

# Bioinformatisches Praktikum

## Evolution von microRNAs

Christian Otto

3. September 2008

### Zusammenfassung

Im Rahmen des bioinformatischen Praktikums im Sommersemester 2008 haben wir uns mit der Evolution von MicroRNAs (miRNA) beschäftigt. Diese noncoding RNA Moleküle bilden eine große Gruppe von kleinen regulatorischen RNAs in Eukaryoten. Basierend auf der miRBase Version 11.0 wurden sämtliche von Hertel et al. identifizierte Cluster in Hinblick auf ihre phylogenetische Entwicklung analysiert [2]. Dabei wurde die Vorgehensweise von Tanzer et al. angewendet [6]. Anhand der gewonnenen phylogenetischen Bäume sowie der genomischen Positionen wurde versucht, daraus die evolutionäre Geschichte zu rekonstruieren.

## 1 Einleitung

MicroRNAs sind Gegenstand relativ junger Forschung und bilden eine große eigenständige Gruppe von noncoding RNA Molekülen in Eukaryoten. Sie bestehen aus einer Precursor Sequenz mit etwa 100 nt Länge, die in einen Stemloop faltet. In diesem Stem befindet sich ein etwa 20 nt langer und hochkonservierter Bereich, der als Mature RNA oder 'reife' miRNA bezeichnet wird. Teilweise existiert im zur Mature Sequenz gegenüberliegendem Stem eine so genannte Mature Star Sequenz. Verschiedene miRNA Familien bilden Cluster, d.h. sie sind in kurzen Abständen im Genom lokalisiert.

Nach Faltung des primären Transkriptes eines Clusters in die typische Sekundärstruktur wird dieses durch ein RNA bindendes Protein erkannt und durch die zellkernspezifische Ribonuklease III, Drosha, zu pre-miRNAs

verkürzt. Diese werden durch den Transporter Exportin-5 ins Cytoplasma transportiert. Dort wird durch das Enzym Dicer der Bereich der reifen miRNA als doppelsträngiges Molekül heraus geschnitten. Im Folgenden wird die Bindung der mature (oder evtl. mature star) Sequenz an den RNA induced silencing complex (RISC) durch Dicer vermittelt. Danach wird der komplementäre RNA-Strang abgebaut. RISC erkennt zur reifen miRNA komplementäre mRNA Ziele und spaltet diese spezifisch. Dadurch kann die Translation dieser mRNAs nicht erfolgreich stattfinden. Insgesamt zählt dieser Mechanismus als 'posttranskriptionelles Gensilencing'.

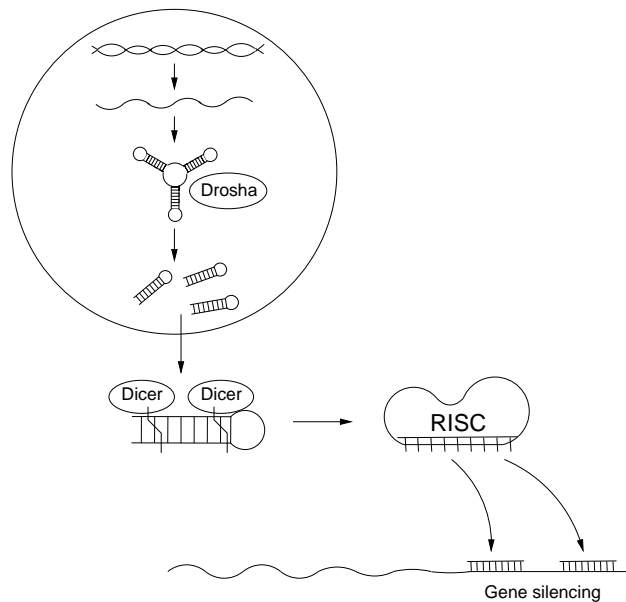


Abbildung 1: Funktionsweise des posttranskriptionellen Gensilencing mit miRNAs

## 2 Durchführung

Im Vorfeld wurden bereits alle zu diesem Zeitpunkt annotierten miRNAs gegen eine umfangreiche Auswahl von Genomen geblastet. Der Blastoutput bildete die Grundlage für das weitere Arbeiten. Anhand dessen wurden chromosomenweise akkumulierte Fenster aus den entsprechenden Genomen heraus geschnitten. Zusätzlich wurde das Fenster um 10 nt an beiden Seiten

erweitert, da die miRBase Sequenzen möglicherweise zu kurz annotiert worden sind. Für jede Familie wurden die gewonnenen Sequenzen zusammen mit den annotierten miRNAs mittels `ClustalW` aligniert. Das Alignment wurde im `Emacs` im `rlee-mode` "geputzt". Dabei wurden Duplikate sowie nicht faltende und nur außerhalb der Mature passende Sequenzen entfernt. Unter Umständen wurde das Alignment noch per Hand verbessert, indem die Sequenzen bzw. Sequenzteile an besser passenden Stellen im Alignment verschoben wurden. Zum Schluß wurde das Alignment an beiden Seiten abgeschnitten, sodass nur die eigentlichen Precursor das Alignment bilden. Aus diesen Alignments wurde mittels `SplitsTree` phylogentische Bäume erstellt. Mit ihrer Hilfe konnte eine vorläufige Einteilung der Familien in Gruppen vorgenommen werden. Außerdem wurden die zurecht geschnittenen Sequenzen nochmals zurück gegen die Genome geblastet um die genauen Genompositionen herauszufinden. Auf Grundlage der Genompositionen und der phylogentischen Bäume aller Familien eines Clusters konnte die evolutionäre Geschichte des Clusters rekonstruiert werden. Basierend auf der Theorie der zwei Genomduplikationen bei Vertebraten und der zusätzlichen Duplikation bei Teleost Fischen konnten die Stammbäume interpretiert werden [3] [5].

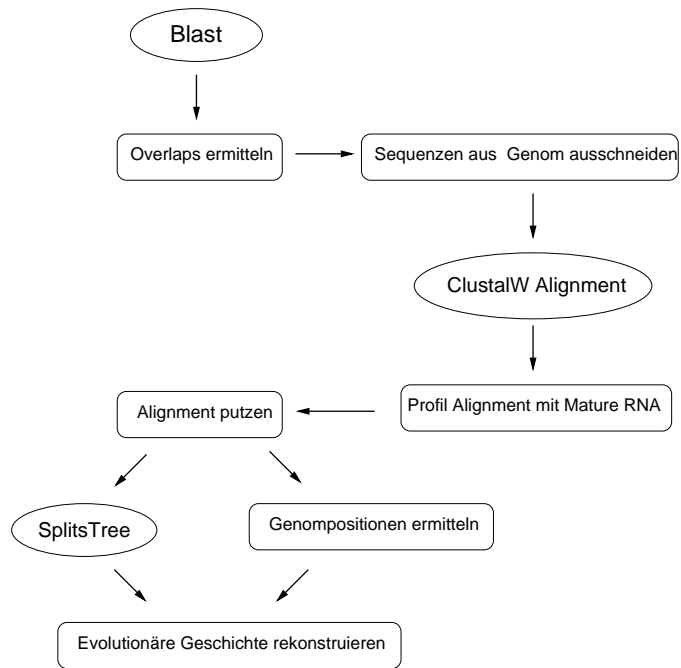


Abbildung 2: Übersicht über Arbeitsablauf

## 3 Ergebnisse

### 3.1 mir-1 Cluster

#### 3.1.1 Übersicht über das Cluster

Das mir-1 Cluster besteht aus 2 Familien mir-1 und mir-133. Man findet das Cluster in allen *Metazoa*. In den *Protostomia*, den *Echinodermata* und den *Urochordata* findet man eine Kopie. Später kam es zu Duplikationen, sodass man in den *Teleostei* bis zu 5 Kopien findet und in den *Arthropoda* bis zu 4 Kopien. Aufgrund von Problemen mit den Genomversionen des *Canis familiaris* ist nicht genau bestimmbar wieviele Kopien der *Canis familiaris* hat. Wahrscheinlich ist, dass er 3 Kopien wie der *Homo sapiens* hat. Besonders an der mir-1 ist, dass die Mature eine palindromische Sequenz hat. So findet man die mir-1 oft auf beiden Strängen im Genom an der selben Position. In der *Mus musculus* ist eine mir-1 Sequenz auf dem antisense, die mmu-mir-1-2-as, annotiert. Jedoch wurde sie experimentell nur in sehr viel geringerer Konzentration gefunden als die mmu-mir-1-2. Diese liegt an der selben genomischen Position jedoch auf dem anderen Strang. Aus diesem Grund vermuten wir, dass es sich bei der experimentell gefundenen mmu-mir-1-2-as um ein Abbauprodukt des Hairpin handelt. Deshalb und weil in keinem anderen Organismus eine mir-1 auf dem antisense Strang annotiert war, wurde sie im Alignment nicht betrachtet.

#### 3.1.2 Evolutionäre Geschichte

Zu Beginn gab es einen Cluster, der aus der mir-1a und der mir-133a bestand. In den *Protostomia* blieb dieser Cluster auch so erhalten. Die *Nematode* verloren aber die mir-133a. Ebenso verloren *B. mori*, *A. gambiae*, *H. robusta* und *A. californica* die mir-133a. *A. californica* duplizierte das Cluster noch einmal. Auch in den *Echinodermata* und *Urochordata* ist das ursprüngliche Cluster noch zu finden. Danach fanden 2 Duplikationen, die zu 4 Kopien des Clusters führten. *P. marinus* verlor alle der Kopien bis auf mir-133a-Ib und mir-133b-IIB. In den *Teleostei* wurden die mir-1a-IA und die mir-133a-IA nochmals dupliziert. *D. rerio* verlor den soeben duplizierten Cluster wieder. Die anderen *Teleostei* wurden die mir-1b-IIB und die mir-133b-IIB. Während *O. latipes* die mir-1a-IA verlor, verloren die anderen die zweite Kopie der mir-133a-IA. *T. rubripes* verlor zusätzlich noch die mir-1a-IB. *X. tropicalis* und *G. gallus* verloren keine der Kopien, die bei den Genomduplikationen in den

Vertebraten entstanden sind. In den *Mammalia* ging die mir-1b-IIB und die mir-133b-IIB verloren. Zusätzlich verloren *P. troglodytes* und *R. norvegicus* die mir-1b-IIA und die mir-133b-IIA. *B. taurus* verlor die mir-133a-IB.

