Centromere localization in medaka fish based on half-tetrad analysis

Masahiro Furuyama, Haruna Nagaoka, Tadashi Sato and Mitsuru Sakaizumi*

Department of Environmental Science, Graduate School of Science and Technology, Niigata University, Ikarashi-Ninocho, Nishi-ku, Niigata 950-2181, Japan

(Received 6 May 2019, accepted 1 June 2019; J-STAGE Advance published date: 10 August 2019)

Gene-centromere (G-C) mapping provides insight into vertebrate genome composition, structure and evolution. Although medaka fish are important experimental animals, no genome-wide G-C map of medaka has been constructed. In this study, we used 112 interspecific triploid hybrids and 152 DNA markers to make G-C maps of all 24 linkage groups (LGs). Under the assumption of 50% interference, 24 centromeres were localized onto all corresponding medaka LGs. Comparison with 21 centromere positions deduced from putative centromeric repeats revealed that 19 were localized inside the centromeric regions of the G-C maps, whereas two were not. Based on the centromere positions indicated in the G-C maps and those of centromeric repeats on each LG, we classified chromosomes as either biarmed or monoarmed; \( n = 24 = 10 \) metacentrics/submetacentrics + 14 subtelocentrics/acrocentrics, which is consistent with the results of previous karyological reports. This study helps to elucidate genome evolution mechanisms, and integrates physical and genetic maps with karyological information of medaka.

Key words: gene-centromere mapping, Oryzias luzonensis, O. latipes, O. sakaizumii, triploid

INTRODUCTION

The identification of centromere positions is essential for the integration of cytogenetic and genetic linkage maps and is also an initial step toward understanding the composition and structure of the centromeric region as well as the whole genome. Mainly due to the lack of well-defined genetic linkage maps using co-dominant markers, centromeres have been located only in very limited fish species so far. Most previous gene-centromere (G-C) map studies on fish species were based on allozyme markers and conducted decades ago. With the advantages and popularity of co-dominant DNA markers, G-C maps have recently been reported for many fish species. Zebrafish, Danio rerio, was the first fish for which all 25 centromeres were localized on genetic linkage maps (Johnson et al., 1996), followed by rainbow trout, Oncorhynchus mykiss (Sakamoto et al., 2000), loach, Misgurnus anguillicaudatus (Morishima et al., 2001), Japanese eel, Anguilla japonica (Nomura et al., 2006), turbot, Scophthalmus maximus (Martínez et al., 2008), large yellow croaker, Pseudosciaena crocea (Li et al., 2008), half-smooth tongue sole, Cynoglossus semilaevis (Ji et al., 2009), walking catfish, Clarias macrocephalus (Poompuang and Sukkorntong, 2011) and bighead carp, Hypophthalmichthys nobilis (Zhu et al., 2013).

Medaka fish (Oryzias latipes, O. sakaizumii and O. sinensis) represent one of the most important experimental animals worldwide, and have been used in fields such as genetics and developmental biology. The first medaka G-C map was made for linkage group (LG) 1, which is the sex LG (Sato et al., 2001). The distance between the marker and centromere is very important for estimating centromeric regions: the shorter the distance, the more accurate the centromeric regions. In this study, we aimed to localize centromeres onto all 24 LGs. The information obtained from this G-C map will be useful for understanding the genome structure and chromosome evolution of this species group.

MATERIALS AND METHODS

Fish All fish strains used in this study were supplied by the Faculty of Science, Niigata University, a subcenter of the National BioResource Project (medaka) supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan, and were reared in 150 × 250 × 105-mm plastic vessels with 2 l of water at 25 °C. Oryzias luzonensis was originally obtained from Luzon (Formacion and Uwa, 1985), and two inbred strains of medaka, HNI-
II and Hd-rR (derived from *O. sakaizumii* and *O. latipes*, respectively), were described by Hyodo-Taguchi and Sakaizumi (1993).

**Crosses**  
Hd-rR strain females were mated with HNI-II males, which resulted in F1 progeny (DNF1). DNF1 females and *O. luzonensis* males were isolated the day before they were mated. Fertilized eggs were immediately collected after spawning; after 2 min and 45 sec, they were exposed to 42 °C water for 2 min to block the second meiotic division (Naruse et al., 1985). The hatching rate of the hybrid embryos is improved by high-temperature treatment, and this improvement has been verified to be the result of triploidization (Sato et al., 2001). Thus, all or most of the hatched fry are expected to be allotriploids.

**Genotyping**  
DNA was extracted from the hatched fry by proteinase K digestion, phenol–chloroform extraction and isopropanol precipitation (Shinoyama et al., 1999). DNA samples were dissolved in TE buffer. To find DNA markers to detect polymorphisms between Hd-rR, HNI-II and *O. luzonensis*, we selected expressed sequence tag (EST) markers from the medaka EST database (http://mbase.nig.ac.jp/mbase/medaka_top.html), and made new markers as necessary (Supplementary Table S1). PCR amplification of genomic DNA was performed as follows: 35 cycles at 95 °C for 30 sec, 55 °C for 30 sec, and 72 °C for 2 min. All PCR products were electrophoresed on polyacrylamide gels as described by Kimura et al. (2004). In this study, we developed an SNP genotyping system between the HNI-II and Hd-rR inbred strains. SNP genotyping was performed using the MassARRAY iPLEX Gold Assay (Sequenom) (Takada et al., 2013). The information used for genotyping is described in Supplementary Table S2.

**Centromere localization**  
The second meiotic division segregation frequency (y), which is a function of the recombination rate between the marker and centromere, was estimated by counting the proportion of progeny heterozygous for Hd-rR and HNI-II alleles in allotriploid hybrids, which allowed us to estimate G-C distance (x) (Johnson et al., 1996). G-C distance may be estimated under different models of chiasma interference (Danzmann and Gharbi, 2001). Assuming 50% chiasma interference, G-C distance was calculated as follows: 

\[
x = \frac{\ln (1 + y) - \ln (1 - y)}{100 / 4}.
\]

**Chromosome type classification**  
Chromosomes were classified according to the criterion of the centromeric index (CI; long arm to short arm ratio) introduced by Levan et al. (1964). Chromosome type and corresponding CI values were as follows: metacentric, 1.0–1.7; submetacentric, 1.7–3.0; subtelocentric, 3.0–7.0; and acrocentric, more than 7.0.

