Genetic Analysis of Vertebral Regionalization and Number in Medaka (Oryzias latipes) Inbred Lines

Tetsuaki Kimura,*1 Minori Shinya,1 and Kiyosi Naruse*†

*Laboratory of Bioresources, National Institute for Basic Biology, Okazaki 444-8585, Aichi, Japan, †Genetic Strains Research Center, National Institute of Genetics, Mishima 411-8540, Japan, and ‡Department of Basic Biology, School of Life Science, Graduate University for Advanced Studies (SOKENDAI), Kanagawa 240-0193, Japan

ABSTRACT Vertebral number is the most variable trait among vertebrates. In addition to the vertebral number, the ratio of abdominal to caudal vertebrae is a variable trait. The vertebral number and the ratio of abdominal to caudal vertebrae contribute to vertebrate diversity. It is very interesting to know how to determine the vertebral number and the ratio of abdominal to caudal vertebrae. In this study, we identify differences in the vertebral number and the ratio of abdominal vertebrae to vertebral number between two inbred lines of medaka, namely, Hd-rRII1 and Kaga. To identify the genetic factor of those differences, we performed quantitative trait locus (QTL) analysis for vertebral number and the ratio of abdominal vertebrae to vertebral number using 200 F2 fish. Our results show a suggestive QTL of the ratio of abdominal vertebrae to vertebral number on chromosome 15, and five QTL of vertebral number on chromosomes 1, 10, 11, 17, and 23. The QTL on chromosome 15 contains hoxDb cluster genes. The QTL of vertebral number include some genes related to the segmentation clock and axial elongation. In addition, we show that the difference in vertebral number between two inbred lines is derived from differences in the anteroposterior length of somites. Our results emphasize that the developmental process should be considered in genetic analyses for vertebral number.

KEYWORDS vertebral number abdominal vertebrae caudal vertebrae genome-wide marker set

Vertebral number is the most variable trait among vertebrates. The adult frog has 6–9 vertebrae, whereas caecilian amphibians have 95–285 vertebrae (Richardson et al. 1998). These variations confer morphological diversity in vertebrates. Therefore, it is developmentally and evolutionarily interesting to elucidate the genetic factors underlying vertebral number. To date, developmental analysis has shown that vertebrae are derived from somites. The resegmentation model describes the vertebrae as structures that are developmentally derived from somites. According to the resegmentation model, each vertebra is formed through recombination of the anterior and posterior halves of two adjacent sclerotomes (Remak 1855; Bagnall et al. 1988; Saga and Takeda 2001). Somites are rhythmically formed from presomitic mesoderm as a periodic pattern along the anteroposterior axis of embryos. This process of somitogenesis is well explained by the clock and wavefront model (Cooke and Zeeman 1976; Dubrulle et al. 2001; Murray et al. 2011). Moreover, developmental analysis has shown that vertebral number is decided by a balance between the pace of the segmentation clock and axial elongation (Gomez et al. 2008; Gomez and Pourquié 2009). The zebrafish mutant hes6, which carries one of the segmentation clock genes, shows decreased vertebral number. Previous pig QTL analysis revealed that NR6A1 and vertmin (VRTN) were underlying vertebral number (Mikawa et al. 2007, 2011), but the role of the genes in the segmental clock and axial elongation was unclear.

In addition to vertebral number, the ratio of abdominal to caudal vertebrae is a variable trait. The vertebral column at least is divided into two parts, namely, abdominal and caudal. Abdominal vertebrae are roughly defined as anterior rib-bearing structures, whereas caudal vertebrae are posterior structures that possess a hemal arch. The ratio of abdominal to caudal vertebrae has also been determined to be species-specific (Richardson et al. 1998; Ward and Brainerd 2007; Gomez and Pourquié 2009). Both the bronze featherback (Notopterus notopterus) and bichir (Polypterus bichir) have a long body, but whereas the bronze featherback has a long tail, the bichir has a long trunk (Ward and Brainerd 2007). In most species, the increase in the total number of vertebrae involves a parallel increase in the abdominal...
and caudal regions (Polly et al. 2001; Ward and Brainerd 2007). In fish, some species show an increase in the number of abdominal vertebrae, whereas others display an increase in the number of caudal vertebrae (Ward and Brainerd 2007). It is now generally accepted that vertebrate anteroposterior axial regionality is specified by the combined expression of the Hox genes known as the Hox cord (Kessel and Gruss 1990; Burke et al. 1995; Carapuço et al. 2005). Moreover, the anterior boundary of hoxd12a can be correlated with the boundary between abdominal and caudal vertebrae in zebrafish (van der Hoeven et al. 1996; Morin-Kensicki et al. 2002). Although hoxd12a is expressed in the boundary, the mechanism involved in defining this boundary is still unclear.

