

1. Einleitung

Das Humane Genomprojekt, das von dem U.S.-amerikanischen 'Department of Energy' sowie den 'National Institutes of Health' koordiniert wird, hat sich zum Ziel gesetzt, alle Gene im menschlichen Genom zu identifizieren, sowie die Sequenz aller 3 Milliarden Basenpaare, aus denen das humane Genom besteht, zu bestimmen (Cantor, 1990). Außerdem sollen alle diese Informationen in Datenbanken gespeichert werden und die sich daraus ergebenden ethischen, legalen und sozialen Probleme diskutiert werden. Das humane Genomprojekt startete 1990 und war mit einer Laufzeit von 15 Jahre geplant. Aufgrund des raschen technologischen Fortschritts wird jedoch erwartet, daß es bis zum Jahre 2003 abgeschlossen werden kann (The Baylor College of Medicine Genome Sequencing Center et al., 1999).

Die nächste große Herausforderung wird dann sein, die Funktion der einzelnen Gene, sowie das Netzwerk, in dem sie untereinander interagieren, zu erforschen. Eine Möglichkeit, die Funktion – oder besser Nicht-Funktion – von Genen systematisch zu entschlüsseln, ist alle Gene eines Organismus zufallsmäßig zu mutagenisieren, und dann ausgehend von den mutanten Phänotypen das betroffenen Genen zu identifizieren. Der Zebrafisch stellt dabei das erste Wirbeltier dar, in dem ein solcher genomweiter, gesättigter Mutagenese-'Screen' durchgeführt wurde.

Diese Doktorarbeit beschreibt zwei verschiedene Aspekte, die bei der Erforschung von komplexen Genomen eine Rolle spielen. Da genetische und physikalische Karten unabdinglich sind, um Gene, die in den mutanten Phänotypen aus einem Mutagenesecscreen betroffen sind, zu klonieren, wird eine Strategie analysiert, mit der effizient eine genetische Kopplungskarte, die Ankerpunkte zu einer physikalischen Karte aufweist, erstellt werden kann. Außerdem wird am Beispiel der Sequenz eines Amphioxus-Cosmids gezeigt, daß sich genomisches Sequenzieren und computergestützter Sequenzanalyse gekoppelt mit EST-Projekten sehr gut ergänzen.

Die Einleitung ist in zwei Abschnitte gegliedert. Im ersten Abschnitt wird die Bedeutung von Modellorganismen, insbesondere von Zebrafisch, Medakafisch, Fugufisch und Amphioxus, erläutert. Der zweite Abschnitt handelt von Methoden zur Identifizierung und Analyse von Genen.

1.1. Modell Organismen

Die Bedeutung von Modellorganismen zur Erforschung der Biologie des Menschen wird in der großen Anzahl verschiedener Organismen, die zur Erforschung verschiedener Aspekte benutzt werden, offensichtlich. Hefen werden zur Erforschung von Eigenschaften, die allen Eukaryonten gemeinsam sind, verwendet. Als Einzeller sind sie einfach in der Handhabung, und in ihrem relativ kleinen Genom von 15-16 Mbp sind alle die Gene, die zum Duplizieren der Chromosomen, zur Zell-Zyklus-Kontrolle oder metabolischen Funktionen dienen, leichter zugänglich. Außerdem steht inzwischen die vollständige Sequenzinformationen von *Saccharomyces cerevisiae* in Datenbanken zur Verfügung (Mewes et al., 1997).

Caenorhabditis elegans ist der erste mehrzellige Organismus, der vollständig sequenziert wurde (*C.elegans* Sequencing Consortium, 1998). Das 97 Mb große Genom ist auf 6 Chromosomen verteilt und enthält mehr als 19.000 Gene, von denen etwa 40% Homologie zu bereits identifizierten Genen anderer Organismen zeigen. *C. elegans* dient besonders zur Erforschung entwicklungsbiologischer und neurobiologischer Fragestellungen. Möglicherweise wird das relativ einfache Nervensystem in Kombination mit etwa 250 in einem Mutagenese 'Screen' identifizierten Verhaltensmutanten Einblicke ermöglichen, wie Gene und Neuronen interagieren müssen, um Verhaltensmuster zu produzieren.

Drosophila melanogaster war der erste komplexe Organismus, in dem ein genomweiter, gesättigter 'Screen' zur Identifikation entwicklungsbiologischer Mutanten durchgeführt wurde (Nüsslein-Volhard und Wieschaus, 1980; Nüsslein-Volhard et al., 1984). Das 165 Mb große Genom ist gründlich kartiert, und es steht bereits sehr viel Sequenzinformation in öffentlichen Datenbanken zur Verfügung. *Drosophila* bietet darüber hinaus eine Reihe von weiteren Vorteilen, wie Einzelsortierungslinien für Transposonmutagenese, achiasmatische Meiose in Männchen und Balancerchromosomen, die das Halten von mutanten Stämmen sehr vereinfachen.

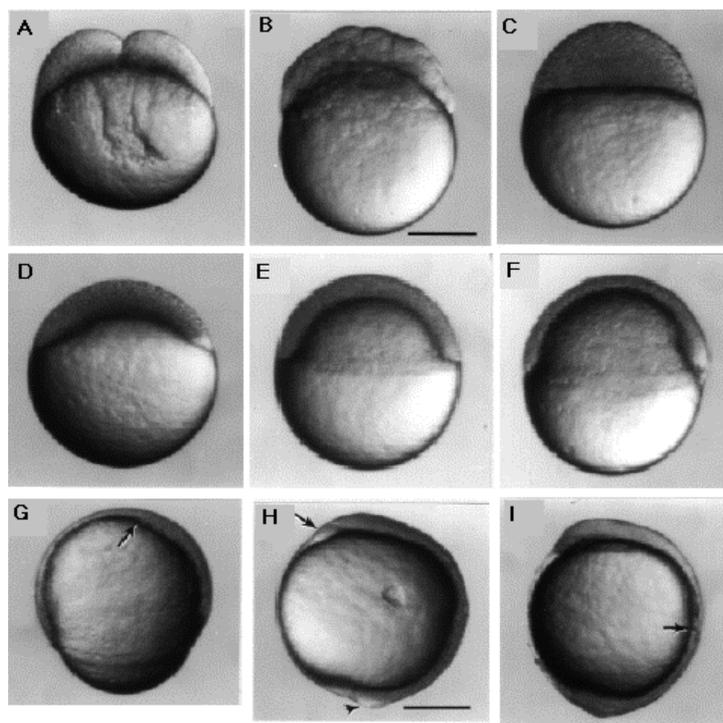
Obwohl gerade auch bei einigen entwicklungsbiologisch relevanten Genen erstaunliche Ähnlichkeit in Sequenz, Funktion und Expressionsmuster zwischen Vertebraten und Nicht-Vertebraten besteht, wie z.B. bei den Pax-Genen und beim Hox-cluster, sind nicht alle Vorgänge, einer Protostomier-Entwicklung direkt auf einen komplexen Deuterostomen wie einen Vertebraten übertragbar. In der Entwicklungsbiologie wurde daher der Krallenfrosch, *Xenopus laevis*, als Vertebratenmodell eingesetzt. *Xenopus* hat transparente Eier, alle Stadien der Entwicklung können direkt am lebenden Objekt beobachtet werden. Die Identifikation von Mutanten sowie andere Manipulationen, wie Transplantationsexperimente und Injektionen sind aufgrund der großen Größe der Eier und Embryonen leicht möglich. Die genetische Analyse in diesem Organismus wird allerdings nicht nur durch dessen großes Genom (3100 Mbp) erschwert (Olmo and Morescalchi, 1978), sondern auch durch eine relativ kürzliche (30 Mio. Jahre) Tetraploidisierung (Bisbee et al., 1977; Shain et al., 1997).

Kleine Nagetiere, wie Mäuse und Ratten, sind seit langem für die genetische und physiologische Forschung etabliert, und als Säugetiere sind sie von allen gängigen Modellorganismen dem Menschen am ähnlichsten. In der Mausgenetik sind ES (=Embryonic Stem) Zellen ein nicht zu unterschätzender Vorteil. Diese können gezielt *in vitro* manipuliert werden und erlauben so 'reverse genetics' (d.h. es kann gezielt ein bestimmtes Gen verändert werden, um dann die phänotypische Auswirkung zu untersuchen). Um frühe entwicklungsbiologische Vorgänge zu erforschen, sind sie allerdings aufgrund der schlechten Zugänglichkeit der Embryonen nicht gut geeignet. Verglichen mit der Anzahl der Nachkommen, die von einem einzigen *Drosophila*-Weibchen, einem Fisch oder von *Xenopus* produziert werden kann, haben sie eine kleine Wurfgröße. Außerdem erfordert das Halten von Mäusen relativ viel Platz. Dennoch wird im Moment ein Mutagenesecscreen, bei dem 12.000 Familien untersucht werden sollen, durchgeführt (Hrabe de Angelis and Balling, 1998). Mehr als 1500 Gene, die durch einen 'gene trap'- Vektor in ES-Zellen unterbrochen wurden, sind bereits identifiziert (Zambrowicz et al., 1998) und weitere 1500 sind auf der 'Genetrap Home-Page' des deutschen Genomprojekts unter http://Public_User:Public_Data@www.rzpd.de/db/html/GeneTrap/main.shtml zugänglich.

Um frühe entwicklungsbiologische Vorgänge in Vertebraten zu untersuchen, erscheinen daher kleine, eierlegende Fische ideal geeignet. Zebrafisch (1.1.1) und Medakafisch (1.1.2) haben in den letzten Jahren zunehmend an Popularität gewonnen. Ein weiterer Fisch, der seit einigen Jahren immer wieder als Modell für ein kompaktes Vertebraten-Genom angeführt wird, ist der Fugufisch (1.1.3). Der Cephalochordate *Branchiostoma* (1.1.4) wird als nächster Verwandter der Wirbeltiere als vereinfachtes Modell für deren Entwicklung und Genomorganisation angesehen, und man kann erwarten, daß sich die Forschung an *Amphioxus* gut mit der an Fischen und anderen Wirbeltieren ergänzen wird, sowie weitere Einblicke in Evolution und Entwicklungsbiologie ermöglichen wird.