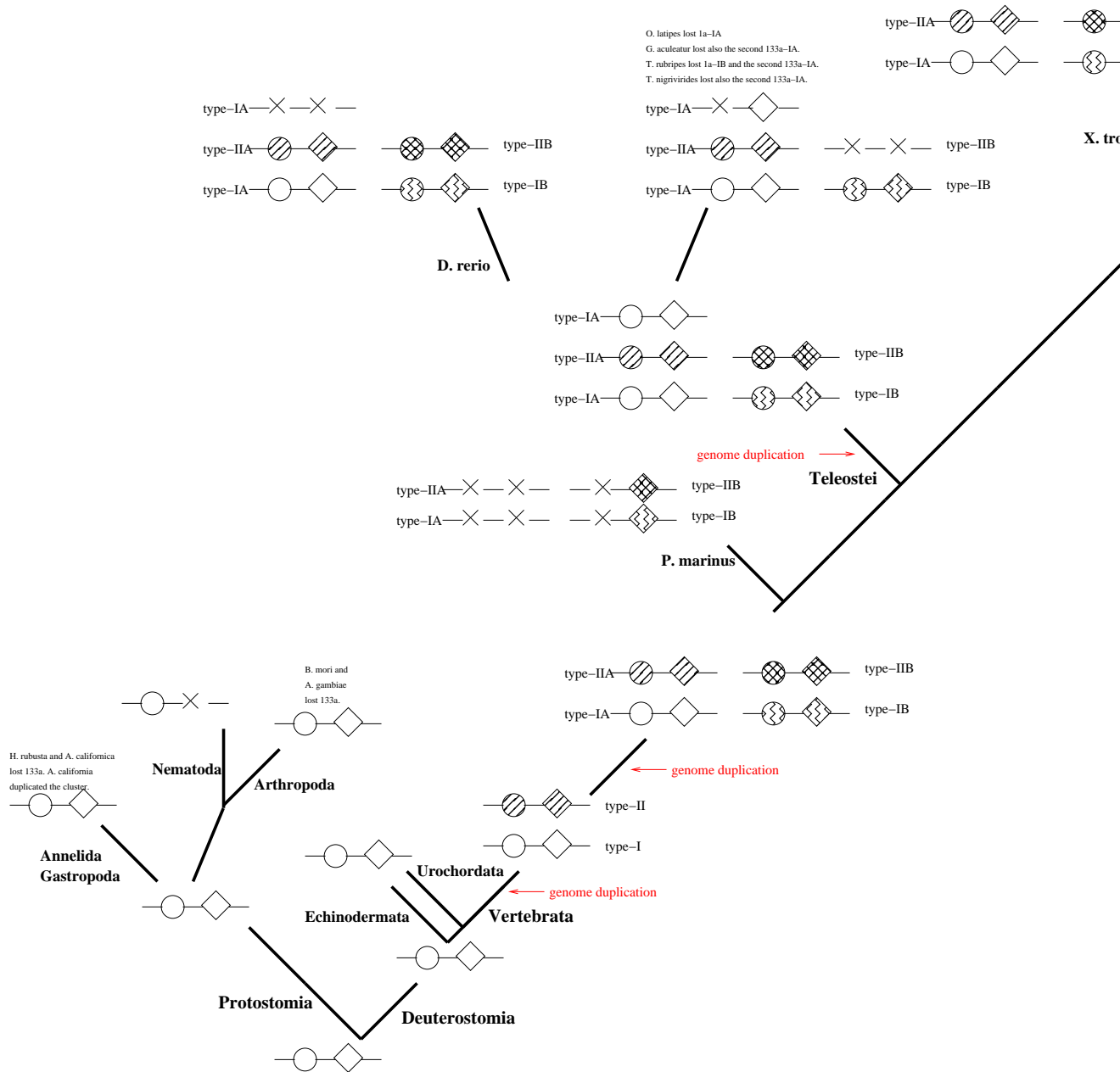


Abbildung 3: Übersicht über die vermutliche evolutionäre Entwicklung des mir-1 Clusters

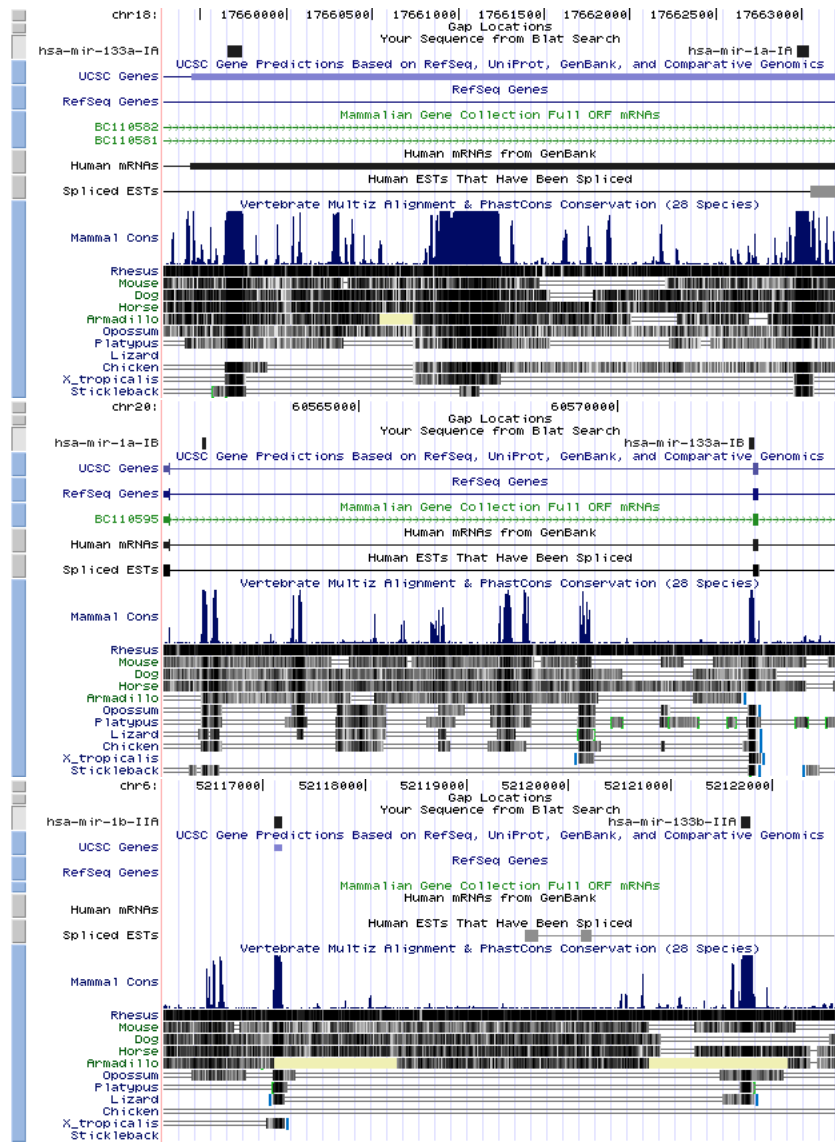


Abbildung 4: Konservierung des mir-1 Clusters in *Homo sapiens*



## 3.2 mir-3 Cluster

### 3.2.1 Übersicht über das Cluster

Das mir-3 Cluster besteht aus 5 Familien mir-3, mir-4, mir-5, mir-6 und mir-286. Das Cluster wurde in *Drosophila* einmal und in *Anopheles gambiae* zweimal gefunden. Während in *Drosophila* alle Familien des Clusters vorhanden sind, sind in *Anopheles gambiae* nur mir-3 und mir-286 vorhanden. Die Familie mir-6 gibt es in 3 Kopien, die auf dem Chromosom hinter einander liegen. Die Familie mir-3 ist in 2 Kopien vorhanden, jedoch hat *Anopheles gambiae* eine Kopie verloren.

### 3.2.2 Evolutionäre Geschichte

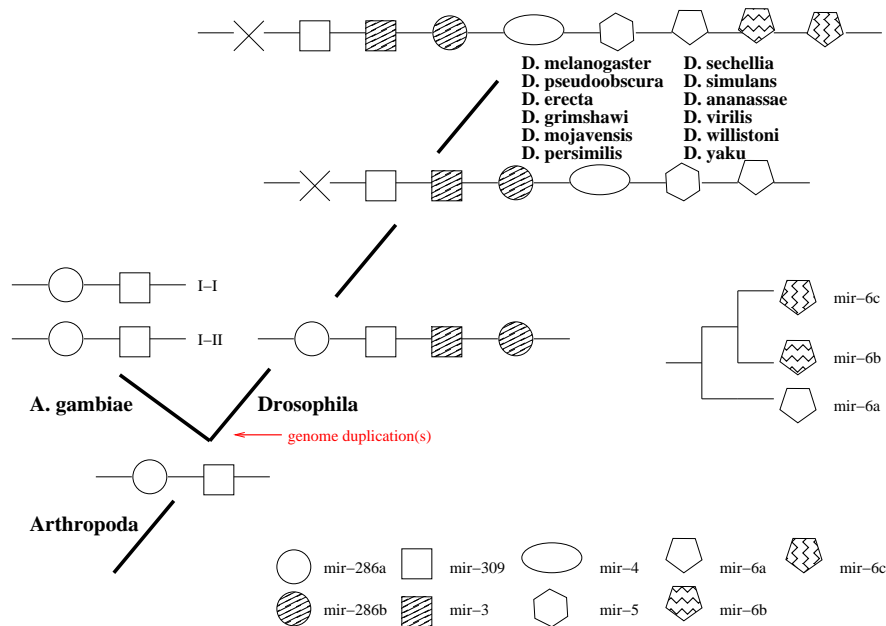


Abbildung 5: Übersicht über die vermutliche evolutionäre Entwicklung des mir-3 Clusters

In den *Arthropoden* war zunächst ein Cluster vorhanden, welcher wahrscheinlich aus einer mir-286a und einer mir-309 (mir-3 Familie) bestand. Dieser Cluster duplizierte sich in *Anopheles gambiae*, sodass dort nun ein Typ I Cluster und ein Typ II Cluster auf 2 unterschiedlichen Chromosomen zu

finden sind. In *Drosophila* kam es zu einer inversen Duplikation. Der Cluster besteht nach dieser aus mir-286a, mir-309, mir-3 und mir-286b in dieser Reihenfolge. Danach entstanden mir-4, mir-5 und mir-6a. Danach duplizierte sich die mir-6a zur mir-6b und schließlich diemir-6b zur mir-6c.

