Progress in Medaka Genomics

Kiyoshi Naruse

Graduate School of Science, University of Tokyo

That you for giving the opportunity to talk about the merits of medaka.

Medaka have been used as specimens in scientific experiments since the beginning of the previous century. The name "medaka" and the fish itself are common familiar to most people, especially in Japan. Not that much is however known about the fish's genome structure and genomic information. Today I would like to talk about how much progress has been made in medaka genome analysis and what can be seen when we flash the light of genomic analysis and information on sex reversal in medaka, the most classic example of endocrine disruption in vertebrates.

This shows the genetic structure of a wild population of medaka. There are four genetically different wild populations of medaka, which are respectively known as the North Japan, South Japan, East Korea and China – West Korea populations. Of the four populations, the North Japan and South Japan populations are the most closely related. If you however compare the genomic DNA base sequences of the two populations, approximately 1% of the amino acid is in the coded area and about 3% is in the uncoded area. This difference is perhaps comparable to the difference between human beings and chimpanzees, or perhaps a little more. Another characteristic is the fact that populations can be mated without restriction despite the large genetic difference between the two populations. The inbred strain from each of the four groups has been established, which enables a large number of individuals as a genetically uniform strain. This is the most effective and useful feature of medaka when the fish is hadled genetically.

First let's compare distinguishing features of genomes. This column contains data of medaka. The data is currently applicable for comparison with human beings and zebra fish, for which genomic analysis has progressed most among all smaller fish. To put genomic features in simple terms, if the size of a medaka genome were said to be "1," a zebra fish would be "2," and a human being twice that amount, or "4." As for number of chromosomes, medaka have 48, human beings 44 + XY, and zebra fish 50. The sex determination system for medaka and human beings is XX – XY; the determination system for zebra fish is unknown. We think that there may be no major sex determination genes.

We are comparing three types of features possessed by genomes. For the first experiment we conducted for comparison, we conducted so called EST analysis of the genetic repertory of medaka. I'm sure many of you are familiar with EST analysis, but to put it simply, a cDNA library is prepared, clones are randomly picked out from the library and the base sequence is usually determined from both 5' and 3'. This time we read the sequence only from the 5' end. We first searched the existing databases for the sequence using a program called "Blast" to see if the determined base sequence is homologous with already registered DNA data of a different type.

At the same time, duplicated genes are taken, so the databases are searched for duplicated clones using a program called "Cluster W," the duplicates are eliminated, and ultimately data sets with different sequences for which there are no duplicates are prepared. This is available to the public as a database linked to a Web site. We are now in the process of creating a database called "M Base," which can be accessed by anyone at http://mbase.bioweb.ne.jp/~dclust/medaka_top.html. Not only is the database available to the public, but it also handles distribution of clones.

Here are the results of EST analysis conducted a little more than a year ago. As a result of reading approximately 8000 clones originating from five libraries and conducting cluster analysis, we get approximately 4600 clusters, i.e., we have identified approximately 4600 unique genes. Although there is some debate, you can more or less think of it like that. If you are wondering what sort of libraries we have

created, the library called “OLa” originates from the male and female body. The other two are created from cultivated medaka cells. The rest are created from the ovaries and a liver.

If you conduct a search with the Blast program, approximately half of these 4600 clusters, or about 2000 base sequences exhibit homogeny with other cDNA sequences in existing databases. Of the 2000, we selected the ones most homogenous with human beings and analyzed where they were located in the genome using a technique called “gene mapping.”

This is the EST mapping method. First we synthesize primer based on the EST sequence or cDNA sequence. We then conduct PCR using genomes DNA of an inbred strain of the previously mentioned South Japan population called “AA2” and genome DNA of an inbred strain of a North Japan population called “HNI” as a template. Next we select a primer with a single band increasing. The increased DNA pieces are cut by restriction enzyme, and a search for so called RFLP is conducted. Using this RFLP as a marker, each individual is gene typed with a back-cross panel, and the linkage of each marker is decided from the typing pattern.

