Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 363 (2005) 24–31
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Genetic linkage map of medaka with polymerase chain reaction length polymorphisms

Tetsuaki Kimura [a], Keiko Yoshida [b], Atsuko Shimada [c], Tomoko Jindo [c], Mitsuru Sakaizumi [d], Hiroshi Mitani [e], Kiyoshi Naruse [c], Hiroyuki Takeda [c], Hidetoshi Inoko [a], Gen Tamiya [a], Minori Shinya [a,f,*]

[a] *Department of Molecular Life Science, Course of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Kanagawa 259-1193, Japan*
[b] *Japan Biological Information Research Center, Japan Biological Informatics Consortium, Koutou-ku, Tokyo 135-0064, Japan*
[c] *Department of Biological Sciences, Graduate School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan*
[d] *Department of Environmental Science, Faculty of Science, Niigata University, Niigata 950-2181, Japan*
[e] *Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba 277-8562, Japan*
[f] *Genetic Strains Research Center, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan*

Received 22 March 2005; received in revised form 15 June 2005; accepted 22 July 2005
Available online 14 October 2005
Received by T. Gojobori

Abstract

With recent improvements in genetic and genomic infrastructures, great interest has been taken in genetic dissection of multi-factorial traits. A genetic map consisting of markers that are highly polymorphic and rapidly genotyped is essential for the genetic mapping of such a complex trait. Medaka, *Oryzias latipes*, is an excellent model system for genetic studies. To promote genetic mapping of complex traits in medaka we developed the first high-throughput and genome-wide marker set in the organism by using its genomic information and the bioinformatic techniques. We tested 545 primer pairs and obtained 265 co-dominant markers between two inbred strains, HNI and Hd-rR. Our map, consisting of 231 uniquely mapped markers, covers 1257.3 centimorgan (cM) of the medaka genome with an average interval distance of 5.4 cM. Furthermore, the newly designed markers were examined for polymorphisms among six medaka inbred strains: HNI, Hd-rR and four additional strains. Most of our markers are simple sequence length polymorphisms (SSLPs) and can be rapidly genotyped by an automated system under a single polymerase chain reaction (PCR) condition. Together with the genotyping data of six medaka inbred strains, our new marker set provides a powerful tool for genome-wide analysis of complex biological phenomena found widely in medaka populations.

Keywords: *Oryzias latipes*; SSLP; Microsatellites; Genetic mapping

1. Introduction

The identification of genes underlying complex traits has become one of the major challenges of current genetic studies. However, both genetic heterogeneity in outbred populations and the significant impact of environmental factors on resulting phenotypes complicate genetic studies of multi-factorial traits. To control these effects and increase the ability to detect susceptible loci, it is effective to use model organisms, particularly inbred strains.

The medaka is a small, egg-laying freshwater fish found in eastern Asia and is suitable for use in genetic studies of both

Abbreviations: cM, Centimorgan; ESTs, Expressed sequence tags; LGs, Linkage groups; MEXT, The Ministry of Education, Culture, Sports, Science and Technology of Japan; NIG, The National Institute of Genetics; PCR, Polymerase chain reaction; PLPs, PCR length polymorphisms; RFLPs, Restriction fragment length polymorphisms; SSLPs, Simple sequence length polymorphisms; WGS, Whole-genome shotgun.

* Corresponding author. Genetic Strains Research Center, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan. Tel./fax: +81 55 981 5849.
E-mail address: mshinya@lab.nig.ac.jp (M. Shinya).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.07.043

Mendelian and complex traits because of its short life cycle, high fertility and the availability of several inbred strains (reviewed in Naruse et al., 2004b). Mutagenesis screens by *N*-ethyl-*N*-nitrosourea have been performed and hundreds of medaka mutants have now been identified in several laboratories (Ishikawa, 2000; Loosli et al., 2000; Furutani-Seiki et al., 2004). The corresponding genes for some of mutants have also now been cloned (Fukamachi et al., 2001; Loosli et al., 2001). Furthermore, genomic information on medaka is still growing rapidly (Kimura et al., 2004; Naruse et al., 2004b).

A marker set with a sufficient density to cover the entire genome is an essential tool for genetic analysis of particular organisms, especially those with complex traits. The first genetic linkage map to cover the whole medaka genome was established using random amplification of polymorphic DNAs, fingerprinting and allozyme analysis (Wada et al., 1995). More recently, several further linkage maps have been developed in medaka (Ohtsuka et al., 1999; Naruse et al., 2000, 2004a), of which the most extensive is a map based on restriction fragment length polymorphisms (RFLPs) that has now become the accepted standard (Naruse et al., 2004a). Notwithstanding the great utility of RFLPs, they still have several major limitations. First, they are mostly bi-allelic and show a lower rate of polymorphism among some inbred strains, making it difficult to analyze crosses between such strains. Second, typing RFLPs is time-consuming and difficult to automate. Furthermore, most RFLPs reported by Naruse et al. (2004a) were designed from expressed sequence tags (ESTs), so that the markers are not expected to distribute evenly. Therefore, despite having the highest density, it is difficult to utilize this map for rapid analysis of the whole medaka genome.

