

1 Physikalische Kartierung von Medaka (*Oryzias latipes*)

Medaka ist ein kleiner Süßwasser-fisch (2-3 cm lang), der in der Region um Japan, Korea und China einheimisch ist. Medaka ist u.a. wegen seines kleinen Genoms (800 Mb) und kurzen Lebenszyklus ein interessantes Modellsystem für die Genetik und die Entwicklungsbiologie. Das Medakagenom besteht aus 24 Chromosomen (haploider Chromosomensatz). Proteinpolymorphismenanalyse zeigte, daß die Wildpopulation von *O. latipes* in vier Gruppen eingeteilt werden kann, welche in Nordjapan, Südjapan, Ostkorea und China/Westkorea vorkommen. Für unsere Arbeiten wurden die Stämme Cab und Hd-rR der süd- und HNI der nordjapanischen Population eingesetzt.

Physikalische Karten sind gute Hilfsmittel für die strukturelle und funktionelle Charakterisierung jedes Gens. Das Ziel war die erste physikalische Karte des Medakagenoms mit *Bacterial Artificial Chromosomes* (BAC) Klonen zu erzeugen. Im Besonderen haben wir dafür die Ähnlichkeit des Medakagenoms zu seinem nahen Verwandten, dem japanischen Pufferfisch (*Takifugu rubripes*) genutzt, dessen Sequenz 2002 veröffentlicht wurde. Die evolutionäre Analyse verschiedener Fischarten zeigte, daß Medaka und Fugu enger verwandt sind als Medaka und Zebrafisch. Die Trennung von Medaka und Fugu fand vor 60-80 Millionen Jahren statt. Medaka und Zebrafisch werden beide bereits seit längerem zu molekularbiologischen Zwecken benutzt, im Vergleich hat der Medaka wegen seines kleineren Genoms (etwa die Hälfte vom Zebrafisch) und Vorhandensein von Inzuchtstämmen Vorteile gegenüber dem Zebrafisch.

Als ersten Schritt für die Erstellung der physikalischen Karte haben wir 103.144 öffentliche Medaka-EST-Sequenzen in 21.121 nicht-redundante Cluster gruppiert. 2.363 Gene wurden für die Erstellung der BAC-Karte ausgewählt, und wir haben zu diesen Sequenzen 35-mer Oligonukleotide konstruiert. Die Oligos wurden gegen 64,500 BAC Klone aus Cab- und Hd-rR-Stämmen hybridisiert, welche das Medakagenom durchschnittlich 14-fach abdecken.

Unsere Daten wurden mit 437 Kartierungsergebnissen aus mit PCR amplifizierten Inserts von Medaka-cDNA-Klonen und BAC-Endfragment-Markern ergänzt. Für die Hybridisierung wurden Mischproben aus 6 bzw. 8 Oligonukleotidsonden eingesetzt. Mit dieser Methode wurde die Anzahl von Hybridisierungen reduziert. Durch VisualGrid wurden die Autoradiogramme der Hybridisierungsergebnisse analysiert, positive Signale wurden manuell identifiziert und die Koordinaten der einzelnen betroffenen Klone berechnet. Für die Konstruktion der physikalischen Karte wurde Wprobeorder benutzt. Wprobeorder benutzt den Simulation-Annealing-Algorithmus

zur genauen Lokalisierung von Markern, basierend auf der Verteilung der Klone zwischen den Markern. Klone, die zwei oder mehr Marker gemeinsam haben, überlappen mit großer Wahrscheinlichkeit. Solche überlappende Klone werden durch Wprobeorder identifiziert und die wahrscheinlichste Klonabfolge in den Contigs berechnet.

Die erste Generation unserer physikalischen Karte besteht aus 902 Segmenten, die 74% des Medakagenoms abdecken. Unsere Karte enthält 2.721 Marker. Davon stammen 2534 von exprimierten Sequenzen (ESTs) und entsprechen nicht-redundante 2328 Loci. 934 Markern (724 unterschiedliche) sind auf der genetischen Karte von Medaka verankert. Durch diese Vernetzung ergibt sich schneller Zugriff auf Klone und Kontiges, was das Auffinden von Kandidaten Genen und deren Charakterisierung vereinfacht.

Unsere Daten wurden für die Sequenzierung des Medaka Chromosoms 22, sowie für die positionelle Klonierung von ENU-indizierten Mutationen benutzt.

6 Summary

Medaka, *Oryzias latipes*, is a small, egg-laying fresh water fish native to Asia that is found primarily in Japan but also in Korea and eastern China. The adult fish are about 3 cm long. Medaka is an interesting model system for genetics and developmental biology because of its small genome size (800 Mb) and the short time of life cycle. The medaka genome consists of 24 sets of chromosomes (haploid). The allozyme studies of medaka divided the wild population into four genetically different groups: The North Japan, South Japan, the East Korean and the China-West Korean population (Naruse *et al.*, 2004). For construction of the physical map of medaka three inbred strains, Cab and Hd-rR both from the southern, and HNI from the northern population were employed.

Physical maps are key tools for both the structural and functional characterisation of individual genes. We have created a first-generation physical map of the medaka genome based on bacterial artificial chromosome (BAC) clones. In particular, we exploited the synteny of the medaka genome to the closely related genome of the pufferfish, *Takifugu rubripes*, for which genomic sequence was already published in 2002. Interestingly, medaka and Fugu are much more related to each other than medaka is to zebrafish. Medaka and Fugu are close relatives that diverged 60-80 million years ago. The use of medaka and zebrafish has become very popular during the last years for molecular biology purposes. In comparison the medaka is preferred for genetical experiments to the zebrafish because of its small genome (one-half of the zebrafish) and inbred strains, which simplifies mapping approaches.

As a first step, we clustered 103,144 public medaka EST sequences to obtain a set of 21,121 non-redundant sequence entities. 2363 genes were selected for the BAC map project. We designed 35mer oligonucleotide probes from the selected genes and hybridized them against 64,500 BAC clones of strains Cab and Hd-rR, representing 14-fold coverage of the medaka genome. Our data set is further supplemented with 437 mapping results generated from PCR-amplified inserts of medaka cDNA clones and BAC end-fragment markers.

To reduce the number of hybridization, probes were pooled according to a 2D and 3D pooling scheme and hybridised against BAC filters. Autoradiograms obtained by hybridization were analysed by VisualGrid. Positive signals were manually detected. VisualGrid evaluated the location of these signals in microtiterplates. For map construction, we utilized the wprobeorder software (Mott *et al.*, 1993 and Grigoriev *et al.*, 1998). Wprobeorder uses a simulated

annealing algorithm to define an optimal order of markers, based on clones that are shared between the markers. The more clones two markers have in common, the closer they are located in the genome.

Our current, edited, first generation medaka BAC map consists of 902 map segments that cover about 74% of the medaka genome. The map contains 2721 markers. Of these, 2534 are from expressed sequences, equivalent to a non-redundant set of 2328 loci. The 934 markers (724 different) are anchored to the medaka genetic map. Thus, genetic map assignments provide immediate access to underlying clones and contigs, simplifying molecular access to candidate gene regions and their characterization.

Our data have been widely used by other groups, e.g. for medaka chromosome 22 sequencing and furthermore for positional cloning of mutations induced by ENU.