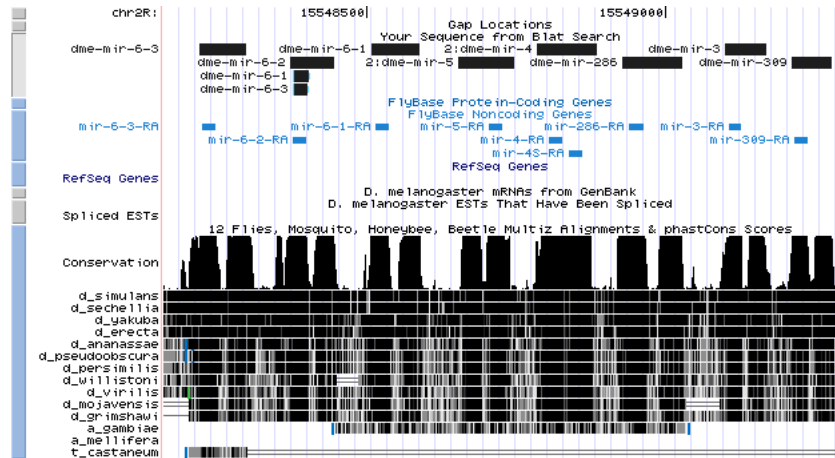


Abbildung 6: Konservierung des mir-3 Clusters in *Drosophila melanogaster*

### 3.3 mir-8 Cluster

#### 3.3.1 Übersicht über das Cluster

Das mir-8 Cluster besteht laut Namensgebung aus vier miRNA-Familien: mir-8, mir-141, mir-200 und mir-429. Allerdings wurde entdeckt dass alle Familien von einander abstammen. Dabei fungiert die mir-8 als ursprüngliche microRNA und alle anderen sind im Laufe der Evolution daraus entstanden. Man findet in allen Vertretern der Deuterostomia Spuren dieser Familien und auch in den Protostomia kommen sie, wenn auch deutlich weniger, vor. Der vorerst fertige, Deuterostomia spezifische, Cluster entstand schon vor den beiden Genomduplikationen, wobei diese keinen Einfluss auf den Aufbau hatten. Nach der Entstehung der Mammalia kam es dann zu einem genomweiten Umbau und zu großen Neuordnungen im Genom sodass eine unvollständige 2. Kopie entstand.

#### 3.3.2 Evolutionäre Geschichte

Die Rekonstruktion der Evolutionsgeschichte gestaltete sich bei diesem Cluster etwas schwierig, da in relativ vielen Arten der basalen Deuterostomia zunächst keine bzw. nur sehr wenige miRNAs dieses Clusters gefunden wurden. Zahlreiche miRNAs wurden per Hand im Trace Archive der NCBI gelistet, wobei diese Genome noch nicht vollständig assembliert sind und es somit noch sein kann, dass die dargestellten Cluster noch nicht der Wirklichkeit entsprechen. Dennoch konnte man anhand des phylogenetischen Baumes des Clusters sehr gut die Verwandtschaftsbeziehungen der einzelnen microRNAs nachvollziehen. So ist deutlich zu erkennen dass die ancestrale RNA eine mir-8 gewesen sein muss, die noch heute in den Protostomia in dieser Form vorkommt. Ausgehend von dieser 'Urform' kam es nach Abspaltung der Hemichordaten zu einer ersten Clusterduplikation an deren Ende ein Cluster mit der mir-429 und mir-200a stand. Dementsprechend hat man in Organismen wie *S.kowalevskii* und *S.purpuratus* nur eine miRNA gefunden, diese hatte sich allerdings schon ein wenig von der Urform entfernt. Die 2.Clusterduplikation, bei der sich die mir-429 duplizierte und die mir-200b entstand, muss vor der Abspaltung der Urchordaten geschehen sein. Denn im *Oikopleura dioica* und in der Gattung *Ciona* tauchen 3 verschiedene microRNAs auf.

Durch die lange individuelle Entwicklung dieser und anderer früher Organismen ist es nicht verwunderlich dass die entdeckten Sequenzen vorallem im Be-

reich des Loops und der Mature-star-Region deutliche genomische Veränderungen aufweisen. Des Weiteren gab es sogar in der Mature Region kleine Mutationen, ohne aber an der Familie der RNA zweifeln zu können. Um trotz dieser teils großen Differenzen eine zuverlässige Aussage über die Art der jeweiligen miRNA treffen zu können wurde noch ein Alignment der Maturesequenzen angefertigt, da diese Bereiche ja einer starken Konservierung unterlegen sind. Anhand dieses Baumes ließ sich vor allem die frühe Evolutionsgeschichte sehr gut rekonstruieren. So erkennt man aber auch deutlich, dass die in der miRBase annotierten csa-mir-200a und cin-mir-200a in Wirklichkeit der mir-200b zugeordnet werden muss. Die Sequenzen der frühen Deuterostomia ließen sich alle einer bestimmten mir-Gruppe zuordnen, so dass man mit relativ hoher Wahrscheinlichkeit davon ausgehen kann dass der Cluster den Grundaufbau, wie nach der 2.Clusterduplikation beschrieben, hatte. Selbst die beiden in der Evolution folgenden Genomduplikationen änderten am Cluster nichts, sodass davon ausgegangen werden muss, dass jeweils die zusätzliche Kopie gleich wieder verloren ging. Dies wird durch die gefundenen Sequenzen im Neunauge und im Hai bestätigt, die jeweils keine Anzeichen eines Duplikats erkennen lassen. Die Teleost Fische durchliefen noch eine spezifische Genomduplikation, sodass es nicht verwundert zumindest im *Danio rerio* eine komplette Kopie zu finden. Alle anderen Teleostei verloren demnach dieses Duplikat wieder. Die wirklich interessante Entwicklung geschah aber erst nach der Trennung der Theria von den Protheria und den Saurpsida, denn in allen Gattungen der Theria taucht der Cluster mit einer unvollständigen Kopie auf einem spermaten Chromosomen auf. Die ist ein erstes Indiz dafür dass diese Kopie ein Ergebnis eines Mammaliaspezifischen genomweiten Rearrangements ist. D.h zu dieser Zeit fanden große Umlagerungen, Verschiebungen und Duplikationen im gesamten Genom statt, sodass durchaus solch eine Entwicklung zu erklären ist. Beim *Homo sapiens* liegen zumal beide Kopien in einer genomischen sehr aktiven Region. Der ursprüngliche Cluster auf dem Chromosomen 1 liegt sehr nah an der Telomerregion, sodass es durchaus sein kann dass sich Teile davon auch an andere Positionen oder Chromosomen kopierten. Die 2.Kopie befindet sich auf Chromosomen 12, auch nah am Telomer und in der Nähe gibt es viele Bereiche die vermutlich von anderen Chromosomen kopiert wurden, was eventuell die Ursache für das Entstehen dieses Duplikats ist. Es bleibt noch zu erwähnen dass in allen Primaten, bis auf den Menschen, die mir-200b verloren gegangen ist. Die letzten Veränderungen des Clusters, nach Abspaltung der Protheria, bedürfen einer genaueren Untersuchung, da über solche Ereignisse noch wenig

bekannt ist.



## 3.4 mir-23 Cluster

### 3.4.1 Übersicht über das Cluster

Das mir-23 Cluster besteht aus den 3 Familien mir-23, mir-24 und mir-27, welche scheinbar nicht miteinander verwandt sind. Das Cluster wurde nur in Vertebraten aufgefunden und ist dort fast ausschließlich in mindestens zwei Kopien vorhanden. In Teleost Fischen sind dabei bis zu 5 Kopien vorhanden, wobei eine Clusterkopie nur in *Danio rerio* zu finden ist. In Mammalia sind fast in allen Organismen jeweils zwei Kopien vorhanden. Die fehlenden Cluster bzw. Familien sind nur bei experimentell gefundenen miRNAs vorgekommen. Außerdem wurde bei *Gallus gallus* nur ein Cluster gefunden. Dies könnte zum einen auf einen evolutionären Verlust oder zum anderen auf unzureichende genomische Information hindeuten.

### 3.4.2 Evolutionäre Geschichte

Am Anfang der Entwicklung der Vertebraten ist ein mir-23 Cluster *de novo* entstanden, welches sich höchstwahrscheinlich vor *Petromyzon marinus* während einer Genomduplikation verdoppelt hat. Alternativ könnte *Petromyzon marinus* jedoch auch eine eigenständige Genomduplikation gehabt haben. Allerdings sind die in dessen Genom gefundenen mir-23 Cluster ähnlich zu den Clustern Typ a und Typ b, auch wenn bei dem zu Typ b ähnlichen Cluster die Familie mir-24 nicht gefunden werden konnte und somit möglicherweise verloren gegangen ist. Vor der Abzweigung der Knorpelfische gab es dann eine weitere Genomduplikation, welche zu den vier verschiedenen Cluster Typ a, b, c und e führte, wobei Typ c die Kopie von Typ a und Typ e die Kopie von Typ b anhand des Verwandtschaftsbaum zu sein scheint. Bei *Callorhinchus mili* ist sowohl das komplette Typ c-Cluster als auch die Familien mir-24 sowie mir-27 des Typs e-Clusters verschwunden. Jedoch ist hier auch möglich, dass die gefundene Typ e mir-23 nur eine nicht lokale Duplikation der Typ b mir-23 Sequenz darstellt. Bei den Teleost Fischen kam es zu einer zusätzlichen Genomduplikation, wobei die Verdopplungen des Typ a sowie des Typ c in keinem verfügbaren Knochenfischgenom identifiziert werden konnten. Die Kopie des Clusters Typ b ist nur im *Danio rerio* vorhanden und die Kopie des Typ e Clusters befindet sich upstream auf demselben Chromosom wie das Typ a Cluster jedoch mit einem Abstand von durchschnittlich 1 Mb und ist auch bei nicht allen Fischen noch komplett vorhanden. Das Auftreten des Typ b-II Clusters nur bei *Danio rerio*

ist durch die lange individuelle evolutionäre Entwicklung des Zebrabärbling nicht ungewöhnlich. Der komplette Verlust der Kopien von Typ a und Typ c sowie der partielle Verlust des Typs e-II Clusters deutet daraufhin, dass bei diesen Kopien keine Funktionsaufteilung erfolgte und somit die Kopien nach einer Zeit wieder verschwanden. Bei *Xenopus tropicalis* ist sowohl das Cluster Typ e als auch die Familie mir-24 des Clusters Typ c nicht mehr vorhanden. Aufgrund der Tatsache, dass sowohl das Cluster Typ c als auch Typ e nach dem Krallenfrosch nicht mehr aufgefunden werden konnte, handelt es sich mit großer Wahrscheinlichkeit um Clusterverluste auf dem kompletten evolutionären Ast. Ob das Cluster Typ e vor oder nach dem Krallenfrosch verloren ging oder ob nur *Xenopus tropicalis* dieses Cluster nicht mehr besitzt, kann nicht mit Sicherheit beantwortet werden. Mit zusätzlichen genomischen Daten von Amphibien könnten in Zukunft dazu noch verlässlichere Aussagen getroffen werden. Mit Ausnahme von *Gallus gallus* konnte in allen gut annotierten Amniota das Cluster Typ a und b komplett lokalisiert werden. Vertreter bei denen das Cluster zunächst nur experimentell bestimmt werden konnte, wurden von der Betrachtung hier ausgeschlossen, da dies auf unvollständige Genome hindeutet. Zusammenfassend wäre somit zu bemerken, dass für eine gesicherte evolutionäre Geschichte dieses Clusters die vorhandenen Genome nicht ausreichen.



