論 文 の 内 容 の 要 旨

水圏生物科学専攻
平成 21 年度博士課程入学
氏　　名　陳　　盈光
指導教員名　浅川　修一

Genetic studies on the upper temperature tolerance of rainbow trout
（ニジマスの高温耐性に関する遺伝学的研究）

Rainbow trout (*Oncorhynchus mykiss*) is a cold-water aquaculture species which is of a considerable economic importance. It is originally a native fish in the North America. It became one of the most popular aquaculture species following the introduction to all continents since 1874. The global annual production of rainbow trout was almost tripled from 259,161 tons in 1989 to 732,432 tons in 2009 (FAO, 2009).

The first introduction of rainbow trout to Japan was in 1877, and its mass production started in 1950s (Matsuda, 1992), when a strain named Donaldson established by Dr. Donaldson in the University of Washington (Donaldson et al., 1957) was introduced to Japan. The Donaldson strain showed several advantages such as increased growth rate, disease resistance and enhanced egg production. Since 1966, Miyazaki Prefectural Fisheries Experimental Station developed a thermally selected strain of rainbow trout based on the Donaldson strain by traditional selective breeding. This strain acquired a degree of upper temperature tolerance as revealed by the fact that the strain grows normally and also feeds actively at 24°C, in contrast to the optimum temperature of water for the normal strain below 20°C. Additionally, the thermally selected strain survived from occasional exposure to heated water at 30~35°C for 1 to 5 min, which is much higher than the upper limit of normal survival water temperature around 24°C (Ineno et al., 1993). A related research indicated that mitochondrial cytochrome *c* oxidase subunit II (COXII) gene was associated with upper temperature tolerance in eggs and embryos (Ikeguchi et al., 2005). However, the molecular mechanism involved in this thermal tolerance has been still unknown.

The objective of the present study was to investigate genes related to upper temperature tolerance in rainbow trout at juvenile and adult stages. First, the expression levels of COXII gene in five different tissues from juveniles were compared between normal and thermally selected strains. Next, thermally selected diploid rainbow trout genome and gynogenetic double haploid genome of the same strain were sequenced by next-generation sequencing (NGS) technologies. Genome-wide microsatellite candidates were exploited with the sequences of diploid genome of thermally selected strain by a GS-FLX sequencer (Roche). Then, genome *de novo* assembly of thermally selected gynogenetic rainbow trout was conducted using Genomic Analyzer II (GAII) and HiSeq 2000 sequencing systems (Illumina).

Moreover, the comparative transcriptome analysis before and after exposure to high temperatures were performed using various tissues from both normal and thermally selected strains to examine the genome-wide gene expression patterns at upper temperatures.

1. COXII mRNA expression analysis

The different expression levels of COXII mRNA between thermally selected and normal strains were analyzed by real-time PCR using brain, heart, liver, muscle, and skin tissues from one year-old juveniles (length, 9.9-13.9 cm; weight,13.2-31.0 g). After triplicated real-time PCR measurements, the expression levels were calculated by the ΔΔCT method with elongation factor-1 alpha gene as an internal standard. When the the data between the two strains were analyzed by student's t-test, no significant differences were found in all tissues examined.

2. Generation of gynogenetic double haplotype fish

In order to generate high quality genomic sequences of rainbow trout, we attempted to use a double haploid fish as a genomic DNA source. For this purpose, a gynogenetic rainbow trout offspring (length, 2.1 cm; weight, 1.3 g) from the thermally selected strain was produced by the fertilization with UV-irradiated masou salmon (*Oncorhynchus masou*) sperms followed by the blockage of the first egg cleavage. Prior to Illumina GAII and HiSeq 2000 sequencing, microsatellite genotypes of the gynogenetic sample were verified. Genomic DNAs of the gynogenetic juvenile and three normal diploid fish were examined using 16 polymorphic microsatellites which were selected from a published linkage map (Rexroad et al., 2008). These microsatellite markers were amplified by PCR, and their genotypes were examined by fragment analysis using an ABI 3100 sequencer and Peak scanner software. As a result, all 16 microsatellites were homozygous for the gynogenetic fish, while most of them in the three diploid fish were heterozygous, indicating that the gynogenetic fish had the double haploid genome.

3. Next-generation sequencing

The rainbow trout genome was sequenced by 3 different next-generation sequencing platforms, GS-FLX, GAII and HiSeq 2000 sequencers. For the GS-FLX sequencing, genomic DNA was extracted from pelvic fin tissues of a three-year-old female of the thermally selected diploid strain. It was used as a genomic reference for genome-wide genetic marker discovery. Five rounds of the 454 GS-FLX sequencing generated 4,634,401single-end reads with total length of 1,531,336,345 bp (average read length, 330bp). Quality filtering (quality score, 20<) and size selection (discarding the sequences of < 100 bp) for the raw reads were performed using FASTX-toolkit software. After the removal of duplicated sequences by CD-HIT, 1.4 Gb of high quality sequences were obtained using a GS-FLX sequencer.

The genomic DNA of the gynogenetic fish (length, 2.1 cm; weight, 1.3 g) was prepared for sequencing with Illumina sequencers. An Illumina GAII sequencer produced 43,722,295,364 bp paired-end data (insert size, 400 bp and 700 bp; read length, 101 bp). Another 201,354,918,016 bp

mate-pair reads were generated by an Illumina Hiseq 2000 sequencer (insert size, 2kb and 5kb; read length, 76 bp). Quality filtering and low quality base masking (quality score, 20; the base with a quality score lower than 20 was masked by N) for the data were performed by FASTX-toolkit software.

