論 文 の 内 容 の 要 旨

水圏生物科学　専　攻
平成 22 年度博士課程　入学
氏　　名　張（ザン）　虹（ホン）
指導教員名　浅川　修一

論文題目　Utilization of a doubled-haploid individual in the generation of a high-quality assembly of the torafugu (*Takifugu rubripes*) genome
（ダブルハプロイド個体を活用した高精度トラフグゲノムアセンブリの作成）

The compact genome of torafugu (*Takifugu rubripes*) (~400 Mb) is a useful model for annotating the vertebrate genomes. However, the currently incomplete assembly of the torafugu genome with 28% of scaffolds, in terms of chromosomal assignment and orientation unknown, has decreased its value for various genetic and genomic analyses. This is partly due to the usage of a natural heterozygous male torafugu as a starting material. As the heterozyogus genome possesses polymorphic sequences or might even contain copy number variations (CNV) between homologous chromosomes, the construction of accurate genome assembling has become extremely difficult.

1. Induction of mitotic gynogenesis in torafugu

For the purpose of reducing the polymorphism levels and improving the quality in genome assembling, we created homozygous torafugus through artificial induction of mitotic gynogenesis. First, diluted milt from 2 males was irradiated with 3 different UV dosages (40, 80 and 160 mJ/cm^2) for a complete inactivation of

sperm genetic materials. For each UV treatment, 150 g of eggs from 1 female were fertilized with 15 mL of irradiated milt mixed from 2 males, followed by an incubation in fresh seawater at 18.0°C. At 180 min post-fertilization, cold shock was applied to approximately 140 g of eggs in each UV treatment by soaking eggs in the icy cold seawater for 45 min. The remaining 10 g of eggs were served as haploid controls. Totally, we generated 7 groups named DG_1 for the diploid gynogenesis and HC_1 for the haploid control treated with a 40 mJ/cm^2 UV dosage; named DG_2 for the diploid gynogenesis and HC_2 for the haploid control treated with a 80 mJ/cm^2 UV dosage; named DG_3 for the diploid gynogenesis and HC_3 for the haploid control treated with a 160 mJ/cm^2 UV dosage; and named NC for the normal control. Among all 3 DG groups, the highest development rate and hatching rate of 28.2% and 4.0% were achieved in the DG_1 group. No fry survived to hatch in all 3 HC groups, meanwhile a microscopic analysis showed that the embryos had typical haploid syndromes, indicating the complete activation of paternal genetic materials by UV treatments.

2. Homozygosity analyses by microsatellite genotyping

Microsatellite genotyping was employed to assess the homozygosity levels of 16 mito-gynogenetic fry randomly sampled from the DG_1 and DG_2 groups (8 fry from either the group). Among the 156 microsatellite loci genotyped in 2 males and 1 female torafugus, a total of 56 microsatellite loci, sharing no common allele between the male and the female were selected to genotype the gynogenetic offspring. Each genotyping was performed by a microsatellite markers of post-PCR fluorescent labeling method and analyzed by capillary electrophoresis in a genetic analyzer. Through comparing the location of genotype peaks in the mito-gynogenetic or the normal offspring with those in the parents, the inheritance patterns clearly revealed that there was no genetic contribution from the paternal genomes to 16 mito-gynogenetic fry. As a result, the homozygosity levels of 16 mito-gynogenetic fry reached 100% at 56 unambiguous microsatellite loci, with an average density of 2-3 loci for every chromosome. The first instance of mitotic gynogenesis induction in torafugu has been successfully established.

3. Deep sequencing of torafugu genomes

Subsequently, we performed deep sequencing of a wild-type and a 5-month old mito-gynogenetic torafugu genomes using 2 Illumina next-generation sequencing (NGS) platforms (Illumina GA IIx and Hiseq 2000). For the wild-type torafugu, the

Illumina GA IIx produced 84,857,156 reads in 101-bp length for 1 paired-end (PE) library of a 400-bp insert size while the Illumina Hiseq 2000 produced 278,642,344 and 244,796,700 reads in 76-bp length for 2 mate pair (MP) libraries of 2-Kb and 5-Kb insert sizes. For the mito-gynogenetic torafugu, the Illumina Hiseq 2000 produced 283,351,680 and 253,179,572 reads for 2 PE libraries of 300-bp and 500-bp insert sizes as well as 248,467,078 and 339,507,094 reads for 2 MP libraries of 2-Kb and 5-Kb insert sizes. The sum of base calls possessed physical coverages of approximately 120× and 280× to the theoritical genome size (~400 Mb) for the wild-type and the mito-gynogenetic torafugu, respecitively.