**RESULTS AND DISCUSSION**

**Construction of centromere-linkage map**  
In this study, we used 112 triploid hybrids and 152 DNA markers (Supplementary Table S1). Hatched fry were all allotriploid hybrids, as determined by genotyping using DNA markers. The map length of each chromosome was 44.3–58.7 cM (average, 50.2 cM). The linkage map spanned a genetic length of 1,204 cM (Table 1), which is 4% shorter than that proposed in a previous study (1,257 cM in female, Kimura et al., 2005); however, we used terminal markers for the physical map of each chromosome in this study. This difference may also be partially due to long interval distances between loci (average interval distance, 9.3 vs. 5.4 cM in Kimura et al., 2005).

| LG | Map length (cM) | Recombination rate (cM/Mbp) |
|----|----------------|-----------------------------|
|    |                | Centromeric region* | Other region |
| 1  | 56.1           | 0.95                       | 1.87         |
| 2  | 51.0           | 0.61                       | 2.24         |
| 3  | 52.6           | 1.14                       | 2.18         |
| 4  | 58.7           | 0.58                       | 2.18         |
| 5  | 56.1           | 0.47                       | 1.96         |
| 6  | 45.9           | 0.60                       | 1.56         |
| 7  | 50.7           | 1.21                       | 1.47         |
| 8  | 47.4           | 0.19                       | 2.53         |
| 9  | 44.5           | 0.70                       | 1.65         |
| 10 | 49.4           | 0.06                       | 2.07         |
| 11 | 44.3           | 2.72                       | 1.50         |
| 12 | 46.2           | 0.32                       | 2.21         |
| 13 | 49.5           | 0.75                       | 1.59         |
| 14 | 48.3           | 0.75                       | 1.85         |
| 15 | 47.1           | 0.48                       | 1.70         |
| 16 | 47.6           | 1.12                       | 1.71         |
| 17 | 50.5           | 0.23                       | 1.71         |
| 18 | 55.2           | 2.01                       | 1.77         |
| 19 | 52.1           | 1.41                       | 2.74         |
| 20 | 44.7           | 1.72                       | 2.15         |
| 21 | 53.4           | 1.14                       | 2.06         |
| 22 | 49.0           | 0.77                       | 1.99         |
| 23 | 52.8           | 2.09                       | 2.24         |
| 24 | 50.6           | 1.31                       | 2.51         |
| Total | 1,203.7     |                             |              |

*Centromeric and other regions are depicted as gray and white boxes, respectively, in G-C maps (Fig. 1).
Fig. 1. Gene-centromere maps of 24 medaka LGs. For each chromosome, the physical map (Hd-rR) is on the left, and the G-C map is on the right (n = 112). Centromere locations determined by half-tetrad analysis are shown as open circles. Centromeric regions are indicated as gray boxes in the physical maps. Centromeric repeats identified by Ichikawa et al. (2017) for Hd-rR and HNI-II are represented by black and white triangles, respectively.
Fig. 1. Continued
meres were successfully positioned onto all 24 LGs of the medaka genetic linkage map (Fig. 1).

In general, centromeric regions possess lower levels of recombination while telomeric regions of the chromosome experience increased crossover events (Choo, 1998; International Human Genome Sequencing Consortium, 2001). We found that almost every centromeric region of the G-C maps, except for those of LGs 11 and 18, had lower recombination frequencies (Table 1). LG11 of HNI-II has a large chromosomal inversion (7–23 Mbp) in the central region and that of Hd-rR does not (Kimura et al., 2005; Ichikawa et al., 2017), which may explain the high recombination frequency in the centromeric region of LG11. For LG18, the relatively higher level of recombination in the centromeric region may be partially due to insufficient narrowing of this region.

### Comparison between centromeric regions in the G-C maps and putative centromeric repeat positions
Putative centromeric repeats have been reported in three inbred medaka strains: Hd-rR, HNI-II and HSOK (Ichikawa et al., 2017). The repeats were identified in one or multiple strains for all LGs except LGs 15 and 24. Based on these results, we predicted the centromere locations on the chromosomes (Table 2). For six LGs (2, 6, 11, 12, 19 and 21), centromeric repeats were identified in Hd-rR and HNI-II; for 17 LGs (1, 3, 4, 5, 7, 8, 9, 10, 13, 14, 15, 16, 17, 18, 20, 22 and 23), they were found in either Hd-rR or HNI-II. The repeats on LG6 were located at 0 Mbp in Hd-rR, but at the opposite end in HNI-II. Furthermore, two centromere positions were predicted for LG21 in Hd-rR.

A comparison between the positions of putative centromeric repeats on 23 LGs and the centromeric regions

### Table 2. Classification of chromosomes

| LG | Centromeric region on G-C map (Mbp) | Chromosome type* judged from | G-C map | Position of centromeric repeats in Hd-rR | Position of centromeric repeats in HNI-II | FISH and/or silver staining | Expected chromosome type |
|----|----------------------------------|-----------------------------|---------|----------------------------------------|------------------------------------------|--------------------------|-------------------------|
| 1  | 23.33–36.66 M/SM/ST/A            | SM                          | SM      | –                                      | SM                                      | Bi (SM)                  |
| 2  | 10.12–13.06 M                    | M                           | M       | M                                      | –                                       | Bi (M)                   |
| 3  | 0.44–14.45 M/SM/ST/A             | ST                          | –       | –                                      | –                                       | Mono (ST)               |
| 4  | 0.04–7.87 ST/A                   | A                           | A       | –                                      | –                                       | Mono (A)               |
| 5  | 0.12–5.87 ST/A                   | –                           | A       | –                                      | –                                       | Mono (A)               |
| 6  | 1.01–4.07 A                      | A                           | A       | A                                      | –                                       | Mono (A)               |
| 7  | 33.83–34.57 A                    | –                           | A       | –                                      | –                                       | Mono (A)               |
| 8  | 15.37–22.52 M/SM/ST              | SM                          | –       | –                                      | –                                       | Bi (SM)                 |
| 9  | 18.24–24.72 M/SM                 | SM                          | –       | SM                                    | Bi (SM)                   |
| 10 | 0.14–7.53 ST/A                   | –                           | ST      | –                                      | –                                       | Mono (ST/A)           |
| 11 | 26.39–28.21 A                    | A                           | A       | A                                      | A                                      | Mono (A)               |
| 12 | 18.25–24.10 M/SM/ST              | SM                          | SM      | –                                      | –                                       | Bi (SM)                 |
| 13 | 29.66–32.77 A                    | A                           | A       | –                                      | A                                      | Mono (A)               |
| 14 | 0.26–6.21 ST/A                   | A                           | A       | –                                      | –                                       | Mono (A)               |
| 15 | 0.44–3.85 ST/A                   | –                           | –       | –                                      | –                                       | Mono (ST/A)           |
| 16 | 0–5.90 ST/A                      | A                           | A       | –                                      | –                                       | Mono (A)               |
| 17 | 0.56–2.50 A                      | A                           | A       | –                                      | –                                       | Mono (A)               |
| 18 | 12.60–21.94 M/SM                 | SM                          | –       | SM                                    | Bi (SM)                   |
| 19 | 11.33–19.13 M/SM/ST              | SM                          | SM      | –                                      | Bi (SM)                   |
| 20 | 0–12.14 M/SM/ST/A                | ST                          | –       | SM (NOR)                              | Bi (SM)                   |
| 21 | 8.60–16.55 M/SM                  | D (A or M)                  | A       | –                                      | –                                       | Bi (M)                 |
| 22 | 23.69–28.97 ST/A                 | A                           | –       | –                                      | –                                       | Mono (A)               |
| 23 | 9.26–15.46 M/SM                  | M                           | –       | –                                      | –                                       | Bi (M)                 |
| 24 | 19.02–23.24 ST/A                 | –                           | –       | –                                      | –                                       | Mono (ST/A)           |