To elucidate genetic factors underlying vertebral number and the ratio of abdominal vertebrae to caudal vertebrae, we performed QTL analysis using medaka, a small, egg-laying freshwater fish from eastern Asia. More than 10 inbred strains of medaka are currently available (Hyodo-Taguchi 1996; Kinoshita et al. 2009). Because an inbred strain is genetically uniform, genetic factors are easily distinguishable from environmental factors. In addition, numerous polymorphisms in the genotype and phenotype of the medaka have been observed among inbred lines. As another advantage, the medaka lays around 20 eggs every morning, and therefore, several hundred progeny from the same parents can be easily obtained for research investigations. The medaka is currently emerging as an important vertebrate model (Wittbrodt et al. 2002; Takeda and Shimada 2010). A high-quality draft of the medaka genomic sequence is available (Kasahara et al. 2007), making it possible to conduct comparative analyses of specific candidate regions with other species and to subsequently narrow down gene regions of interest. Thus, it is ideal to use the medaka as an animal model for the elucidation of genetic factors underlying complex traits (Kimura et al. 2007).

In this study, we have shown differences between the vertebral number and the ratio of abdominal vertebrae to vertebral number between the Hd-rR-III strain and the Kaga strain. We have also performed QTL analyses for those traits by using 200 F2 progenies.

MATERIALS AND METHODS

Fish strains

The vertebral number in fish is influenced by temperature during its ontogeny (Hubbs 1940; Gabriel 1944). We incubated all eggs at 28°C until hatch to minimize the relative effects of temperature. The medaka adults were maintained at 26°C on a 14-hr light/10-hr dark cycle. Hd-rR-III and Kaga strains (Hyodo-Taguchi 1996; Naruse et al. 2004a) are inbred medaka strains established from a southern and a northern Japanese population, respectively (Takehana et al. 2003). Two pairs of Hd-rR-III and Kaga strains were crossed to generate the F1 progeny. Sex combination was reciprocal. A total of 200 F1 progeny were obtained from intercrossing 11 pairs of the F1 progeny.

Skeletal preparation and vertebral count

After four months, fish were killed. We removed the eyeballs and fixed them in 100% ethanol for DNA extraction. Fish were fixed in 3.7% formaldehyde/PBS overnight and washed with water. After the scales were removed, the samples were rinsed with 0.5% potassium hydroxide (KOH) solution and stained with ethanol-saturated 20% alizarin red S/0.5% KOH for 1–2 days. After staining, the samples were washed in water and then transferred to 20% glycerol containing 0.1% hydrogen peroxide to bleach any pigments. The samples were then dehydrated using an ascending glycerol series.

We counted the number of vertebrae in the images of prepared skeletal samples according to a study by Iwamatsu (1997). Abdominal vertebrae were defined as the vertebrae running from the atlas to edge of the rib cage (Figure 1). Caudal vertebrae were defined as structures running from the first hemal spine to the first preural column. The urostyle was excluded in the analysis. The differences between Hd-rR-III strain and Kaga strain were examined using Welch’s t-test. Correlations among the vertebral number, abdominal vertebrae, and caudal vertebrae were calculated using Spearman’s rank correlation coefficients in the R software.

Somite length and trunk length measurement

To measure the anteroposterior length of two to four somites, four to six somite embryos were photographed in 2% methylcellulose using a Leica M125 stereomicroscope equipped with an Olympus DP71 universal camera. To measure the trunk length, an egg envelope from a 7-day-old embryo was removed using a hatchling enzyme and then photographed. Measurement was performed using the ImageJ software.

Genotyping

We designed 147 PCR product length polymorphism (PLP) primers, as described by Kimura and Naruse (2010), for QTL analysis (supporting information, Table S1).