1.1.1. Der Zebrafisch

Der Zebrafisch (*Danio rerio*) ist ein kleiner, tropischer Süßwasserfisch, der in die Gruppe der Karpflinge (Cyprinidae) gehört. In den letzten Jahren hat er zunehmend an Popularität bei der Erforschung früher entwicklungsbiologischer Vorgänge gewonnen. Er hat viele Eigenschaften, durch die er sehr attraktiv für die systematische Analyse entwicklungsbiologischer Vorgänge in großem Maßstab ist: Er ist leicht zu halten und ein einzelnes, gut legendes Weibchen produziert wöchentlich hunderte von Eiern, die auch *in vitro* befruchtet werden können. Die extrakorporal befruchteten Eier entwickelt sich innerhalb weniger Tage zu freischwimmenden Larven und innerhalb von 3 Monaten zum geschlechtsreifen Tier. Einen guten Eindruck über die Geschwindigkeit der Entwicklung vermittelt Abbildung 1. Die Zebrafischeier besitzen ein transparentes Chorion, so daß alle Entwicklungsstadien direkt am lebenden Objekt betrachtet werden können. Der Embryo ist relativ groß, was Manipulationen wie Injektionen von RNA, Antikörpern oder Farbstoffen, Transplantationsexperimente, sowie die Identifizierung mutanter Phänotypen vereinfacht.



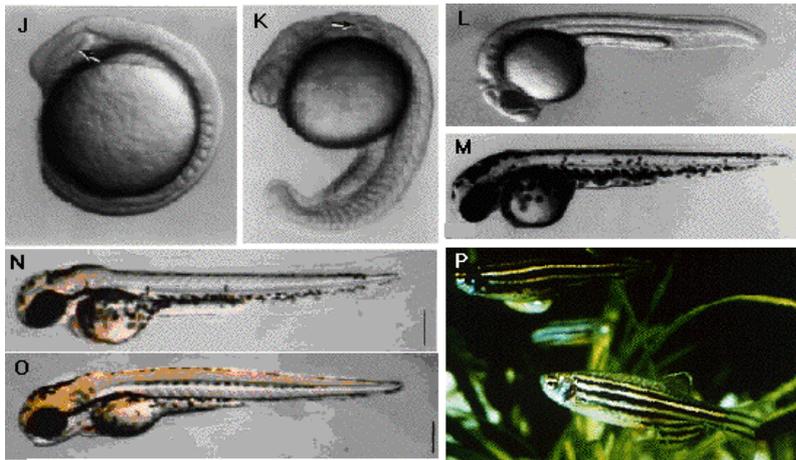


Abb. 1: Embryonalentwicklung des Zebrafisch (aus 'The Zebrafishbook', 1995).

A: Zweizellstadium (45 min.) B: 64-Zellen Stadium (2 h) C: Übergang vom 'oblong-' zum 'sphere-stage' (3,8 h) D: 'Dome-Stage' (4,3 h) E: 50% Epibolie-Stadium (5,25 h) F: 'Shield-Stage' (6 h) G: 70% Epibolie der Pfeil zeigt auf den axialen Hypoblast (7,7 h) H: Knospenstadium; die Pfeilspitze zeigt auf die Schwanzknospe, der Pfeil auf das sog. Polster (10 h) I: Zwei Somiten-Stadium, der Pfeil zeigt auf die Grenze zwischen 1. und 2. Somite. (10,7 h) J: 8-Somiten-Stadium, der Pfeil zeigt auf die sich entwickelnde Augenanlage (13 h) K: 20-Somiten-Stadium, der Pfeil zeigt auf das Ohr-Bläschen (19 h) L: 24-Stunden-Embryo, das Gehirn ist gut sichtbar untergliedert, die Melanisierung hat begonnen, ist aber bei niedriger Vergrößerung noch nicht sichtbar. M: 42-Stunden-Embryo, die Pigmentierung dehnt sich bereits über das ganze Tier aus. N: 48-50-Stunden O: 72-Stunden-Larve, zunehmend wird auch die gelbe Pigmentierung sichtbar. P: Adulter Fisch.

Ein weiterer Vorteil des Zebrafisches ist es, daß bis zum 5. Tag haploide Tiere lebensfähig sind. Sie erhält man, indem man Eier mit UV- behandelten Spermien aktiviert. Ferner kann man gynogenetische diploide Fische erzeugen (Streisinger et al., 1981), indem man entweder die zweite meiotische Teilung durch 'early pressure' verhindert oder die erste mitotische Teilung durch einen 'heatshock'. Durch 'early pressure' erzeugte Nachkommen sind teilweise und durch 'heatshock' erzeugte vollständig homozygot (Abb. 2).

Durch haploide Entwicklung, 'heatshock' oder 'early pressure' ist es z.B. möglich, bei einem klassischen 3 Generationen Mutagenesescreeen eine Generation zu überspringen, einen 'Screen' für maternelle Mutanten durchzuführen, oder genetische Kopplungsanalysen an Kreuzungen von nicht vollständig homozygoten Stämmen, die es im Zebrafisch noch nicht gibt, durchzuführen. Die erste genetische Kopplungskarte wurde 1994 von Postlethwait et al. veröffentlicht und beruht auf RAPD-Markern, die auf haploiden Nachkommen eines C32xSJD-Weibchen kartiert worden sind. Das Zebrafisch Genom besteht aus 25 Kopplungsgruppen, die zusammen etwa 2600 cMorgans überspannen. Da das gesamte Genom eine Größe von $1,7 \cdot 10^9$ bp hat, entspricht ein cM durchschnittlich etwa 625 kb.

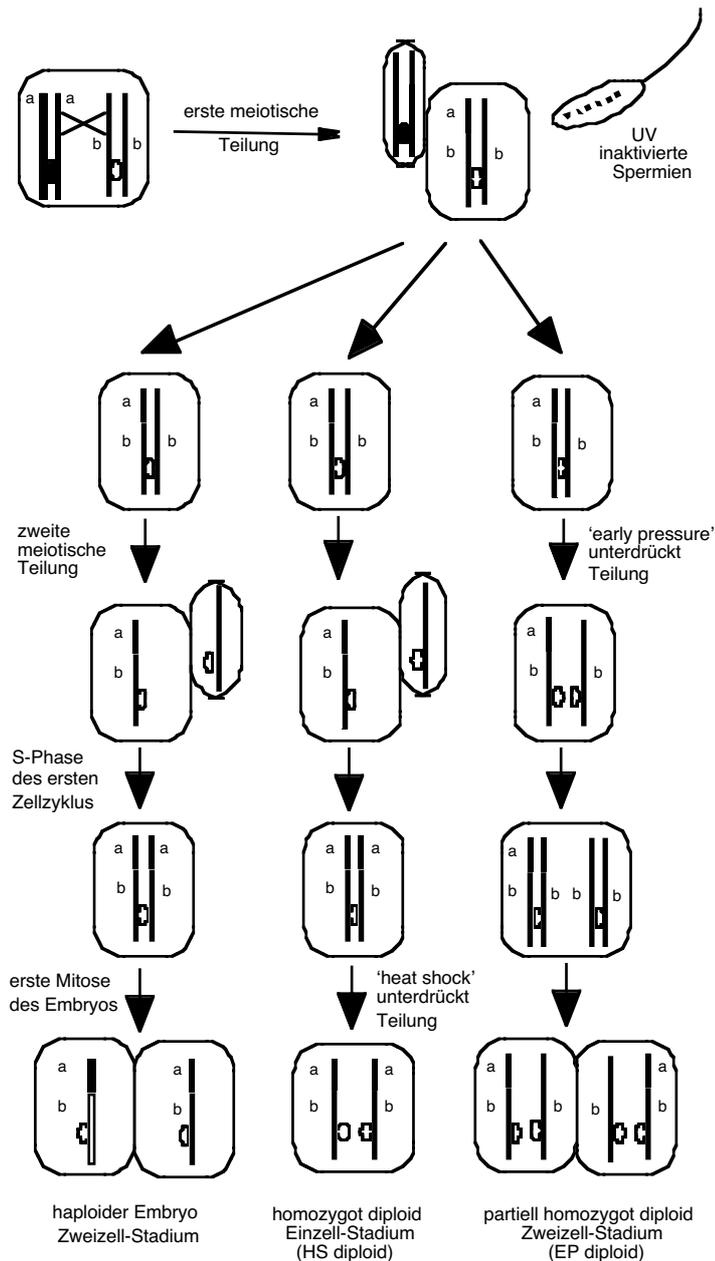


Abb. 2: Parthenogenese-Methoden im Zebrafisch.

Der Zebrafisch Embryo kann sich etwa bis zum 5. Tag haploid entwickeln. Wird die erst mitotische Teilung des Embryos durch einen Hitzeschock unterdrückt, entwickeln sich homozygote, diploide Tiere. Dazu werden die Eier 13 Minuten, nachdem sie durch UV bestrahlte Spermien aktiviert wurden, für kurze Zeit in ein warmes Wasserbad überführt. Etwa 5% der so erhaltenen, homozygoten Embryonen entwickeln sich bis zur Geschlechtsreife. Eine weitere Möglichkeit ist es, durch einen gepulsten des hydrostatischen Druck (french press) 1,5 min nach 'Befruchtung' die zweite meiotische Teilung zu verhindern. Durch den Druckpuls kollabiert der Spindelapparat und der Polkörper wird nicht ausgestoßen. Die so gewonnenen Embryonen sind aufgrund der Rekombinationen in der ersten meiotischen Teilung nicht vollständig homozygot.

1.1.2. Der Medakafisch

Der Medakafisch (*Oryzias latipes*) gehört in die Gruppe der eierlegenden Zahnkarpfen (Cypridonitiformes). Er ist dem Zebrafisch in Größe und Entwicklung sehr ähnlich (Abb. 3). Seine Embryonalentwicklung dauert knapp 10 Tage, und er erreicht bereits nach 8-9 Wochen Geschlechtsreife. Er ist leicht zu halten und stellt keine hohen Anforderungen an die Wasserqualität. Sein natürliches Verbreitungsgebiet sind die Reisfelder Süd-Ost-Asiens. Er wird für genetische Studien genutzt und dazu schon seit Jahrzehnten im Labor gezüchtet. Es gibt gut getrennte, fast homozygote Inzuchtlinien, was sein wichtigster Vorteil für das effiziente, genetische Kartieren von Mutanten ist. In ihrer reproduktiven Phase werden sie in einem hell/dunkel-Rhythmus von 16 Stunden zu 8 Stunden gehalten. Nach 9 Monaten schließt sich dann eine dreimonatige Ruhepause mit einem hell/dunkel-Rhythmus von je 12 Stunden an. Von einem einzelnen, gut legenden Weibchen kann man täglich etwa dreißig Eier gewinnen, was einen geringen Mehraufwand im Vergleich zum Zebrafisch darstellt, da ein Pärchen mehrere Tage hintereinander 'angesetzt' werden muß, um z.B. bei einem Mutantenscreen eine genügend große Anzahl an Nachkommen von einem Elternpaar zu erhalten.