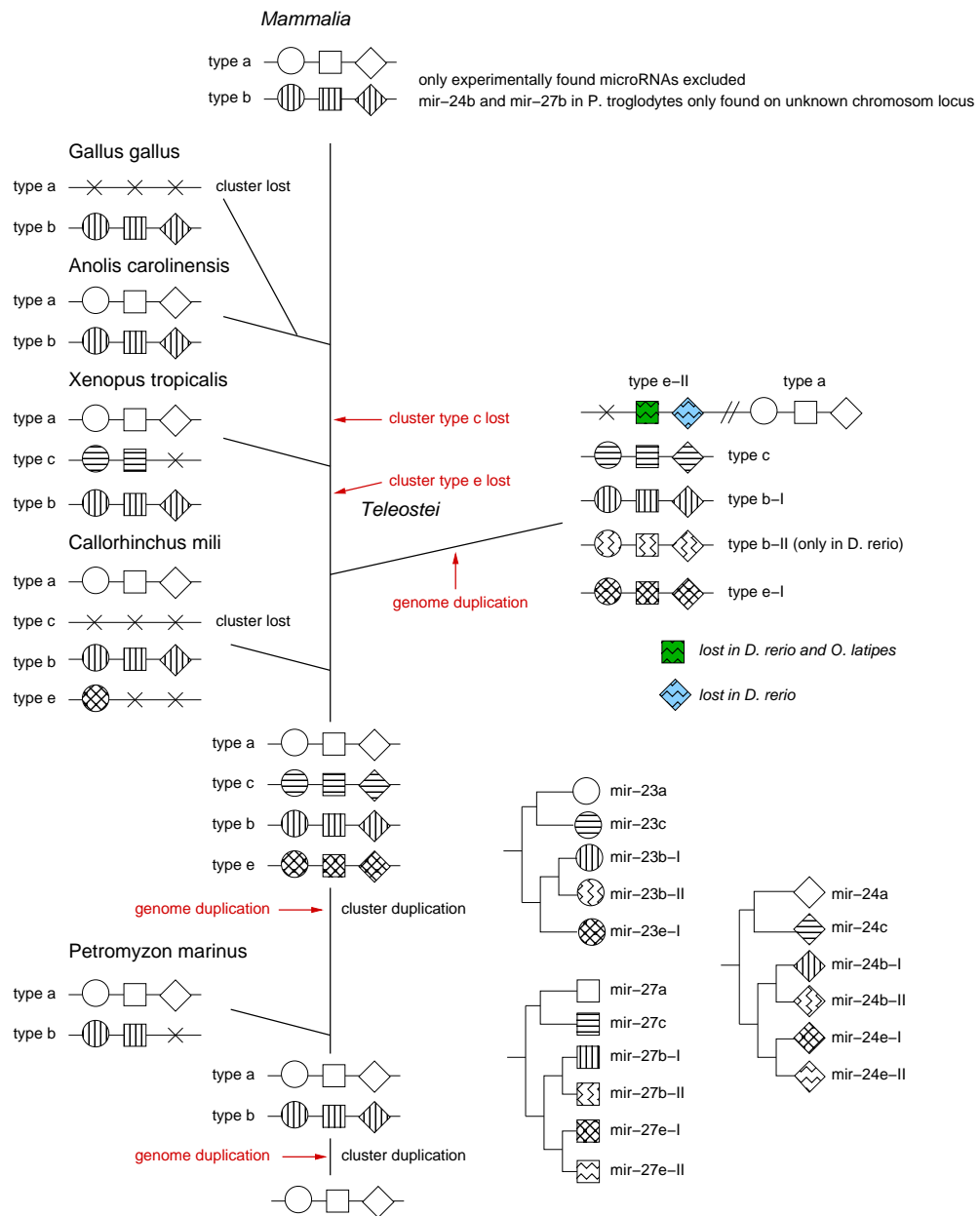


Abbildung 8: Übersicht über die vermutliche evolutionäre Entwicklung des mir-23 Clusters

### 3.5 mir-58 Cluster

Der mir-58 Cluster besteht aus den beiden Familien mir-58 und mir-270. Die letztere ist als 'nonfamilied' gekennzeichnet, d.h. sie ist in noch keinen anderen Organismen gefunden worden. Die bereits annotierten Sequenzen wurden jeweils an ihrer assemblierten Position wiedergefunden. Des Weiteren ist es nicht gelungen die mir-270 in einem anderen Organismus als in *C.elegans* zu lokalisieren.

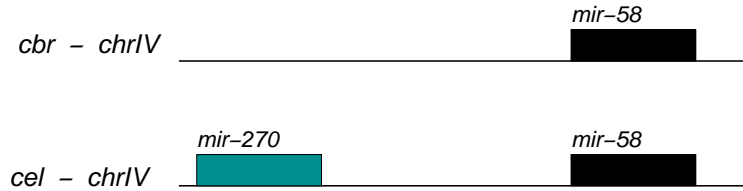


Abbildung 9: Aufbau des mir-58 Clusters

Eine Rekonstruktion der evolutionsspezifischen Geschichte ist nun relativ trivial, da die mir-58 nur in Vertretern der Gattung *Ceanorhabditis* vorkommen, d.h. sie sind wahrscheinlich nach der Abspaltung der Strongyloidea entstanden. Des Weiteren scheint es so, als sei die Entwicklung der mir-270 ein rein *C.elegans* spezifischer Vorgang gewesen, denn beim blast gegen die weiteren Genome dieser Gattung, entstand bezüglich der mir-270 kein weiterer Hit. Um den gesamten Clusteraufbau genauer abzuklären müsste man die Sequenzen noch gegen verschiedene andere Nematodengenome im Trace Archive blasten.

Jedoch ist die Wahrscheinlichkeit, dass es sich um hochspezifische microRNAs handelt hoch, sodass man davon ausgehen kann dass diese Familie *Ceanorhabditis*-spezifische Entwicklungsprozesse steuert. Allerdings wurde in der Tarbase und Wormbase kein Hinweis auf eine bisher bekannte Aufgabe gefunden.

### 3.6 mir-61 Cluster

Das mir-61 Cluster besteht aus den zwei Familien mir-61 und mir-250. Beide wurden nur in mehreren Genomen von *Caenorhabditis* lokalisiert. Die Cluster in *Caenorhabditis elegans* und *Caenorhabditis briggsae* sind bereits in der mirBase annotiert gewesen und wurden mit Hilfe von **Blast** auch dort wieder gefunden. Zusätzlich wurde das Cluster noch in *Caenorhabditis brenneri* (cbn), *Caenorhabditis japonica* (cja), *Caenorhabditis novel* (cno) sowie in *Caenorhabditis remanei* (cre) lokalisiert. Allerdings konnte in *Caenorhabditis japonica* nur die Familie mir-250 entdeckt werden. Diese Familie ist in *Caenorhabditis japonica* höchstwahrscheinlich verloren gegangen, da sie sich etwa 2000 nt vor dem Ende des Contigs befindet.

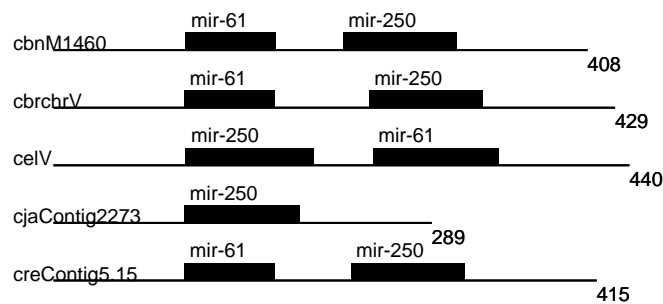


Abbildung 10: Aufbau des mir-61 Clusters

Die Rekonstruktion der evolutionären Geschichte ist in diesem Fall nicht eindeutig möglich, da sich die Organismen alle in der gleichen Gruppe befinden. Man kann nur die Aussage treffen, dass die Genome von *Caenorhabditis brenneri* und *Caenorhabditis novel* sehr nah verwandt sind. Die Reihenfolge der Familien innerhalb des Clusters ist nicht aussagekräftig, da die Annotierung des plus- bzw. des minus-Stranges willkürlich erfolgt, solange keine Referenzgene oder zirkuläre Erbmaterialien vorliegen. Eine Konservierung über diese Gruppe hinweg konnte nicht entdeckt werden. Es handelt sich somit um eine microRNA, welche höchstwahrscheinlich wurmspezifische Entwicklungsprozesse steuert und deshalb in keiner anderen Organismengruppe wiedergefunden werden konnte. Ein von der Tarbase genanntes Ziel ist z.B. das Gen *vav-1*, welches auf die Translationshemmung des dazugehörigen Proteins *vav-1* einwirkt. In dem Paper von Norman et al. wird das Protein *vav-1* als Rho/Rac-Family guanine nucleotide exchange factor funktionell eingeordnet [4]. Es soll dabei in *Caenorhabditis elegans* für die Regulation von

grundlegenden rhythmischen Verhaltensweisen verantwortlich sein.