*Classification of chromosomes according to Levan et al. (1964): M, metacentric; SM, submetacentric; ST, subtelocentric; A, acrocentric; Bi, biarmed; Mono, monoarmed; D, dicentric.
in the G-C maps revealed that repeat positions on 19 LGs (1–5, 7–12, 14, 16–20, 22 and 23) were inside their respective centromeric regions in the G-C maps, whereas those of LGs 6 and 13 were not. The putative centromeric repeats on LG6 were at 0 Mbp in Hd-rR, but at the opposite end in HNI-II, and those on LG13 may not be functional. For LG 21, two positions (2.5 and 14.5 Mbp) of centromeric repeats were found in Hd-rR, whereas only one (2.5 Mbp) was found in HNI-II; additionally, the repeats at 14.5 Mbp were located inside the centromeric region in the G-C map, which indicates that these repeats have centromeric function. The repeats at 2.5 Mbp may be silenced in both Hd-rR and HNI-II. The centromere of HNI-II is assumed to be located at the same position as that of Hd-rR. Although centromeric repeats were not identified on LG24, the centromere was mapped at one end (19–23 Mbp) of the chromosome.

**Chromosome classification** Oryzias latipes and O. sakaizumii have 24 pairs of chromosomes (2n = 48) that were classified into two types: 10 pairs of biarmed (metacentric and submetacentric) and 14 pairs of monoarmed (subtelocentric and acrocentric) chromosomes (Uwa and Ojima, 1981). Previous fluorescence *in situ* hybridization (FISH) studies showed that LGs 1, 9, and 18 are biarmed (Matsuda et al., 1998; Brunner et al., 2001; Kondo et al., 2002), whereas LGs 11 and 13 are monoarmed (Myosho et al., 2012). Silver staining analysis by Uwa and Ojima (1981) revealed nucleolus organizer regions (NORs) on the tip of the short arms of a satellited chromosome pair. This result and our FISH analysis demonstrated a NOR on LG 20 (data not shown), which indicates that LG 20 is biarmed and bears a NOR.

Based on the definition of chromosome types (Levan et al., 1964) and our G-C maps, we classified LGs 2, 9, 11, 21 and 23 as biarmed chromosomes, and LGs 4–7, 10, 11, 13–17, 22 and 24 as monoarmed chromosomes. The chromosome types of the five remaining LGs were not determined from the G-C maps alone, because their centromeric regions could not be narrowed enough. Based on the position of the putative centromeric repeats, we classified Hd-rR chromosomes into two types: nine biarmed LGs (1, 2, 8, 9, 12, 18, 19, 21 and 23) and 10 monoarmed LGs (3, 4, 6, 11, 13, 14, 16, 17, 20 and 22). For HNI-II, LGs 2, 12 and 19 were biarmed, whereas LGs 5–7, 10, 11 and 21 were monoarmed (Table 2).

We classified all 24 LGs into either biarmed (1, 2, 8, 9, 12, 18–21 and 23) or monoarmed chromosomes (3–7, 10, 11, 13–17, 22 and 24). Although the chromosome type of LG 20 was determined to be monoarmed (subtelocentric) based on the centromeric repeats of Hd-rR, this chromosome was expected to be biarmed, because this NOR-bearing chromosome was classified as submetacentric (Uwa and Ojima, 1981).

The results of this study integrate the centromere map and genetic linkage map in O. latipes/O. sakaizumii, which provides valuable information that will help to consolidate the genetic and physical maps in the near future. These findings will also be useful for studies on genome structure, chromosome evolution and positional cloning of genes in this species complex.

We are grateful to Prof. T. Shiroishi and Dr. T. Takada of the National Institute of Genetics for their generous help with SNP genotyping. We are also grateful to Prof. Y. Matsuda and Dr. Y. Uno of Nagoya University, and Dr. Y. Takehana of Nagahama Institute of Bio-Science and Technology, for their invaluable advice regarding FISH analysis. We thank Dr. Mallory Eckstut from Edanz Group for editing a draft of this manuscript.

**REFERENCES**

Brunner, B., Hornung, U., Shan, Z., Nanda, I., Kondo, M., Zend-Aajusch, E., Haaf, T., Ropers, H.-H., Shima, A., Schmid, M., et al. (2001) Genomic organization and expression of the doublesex-related gene cluster in vertebrates and detection of putative regulatory regions for DMRT1. Genomics 77, 8–17.

Choo, K. H. A. (1998) Why is the centromere so cold? Genome Res. 8, 81–82.

Danzmann, R. G., and Gharbi, K. (2001) Gene mapping in fishes: a means to an end. Genetica 111, 3–23.

Forsman, M. J., and Uwa, H. (1985) Cytogenetic studies on the origin and species differentiation of the Philippine medaka, Oryzias luzonensis. J. Fish Biol. 27, 285–291.

Hyodo-Taguchi, Y., and Sakaizumi, M. (1993) List of inbred strains of the medaka, Oryzias latipes, maintained in the Division of Biology, National Institute of Radiological Science. Fish Biol. J. Medaka 5, 29–30.

Ichikawa, K., Tomioka, S., Suzuki, Y., Nakamura, R., Doi, K., Yoshimura, J., Kumagai, M., Inoue, Y., Uchida, Y., Irie, N., et al. (2017) Centromere evolution and CpG methylation during vertebrate speciation. Nat. Commun. 8, 1833.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Ji, X. S., Chen, S. L., Liao, X. L., Yang, J. F., and Xu, T. J. (2009) Microsatellite-centromere mapping in Cynoglossus semilaevis using gynogenetic diploid families produced by the use of homologous and non-homologous sperm. J. Fish Biol. 75, 422–434.

Johnson, S. L., Gates, M. A., Johnson, M., Talbot, W. S., Horne, S., Baik, K., Rude, S., Wong, J. R., and Postlethwait, J. H. (1996) Centromere-linkage analysis and consolidation of the zebrafish genetic map. Genetics 142, 1277–1288.