Genomic DNA was extracted from fish eyeballs fixed in 100% ethanol. The samples were suspended in 500 μl of lysis buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 200 μg/ml proteinase K) and incubated at 55°C for 3 hr. Thereafter, the samples were incubated at 95°C for 10 min, after which 500 μl of DNase-free water was added. All genomic DNA samples were stored at −20°C until use.

PCR reactions were performed as described previously (Kimura and Naruse 2010). The protocol was modified to use KAPA Taq Extra DNA polymerase (Kapa Biosystems, Woburn, MA). PCR products were analyzed by using a DNA-500 kit on MCE-202 MultiNA (SHMDZU, Kyoto, Japan). In total, 200 F2 fish were genotyped and used in QTL analysis (File S1).

QTL analysis

A linkage map was constructed by using Mapmaker/EXP 3.0b (Lander et al. 1987). Recombination frequencies were converted to map distance (cM) by using Haldane’s map function. A logarithm of the odds score of 3.0 was used to determine the order of the loci. After the ordering of markers within each linkage group, the RIPPLE command allowed the testing of robustness of the map obtained.

QTL analysis was performed using R/qtl (Broman et al. 2003; Broman and Sen 2009). Nonparametric interval mapping is performed with the scanone function, using the argument “model = np.” The logarithm of odds (LOD) significance threshold for interval mapping was determined by 10,000 iterated permutations. Two-QTL scans were performed using Haley-Knott regression (Haley and Knott 1992). The LOD score significance thresholds for two-QTL scans were determined by using 1000 iterated permutations. The genome-wide significance thresholds were set at α = 0.05. The bayesint function was used to calculate the Bayes credible interval (95%) for interval estimation of QTL location. The fitqtl function was used to calculate the percentage variance explained by the peak marker for each QTL.

RESULTS AND DISCUSSION

Genetic linkage map

A total of 147 PLP markers based on the medaka genome sequence were developed, from which a linkage map was constructed on the
basis of the data generated from 200 F2 progeny. The map had 24 linkage groups (same number as the number of chromosomes), and the total length of the map was estimated at 1493.7 cM, with the average distance between markers calculated as 12.1 cM (Table S1). Our map shows almost the same length as previous maps (Naruse et al. 2004b; Kimura et al. 2005), and thus, we concluded that this map is sufficient for QTL analysis. Because our marker set covers 700 M bases, 1 cM corresponds to 469 K base pairs.

**Vertebral counts**

To identify the genetic factors controlling vertebral number in the medaka, we compared the vertebral number of Hd-rR-II1 with Kaga (Table 1). The significant difference was observed in average of vertebral number using Welch’s t-test ($P = 7.5 \times 10^{-30}$). Hd-r-R-II1 strain has a larger vertebral number than does the Kaga strain. Because previous publications reported that, in medaka, an average of the vertebral number of laboratory strains and wild populations were 30 vertebrae (Ogawa 1971; Ali and Lindsey 1974; Yamahira and Nishida 2009), the vertebral number of Kaga strain was the lowest number of medaka. Because all loci were shown to be homozygous, it was relatively easy to show that the recessive traits observed in the inbred lines were hidden in the wild population and that extreme phenotype could actually appear in the inbred lines. To generate an F1 progeny, we crossed Hd-rR-II1 and Kaga strains reciprocally. The F1 progeny showed almost the same vertebral number as the Hd-r-R-II1 strain, although the mean value shifted downward (Table 1). The vertebral numbers show a greater degree of variation in the F2 progeny than in the three isogenic populations (Hd-rR-II1, Kaga, and F1). The range of the F2 fish traits ranged from those observed in the Kaga strain to those observed in the Hd-r-R-II1 strain (Table 1).

