2 Introduction

2.1 Teleosts

The use of fish models has become very popular during the last years for various fields of biological and medical research like genetics, developmental biology, evolution, sex determination, diseases and tumorigenesis. Teleosts such as zebrafish (*Danio rerio*), the pufferfish (*Takifugu rubripes*), and medaka (*Oryzias latipes*) are important model systems for analyzing gene function in vertebrates because of their suitability for forward genetics and their relatively small genome size.

The compact genome of pufferfish allowed molecular studies of human genes. Most human genes originated prior to vertebrate evolution and homology exists in the teleost genomes. The genome size of *Takifugu rubripes* is only 365 Mb (Mega base) (Table 2-1). This makes fugu attractive for genome analysis but unfortunately no experimental tools are available in this species (Table 2-1). In comparison, zebrafish and medaka have larger genome sizes but many experimental tools are available in these species. Of these two, the genomes of medaka is one-half of the zebrafish genome. Furthermore, until now, nothing is known about sex determination in zebrafish and none of the markers of the genetic zebrafish map have been found to be sex-linked. On the other hand, there are highly polymorphic inbred medaka strains available that can be used for both mutagenesis screens and genetic mapping (Wittbrodt *et al.*, 2002). Taken together, these features make medaka an ideal teleost model for genetic research (see Table 2-1).

	Zebrafish	**Medaka**	**Pufferfish**
Genome size	1700 Mb	800Mb	365Mb
Chromosome number of 2n	50	48	44
Sex determination	?	XY type	?
Crossing in laboratories	yes	yes	no
Linkage map	yes	yes	no
The number of mutant strains	2000	120	0
Transgenic fish	yes	yes	no
ES-like cells	yes	yes	no
Inbred strains	none	many	none

Table 2-1: Biological characteristics and availability of experimental tools in three species of teleosts.

2.1.1 Taxonomy

The zebrafish, medaka and pufferfish belong to the subdivision Euteleostei or division Teleostei. Medaka and pufferfish belong to the superorder Acanthopterygii as zebrafish pertains to the Ostariophysi superorder. The evolutionary relationships between fish models show (Figure 2-1) that the last common ancestor of medaka and zebrafish lived more than 110 million years ago. Medaka and fugu are close relatives that diverged 60-80 million years ago. Consequently medaka and fugu are more related to each other than medaka is to zebrafish.

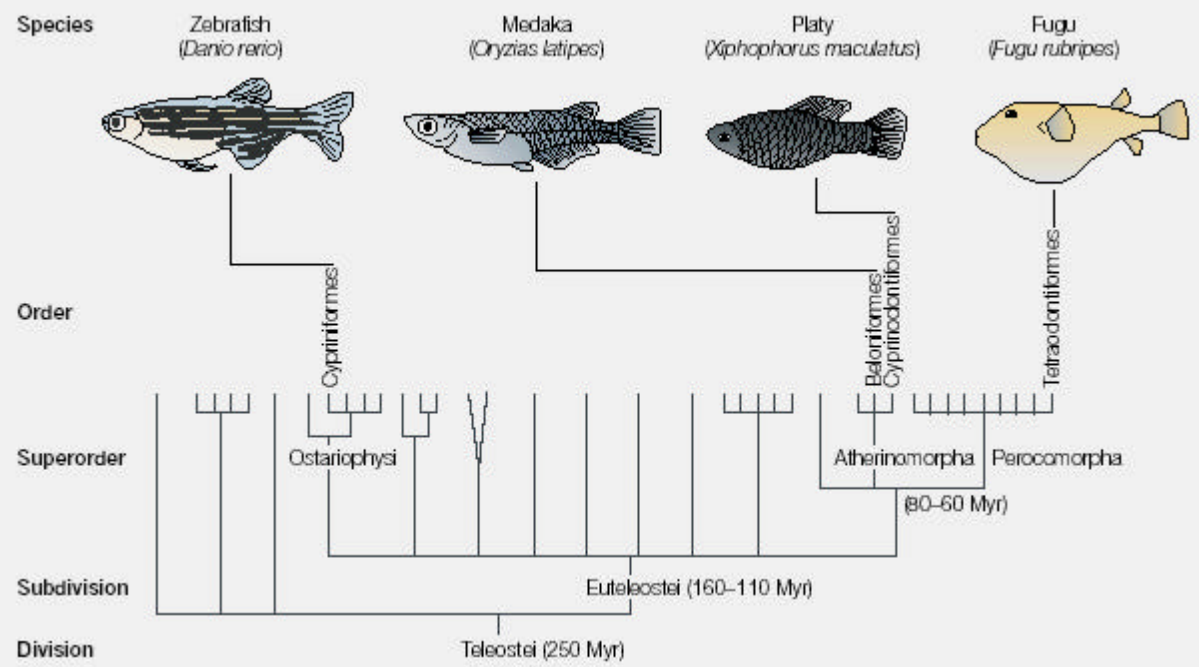

Figure 2-1. Evolutionary relationships between fish models (Wittbrodt *et al.*, 2002).

2.2 Medaka (*Oryzias latipes*)

The Japanese medaka, *Oryzias latipes*, is a small, egg lying, freshwater fish which can be found in Asia, especially in Japan, Korea and eastern China. The adult fish is 2-3 cm long, weighs 0.3 g and can tolerate a wide range (4°C - 40°C) of temperatures (Iwamatsu, 1993). Medaka is easy to breed and resistant against common fish diseases. The female medaka can be easily distinguished from the male by its external morphology and, once fertilized, the female lays a cluster of eggs every day. The spawning correlates with the light cycle and optimal temperature; about 30-50 eggs are spawned every day, and up to 3,000 eggs are produced during a mating

season. The large and transparent eggs are suitable for embryonic manipulation. As the chorion and embryos are transparent, the phenotype can easily be interpreted (Ishikawa, 2000). The eggs hatch 7 days after fertilization and the newly hatched fry grow and will be sexually matured in 3 months. The maximal survival time of medaka maintained outdoors was 1838 days, about 5 years, (Egami and Etoh, 1969) while in the laboratory at 27°C they live one year.

