

4. Diskussion

4.1. Wahl der Organismen

Fische stellen die größte Gruppe der Wirbeltiere dar, sie kommen in allen Gewässern der Erde vor und sind an alle möglichen Umweltbedingungen angepaßt. Es sind weltweit mehr als 24.000 Fischarten bekannt und am Fuße der Vertebratenevolution spielen sie eine bedeutende Rolle. Es ist daher durchaus erstrebenswert mehrere Fischarten in die vergleichende, molekulargenetische und entwicklungsbiologische Forschung einzubeziehen. Die hier vorgestellten Arten - Zebrafisch und Medakafisch - haben beide ihre Stärken und Schwächen. Der größte Vorteil des Zebrafischs ist es sicherlich, daß weltweit viele Arbeitsgruppen mit seiner Erforschung befaßt sind und so bereits einige aufwendige Projekte durchgeführt werden konnten. Demgegenüber bearbeitet nur eine kleine Anzahl von Forschern den Medakafisch. Dessen wichtigster Vorteil gegenüber dem Zebrafisch ist, wie diese Arbeit zeigt, die Existenz von gut getrennten Inzuchtstämmen. Zusammen mit dem kleineren Genom wird dadurch die positionelle Klonierung von Mutationen erheblich vereinfacht. Beide Fische gehören zu den Teleostei (echte Knochenfische), sind aber innerhalb dieser Abteilung weit getrennt (Abb. 45) und sollten sich daher aufgrund ihrer entfernten Verwandtschaft sehr gut ergänzen.

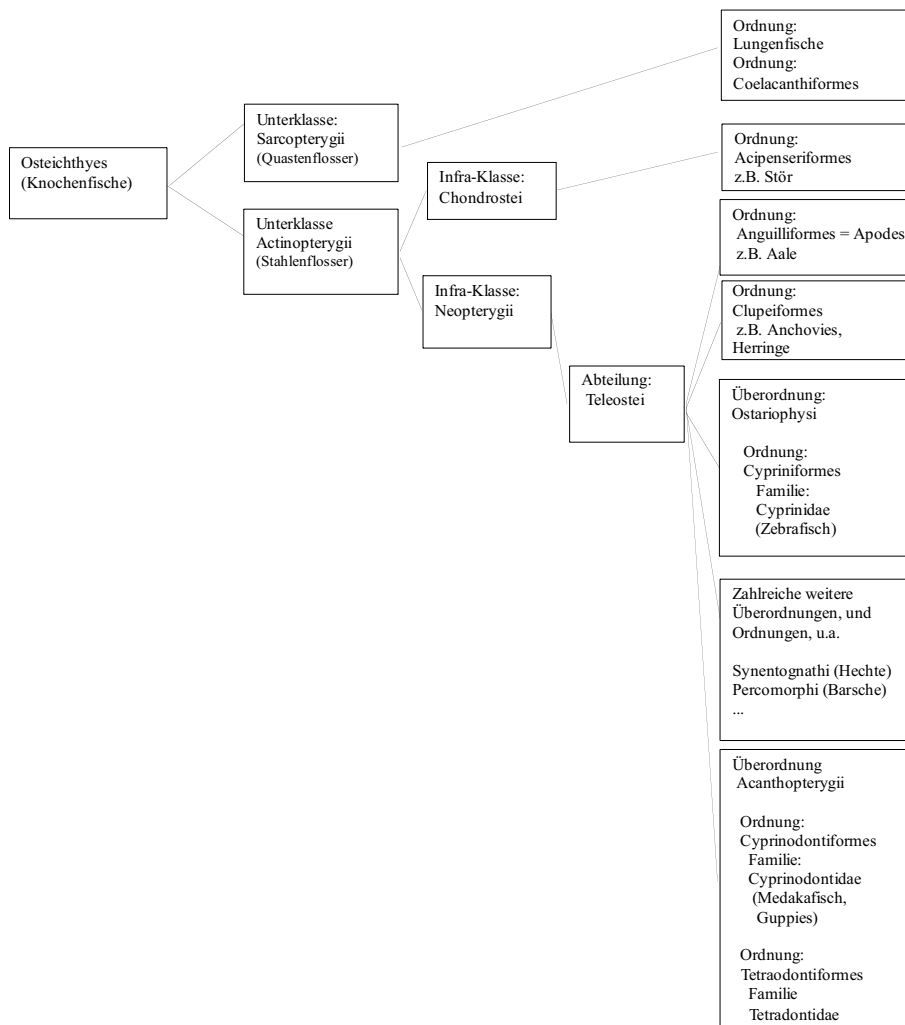


Abb. 45: Verwandtschaftliche Beziehung der Knochenfische nach Nelson (Nelson, 1994).

4.2. Genetische Kartierung im Zebrafisch

4.2.1. Wahl der 'Stämme' für das Kartierungs-Panel

Durch RAPD-PCR Vergleich war ein recht hoher Polymorphie-Grad zwischen Zebrafischen verschiedener 'Stämme' festgestellt worden. Dieser lag zwischen den analysierten AB-/Wik- und AB-/Tue-Fischen etwa gleich hoch bei 56-58% und damit deutlich höher als zwischen den Tue-/Wik-Fischen, bei denen nur 38% der Banden spezifisch für einen der beiden Fische waren. Da sich im Laufe der Arbeit aber herausstellte, daß das, was unter den Zebrafisch-Forschern als 'Stamm' bezeichnet wird, deutlich verschieden von der unter Genetikern sonst üblichen Bezeichnung, nämlich einer reinen Linie ist, war die vergleichende RAPD-Analyse von drei Fischen zwar schnell und einfach, nicht aber repräsentativ. Im weiteren Verlauf der Arbeit wurde eine hohe Variabilität innerhalb der einzelnen 'Stämme' festgestellt; damit sind die Fische auch an vielen Loci heterozygot. Selbst die durch 'early pressure' gewonnenen SJD- und C32-Linien wiesen noch eine deutliche Heterogenität innerhalb der Linien auf, wodurch das Kartieren von 'inter-mermaid'-Markern, die bei Hybridisierung einen +/- -Polymorphismus zeigen, erheblich erschwert wurde. Die innerhalb dieser Arbeit verwendete Methode erlaubte es jedoch, eine große Anzahl von Tieren parallel zu untersuchen. Daher wurden alle zur Verfügung stehenden Kreuzungen in die Analysen mit einbezogen.

4.2.2. IRS-PCR-Strategie

Innerhalb dieser Arbeit wurde getestet, ob es möglich wäre, eine auf Hybridisierungen basierende Strategie für das genetische Kartieren im Zebrafisch anzuwenden. Die technisch robusteste Methode dafür ist es, im Genom eingestreute, repetitive Elemente auszunutzen, um die zwischen ihnen liegende, nicht repetitive DNA zu amplifizieren (IRS-PCR). Auf solchen repetitiven Elementen lastet in der Regel kein Selektionsdruck, und so können zwischen verschiedenen Stämmen Variationen auftreten, die darin resultieren, daß ein Fragment in einem Stamm amplifiziert werden kann, in einem anderen aber nicht. Hybridisierung eines einzelnen, klonierten IRS-PCR-Produktes auf IRS-PCR-Produkte verschiedener Stämme kann also in einem +/- -Polymorphismus resultieren.