Previous studies using the medaka suggested that the number of abdominal and caudal vertebrae were influenced by different genetic factors (Ali and Lindsey 1974; Yamahira and Nishida 2009; Yamahira et al. 2009; Kiso et al. 2011). And the vertebral number can increase independently in caudal and abdominal regions (Ward and Brainerd 2007). To estimate a relation among the vertebral number, abdominal vertebral number, and caudal vertebral number, we calculated the correlations. The correlation between the number of abdominal and caudal vertebrae was negative in all generations; the correlation coefficients were $-0.40$, $-0.83$, $-0.96$, and $-0.30$ for Hd-r-R-II1, Kaga, F1, and F2, respectively. The correlation between the number of abdominal vertebrae and total vertebral number was not consistent among all the generations; the correlation coefficients were 0.00, 0.28, $-0.06$, and 0.42 for Hd-r-R-II1, Kaga, F1, and F2, respectively. The correlation between the number of caudal vertebrae and total vertebral was positive in all the generations; the correlation coefficients were 0.91, 0.28, 0.33, and 0.71 for Hd-r-R-II1, Kaga, F1, and F2, respectively. In the F2 progeny, the negative correlation (between the number of the abdominal and caudal vertebrae) and the positive correlations (between the total vertebral number and the number of both the abdominal and caudal vertebrae) suggested that although both abdominal and caudal vertebrae were able to contribute to the increase in the number of vertebrae, the caudal vertebrae had contributed more. No traits were correlated with sex in the parental strains.

**QTL analysis of the ratio of abdominal vertebrae to vertebral number**

Because F1 fish have varying numbers of vertebrae, we considered the ratio of abdominal vertebrae to total vertebral number as a trait. We found that the ratio between Hd-r-R-II1 and Kaga strains was significantly different (judged by Welch’s t-test; $P = 0.004$) (Table 1). We then performed QTL analysis to elucidate the genetic factors underlying this trait.

Because the distribution of the phenotype was not normal, we performed both parametric and nonparametric QTL analysis (Kruglyak and Lander 1995) (Figure 2 and Table 2). In the parametric analysis, we identified one significant QTL on chromosome (Chr) 15. The LOD peak of the QTL was 3.65, and the genome-wide LOD score significance threshold was 3.62. In the nonparametric analysis, we were not able to detect significant QTL. The genome-wide LOD score significance threshold was 3.44 ($\alpha = 0.05$) and 3.10 ($\alpha = 0.1$), and the highest LOD score was 3.40 on Chr15. Because this method extends rank-based test statistics (extension of Kruskal-Wallis test), this analysis does not work well in the case of many ties in the phenotypes. Therefore, the power of the nonparametric analysis was decreased in our case. However, as QTL on Chr15 was not detected by nonparametric analysis, we thought it was suggestive QTL. Subsequent analyses were parametric analyses. The Bayes credible interval suggested that the QTL spanned 40 cM, which corresponds to 18.8 Mb, located between markers MID1511 and MID1522. Contrary to the expectations from the parental strain phenotype, the Kaga allele from the QTL was associated with a decrease in the ratio of abdominal to total vertebral. The QTL reflected the 7.9% phenotypic variance observed. We did not detect any QTL.
QTL analysis for total vertebral number

Because the distribution of the phenotype was not normal as the total vertebral number, we performed both parametric and nonparametric QTL analysis. In the parametric analysis, vertebral number, we performed both parametric and nonparametric analyses, respectively. The significance LOD threshold was determined by 10,000 iterated permutations. The black dashed line indicates genome-wide significance for the vertebral number at threshold $\alpha = 0.05$ (LOD = 3.62). The blue dashed line indicates genome-wide significance for the ratio of the number of abdominal vertebrae to total vertebral number at threshold $\alpha = 0.05$ (LOD = 3.44). The black bar indicates Bayes credible interval (95%).
| Trait | Chr | Position LOD Score Threshold | Parametric Nonparametric Parametric Nonparametric Parametric Nonparametric Closest Marker PVE Hd-rR-II1 Heterozygous Kaga |
|-------|-----|-----------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Vertebral number 1 | 38.0 | 39.0 | 4.44 | 4.26 | 3.59 | 3.49 | MID0122 | 2.6 | 28.96 |
| Vertebral number 10 | 8.0 | 7.0 | 14.57 | 12.81 | 3.59 | 3.49 | MID1025 | 22.3 | 28.84 |
| Vertebral number 17 | 40.0 | 40.0 | 4.28 | 4.11 | 3.59 | 3.49 | MID1718 | 5.4 | 28.51 |
| Vertebral number 23 | 34.2 | 34.1 | 4.39 | 4.46 | 3.59 | 3.49 | MID2313 | 4.6 | 28.77 |

Threshold indicates the LOD score for a genome-wide significance threshold of $\alpha = 0.05$ determined by 10,000 iterated permutations. Position indicates QTL peak position on chromosome (cM). The mean SEM of the phenotype of F$_2$ sh for the different three genotypic classes of closest marker, Hd-rR-II1, heterozygous, and Kaga, are shown. The numbers of nearest neighbors are $11$ for vertebral number 1, $12$ for vertebral number 10, $11$ for vertebral number 17, and $15$ for vertebral number 23. Although one QTL was detected for the ratio of abdominal vertebrae to vertebral number, ratios for vertebral numbers 1, 10, 17, and 23 were not significantly different from 1 by Student’s t-test.