2.2.1 The history of medaka

The breeding of medaka, which in Japanese means a tiny fish with big eyes, has been one of the traditional hobbies in Japan in the last centuries.

The medaka was first scientifically called *Poecilia latipes*. Jordan and Synder renamed this species to *Oryzias latipes* in 1906, which reflects the preferred habitat of medaka, the rice fields (Wittbrodt *et al.*, 2002).

Medaka was the first fish in which the Mendelian laws were proven to be valid (Toyama, 1916). As early as in 1921, Aida showed that the male-determining factor gene (Y) was linked with a pigment gene R and suggested the existence of X and Y sex chromosomes in this fish. It was also the first animal in which complete reversal of sex differentiation in both directions was successfully induced by administration of sex hormones during larval life (Yamamto, 1958). In 1961, Yamamato showed (in Japanese language) crossover between the X and Y chromosomes (Kondo *et al.*, 2001).

Afterwards, genetic studies in medaka have focused on pigmentation and sex determination. In recent years, the study of genes and their functions in medaka has become focus of attention and in this respect different projects are in progress (Wittbrodt *et al.*, 2002; Shima *et al.*, 2003), e.g. ENU-screens to uncover genes important for particular development pathways, mutagenesis screening to establish the condition for systematic mutagenesis, genetic approach to understand gene function (for instance the study of major histocompatibility complex), medaka embryonic stem cells to provide a system for studying cell differentiation.

2.2.2 Geographic variation and strains

The genus Oryzias contains 23 species (http://biol1.bio.nagoya-u.ac.jp:8000/wildtype.html). Out of these, medaka (*Oryzias latipes*) is the only species distributed in the temperate zone.

Until now at least 15 inbred strains of medaka have been established. To establish an inbred strain one male and one female fish born from one pair of parents are selected and mated in each generation. A list of inbred strains of medaka (Table 2-2) was published by Naruse *et al.*, 2004.

The allozyme studies of medaka divided the wild population into four genetically different groups: The North Japan (HNI-I, HNI II, Kaga), South Japan (HO4C, HO5, HB32C, HB12A, HB11A, HB11C, Hd-rR, Hd-rr, Cab and AA2), the East Korean (HSOK) and the China-West Korean population (Naruse *et al.*, 2004). The results of the allozymic analysis have been confirmed by restriction fragment length polymorphism (RFLP) analysis of the mitochondrial DNA of medaka (Sakaizumi *et al.*, 1983; Takehana *et al.*, 2003. The genetical analysis of the *Oryzias* wild population confirms that the South and North Japan population are genetically divergent and that they are more related to each other than to the other groups. The average sequence divergence between the Southern and Northern population is about 0.8% in the coding regions and 2.6% in the intron regions (Naruse *et al.*, 2000). Considering that the difference between genomes of human and great ape is 1-2% (Fujiyama *et al.*, 2002), the SNP (single nucleotide polymorphisms) frequency between these two medaka populations is high. Therefore, they can be used for genetic mapping of genes and mutations from medaka inbred lines.

Strain	Genetic background	Origin	Special features	Reference
HO4C	Southern population	Commercial strain in Japan	Orange-red color (*b/b*)	Hyodo-Taguchi, 1996
HO5	Southern population	Commercial strain in Japan	Orange-red color (*b/b*)	Hyodo-Taguchi, 1996
HB32C	Southern population	Wild population in Chiba	Wild type	Hyodo-Taguchi, 1996
HB32D	Southern population	Wild population in Chiba	Wild type	Hyodo-Taguchi, 1996
HB12A	Southern population	Wild population in Chiba	Wild type	Hyodo-Taguchi, 1996
HB11A	Southern population	Wild population in Chiba	Wild type	Hyodo-Taguchi, 1996
HB11C	Southern population	Wild population in Chiba	Wild type	Hyodo-Taguchi, 1996
Hd-rR	Southern population	Stock at Nagoya Univ.	Orange-red in male and white in female	Hyodo-Taguchi, 1996
Hd-rr	Southern population	Stock at Nagoya Univ.	White in both male and female	Hyodo-Taguchi, 1996
Cab	Southern population	Commercial strain from Carolina Biological Supply	Variegated body color	Loosli et al., 2001
AA2	Southern population	Mutant stocks at Nogoya Univ.	*b/b*, *gu/gu* and *ff/ff* genotype	Shimada and Shima, 1988
HNI-I	Northern population	Wild population in Niigata	Wild type	Hyodo-Taguchi, 1996
HNI-II	Northern population	Wild population in Niigata	Wild type	Hyodo-Taguchi, 1996
Kaga	Northern population	Wild population in Kaga	Wild type	Loosli et al., 2001
HSOK	East-Korea population	Wild population in Sokcho, Korea	Wild type	Hyodo-Taguchi, 1996

Table 2-2: Medaka inbred strains and their features (Naruse *et al.*, 2004)

To create the physical map of medaka three inbred strains: Cab (originally obtained from Carolina Biological Supply and established as an inbred strain by Wittbrodt's group. the Cab strain is currently in use for mutagenesis using ethylnitrosourea (ENU)), Hd-rR (a target of the genome sequencing project) both from the Southern, and HNI from the Northern population

were employed. In this way we have the possibility to compare the genomes of the Southern populations against each other and against the Northern population.

2.2.3 See-through medaka fish

Lately, the transparent fish, the see-through medaka, was created in Japan. This fish is transparent in the adult stage as well as in embryonic stages and was generated by genetically removing pigments from the skin and the peritoneum by the crossing of color mutants. The see-through is healthy and fertile (Figure 2-2).