4. Genome-wide homozygosity analyses

Since the microsatellite genotyping at limited loci failed to reveal the whole-genome homozygosity levels of the mito-gynogenetic torafugu, our research established a genome-wide single nucleotide polymorphism (SNP) genotyping method based on the NGS data. After the pre-processing to raw reads, a total of 74,235,872 high-quality and de-duplicated reads from either the wild-type or the mito-gynogenetic torafugu DNA libraries were mapped against the reference sequences (the fifth fugu genome assemblies) with 1 mismatch allowed by Bowtie. The potential SNPs (pSNPs) between the reference and the mapped NGS reads (known as inter-pSNPs) were called with variant frequency of 100% at 26-34× depth of coverage by the CLC genomics workbench. A total of 39,499 inter-pSNPs were detected in the wild-type torafugu genome while a total of 93,525 inter-pSNPs were detected in the mito-gynogenetic torafugu genome. These 2 individuals shared 8,609 inter-SNPs in common. Among the remaining 30,890 (39,499-8,609) and 84,916 (93,525-8,609) inter-pSNPs in 2 individuals, the SNPs between homologous chromosomes within 1 individual (known as intra-SNPs) were called at the variant frequency ranged from 40%-60% at 26-34× depth of coverage as well. As a result, no intra-SNPs was detected in the mito-gynogenetic torafugu genome while 5,621 intra-SNPs were detected in the wild-type torafugu genome. Thus, it has been confirmed that the mito-gynogenetic torafugu was a doubled-haploid individual.

5. *De novo* assembly of torafugu genomes

According to the results of genome-wide SNP analyses, we estimated the effect of homozygosity levels on genome assembling by importing pre-processed reads containing nearly the same number of bases from the wild-type (5,880,348,666 bp) and the doubled-haploid (5,880,348,674 bp) torafugu DNA libraries into the

SOAPdenovo genome assembler with the same parameter settings. The N50 size and the maximum length of contigs constructed from the reads of the doubled-haploid torafugu libraries were increased by more than 5 times and 2 times compared with the values of the wild-type torafugu. This has strongly suggested that different genome-wide homozygosity levels had great effects on genome assembling and the utilization of a doubled-haploid genome has improved the assembling quality by decreasing the polymorphic pathways of *de Bruijn* graphs during k-mer overlapping connection.

To achieve a good-quality assembling of the doubled-haploid torafugu genome, we employed 3 genome assemblers (SOAPdenovo, IDBA-UD and an embedded *de novo* assembler of the CLC genomics workbench) to investigate their different performance. The setting of multiple k-mer values for SOAPdenovo was ranged from 63-77, with a gradient of 2-mers every assembling while for the CLC genomics workbench, k-mer values of 53, 57, 61 and 64 were assigned. For IDBA-UD, genome assembling was performed under a gradient k-mer values ranged from 31-81, with a step addition of 25-mer. Among 3 assemblers, SOAPdenovo generated the largest contig of 75,131 bp and scaffold of 5,484,273 bp under the k-mer value of 71 and 73, respectively. After removal of potential sequencing errors with the k-mers of single-occurrence, the torafugu genome was assembled into 185,868 scaffolds containing estimated 380,443,322 bp of residues with a maximum scaffold size of 4,727,828 bp by SOAPdenovo. A further gap closure to the gap regions of the scaffolds increased the number of residues without Ns to 361,764,858 bp, and closed 47,203 out of 130,146 gaps. Finally, an alignment of the largest scaffold against 22 sequences of the fifth fugu genome assembly, showed a highly-identical match with a sequence on chromosome No. 9. Thus, the accuracy of *de novo* assembled torafugu sequences have been proved.

6. Conclusions

We confirmed the complete homozygosity of the mito-gynogenetic torafugu by microsatellite genotyping and genome-wide SNP analyses. With the utilization of the doubled-haploid torafugu as a starting material, we obtained an approximately 378 Mb of sequence assembly for the torafugu genome with an estimated coverage depth of 50× from the Illumina paired-read sequences in less than 5 months. The quality of the torafugu genome assembled from short reads was proved to be valid.