Fish genomes flying

Symposium on Medaka Genomics

Akihiro Shima¹, Heinz Himmelbauer², Hiroshi Mitani¹, Makoto Furutani-Seiki³,
Joachim Wittbrodt^{4*} & Manfred Scharf⁵

¹University of Tokyo, Kashiwa City, Japan, ²Max Planck Institute of Molecular Genetics, Berlin, Germany, ³ERATO, Kondoh Differentiation Signalling Project, Kyoto, Japan, ⁴European Molecular Biology Laboratory, Heidelberg, Germany, and ⁵University of Würzburg, Würzburg, Germany

More than 50 researchers from all over the world gathered at EMBL in Heidelberg, Germany, to participate in the first meeting of the Medaka Genome Initiative. The symposium was held from 31 July to 1 August 2002.

EMBO reports 4, 121–125 (2003)
doi:10.1038/sj.embor.embor743

Introduction

Fish models have become very popular during the past decade (see Fig. 1 for phylogenetic relationships). Recently established models for genomic studies (the Japanese and freshwater pufferfish, *Fugu rubripes* and *Tetraodon nigroviridis*, respectively) and for developmental biology (the zebrafish, *Danio rerio*) are now the focus of interest of many biologists, although 'old' models are still useful for the study of tumorigenesis (platyfish, *Xiphophorus*) and evolution, radiation and speciation (African cichlids), as well as for the evolution of sex determination and developmental genetics (the medaka, *Oryzias latipes*). Although the zebrafish and its features as an

experimental system are well known among scientists within and outside the field, knowledge of the medaka has so far been restricted (with a few exceptions) to its home range, Japan, and some other countries in the Far East.

In Japan, the breeding of medaka colour variants for ornamental purposes is a tradition that goes back several centuries. In the early twentieth century, the medaka was recruited as a model species for biological and genetic research. In fact, it was the study of the inheritance of body colour in the medaka by Toyama in 1916 that proved that Mendelian laws also apply to fish (Toyama, 1916). Another milestone achievement was the discovery of Y-chromosome-linked inheritance by Aida in 1921 (Aida, 1921). Within the past seven years, the number of publications on studies that involve the medaka have doubled each year, illustrating its increasing popularity as a model species.

One of the scientific highlights of 2002 was the release of the draft version of the first whole-genome sequence of a teleost, that of the pufferfish *Fugu rubripes*. An advanced genome draft sequence is also now available for the related pufferfish species *Tetraodon nigroviridis*. However, neither species is used as an experimental system; the primary value of each is its exceptionally small genome size, which is an advantage in genome analysis. Phylogenetic studies place the medaka close to both pufferfish species, with an estimated divergence time of 60–80 million years (Myr) ago, which is less than the evolutionary distance between man and mouse. Medaka and zebrafish, now the most widely used fish model species, are relatively distant cousins that have evolved separately for at least 135 Myr. The key technologies that have made the zebrafish such a successful model species are fully applicable to the medaka. However, genomics in the medaka offers several advantages, such as the availability of divergent, perfectly inbred strains and a genome of only 800 million bases, half the size of the zebrafish genome. In addition, a practical aspect is that the sexes are easily distinguished in medaka, in contrast to zebrafish, facilitating breeding in general and genetic studies in particular.

The recent symposium, Medaka Genomics 2002, was inaugurated with an introductory talk by A. Shima (Kashiwa City, Japan), who provided a historical overview of the Medaka Genome Initiative. The inception of this project was a European-Union-sponsored practical

¹University of Tokyo, School of Science, 5-1-5 Kashiwanoha, Kashiwa City 277-8562, Japan

²Max Planck Institute of Molecular Genetics, Ihnestr. 73, 14195 Berlin, Germany

³ERATO, Kondoh Differentiation Signalling Project, 606-8305 Kyoto, Japan

⁴European Molecular Biology Laboratory, Developmental Biology Programme, Meyerhofstr. 1, 69012 Heidelberg, Germany

⁵University of Würzburg, Biocenter, Am Hubland, 97074 Würzburg, Germany

*Corresponding author. Tel: +49 6221 387 576; Fax: +49 6221 387 166; E-mail: jochen.wittbrodt@embl.de

Received 7 November 2002; accepted 23 December 2002

course held at the Biocentre of the University of Würzburg in 1993, and it has successfully developed since then (Wittbrodt *et al.*, 2002), with the underlying philosophy that a genome project is much more than whole-genome sequencing. This symposium attempted to give an overview of the progress that has been made towards both the analysis of the entire genome of the medaka and an understanding of gene function in this model organism. It was also designed to foster interactions and exchange, and to define priorities for rapid progress towards the common goal of understanding how its genome works. The talks demonstrated the broad spectrum of expertise that has been united in this initiative, with topics ranging from large-scale mutagenesis screens and the current status of the mapping and sequencing projects, to transgenic approaches, functional genomics and the evolutionary analysis of selected genes (see Fig. 2). It was impressive to see how much the knowledge and technology in the field have advanced through the efforts of a relatively small community, owing to the open, friendly and truly interactive nature of the initiative and the understanding that it is the concerted generation of resources that will eventually contribute to our understanding of how a genome works.