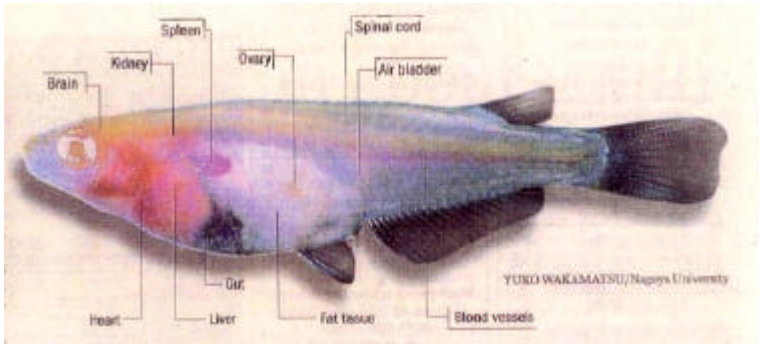

Figure 2-2: See-through fish created in Japan (Wakamatsu, Nagoya University, Japan)

In this fish, the main organ systems are visible from the outside of the body of living fish by the naked eye. This advantage helps the study of the alteration of the internal structures at the later stage of life such as organogenesis in postembryonic stages, growth and aging, carcinogenesis and genesis of other diseases, and inflammation or other alteration caused by exposure to chemical substances including endocrine disrupters (Wakamatsu *et al.*, 2001).

The use of this fish in theory provides the possibility of large-scale mutagenesis to screen mutations affecting postembryonic or late-onset biomedical phenomena because of the transparency in the adult stage. Using the see–through fish decreases the number of animals to be used for experiments because many types of experiments can be conducted without killing the animals. Furthermore, this fish can be used for classroom lessons as anatomical, physiological and etc. subject.

These features improve the potential of medaka as an experimental model.

2.3 Genome study among the teleosts and human provides the knowledge of their synteny and phylogeny

2.3.1 Synteny conservation among medaka, zebrafish and human

The medaka map can be anchored to the maps and genome sequence of other vertebrate species by establishing orthology relationships between the genetically mapped medaka genes and other vertebrates.

The genome-wide comparison of linkage relationships among the medaka and human genomes by the BLASTX searches indicated that 818 pairs of genes and ESTs seem to be orthologous in these species. The degree of synteny between medaka and human was analyzed and the result showed that the distribution of orthologous genes in these two species is dispersed but it is obviously not random (Naruse *et al.*, 2004). However, detection of clusters of orthologous gene pairs in these species shows that the medaka and human genomes share many conserved syntenic segments even after more than four hundred million years divergence from a common ancestor (Kumar and Hedges, 1998). Medaka and human share 104 conserved syntenic segments involving at least three orthologous gene pairs in the data set. 255 orthologous gene pair clusters are also found in the medaka/zebrafish matrix. Naruse and colleagues (2004) reported that the degree of synteny conservation between human and either medaka or zebrafish is almost the same.

2.3.1.1 HOX gene clusters

The Hox genes are a cluster of homeobox-containing genes that are required to pattern the principal axis of the animal body. They are master genes for patterning the anterior and posterior axis in both vertebrates and invertebrates. 31 Hox genes in four clusters were found in pufferfish, whereas 47 Hox genes in seven clusters exist in the zebrafish, suggesting a Hox cluster duplication in an ancestor of the zebrafish lineage (Amores *et al.*, 1998). In medaka 22 Hox genes were mapped to seven different linkage groups. It suggests that medaka may also have undergone additional duplication of the Hox clusters (Ishikawa, 2000; Naruse *et al.*, 2000; Naruse *et al.*, 2004). It seems that the additional duplication of Hox clusters happened at least before the divergence of medaka and zebrafish.

Mammals have about 40 Hox genes, which are organized in four clusters on different chromosomes (Scott, 1992).

2.3.1.2 Major Histocompatibility Complex (MHC)

The primary immunological function of Major Histocompatibility Complex (MHC) molecules is to bind and present antigenic peptides on the surfaces of cells for recognition (binding) by the antigen-specific T cell receptors (TCRs) of lymphocytes. The MHC is one of the interesting regions of vertebrate genome, because of existence of important immunologically relevant genes in this region. The MHC is divided into three subregions, class I, II and III. Genetic analysis indicated that the linkage among MHC class I, II and III genes is ancestral because it can be found in many vertebrates from cartilaginous fish to mammals (Terado *et al.*, 2003). However, in all bony fish the MHC class I and II genes map to multiple chromosomes (Flajnik and Kasahara, 2001). In addition, in medaka and zebrafish, the mammalian MHC class III complement genes are linked neither to class I nor to class II genes (Kuroda *et al.*, 2000), indicating that extensive chromosomal rearrangements in the bony fish lineage dispersed the MHC genes to several chromosomes. Nevertheless, a core MHC exists in medaka, contains class I genes, TAP-transporters and proteasome genes (both required for peptide processing and transport), plus several framework genes (e.g. Col11a2, Daxx) that are also present in the human MHC. Nucleotide sequence analysis of two Hd-rR BAC clones (see 2.4.6.1) from Southern Japanese population, which together cover about 430 kb of the MHC class I region, indicated 22 putative genes and 3 pseudogenes (Matsuo *et al.*, 2002). Except of three genes whose human orthologos are mapped to different chromosome, 19 of the 22 genes have their orthologos in the human MHC, indicating a high degree of synteny conservation of these genes between mammals and bony fish. The MHC class I region was also analyzed in the HNI strain from Northern population and polymorphism between these two strains was observed. For instance, a segment of about 100 kb encompassing two class I genes and proteasome subunit genes Psmb8 and Psmb9 could not be aligned between the two strains. In the rest of the compared region the Hd-rR and HNI MHC showed about 95% nucleotide identity (Tsukamoto *et al.*, 2005).

The study of MHC and Hox genes in medaka, zebrafish and mammals shows that the conserved synteny can be found between these species. As the positions of these genes on linkage groups or chromosomes are different it suggested rearrangements in gene order between medaka, zebrafish and mammals (Naruse *et al.*, 2004).

2.3.2 Gene duplication in teleost fish

The comparison of the genome between the species is an important tool for investigation of gene function and regulation as well as gene evolution. Comparative genomics has shown that in teleost fish many examples of duplicated genes exist, whereas in mammals only one copy of that gene is present (Wittbrodt *et al.*, 1998). Sequence analysis of seven Hox gene clusters found in zebrafish (see 2.3.1.1) revealed that one of them is the single fish orthologue of one of the mammalian Hox clusters and the remaining six appeared to be the result of a duplication of the other three that are found in mammals. It was suggested that the major parts of the zebrafish genome are present in duplicate, possibly of whole genome duplication in an ancestor of zebrafish. That one cluster is not present in two versions was explained by a later loss of the second copy in the lineage leading to the zebrafish. Many genes in fish are present as a single copy, and therefore most of the duplicated version of genes must have degenerated since the initial duplication event (Taylor *et al.*, 2003). However, the redundant gene produced by genome duplication might have evolved new functions that were necessary for fish diversity. A molecular phylogenetic analysis of these genes supports an alternative theory that tandem gene duplications are the source of the additional genes. Gene duplication is the major force of evolution (Ohno, 1970). As many genes in fish are present as single copy, it was suggested that after whole genome duplication, many of the duplicated versions of genes were degenerated or silenced (Van de Peer *et al.*, 2001).