## 3.7 mir-90 Cluster

### 3.7.1 Übersicht über das Cluster

Das mir-90 Cluster besteht aus drei miRNA-Familien: der mir-90, der mir-80 und der mir-238. Alle drei Familien treten ausschließlich in der Gattung *Caenorhabditis* auf und der Clusteraufbau ist weitgehend über alle Arten konserviert. Die vorher als 'nonfamilied' gekennzeichnete mir-238 konnte in allen fünf lokal vorliegenden Genomen wiedergefunden werden. Des Weiteren sind die einzelnen Familien untereinander nicht verwandt.

### 3.7.2 Evolutionäre Geschichte

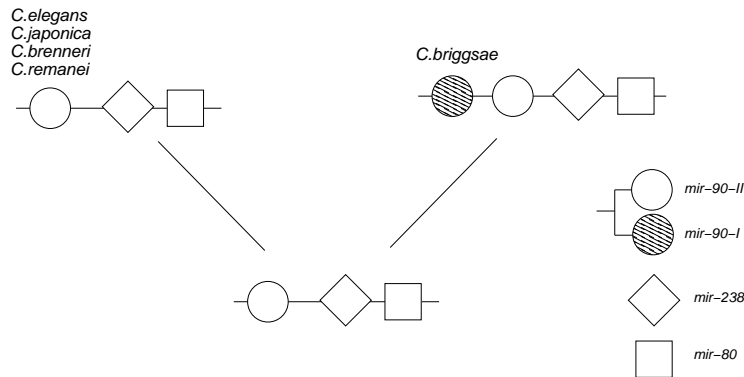


Abbildung 11: Übersicht über die vermutliche evolutionäre Entwicklung des mir-90 Clusters

Alle drei Familien sind *Caenorhabditis* spezifisch, da in den lokal vorhandenen Genomen und im Trace Archive kein Hinweis auf ein Vorkommen in Organismen wie *Haemonchus contortus* und *Ancylostoma canium* gefunden wurde. D.h. alle drei miRNAs sind nach der Abspaltung der Strongyloidea in etwa zur gleichen Zeit entstanden, wobei es nicht möglich ist zu differenzieren welche der Familien sich zuerst entwickelt hat. Im Folgenden blieb der Cluster in seiner Form erhalten, was darauf schließen lässt, dass alle miRNAs eine regulatorische Funktion ausüben und somit von größeren Mutationen verschont geblieben sind. Über die genaue Evolution des Clusters kann man

nur spekulieren, da er nur in einer einzigen Gattung vorkommt und fast allen Arten denselben Aufbau haben.

Auffällig ist lediglich die zusätzliche Kopie der mir-90 im *Caenorhabditis briggsae*, die sich vor den anderen Familien des Clusters angelagert hat. Aus dem Baum, welcher den gesamten Cluster zeigt, kann man vermuten dass die mir-90-2 die ursprüngliche microRNA ist, wohingegen die mir-90-1 aus einer Tandemduplikation entstanden ist. Der grobe Aufbau des Clusters stellt sich anhand der gewonnenen Genompositionen folgendermaßen dar: Am upstream Ende des Clusters befindet sich die mir-90, an die sich downstream in allen Arten eine relativ große Lücke von bis zu 14kb anschließt. Danach folgen die mir-238, die trotz Ihres 'nonfamilied' Status in allen Vertretern der Gattung gefunden wurde, und die mir-80. Ein genauer Rückschluss der Evolutionsgeschichte auf die jeweiligen Artspezifischen Abstände zwischen den Familien ist leider nicht möglich, da *C.japonica*, *C.brenneri* und *C.remanei* noch nicht vollständig sequenziert sind und somit nicht eindeutig geklärt werden kann an welchen exakten Positionen die miRNAs letztendlich liegen. Daher kann auch der zuvor geschilderte Aufbau des Clusters nur als Vermutung im Hinblick auf die schon fertigen Genome betrachtet werden. Die Duplikation der mir-90 läßt nun einigen Raum um die Evolution ein wenig zu interpretieren. So ist es zum Beispiel möglich dass die Duplikation erst nach dem Abzweig des *C.briggsae* stattfand, und sie somit ein individuelles Ereignis war. Auf der anderen Seite kann es aber auch sein, dass die mir-90 schon vor der Abspaltung der einzelnen Arten als 2. Kopie vorlag, sie dann aber in allen Vertretern bis auf *C.briggsae* verloren ging. Des Weiteren kann man darüber spekulieren ob die Duplikation upstream, d.h. vor den Cluster, oder downstream, d.h. vor die 2.miRNA, stattfand. Nur kann leider niemand die genaue Entwicklung mit Sicherheit beweisen noch die anderen wiederlegen.

Da alle im Cluster enthaltenen microRNAs ausschließlich in dieser einen Gattung vorkommen, kann man vermuten dass sie ganz spezielle regulatorische Aufgaben ausführen. Diese konnten aber mit der Tarbase und der Wormbase nicht genauer spezifiziert werden.

## 3.8 mir-96 Cluster

### 3.8.1 Übersicht über das Cluster

Das mir-96 Cluster besteht aus den 3 Familien mir-96, mir-182 und mir-183, welche scheinbar nicht verwandt sind. Das Cluster wurde nur in Deuterostomia identifiziert, wobei es in Echinodermaten nicht gefunden wurde. In Urchordaten wurde nur mir-183 entdeckt, in *Branchiostoma floridae* zusätzlich dazu noch die mir-96. Die dritte Familie, mir-182, konnte zuerst in *Callorhinchus mili* identifiziert werden. Alle weiteren untersuchten Vertebratengenome bis auf *Petromyzon marinus* sowie die Sauropsida enthielten das komplette mir-96 Cluster. Dabei konnte im Genom von *Gallus gallus* sowohl die mir-96 als auch die mir-182, in *Anolis carolinensis* die mir-96 und in *Taeniopygia guttata* die mir-182 nicht lokalisiert werden. Schlussfolgerung sind durch die sehr frühen Genomversionen daraus zunächst nicht zu treffen, wobei ein teilweiser oder kompletter Verlust der MicroRNAs mir-182 und mir-196 durchaus möglich wäre.

### 3.8.2 Evolutionäre Geschichte

Bei der evolutionären Geschichte sind zwei Varianten möglich. Zum einen könnten das komplette Cluster schon bei der Abzweigung der Urchordaten vorhanden gewesen sein und nur bei *Ciona intestinalis* bis auf eine Familie verloren gegangen sein. Außerdem hätte dann auch *Branchiostoma floridae* eine Familie während der eigenständigen evolutionären Entwicklung verloren. Zum anderen könnte es sich bei den Familien mir-96 sowie mir-182 um *de novo* Innovationen handeln, die vor der Abzweigung der Knorpelfische entstanden sind. Jedoch lässt sich aufgrund von fehlenden Informationen nicht genau sagen, zu welchem Zeitpunkt die Familien entstanden sind, da zu wenig genomische Daten von basalen Chordaten vorhanden sind. Die aktuellen Daten würden dafür sprechen, dass die erste *de novo* Innovation vor den Urchordaten, die zweite vor den Cephalochordaten und die dritte vor den Vertebraten stattgefunden hat. Mit mehr genomischen Daten könnten diese Zeitpunkte in Zukunft noch präzisiert werden. Als gesichert hingegen angenommen werden, dass die Vertebraten das komplette Cluster besitzen und sich dieses Cluster bei der Teleost eigenen Genomduplikation dupliziert hat. Allerdings ist in *Danio rerio* diese Clusterkopie im Laufe seiner individuellen Entwicklung verloren gegangen.

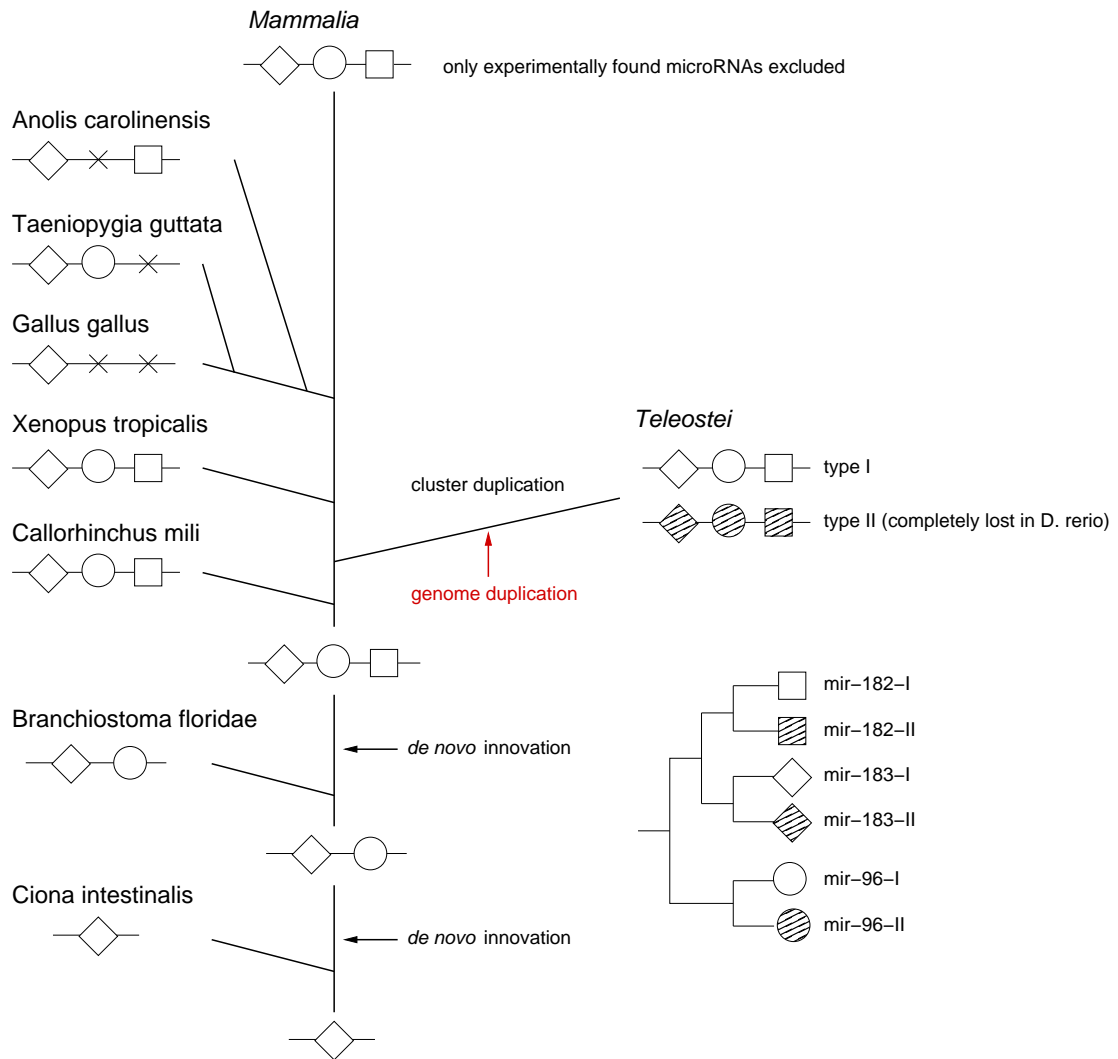


Abbildung 12: Übersicht über die vermutliche evolutionäre Entwicklung des mir-96 Clusters



## 3.9 mir-181 Cluster

### 3.9.1 Übersicht über das Cluster

Das mir-181 Cluster besteht aus nur einer Familie. Man findet ihn in mindestens je 4 Kopien in den *Teleostemi*. In den *Teleostei* kam es zunächst zu einer Duplikation und danach zu Verlusten in den verschiedenen Arten. Auch in den *Mammalia* kam es zu einer Duplikation.