Mutagenesis and genomic resources

In vertebrates, genomic functions are often safeguarded by functional redundancy. This is especially true in teleost fish, in which the genome seems to have been duplicated at the base of the clade's radiation. Thus, a mutagenesis screen in one species might not suffice to uncover all potential gene functions. The medaka was chosen to complement the zebrafish and other vertebrate species in systematic mutagenesis screens for determinants of embryonic patterning (Loosli *et al.*, 2000) and other aspects of development, organogenesis and behaviour because it is evolutionarily closer to the genomic model system *Fugu* than to zebrafish and its genome is only half the size of that of zebrafish. The availability of different inbred strains and an established genetic map are specific advantages in using this species. M. Furutani-Seiki (Kyoto, Japan) gave an update on the status of a continuing international large-scale mutagenesis project in

which more than 12 groups are participating. The classical F_3 screen (study of mutants in the third generation after mutagenesis; Fig. 3), uses *in situ* hybridization and antibody staining in addition to morphological characterization, and the frequent occurrence of a temperature-sensitive phenotype is an additional benefit. The international team has so far analysed 628 F_2 families, obtained 1,505 embryonic lethal mutations and identified 352 of those causing specific embryonic patterning defects. In the mutants, many specific structures and organs are affected, including the early axis, central nervous system, eye, sensory neural network, developing gonads and thymus. Whereas several phenotypes are strikingly similar to those of known zebrafish mutants, many medaka mutants exhibit novel phenotypes to those that so far have been uncovered in zebrafish. Many medaka mutants have already been identified. Thus, genetic and physical mapping approaches are of crucial importance for isolating the corresponding genes.

Physical maps based on bacterial artificial chromosome (BAC) clones are key tools for both the structural and functional characterization of individual genes. In addition to providing templates for the genomic sequencing of complex genomes, BACs can be used for the positional cloning of genes that cause particular mutant phenotypes, for the rescue of mutants once candidate genes have been mapped to them, and for the establishment of links between physical, genetic and cytogenetic maps through fluorescence *in situ* hybridization studies. Another strategy, the whole-genome shotgun (WGS) approach, is complementary to the BAC approach in terms of whole-genome sequence analysis. With WGS, it is straightforward to produce a large body of sequence information, but assembly can be a major hurdle in repeat-rich regions of the genome. In contrast, BAC sequencing is hampered by the necessity to generate, and perform quality control on, multiple shotgun libraries, but reduces the assembly problem by reducing complexity. Proposals to integrate these strategies include: first, large-scale preparation of BAC end-sequences with which to put contiguous WGS sequences (contigs) into the context of larger scaffolds, and second, the sequencing of BACs at low coverage, followed by the detection of matching sequences from a pool of WGS reads by

Fig. 1 | Evolutionary relationships in various organisms that are used as model systems to study processes such as development, disease, sex determination, tumorigenesis and evolution. The fish shown is an adult medaka male.

Fig. 2 | Flow chart illustrating the interrelations of different approaches and data sets that are integrated in the Medaka Genome Initiative and associated with it. Beige represents the current activities of the Medaka Genome Initiative; blue represents the whole-genome shotgun approach that has been initiated. Both projects are merging efforts to release soon the first draft of the medaka genome. BAC, bacterial artificial chromosome; EST, expressed sequence tag; FISH, fluorescence *in situ* hybridization; RH, radiation hybrid; WGS, whole-genome shotgun.