This is an example of primer screening. The name of the EST sequence is here. Primer is synthesized from the respective EST sequences. With cDNA, primer is created by separating approximately 300 base pair. Because there is intron, larger DNA pieces or those of approximately the same size are amplified. Thus we select primer with only one band increasing. With this primer, the lengths of DNA pieces amplified by AA2 and HNI genomes are different. This is surmised to be due to difference in intron size between the systems. We cut those with same amplified size by eight types of restriction enzyme and check if they exhibit RFLP. There is a difference in the way all DNA pieces amplified by this primer are cut by restriction enzyme. Even if you look at this, you will see just how different genome sequences of North Japan and South Japan medaka are.

This is a linkage map created in this way. A little while ago I said that medaka have 48 chromosomes, but we have obtained linkage groups with 24, half that amount. I think that each of the linkage groups corresponds to a chromosome. This data is a bit old, but about 500 genes are mapped here. The next figure shows a comparison of this gene map and one of a human being. The horizontal axis represents human 1 – 22 and XY chromosomes; the vertical axis represents medaka linkage groups. With linkage group 1 (LG1) as example, we are studying on which number of human chromosomes the genes of medaka LG1 go on. As we mark them with dots, we begin to see the area where the dots are concentrated.

This shows that multiple homogenous genes are on the same linkage group or on the same chromosome in the case of medaka and human beings. We call this “maintaining synteny”; we can see several areas where synteny is maintained. The reason for this is still unclear, but you can clearly see that synteny is maintained for the linkage group carrying the Hox gene.

The example we mentioned a little while ago is an example where 377 genes are compared with each other. Let’s take a closer look at this in a certain linkage group. This is the result obtained during the process of positional cloning of a mutant called “RS3.” RS3 is a mutant of medaka that has no scales. When we mapped this mutant, we found it was linked with the Hoxd gene of LG21. In the case of human beings, the Hoxd gene would be on the 2q chromosome. A gene called “nebulin” was obtained when conducting EST analysis was coincidentally contained in this area, so we mapped it. We then found that the nebulin gene was similarly on LG21 of medaka.

Thus, thinking that genes may be preserved to an extreme degree in LG21 of medaka and 2q of human beings, or in other words, thinking that synteny is maintained in LG21 of medaka and 2q of human beings, we took a sample of the gene on 2q from the human beings and mapped it on medaka. We then found that, although some of it went to different linkage groups, a large potion went on LG21 and LG2. In the case of zebra fish, these genes are on LG9. This means that synteny is maintained to a large degree in this interval.

We ultimately found the casual gene of RS3 mutant with no scales is EDAR, which is a member of the TNF receptor super gene family, and we furthermore found it contained a mutant. In other words, we found out which gene is the casual gene. As for the EDAR gene, mutants are known for mice and human beings. The phenotype in this case is absence of hair.

In other words, the same genetic mutant in the case of human beings or mice would be characterized by absence of hair, and in the case of medaka, would be characterized by absence of scales. This tells us that medaka and other fish may be distant relatives of human beings. A footnote to this however would be the order of the genes is messy, and in a sense, is not maintained at all. This oppositely occurs frequently; compared with this, translocation does not exist to a large extent. If you consider various aspects, there are of course many dissimilarities, but there are also quite a few similarities between medaka and human beings.

This is the current status of medaka genome analysis. We tried applying the information obtained by such analysis to the classic phenomena of sex reversal in medaka. Artificially induced sex reversal in medaka was studied by Prof. Tokio Yamamoto in 1953. Prof. Yamamoto used the d-rR strain mentioned yesterday. In his experiment, the males were red and the females were white. In other words, the strain was such that the genes that determine red are on the Y chromosome. If the specimens of this strain are fed ordinary food, males and females are produced proportionally (1 to 1), with all males being red and all females being white.