Kimura, T., Jindo, T., Narita, T., Naruse, K., Kobayashi, D., Shin-I, T., Kitagawa, T., Mitani, H., Shima, A., Kohara, Y., et al. (2004) Large-scale isolation of ESTs from medaka embryos and its application to medaka developmental genetics. Mech. Dev. 121, 915–932.

Kimura, T., Yoshida, K., Shimada, A., Jindo, T., Sakaizumi, M., Mitani, H., Naruse, K., Takeda, H., Inoko, H., Tamiya, G., et al. (2005) Genetic linkage map of medaka with polymerase chain reaction length polymorphisms. Gene 363, 24–31.

Kondo, M., Froschauer, B., Kitano, A., Nanda, I., Hornung, U., Vollf, J.-N., Asakawa, S., Mitani, H., Naruse, K., Tanaka, M., et al. (2002) Molecular cloning and characterization of DMRT genes from the medaka Oryzias latipes and the platyfish Xiphophorus maculatus. Gene 295, 213–222.
Gene-centromere mapping of medaka

Levan, A., Fredga, K., and Sandberg, A. A. (1964) Nomenclature for centromere position on chromosomes. Hereditas 52, 201–220.

Li, Y., Cai, M., Wang, Z., Guo, W., Liu, X., Wang, X., and Ning, Y. (2008) Microsatellite-centromere mapping in large yellow croaker (Pseudosciaena crocea) using gynogenetic diploid families. Mar. Biotechnol. 10, 83–90.

Martínez, P., Hermida, M., Pardo, B. G., Fernández, C., Castro, J., Cal, R. M., Alvarez-Dios, J. A., Gómez-Tato, A., and Bouza, C. (2008) Centromere-linkage in the turbot (Scophthalmus maximus) through half-tetrad analysis in diploid meiogonetics. Aquaculture 280, 81–88.

Matsuda, M., Matsuda, C., Hamaguchi, S., and Sakaizumi, M. (1998) Identification of sex chromosomes of the medaka, Oryzias latipes, by fluorescence in situ hybridization. Cytogenet. Cell Genet. 82, 257–262.

Nomura, K., Morishima, K., Tanaka, H., Unuma, T., Okuzawa, K., Ohta, H., and Arai, K. (2006) Microsatellite-centromere mapping in the Japanese eel (Anguilla japonica) by half-tetrad analysis using induced triploid families. Aquaculture 257, 53–67.

Poompuang, S., and Sukkorntong, C. (2011) Microsatellite-centromere mapping in walking catfish Clarias macrocephalus (Günther, 1864) using gynogenetic diploids. Aquacult. Res. 42, 210–220.

Sakamoto, T., Danzmann, R. G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S. K., Woram, R. A., Okamoto, N., Ferguson, M. M., Holm, L. E., et al. (2000) A microsatellite linkage map of rainbow trout (Oncorhynchus mykiss) characterized by large sex-specific differences in recombination rates. Genetics 155, 1331–1345.

Sato, T., Yokomizo, S., Matsuda, M., Hamaguchi, S., and Sakaizumi, M. (2001) Gene-centromere mapping of medaka sex chromosomes using triploid hybrids between Oryzias latipes and O. luzonensis. Genetica 111, 71–75.

Shinomiya, A., Matsuda, M., Hamaguchi, S., and Sakaizumi, M. (1999) Identification of genetic sex of the medaka, Oryzias latipes, by PCR. J. Exp. Zool. 280, 53–67.

Uwa, H., and Ojima, Y. (1981) Detailed and banding karyotype analysis of the medaka, Oryzias latipes, in cultured cells. Proc. Japan Acad. 57, 39–43.

Zhu, C., Sun, Y., Yu, X., and Tong, J. (2013) Centromere localization for Bighead Carp (Aristichthys nobilis) through half-tetrad analysis in diploid gynogenetic families. PLoS One 8, e82950.
Supplementary Table S1. Marker catalog used in this study