The database Wanda can be used for more information about duplicated genes. Wanda is a database for more investigation in duplicated genes and provides lists of groups of orthologous genes with one copy from man, mouse, and chicken, one or two from tetraploid Xenopus and two or more ancient copies from ray-finned fish. This page also contains the sequence alignment and phylogenetic trees that were necessary for determining the correct orthologous and paralogous relationships among genes. Wanda is available at http://www.evolutionsbiologie.uni-konstanz.de/Wanda/.

2.4 Medaka as a genetic tool

Medaka is widely used for carcinogenesis studies and also for physical and chemical mutagenesis experiments. Until now more than 80 spontaneous mutations were identified in medaka, which were maintained at the medaka stock center in Nagoya University, Japan. The mutants display defects in pigmentation, body size, fin or the scale morphology. Most of

mutations affect pigmentation, but other mutants such as *double anal fin* (*Da*) (Ohtsuka *et al.*, 1999) or *pectoral fin-less* (*pl)* were also observed in medaka. *Da* is a mutant that has an autosomal semidominant effect and makes a mirror-image duplication of the ventral body structures. Loosli and colleagues (2000) reported the first systematic mutagenesis approach to isolate embryonic-lethal developmental mutants in medaka. Large-scale mutagenesis projects are now being conducted by several groups in Japan and are delivering a vastly expanded pool of medaka stocks. Recently, the ERATO (Exploratory Research for Advanced Technology) project conducted by H. Kondo and M. Furutani-Seiki started a large-scale genome-wide screening by ENU (N-ethyl-N-nitrosourea) mutagenesis, which is comparable to the studies done in zebrafish (Driever *et al.*, 1996; Furutani-Seiki *et al.*, 2004).

The recent aim of these projects is the identification of novel genes and pathways and thereby obtaining new insights into gene function in vertebrates.

2.4.1 Medaka genetic and linkage map

A genetic map is a representation of the order of genes on a chromosome, in which the distance between adjacent genes is proportional to the rate of recombination between them. The map unit is defined as that distance between genes for which one product of meiosis out of 100 is recombinant and will be referred as centimorgan (cM). Genetic maps have been used for various kinds of biological analyses, such as map based cloning, comparative vertebrate genomics and detection of radiation–induced DNA mutations.

When a new mutant gene is discovered, the first step in genetic analysis is usually genetic mapping to determine its position in the genome.

Naruse and his colleagues established a genetic linkage map of medaka using 633 marker including 488 AFLPs (amplified fragment length polymorphism), 28 RAPDs (random amplified polymorphic DNA), 34 IRSs (interspersed repetitive sequences), 75 ESTs (expressed sequence tags), 4 STSs (sequence tagged sites), and 4 phenotypic markers (Naruse *et al.*, 2000). They constructed a genetic map with 24 linkage groups, corresponding to the haploid number of medaka chromosomes.

The map length of all linkage groups (LG) in medaka male meiosis is 1400 cM. As the total genome size of medaka is 800 Mb, it was estimated that the physical length of each LG would be 19 to 59 Mb. Then, the map density would be about one marker per Mb or one marker every 1.75 cM (= 1400 cM /800 Mb). This map indicates a high degree of polymorphism between the two inbred strains, AA2 and HNI, derived from southern and northern Japanese populations.

Comparisons of the marker distributions of anonymous DNA markers, such as ESTs show that distributions of genes are not equal in all LGs. For example, the gene density of LG22 is 4.3 times higher than that of LG2 (Naruse *et al.*, 2004). On this map the largest LG, based on the estimated physical length is the LG1 (59 Mb) whereas the longest LG based on recombination mapping is LG4 (104 cM). Additionally, in medaka the male-determined factor Y was mapped to LG1. LG1 had a total map length of 44.2 cM, and 49 mapped loci. That means the largest number of markers were detected on the LG1.

To study the linkage map and map information see the website of Mbase (http://mbase.bioweb.ne.jp/~dclust/medaka_top.html).

2.4.2 Sex determination

Fish are an interesting group of organisms to study sex determination from the evolutionary point of view, because this group has a complete range of various types of sexuality from hermaphroditism to gonochorism and from environmental to genetic sex determination (Devlin and Nagahama, 2002). The comparison of structure and expression of sex-determination genes in different fish species helps to know, which genes are important in the sex-determination cascade and additionally their conservation in evolution can be studied.

Until now there is no information about sex determination, sex chromosome and the process of sex differentiation in both pufferfish species, *Tetraodon nigroviridis* and *Takifugu rubripes*, even though their genome sequences have been published.

In zebrafish the mode of sex determination is unknown. In its larvae was observed that some of the undifferentiated gonads develop into an ovary-like gonad and become ovaries and in the remaining ones, the oocytes undergo apoptosis and a testes develops from undifferentiated parts of the gonads (Hsiao and Tsai, 2003). How these processes happen and which factors are participating is still unclear. Sex-linked markers of the zebrafish genomic map and mutants were not reported up to now.

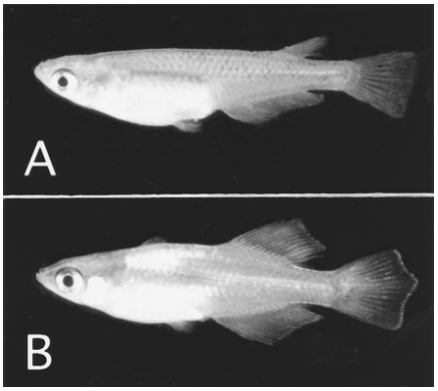

Figure 2-3: The male and female of medaka.
A: female fish, B: mal fish.