When fish of this strain are fed food containing estrone, the fish are more or less produced at the ratio of 1 to 1 in terms of red to white, but red females appear. In other words, genetically speaking, XY, that is those that are supposed to become males, when fed food containing estrone, became females. In a certain sense, this could be thought of as a typical case of endocrine disruption. This sort of phenomenon was discovered. We were lucky to discover such a phenomenon in a number of senses. This slide illustrates what I mean. The slide shows difference in the rate of sex reversal according to strain. As for the d-rR strain we talked about a little while ago, this is the rate of sex reversal of the Hd-rR inbred strain established from the d-rR strain. This is the rate of sex reversal of inbred strain called “HNI” originating in North Japan medaka. The rate of sex reversal is measured by immersing eggs in water containing 0.25 μ g/ml of E2. If the eggs are bred until maturity, approximately 90% of the individuals of the Hd-rR strain reverse sex and change to females, just as Prof. Yamamoto’s experiments. There was almost no sex reversal for the HNI strain on the other hand, even when treated in the same manner with the same concentration. The rate of sex reversal was about 5%.

Thus we can say that if Prof. Yamamoto had used North Japan medaka, his sex reversal experiment would not have gone well. We must therefore raise the concentration of E2, but if we raise the concentration, we still will not get individuals with sex reversal, but rather the embryos of both strains will die.

We must therefore find out where the gene that produces this phenomenon, in other words, this difference is. For this analysis, we created a congenic strain. By congenic strain, we mean a strain having only the Y chromosome of another strain such as HNI in an Hd-rR background. The black part is a Y chromosome originating from HNI; the entire Y chromosome originates from HNI. Only a portion of the Y chromosome of the next strain originates from HNI. Even a smaller area of this strain originates from the Y chromosome of HNI. The gene that determines the male sex of medaka is actually in this area.

This green part is the East Korea population mentioned a little while ago, and is the sex chromosome map of the congenic strain having a Y chromosome originating in the East Korea population. This is the sex chromosome map of the congenic strain having a Y chromosome originating in the China – West Korea population. It takes back mating of about ten generations to create a congenic strain. Each generation lasts about three months, so it is quite a job just to create the various strains.

Using these congenic strains, we studied how the rate of sex reversal differs among the various strains. This is the same slide we saw a little while ago. The pink bar represents the rate of sex reversal for the Hd-rR strain. The black bar represents the rate of sex reversal for the HNI strain. Although most of the genes originate in the Hd-rR strain, the white bar represents the rate of sex reversal for the strain whereby only a small portion of the Y chromosome originates in the HNI strain. If you compare the pink and white bars, you can see that the rate of sex reversal varies dramatically according amount of E2. On the other hand, even if it has a Y chromosome of the East Korea or China – West Korea population, sex reversal occurs frequently but efficiency differs somewhat from that of Hd-rR.

Now let's take a close look at the map of the medaka sex chromosome. The fact that the strain represented by the white bar has about the same sensitivity to estrogen as the HNI strain is because this area originating in the Y chromosome of the HNI strain controls sensitivity to estrogen. Under ordinary circumstances, sensitivity to estrogen would probably be controlled by the estrogen receptors, or ER. We currently know of three types of ER in medaka: α, β and γ. We currently have data on ER α and ER γ but they are not in this area. We still do not know about ER β, but Professors Nagahama and Sakaizumi's group is conducting positional cloning of genes that determine sex. The base sequence has almost been completely determined. In other words, genomic analysis of this area is almost complete.

We are studying the genomic sequence of this area, but we have yet to find a gene that acts like ER. In conclusion, we can say that ER has nothing to do with sensitivity to estrogen. Anyway, it is possible. Then what is controlling sensitivity to estrogen? We still do not have enough data, but the polymorph of the gene that determines sex itself may decide sensitivity to estrogen. Experiments are currently being conducted to determine this.

Although it is little bit early, corroborators and contributors have instructed me not to give their names, so I will close my presentation here. Finally I would like to say that human beings, zebra fish and Japanese puffer fish appearing on three slides were the source of genomic analysis conducted of sequence level. We are determined to use medaka to conduct genomic analysis of sequence level. Even so, it takes quite a bit of money to conduct genomic analysis of sequence level, and if you take the speed at which research at large is conducted, the genome sequence itself does not take that much time. I would like to end my presentation with request for spiritual and financial backing of medaka genome sequence from the standpoint of propagating original Japanese genomic information. Thank you very much.