Polymerase chain reaction (PCR) length polymorphisms (PLPs) are very effective markers for analyzing complex traits because they can be directly assayed by the differences in PCR fragment lengths and are amenable to the development of high-throughput genotyping. Simple sequence length polymorphisms (SSLPs), one of the PLP subtypes, consist of short tandem repeats whose units are generally 2-6 bp in length. They are known to distribute somewhat uniformly throughout the genome and to allow for a greater information content per locus (reviewed in Bruford and Wayne, 1993). Because of these advantages, SSLPs have been used for comprehensive linkage maps in human (Dib et al., 1996), mouse (Dietrich et al., 1992) and many other organisms (Jacob et al., 1995; Knapik et al., 1998).

In the present study, we describe the first construction of a genetic linkage map of the medaka using PLP markers (mostly SSLPs), all of which can be genotyped in an automated genotyping system. A total of 265 co-dominant markers between HNI and Hd-rR inbred strains (Hyodo-Taguchi, 1996) were mapped with 368 informative meioses, resulting in a map with a genetic length of 1257.3 centimorgan (cM), 231 uniquely mapped markers and an average marker distance of 5.4 cM. Our new map will thus facilitate whole-genome scans as a first step to elucidate loci responsible for complex traits in medaka.

2. Materials and methods

2.1. Strains and genetic crosses

The AA2 (Shimada and Shima, 1998), Cab (Loosli et al., 2000), and Hd-rR strains are inbred strains established from a southern Japanese population. The HNI and Kaga (Naruse et al., 2004b) strains are inbred strains from a northern Japanese population. The HSOK (Hyodo-Taguchi, 1996) strain is an inbred strain generated from an east Korean population. Kunming (Takehana et al., 2004) is a wild stock derived from the China–west Korean population. Fish were maintained in an in-house facility in a constant re-circulating system at 26 °C on a 14 h light and 10 h dark cycle. A total of 184 F_2 progeny were obtained by intercrossing F_1 (HNI × Hd-rR), 48 F_2 progeny were obtained by intercrossing F_1 (Kaga × Hd-rR) and 44 backcross progeny were obtained from the cross of Hd-rR female and (Hd-rR × Kunming) F_1 male and the cross of (Hd-rR × Kunming) F_1 female and Hd-rR male.

2.2. Design of primer pairs

Medaka genome sequences were obtained from whole-genome shotgun (WGS) sequences of the Hd-rR genome produced by the Academia Sequencing Center of the National Institute of Genetics (NIG) in Mishima, Japan and the RIKEN Institute under the support of the National Bioresource Project of the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) (available at http://shigen.lab.nig.ac.jp/medaka/genome/top.jsp). DNA sequences with quality values of ≥20 assigned by the base-call program *phred* (Ewing et al., 1998; Ewing and Green, 1998), were analyzed with *Sputnik* to detect CA repeats (available at http://espressosoftware.com/pages/sputnik.jsp). By using *PrimerExpress* software (Applied Biosystems, Foster City, CA), we designed and selected primer pairs of 18–23 mer each, which generate a product of 200–300 bp with $(CA)_{8–20}$ and whose predicted melting temperatures are in the range of 57–58.5 °C and with differences of ≤0.5 °C between pair. In addition, using the e-PCR program (available at http://www.ncbi.nlm.nih.gov/sutils/e-pcr/), the selected primer pairs were verified to amplify a single copy on the WGS. A total of 379 primer pairs were synthesized commercially (GenSet KK, Kyoto, Japan).

To fill the 33 large gaps, we first found markers from Naruse map (Naruse et al., 2004a) that had been placed in the target regions. As most markers in the Naruse map originated from ESTs, we could obtain EST sequences from their marker names. Using BLASTN, we compared the EST sequences with the scaffold sequences assembled from 9-fold genome coverage sequences of Hd-rR produced in the medaka genome project carried our at the Academia Sequencing Center of the NIG in Mishima, Japan under the support of the Grants-in-Aid for Scientific Research in Priority Area ‘Genome Science’ from the MEXT (available at http://dolphin.lab.nig.ac.jp/medaka/index.php). The scaffold sequences closely linked to the EST sequences were analyzed with *Sputnik* to find CA repeats ≥5. Primer pairs flanking the repeats were designed