Im Zebrafisch sind bisher sieben repetitive Elemente veröffentlicht, von denen die mermaid- (=Dana), oops- (=angel), TdrI- und B13-Elemente verstreut im Genom vorkommen (Tab. 1, Seite 20). Von diesen ist aufgrund seiner Kopienzahl lediglich das 'mermaid'-Element für eine IRS-PCR-Strategie geeignet (Abb. 20, Seite 53). Da über Anzahl der 'mermaid'-Elemente beim Zebrafisch äußerst widersprüchliche Angaben bestanden, wurde zu Beginn der Untersuchungen zunächst die Anzahl der 'mermaid'-Elemente im Zebrafischgenom bestimmt (Seite 54). Aus der Anzahl der Elemente im Genom kann man unter der Voraussetzung, daß sie annähernd zufallsmäßig über das Genom verteilt sind, die Anzahl der möglichen 'inter-mermaid'-PCR-Produkte abschätzen. Unter den verwendeten PCR-Bedingungen wurden Fragmente bis zu einer Größe von etwa 2 kb amplifiziert. Bei 53.000 'mermaid'-Elementen in einem Genom von $1,7 \cdot 10^9$ bp ist die Wahrscheinlichkeit, daß zwei Elemente dichter als 2 kb beieinander liegen:

$$p = 2000 \cdot 53000 / 1,7 \cdot 10^9 = 0,062$$

Bei 53.000 Elementen sollten somit 3.200 im Abstand von weniger als 2 kb, und 1600 im Abstand von weniger als 2 kb und in der richtigen Richtung zueinander liegen. Diese Abschätzung stimmt hervorragend mit den Ergebnissen aus den Rückhybridisierungs-Experimenten (Seite 56) und den Hybridisierungen auf die Cosmid-Filter (Seite 54) überein. Das 'Oligo-fingerprinting' legt zwar den Schluß nahe, daß bis zu 6.800 verschiedene Klone in den Marker-Banken enthalten sein können, dies ist aber sicherlich zum einen auf die begrenzte Anzahl an Hybridisierungsexperimenten mit lediglich 100 verschiedenen Oligonukleotiden zurückzuführen. Bei einer größeren Anzahl von Hybridisierungen könnten technische Schwierigkeiten, die bei dieser Methode auftreten und die Einfluß auf das Ergebnis haben können, wie Ausbeute der PCRs, Gleichmäßigkeit des 'Spotting' und der Hybridisierung besser gemittelt werden. Zum andern mußten bei dem 'Clustering' relativ strenge Kriterien angelegt werden, da durch Rückhybridisierung einzelner Klone auf die Markerbanken festgestellt wurde, daß bei weniger stringentem 'Clustering' Klone verschiedener Sequenz in einem Cluster zusammengefaßt wurden. Obwohl das 'Oligo-fingerprinting' weniger genau zu sein scheint als die Analyse durch eine Serie von Rückhybridisierungen, sprechen dennoch einige entscheidende Vorteile für diese Methode, um die große Anzahl von Klonen der Markerbanken zu analysieren: Zum einen ist das 'Oligo-fingerprinting' sehr gut parallelisierbar. Während bei einer Serie von Rückhybridisierungen die Hybridisierungen nacheinander durchgeführt und zwischen den einzelnen Hybridisierungen vollständig ausgewertet werden müßten, um die gewonnenen Informationen maximal ausnutzen zu können, können für die Analyse durch 'Oligo-fingerprinting' die Hybridisierungen parallel durchgeführt und anschließend ausgewertet werden. Außerdem sind für das 'Oligo-fingerprinting' nur ein Bruchteil der Hybridisierungen notwendig. Die Auswertung der Bilddateien ist gut automatisierbar, und in einem Labor, in dem solche Experimente routinemäßig durchgeführt werden, auch sehr effizient: Das gleiche Set von Oligos wird parallel zur Analyse von mehreren Marker- und cDNA-Banken verwendet.

Die Anzahl der IRS-PCR-Produkte kann theoretisch durch Kombination zweier Primer, die in entgegengesetzter Richtung an das Element binden, verdoppelt werden. Allerdings muß hierbei berücksichtigt werden, daß das 'mermaid'-Element aus verschiedenen konservierten Bereichen besteht (Abb. 10, Seite 21), wobei der Bereich zwischen Nukleotid 94-119, der im 5'-Bereich von Box2 liegt, am konserviertesten ist. In diesem Bereich binden auch die Primer MMA und MMB. Bei Verwendung des MMA-Primers werden daher, sofern vorhanden, an jedem Ende des IRS-PCR-Produktes 64 bp der konservierten Box1 amplifiziert. Bei Verwendung des in die entgegengesetzte Richtung weisenden Primers MMB(rev) stellt sich die Situation noch drastischer dar. Hier werden außer dem 3'-Bereich der Box2 noch ein konservierter (TG)_n-Bereich, Box3 und die seltener vorhandene Box4 amplifiziert. Dies führt zu einer erheblichen Erhöhung des Hintergrundes bei den nachfolgenden Hybridisierungen und in extremen Fällen zu einem starken, positiven Hybridisierungsergebnis bei allen untersuchten Tieren.

4.2.3. Identifikation polymorpher Klone in den Markerbanken

Für die Verwendbarkeit der 'inter-mermaid'-PCR-Produkte als Marker bei der genetischen Kartierung spielt außer ihrer Anzahl und Verteilung im Genom, auch der Anteil an polymorphen PCR-Produkten eine Rolle. Bei einer Rate von 5-10%, welches 80-160 zusätzlichen Markern entspricht, würde sich der Aufwand für die

Herstellung und Analyse der Markerbanken sowie der Kartierungsfilter lohnen. Da das 'Oligo-fingerprinting' und 'Clustering' erst gegen Ende dieser Arbeit durchgeführt werden konnten, wurden polymorphe Klone hauptsächlich durch Hybridisierung auf kleinen Southern-Blots mit IRS-PCR-Produkten verschiedener Zebrafisch-'Stämme' identifiziert (Seite 58). Bei der Analyse von 800 Klonen zeigten insgesamt 180 einen +/- - Polymorphismus. Dabei entfielen 85 auf die Stämme AB und Tue, zwischen Wik und Tue zeigten 67 der untersuchten Marker und zwischen AB und SJD 139 einen +/- -Polymorphismus. Allerdings waren später 66% der zuvor als polymorph identifizierten Marker auf den Kartierungs-Kreuzungen nicht informativ.

Bei knapp 1000 Hybridisierungen aus einer Markerbank, die eine Komplexität von etwa 1.700 verschiedenen Klonen aufweist, war es unvermeidlich, daß teilweise Klone mit gleicher Sequenz mehrfach untersucht wurden. Daher wurde die Analyse nur wenig über die ersten 800, zufällig ausgewählten Klone hinausgeführt. Um die Banken voll auszuschöpfen zu können, hätte das Ergebnis des 'Clusterings' abgewartet werden müssen, oder alle untersuchten Klone hätten auf die Markerbanken zurück hybridisiert werden müssen, was einen nicht unerheblichen Aufwand dargestellt hätte. In einem Pilotversuch wurde getestet, ob es möglich wäre, die Marker in Pools von 6 Klonen auf die Banken zurück zu hybridisieren, um so den nötigen Aufwand zu reduzieren: Selbst bei diesen kleinen Pools war es nicht immer möglich, die verwendeten Klone eindeutig zu identifizieren und die Signale vom Hintergrund zu trennen. Dies ist zum einen auf die Struktur der 'mermaid'-Elemente und der Tatsache, daß bei Verwendung des MMA-Primers jeweils etwa 70 bp an den Enden der IRS-PCR-Produkte konserviert sind, zurückzuführen. Wenn man außerdem berücksichtigt, daß die durchschnittliche Insertgröße bei etwa 500 bp liegt, wird offensichtlich, daß bei einem Pool von 6 Klonen, die Probe einen nicht unerheblichen 'repetitiven' Anteil hat. Ein weiterer Störfaktor ist das Vorhandensein von $(CA)_n$ -Wiederholungen, die in etwa 25% der sequenzierten 'inter-mermaid'-Produkte vorkommen. Die einzige Methode, die sinnvoll und effizient erscheint, um die Markerbanken vollständig auszunutzen, ist die konsequente Durchführung des 'Oligo-fingerprintings' und anschließendes 'Clustering'.