### 3.9.2 Evolutionäre Geschichte

Am Anfang gab es wahrscheinlich eine einzelne mir-181. Durch Genomduplikation entstand daraus ein Cluster bestehend aus mir-181a und mir-181b. Bei einer zweiten Genomduplikation entstand duplizierte sich der komplette Cluster auf ein anderes Chromosom. Nach diesen beiden Duplikationen, welche noch vor der Abspaltung der *Teleostei* erfolgten, gab es mir-181a-1 und mir-181b-1 auf einem Chromosom und auf einen anderen mir-181a-2 und mir-181b-2. Welches von den beiden die ursprüngliche war, ist aus dem Alignment nicht ersichtlich. In den *Teleostei* trat eine dritte Genomduplikation auf, bei der sich das komplette Cluster nocheinmal duplizierte. *D. rerio* duplizierte zusätzlich nocheinmal die mir-181a-1 Typ II und die mir-181b-1 Typ II, verlor aber später die duplizierte mir-181b-1 Typ II. *O. latipes* verlor 5 der 8 miRNAs: mir-181b-1 Typ I und Typ II, mir-181a-2 Typ I und Typ II und die mir-181b-2 Typ II. *G. aculeatus* verlor nix und duplizierte auch nix. *T. rubripes* und *T. nigroviridis* verloren mir-181a-1 Typ II und mir-181b-2 Typ II. *X. tropicalis* und *G. gallus* behielten die vor den *Teleostei* vorhanden 4 miRNAs in diesem Cluster. In den *Mammalia* kam es zu einer Duplikation der mir-181a-1 und mir-181b-1 zu mir-181c und mir-181d.

s

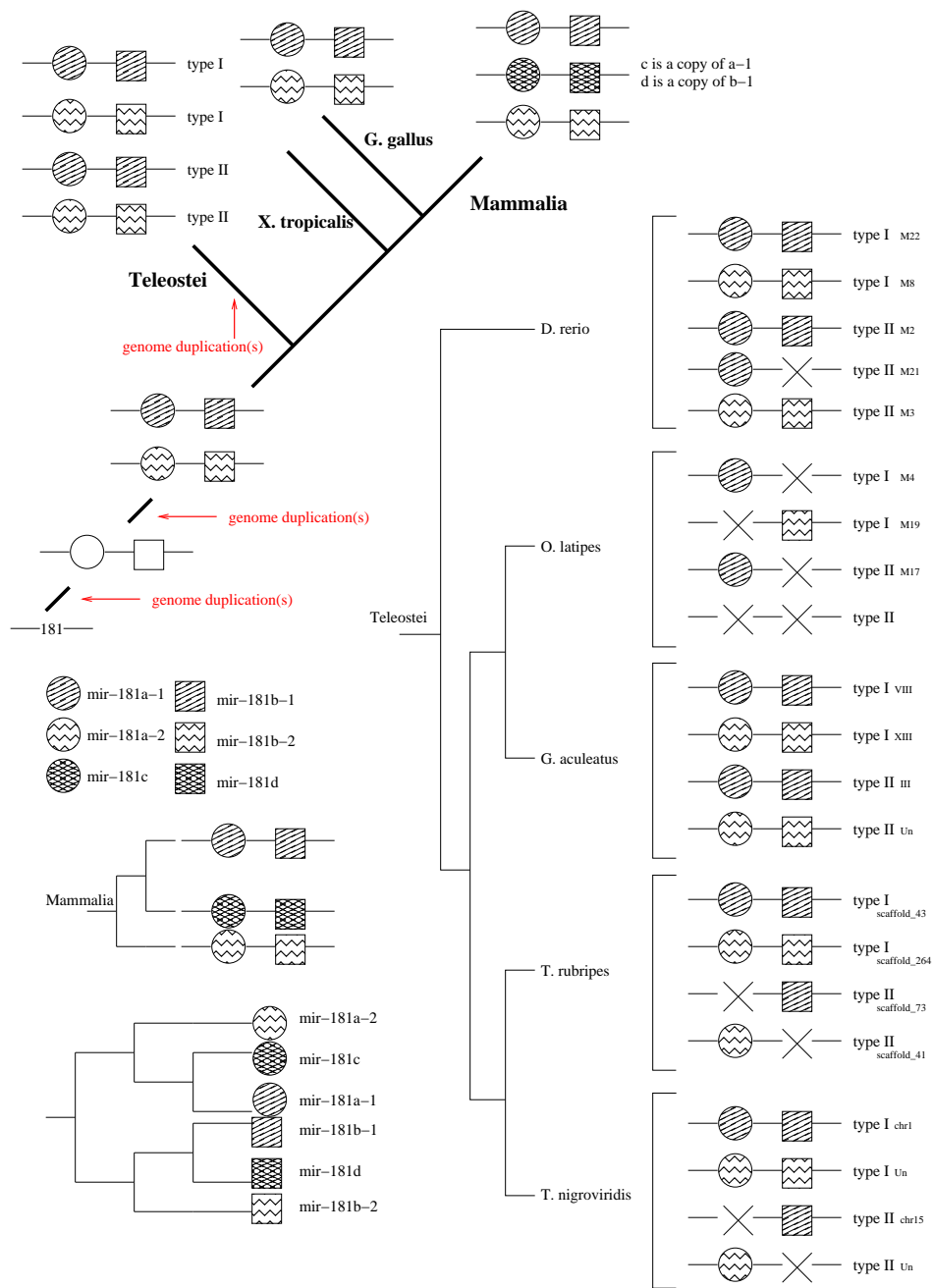


Abbildung 13: Übersicht über die vermutliche evolutionäre Entwicklung des mir-181 Clusters

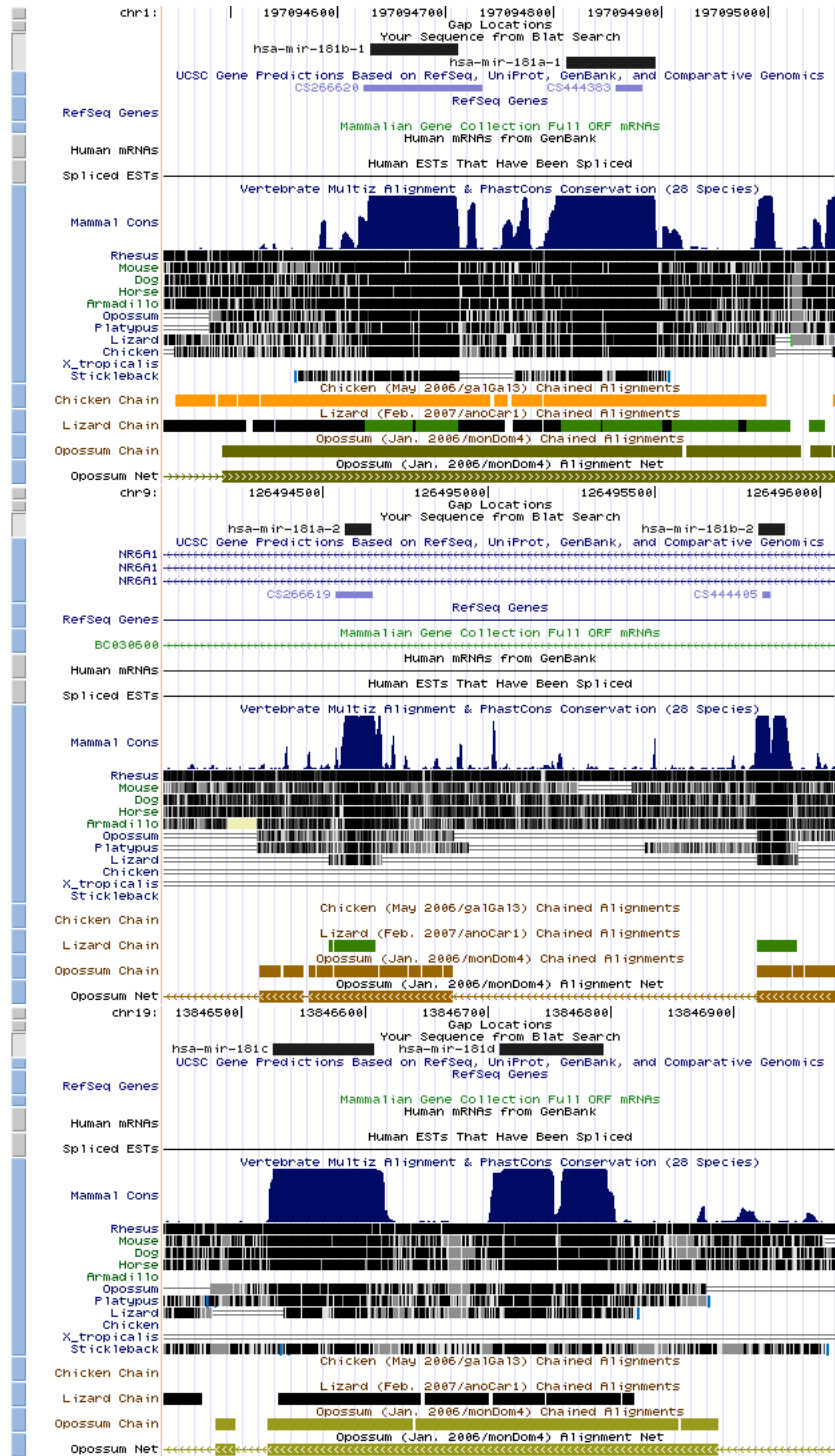


Abbildung 14: Konservierung des mir-181 Clusters in *H. sapiens*

## 3.10 mir-191 Cluster

### 3.10.1 Übersicht über das Cluster

Das mir-191 Cluster besteht aus 2 Familien mir-191 und mir-425. Es ist in allen *Tetrapoda* einmal zu finden. Das revere Komplement der dre-mir-462 passt sehr gut zu der mir-191 Familie.