| LG | Marker name | Mbp (ver 2) | Note | Forward primer | Reverse primer | Polymorphism |
|----|-------------|-------------|------|----------------|----------------|--------------|
| 1  | MA01-1      | 0.25        |       |                |                | SNP          |
| 1  | MA01-2      | 11.07       | Mass array marker* |                |                | SNP          |
| 1  | SL1         | 18.27       | Matsuda et al. (1999) | CCGAATGGGAAATTATTCTGCTC | CTTTTGTGCTTTGGTTATGAAACGATG | In/del       |
| 1  | MA01-3      | 23.33       | Mass array marker |                |                | SNP          |
| 1  | SL2         | unkown      | Matsuda et al. (1999) | GCCATCTTGAGGTAGCCCAT | CTTTTGTGCTTTGGTTATGAAACGATG | RFLP (DraI & RsaI digest) |
| 1  | MEDAKA081715| 27.17       | This study | ACCCCCCTTCTGCTGCTT | ACTGTGACCTGCTTTCCATC | Heteroduplex |
| 1  | MA01-4      | 36.66       | Mass array marker |                |                | SNP          |
| 2  | MA02-1      | 0.20        | Mass array marker* |                |                | SNP          |
| 2  | MF01SSA007A05| 10.12     | EST | GGCACACACAAAATGTAAGA | TGCTTTTAAGGCTTTGGTTATGAAACGATG | In/del |
| 2  | MF01SSA051F03-2 | 11.09    | EST | TCCGGATCTGAGCTGATCTC | CCGAATCCACAGGAGACGACGAGTC | In/del |
| 2  | OLa1510f    | 13.06       | EST | GCCATCTTGAGGTAGCCCAT | CTTTTGTGCTTTGGTTATGAAACGATG | SNP          |
| 2  | MA02-2b     | 17.60       | Mass array marker |                |                | SNP          |
| 2  | MA02-4c     | 25.08       | Mass array marker |                |                | SNP          |
| 3  | MA03-1      | 0.44        | Mass array marker* |                |                | SNP          |
| 3  | MF01SSA020C07 | 4.53     | EST | TCTACAGAAACATCCAAACAG | TTGAGCTGTATAACGGAAGAG | In/del |
| 3  | MF01SSA034A09 | 8.57      | EST | CCGAATCCACAGGAGACG | CCCAATCCAAACTCCACAGTC | In/del |
| 3  | AU171481    | 9.83        | EST | TGGGTGCTTCTGCGNATGTTCC | ATGCCTGTACGAGCACAGATGA | SNP |
| 3  | OLa1204g    | 14.45       | EST | GCCATCTTGAGGTAGCCCAT | CTTTTGTGCTTTGGTTATGAAACGATG | SNP          |
| 3  | MAPCR03-2   | 18.68       | This study | TGGGAATGCTTCAACGCTGTC | CACTGTGACAGAGGCCTGAAAGA | Heteroduplex |
| 3  | AU171644    | 26.25       | EST | CAGGCTGAGAAAATACGAGCCGGAGG | GCATTCAAGATACGAGGGACCGACCTGAGG | In/del |
| 3  | MA03-4      | 32.66       | Mass array marker |                |                | SNP          |
| 4  | MF01SSA026F02-5 | 0.04    | EST | CGTAAATATGGGAAATGCTGACAGG | GCATTCAAGATACGAGGGACCGAC | SNP |
| 4  | MF01SSA037B05 | 7.87      | EST | ATGCTGGATGCGCAGCAGGCAGAG | ACACCAAGCCCAACACT | In/del |
| 4  | MF01SSA034A09 | 13.93     | EST | CGGCGAGGTTGCTGTC | TGGCAATACGAGGACG | Heteroduplex |
| 4  | MAPCR04-3   | 17.99       | This study | TTCAGTTCTGATGCTTTTCG | TGGCAATACGAGGACG | Heteroduplex |
| 4  | MAPCR04-2   | 27.54       | This study | TTCAGTTCTGATGCTTTTCG | TGGCAATACGAGGACG | Heteroduplex |
| 4  | MA04-4      | 32.66       | Mass array marker |                |                | SNP          |
| 5  | MAPCR05-1   | 0.12        | This study | ACGAGTTTCTCAGGCCCTCATC | GTGTTTCCATGCTTGATGCT | In/del |
| 5  | OLa2212b    | 5.87        | EST | AAGAAGAAGCGGATGCGCTGCTGT | TGGCAATACGAGGACG | Heteroduplex |
| 5  | MF01SSA044H11 | 12.91     | EST | AGAGAAAGGTTGAGCAAGAGA | CTTGCAATACGAGGACG | Heteroduplex |
| 5  | MA05-2      | 18.13       | Mass array marker |                |                | SNP          |
| 5  | OLa0607f    | 25.38       | EST | AGCGACTCTGAGCCCTTACC | TGGCAATACGAGGACG | Heteroduplex |
| 5  | MA05-4      | 32.94       | Mass array marker |                |                | SNP          |
| 6  | MA06-1      | 1.01        | Mass array marker |                |                | SNP          |
| 6  | MF01SSA047D04 | 1.28       | EST | GATCGTTCTCAGGCTTTTTGT | AGCATTAACCTGTTGAAACCT | In/del |
| Gene ID | Score | Type       | Description                  | Sequence 1                           | Sequence 2                           | In/del     |
|--------|-------|------------|------------------------------|-------------------------------------|-------------------------------------|------------|
| OLe1804f | 2.77  | EST        | CTCATGCTCCCGCAGGCTCAAGAA     | TTGGTTTCGACAGACCCGGGTAC             | In/del                              |
| MEDAKA045681 | 4.07 | This study | AGTGGTGCCGCTCAGTTAAGG       | TTGGTGAAGGTCATGGTGAAGG             | In/del                              |
| MF01SSA061H06 | 8.34 | EST        | GCCATGCTGAAGCGCTGCGAGA      | TCCCAACATTTCACGACTCCAA             | In/del                              |
| MF01SSA044D08 | 13.96 | EST       | CGTGGGAAAGAAGTGTGTTGGA      | GTCGTTGTTGTGCTGTTGACC             | In/del                              |
| MA06-2  | 16.10 | Mass array marker |               |                                      |                                      | SNP        |
| MF01SSA046F03 | 20.22 | EST       | ATTTTGATGTTGGGCTGATGCT    | TCACGCTACCTGCTGTCATTGCTG           | In/del                              |
| MF01SSA031E04 | 26.79 | EST       | GTTTCTCCACCATCATGG         | CTTTTTCGAGATTGTTGAC              | In/del                              |
| MAPCR06-4 | 32.15 | This study | TTGCACTTCTGACCTGACATTC     | TGAAAATGACCACCCCTATTACC           | In/del                              |
| MAPCR07-1 | 0.13  | This study | TGCAACGTTGAGTTGTCAG    | GGCAGATTCCACATTACTG               | In/del                              |
| MF01SSA051B01 | 9.82 | EST        | CTCAATAAAGAAGGCGGACTC      | CTCATACATTGGACCTG                 | In/del                              |