4.2.4. Kopplungsanalyse

Die 180 als polymorph identifizierten (Abb. 26, Seite 59) Klone, sowie weitere 50, durch das 'Clustering' als potentiell polymorph eingestuften Klone wurden auf Kartierungsfilter hybridisiert. Dabei stellte sich heraus, daß trotz dieser Vorselektion nur etwa ein Drittel der Hybridisierungen ein auswertbares Segregationsmuster in einer der Kreuzungen ergab. Dabei lag das aus haploiden Nachkommen eines SJDxC32-Weibchen gewonnene Panel mit etwa 80 auswertbaren Hybridisierungen an der Spitze. Die Auswertung der Hybridisierungen erfolgte nach manueller Eingabe der Segregationsmuster durch MapmanagerXPb09. Unter Einbeziehung aller 96 Nachkommen in die Analyse bleiben von etwa 80 Markern bei einer Irrtumswahrscheinlichkeit von $p = 0,0001$ lediglich acht Marker ungekoppelt und 39 Marker häuften sich auf einer Kopplungsgruppe. Bei der Auswertung der Autoradiogramme fiel bereits auf, daß einige Tiere mit fast allen Markern ein positives Hybridisierungsergebnis lieferten, während andere Tiere vorwiegend negativ mit allen verwendeten Markern waren. Durch diese Tiere wird die Karte künstlich zusammen gehalten. Es war daher offensichtlich, daß die Rohdaten bereinigt werden mußten. Hierzu wurden alle Tiere von der Analyse ausgeschlossen, die bei der Hybridisierung mit den 'inter-mermaid'-Markern vorwiegend ein positives bzw. negatives Ergebnis zeigten. Die genetische Karte wuchs dadurch von rund 1400 cM in 12 Kopplungsgruppen auf ca 2200 cM in 11 Kopplungsgruppen. Für die 33, selbst bei weniger stringenten Bedingungen

(Irrtumswahrscheinlichkeit von $p=0,001$) ungekoppelten Marker wurden dabei jeweils 50 cM berechnet. Die bestehende Häufung von Markern auf die erste Kopplungsgruppe sowie eine Reihe von 'Doppel-Crossing-Overs', die im Zebrafisch normalerweise nicht vorkommen, zeigt allerdings, daß die Bereinigung noch nicht ausreichend war und noch weitere Marker fälschlicher Weise dieser Gruppe zugeordnet wurden. Es ist allerdings die Frage zu stellen, warum überhaupt Tiere von der Auswertung ausgeschlossen werden müssen, denn bei haploiden Nachkommen, die aus einer Kreuzung von Tieren gynogenetisch etablierter 'Stämme' entstanden sind, sollte man bei jedem Marker ein ideales Verhältnis von 1:1 an positiven und negativen Phänotypen erwarten können. Es konnte in Erfahrung gebracht werden, daß dieses 'Panel' nicht aus einem einzigen Pärchen generiert wurde, sondern daß verschiedene Pärchen Nachkommen beigesteuert haben. Das, berücksichtigt man die relativ hohe Rate an Polymorphismen innerhalb der beiden 'Stämme', könnte für einen Teil der Diskrepanzen verantwortlich sein. Vorwiegend positive Nachkommen können auch durch Kreuzkontaminationen zwischen verschiedenen Tieren entstehen, die trotz größter Vorsicht nicht auszuschließen sind, wenn DNAs im Mikrotiterplattenformat verschickt werden. Vorwiegend negative Tiere hingegen könnten durch geringe DNA-Konzentrationen bei der IRS-PCR verursacht werden, was eine geringere Komplexität der Produkte zu Folge haben kann.

Bei der Kopplungsanalyse fallen vier Markerpärchen auf, die lediglich 2-3 cM auseinanderliegen (Abb. 28 und 29). In dem einen Fall, wo beide Klone von ihnen sequenziert wurden (WIK1M19, WIK7L12), waren sie nicht sequenzidentisch und auch in den Fällen, in denen beide durch das 'Oligo-fingerprinting' einem Cluster zugeordnet werden konnten (WIK1L19 und AB25M2 bzw. AB2E7 und IND18H4), waren sie in unterschiedlichen Clustern platziert. Hingegen in dem einen Fall, wo zwei Marker (IND18H4, WIK3G21), auf exakt die gleiche Position kartieren (Abb. 28, Gruppe3), konnten auch beide dem gleichen Cluster zugeordnet werden. Dies stellt eine wichtige interne Kontrolle für das ganze Projekt dar.

4.2.5. Integration von genetischer und physikalischer Karte

Der größte Vorteil einer auf Hybridisierung basierenden Kartierungsstrategie ist es, daß ohne erheblichen Mehraufwand physikalische Ankerpunkte für eine genetische Karte erzeugt werden können, sowie die genetischen Marker zytogenetisch lokalisiert werden können. Die Marker werden dazu gleichzeitig mit der Hybridisierung auf das Kartierungs-Panel auch auf Filter hybridisiert, die IRS-PCR-Produkt von dreidimensionalen PAC-Pools, Irradiationshybriden und somatischen Zellhybriden enthalten. Dabei fiel auf, daß Marker, die auf dem Kartierungspanel ein auswertbares Segregationsmuster zeigten, in der Regel nur wenige oder gar keine Hybridisierungssignale auf den 'physikalischen' Kartierungsressourcen lieferten, bzw. umgekehrt, Marker, die mit den PAC-Pools und Zellhybriden auswertbare Hybridisierungssignale ergaben, waren in allen getesteten Zebrafischstämmen positiv. Für eine abschließende Bewertung konnten bisher zu wenige Marker ausgewertet werden, zumal für diese Arbeit ausschließlich unvollständig charakterisierte Hybrid-Panel zur Verfügung standen, die keine Integration mit Kartierungsdaten anderer Laboratorien ermöglichten.

Da die meisten getesteten IRS-Marker auf allen Zebrafischen positiv sind, sollten in den Markerbanken mehr als 1000 verschiedene Marker enthalten sein, die zwar keine Intergration der physikalischen und der genetischen Karte ermöglichen, wohl aber eine Integration der physikalischen mit der 'radiationhybrid-map'. Dies ist um so bedeutender, als aufgrund der hohen Variabilität innerhalb der Zebrafisch-Stämme die meisten Arbeitsgruppen dazu übergegangen sind, Gene nicht mittels Kopplungsanalyse auf eine Referenzkreuzung

sondern durch PCR-Analyse auf ein Irradiationshybrid-Panel zu kartieren, zumal innerhalb der letzten Jahre auf beide Panel in einem großen, mehrere Arbeitsgruppen umfassenden Aufwand jeweils rund 1000 Marker positioniert wurden, die CA-'Repeats', die eine Verknüpfung mit der genetischen Karte ermöglichen, ESTs und klonierte Gene umfassen, und die in der zweiten Hälfte dieses Jahres veröffentlicht wurden (Geisler et al., 1999 und Hukriede et al., 1999). Auch 'inter-mermaid'-Marker, die sich nicht durch Hybridisierung auf das Panel kartieren lassen, können durch PCR kartiert werden. Für zwei Marker, AB25C1 und AB25O1, wurden Primer synthetisiert und das kommerziell erhältliche Irradiationhybridpanel 'gescreent'. Der Klon AB25C1, der auf dem unbereinigten, haploiden (C32xSJD)-Panel in Kopplungsgruppe 8 erscheint, konnte so 41,3 cM von der Spitze von Chromosom 17, und AB25O1, der in Gruppe 1 erscheint, zwischen 14,1 und 21,8 cM von der Spitze von Chromosom 5 kartiert werden (G. Otto, persönliche Mitteilung). Diese Art der Kartierung stellt allerdings durch Sequenzierung, Primer-Synthese, rund 100 PCR-Reaktionen, die auf einem Gel untersucht werden müssen, einen erheblich höheren Kosten- und Arbeitsaufwand im Vergleich zu einer Hybridisierung dar.

