder with a high-speed homogenizer, and DNA was extracted and purified by the Cetyltrimethyl Ammonium Bromide (CTAB) method (Sambrook and Russell, 2001). DNA concentrations were determined with a DU-600 spectrophotometer (Beckman Coulter, Fullerton, CA, USA) and adjusted to 5 ng/μL for PCR amplification.

SSR markers analysis A set of 100 different primer combinations of SSR, which were developed from transcriptome data of *P. haitanensis* (Xie et al., 2013), were first tested in a representative sample of 10 cultivars, and 30 different primer combinations (Additional file 2). This approach can obtain highly polymorphic, specific, and reproducible banks for genotyping accessions from the entire initial collection. The PCR amplifications for SSR analysis were performed in a 25 μL PCR reaction mixture containing 2.5 μL of 10 × PCR Buffer, 5 ng of genomic DNA, 1.0 U of Taq polymerase (Takara Biotechnology (Dalian, China) Co., Ltd.), 0.2 M of forward primer (Takara), 0.2 M of reverse primer (Takara), and 200 μM of dNTP (Takara). The amplifications were performed in a MT programmable thermal controller PTC-200 (MJ Research, Inc.) with a primary denaturation at 95 °C for 5 min, followed by 35 cycles of three steps (40 s of denaturing at 94 °C, 60 s of annealing

Table 1
The core subsets based on the random sampling strategy.

Sample sizes	ND	SM	JD
30%	1, 8, 9, 10, 11, 13, 17, 18, 22, 23, 32, 33, 39, 42, 43, 47, 50, 51, 63, 68, 70, 71, 77	1, 2, 6, 8, 10, 13, 15, 19, 29, 30, 32, 33, 36, 39, 44, 45, 52, 59, 63, 67, 70, 73, 77	1, 3, 6, 7, 9, 10, 13, 15, 17, 29, 31, 36, 39, 42, 43, 47, 49, 56, 59, 63, 68, 75, 76
25%	1, 8, 9, 10, 11, 17, 18, 23, 32, 39, 42, 47, 50, 51, 63, 68, 71, 70, 77	1, 2, 6, 8, 10, 13, 15, 19, 29, 33, 36, 45, 55, 59, 63, 67, 70, 73, 77	1, 3, 7, 9, 10, 15, 17, 29, 36, 39, 42, 43, 47, 49, 59, 63, 68, 75, 76
20%	1, 9, 10, 11, 17, 18, 23, 32, 39, 42, 47, 50, 63, 68, 77	1, 6, 8, 10, 13, 19, 29, 36, 45, 55, 59, 63, 67, 73, 77	1, 3, 7, 9, 10, 15, 17, 29, 36, 42, 43, 47, 59, 63, 75
15%	1, 9, 10, 11, 17, 23, 32, 39, 42, 50, 63, 77	1, 6, 8, 19, 29, 36, 45, 59, 63, 67, 73, 77	1, 3, 9, 10, 15, 17, 36, 42, 43, 47, 59, 63
10%	1, 11, 18, 23, 32, 39, 50, 77	1, 6, 8, 29, 45, 59, 63, 73, 77	1, 3, 10, 15, 36, 42, 43, 63

ND: Genetic distance using Nei & Li genetic similarity coefficient, SM: Genetic distance using simple matching coefficient, JD: Genetic distance using Jaccard genetic similarity coefficient.