using appropriate software. H. Himmelbauer (Berlin, Germany) reported on progress towards the generation and completion of a BAC contig map of the medaka genome, based on his previous experience and on technology developed for a yeast artificial chromosome/BAC map of the mouse. The medaka BAC map will encompass segments of all the available BAC libraries (northern strain HNI and southern strains CAB and Hd-rR) and will therefore permit comparisons across strains and genotypes. A total of 10,000 probes are now being mapped to BACs, representing a 15-fold coverage of the medaka genome. Probes are being derived from medaka genes, expressed sequence tags (ESTs) are being generated and genetically mapped in the University of Tokyo Medaka Genetic Mapping Project (H.M., K. Naruse and A.S.) and pairs of end probes are being generated from CAB and Hd-rR BACs and sequenced at Keio Medical School (S. Sasaki, S. Asakawa and N. Shimizu). Additionally, information on synteny (the location of genes on the same chromosome or chromosome segment) from the recently completed *Fugu* and *Tetraodon* genome sequences is being used in the rapid generation of contigs across large intervals. This project is expected to provide a deeper understanding of the structure and large-scale organization of the medaka genome and of the genes encoded therein.

Genetic linkage maps have been used for various kinds of analysis, such as positional cloning, quantitative trait locus analysis, comparative vertebrate genomics and the detection of radiation-induced DNA mutations. H. Mitani (Kashiwa City, Japan) reported that, in an effort to expand the existing map, the group at the University of Tokyo has derived 2,400 pairs of mapping primers based on ESTs that have a significant degree of similarity to previously identified genes from other species. About 800 of these were polymorphic between the mapping strains and thus were incorporated into the existing genetic map. Studies of the zebrafish genome had determined that the conservation of synteny between the zebrafish and human genomes is extensive. Given the evolutionary relationships between humans, zebrafish and medaka, this synteny was also expected to be conserved between the medaka and human chromosomes, and this has now been confirmed by a comparison of linkage relationships between orthologous genes in these three organisms. Although some conserved syntenies are extensive, even conserved groups of genes can have changes in gene order, apparently reflecting the relatively frequent occurrence of inversions and other intra-chromosomal rearrangements. In most analysed medaka and zebrafish linkage groups, the distribution of

duplicated genes suggests that two chromosomes in the medaka and the zebrafish are derived from each ancestral chromosome. This and extra-Hox clusters in the teleost lineage reported previously support the argument that a 'whole-genome duplication' was the basis of the teleost radiation.

Two laboratories are working on large-scale whole-mount *in situ* screens to complement the mapping approaches discussed above. So far, randomly picked cDNA clones from libraries generated from different developmental stages have been analysed by sequencing and whole-mount *in situ* analysis by both laboratories, but with a slightly different emphasis: whereas the Bourrat group (F. Bourrat, Paris, France) focuses on genes expressed in the proliferative area of the optic tectum by screening brain-specific libraries of corresponding stages (Nguyen *et al.*, 2001), the Wittbrodt group (J. Wittbrodt, Heidelberg, Germany) is interested in early patterning markers and genes expressed in the developing and differentiating eye. In total, the two groups have identified more than 1,000 non-redundant genes and characterized their expression patterns, and most of the information is available in a publicly accessible database generated in collaboration by J. Wittbrodt and T. Henrich (Kyoto, Japan) (Henrich *et al.*, 2002). The EST information is linked to the Medaka EST database, Mbase (A. Shima and H. Mitani), to allow direct access to an expression pattern from an EST sequence in which the pattern is already established. This database will be extended to 5,000 genes by the end of 2003, and a normalized Unigene set of more than 15,000 individual genes has been established (H. Himmelbauer). Further extension to include all the EST clusters generated by the Tokyo group should make it possible to complete the expression analysis of (eventually all) medaka genes in a collaborative effort using established resources. The whole-mount expression analysis will be complemented with quantitative data derived from cDNA microarray analysis, which will be performed in parallel with the same resources.

Insights from the outside

Autonomous retrotransposable elements certainly have a major role in the evolution of eukaryotic genomes, and might have been involved in generating the high genetic diversity observed in fish. Such elements can duplicate and integrate into new genomic locations in which they might disrupt genes, recombine ectopically to produce genomic rearrangements, modify the expression patterns of flanking genes, and duplicate different types of non-autonomous sequence. Retrotransposable elements can be used as molecular tools for genome analysis, but can also be a source of problems during the assembly of genome projects. J.-N. Volff (Würzburg, Germany) reported that recent analyses, particularly those involving medaka, have strongly suggested that fish genomes contain an amazing diversity of active retrotransposons that is not observed in mammals. Retrotransposable elements atypical of the vertebrate lineage were first identified in the medaka. These include the first vertebrate retrotransposons found to encode a restriction-enzyme-like endonuclease or an endonuclease related to certain mobile DNA introns. Numerous other elements, including retrotransposons with or without long terminal repeats and short interspersed nuclear elements (SINEs), are also present in the genome of medaka. Thus, the medaka has the potential for development into an outstanding model for functional and evolutionary studies of retrotransposable elements that are absent from the genomes of higher vertebrates.