Eine Alternative zur Kartierung der mehr als 1000 verschiedenen Marker auf die somatischen Zellhybride bzw. Irradiationshybride, wäre die IRS-Produkte der jeweils rund 100 Zelllinien auf die Filter der Markerbanken zu hybridisieren, um so die Marker bereits einer Kopplungsgruppe zuordnen zu können. Die Sequenzanalyse von ca. 200 Klonen zeigt jedoch, daß etwa 15% der 'inter-mermaid'-PCR-Produkte eine (CA)_n-Wiederholung enthalten, die zusammen mit dem relativ großen Anteil an konservierter Sequenz (konservierte Box1) einen hohen Hintergrund verursacht. Durch Konkurrenz mit gescherter Zebrafisch-DNA konnte das Signal bei der Hybridisierung von 'inter-mermaid'-PCR-Produkt verschiedener somatischer Zellhybride auf PAC-Koloniefilter fast vollständig gelöscht werden. Hier bietet aber der Einsatz spezifischer Kompetitoren, wie (CA)_n-Oligonukleotide oder PCR-Produkt der konservierten Box1, noch Raum zur Optimierung.

Innerhalb dieser Arbeit konnte gezeigt werden, daß es im Prinzip möglich ist, durch Hybridisierung von 'inter-mermaid'-Markern genetische, physikalische und zytogenetische Karte miteinander zu verbinden. Die Hauptschwierigkeit bei der vollständigen Ausnutzung dieser Ressourcen dürfte auch hier, wie schon bei der genetischen Kartierung, in der Variabilität innerhalb der Stämme liegen, die die Grundlage für die verschiedenen Banken und Zelllinien bildeten.

4.2.6. Alternative Hybridisierungsstrategie - RDA

Die repräsentative Differenzanalyse (RDA) ist eine Methode, mittels derer Restriktionsfragementlängenpolymorphismen angereichert und in hybridisierbare Marker umgewandelt werden können. In der Maus konnten so von beinahe isogenen Stämmen DNA-Fragmente isoliert werden, die in die Nähe des *jcpc*-Lokus auf Chromosom 10 des Mausgenoms kartieren (Baldocchi et al., 1996). Im Zebrafisch sollten mit Hilfe dieser Methode Restriktionsfragmente angereichert werden, die spezifisch für einen Stamm sind und nahe bei einer in den 'Mutantenscreens' identifizierten Mutationen liegen. So sollte durch Hybridisierung dieser Marker auf eine Referenzkreuzung gleichzeitig eine genetische Karte entstehen und die Mutanten kartiert werden. Um diese relativ neue und aufwendige Methode zu testen, wurde die DNA von jeweils 3 Individuen des Tue- und des AB-Stammes gepoolt und die Tue-'Tester'-DNA wurde mit Lambda-DNA kontaminiert, die als Kontrolle für eine spezifische Anreicherung dienen sollte. Während die anfängliche Kontamination unterhalb des Detektionslimits lag, war sie bereits nach einer Runde subtraktiver, kinetischer Anreicherung deutlich sichtbar und nach der dritten Runde die dominierende Bande (Abb. 31, Seite 64). Die klonierten Differenzprodukte wiesen bei Hybridisierung auf Amplikons von Tue- bzw. AB-DNA keinen +/- -Polymorphismus auf, ca. 90% der

analysierten Klone enthielten das Lambda-Fragment. Daraus kann man schließen, daß Unterschiede zwischen zwei DNA-Population mittels dieser Methode effizient angereichert werden können, aber auch, daß der Unterschied zwischen den untersuchten AB- und Tue-'Stämmen' nicht oder kaum größer ist als die Variabilität innerhalb der Stämme. Diese Methode ist zwar aufwendig und technisch anspruchsvoll, ließe sich aber zur genetischen Kartierung von Mutationen, die in einen gut getrennten Stamm gekreuzt wurden, durchaus einsetzen. Besonders in einem System, wie dem Zebrafisch, über dessen Genom zu Beginn dieser Arbeit nur sehr wenig bekannt war, und in dem annähernd 2000 entwicklungs-genetisch relevante Mutationen erzeugt wurden, die 400-500 verschiedene Gene betreffen (Haffter et al., 1996), wäre so gleichzeitig mit der Analyse der Mutanten eine genetische Karte entstanden.

4.2.7. Abschließende Bewertung und Ausblick

Der Zebrafisch ist das erste Wirbeltier, in dem eine systematische Suche nach Genen unternommen wurde, die in der Embryonalentwicklung von Bedeutung sind. Aufgabe dieser Arbeit war es, an einer Infrastruktur zu arbeiten, in der die in einem 'Mutagenesecreeen' identifizierten, rund 500 Gene kartiert werden könnten. Da das genetische Kartieren durch Hybridisierung bei einer großen Anzahl von Tieren und Loci deutlich effizienter ist, als eine Strategie, bei der einzelne PCR-Produkte auf einem Gel analysiert werden müssen, wurden RDA und IRS-PCR auf ihre Tauglichkeit getestet. Beide Strategien leiden unter der hohen Rate an Polymorphismen innerhalb der 'Stämme' und dem damit vergleichsweise geringen Grad an Polymorphismen zwischen verschiedenen 'Stämmen'. Die RDA-Strategie kann unter diesen Bedingungen als undurchführbar angesehen werden, während die 'inter-mermaid'-PCR durchaus einige auswertbare Marker liefern konnte. Allerdings ist die 'inter-mermaid'-Strategie im Zebrafisch nicht annähernd so erfolgreich, wie die Kartierung durch inter-B1-Marker in der Maus. Dort wurde die Komplexität einer inter-B1-Markerbank auf etwa 30.000 verschiedene Klone geschätzt (McCarthy et al., 1995). Das sind in dem doppelt so großen Genom 17 mal mehr IRS-PCR-Produkte, als mit Zebrafisch-DNA erzielt wurden. In der Maus ist außerdem die Rate an +/- - Polymorphismen zwischen den Spezies *Mus musculus* und *Mus spretus*, die sich kreuzen lassen und von denen homozygote Stämme existieren, mit ca. 26% deutlich höher als beim Zebrafisch, bei dem weniger als 8% der Marker-Klone auf einer der Referenzkreuzungen ausgewertet werden können. Zwischen verschiedenen Stämmen der Labormaus ist mit 2% der etwa 7000 erhaltenen IRS-PCR-Marker ebenfalls eine recht geringe Rate an IRS-PCR-Produkten polymorph (M. Knoblauch, pers. Mitteilung). Aber auch die nicht polymorphen Marker können wichtiges Instrument bei der Konstruktion einer physikalischen Karte sein, die mittels polymorpher Marker mit der genetischen Karte verankert wird. Zur vollständigen Ausnutzung der 'inter-mermaid'-Markerbank sollte die Zuordnung der einzelnen Klone in Cluster durch weitere 'Oligo-fingerprints' verbessert werden, da durch Sequenzanalyse von Repräsentanten einiger Cluster festgestellt werden konnte, daß im Moment noch gleiche Klone verschiedenen Clustern zugeordnet werden, aber auch umgekehrt, sequenzverschiedene Klone in einem Cluster zusammengefaßt sind. Wichtigstes, längerfristiges Ziel sollte aber die Etablierung von homozygoten Stämmen sein. Eine große Anzahl der Schwierigkeiten, die bei der Kartierung auftraten, wären bei Verwendung von echten Stämmen nicht entstanden: Nur knapp 30% der auf Southern-blots vorselektionierten Marker waren auf einer der Referenzkreuzungen informativ; sowohl das haploide, als auch das durch 'heat shock' generierte (C32xSJD)-Panel sind nicht aus einem Elternpaar entstanden, was aufgrund der großen Variabilität innerhalb der Stämme bedeutet, daß Marker in der Regel nur auf einem der beiden Panel informativ sind. Ferner ist bei einigen

Markern kein klarer +/- -Polymorphismus zu erkennen, sondern die Phänotypen variieren auf den Autoradiogrammen von hellgrau zu dunkelgrau und sind schwierig auszuwerten.