Because the epaxial axis intervals possess a shorter segment length, the NOD score may explain the difference in vertebral number.

Therefore, the number of vertebrae is related to the anterior-posterior position on chromosome (cM). The mean SEM of the phenotype of F$_2$ sh for the different three genotypic classes of the closest marker, Hd-rR-II1, heterozygous, and Kaga, are shown. The PVEs and means are values of parametric analysis. Chr, chromosome; LOD, logarithm of odds; PVE, percentage variance explained.

**Table 2 QTL position and effect**

| Trait | Position LOD Score Threshold | Parametric Nonparametric Parametric Nonparametric Parametric Nonparametric Closest Marker PVE Hd-rR-II1 Heterozygous Kaga |
|-------|-----------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Vertebral number 1 | 38.0 | 39.0 | 4.44 | 4.26 | 3.59 | 3.49 | MID0122 | 2.6 | 28.96 |
| Vertebral number 10 | 8.0 | 7.0 | 14.57 | 12.81 | 3.59 | 3.49 | MID1025 | 22.3 | 28.84 |
| Vertebral number 17 | 40.0 | 40.0 | 4.28 | 4.11 | 3.59 | 3.49 | MID1718 | 5.4 | 28.51 |
| Vertebral number 23 | 34.2 | 34.1 | 4.39 | 4.46 | 3.59 | 3.49 | MID2313 | 4.6 | 28.77 |
among the inbred lines, but the results of this study suggest that it is not the only factor.

In this study, we detected one suggestive QTL associated with the ratio of abdominal vertebrae to total vertebral number and five QTL associated with vertebral number. Additionally, we found that the anteroposterior length of somites explains the difference in the total vertebral number among the inbred lines. Thus, the developmental process provides important clues for genetic analysis. The identification of genes within these QTL can help in defining mechanisms underlying the establishment of vertebrate body plans.

ACKNOWLEDGMENTS

We thank the members of our laboratories for their useful discussion. We thank the National Institute for Basic Biology (NIBB) Core Research Facility and the NIBB BioResource Center for excellent technical service involving fish husbandry and stocks. The Hd-rR-II1 and Kaga strains were obtained from the National BioResource Project Medaka (NBRP Medaka), which is supported by the Ministry of Education, Culture, Sports, Science and Technology in Japan. This work was supported by the NIBB Cooperative Research Program (8-202, 9-203, 10-202).

Figure 3 QTL associated with the vertebral number. Black and blue lines indicate the logarithm of odds (LOD) curves of interval mapping for the parametric and the nonparametric analyses, respectively. The significance LOD threshold was determined by 10,000 iterated permutations. The black dashed line indicates genome-wide significance for the parametric analysis at threshold $\alpha = 0.05$ (LOD = 3.59). The blue dashed line indicates genome-wide significance for the nonparametric analysis at threshold $\alpha = 0.05$ (LOD = 3.49). The black and blue bars indicate Bayes credible interval (95%) for the parametric and nonparametric analyses, respectively.

Figure 4 Anteroposterior length of somites two to four differs between Hd-rRII1 and Kaga strains. Anteroposterior length of somites two to four was measured in stage 20–21 embryos (four to seven somite stage). (A) Five somite stage Hd-rRII1 embryo. (B) Five somite stage Kaga embryo. Vertical red lines show anteroposterior length of each somite. The scale bars represent 100 $\mu$m. (C) Box plot of anteroposterior length of somites two to four. The mean $\pm$ SEM of the anteroposterior length of somites two to four are 99.6 $\pm$ 1.5 in Hd-rRII1 ($n = 34$) and 116.8 $\pm$ 1.4 in Kaga ($n = 33$). Kaga strain has approximately 17% longer somites than Hd-rRII1. The somite length showed significant difference based on Welch’s t-test ($P = 2.64 \times 10^{-12}$).