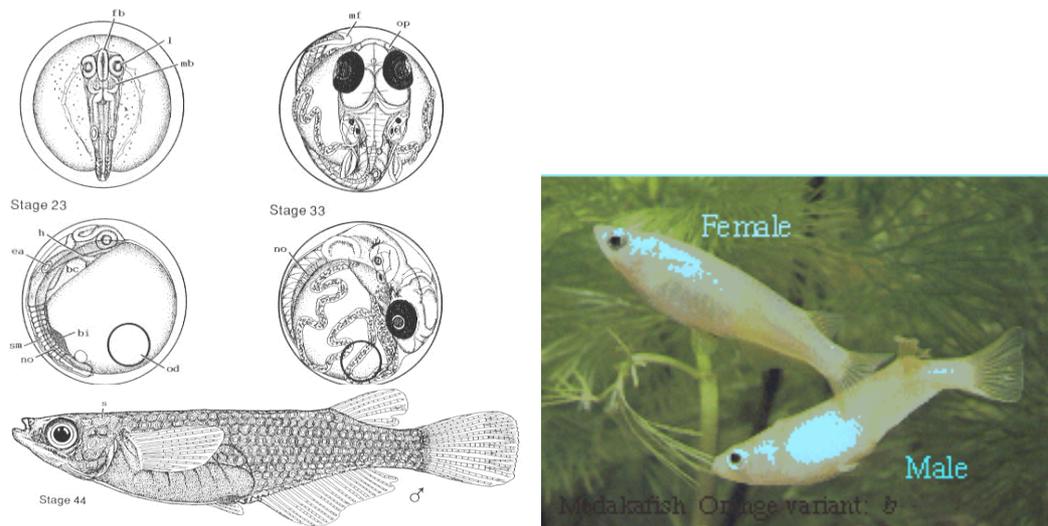


Abb. 3 Links: Dorsale und laterale Sicht von Stadium 23 (1 Tag, 17 Stunden) und Stadium 33 (4 Tage, 10 Stunden) Medaka-Embryonen, sowie eines adulten Männchen (aus (Iwamatsu, 1994)). Rechts: Fotos von der Farbmутante orange, die als Tester-Stamm für Mutagenese-Experimente verwendet wird.

Als Modellorganismus für Wirbeltierentwicklung ist er aufgrund der raschen Entwicklung, der einfachen Handhabung, dem transparenten, allerdings etwas harten Chorion und der relativ großen Anzahl von Nachkommen ebenfalls sehr gut geeignet. Alle Manipulationen, die am Zebrafisch vorgenommen werden können, - einschließlich der Parthenogenese - sind auch am Medakafisch möglich. Im Medakafisch gibt es darüber hinaus eine ES-Zell-Linie (Hong et al., 1998), bei der allerdings noch nicht gezeigt werden konnte, daß sie auch zur Keimbahn differenzieren kann. Das Medaka-Genom hat haploid eine physikalische Größe von weniger als $1,0 \cdot 10^9$ bp (Tanaka, 1995; Uwa and Itawa, 1981) und ist damit mindestens 40% kleiner als das des Zebrafisches. Die genetische Karte mit 227 RAPD Markern besteht aus 28 Kopplungsgruppen und hat eine

Größe von 2480 cM (Wada et al. 1995). Da der Karyotyp 24 Chromosomenpaare (Uwa and Iwata, 1981) enthält, ist die Karte somit noch nicht vollständig, sondern weist noch 4 Lücken auf.

Der Medakafisch gehört in die gleiche Überordnung (Acanthopterygii) wie die Pufferfische; mit dem Zebrafisch, der in die Gruppe der Ostariophysi gehört, ist er nur entfernt verwandt.

1.1.3. Der Fugufisch

Der Fugufisch (*Fugu rubripes*) hielt Einzug in die Molekulargenetik, als er von S. Brenner 1993 als Modellorganismus für die Genomanalyse vorgeschlagen wurde (Brenner et al., 1993). Das Fugu-Genom ist nur 400 Mb groß (Hindgardener et al. 1970), und somit 7,5 mal kleiner als das des Menschen. Nach Sequenzierung einer Reihe zufällig ausgewählter, kurzer genomischer Klone wurde das Genrepertoire als vergleichbar mit dem anderer Wirbeltiere geschätzt. Das kompaktere Genom wird vor allem durch eine geringere Anzahl repetitiver Sequenzen sowie durch eine Reduktion der Introngrößen und intergenischer Sequenzen erreicht.

Seit 1993 wurden eine Reihe von Genen im Fugu sequenziert, von denen einige eine deutliche Reduktion in der genomischen Größe aufwiesen. Als Beispiele seien hier nur das Chorea Huntington-Gen (Baxendale, et al., 1995), sowie die HoxA und HoxC-Cluster genannt (Apparicio et al., 1997). Einen eindrucksvollen Vergleich der Introngrößen von Mensch, Zebrafisch und Fugufisch bietet auch das Beispiel der Sperminsynthase (Abb. 4) (Boeddrich et al., 1999). Das HoxB-Cluster, sowie das homologe Gen zum PKD1-Gen sind dagegen in ihrer Größe kaum reduziert. Durch DNA- oder Protein-Sequenzvergleich von Fugu und anderen Vertebraten ist es möglich, konservierte und damit möglicherweise funktionell bedeutende Domänen zu identifizieren (Aparicio et al., 1995). Außerdem bieten Bereiche konservierter Syntenie zum Säugetiergenom einen Vorteil bei der positionellen Klonierung von Genen: Regionen, die im Menschen oder Zebrafisch Megabasen umfassen, sind im Fugu-Genom auf wenige 100 Kilobasen reduziert. Ein chromosomaler Abschnitt einer Art wird als synten zu einem chromosomalen Abschnitt einer anderen Art bezeichnet, wenn die beiden Abschnitte zwei oder mehr homologe Gene tragen. Ein Beispiel für Syntenie, konservierte Genordnung und gleichzeitig eine deutliche Reduktion in der Größe bietet der auf dem menschlichen Chromosom 14q24.3 lokalisierte 'Alzheimer Disease Locus' (AD3), der die Gene cFos, S31iii und S20i15 umfaßt, und eine genomische Region von 600 kb umspannt. Im Fugu-Genom liegen diese drei Gene in einer genomischen Region von lediglich 12,4 kb (Trower et al., 1996). Weitere Beispiele sind die Region des 'platelet derived growth factor receptor-β' (PDGFRβ) und des 'macrophage colony stimulating factor 1 receptor' (CSF1R) (How et al., 1996), die Region, die das 'tuberous sclerosis 2'-Gen (TSC2), das 'polycystic kidney disease 1'-Gen (PKD1) und das 'somatostatin type 5 receptor'-Gen (SSTR5) umfaßt (Sandford et al., 1996), sowie die homologe Region zu Xp22.1-Xp22.2 (Brunner et al., 1999), die in einem großen Bereich konservierte Syntenie aufweist, in der Region um das Sperminsynthase-Gen, das ebenfalls im humanen Xp22.2 lokalisiert ist, jedoch nicht (Boeddrich et al., 1999).

Pufferfische können in Gefangenschaft in der Regel nicht gezüchtet werden – es gibt lediglich einen Bericht von einem Züchtungserfolg in Gefangenschaft bei *Tetraodon fluviatilis* (Sterba, 1983). Der Einsatz des Pufferfisches ist daher auf die vergleichende Sequenzanalyse beschränkt, was ihm aufgrund des außerordentlich kleinen Genoms dennoch eine bedeutende Stellung einräumt.

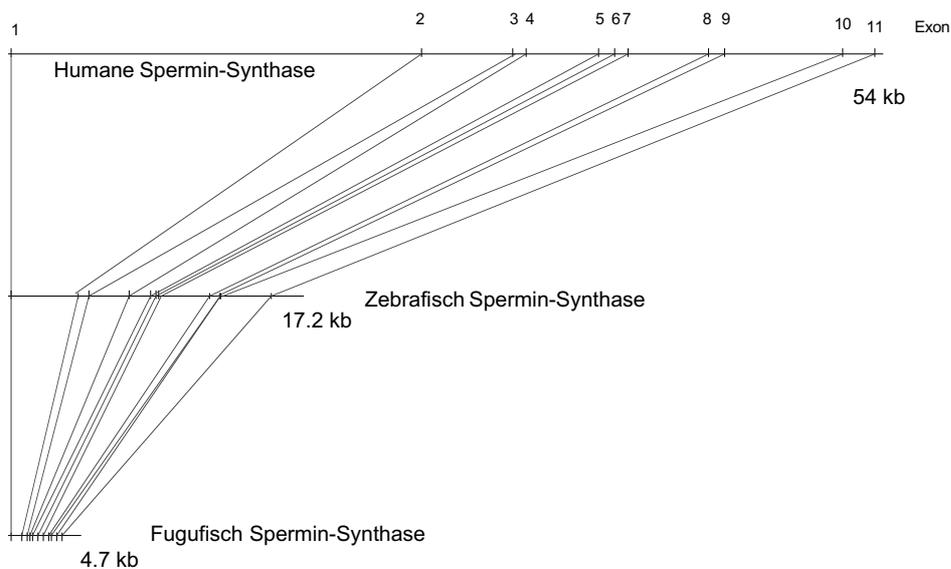


Abb. 4 Größenvergleich des Spermin-Synthase-Gens aus Mensch, Zebrafisch und Fugufisch; die sich entsprechenden Exons sind jeweils durch eine Linie verbunden.