Die IRS-PCR wird außer zur genetischen Kartierung auch zur Amplifikation von speziesspezifischer DNA, z.B. aus Irradiations-Hybriden oder YACs, verwendet. Mit inter-Alu-PCR kann so humane DNA aus Zelllinien, die aus einer Fusion mit Hamsterzellen entstanden sind, amplifiziert werden. (Ledbetter et al., 1990). Solche Amplikons können zur physikalischen und zytogenetischen Kartierung eingesetzt werden. Auch im Zebrafisch ist dies, wie andeutungsweise gezeigt werden konnte, möglich. Großer Vorteil hier ist die Tatsache, daß unter den verwendeten PCR-Bedingungen keine Amplifikation von Produkten aus Hamster- oder Maus-DNA stattfindet. Die technische Schwierigkeit bei der Ausnutzung der 'inter-mermaid'-PCR-Produkte für das physikalische Kartieren im Zebrafisch liegt in der Tatsache begründet, daß ein relativ großer Anteil konservierter Sequenz amplifiziert wird, der bei komplexen 'inter-mermaid'-Hybridisierungsproben für einen hohen Hintergrund verantwortlich ist. Hier kann durch Optimierung der Kompetitionsbedingungen möglicherweise Abhilfe geschaffen werden und so eine Ressource, die eine effiziente Verknüpfung von genetischer, physikalischer und zytogenetischer Karte erlaubt, ausgenutzt werden.

4.3. Genetische Kartierung im Medaka

Ein weiterer Kandidat für einen genomweiten, gesättigten Mutagenese-'Screen', auf entwicklungs-genetisch relevante Gene ist der Medakafisch. Daher wurde innerhalb dieser Arbeit untersucht, ob effiziente, auf Hybridisierung basierende Kartierungsstrategien, die eine rasche Verankerung einer neuen Referenzkreuzung mit der bestehenden genetischen Karte, sowie die Verknüpfung von genetischer und physikalischer Karte erlauben, auch im Medakafisch eingesetzt werden können.

4.3.1. IRS-PCR im Medaka

Das 800-1000 Mb große Medakagenom enthält nur rund 3.400 'mermaid'-Elemente, die bei Verwendung des MMA-Primers eine Amplifikation von lediglich 8-11 'inter-mermaid'-PCR-Produkten ermöglichen. Daher wurde für die Konstruktion der Marker-Banken eine Kombination von drei verschiedenen Primern, die jeweils an verschiedene repetitive Elemente binden, verwendet, die zusammen ein komplexes PCR-Produkt von etwa 170 verschiedenen Fragmenten ergeben. Die Analyse von 96 Hybridisierungen ergab, daß etwa ein Drittel der IRS-PCR-Produkte einen +/- - Polymorphismus zwischen Tieren verschiedener Stämme aufwiesen. Im Vergleich zum Zebrafisch, wo höchstens 17,4% der Klone im direkten Vergleich von zwei verschiedenen Stämmen polymorph waren (Seite 58), ist dies eine deutlich höhere Rate. Ein weiteres Kriterium für die Anwendbarkeit der IRS-PCR-Marker ist ihre Übertragbarkeit innerhalb und zwischen den verschiedenen Stämmen und Referenzkreuzungen. Im Vergleich zum Zebrafisch zeigt sich eine deutlich höhere Homogenität innerhalb der Stämme darin, daß immer beide untersuchten Vertreter eines Stammes positiv oder negativ mit einer Hybridisierungsprobe waren, und zum anderen darin, daß mehr als 50% der durch 'Southern-blot' vorselektionierten Marker auch auf der Referenzkreuzung ausgewertet werden konnten, während dies im Zebrafisch nur bei knapp einem Drittel der untersuchten Marker möglich war.

Eine Schwierigkeit beim Medakafisch stellte sich jedoch darin dar, daß unterschiedliche Stämme zur Erstellung der beiden Referenzkreuzungen benutzt wurden. Da aber sowohl der AA2- und Cab-Stamm der Süd-Population entstammten und die HNI- und Kaga-Tiere aus der Nord-Populationen kamen, ließen sich immerhin 5 der 10 gefundenen IRS-Marker gleichzeitig auf beiden Paneln kartieren. Die andere Schwierigkeit lag darin, daß selbst bei der Verwendung von drei verschiedenen Primern die Anzahl der möglichen IRS-PCR-Produkte lediglich bei etwa 170 lag. Da sich dadurch bereits innerhalb der ersten sechsundneunzig Klone ein hoher Anteil an Wiederholungen ergab, wurde die Analyse auf polymorphe Klone nicht darüber hinaus weitergeführt. Im Vergleich dazu war es in dem etwa doppelt so großen Zebrafisch-Genom möglich, ein etwa 10 mal komplexeres IRS-PCR-Produkt zu amplifizieren. Im Zebrafisch war die 'inter-mermaid'-PCR auf die Verwendung eines Primers beschränkt (Seite 81). Es wurde innerhalb der Arbeit nicht getestet, ob im Medakafisch durch den Einsatz weiterer Primer, die in die entgegengesetzte Richtung an die 'Repeat'-Elemente OLR1 und OLR2 binden, die Komplexität des IRS-PCR-Produktes weiter erhöht werden könnte. Theoretisch müßte es möglich sein, durch die Verwendung von zwei weiteren Primern die Anzahl der verschiedenen IRS-PCR-Produkte auf etwa 300 zu erhöhen. Unter der Voraussetzung, daß die Rate polymorpher IRS-PCR-Produkte dann weiterhin bei etwa 30% liegt, wäre so die Generierung von 50 -100 neuen Markern möglich, deren Segregationsmuster darüber hinaus schnell und effizient in einer großen Anzahl von Tieren ausgewertet werden könnte. Dazu müßten allerdings zuvor neue Markerbanken konstruiert und durch 'Oligo-fingerprinting' analysiert werden.

4.3.2. Anwendung einer modifizierten AFLP-Strategie im Medaka

Da die Anzahl der zu erwartenden IRS-Marker im Medakafisch auch nach Einsatz der Primer Olrep1 und Olrep2 lediglich bei 170 liegt, sollte eine weitere Methode, die unabhängig von repetitiven Elementen funktioniert und darüber hinaus eine direkte Verknüpfung von genetischer Karte und genomischen Klonen, ermöglicht, ausprobiert werden. Diese genomischen Klone könnten als Ausgangspunkte für eine physikalische Karte dienen. DNA des Cab- und des Kaga-Stammes wurde mit zwei verschiedenen Restriktionsendonuklasen vollständig verdaut. Durch Ligation von Adaptern wurde die Amplifikation mit einem komplementären Primer ermöglicht, und durch Verlängerung dieses Primers um zwei spezifische Nukleotide konnte die Komplexität der entstehenden Amplikons kontrolliert werden (Abb. 36, Seite 69). Im direkten Vergleich der Amplikons von Kaga- und Cab-DNA wurde bereits der hohe Grad an Polymorphismen zwischen den beiden Stämmen deutlich. Die Amplikons, die aus den verschiedenen Stämmen generiert wurden, dienten als Hybridisierungsprobe auf jeweils einen Cosmid-Filter. Die überwiegende Anzahl der dabei identifizierten Klone war entweder nur mit dem Kaga- oder dem Cab-Amplikon positiv. Von den 96 differentiell hybridisierenden Cosmiden, die aus den Glycerinsstocks der Genbank isoliert wurden, gelang es bei 21% eine Probe zu isolieren, deren Segregationsmuster auf der (Cab x Kaga)-Geschwisterkreuzung ausgewertet werden konnte.