| ncoa6   | 14.77 | This study | AGTGGCAGAGTTGTAAGG       | AAACACTTACATTCAACTGGG             | In/del                              |
| OLa1008g | 21.84 | EST        | AAGATGCGACGGCCGAAACACTT   | GGTAAAACGTGCTGACAGGGA             | In/del                              |
| MAPCR07-4 | 33.38 | This study | CTTCCCACCTTCTGATGTT       | GAAGCCACCAGCTGTTAAATG             | Heteroduplex |
| OLa0405c | 0.39  | EST        | CTCTCCCCTCTTTGAGAAGA    | CCAAACCTGGAGAGAGCTG              | In/del                              |
| MAPCR08-2 | 4.15  | This study | TTAAGGCAAAAGAAGGAGACAG    | CACTGTTTTGAAATACTG               | In/del                              |
| MF01SSA022B03 | 8.37 | EST        | AAAAGACGCTCCACCATAATTT    | CTATGAAATGAGAGTCGAAGGC            | In/del                              |
| MF01SSA159B11 | 13.45 | EST        | CGTGTAGGAGGGAGCTGCTCAGA   | TTGTTCCCTCCACACGAGGTTTATGAC       | In/del                              |
| OLB2807d | 15.37 | EST        | AGTGTGTTAGTATGGGCGGCAGC    | TTCTGCGTGTCTCCTATGGG             | In/del                              |
| OLB1911b | 19.29 | EST        | GCTGTCGGAAGGTTGTTTATTC    | AAGAGGTGCTGATGATGACGA            | In/del                              |
| MAPCR08-3 | 22.52 | This study | GATTCAGTGGGCAAGAGACAG    | CTTTGCCAAGAGTCGATATT            | In/del                              |
| MA08-4  | 25.69 | Mass array marker |               |                                      |                                      | SNP        |
| MF01SSA038G10 | 1.27  | EST        | CTGGTCTCCATTTGAAACTT      | GGTGTCTGGTGTTGGTAC              | In/del                              |
| MA09-2c  | 5.31  | Mass array marker |               |                                      |                                      | SNP        |
| OLB1509f | 12.23 | EST        | TAAAGGCCCTTCTCCTGCTTCC    | ATTTTTCCACATGCTGACGCAGGCC         | In/del                              |
| OLB2107h | 17.10 | EST        | GTGTGCCAGCTCCCGAATTGAC    | CTCCTCCAGTGCTGGATACAGGTTG        | In/del                              |
| AU170636 | 18.24 | EST        | GCAAAGGAGACTAAACCCCTAGGACCT    | GGCACCCATTTGAGCCTATTAAGGCC         | In/del                              |
| MEDAKA036243-1 | 24.72 | EST        | AAGAGGAGCAGTGCTTTGGAC     | TTGTTGCAGGCCAGATTAAA           | In/del                              |
| MA09-4b  | 31.87 | Mass array marker |               |                                      |                                      | SNP        |
| MA10-1  | 0.14  | Mass array marker |               |                                      |                                      | SNP        |
| MEDAKA085272 | 7.53  | This study | CTTGGAGACGGCGTTGTTATT    | ATCAGCTCAGAAACGCCCTCA             | In/del                              |
| OLB2511e | 13.95 | EST        | ACACGGGAAATGGCAGACCGCTAG  | GTGACGGCGAGATGTTGACATCAAACATG     | In/del                              |
| MA10-2  | 18.22 | Mass array marker |               |                                      |                                      | SNP        |
| Casp3A  | 24.53 | EST        | TGATATACGGCGACTGATGCTGCG    | GGTGCTCCTCCACCGAGTAATAGC          | Heteroduplex |
| MAPCR10-4 | 31.08 | This study | CAAATAATCTTCAAAACAGGACCCAGG | TTTTGCAGAGCCAGATTTAAC             | In/del                              |
| MA11-1  | 0.18  | Mass array marker |               |                                      |                                      | SNP        |
| MA11-2b  | 4.41  | Mass array marker |               |                                      |                                      | SNP        |
| OLB1508c | 11.21 | EST        | GAGGCCAAGAAGGCGAAGAGG     | CTTTTTAACAGCGAGGGCTTCTCC          | Heteroduplex |
| MA11-3a  | 19.33 | Mass array marker |               |                                      |                                      | SNP        |
| 11 | MA11-4 | 26.39 | Mass array marker | SNP |
|----|--------|-------|-------------------|-----|
| 12 | MAPCR12-1 | 0.04 | This study | GCCCTTCTACACACACAGCAC | TAAAGCGAAAGCCCCAAGG | In/del |
| 12 | slc45a2 | 10.57 | This study | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 12 | MF01SSA009C11 | 12.27 | EST | CGTATATTCAGAATTTGCCCA | TGGAAAGTACACCGAGATGA | In/del |
| 12 | MF01SSA180G12 | 15.81 | EST | CGAAGCGCGAGTGTGAAAGGG | AGTCTCCCATTGGAATACGCG | In/del |
| 12 | AU171862 | 18.25 | EST | CGTGGTGAGTACAGGTGCAACGGCA | CGCTGGAAGATGTCCTAAGGGTTGCC | In/del |
| 12 | MEDAKA080719-3 | 22.63 | This study | GCCCTTCTACACACAGCAC | TAAAGCGAAAGCCCCAAGG | In/del |
| 12 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 13 | MAPCR13-4 | 32.77 | This study | TGGCAAGGCGACGACAGAAAG | TAAAGCGAAAGCCCCAAGG | In/del |
| 13 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 13 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 13 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 13 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 14 | MAPCR13-4 | 32.77 | This study | TGGCAAGGCGACGACAGAAAG | TAAAGCGAAAGCCCCAAGG | In/del |
| 14 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 15 | MAPCR13-4 | 32.77 | This study | TGGCAAGGCGACGACAGAAAG | TAAAGCGAAAGCCCCAAGG | In/del |
| 15 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 15 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 15 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 15 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 15 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 16 | MAPCR13-4 | 32.77 | This study | TGGCAAGGCGACGACAGAAAG | TAAAGCGAAAGCCCCAAGG | In/del |