### 3.10.2 Evolutionäre Geschichte

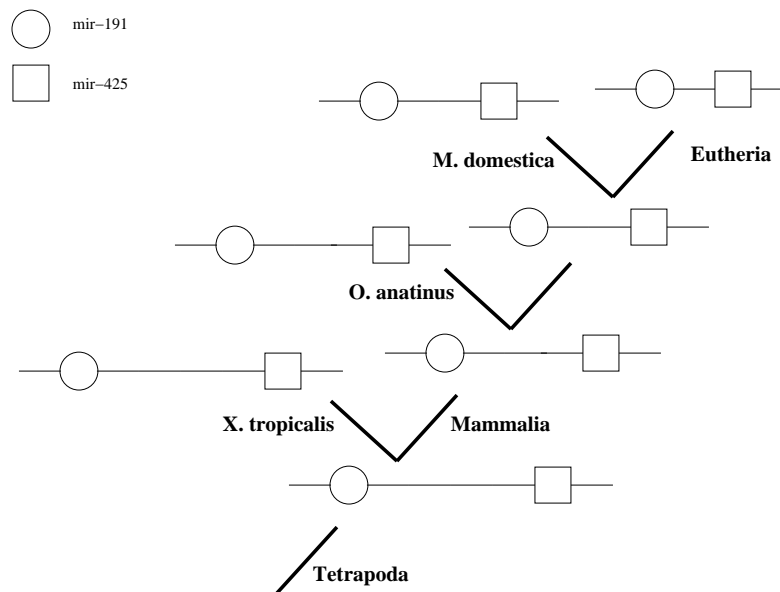


Abbildung 15: Übersicht über die vermutliche evolutionäre Entwicklung des mir-191 Clusters

Im Cluster lagen zunächst die mir-191 und die mir-425 getrennt durch mehrere tausend Basen getrennt nebeneinander. Dieser Abstand verkürzte sich immer weiter. So beträgt der Abstand im *X. tropicalis* rund 3500 Basen. Im *O. anatinus* ist er auf rund 2700 Basen verkürzt und im *M. domestica* sogar auf 2600 Basen. Bei diesen 3 *Tetrapoda* beinhalten die Zwischenstücke die repeat regions. In den *Eutheria* sind zwischen 762 und 779 Basen zwischen der mir-191 und der mir-425. Das Zwischenstück beinhaltet keine annotierten Bereiche.

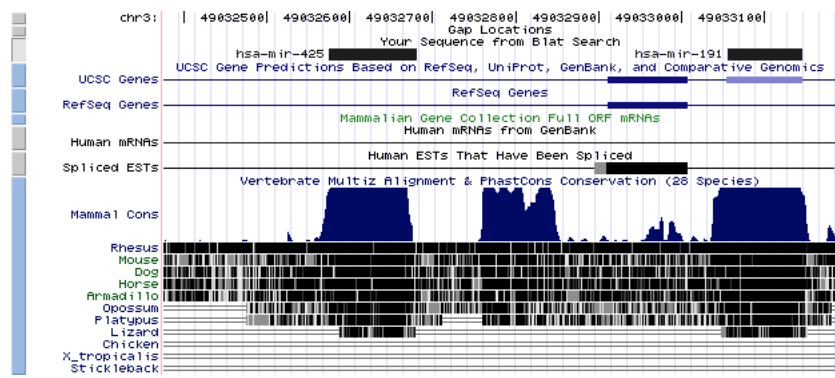


Abbildung 16: Konservierung des mir-191 Clusters in *Homo sapiens*

## 3.11 mir-193 Cluster

### 3.11.1 Übersicht über das Cluster

Das mir-193 Cluster besteht aus den zwei Familien mir-193 und mir-365, welche scheinbar nicht miteinander verwandt sind. Das Cluster wurde nur in den Vertebraten in zwei Clusterkopien ab *Callorhinchus mili* gefunden, wobei dabei die Familie mir-193 nur einmal gefunden wurde. Bei den Teleost Fischen hat *Danio rerio* drei Kopien, wohingegen die anderen Knochenfische nur ein Cluster noch besitzen. Sowohl in *Xenopus tropicalis* als auch *Anolis carolinensis* ist nur noch jeweils eine Kopie vorhanden. Bei Mammalia konnten bei einzelnen Familien bestimmte miRNAs nicht lokalisiert werden und bei *Rattus norvegicus* ist der komplette Typ a Cluster verschwunden.

### 3.11.2 Evolutionäre Geschichte

Als Resultat der ersten zwei Clusterduplikationen konnten die zwei Clusterkopien Typ a und Typ b in den meisten Vertebraten wieder gefunden werden. Dabei sind teilweise Cluster partiell oder komplett verloren gegangen. Hauptsächlich bei partiellen Verlusten in Mammalia ist es wahrscheinlicher, dass diese Verluste auf unvollständige Genome zurückzuführen sind. In den Knochenfischen erfolgte eine weitere Clusterduplikation. Das Clusters Typ a wurde dabei auf das gleiche Chromosom wie das Cluster Typ b kopiert jedoch etwa 8 Mb upstream kopiert. Allerdings ist das ursprüngliche Cluster Typ a und b nur noch im *Danio rerio* vorhanden und die Kopie des Clusters Typ b ist in allen Teleost Fischen verloren gegangen. Zu den nicht gefundenen Clusterkopien in *Anolis carolinensis* sowie in *Xenopus tropicalis* lässt sich nicht mit Bestimmtheit sagen, ob es sich dabei um verlorene Cluster oder noch nicht gefundene Cluster handelt. Jedoch ist durch die großen Zeiträume evolutionärer Entwicklung durchaus denkbar, dass die vorhandenen Cluster die Hauptaufgaben dieser miRNAs übernommen haben und die jeweils anderen sich dann unterschiedlich entwickelt haben. Dadurch könnte ihre miRNA Funktionalität vollständig verloren gegangen sein oder sich verändert haben, wodurch das Cluster dann nicht mehr durch Blast gefunden wird.

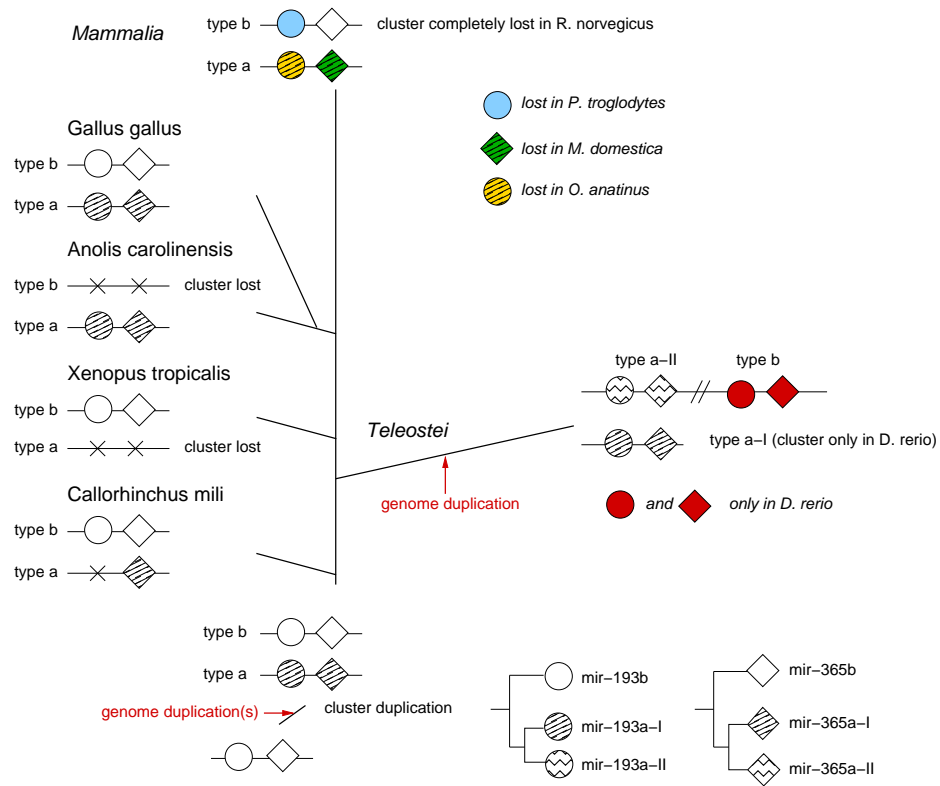


Abbildung 17: Übersicht über die vermutliche evolutionäre Entwicklung des mir-193 Clusters

## 3.12 mir-221 Cluster

### 3.12.1 Übersicht über das Cluster

Das mir-221 Cluster besteht nur aus der Familie mir-221, wobei diese sich in die bereits in der miRBase annotierten Unterfamilien mir-221 und mir-222 aufspaltet. Bei einer der Genomduplikationen hat sich dabei die ursprüngliche miRNA auf das gleiche Chromosom dupliziert, wodurch nun das Cluster aus den zwei nebeneinanderliegenden Unterfamilien besteht. Das Cluster konnte erst in den Vertebraten ab *Callorhinchus mili* gefunden werden und ist dann in fast allen Vertebraten vorhanden. Die nicht entdeckte mir-221 in *Callorhinchus mili* könnte auf einen evolutionären Verlust während der separaten Entwicklung der Knorpelfische hinweisen. Dies müsste jedoch noch mit zusätzlichen genomischen Daten dieser Klasse verglichen werden, um diese Vermutung zu untermauern. Die nicht gefundene mir-222 in *Pan troglodytes* lässt sich höchstwahrscheinlich durch das unvollständige Genom erklären.