Obwohl diese Methode - im Vergleich zur IRS-PCR - relativ viele Arbeitsschritte beinhaltet, lohnt sich der Aufwand, da die Ligationsreaktionen durch Verwendung von Primern mit unterschiedlicher Verlängerung zur Herstellung verschiedener Amplikons genutzt werden können. Dabei ist allerdings zu beachten, daß verschiedene Basenabfolgen unterschiedlich starke Mutationsraten haben; besonders anfällig für Mutationen ist die Basenabfolge CpG. In dieser Arbeit wurden beispielsweise die Restriktionsenzyme *Bam*HI und *Acc*65I verwendet, die beide als letztes doppelsträngiges Basenpaar ein C/G stehen lassen. Bei Verwendung eines Primers, der auf seiner 3'-Seite um ein -GN verlängert ist, wird so die kritische Sequenz CG analysiert. Daß dies

eine wichtige Rolle spielt, zeigt sich in der Tatsache, daß bei Verwendung eines Primers, der eine AG-Verlängerung trägt, Amplikons entstehen, die eine geringere Rate an differentiell hybridisierenden Cosmiden aufweisen. Mit verschiedenen Kombinationen von Restriktionsenzymen und Primern sollte dennoch die Identifikation von einigen hundert bis tausend neuen Markern möglich sein. Abgesehen von dem höheren Arbeitsaufwand, den diese Methode im Vergleich zur IRS-PCR erfordert, hat sie noch einen weiteren Nachteil. Es wird von jedem Tier der Referenzkreuzung genügend DNA guter Qualität benötigt, um verschiedene Restriktionsverdau durchzuführen und sie auf einem Gel zu kontrollieren; daher war es beispielsweise nicht möglich, das bereits charakterisierte Rückkreuzungspanel von 39 AA2 x (HNIxAA2)-Tieren in die Analyse einzubeziehen, da hiervon nur jeweils ca. 200 ng DNA zur Verfügung stand.

4.3.3. Abschließende Beurteilung und Ausblick

Der Medakafisch zeichnet sich durch ein relativ kleines Genom, von lediglich 800-1000 Mb aus, was etwa der doppelten Größe des Fugu- bzw. der halben Größe des Zebrafisch entspricht. Alle Methoden, die im Zebrafisch zur Analyse der Embryonalentwicklung zur Verfügung stehen können auch im Medakafisch angewendet werden: Mutagenese-Screens, entweder in der F3 nach Mutagenisierung mit ENU, Psoralen oder Röntgenstrahlung, bzw. bereits in der F2, wenn Gynogenese-Methoden ausgenutzt werden, sowie 'whole mount in-situ'-Hybridisierung mit cDNAs, Antikörperfärbung, Injektionsexperimente, die eine vorübergehende Miß- bzw. Überexpression oder das Stilllegen eines Gens z.B. durch 'anti-sense' RNA- Injektion bewirken, und Transplantationsexperimente. Es existieren im Medaka darüber hinaus gut getrennte Inzuchtstämme, die eine genetische Kartierung der Mutanten ermöglichen. Die Bedeutung von gut getrennten Inzuchtstämmen wurde im Vergleich der IRS-PCR-Strategie im Zebrafisch und Medaka deutlich: Im Medaka war nicht nur ein höherer Prozentsatz der untersuchten Klone polymorph, sondern es ließ sich auch das Segregationsmuster eines höheren Prozentsatzes der als polymorph identifizierten Klone in einer der Referenzkreuzungen auswerten. Noch drastischer stellt sich der Vergleich der modifizierten AFLP-Methode zwischen Medaka- und Zebrafisch dar. Im Medaka wurde ein sehr hoher Prozentsatz an differentiell hybridisierenden genomischen Klonen detektiert, während im Zebrafisch so keine neuen Marker generiert werden konnten (G. Otto, pers. Mitteilung).

Trotz der offensichtlichen Vorteile des Medakfischs für die Genomanalyse ist es bislang erst an 5 Punkten gelungen, die neue (CabxKaga)-Geschwisterkreuzung mit der alten Referenzkreuzung zu verankern. Auf der anderen Seite sind die beiden Stämme so polymorph zueinander, daß es möglich sein wird, eine vollständige Karte von der (CabxKaga)-Kreuzung unabhängig von einer bereits vorhandenen Karte zu erstellen. Kritisch zu beleuchten ist jedoch die Frage, ob eine genetische Karte sinnvoll ist, und auch in der Zukunft von Bedeutung sein wird:

Die Alternative zum Kartieren von interessanten Mutationen auf eine genetische Karte, gefolgt von einer physikalischen Karte für diesen Bereiche, ist der Versuch, global das gesamte Genom eines Organismus in *Conitgs* überlappender genomischer Klone zu ordnen auf die dann cDNAs plaziert werden. Diese stellen dann zum einen die Grundlage für Sequenzierungsprojekte, wie sie für einige Modellorganismen initiiert worden sind, dar zum anderen bieten sie aber auch die Möglichkeit Kandidatengene aus einer relativ großen Region zu isolieren, und gezielt auf Kopplung mit einem mutanten Phänotypen zu untersuchen. Hauptvorteil einer solchen genomweiten Strategie ist es, daß durch eine Vielzahl von Startpunkten die Bemühungen parallelisiert werden

und so der Aufwand für eine einzelne Megabase im Vergleich zum einem 'chromosomal walk' deutlich sinkt. Da im Zebrafisch eine recht große Anzahl von Mutanten kloniert werden sollte, wird derzeit versucht, eine physikalische Karte durch Restriktionsverdau-'Fingerprinting' von PAC-Klonen sowie durch Sequenzierung von BAC-Enden einer noch zu etablierenden BAC-Bank zu erstellen. Diese Methode ist sicherlich nicht die effizienteste: allein durch Hybridisierung von cDNAs (z.B. den 30.000, die in den verschiedenen Zebrafisch-EST-Projekten bisher identifiziert werden konnten), auf PAC-Clone sollte es möglich sein, den größten Anteil des Genoms in *Contigs* genomischer Klone abzudecken - besonders, wenn man auf genomische Banken mit längeren Inserts zurückgreifen könnte.

Im Medakafisch werden wahrscheinlich nie die Mittel und Kapazitäten zur Verfügung stehen, solch groß angelegten physikalischen Kartierungsprojekte, wie sie für den Zebrafisch bereits in Planung sind, durchzuführen - Aber auch hier wird es durch eine Kombination von neuen Mutagenisierungstechniken - es konnte in diesem Jahr bereits gezeigt werden, daß erfolgreich durch gleichzeitige Injektion des 'sleeping beauty' Transposons (Ivics, et al. 1997) und Transposase mutagenisiert werden kann (Henrich, 1999) - mit neuen molekulargenetischen Methoden wie RDA, möglich sein, die Mehrzahl der Gene, die in einem mutanten Phänotypen betroffen sind, in naher Zukunft zu klonieren, und dies unabhängig davon, ob eine vollständige genetische oder physikalische Karte zur Verfügung steht. Eine Mutagenese durch Insertion von fremder DNA, deren Sequenz bekannt ist, stellt dabei insofern einen Vorteil dar, als durch einfache 'Vektorette'-PCR physikalische Marker, die direkt neben der Insertionsstelle - und damit auch in unmittelbarer Nähe des betroffenen Gens - liegen, isoliert werden können. Eine weitere Möglichkeit ist es, die inserierende DNA dahingehend zu verändern, daß ein 'plasmid rescue' möglich wird. Diese theoretischen Überlegungen über den Nutzen einer genetischen und physikalischen Karte liegen darin begründet, daß der Medakafisch bislang erst für die Analyse einer kleinen Gruppe von Genen, nämlich solchen, die die Augenentwicklung betreffen, verwendet wird. Sollte sich aber hier das Interesse auf andere Organe und Mutationen ausweiten, und dafür sprechen nicht allein die im Vergleich zum Zebrafisch meist klareren Phänotypen, wird eine gute genetische Karte, sowie eine möglichst große Anzahl von robusten, effizient auszuwertenden Markern, zu denen diese Arbeit einen ersten Beitrag leisten konnte, von großem Nutzen sein.