In total, 246.4 Gb sequence data were obtained by the three sequencing platforms, which covered 102.7x of the estimated rainbow trout genome size, 2.4 Gb, which will be described later in more details.

4. Massive genetic marker discovery

To discover the microsatellite genetic markers on the rainbow trout genome, the sequenced reads using a GS-FLX sequencer were loaded into MIcroSAtellite identification tool (MISA) to search for the potential microsatellite candidates. The microsatellite candidates met the requirement of repeat unit (2~8 bp) over 3 times and had the minimum length of 20 bp. Total 215,024 microsatellites were detected, which accounted for 0.65% of the 1.4 Gb rainbow trout genomic sequences. Dinucleotide microsatellite candidates that accounted for 70.7% of all discovered candidates were dominated among all kinds of repeats, which were followed by tetra- (12%) and pentanucleotide (5.7%) microsatellite candidates. The primers of all candidates that had sufficient flanking regions were designed by the integration of MISA and Primer3.

5. Genome *de novo* assembly

The accurate estimation of the rainbow trout genome size is preferable for the following *de novo* assembly step. To evaluate the genome size, k-mer distribution histogram was useful. As k-mers represent k nucleotides generated from sequencing reads iteratively, the occurrence of k-mers reflects the sequencing depth. Thus, the actual sequencing coverage is observed in the k-mer distribution histogram. Generations of k-mers (k = 20, 30) were carried out and, then, their frequencies were calculated by Jellyfish software. The peak frequencies of 20-mer and 30-mer for the sequencing reads of the gynogenetic fish were estimated to be 6.7 and 5.4, respectively. The rate of sequencing errors was also estimated from the histogram. The genome size was calculated with the total number of k-mers, the peak frequencies of k-mer, and the error rate. As a result, the estimated genome size of rainbow trout was about 2.4 Gb.

All of pair-end (400 bp and 700 bp insert size) and part of mate-pair (2kb insert size) sequencing data from the gynogenetic fish were used to perform genome *de novo* assembling using SOAPdenovo software. In order to optimize settings of *de novo* assembly, k-mers of 51~71 were selected to generate contigs by de Bruijn graph algorithm. Subsequently, the distance information of all pair-end and mate-pair reads (400 bp, 700 bp and 2kb) were loaded into Soapdenovo to perform scaffolding. These processes were performed step by step from short (400 bp) to long (2kb) sizes.

When the k-mer of 59 was employed, the longest total contig size of 1,861,100,116 bp was obtained, which accounted for 77.5% of the genome size of rainbow trout, whereas the 51-mer showed the highest coverage of scaffold (2,239,632,961 bp, accounting for 93%). The N50 lengths of contigs

and scaffolds of this assembly were 577 bp and 15,046 bp, respectively. There were still about 0.5 Gb gaps in the assembled sequences.

6. *De novo* transcriptome assembly of normal and thermally selected strains

Total RNAs were extracted from brain, heart, liver, muscle and gill of both normal and thermally selected strains before and after exposure to high temperature at 26°C for 30min. cDNA libraries were constructed using these total RNAs and subsequently sequenced with a HiSeq 2000 sequencer using the pair-end method with 150 bp insert size. The total sequence lengths of the samples ranged from 8,081,204,488 bp (gill tissues of thermally selected strain after exposure to high temperatures) to 11,699,637,448 bp (muscle tissues of the normal strain before exposure to high temperatures). To obtain high quality sequences, quality selections to filter out low quality reads were performed. For each sample, 50M reads data were randomly selected to perform further analysis. These selected 50 M paired sequences were applied to Velvet and Oases software packages to generate the contigs of transcripts. Among five different tissues, brain tissues produced the most varieties of contigs of transcripts : 55k, 52k, 52k, and 49k contigs of 1000-bp or more in size were obtained in normal-before, normal-after, thermally-selected-before, and thermally-selected-after fish, respectively. On the contrary, the numbers of contigs from muscle tissues were 13k, 11k, 12 and 11k, respectively.

The homology search of these *de novo* assembled transcripts was conducted using BLASTX (E-value = 1E-10) against the NCBI protein database. Then, they were annotated by Blast2Go to assign gene ontology (GO) terms and were mapped on the known KEGG pathway. There were also found many novel genes that showed low similarity with known genes in the NCBI database. Consequently, detailed expression profiles of annotated and unannotated genes for each tissue were constructed by mapping quality-filtered raw reads back on these *de novo* assembled transcripts using Bowtie software. Liver tissues showed 673 known genes that exhibited at least two-fold changes of expression after exposure to high temperatures, including many heat shock protein family genes and BAG family protein regulator genes. These results are very useful to find genes responsible for upper temperature tolerance of rainbow trout.

7. Conclusions

In this study, both genomic and transcripome analyses of the thermally selected rainbow trout were investigated. For the survey of COXII mRNA expression, no significant differences were observed in brain, heart, liver, muscle and skin tissues between the normal and thermally selected juveniles in contrast to the significant expressions observed in embryos at 2 to 4 cell stages of the thermally selected strain (Ikeguchi et al., 2005). The discovered microsatellite candidates will be helpful to assist future marker-assist selections, because there has been still no useful genome reference of rainbow trout. Thus, the newly assembled draft genome sequence of rainbow trout in this study can be utilized as a reference to assist further genetic and genomic studies of rainbow trout. By the comparative gene expression analysis, lists of candidate genes for the thermal tolerance such as heat

shock protein 90 alpha gene and DNAj homolog subfamily b member gene were presented. Future studies based on these results will provide more detailed information for the genes and mutations that are responsible for upper temperature tolerance in fish.