M. Nonaka (Tokyo, Japan) discussed evolutionary aspects that have emerged from his extensive study of the structure of the medaka's major histocompatibility complex (MHC). This is a genomic region that harbours several genes essential for adaptive immune recognition (class I antigen presentation) and is specific to jawed vertebrates. The MHC region serves as a model for the survey of the evolution of the vertebrate genome. In contrast to its centralized structure in cartilaginous fish (Chondrichthyes), amphibians, birds and mammals, in the medaka and other teleosts the MHC genes are dispersed over several chromosomal loci, suggesting extensive genomic reorganization in this lineage. However, several genes involved directly in class I antigen presentation form a tight cluster in the medaka genome, and although the products of these genes are structurally unrelated, they are intimately linked at a structural level, indicating that adaptive processes have maintained the linkage between these genes throughout vertebrate evolution.

Although the genes that determine the development of the male or female sex are known in *Caenorhabditis elegans*, *Drosophila melanogaster* and most mammals, the identity of the sex-determining factors in several species remains unknown. M. Scharl (Würzburg, Germany) reported the results of the joint efforts of his group, together with the teams of N. Shimizu and A. Shima, in cloning the medaka sex-determining gene. They found that the Y-chromosome-specific region spans only about 280 kilobases, a region that has been entirely sequenced, and that it contains a duplicated copy of the autosomal *DMRT1* gene, named *DMRT1Y*. This is the only functional gene on this chromosome segment and it maps precisely to the male sex-determining locus. The gene is expressed during male embryonic and larval development and in the Sertoli cells of the adult testes. These features make *DMRT1Y* a highly likely

Fig. 3 | Example of an F₃ mutagenesis screen scheme. Males were mutagenized with *N*-ethyl-*N*-nitrosourea (ENU) and crossed with wild-type females. Mutations induced by the treatment are eventually uncovered in the F₃ generation after random intercrosses within the F₂ families. Carriers for mutant genes are highlighted in red.

candidate for the medaka male sex-determining gene (Matsuda *et al.*, 2002; Nanda *et al.*, 2002).

Novel technologies

Numerous vertebrate genes that have been identified by various approaches now await functional characterization, ideally in mutants. Besides classical forward genetics, gene-driven approaches will be necessary to achieve this ambitious goal. Mouse knockout technology is not suitable on a genome-wide scale, but the high efficiency of chemical mutagens such as ethyl nitrosourea (ENU) can be used to establish an alternative reverse-genetics approach. The screening of mutated F_1 animals for alterations in a gene of interest has been performed successfully for various vertebrate species. Although, of these, the fish systems have proved particularly well suited for the large-scale analysis of mutants because of their numerous offspring and extra-uterine development, an efficient detection method is crucial if this type of screen is to be routinely applied. T. Czerny (Vienna, Austria) presented a novel approach for the identification of nonsense mutations based on the fact that, in contrast to missense mutations, protein truncations are likely to result in a null phenotype. His strategy of identifying premature stop codons, called 'screen-out', enables the rapid identification of loss-of-function mutants. This approach is based on the in-frame fusion of an exonic sequence with a toxin. After transformation, only those bacteria that received a mutant copy of the exon survive. As part of the Medaka Genome Initiative, the group will generate a stock of medaka mutants, which will be preserved by sperm freezing. Application of the screen-out protocol will then permit the large-scale recovery of mutants from this stock.

Using the classical microinjection technique, M. Kinoshita (Kyoto, Japan) has established a transgenic medaka line that expresses green fluorescent protein (GFP) in primordial germ cells. Using the unique 'see-through' medaka line—a multiple recessive pigmentation mutant strain that is transparent even in the adult stage—as the genetic background, the group was able to track primordial germ cells and their differentiated descendants *in vivo*, not only during embryonic and early larval development but also in adult fish (Tanaka *et al.*, 2001; Wakamatsu *et al.*, 2001). Such transgenic medaka can be used for *in vivo* studies of germline development or in the adult stage for the easy detection of endocrine-disrupting substances.