1.1.4. Lanzettfisch

Der Amphioxus oder Lanzettfisch (*Branchiostoma floridae*) ist für das Verständnis der generellen Wirbeltierorganisation ein wichtiger Modellorganismus. Als Cephalochordate und Deuterostomier ist er den Wirbeltieren in der grundsätzlichen Anordnung von dorsal liegendem Nervensystem und ventral liegendem Darm ähnlich. Er ist allerdings deutlich primitiver als die Wirbeltiere organisiert; er hat z.B. nur eine einschichtige Epidermis, es liegen Sinnesorgane (Rückenmarksaugen) im Neuralrohr, der Kiemendarm enthält zahlreiche Kiemenspalten, die blindsackartige Leber ist hohl (Abb. 5). Auch in der Wirbeltier-Evolution kommt dem Amphioxus eine bedeutende Stellung zu. Die gängige Vorstellung ist, daß durch zwei Genomduplikationen in der Linie zwischen Amphioxus und den Wirbeltieren (Abb. 6, Punkte möglicher Duplikation sind durch ein D! gekennzeichnet) - das Genom Redundanz erlangte, und so Mutationen und Weiterentwicklung ermöglichte (Ohno, 1970). Das Amphioxusgenom sollte also eine Art ursprüngliches Vertebratengenom darstellen. Tatsächlich findet man einige Gene, die in Vertebraten in vierfacher Ausfertigung auftreten beim Amphioxus nur einmal (z.B. das *Hox*-cluster (Holland and Garcia-Fernandez, 1996), *Cdx* (Gamer and Wright, 1994), *MyoD* (Atchley et al., 1994), *Wnt*-Genfamilie (Sidow, 1992)). Gegen die Hypothese zweier, sukzessiver Genomduplikationen spricht nach Hughes jedoch der Befund, daß entwicklungsbiologisch bedeutende Gene in Vertebraten selten in der Form ((AB) (CD)) auftreten (Hughes, 1999), wobei jeweils A und B bzw. C und D eine größere Homologie untereinander aufweisen, als zu einem Mitglied der jeweils anderen paralogen Gruppe. Das Amphioxus-Genom wird auf 1,28 pg geschätzt (Atkin and Ohno, 1967), was einer haploiden Genomlänge von etwa 500 Mbp entspricht und ist damit verglichen mit Fugu (480 Mbp), und Medaka (800-1000 Mbp) relativ groß.

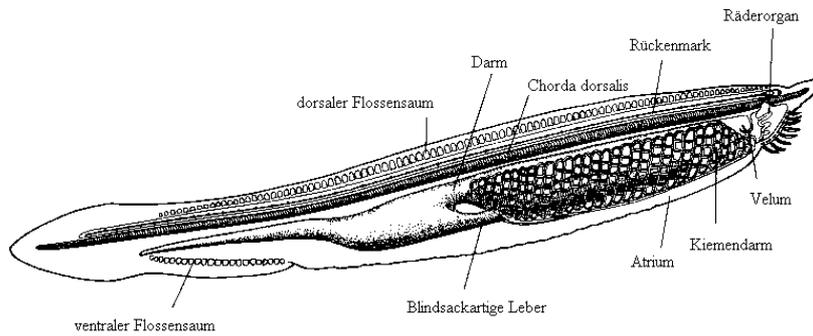


Abb. 5 Amphioxus (nach Kershaw, 1983). Die Gonaden sind zur besseren Übersicht nicht dargestellt.

Obwohl es nicht möglich ist, den vollständigen Lebenszyklus des Amphioxus in Gefangenschaft zu halten, läßt er sich dennoch als vereinfachtes 'Vertebratenmodell' für die Entwicklungsbiologie nutzen, da er während der Paarungszeit im Labor zur Ei- und Spermaablage gebracht werden kann. So können nach *in vitro*-Fertilisation synchrone Embryos und Larven für Genexpressionsstudien durch 'whole mount'- *in situ*-Hybridisierung und cDNA-Analysen sowie für morphologische Untersuchungen gewonnen werden. Außerdem können durch Sequenzanalyse mit Hilfe von Exon-Vorhersageprogrammen aus genomischer Sequenz Genstrukturen vorhergesagt werden (Boeddrich et al., 1999), die, sobald genügend Sequenzdaten zur Verfügung stehen, die Hypothese von eines Ur-Vertebratengenoms belegen oder widerlegen können.

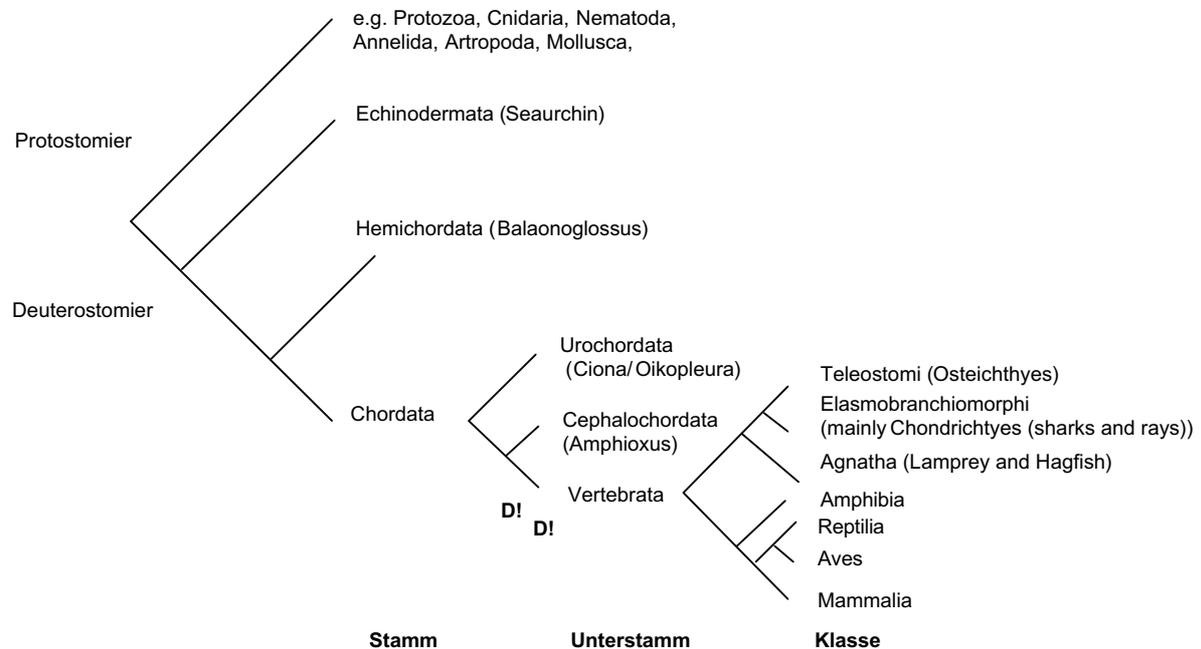


Abb. 6: Cladogramm der verwandtschaftlichen Beziehungen der Deuterostomier (nach Sidow, 1996).

1.2. Isolierung und Analyse von Genen

Es wird geschätzt, daß das durchschnittliche Vertebraten-Genom 80.000 - 100.000 transkribierte Gene enthält, die allerdings nur 0,5-5% des gesamten Genoms darstellen. Dieser kleine Teil wird gewöhnlich als kodierende DNA bezeichnet und auf ihn ist das Hauptinteresse der Genomforschung fokussiert. Um Gene zu identifizieren und zu analysieren, kann man grundsätzlich zwei verschiedene Strategien verfolgen:

- Die klassische Situation ist, daß man einen interessanten (mutanten) Phänotyp hat, und man daran interessiert ist, das betreffende Gen zu isolieren. Diese Strategie wird auch als 'forward genetics' bezeichnet.
- Die andere Möglichkeit ist, daß man ein Gen im Rahmen eines Sequenzierungs-Projekts identifiziert hat und dann wissen will, was die Funktion dieses Gens ist, bzw. was passiert, wenn man dieses Gen ausschaltet, zum falschen Zeitpunkt einschaltet, verstärkt oder vermindert exprimiert. Dieses Vorgehen wird auch als 'reverse genetics' bezeichnet.

Da bei den Zebrafisch- bzw. Medakafisch-'Screens' (Seite 15), die ja den Ausgangspunkt für diese Doktorarbeit bildeten, das Genom zufallsmäßig mutagenisiert wurde, die Analyse der Gene also von einem Phänotyp ausgeht, und die Werkzeuge für die 'reverse genetics' in Medaka und Zebrafisch momentan nur in beschränktem Ausmaß zur Verfügung stehen, werde ich nachfolgend auch den Schwerpunkt auf die Möglichkeiten der 'forward genetics' setzen.

Zunächst werden durch Kopplungsanalysen Marker bestimmt, die möglichst dicht mit dem entsprechenden Phänotypen gekoppelt vererbt werden (siehe 1.2.2. Genetisches Kartieren, Seite 16). Ausgehend von den flankierenden genetischen Markern wird dann eine physikalische Karte (siehe 1.2.3. Physikalisches Kartieren, Seite 23) erstellt, die die gesamte Region überspannt. Je nach Auflösung und Markerdichte der genetischen Karte, sowie der Rekombinationsfrequenz in der Region kann die entsprechende physikalische Karte noch mehrere Mb betragen. Durch weitere Marker wird die Kandidatenregion möglichst weit eingegrenzt, und es werden durch vergleichendes Kartieren, computergestützte Sequenzanalyse (siehe 1.2.4. Sequenzanalyse, Seite 24), 'Exon trapping' oder 'cDNA selection' Kandidatengene isoliert. Diese werden dann auf Mutationen analysiert. Durch einen 'Mutant-rescue' kann schließlich der Zusammenhang von Gen und Phänotyp bewiesen werden. Dabei wird analysiert, ob sich der Phänotyp der Mutante durch Einbringen von nicht mutanter DNA oder RNA in den Wildtyp konvertieren läßt.