Medaka has an XX-XY sex determination system like mammals, with the male determination locus on the Y chromosome. In medaka the male can be easily distinguished from female by a number of secondary sex characters like the shape and size of the dorsal and anal fin (Figure 2-3).

The fluorescent *in situ* hybridization (FISH) analysis showed that the medaka sex chromosome is one of medaka's largest chromosomes. The morphology of the X and Y chromosomes showed no difference between them (Matsuda *et al.*, 1998). It can therefore be assumed that medaka has a genetic sex-determination system in a very early stage of evolution. In medaka the male-determined factor, Y, was mapped to linkage group 1 (LG1). The genetic map of the sex chromosomes evidenced that these two chromosomes share a highly homologous region (Wada *et al.*, 1995; Naruse *et al.*, 2000).

Additionally, crossing-over is possible over almost the entire length of the sex chromosome. The linkage map of the sex chromosome contains in large part an even distribution of markers, like on an autosome. However, around the sex determination locus, a pronounced clustering of markers indicates a region of reduced recombination (Kondo *et al.*, 2001; Matsuda *et al.*, 1999).

2.4.2.1 Dmrt1bY is the male sex-determining gene of medaka

The origin of the sex-determining locus of the medaka is a region at the top of linkage group 9, which contains the *dmrt1* (doublesex/Mab-3-related transcription factor 1) gene. The chromosomal segment containing *dmrt1* on the LG 9 became duplicated and the duplicated version was inserted into the Y chromosome and designated as *dmrt1bY* to distinguish it from its autosomal ancestor *dmrt1a* (Nanda *et al.*, 2002). This segment is 260 kb on the Y and is absent from the X. As pairing and crossing-over is inhibited in this region it leads to recombination isolation of the sex-determination gene. This recombination suppression is the character of the Y as the male determination chromosome. With the exception of the Y-specific region, sequences on the X and Y are very similar. It seems that *dmrt1bY* is the only functional gene in the Y-specific region (Nanda *et al.*, 2002). The coding sequence of this gene consists of five exons and covers about 50 kb of genomic sequence. Naturally occurring mutation in this gene caused XY females (Matsuda *et al.*, 2002).

The *dmrt1* gene is highly conserved during evolution. Its mammalian homologue is involved in sex determination as well and deletion of one copy of human *DMRT1* in XY males leads to female sex reversal (Nanda *et al.*, 2003). The *DMRT1* was also observed in birds and the corresponding homologous gene with the function of sex determination was detected in

Drosophila melanogaster and in *C. elegans* (Zarkowa, 2001). There are no genes in the Y-specific region, which are essential for the fish, not even for the function of the testes or sperm. The investigation of *dmrt1bY* in other fishes indicated that pufferfish and zebrafish have only a single copy of *dmrt1*, which is the orthologue of the autosomal *dmrt1a* of medaka. In *Oryzias curvinotus*, a sister species to the medaka, which also has an XX-XY sex determination system, the *dmrt1bY* was observed (Kondo *et al.*, 2003). Southern blot analysis, linkage analysis and FISH analysis with dmrt1 probes indicated that the Y chromosomes of *O. latipes* and *O. curvinotus* are homologous (Kondo *et al.*, 2004)

2.4.2.2 Sex-reversal in medaka

The study of sex reversals in medaka was reported first by Yamamoto in 1958. In contrast to higher vertebrates, full sex reversals can be obtained in medaka, using steroid sex hormones during the larval period. Treatment with these hormones generated XY females, XX males, YY males and YY females. Interestingly, sex reversed fishes are fully fertile. In experiments, by crossing sex reversed XY female to normal male, fully viable and fertile YY males can be obtained and if these YY males were treated with hormones, became YY females (Yamamoto, 1975).

It was observed, when XY embryos are feminized by estrogen treatment that the expression of *dmrt1bY* is not affected, and that the gene is even transcribed in the ovaries of the adult sex reversed fish (Nanda *et al.*, 2002). An analysis of several laboratory strains shows, that there is a high percentage of males with XX sex chromosomes, lacking *dmrt1bY*. These observations make it clear that *dmrt1bY* is not necessary for male development in every case. Obviously the sex determination system of medaka is not as stable as it appeared, and other factors may substitute for the *dmrt1bY* function.

2.4.3 Medaka and transgenesis

Transgenesis means the injection of foreign DNA into the cytoplasm of medaka 1 or 2-cell-stage embryos. The expression from the injected foreign DNA starts immediately after injection and lasts up to somitogenesis and early organogenesis stages (Loosli *et al.*, 1999). The injection of foreign DNA into germinal –vesicle stage oocytes in medaka created the first stable transgenic fish lines and later it was performed by cytoplasmic injection into one-cell-stage embryos (Wittbrodt *et al.*, 2002). 1-5% of medaka embryos, which were injected with plasmid DNA,

stably integrated the foreign DNA to be used as transgenic founders. One example of a stable transgenic line is a strain that expresses a GFP gene (Green fluorescent protein) from the *vasa* promoter. The endogenous maternal vasa mRNA is present in all cells of the early embryo until the midblastula transition (Tanaka *et al.*, 2001), useful for studying germ-cell development and, when used in mutagenesis screens, they could help to identified genes that are involved in germ-cell-fate determination (Wittbrodt *et al.*, 2002).