The use of teleost species as genetic model systems is still limited by a low transgenesis efficiency and the strong mosaic fashion in which the injected DNA is distributed in the recipient embryo, even after injection into the one-cell stage. C. Grabher (Heidelberg, Germany) presented two strategies to overcome both problems. One method was developed in close collaboration with J.S. Joly (Paris, France). A GFP reporter gene was flanked with the meganuclease *I-SceI* recognition sites, and co-injected with the *I-SceI* enzyme into medaka embryos (Thermes *et al.*, 2002). Besides efficient promoter-dependent expression in the parental (P) generation, the method facilitates efficient transgenesis, probably owing to integration during early cleavage stages. The high transgenesis frequency and germline transmission rate permit the rapid establishment of transgenic lines with a limited F_1 screening effort. In an alternative approach, a GFP reporter gene controlled by a ubiquitous promoter was flanked by the recognition sequences of the artificially reconstituted *Sleeping Beauty* (SB) transposase (Ivics *et al.*, 1997) and co-injected with SB mRNA. More than 35% of the parental fish screened gave rise to stable transgenic lines. This approach there-

fore represents the most potent transgenesis approach so far. Interestingly, more than 15% of the transgenic animals showed enhancer trapping effects, namely novel patterns that are probably due to the integration of the transgene in the vicinity of enhancer elements. Transgenic lines with restricted expression patterns provide useful tools for a variety of applications, including four-dimensional confocal analysis of a tissue of interest, mutant analysis, enhancer or gene traps, fluorescence-activated cell sorting of labelled cells and the production of specific cDNA libraries, all of which are now being applied.

Perspectives

This exciting meeting provided an excellent opportunity for scientists from around the world to discuss recent progress in the field and, more importantly, to coordinate approaches strategically to future challenges. All of the participants were excited about the prospect that, in a little less than a year, the entire medaka genome will be available and, because of the interactive structure provided by the Medaka Genome Initiative, will be a powerful tool for comparative functional genomics. This is even more likely in view of the recent publication of the *Fugu* and *Tetraodon* genomes. With the release of the medaka and zebrafish sequences expected soon, almost 200 Myr of evolution will be experimentally available in these fully developed model systems, a unique position among the vertebrates.

ACKNOWLEDGEMENTS

We thank E. Fassmann for excellent organization of the meeting, A. Berger for the calculation and drawing of the phylogenetic tree, and F. Loosli, J.-R. Martinez-Morales and M. Kondoh for their critical reading of the manuscript.

REFERENCES

- Aida, T. (1921) On the inheritance of colour in a freshwater fish *Apllocheilus latipes* Temminck and Schlegel, with special reference to sex-linked inheritance. *Genetics*, **6**, 554–573.
- Henrich, T., Ramialison, M., Quiring, R., Wittbrodt, B., Furutani-Seiki, M., Wittbrodt, J. & Kondoh, H. (2003) MEPD: a medaka gene expression pattern database. *Nucleic Acids Res.*, **31**, 72–74.
- Ivics, Z., Hackett, P.B., Plasterk, R.H. & Izsvak, Z. (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell*, **91**, 501–510.
- Loosli, F. *et al.* (2000) A genetic screen for mutations affecting embryonic development in medaka fish (*Oryzias latipes*). *Mech. Dev.*, **97**, 133–139.
- Matsuda, M. *et al.* (2002) *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, **417**, 559–563.
- Nanda, I. *et al.* (2002) A duplicated copy of *DMRT1* in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proc. Natl Acad. Sci. USA*, **99**, 11778–11783.
- Nguyen, V., Joly, J.S. & Bourrat, F. (2001) An *in situ* screen for genes controlling cell proliferation in the optic tectum of the medaka (*Oryzias latipes*). *Mech. Dev.*, **107**, 55–67.
- Tanaka, M., Kinoshita, M., Kobayashi, D. & Nagahama, Y. (2001) Establishment of medaka (*Oryzias latipes*) transgenic lines with the expression of green fluorescent protein fluorescence exclusively in germ cells: a useful model to monitor germ cells in a live vertebrate. *Proc. Natl Acad. Sci. USA*, **98**, 2544–2549.
- Thermes, V. *et al.* (2002) *I-SceI* meganuclease mediates highly efficient transgenesis in fish. *Mech. Dev.*, **118**, 91–98.
- Toyama, K. (1916) On some Mendelian characters. *Rep. Jap. Breed. Soc.*, **1**, 1–19. (In Japanese.)
- Wakamatsu, Y., Pristayzhnyuk, S., Kinoshita, M., Tanaka, M. & Ozato, K. (2001) The see-through medaka: a fish model that is transparent throughout life. *Proc. Natl Acad. Sci. USA*, **98**, 10046–10050.
- Wittbrodt, J., Shima, A. & Scharl, M. (2002) Medaka—a model organism from the Far East. *Nature Rev. Genet.*, **3**, 53–64.