1.2.1. Mutagenese-Screen

Ziel eines genomweiten, gesättigten Mutagenese-Screens ist es, jedes Gen eines Organismus durch ein Mutagen zu treffen, durch Kreuzung zur Homozygotie zu bringen und dann den Phänotypen zu analysieren. Als Mutgene kommen zum einen Röntgen- oder γ -Strahlen in Frage, die jedoch meist Chromosomenbrüche und

größere Deletionen zur Folge haben. Weitere potente Mutagene, die darüber hinaus den Vorteil haben, daß sie das getroffene Gen mit einem Etikett (tag) versehen, sind Transposons oder Retroviren. Sowohl im Medakafisch als auch im Zebrafisch konnten durch co-Injektion von einem Transposon (sleeping beauty) und Transposase mutante Phänotypen erzeugt werden (Henrich und Wittbrodt, pers. Kommunikation, Ivics et al., 1997). Aber weder im Zebrafisch noch im Medaka sind bisher Methoden, die eine effiziente und kontrollierte Anwendung der Transposon-Mutagenese ermöglichen würden, etabliert. Im Zebrafisch wurde außerdem erfolgreich durch den Vesicular Stomatitis-Virus mutagenisiert (Amsterdam et al., 1997). Allerdings ist, wie von den Autoren eingeräumt wird, die Mutationseffizienz um etwa den Faktor 7-10 geringer als bei der chemischen Mutagenese.

Die chemische Mutagenese hat den Vorteil, daß sie ein höchstes Maß an Zufällmäßigkeit bietet, während die 'biologischen' Mutagenesesysteme bestimmte Chromatinstrukturen bevorzugen. Für die chemische Mutagenese stehen verschiedene Agenzien zur Verfügung, von denen sich zwei in den letzten Jahren besonders durchgesetzt haben. Zum einen können durch Substanzen, die eine Quervernetzung der DNA bewirken, wie z.B. Psoralen (Abb. 7), Deletionen von einigen hundert Basenpaaren hervorgerufen werden. Und zum anderen können durch alkylierende Substanzen, wie ENU (Abb. 7) oder EMS die Paarungseigenschaften der Basen dahingehend verändert werden, daß eine Punktmutation entsteht.

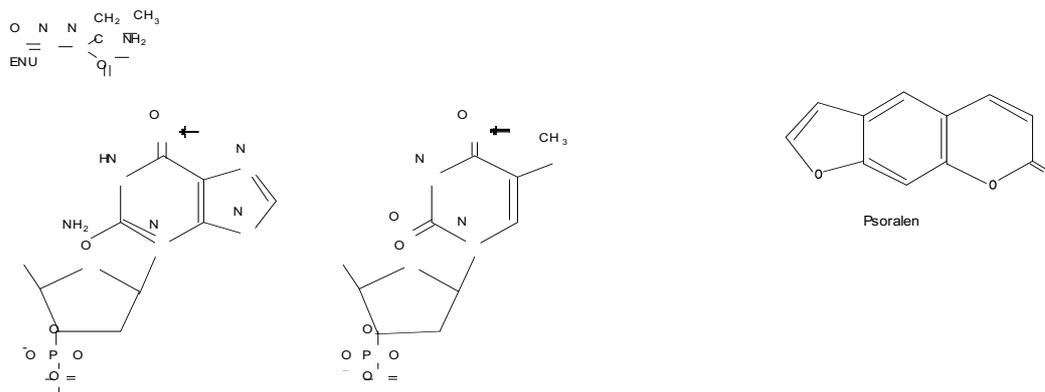


Abb. 7: ENU und Psoralen. - Die Pfeile kennzeichnen die Stellen, die nach Alkylierung eine Änderung der Paarungseigenschaften der Basen zur Folge haben. O⁶-Methylguanin kann mit Thymidin und O⁴-Methylthymidin kann mit Guanin paaren. Psoralen verursacht eine Quervernetzung der DNA und damit Deletionen.

Die beiden großen Zebrafisch-'Screens' in Boston und Tübingen wurden durch ENU-Mutagenese durchgeführt. Dazu wurden adulte Wildtyp Männchen mehrere Male im Abstand von einer Woche für einige Stunden in ENU-haltiges Wasser gesetzt und anschließend mit Wildtyp-Weibchen ausgekreuzt. Die Nachkommenschaft ist dann heterozygot für eine Reihe von Mutationen. Aus der F₁ werden Geschwister verpaart, die jeweils für ein mutagenisiertes Genom haploid sind – der Einfachheit halber ist in Abb. 8 nur ein

Genom als mutagenisiert dargestellt. Eine Mutation, ist dann bei jeweils 50% der F₂-Nachkommen heterozygot. Von einer Reihe von F₂-Verpaarungen werden die F₃-Embryonen analysiert; waren beide Eltern heterozygot (in etwa 25% der Testkreuzungen) für die gleiche Mutation, wird ein Viertel der F₃-Nachkommenschaft den mutanten Phänotyp zeigen. Insgesamt wurden in beiden 'Screens' etwa 5000 F₂-Familien aufgezogen und mehr als 6500 Mutationen identifiziert, von denen etwa 1900 spezifische Mutationen in der Embryonalentwicklung zeigten (Haffter et al., 1996; Driever et al., 1996). Von den 1200 Mutanten des Tübingen-Screens wurden knapp 900 auf Komplementationen untersucht und konnten so 372 Genen zugeordnet werden (Haffter et al., 1996).

Das Klonieren der betroffenen Gene soll nun durch eine kombinierte Strategie von Kandidatengen-Versuchen und klassischem, positionellen Klonieren erfolgen. Für den 'candidate-gene-approach' werden Gene, die bereits in anderen Spezies kloniert sind, und von denen man aufgrund eines ähnlichen Phänotyps und Expressionsmuster in andern Organismen die Vermutung hat, daß sie betroffen sein könnten, direkt getestet, während bei positioneller Klonierung zunächst eine genetische Karte erstellt werden muß.

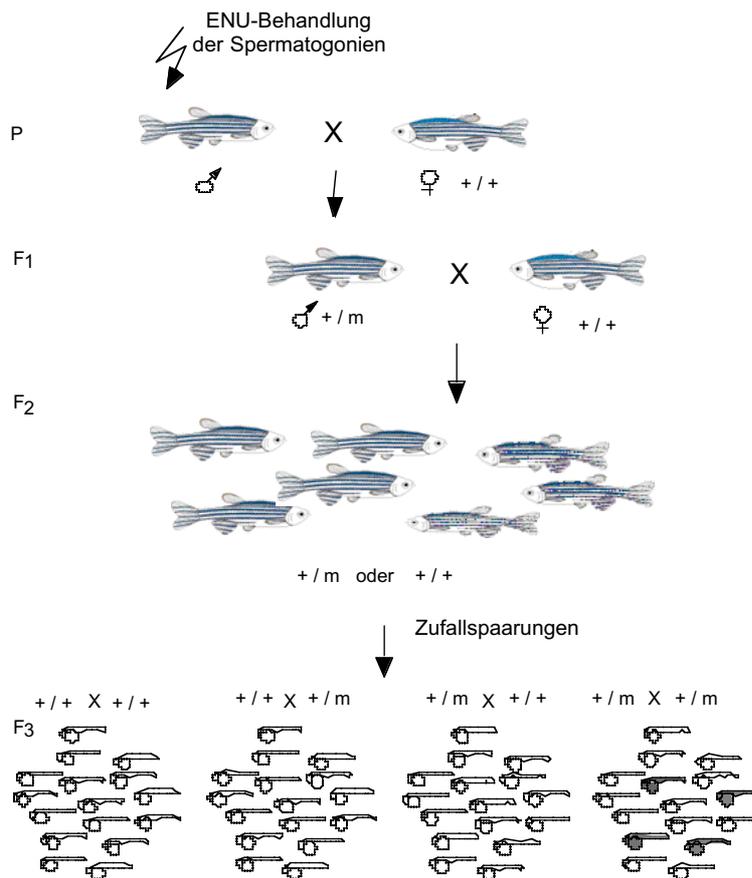


Abb. 8: Drei-Generationen-Mutagenese-Screen nach (Haffter et al., 1996). Für Details siehe Text.

1.2.2. Genetisches Kartieren

Genetische Kopplung zwischen zwei Merkmalen wurde entdeckt, lange bevor bekannt war, daß die Träger der Erbinformationen lineare DNA-Moleküle sind. Schon 1911 fiel T.H. Morgan und A. Sturtevant auf, daß bei *Drosophila melanogaster* die Mutationen vestigal und purple nicht der Mendelschen Regel von der freien Rekombinierbarkeit der Gene gehorchten und meist gekoppelt vererbt werden. Der Prozentsatz, mit dem auf einem Chromosom gekoppelte Marker nicht zusammen vererbt werden, ergibt deren Entfernung in centiMorgan (cM) auf einer genetischen Karte. So wurden im Laufe des letzten Jahrhunderts mutante Loci, hauptsächlich in *Drosophila*, der Maus und einigen Pflanzen durch genetische Kopplungsanalyse kartiert, ohne jedoch dadurch die biochemischen Grundlagen dieser Defekte zu erhellen. Klonierte Gene hingegen beinhalten wertvolle Informationen und sind unabdingbar für vergleichendes Kartieren, in Kreuzungen nahe verwandter Stämme sind sie aber oft nur wenig polymorph und daher nur sehr aufwendig zu kartieren. Die häufigsten Marker auf genetischen Karten sind daher anonyme DNA-Fragmente, wie Restriktionsfragments-Längenpolymorphismen (RFLPs), Mini- und Mikrosatelliten-Polymorphismen, oder zufallsmäßig von kurzen Primern amplifizierte polymorphe DNAs (randomly amplified polymorphic DNA = RAPD) und polymorphe PCR-Produkte, die von eingestreuten, repetitiven Sequenzen ausgehend amplifiziert wurden (interspersed repeat sequence PCR = IRS-PCR). Diese anonymen DNA-Fragmente sind häufig auch zwischen nahe verwandten Stämmen polymorph und sind daher in groß angelegten Kartierungsprojekten die Marker der Wahl.

Einen Sonderfall der genetischen Kopplungskarte stellt die vergleichende Karte dar. Durch konservierte Syntenie ist es möglich, ganze Chromosomenabschnitte verschiedener Organismen miteinander zu korrelieren. Je näher zwei Arten miteinander verwandt sind, desto größere Chromosomenabschnitte weisen konservierte Genordnung auf (siehe auch 1.1.3). Durch die vergleichende Kartierung ist es möglich, Informationen von einer gut kartierten Spezies (z.B. Mensch und Maus) zu einer weniger gut kartierten Spezies (verschiedene Haustiere) zu transferieren, aber auch umgekehrt Merkmale, die im Menschen nur schlecht untersucht werden können, von einer schlecht kartierten Art, in der die Merkmale aber gut untersucht werden können, auf die menschliche Genkarte zu übertragen.