4.4. Analyse eines Amphioxus Cosmids

4.4.1. Bedeutung des Amphioxusgenoms für die Analyse des Vertebratengenoms

Nach den heute gängigen Theorien ist das Genom komplexer Vertebraten durch zwei sukzessive Duplikationen eines einfachen Chordaten-Genoms entstanden (Holland et al., 1994). Demnach ist die Analyse eines Chordaten-Genoms, das keine Anzeichen einer Genomduplikationen aufweist, wie z.B. das des Amphioxus (Schmidtke et al., 1977), eine Möglichkeit, Einblicke in ein Genom zu erhalten, welches dem hypothetischen Ur-Vertebratengenom ähnlich ist. Innerhalb dieser Arbeit wurde ein Amphioxus-Cosmid mit einer Insert-Größe von ca. 40 kb, was knapp 0,1 ‰ des Genoms entspricht, detailliert untersucht. Dies ist das größte bisher veröffentlichte *Contig* dieses Organismus. Das Cosmid kann als annähernd zufällig ausgewählt betrachtet werden, da die Hybrisierungsprobe (Spermin-Synthase des Zebrafisch), die zum 'Screening' der Genbankfilter verwendet wurde, zwar genügend Homologie zu einem kleinen genomischen Abschnitt aufwies,

um ein positives Signal zu verursachen, nicht aber zur Identifikation eines Amphioxus Spermin-Synthase-Gens geführt hat.

4.4.2. Basenzusammensetzung des Cosmids

Das Cosmid weist einen G+C-Gehalt von 41,35 % auf und liegt damit innerhalb des für Eukaryonten üblichen 37-47% Bereichs (Bernardi, 1993; Vinogradov, 1994) und sehr dicht an dem von Schmidtke et al. durch Schmelzpunktanalyse bestimmten G+C-Anteil von 41,5% für eine nicht näher bestimmte Amphioxus-Art (Schmidtke et al., 1979). Die gesamten bisher veröffentlichten Sequenzen von *Branchiostoma floridae*, die hauptsächlich aus cDNAs oder partiellen cDNAs bestehen, weisen im Gegensatz hierzu einen G+C-Gehalt von beinahe 49% auf - und liegen damit deutlich über dem Durchschnitt des Gesamtgenoms. Es ist schon lange bekannt, daß in warmblütigen Wirbeltieren Gene einen etwa 10% höheren G+C-Gehalt aufweisen, als das Isochor, in dem sie gefunden werden, und außerdem Gene preferentiell in GC-reicheren Abschnitten liegen (Bernardi, et al., 1985). In wechselwarmen Wirbeltieren sind bisher keine groß angelegten, systematischen Vergleiche von genomischen G+C-Gehalt und dem G+C-Gehalt von Genen veröffentlicht; allerdings wird in der gleichen Veröffentlichung erwähnt, daß beispielsweise das *Xenopus* β -Globin-Gen im Vergleich zu dem von warmblütigen Organismen einen niedrigen GC-Gehalt aufweist, ebenso wie das Insulin-Gen im Schleimfisch (Myxini). Das etwa 15 kb große Mitochondrien-Genom von Amphioxus hat einen relativ niedrigen GC-Gehalt von 37,3%.

Die Basenabfolge CpG kommt innerhalb des Cosmids etwa halb so oft vor, wie man aufgrund seiner Basenzusammensetzung erwarten würde. Ausführliche Untersuchungen zur Verteilung der CpG-Sequenz existieren bisher hauptsächlich im Säugetiergenom. Dort ist das Auftreten dieser besonders mutationsanfälligen Basenabfolge um etwa den Faktor fünf reduziert. Sie tritt in der Regel gehäuft in sogenannten CpG- Inseln am 5' Ende von Haushaltsgenen auf, deren Transkription in der Regel durch Methylierung der Cytosin-Reste supprimiert wird. Durchschnittlich findet man im Säugetiergenom alle 100 kb eine CpG-Insel von etwa 1-2 kb Länge vor. Innerhalb der 40 kb Amphioxussequenz werden von Grail bereits sieben potentielle CpG-Inseln angezeigt, über deren funktionelle Bedeutung allerdings noch keine Aussage gemacht werden kann. Es gibt nur wenige Untersuchungen über das Methylierungsmuster von Genen außerhalb der Klasse der Säugetiere. Bei Amphioxus wurde bisher das Methylierungsmuster von 17 Genen untersucht (Tweedie et al., 1997). Sieben davon ergaben einen klaren Hinweis auf m⁵C-Methylierung, da bei Southernblot-Analyse von DNA, die mit *MspI* verdaut wurde, im Vergleich zu Verdauung mit dem methylierungs-sensitiven Isoschizomer *HpaII* ein unterschiedliches Hybridisierungsmuster auftrat. Auffällig ist hierbei, daß klassische Haushaltsgene, wie ribosomale Proteine P2 und P6, im Amphioxus methyliert sind, während andere Gene, bei denen man von einer gewebespezifischen Expression ausgehen kann, wie 'insuline-like Protein' oder 'alkaly myosin light chain', in der nicht methylierten Fraktion des Genoms enthalten sind. Bei dem mit Amphioxus nahe verwandten Tunikaten *Ciona* ergeben erste Analysen ebenfalls den Status: Ubiquitär exprimierte Gene sind methyliert, während z. B. Transposons nicht methyliert sind (Simmen et al., 1999). In beiden Genomen erscheinen die Methylierungsmuster stabil zu sein. Ob die Methylierung in Amphioxus eine Rolle bei der Transkriptionskontrolle spielt oder ob sie eine andere Funktion, wie zum Beispiel bei der Replikationskontrolle, hat, kann zum jetzigen Zeitpunkt noch nicht geklärt werden; genauso wenig, ob die methylierten Gene eventuell mit nicht-methylierten CpG-Inseln assoziiert sind.

4.4.3. Gengehalt des Amphioxuscosmids

Innerhalb der 40 kb genomischer Amphioxussequenz konnten sicher zwei Gene nachgewiesen werden, deren Expression durch RT-PCR bzw. durch Vergleich mit der internen EST-Datenbank (Panopoulou, 1999) bewiesen werden konnte. Eines davon ist eine Aldoketo-Reduktase, die zwischen Nukelotid 19107 und 23169 lokalisiert ist und 9 Exons umfaßt. Des weiteren wurden von allen drei Exonvorhersageprogrammen zwischen den Nukleotiden 25476-23893 drei nahe beieinander liegende Exons identifiziert, die alle innerhalb des gleichen Amphioxus ESTs (GeneBank Zugriffsnummer AI391437) liegen, und die signifikante Ähnlichkeit zu einem 'low density' Lipoprotein Rezeptor verwandten Protein haben. Außerdem wurde von allen drei verwendeten Exonvorhersageprogrammen ein weiteres, potentiell Exon vorhergesagt, welches eine signifikante Ähnlichkeit zu $\alpha(1a, 1b, 2a, 2b)$ adrenergen Rezeptoren bzw. zu Serotonin Rezeptoren aufweist. Es ist zwischen den Nukleotiden 2878 und 3276 des reversen Stranges lokalisiert. Die weiteren, in den Bereichen von Nukleotid 10450-11850 und von 13500-14000 gefundenen Homologien zu drei verschiedenen Klonen der Amphioxus EST-Datenbank (GenBank Zugriffsnummer.: AI391435, AI391436 und AI391434) (Abb. E32) wurden von keinem der Exonvorhersageprogramme als Exons identifiziert, zeigen keine Homologie zu irgendeinem bekannten Protein und weisen teilweise auch keine längeren offenen Leseraster auf. Außerdem scheint der EST-Klon 026K17BFLG RRset39 ein größeres 'Intron' zu besitzen, als die entsprechende genomische Sequenz. Bei den EST-Klonen MPIMGBFLG_32D13 und 023B11_BFGL RRset39 fällt im Vergleich mit der genomischen Sequenz auf, daß sie an einigen Stellen mehrere Extra-Nukleotide aufweisen, was nur teilweise mit der Sequenzierungsgenauigkeit der ESTs, die jeweils nur mit einer Sequenzreaktion und einem 'read' erhalten wurden, erklärt werden kann. Eine mögliche Erklärung wäre, daß es sich bei diesen ESTs um genomische Verunreinigungen der cDNA-Bank handelt, die trotz poly-A⁺-Selektion der RNA kloniert wurden. Alternativ könnte es sich hierbei auch um tatsächlich transkribierte Abschnitte handeln, die nicht unbedingt Teil eines Gens sein müssen, die aber aufgrund von längeren (A)_n-Abschnitten, die alle diese ESTs enthalten, ebenfalls selektioniert wurden. Es könnte sich dabei aber auch um untranslatierte Bereiche von echten Genen handeln, die jedoch mit den verwendeten Exonvorhersageprogrammen nicht identifiziert werden konnten, da die Vorhersage von terminalen Exons im allgemeinen ungenauer ist, als die von internen, dagegen spricht allerdings, daß in der Nähe keine weiteren, internen Exons, die ebenfalls Homologie zu einem Teil des ESTs aufweisen, gefunden wurden. Auch der Klon MPIMGBFLG498_45P2, der mit vorhergesagten Exons übereinstimmt und Homologie zu einem Datenbankeintrag zeigt, scheint Tabelle 5 (Ergebnisse, Seite 74) nach eine Unstimmigkeit aufzuweisen: Der Bereich von Nukleotid 43-155 hat zwei homologe genomische Abschnitte. Dabei ist allerdings der zweite von Nukleotid 43-152 nicht gut konserviert, und die Konsensus-Splice-Stellen 5'-GT und AG-3' befinden sich im entsprechenden genomischen Abschnitt genau im Übergang von Nukleotid 151 zu 152 des ESTs.