2.4.4 ES-like cells

Embryonic stem cells are undifferentiated cells that are unlike any specific adult cell. However, they have the ability to form any adult cell. Embryonic stem (ES) cell lines directly derived from early embryos offer an *in vitro* system to study the molecular mechanism of cell commitment, determination and differentiation during embryogenesis. The ES cell has the attribute to retain its full developmental potential and after retransplantation into the host embryo they can differentiate into all cell types (except the trophoblast in mouse). They can also contribute to the germ line and can pass to the next generation. Mouse ES cells have widely been exploited for gene targeting for the production of knockout animal models for the analysis of gene function and human disease. Fish ES cell cultivation began approximately 10 years ago in zebrafish and medaka. Wakamatsu *et al.* (1994) adopted the feeder layer technique and reported the first medaka ES-like cell line, OLES1. The ES-like cells from zebrafish and medaka can contribute to the germ line after a short period of *in vitro* culture (Ma *et al.*, 2001). Long-term culture conditions, which give enough time for the genetic manipulation of cells, have been established only in medaka. These cell lines show all characteristics of mouse ES cells *in vitro* and have the competence for chimera formation. They have the ability to differentiate into many cell types like melanocytes, neurons and muscle cells *in vitro*. Medaka ES-like cells also remain pluripotent *in vivo* and after transplantation into host embryos they can differentiate into highly specialized cell types, such as the epithelial cells of the undulating fins, contracting muscle cells of the heart, neurons and retinal cells (Hong *et al.*, 1998). The crucial step in establishing ES technology is transmission of the retransplanted ES cells through the germline. Although several medaka ES lines have proved to be fully competent to contribute to all somatic cell lineages and organs of the adult fish, germline transmission has not been obtained so far. However, the ES-like cell lines in medaka provide a system to study many aspects of cell differentiation *in vitro* and in chimeric embryos in fish as a lower vertebrate system.

2.4.5 Nuclear transplantation in the medaka

Nuclear transplantation is a key technique for genetic modification and cloning. The first successful nuclear transplantation in vertebrates was done by Briggs and King in 1952. They transplanted the blastula nuclei from tadpoles into enucleated eggs of *Rana pipiens* and they were able to create normal tadpoles. Afterwards different groups have been trying to develop the nuclear transplantation methods. Wakamatsu and her colleagues are one of the groups, which are working on the nuclear transplantation in fish. First they were able to transplant the nuclei from blastula-stage embryos into non-enucleated unfertilized eggs of the medaka. They established that nuclei taken from blastula- or gastrula-stage embryos exhibit a higher potential for normal development in nuclear transplants than those taken from older embryos. They obtained nuclear transplants that developed into adult fish but were triploid and infertile. The nuclear transplant fish with wild-type body color, which is specific to the donor strain, had 72 chromosomes, which is a triploid chromosome number. These triploids might originate from the fusion of diploid donor nuclei and haploid recipient pronuclei. The presence of haploid pronuclei in the recipient eggs is considered to have almost no effect on the development of the resultant triploid embryos or to only slightly improve the progress of embryonic development to later stages (Niwa e*t al.*, 1999).

Secondly, the embryonic cell nuclei from transgenic fish carrying the GFP (green fluorescent protein) gene were transplanted into unfertilized eggs irradiated by X-rays. Six (2-3%) of nuclear transplants grew to the adult stage. They were all diploid and fertile, and homozygous for the natural and introduced marker genes from the donor nuclei. These genetic markers were transmitted to the subsequent generation in a mendelian fashion (Niwa e*t al.*, 1999; Wakamatsu e*t al.*, 2001; Ju *et al.*, 2003).

In a third step the nuclei of cultured cell from transgenic fish were transplanted into unfertilized eggs, without enucleation, in two experiment series. In the first experiment series, fibroblast cells cultured from the adult caudal fin were used as donors. From more than thousand transplanted eggs, a considerable number of nuclear transplants developed to various embryonic stages. Only 0.3% of them hatched and survived, for 3 weeks at the longest. In the second experiment series the cells cultured from 6-day-old embryos were used as donors and from thousand transplanted eggs, only 5 (0.5%) hatched. In both series the results showed that the donor nuclei contributed to the development of the nuclear transplants. Also the chromosome number varied among cells in a single transplant embryo in both experiences. One possibility to explain the chromosomal aberration is that the cell cycle of the donor nuclei might have failed to synchronize with that of

recipient eggs. That means the cell cycle of the donor nuclei was not in synchrony with that of the developing embryos, leading to chromosomal aberration. Such non-synchronous cell cycles of cultured cells and early embryos were not occurred in mammalian nuclear transplantation, because the cell cycle of early embryos presumably is the same as that of cultured cells (Ju *et al.*, 2003).

2.4.6 Medaka genomic resources

2.4.6.1 BAC libraries

The BAC (bacterial artificial chromosome) libraries used were established from southern (Cab and Hd-rR) and northern (HNI) strains of medaka, *Oryzias latipes*. All three BAC libraries were stabilized by chloramphenicol resistance. The Cab library is available at the RZPD (Resource Centre/Primary Database, http://www.rzpd.de/products/sets_libraries, library number 756). This library consists of 55,296 clones in 144 384-well microtiter plates. The medaka genomic DNA was prepared from fresh brains of more than 20 fish. After partial digestion with EcoRI the DNA fragments were ligated into the pBACe3.6 (Figure 4-2) vector and transformed into *Escherichia coli* strain DH10B host cells. The average insert size of a Cab BAC clone is 150 kb. The total library therefore provides a 10-fold coverage of the medaka genome.

The Hd-rR BAC library consists of approximately 92,160 clones, arrayed into 240 384-well microtiter plates. The medaka genomic DNA was prepared from sperm of 20 Hd-rR males. The DNA fragments were ligated into the pBAC-Lac (Figure4-1) vector (Asakawa *et al.*, 1997) and transformed into DH10B cells. The average insert size of these clones is 210 kb. The library thus covers the medaka genome 24 times (Matsuda *et al.*, 2001).

The HNI BAC library was prepared from the DNA of primary cultured cells from the tail fin of a HNI strain fish, cloned into the vector pBAC-Lac (Asakawa *et al.*, 1997). The average insert size of the library is 160 kb. The library is expected to cover the medaka genome 20 times (Kondo *et al.*, 2002).

2.4.6.2 Cosmid libraries

Two cosmid libraries from Cab and Kaga strains are available at the RZPD (RZPD: Cab, library number 74; Kaga, library number 75). The average insert size of these clones is 40 kb.

A cosmid HNI library was constructed in Japan (Ohtsuka *et al.*, 2002). This library consists of approximately 120,000 clones, which represents a six-fold genomic coverage.