1.2.2.1. Genetische Karten im Zebrafisch

Die genetische Kartierung im Zebrafisch geht bereits auf 1986 zurück (Streisinger et al., 1986). Es wurde gezeigt, daß die Mutationen *gol-1*, *gol-2*, *alb-1* und *spa-1* nicht gekoppelt sind. Außerdem wurde ihre Entfernung vom Centromer durch Halb-Tetraden-Analyse bestimmt. Dazu werden Eier heterozygoter Weibchen mit UV-bestrahlten Spermien aktiviert und dann einem 'early pressure' (Abb. 2) ausgesetzt. Die so entstehenden Halbtetraden zeigen den parental Phänotyp, wenn die Gene gekoppelt sind. Sind sie nicht gekoppelt, zeigen die so erzeugten Nachkommen genauso oft den parental Phänotyp wie den nicht parental. S. Johnson schlug 1995 vor, Mutationen relativ zum Centromer zu kartieren. Dazu wird ihre Lage relativ zu centromernahen Markern, die mit wenigen einfachen PCR-Reaktionen, wie es bei RAPD- oder (CA)_n-repeat Markern möglich ist, bestimmt (Johnson et al., 1995; Kauffman et al., 1995).

RAPDs sind die Marker, die am schnellsten und preiswertesten hergestellt werden können. Ein kurzer Primer mit willkürlicher Sequenz amplifiziert unter wenig stringenten PCR-Bedingungen genomische Fragmente, die über das gesamte Genom verteilt sind. Ein gewisser Teil dieser DNAs ist zwischen verschiedenen Stämmen polymorph und spiegelt damit Sequenzunterschiede wieder, die die Primerbindungsstellen betreffen. Hauptnachteil der RAPD-Marker ist die hohe Variabilität der Produkte, da die PCR-Reaktionen unter wenig stringenten Bedingungen durchgeführt werden, und so in verschiedenen Laboratorien zu unterschiedlichen Ergebnissen führen können. Ein weiteres Problem der RAPD-Marker ist, daß sie in der Regel dominant sind, und es daher nicht möglich ist, zwischen homozygoten und heterozygoten Tieren zu unterscheiden. In einer F_1 -Geschwisterkreuzung (inter-cross) sind daher nur durchschnittlich ein Viertel der Tiere, nämlich die mit dem Null-Allel informativ. Bei der ersten veröffentlichten genetischen Kopplungskarte des Zebrafisch (Postlethwait et al., 1994) wurde dieses Problem umgangen, indem die Marker auf haploiden Nachkommen eines heterozygoten ABxDAR-Weibchens kartiert wurden. Diese erste Karte gruppierete etwa 400 informative RAPDs in 29 Kopplungsgruppen und wies damit noch 4 Lücken auf, da das haploide Genom des Zebrafisch aus nur 25 Chromosomen besteht (Daga et al., 1996; Endo and Ingalls, 1968). Auf einem weiteren haploiden 'Mapping-panel' von einem C32xSJD-Weibchen konnten durch zusätzliche Marker die Lücken geschlossen (Johnson et al., 1996) und die Länge der genetischen Karte auf 2900 cM bestimmt werden. Es werden im Moment Anstrengungen unternommen, zumindest einen Teil der RAPDs zu klonieren, zu sequenzieren und so in STS (=sequence tagged sites) umzuwandeln.

Die zweite genetische Kopplungskarte, die für den Zebrafisch erstellt wurde, beruht auf CA-repeat Markern (Kauffman et al., 1995; Knapik et al., 1996; Knapik et al., 1998). Goff et al. konnten 1992 zeigen, daß diese Marker auch erfolgreich im Zebrafisch angewendet werden können (Goff et al., 1992). Diese Mikrosatelliten-Marker beruhen auf der Tatsache, daß CA-'Repeats' in ihrer Länge hochvariabel sind. Sie werden hergestellt, indem aus einer genomische Bank mit relativ kleinen Inserts diejenigen isoliert werden, die mit einem $(CA)_n$ -Primer hybridisieren. Bei den positiven Klonen werden dann die Sequenzen, die die repetitive Basenfolge flankieren, bestimmt und Primer generiert, die das repetitive Element überspannen. Elektrophoretisch können dann die Längenunterschiede der repetitiven Sequenz bestimmt werden. Die CA-repeat-Marker sind relativ aufwendig in der Herstellung: Es müssen DNA-Fragmente kloniert und sequenziert werden, Primer müssen synthetisiert und getestet werden. Häufig lohnt sich jedoch der Aufwand, da die Marker zwischen verschiedenen Kreuzungen transferiert werden können. So sind 55% der CA-repeat Marker, die zwischen den Stämmen IND und AB informativ sind, es auch zwischen AB und TUE (Knapik et al., 1998). Von 1777 Primer-Paaren, die innerhalb des Kartierungs-Projektes bisher synthetisiert worden sind, waren 690 informativ – der Rest ergab entweder kein PCR-Produkt der erwarteten Länge (582), segregierte nicht nach den Mendelschen Regeln (153), war nicht polymorph (240) oder wies ein oder mehr gemeinsame Allele zwischen den Stämmen auf (112). Die CA-repeat Karte ist mit etwa 2300 cM deutlich kleiner als die erste RAPD-Karte, die 2700 cM überspannt. Das kann dadurch erklärt werden, daß die weibliche Meiose, auf der die RAPD-Karte beruht, normalerweise mehr Rekombinationen zeigt, als die weibliche, während die CA-repeat Karte männliche und weibliche Meiosen gleichermaßen berücksichtigt.

Um das Zebrafisch-System vollständig nutzen zu können, ist es erforderlich auch Gene (cDNAs) auf den vorhandenen Kopplungskarten positionieren zu können. Dazu wird die Tatsache ausgenutzt, daß cDNAs in ihrer 3'-UTR häufig weniger konserviert sind als in ihrer kodierenden Sequenz und sich selbst kleine

Sequenzunterschiede einzelsträngiger DNAs auf einem nicht denaturierenden Gel nachweisen lassen (SSCP = single-strand conformation polymorphism). So waren von mehr als 100 bisher auf SSCP-Polymorphismus getesteten Zebrafisch-cDNAs 31% zwischen den Stämmen AB und IND und 41% zwischen den Stämmen C32 und SJD polymorph. Zehn cDNAs konnten so auf einer der beiden oben beschriebenen Karten plaziert werden (Fornzler et al., 1998), sowie weitere 105 SSCPS auf einer weiteren Karte, die aus 48 haploiden Nachkommen eines TUE x TL –Weibchen besteht (Gates et al., 1999). SSCP-Marker sind ebenfalls aufwendig in ihrer Herstellung und Handhabung; allerdings stehen durch die verschiedenen EST-Projekte im Zebrafisch bereits Sequenzinformationen von einer Reihe von cDNAs zur Verfügung, die so zur Kartierung ausgenutzt werden können, und die so kartierten Loci tragen mehr biologische Information als anonyme DNA-Sequenzen.

1.2.2.2. Genetische Karte im Medakafisch

Der entscheidende Vorteil des Medakafisches gegenüber dem Zebrafisch ist die Existenz von gut getrennten Stämmen, die nach mehr als 17 Generationen aufeinanderfolgender Geschwisterverpaarungen annähernd als Inzuchtlinien betrachtet werden können.

Die erste genetische Kopplungskarte im Medaka wurde, wie auch im Zebrafisch hauptsächlich auf der Grundlage von RAPD-Markern erstellt (Wada et al., 1995). Mit 161 RAPD-Markern, 3 Farbmutationen und 5 Enzymloci konnte das Genom in 28 Kopplungsgruppen, die insgesamt 2480 cM überspannen, geordnet werden. Da das Medakagenom aus 24 Chromosomen besteht (Uwa and Itawa, 1981), weist diese Karte noch 4 Lücken auf. Das Medakagenom wird auf weniger als $1,0 \times 10^9$ bp geschätzt. Damit entspricht ein cM etwa 325 kbp - zum Vergleich: im Zebrafisch entspricht ein cM etwa 625 kb (Postlethwait et al., 1994). Das ist um so bemerkenswerter, als die Medaka-Karte ausschließlich auf männlichen Meiosen basiert. Ein F₁-Männchen aus einer Kreuzung von einem Südstamm-Weibchen (Hd-rR) und einem Nordstamm-Männchen (HNI) wurde gegen Südstamm-Weibchen zurückgekreuzt, und jeweils 20-150 Nachkommen wurden untersucht. Diese erste Medaka-Karte hat damit eine Auflösung von 0,7-5 cMorgan.

1.2.2.3. Kartierung durch Hybridisierung

Alle oben erwähnten Strategien, eine genetische Karte zu erstellen beruhen auf PCR-Technologie, bei der anschließend die Produkte auf einem Polyacrylamid- oder Agarose-Gel analysiert werden. Da auf einem Gel immer nur eine begrenzte Anzahl von Spuren zur Verfügung stehen, können auch nur immer ein begrenzte Anzahl von Tieren gleichzeitig analysiert werden. Eine Alternative dazu ist ein Methode, die auf Hybridisierung beruht. Hier können im Prinzip unbegrenzt viele Tiere gleichzeitig untersucht werden, und zudem können viele Hybridisierungen simultan durchgeführt werden. Drei Möglichkeiten für eine Kartierungsstrategie auf Hybridisierungsgrundlage sind bisher in Genomprojekten realisiert worden. Die erste ist die Detektion von RFLPs, bei der Southernblots von mit einem Restriktionsenzym behandelten DNAs mit (anonymen) Proben untersucht werden. Diese scheidet im großen Maßstab bei den kleinen Fischen allerdings aus, da zur Herstellung der Southernblots relativ viel DNA benötigt wird und ein Southernblot selbst bei vorsichtiger Handhabung nur eine begrenzte Lebensdauer hat. Die zweite Möglichkeit beruht auf IRS-PCR und

wurde das erste Mal in großem Maßstab in dem Maus-Projekt eingesetzt (McCarthy et al., 1995). Und die dritte ist eine modifizierte AFLP-Technologie, die ebenfalls zum ersten Mal in der Maus eingesetzt wurde (Himmelbauer et al., 1998).