Innerhalb dieses Cosmids wurden demnach sicher zwei Gene nachgewiesen, und ein weiteres Exon kann mit hoher Wahrscheinlichkeit angenommen werden. Dies ergibt eine durchschnittliche Gendichte von einem Gen pro 17-20 kb. Innerhalb des 580 Mb großen Amphioxusgenom werden etwa 20.000 Gene vermutet. Diese Zahl ergibt sich hauptsächlich daraus, daß das Genom höherer Vertebraten etwa 80.000-100.000 Gene beinhaltet, und der Genomduplikationstheorie nach im Amphioxus etwa ein Viertel davon vorhanden sein müssen. Außerdem unterstützen einige Untersuchungen von Genfamilien wie den Hox-Genen (Holland and Garcia-Fernandez, 1996) diese Hypothese weitgehend. Und zum anderen durch die ausführliche Analyse einer

Amphioxus cDNA-Bibliothek, die 54.500 Klone umfaßt, durch 'Oligo-fingerprinting', wodurch bereits 11.000 verschiedene ESTs nachgewiesen werden konnten. Bei Analyse einer größeren Anzahl an cDNAs kann man erwarten, daß sich die Zahl auch auf mindestens 20.000 verschiedene Klone erhöhen wird. Bei gleichmäßiger Verteilung der Gene über das gesamte Genom sollte man durchschnittlich alle 29 kb ein Gen erwarten. Damit liegt die Gendichte innerhalb des analysierten Cosmids etwa 50% über dem erwarteten Durchschnittswert. Obwohl man von dieser hier untersuchten, relativ kleinen, statistisch nicht signifikanten Sequenzlänge nicht generalisieren kann und die Gene in den bisher untersuchten Eukaryonten-Genomen auch nicht gleichmäßig verteilt sind, ergibt die Analyse dieses Cosmids noch weitere Hinweise auf eine möglicherweise höhere Gendichte: Zum einen könnten die auffällig übereinstimmenden Exonvorhersagen aller drei Programme im Bereich von 11-19 kb des reversen Strangs, die allerdings weder Homologie zu bereits bekannten Genen aufwiesen, noch innerhalb der vorhergesagten Grenzen mit einem bereits identifizierten EST übereinstimmten, tatsächlich Teile weiterer Gene sein, die die Gendichte in dem analysierten Cosmid weiter erhöhen würden. Zum anderen scheint auch die genomische Größe von Genen im Vergleich zum Menschen deutlich reduziert. Zum Beispiel umfaßt die humane 3α -Hydroxysteroid-Dehydrogenase eine genomische Region von etwa 20 kb, während das hier beschriebene homologe Gen im Amphioxus eine genomische Region von lediglich 4 kb überspannt. Eine durchschnittliche Reduktion von inter- und intragenischen genomischen Abschnitten um etwa den Faktor 8-10 kann man beispielsweise im Fugu, dessen Genom eine vergleichbare Größe wie das des Amphioxus hat, beobachten. Extrapoliert man von dem hier gefundenen Faktor 5, dann könnte das Amphioxus-Genom etwa halb so viele Gene wie das Vertebraten-Genom beinhalten, doppelt so viele, wie derzeit erwartet. Auf alle Fälle wäre eine statistisch relevante Erhebung, ähnlich wie sie bei *Ciona intestinalis* durchgeführt wurde (Simmen et al., 1998), auch im Amphioxus-Genom erstrebenswert.

4.4.4. Repetitive Elemente innerhalb des Amphioxuscsmids MPMGc117B0533

Innerhalb des Cosmides wurden verschiedene repetitive Elemente in unmittelbarer Nachbarschaft zu offensichtlichen 'single copy' Sequenzen, wie Genen gefunden. Neben den beiden auffälligen repetitiven Elementen von 97 bp Länge werden als weitere repetitive Elemente einfachere, kurze, direkte Wiederholungen wie $(A)_n$, $(GAAA)_n$, $(TA)_n$, $(CA)_n$, und $(AAAT)_n$, sowie eine kurze invertierte Wiederholung gefunden. Dies deckt sich mit der Beschreibung von Schmidtke et al. (1979), der bei vergleichender Hydroxylapatit-Chromatographie von 250 bp und 1200 bp langen DNA-Fragmenten beobachtete, daß selbst bei sehr niedrigen Cot-Werten ($C_0t < 10^{-3}$) die Bindung der längeren DNA-Fragmente an Hydroxylapatit deutlich, um den Faktor 4,3 erhöht war. Daraus zog er den Schluß, daß hochrepetitive oder palindromische Sequenzen im gesamten Genom verteilt vorkommen und so in den längeren DNA-Fragmenten häufiger bereits nach kurzer Renaturierungszeit Abschnitte doppelsträngiger DNA, die an Hydroxylapatit binden können, entstehen (Schmidtke et al., 1979).

4.4.5. Kombination von EST-Projekt und genomischer Sequenzanalyse

In dieser Arbeit wurde ein Amphioxus-Cosmid von 40,8 kb Länge detailliert analysiert. Dazu wurden drei unterschiedliche Computerprogramme (Grail, GenScan und Mzef) benutzt, um die Sequenz auf mögliche Exons

zu untersuchen. Potentielle Exons wurden durch Blast mit öffentlichen Datenbanken, wie GeneBank und SwissProt verglichen. Bei der Analyse konnte aber auch auf ein internes EST-Projekt zurückgegriffen werden (Panopoulou, 1999), bei dem eine 27.600 Klone umfassende cDNA-Bank durch 'Oligo-fingerprinting'-Analyse in rund 11.000 verschiedene Cluster unterschiedlicher Sequenzen unterteilt wurde. Diese Arbeit zeigt, daß die Kombination von EST-Projekt und umfassender genomischer Sequenzanalyse eine effiziente Möglichkeit sein kann, um zum einen die Expression potentieller Gene *in silico* zu bestätigen, die in großen Genomprojekten noch weiter an Bedeutung gewinnen kann. Zum anderen wurde aber auch gezeigt, daß durch vorsichtige, vergleichende Analyse der Sequenzen die verwendete cDNA-Bank charakterisiert werden kann. Hier kann beispielsweise eine genomische Kontamination der verwendeten cDNA-Bank nicht ganz ausgeschlossen werden. Auf alle Fälle wird deutlich, daß in gewissem Umfang Exonvorhersagen mit den vorhandenen Programmen auch in genomischer DNA des Amphioxus möglich sind und daß das analysierte Cosmid, nach allem, was bisher über das Amphioxus-Genom bekannt ist, repräsentativ zu sein scheint. Eine weitere, statistisch relevante, Analyse des Amphioxus-Genoms, wozu diese Arbeit einen ersten Beitrag leistet, wäre sicherlich sinnvoll, um weitere Einblicke in die Vertebraten-Evolution zu erhalten.