2.4.6.3 Physical mapping of medaka genome

Physical maps provide a view of how the clones from the genomic clone libraries are distributed throughout the genome. In other words, physical maps are the maps of cloned genomic DNA. A physical map provides knowledge of the exact position of a marker or gene on the chromosome. The distance of marker or gene to another locus on the same chromosome is expressed by number of base pairs (bp), a physical equivalent. The first stage in the analysis of complex genomes is usually the production of a physical map, which provides a diagram of the physical location of various landmarks (a landmark is the location of a particular DNA sequence, such as a coding region). Physical maps are key tools for both structural and functional characterization of individual genes and gene complexes. At the time when this thesis project was initiated, no physical map of the medaka genome existed. For the reasons outlined above, a physical map is necessary for many applications including genomic sequencing and the straight-forward identification of genes that are responsible for interesting phenotypes discovered in ENU-mutagenised fish. A description of the medaka physical map obtained will follow in the Results section.

2.4.6.4 Medaka expressed sequence tags

An EST (expressed sequence tag) is a short DNA sequence obtained from a cDNA (complementary DNA) clone that corresponds to a mRNA. Each EST represents a part of a gene. The EST approach is a powerful technique for large-scale transcriptome characterization. EST projects carried out in Japan and Germany provided sequence information on the majority of medaka genes, by sampling gene expressed in various embryonic stages and adult tissues (Kimura *et. al*, 2004; Berger *et al.*, unpublished).

A group of medaka ESTs was established from four cDNA libraries that were prepared from the whole body, liver, ovary and a cultured cell line of the HNI strain. These cDNA libraries were created by Prof. H. Mitani from the university of Tokyo, Japan (Table 3-1) and are available at http://mbase.bioweb.ne.jp/~dclust/lib_info.htm. A medaka EST database with mapping information is available at http://mbase.bioweb.ne.jp/. The 5'-terminal DNA sequences of 13,000 cDNA clones were determined and ~50% of medaka ESTs which are bigger than 300 bp showed homology with previously identified genes from other species (Wittbrodt *et al.*, 2002).

The first large-scale isolation of ESTs from developing medaka embryos was reported by Kimura and his colleagues. They have sequenced both ends of approximately 65,000 clones from cDNA libraries of segmentation to fry stage embryos with a total of 132,082 EST sequences analyzed. Clustering analysis with 3-prime sequences finally identified a total of 12,429 clusters. Using this sequence information, a 60-mer oligonucleotide microarray with 8,091 unigenes (clusters) was constructed and tested for its usefulness in expression profiling. Furthermore, 924 clusters were subjected to *in situ* hybridization to determine the spatial localization of their transcripts (Kimura *et al*., 2004). The EST sequence information, images of *in situ* hybridization results, a list of the cluster names selected for both *in situ* hybridization (924) and more other information about these ESTs, are available at http://medaka.lab.nig.ac.jp. By March 2004, about 150,000 entries of medaka ESTs had been submitted to public databases.

Research on the medaka transcriptome started with the oligonucleotide fingerprinting approach (OFP) as a method to reduce redundancy of medaka cDNA libraries in group of Dr. Himmelbauer, MPI, Germany. They used for their analysis 137,472 cDNA clones from different adult organs and developmental stages (ovary, gastrula, neurula and organogenesis). From a reduced redundancy set they generated 5' ESTs and obtained 10,542 unique sequences. These clones are used as a resource for further experiments, including expression analysis (Berger *et al.*, unpublished).

2.5 Teleost genome sequencing

In the past years the genome sequence of different vertebrate model system like mouse (Gregory *et al.*, 2002), *Takifugu rubripes* (Aparicio *et al.*, 2002), tetraodon (Jaillon *et al.*, 2004) were published. The draft of the whole genome sequence of fugu has greatly influenced genomic research in medaka, because medaka and fugu are evolutionarily close to each other (Figure 2-1). For sequencing large complex genomes like medaka there are two main approaches, first, the whole-genome shotgun (WGS), i.e. shotgun sequencing of the entire genome, and second, hierarchical shotgun, i.e. shotgun sequencing of BAC clones or contigs arranged by fingerprinting or hybridization. In the WGS approach, the shotgun clones will be randomly chosen and their end sequence will be read without any information about the position of these clones in the map. After sequencing, the overlapping clones will be identified and a map will be constructed.

In the hierarchical shotgun, overlapping clones (obtained by using fingerprinting or hybridization or STS-based approach) are first used to build a physical map. Subsequently, individual clones will be chosen from a minimal tile path and sequenced.

Of these two approaches the WGS has the advantage of the simplicity and rapid early coverage of the whole genome and its usefulness for identifying genes. However, if the genome contains highly repetitive sequences, such as the human genome, the use of the WGS approach is difficult. Therefore, to complete and to finish the genome sequence, the hierarchical approach is required.

2.5.1 Zebrafish

Zebrafish is one of the most successful model systems due to the ease of genome-wide screening for mutants showing specific phenotypes (Driever *et al.*, 1996). The zebrafish will play a key role in finding genes in the other genomes. Its genome size is 1700 Mb, only half the size of that of mouse or human. In 2001, sequencing of the whole genome of zebrafish (*Danio rerio* sequencing project) was launched at the Sanger Center, UK, (http://www.sanger.ac.uk/Projects/D_rerio/) following two strategies: clone mapping and sequencing from BAC and PAC libraries and whole genome shotgun sequencing. For the clone based mapping, clones from different libraries were fingerprinted by digesting clones with HindIII. After band separation by electrophoresis on agarose, the images of bands were processed by IMAGE and FPC software (http://www.sanger.ac.uk/Projects/D_rerio/WebFPC/zebrafish/large.shtml) and the clones are contiguated on the basis of shared bands. The map is providing a template for clone tile path selection, as a prerequisite for sequencing, which is taken place at Sanger Institute. By January of 2005, there are 212839 fingerprints assembled into 3775 contigs. Sequencing data continues to be updated frequently on the database web sites (http://www.ensembl.org/Danio_rerio/). The DNA of ~1000 zebrafish 5 days embryos was used for the whole genome shotgun sequencing approach.

The fourth assembly Zv4 of the zebrafish genome was released. The assembly comprises a total sequence length of 1,560,480,686 bp in 21,333 fragments. This assembly has been tied to the FPC map (data freeze 17th of May, 2004). This assembly contains 443 Mb from 2,828 finished clones, and 121 Mb from 1,272 unfinished clones. The FPC contigs provided a template to place

the (un) finished sequence. A new mix strategy of sequence alignment and BAC end position was used to fill the remaining of the sequence with WGS assembly contigs.