Die IRS-PCR wird benutzt, um zwischen zwei Mitgliedern einer hochrepetitiven, verstreuten DNA-Familie liegende, noch uncharakterisierte Sequenzen zu vermehren (Nelson et al., 1989). Gut charakterisierte Beispiele solcher Familien findet man im Menschen mit der Alu-Familie und in der Maus mit der entsprechenden B1-Familie. Hat eine Familie genügend viele Kopien im Genom, kann man mit einem einzigen Primer dazwischen liegende, nicht repetitive Sequenzen vermehren; die Primersequenz muß hierbei der Konsensus-Sequenz der repetitiven DNA-Familie entsprechen und sollte möglichst an einem Ende der repetitiven Einheit liegen. Eine solche Ein-Primer-PCR ist natürlich nur dann möglich, wenn Paare nahe beieinander liegender Wiederholungen entgegengesetzt orientiert sind, so daß konvergierende Primer binden können. Bei einer IRS-PCR von einem komplexen Genom entsteht so eine Vielzahl von Produkten, die entweder noch als einzelne Banden nach elektrophoretischer Auftrennung sichtbar sind, oder einen kontinuierlichen Schmier bilden. Viele der kurzen, eingestreuten, repetitiven Elemente (SINEs) enthalten ähnlich wie tRNA-Gene einen *poIII*-Promotor, und können *in vitro* transkribiert werden (Fuhrman et al., 1981). Daher wird vermutet, daß sie durch Transkription und reverser Transkription erneut ins Genom integrieren können. Zwischen zwei Genomen führen sowohl Unterschiede, die die Primerbindungsstelle betreffen z.B. durch eine Punktmutation oder eine kleine Deletion, als auch solche, die die Position der SINEs betreffen, ähnlich wie bei den RAPDs zu einem anwesend/abwesend (+/-)-Polymorphismus (Abb 9).

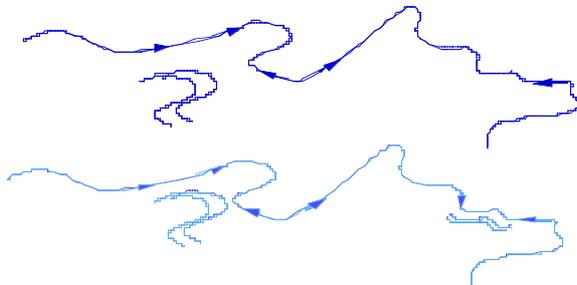


Abb. 9: IRS-PCR von zwei verschiedenen Genomen. Es werden nur Produkte amplifiziert, die zwischen zwei konvergierenden, bis zu 3 kb entfernten Elementen liegen.

Sowohl im Zebrafisch- als auch im Medakafisch-Genom sind inzwischen einige repetitive Elemente bekannt. Einen Überblick geben die beiden Tabellen 1 und 2. Nicht alle Elemente sind allerdings als Grundlage für eine IRS-PCR geeignet. Die Alu-Elemente des Zebrafischs haben beispielsweise nichts mit den menschlichen Alu-Elementen zu tun. Es handelt sich hauptsächlich um tandemartig hintereinanderliegende Elemente, die im centromernahen Heterochromatin vorkommen, und nur vereinzelt verstreut liegen. Bei den AluI Typ Ia-Elementen und der Satelliten-Familie Ia handelt es sich um die gleiche Sequenz, die unabhängig voneinander veröffentlicht wurde. Bei der Dana- und der Mermaid-Familie handelt es sich ebenfalls um die gleiche Sequenz, die zeitgleich, unabhängig voneinander veröffentlicht wurde. Die Mermaid-Elemente kommen sowohl im Zebrafisch- als auch im Medaka-Genom vor. Im Medaka-Genom sind zwei weitere verstreut

liegende repetitive Elemente veröffentlicht, die als *Oryzias latipes* repeat 1 (OLR1) und Medaka repeat element (MRE = OLR2) bezeichnet werden (Tabelle 2).

Name	Struktur	Weitere charakteristische Eigenschaften	Kopien/haploidem Genom	% des Genoms	Referenz
AluI repeats Type Ia Type II	tandemartig und eingestreut 195bp repeat 190bp repeat	A+T-reich (65%) G+C-reich (65%)	500.000 90.000	4-5% 0,5%	He et al. (1992)
'Satellite family' Type Ia Type Ib	tandemartig 200 bp repeat eingestreut und tandemartig 180 bp repeat	A+T-reich (65%) A+T-reich (65%)	800.000 22.000	8 % 0,2%	Ekker et al. (1992)
Mermaid	ca. 300 bp eingestreut	Homologie zu tRNA-Genen Auch in anderen Vertebraten	12.000	ca. 1%	Shimoda et al. (1995)
Dana Retroposon	eingestreut	Homologie zu tRNA-Genen Danio spezifisch	400.000	ca. 10%	Izsvak et al. (1995)
TcI-like transposon	TdrI TdrII eingestreut 1250 bp 1500 bp		1000	ca. 0,1%	Izsvak et al. (1995)
Oops = Angel		fold-back ca 350 bp	?	?	Izsvak et al. (1995)
B13	eingestreut 350 bp		?	?	S. Johnson (pers. Mitteilung)

Tabelle 1: Repetitive Sequenzen im Zebrafisch-Genom.

Name	Struktur	Weitere charakteristische Eigenschaften	Kopien/haploidem Genom	% des Genoms	Referenz
Mermaid	eingestreut	t-RNA-Struktur	3370	0,12%	Burgtorf et al. (in Vorbereitung)
OLR1 = Olrep1	eingestreut	160 bp	5000	0,11%	Naruse et al. 1992
OLR2 = MRE	eingestreut	220 bp	4800	0,14%	Uchiyama et al 1996

Tabelle 2: Repetitive Elemente im Medakafisch-Genom.

Die 'mermaid'-Elemente weisen eine charakteristische Struktur von bis zu vier konservierten Domänen mit dazwischen liegenden variablen Sequenzen auf (Abb. 10). Oft sind aber auch nur die Abschnitte von konservierter Box #1 bis konservierter Box #2 vorhanden. In der Box #1 befinden sich BoxA und BoxB die charakteristisch für einen RNA-PolymeraseIII-Promotor sind. Ein Sequenzvergleich der bisher in Genebank

veröffentlichten 'mermaid'-Elemente, sowie der OLR1- und OLR2-Sequenzen befindet sich im Anhang dieser Arbeit (Seite 104-106).

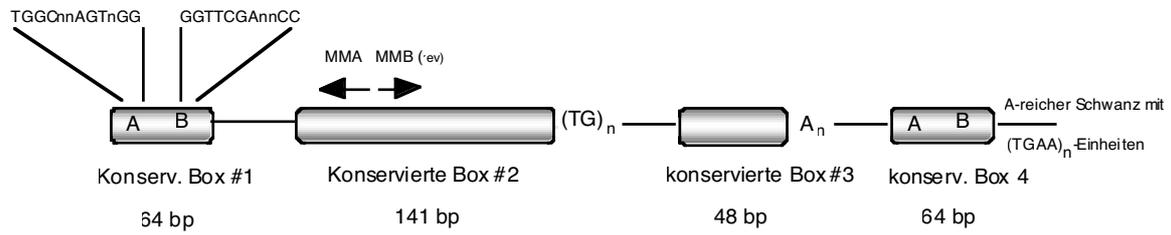


Abb. 10: Schematische Darstellung eines Mermaid-Elements.

Klonierte IRS-PCR-Produkte stellen in der Regel 'single-copy'-Proben dar und können daher durch Hybridisierung kartiert werden. Sie werden dazu auf Filter hybridisiert, auf die zuvor IRS-PCR-Produkt von einzelnen Tieren einer Rück- oder Geschwisterkreuzung gespottet wurden (Abb. 11).

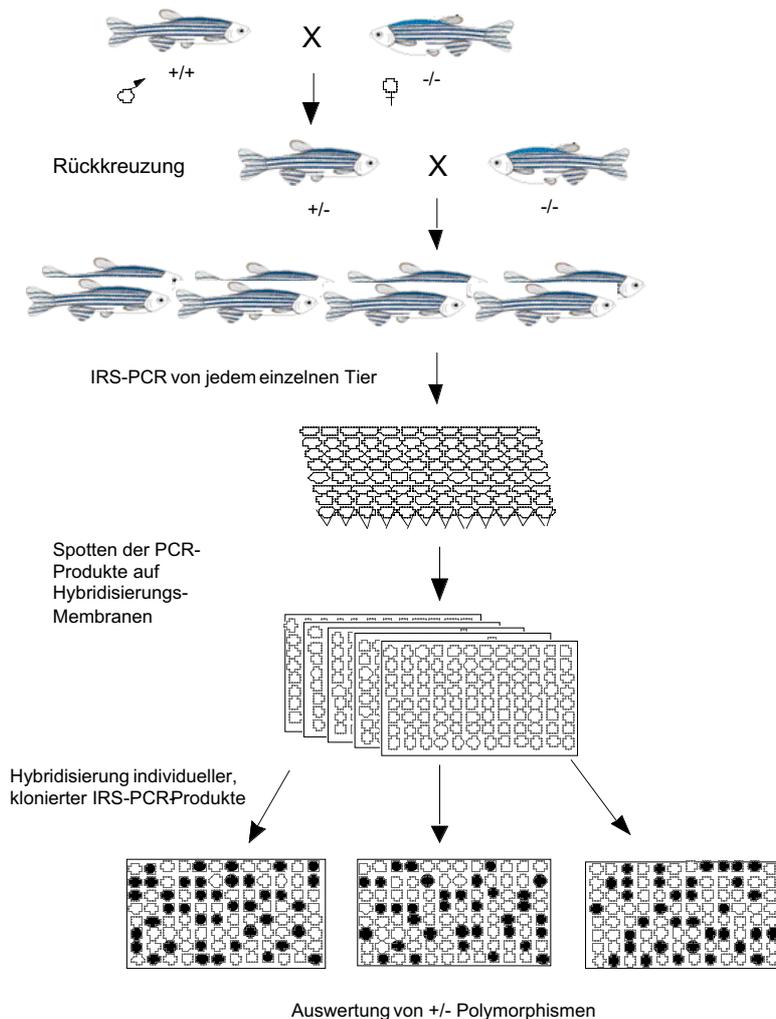


Abb. 11: Kartierung durch Hybridisierung von IRS-PCR-Produkten

Die selektive PCR-Amplifikation erfolgt nach Anligieren von Adaptoren, mittels Primern, die über die Restriktionsstelle hinaus einige Basen in das Restriktionsfragment reichen. In der ursprünglichen Version wird dabei durch den spezifischen Primer die Anzahl der Produkte soweit reduziert, daß von der gesamten genomischen DNA nur noch ca. 50-70 Fragmente amplifiziert werden, die dann auf einem Polyacrylamid-Gel analysiert werden können (Vos et al., 1995).