| 16 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 16 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 16 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 16 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 16 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 17 | MAPCR13-4 | 32.77 | This study | TGGCAAGGCGACGACAGAAAG | TAAAGCGAAAGCCCCAAGG | In/del |
| 17 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 17 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 17 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 17 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
| 17 | MF01SSA009C11 | 12.27 | EST | AAAAGGTGTAATCGTCTGCC | AAAATTCTGTGTCGTTGCT | In/del |
|   | ID                | Score | Type          | Reference   | Description             | Method       | SNP    |
|---|------------------|-------|---------------|-------------|--------------------------|--------------|--------|
| 17| MAPCR17–2        | 17.68 | This study    | CTTGTTTCCCTCACAACCTG | CCTGTTTCTGCTGTACCTTCG | In/del       |        |
| 17| OLc5811d         | 19.34 | EST           | CTTACAGTCCAGCGGCTACACTGAGAAGCTTGA | TTTCGCTGGAGGACTCGAAGGCA   | In/del       |        |
| 17| MAPCR17–4        | 31.64 | This study    | GCAGCTCAAGGGACTGGTTTGAAGGCAAGG | GGCAGCCCTGGATAATCAGA   | In/del       |        |
| 18| MA18–1a          | Unknown | Mass array marker | SNPs       |                         |              | SNP    |
| 18| MA18–2a          | 4.49  | Mass array marker | SNPs       |                         |              | SNP    |
| 18| OLa1011h         | 12.60 | EST           | ATATACCATGCAGAGGGTGTAGG | TTCTTCCCCCAATAGGTGTTGA | In/del       |        |
| 18| MAPCR18–3        | 19.61 | This study    | GCCAGCTCAAGGGACTGGTTTGAAGGCAAGG | GGCAGCCCTGGATAATCAGA   | In/del       |        |
| 18| OLb3105a         | 21.94 | EST           | TCAAGGCAGGAAGGACGGCAAAAGG | TQATGACATCACTGAACCCAGCAGG | In/del       |        |
| 18| MA18–4           | 30.48 | Mass array marker | SNP       |                         |              | SNP    |
| 19| MAPCR19–1        | 0.4   | This study    | TACAGGGCAGAACACGTGAC | AATATTGACTGGCCCAGGTG | In/del       |        |
| 19| OLa2308e         | 7.48  | EST           | TCCCTGCCAGATCCAGAGATGCTCAA | AACACTGGGACAGTGACTCGACCT | In/del       |        |
| 19| AU169836         | 11.33 | EST           | ATGTGTTATTTGTGGGTGTTAATGAGGC | CACTGAGAAAAAGATTGGAGAGCAGAC | In/del       |        |
| 19| Hoxb6b           | 19.13 | EST           | CCACCTCATCCTGTACCAGCATCAGAAG | CAATGGCAGATCCTGAAAGTCTGGGCA | In/del       |        |
| 19| OLa3002f         | 23.16 | EST           | TCTCCGCTCCTCCTGCTGAAGGCG | ATCTGCTGACCTTACAGGGCG | In/del       |        |
| 20| MF01SSA142H12    | 3.40  | EST           | TCGCTTCAAGGGCTCCACCAACA | CAGATCCTTTCAGGGCTCGACTCT | In/del       |        |
| 20| OLa1111f         | 12.14 | EST           | ATGTGCCAGCAGCCAGTGATCGGCAAAG | ATCCCGCGGATTGGTTCCAAAGCGC | In/del       |        |
| 20| MAPCR20–2        | 14.17 | This study    | TTCTCAGTCTGGAATCTCTTTCA | GTCAGGATTTGCGGAGAAGA | In/del       |        |
| 20| MA20–3           | 20.49 | Mass array marker | SNP       |                         |              | SNP    |
| 20| MA20–4b          | 25.89 | Mass array marker | SNP       |                         |              | SNP    |
| 21| MAPCR21–1        | 0.1   | This study    | GGAGAGCATCCTTGAAGTCG | CCACTCGTCAAGGAAACACACG | Heteroduplex |        |
| 21| OLa0706h         | 8.60  | EST           | CGGAAATCATGAAACTGTCACCCG | AGCCATAGGGGCAAAACCGAGTA | Heteroduplex |        |
| 21| AU170530         | 16.55 | EST           | CGCTCTCTTCTTTTTAAACAAATTATGCTGCTCA | GCTGCTGTTACAGTTCAAAACAACA | Heteroduplex |        |
| 21| AU169483         | 20.9  | EST           | ATGCTGGGAGGACTGCTGAATCTAC | GTGTGCTCCTGTAAGCAGCATGAA | In/del       |        |
| 21| MA21–3           | 23.99 | Mass array marker | SNP       |                         |              | SNP    |
| 21| MA21–4a          | 29.50 | Mass array marker | SNP       |                         |              | SNP    |
| 22| MA22–1           | 0.23  | Mass array marker | SNP       |                         |              | SNP    |
| 22| MF01SSA003A16–4  | 1.73  | EST           | TTGTTGGCCACAAACAAACAGGACTGGAC | GATTTCGCCGAGAAACCCCGTGACCA | Heteroduplex |        |
| 22| MAPCR22–2        | 6.89  | This study    | CACAAATGGTCTGGTTCGATG  | CTGAAAGCTGGCATGACTGCTA | In/del       |        |
| 22| OLb0103g         | 15.46 | EST           | AGCTTTTTCTTTGCTGCAAGGCTCAGG | CGATACACTGGGCACTTGTCAAAAT | In/del       |        |
| 22| MA23–4a          | 24.36 | Mass array marker | SNP       |                         |              | SNP    |
| 23| MAPCR23–1        | 1.1   | This study    | TGAAAGCTGATCATTCTGCTG | TTTACTCCCACAAACTATTTTCTTCTC | In/del       |        |
| 23| MF01SSA042C04    | 10.16 | EST           | TCTACAGGATTTGCGGTTGGGAGAAGC | AGATCAAGCTGCTGGTCCAAAC | Heteroduplex |        |
| 24| OLa2403g         | 14.27 | EST           | TGGTTACGGACCACATATTGAGGCA | TGGATTCAAGCTTGGACAAACACTCT | In/del       |        |
|   | Accession |   | Method                | Primer 1 | Primer 2    | Type  |
|---|-----------|---|-----------------------|----------|-------------|-------|
| 24| MAPCR24-3 | 17.14 | This study           | ATGAGGAGGCAGAATGACA | CCATCGTCAACGACTTTGC | In/del |
| 24| MEDAKA001666 | 19.02 | This study           | AAAGATGACTTGGACTAAACACTTG | TGAAAGGGGCTGTTCTTTCAC | In/del |
| 24| MAPCR24-4 | 23.24 | This study           | CCACAGTCACAGCGGAAGT | CATGCTGTTCACAGGGTTG | In/del |