2.5.2 Tetraodon nigroviridis

Tetraodon is native to Malaysia, Indonesia and India. Its genome size is 385 Mb, which is 8 times smaller than the human genome. The tetraodon genome consists of 21 chromosome pairs. This fish is an attractive tool for genetic researches because of its small genome size. The small size of the tetraodon genome was not due to a large reduction in the number of genes compared to mammals, but to a reduction in non-coding sequences. Interestingly from one species of vertebrate to another, even for those as far apart as a fish and a mammal, the same genes are present for the most part. Although the length of the exons is similar in human and tetraodon, the size of the introns and the intergenic sequences is greatly reduced in this fish. Indeed the tetraodon genome has a remarkably low content of repetitive DNA. Transposable elements, which constitute the majority of repeated sequences, represent 45% of the sequence of the human genome, whereas they represent only 3.8% of the tetraodon genome. The low number of repeated sequences makes a random shotgun sequencing strategy possible, which is rapid and economic. Considering WGS is not applicable for the human genome with large number of repetitive elements.

The sequencing project of tetraodon started at Genoscope in 1998 using the WGS approach. In June 2001, the Whitehead Institute for Genomic Research (WIGR) joined Genoscope in this project, to accelerate the sequencing of the tetraodon genome. The version 7 of tetraodon sequenced created in April 2004 contains 49,609 contigs (A set of overlapping clones without gaps called a contig) representing 312 Mb of genome sequence. Arachne has ordered these contigs using clone links to form 25773 supercontigs or scaffolds (Scaffold is a larger contig). Arachne is a tool for assembling genome sequences from whole genome shotgun reads, mostly in forward-reverse pairs obtained by sequencing clone ends. These cover 342 Mb, which signifies that there are 30 Mb of gaps which are covered by one or more clones in the scaffolds (Jaillon *et al.*, 2004). The Tetraodon sequences are available at (http://www.genoscope.cns.fr/externe/English/Projets/Projet_C/organisme_C.html).

2.5.3 Takifugu rubripes

This teleost fish belongs to the order *Tetraodontiformes* and family *Tetraodontidae*. Its natural habitat spans the Sea of Japan, the East China Sea, and the Yellow Sea. The genome size of the fugu is 365 Mb which is distributed among 22 chromosomes.

In 2002, the compact genome of fugu has been sequenced to over 95% coverage, and more than 80% of the assembly is in multigene-sized scaffolds (Aparicio *et al.*, 2002). The 3.71 million passing reads were assembled into 12,381 scaffolds longer than 2 kb, for a total of 332.5 Mb. These scaffolds contain 45,024 contigs. 745 scaffolds longer than 100 kb account for 35% of the assembled sequence (119.5 Mb); 1908 scaffolds longer than 50 kb account for 60% of the assembly (200.8 Mb); 4108 scaffolds longer than 20 kb account for 81% of the assembly (271 Mb) (Aparicio *et al.*, 2002).

2.7% of the fugu genome accounted for the repetitive sequences, whereas they represent 45% of the sequence of the human genome. Furthermore, the average gene density in fugu is one gene locus per 10.9 kb of genomic sequence. In the human for example the sequence data suggest that chromosome 19 has an average of 23 genes per Mbp, whereas chromosome 4 averages only 6 genes per Mb (Aparicio *et al.*, 2002; Venter *et al.*, 2001).

Aparicio and colleagues examined the similarities and differences between the human and the fugu and they showed that three-quarter of the human proteins have a strong match in fugu. On the other hand, this implies that 25% of the human proteins have no pufferfish ortholog.

2.5.4 Medaka genome sequencing project

The Hd-rR strain from the south population has been selected as the reference genotype to determine the complete genomic sequence of the medaka, using both WGS approach and BAC sequencing. Sequencing is being carried out at the Academia Sequencing Centre of the NIG in Mishima, Japan. In May 2004, the sequence had a 8.9-fold genome coverage (http://dolphin.lab.nig.ac.jp/medaka/index.php).

2.6 Objective

Medaka is an interesting vertebrate model system for different fields of biology and medical research because of its virtues (Table 2-1), especially its small genome size make this fish a

fascinating tool for genetic approaches. This work aimed to create a physical map of the medaka genome. Physical maps are key tools for both functional and structural characterization of individual genes.

In preparing physical maps of genomes, vectors that can carry very large inserts are the most useful because their use in cloning requires breaking the genome up into fewer pieces than would be necessary for other vectors. That means fewer clones are required to represent the genome. Cosmids, YACs (yeast artificial chromosome), BACs (bacterial artificial chromosome), and PACs (phage P1-based artificial chromosome) are mainly used. For the physical mapping of medaka genome the use of BACs vectors (see 2.4.6.1) was preferred to the other vectors because in addition to their ability to contain large insert, they can be amplified in bacteria and isolated and manipulated simply with basic bacterial plasmid technology.

A physical map can be considered as a set of overlapping clones that spans the genome. To find the overlapping clones, a hybridization strategy was utilized. Known pieces of DNA (probes) were hybridized against the BAC libraries and positive hybridized signals indicated the location of the DNA in clone libraries. Then, in the first part of the work, different probes were produced for hybridizing. These probes were derived from different sources, i.e. BAC end-fragments, cDNA inserts and oligonucleotide probes, generated from a comparison of medaka cDNA sequence with the fugu genome. In the second part of the work the probes were hybridized against three different BAC libraries, Cab and Hd-rR libraries from the south and HNI from the north population. In the third part of the work the hybridized signals were identified by different computer software. At the end a list of overlapping clones were produces, which were utilized for map construction.

This project is a part of the medaka Genome Initiative (MGI) (http://gardy.dsp.jst.go.jp/MedGenln). MGI is a consortium, which consist of various research groups in Japan and Germany who are cooperating to utilize the medaka as a model system in functional genomics.

This work describes the way of assembling a medaka physical map. Our data will be used for chromosomal sequencing of medaka chromosome 22 and by other groups in Japan and Germany, which are working on positional cloning of mutations.