Bei der modifizierten AFLP-Technik (Himmelbauer et al., 1998) ist man nicht mehr darauf angewiesen, die amplifizierten Fragmente auf einem Gel auflösen zu können. Durch die Kombination von Restriktionsenzym und spezifischen Primern wird ein Amplikon hergestellt, das in seiner Komplexität verglichen mit gesamter genomischer DNA so weit reduziert ist, daß es effizient sowohl als Hybridisierungsprobe auf Filtern mit genomischen Genbanken als auch als Hybridisierungs-'Target' auf Filtern mit Amplikons von Tieren einer Rückkreuzung eingesetzt werden kann. Im Gegensatz zur IRS-PCR-Methode wird hier bereits durch den polymorphen Marker (Restriktionsfragment eines genomischen Klonen) eine Verknüpfung von genetischer und physikalischer Karte erreicht.

1.2.3. Physikalisches Kartieren

Hat man eine Mutation auf einer genetischen Kopplungskarte kartiert, ist der nächste Schritt zur positionellen Klonierung des Gens die Erstellung einer physikalischen Karte, die sich über den Bereich der beiden flankierenden Marker erstreckt. Die Entfernungen auf physikalischen Karten werden in Basenpaaren (bp) oder Vielfachen davon (kb, Mb) gemessen. Abhängig davon, wie weit die flankierenden Marker auseinander liegen, wird zunächst eine Karte relativ niedriger Auflösung erstellt. Dazu werden überlappende Klone aus Genbanken mit großer Insertgröße, wie YAC-Banken isoliert, die die gesamte Region überspannen. Man benutzt dazu Proben, die man aus den Klonen isoliert hat, oder die zuvor bereits auf den Klonen lokalisiert wurden, als Sonde und 'screened' damit erneut die Genbanken. Die so identifizierten Klone werden wiederum benutzt, um weitere Klone zu isolieren und so ein *Contig* zu erstellen.

Bei einer Genomgröße von $1,7 \cdot 10^9$ bp und einer geschätzten Anzahl von 100.000 Genen im Zebrafisch beginnt durchschnittlich alle 17 kb ein neues Gen. In einem *Contig* von 625 kb (durchschnittlich 1 cM) können also mehr als 36 Gene liegen. Daher versucht man die Region, in der das mutante Gen liegt, durch die Isolation weiterer Marker so weit wie möglich einzugrenzen. Schließlich wird eine hochauflösende physikalische Karte mit überlappenden Klonen aus Genbanken mit kleinerer Insertgröße (PACs, BACs oder Cosmide), die in ihrer Handhabung einfacher sind als YACs, erstellt, und es werden cDNAs auf der Karte positioniert. Dazu können die genomischen Klone in 'Exon-Trapping'- oder 'cDNA-Selection'-Experimenten sowie als Sonden auf cDNA-Banken eingesetzt werden (Del Mastro et al., 1995; Duyk et al., 1990; Evin et al., 1990). Das Ziel ist es letztendlich die Basenabfolge der DNA in den Bereichen, die für die Mutation relevant sein können, zu bestimmen. Dazu können genomische Klone auch direkt sequenziert werden, und durch Computeranalyse kodierende Bereiche identifiziert werden (Seite 24).

Für die physikalische Kartierung im Zebrafisch und Medakafisch stehen bereits eine Reihe von Ressourcen zur Verfügung. Für das Zebrafisch-Genom existieren zwei YAC-Banken (Zhong et al., 1997);

Silverman, nicht publiziert), eine BAC-Bank, eine PAC-Bank (Amemiya and Zon, 1999) und eine Cosmid-Bank (Burgtorf et al., 1998), die entweder über das Ressourcen Zentrum in Berlin (YACs, PAC, Cosmid), Research Genetix (Silverman-YAC) oder Genome Systems (BAC, PAC, Zhong-YAC) erhältlich sind. Zusammen entsprechen diese Banken etwa zwanzig Zebrafisch-Genom-Äquivalenten. Im Medakafisch stehen erst seit kurzem zwei Cosmid-Banken von etwa vier Genom-Äquivalenten (Burgtorf, in Vorbereitung) und eine BAC-Bank von etwa acht bis zehn Genom-Äquivalenten (Amemiya, in Vorbereitung) zur Verfügung. Diese sind über das Ressourcen Zentrum in Berlin erhältlich.

Als weitere Möglichkeit können zur physikalischen Kartierung somatische oder Radiations-Zellhybride ausgenutzt werden. Bei somatischen Zellhybriden werden jeweils ein oder wenige Chromosomen in eine Zelllinie transferiert. Für den Zebrafisch gibt es bereits ein relativ gut charakterisiertes 'Panel' (Chevrette et al., 1997). Radiations-Hybride bieten die Möglichkeit kleinere Chromosomenfragmente, die durch γ -Strahlung erzeugt wurden, und dann durch Fusion mit einer Zelllinie kloniert wurden, zu untersuchen. Auch hier gibt es im Zebrafisch ein 'Panel' (Kwok et al., 1998), das im Moment als wichtigste Grundlage für die Kartierung von cDNAs genutzt wird. Daß die Methode des positionellen Klonierens insgesamt sehr schwierig und aufwendig ist, zeigt sich in der Tatsache, daß selbst 6 Jahre nach den großen 'Screens' im Zebrafisch lediglich die Mutationen *oep* (one eyed pinhead) (Zhang et al., 1998) und *sau* (sauternes) (Brownlie et al., 1998) durch diese Strategie kloniert worden sind.

Die Alternative zum Erstellen einer physikalischen Karte lediglich für die Bereiche, in denen zuvor eine interessante Mutation identifiziert worden ist, ist der Versuch, global das gesamte Genom eines Organismus in *Conitgs* überlappender genomischer Klone zu ordnen. Diese stellen dann auch die Grundlage für Sequenzierungsprojekte dar, die für einige Modellorganismen initiiert worden sind. Hauptvorteil einer solchen genomweiten Strategie ist, daß durch eine Vielzahl von Startpunkten die Bemühungen parallelisiert werden und so der Aufwand für eine einzelne Megabase im Vergleich zum 'chromosomal walk' deutlich sinkt. Im Zebrafisch soll daher versucht werden, eine physikalische Karte durch Restriktionsverdau-Fingerprinting von PAC-Klonen sowie durch Sequenzierung von BAC-Enden einer weiteren BAC-Bank zu erstellen.

1.2.4. Sequenzanalyse

Organismen, die einer genetischen Analyse durch Kreuzungen und Mutagenese-'Screens' nicht zugänglich sind, da sie im Labor nicht gezüchtet werden können, die aber aufgrund ihrer phylogentischen Stellung (*Amphioxus*) oder aufgrund ihres kompakten Genomes (*Fugu*) analysiert werden sollen, können durch Sequenzanalyse, sowohl von genomischen Klonen, als auch von cDNAs analysiert und verglichen werden.

Die Herausforderung bei der Analyse von cDNAs ist es, stark exprimierte Gene, die somit in cDNA-Banken stark repräsentiert sind, zu identifizieren, um sie nicht mehrfach zu sequenzieren. Die Methode der Wahl ist hierzu das Oligo-Fingerprinting von ganzen cDNA-Banken. Dazu wird eine ganze Anzahl kurzer Oligonukleotide nacheinander auf eine geordnete cDNA-Bank hybridisiert. Anschließend werden die Klone identifiziert, die ein gleiches Hybridisierungsmuster haben und in sogenannten Clustern zusammengefaßt (Meier-Ewert et al., 1993). Diese Methode wurde bisher sowohl auf einige Zebrafisch-cDNA-Banken (Clark, et al. in Vorbereitung) als auch auf zwei *Amphioxus*-cDNA-Banken (Panopoulou et al. in Vorbereitung)

angewendet. Datenbanken solcher ESTs können für die Analyse von genomischen Sequenzen wichtige Informationsquellen sein.

Eine der Schwierigkeiten bei der Analyse genomischer Sequenzen ist die Identifizierung von Genen. Es stehen aber inzwischen einige computergestützte Algorithmen zur Verfügung, die offene Leseraster (ORFs) in Bereichen identifizieren, die von konservierten Splice-Stellen flankiert werden (Solovyev, 1994; Uberbacher, 1991). Bereiche mit offenem Leseraster können dann direkt translatiert werden und anschließend Datenbanken mit BLAST (Altschul, 1990) nach Sequenzhomologien durchsucht werden. Der Beweis, daß solche putativen Gene tatsächlich exprimiert werden, kann dann entweder durch die direkte Isolierung aus cDNA-Populationen durch RT-PCR oder durch die Isolierung eines entsprechenden cDNA-Klons geführt werden.

Mit dem Fortschritt der zur Verfügung stehenden Sequenzierungsdaten wird immer häufiger der Fall eintreten, daß man ein Gen identifiziert und dessen Bedeutung durch einen 'knock out' analysieren will. Dieses Vorgehen wird auch als 'reverse genetics' bezeichnet. Für das Ausschalten von Genen stehen einem eine Reihe verschiedener Möglichkeiten zu Verfügung. Die klassische ist ein 'targeted knock out', in dem die genomische Sequenz des Gens gezielt manipuliert wird, und dann das Konstrukt in die Keimbahn des betreffenden Organismus gebracht wird. Weitere Möglichkeiten, eine Genfunktion *in vivo* zu untersuchen sind 'chemical knock outs', bei der Antikörper gegen das entsprechende Genprodukt injiziert werden, bzw. durch antisense RNA-Injektion die Translation des entsprechenden Gens verhindert wird.

Für die *in-vitro* Analyse von Genfunktionen stehen heute neben den klassischen, biochemischen Methoden auch das 'yeast two hybrid' (Fields and Song, 1989) bzw. 'yeast three hybrid' (Drees, 1999) System zur Verfügung. Außerdem wächst in den Datenbanken stetig die Zahl der Peptid-Motive, denen ein Funktion zugeordnet werden kann, was bedeutet, daß in der Zukunft reine 'in silico' Experimente möglich werden.