*Primer extension/mass spectrometry method.*
### Supplementary Table S2. Massarray markers used in this study

| Marker name | Genotype | Forward primer | Reverse primer | Primer for extension |
|-------------|----------|----------------|----------------|----------------------|
| MA01-1      | T        | 253,629        | MA06-1b        | GCGAGGTGAACACGCTATGCA |
| MA01-1      | G        | 11,071,848     | MA04-1c        | CAGACACGCTGCGTATGCA  |
| MA01-3      | A        | 23,338,750     | MA02-4c        | CGGAGGTCACGCTGCAAGG  |
| MA01-4      | G        | 36,366,542     | MA02-4c        | CGGAGGTCACGCTGCAAGG  |
|              |          |                | MA02-4c        | CGGAGGTCACGCTGCAAGG  |
| MA02-1      | C        | 207,635        | MA06-1a        | CAGACACGCTGCGTATGCA  |
| MA02-2a     | G        | 17,601,653     | MA04-1c        | CAGACACGCTGCGTATGCA  |
| MA02-2b     | C        | 17,601,653     | MA04-1c        | CAGACACGCTGCGTATGCA  |
| MA02-2c     | G        | 21,945,522     | MA02-3a        | CAGACACGCTGCGTATGCA  |
| MA02-3a     | A        | 21,945,522     | MA02-3a        | CAGACACGCTGCGTATGCA  |
| MA02-3d     | C        | 21,945,522     | MA02-3d        | CAGACACGCTGCGTATGCA  |
| MA02-4b     | C        | 25,084,281     | MA02-4b        | CAGACACGCTGCGTATGCA  |
| MA02-4c     | A        | 25,084,281     | MA02-4c        | CAGACACGCTGCGTATGCA  |
| MA03-1      | T        | 445,916        | MA03-2         | CAGACACGCTGCGTATGCA  |
| MA03-2      | G        | 18,685,495     | MA03-2         | CAGACACGCTGCGTATGCA  |
| MA03-3      | C        | 32,622,177     | MA03-3         | CAGACACGCTGCGTATGCA  |
| MA03-4      | unknown  | 25,084,281     | MA03-4         | CAGACACGCTGCGTATGCA  |
| MA04-1      | C        | 1,645,815      | MA04-1         | CAGACACGCTGCGTATGCA  |
| MA04-1c     | C        | 1,645,815      | MA04-1         | CAGACACGCTGCGTATGCA  |
| MA04-2      | G        | 17,993,699     | MA04-2         | CAGACACGCTGCGTATGCA  |
| MA04-3      | A        | 27,545,988     | MA04-3         | CAGACACGCTGCGTATGCA  |
| MA04-4      | G        | 32,622,177     | MA04-4         | CAGACACGCTGCGTATGCA  |
| MA05-1      | T        | 125,882        | MA05-1         | CAGACACGCTGCGTATGCA  |
| MA05-2      | G        | 18,134,501     | MA05-2         | CAGACACGCTGCGTATGCA  |
| MA05-3      | A        | 31,705,135     | MA05-3         | CAGACACGCTGCGTATGCA  |
| MA05-4      | T        | 32,942,944     | MA05-4         | CAGACACGCTGCGTATGCA  |
| MA06-1      | G        | 1,016,816      | MA06-1         | CAGACACGCTGCGTATGCA  |
| MA06-1a     | A        | 1,017,001      | MA06-1a        | CAGACACGCTGCGTATGCA  |
| MA06-1b     | C        | 1,016,657      | MA06-1b        | CAGACACGCTGCGTATGCA  |
| MA06-2      | G        | 16,107,284     | MA06-2         | CAGACACGCTGCGTATGCA  |
| MA06-3      | A        | 26,797,986     | MA06-3         | CAGACACGCTGCGTATGCA  |
| MA06-4      | T        | 32,153,016     | MA06-4         | CAGACACGCTGCGTATGCA  |
| MA07-1      | C        | 133,624        | MA07-1         | CAGACACGCTGCGTATGCA  |
| MA07-2      | C        | 3,585,834      | MA07-2         | CAGACACGCTGCGTATGCA  |
| MA07-3      | C        | 10,473,531     | MA07-3         | CAGACACGCTGCGTATGCA  |
13 MA13-2 3,395,104 T C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GAGTCAGTGAAAGCTTCTTCTG
13 MA13-3 23,714,499 T C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GTTCAGGAAACATTCCTCAGTGG
13 MA13-4 32,775,074 A G ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG CTAGGAAACATTCCTCAGTGG
14 MA14-1 265,374 C T ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG CAGATACATTTCAGTCTG
14 MA14-2 8,785,003 G A ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG TAAGCTGAGAACCCCTCATAAAG
14 MA14-3 24,986,580 T C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGACAGGAAACATACG
14 MA14-4 30,282,384 A T ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGAGTCAGTGTTGCTTACCT
15 MA15-1 43,581 C G ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGTAAGG
15 MA15-1a 44,072 G A ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG CAGGGGTGTCAACATA
15 MA15-2 14,798,836 T C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG ATGGGAGTCTGAGG
15 MA15-3 27,027,724 T C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG CCGGACTTTAAGGTCGCGCGGG
15 MA15-4 30,336,373 A T ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGAGTCAGTGTTGCTTACCT
16 MA16-1 4,141,407 A C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GCAAGTCAACAGGTA
16 MA16-2 20,435,475 A T ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGAACATTTATTACAGGAAAAG
16 MA16-2a 20,435,359 G A ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG AACACTCATTACTGACACATGT
16 MA16-2b 20,435,359 T C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG ATGGGAGTCTGAGG
16 MA16-3 28,965,153 A C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG CCGGACTTTAAGGTCGCGCG
16 MA16-4 32,748,201 G T ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGAGTCAGTGTTGCTTACCT
17 MA17-1 569,122 C T ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GCGCCCATACCTCCATATAC
17 MA17-2 17,687,463 T A ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG cccGGACACATTGACAGAACAT
17 MA17-3 25,940,819 A C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GCCCTCTGAGGTCGCTCTT
17 MA17-4 31,650,745 A G ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GCGGACTTTAAGGTCGCGCG
17 MA17-4c 31,650,869 A C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGGGATGTCAGTCACTGCTT
18 MA18-1a unknown T C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGGGATGTCAGTCACTGCTT
18 MA18-1c unknown G A ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG TATACCTTCTTCTCTCTTCT
18 MA18-2 4,497,261 T G ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGTGTCTGAGGTCGCTGTAC
18 MA18-2a 4,496,941 T C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GCCTGATACGACGAAAA
18 MA18-3 16,910,740 G A ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGGGATGTCAGTCACTGCTT
18 MA18-4 30,484,389 G G ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG CCTGCAACATACGAC
19 MA19-1 402,674 C T ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG CCTCCATAGGTACTGAGGTC
19 MA19-2 4,908,457 A T ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGAGGCGGGCTGTTGCTGCG
19 MA19-3 12,645,659 A G ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG TGTCGCTCTGAGGGAAG
19 MA19-4 24,786,905 T C ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG CCCACATGTCAGGCTG
20 MA20-1 410,661 A G ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG GGGGATGTCAGTCACTGCTT
20 MA20-2 14,179,255 T G ACGTTGAGTGCTTGGACAGGTTAAGG ACGTTGAGTGCTTGGACAGGTTAAGG TGGCCATCCGAAACAA
| Location | Sequence ID | Start Position | End Position | Base 1 | Base 2 | Sequence 1 | Sequence 2 |
|----------|-------------|----------------|--------------|--------|--------|------------|------------|
| 20 MA20-3 | 20,498,732  | G C             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | CGCCAAACATCGTGAGGAACAT |
| 20 MA20-4 | 25,891,725  | G A             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GAGGGAGGGGTGCCGACTAA |
| 20 MA20-4b | 25,891,212  | C G             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | TCCACTCGCTGGCATT |
| 20 MA20-4c | 25,891,060  | T G             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | ATAACTCCATTGGAGATCTTTT |
| 21 MA21-1 | 100,766     | A G             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | TTAACAAAGGAAGACTGTTAAAACAC |
| 21 MA21-2 | 4,290,747   | A T             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GAGGGAGGGGTGCCGACTAA |
| 21 MA21-3 | 23,999,437  | C A             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCAAATGGGATTCCAGGAG |
| 21 MA21-4 | 29,508,028  | G A             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGCAGGAATCAGCAACAGAG |
| 21 MA21-4a | 29,508,743  | C A             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | ATACCTCCATTGGAGATCTTTT |
| 21 MA21-4b | 29,508,564  | A G             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GCGCAAGGAATCAGCAACAGAG |
| 22 MA22-1 | 232,558     | G T             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GAGGGAGGGGTGCCGACTAA |
| 22 MA22-2 | 3,938,426   | G A             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 22 MA22-3 | 20,007,790  | C G             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 22 MA22-4 | 26,629,150  | C T             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 23 MA23-1 | 257,391     | A G             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 23 MA23-2 | 6,893,492   | G C             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 23 MA23-3 | 20,116,004  | T C             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 23 MA23-4 | 24,363,992  | A T             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 23 MA23-4a | 24,363,433  | A G             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 24 MA24-1 | 1,102,453   | C G             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 24 MA24-2 | 5,886,699   | T C             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 24 MA24-3 | 17,144,745  | G C             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |
| 24 MA24-4 | 23,366,242  | G C             | ACGTTGGATGCGCTGGAGGCTATGGAATG | ACGTTGGATGCGCTGGAGGCTATGGAATG | GGGCCACCCCAAGGATAGG |