### 3.12.2 Evolutionäre Geschichte

Während der Genomduplikationen kam es zu einer Tandemduplikation der ursprünglichen miRNA Familie, welches das Cluster aus den Unterfamilien mir-221 und mir-222 bildete. Die Teleost Fische haben durch ihre zusätzliche Genomduplikation eine Kopie dieses Clusters, wobei in *Oryzias latipes* das Type-I Cluster und in *Tetraodon nigroviridis* das Type-II Cluster verschwunden ist. Da beide Cluster nach der Duplikation noch die selbe Funktion in der Zelle erfüllen, ist dieser Verlust eines der Cluster nicht weiter ungewöhnlich. Ab *Xenopus tropicalis* konnten beide Clusterkopien außer in *Pan troglodytes* in allen untersuchten Genomen vollständig lokalisiert werden. Als Target der MicroRNA 221 und 222 wurde das Gen *Kit* und die dadurch herunterregulierte Translation des Rezeptors *Kit* in der Tarbase annotiert. Diese Translationshemmung hat laut dem Paper von Felli et al. eine Inhibition der Erythropoese und des normalen Wachstums von Erythroleukämiezellen zur Folge [1], wobei das zweite das Potential einer Krebstherapie mit Hilfe von miR-221 und miR-222 bieten könnte.



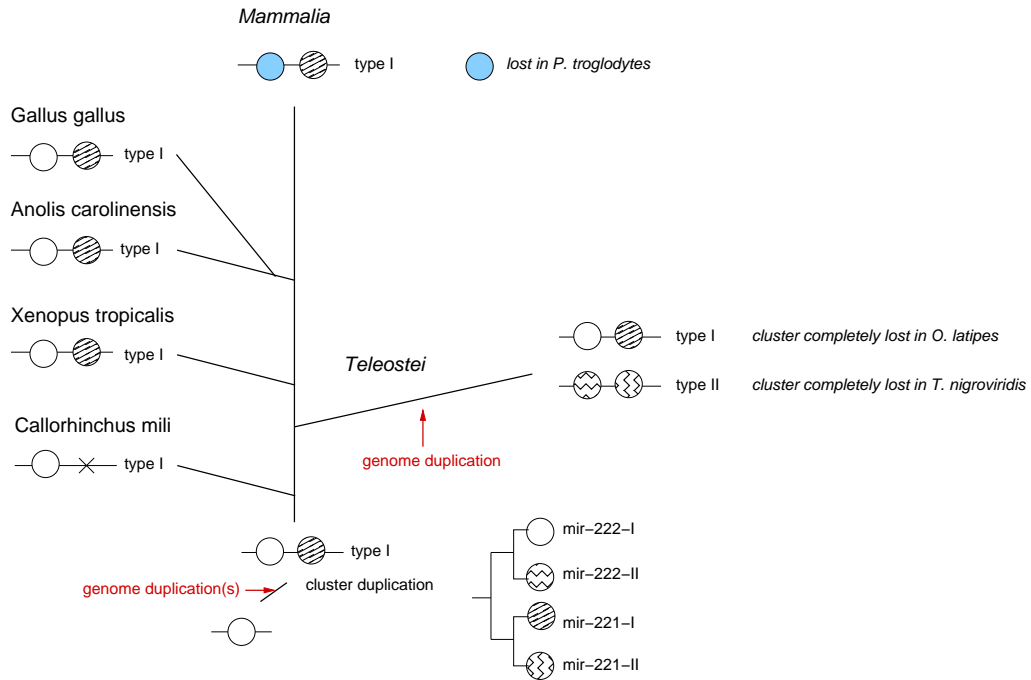


Abbildung 18: Übersicht über die vermutliche evolutionäre Entwicklung des mir-221 Clusters

### 3.13 mir-232 Cluster

Der mir-232 Cluster ist kein Cluster im eigentlichen Sinne, da er nur aus einer microRNA besteht. Allerdings bestand in der früheren miRBase-Version ein doppelter Eintrag für die mir-232, sodass man davon ausging es handele sich um mehrere Familien. Da dieser Irrtum nun ausgeräumt ist, hat man nur noch eine einzelne miRNA-Familie. Die gegebenen Inputsequenzen wurden nur auf den korrespondierenden Chromosomen wiedergefunden. D.h. die microRNA ist *Caenorhabditis* spezifisch und kommt in keinen weiteren Organismen vor. Beide Sequenzen sind schon in der miRBase annotiert.

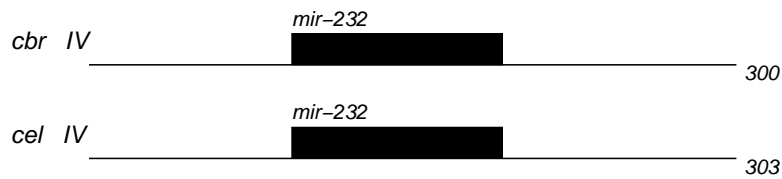


Abbildung 19: Aufbau des mir-232 Clusters

Eine Rekonstruktion der evolutionsspezifischen Geschichte ist nun relativ trivial, da beide miRNAs nur im *C.elegans* und *C.briggsae* vorkommen, d.h. sie sind wahrscheinlich nach der Abspaltung der Strongyloidea entstanden. Um dies genauer abzuklären müsste man die Sequenzen noch gegen verschiedene Genome der Gattung *Caenorhabditis* bzw andere Nematoden blasten (vorrangig im Trace Archive).

Jedoch ist die Wahrscheinlichkeit, dass es sich um hochspezifische microRNAs handelt hoch, sodass man davon ausgehen kann dass diese Familie *Caenorhabditis*-spezifische Entwicklungsprozesse steuert. Allerdings wurde in der Tarbase und Wormbase kein Hinweis auf eine bisher bekannte Aufgabe gefunden.

## 3.14 mir-275 Cluster

### 3.14.1 Übersicht über das Cluster

Der mir-275 Cluster besteht aus den Familien mir-275 und mir-305. Beide wurden in verschiedenen Vertretern der Mandibulata wiedergefunden, wobei jeweils nur ein Cluster pro Genom gefunden wurde. Es wurden alle in der miRBase annotierten miRNAs gefunden; darüber hinaus wurden die Sequenzen in den Arten *Daphnia pulex* und *Tribolium castaneum* als kompletter Cluster entdeckt.

### 3.14.2 Evolutionäre Geschichte

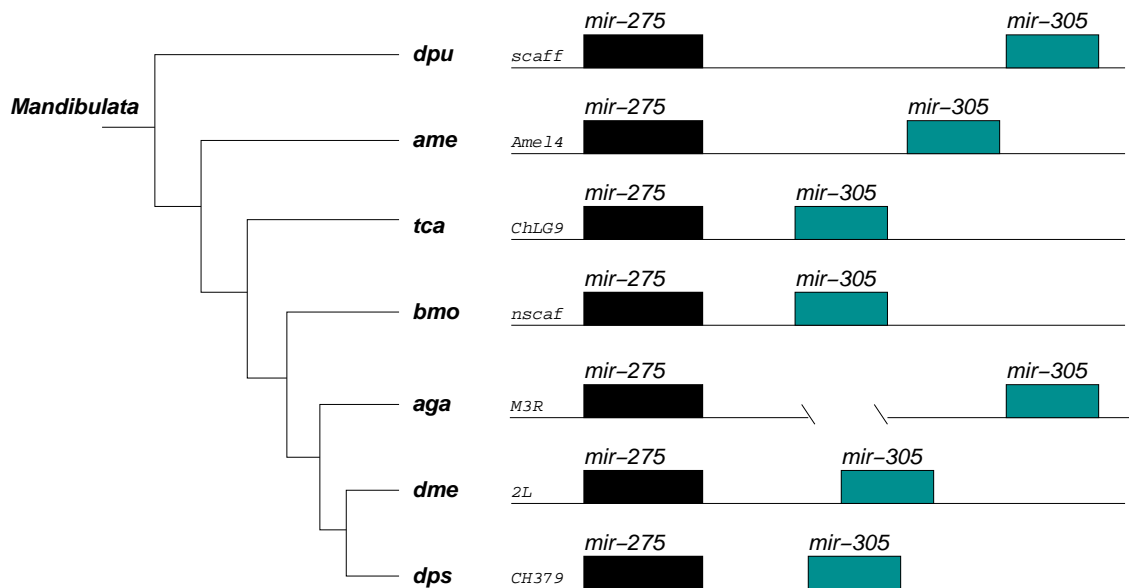


Abbildung 20: Aufbau des mir-275 Clusters

Die Rekonstruktion der Evolutionsgeschichte ist relativ einfach, da das gesamte Cluster höchst wahrscheinlich nach dem Entwicklung der Mandibulata entstanden ist. Um dies genauer in Erfahrung zu bringen müsste noch gegen *Ixodes scapularis* geblastet werden, wobei dieses Genom noch nicht

richtig assembliert ist und deswegen auch keine vollstaendige Auflösung bringen kann. Der Versuch die Sequenzen in Nematoden beziehungsweise in Mollusken oder Anneliden zu finden schlug fehl. D.h. es sind mit hoher Sicherheit sehr spezifische microRNAs. Dies erkennt man auch am Clusteraufbau und dem Clusterbaum, denn die Sequenzen zeigen eine hohe Konservierung und der Clusteraufbau ist über alle Vertreter der Mandibulata konsistent. Lediglich bei *Daphnia pulex* wurden beide miRNAs auf verschiedenen Contigs gefunden, da aber das Genom noch nicht vollständig ist, auf vielen einzelnen Contigs verstreut liegt und bei allen anderen Arten ein einheitlicher Aufbau zu erkennen ist liegt die Vermutung nahe, dass dies auch hier der Fall ist. Einen Aufschluss über eine mögliche regulatorische Verwendung konnte die Tarbase leider nicht liefern.

## 4 Clusteraufteilung

Clustername	Verantwortlicher
mir-1	Lydia
mir-3	Lydia
mir-8	Sebastian
mir-23	Christian
mir-34	Axel
mir-58	Sebastian
mir-61	Christian
mir-90	Sebastian
mir-96	Christian
mir-130	Axel
mir-181	Lydia
mir-191	Lydia
mir-193	Christian
mir-221	Christian
mir-232	Sebastian
mir-249	Axel
mir-275	Sebastian
mir-302	Axel

## Literatur

- [1] N Felli, L Fontana, E Pelosi, R Botta, D Bonci, F Facchiano, F Liuzzi, V Lulli, O Morsilli, S Santoro, M Valtieri, G A Calin, C G Liu, A Sorrentino, C M Croce, and C Peschle. Micornas 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation. *Proc Natl Acad Sci U S A*, 102(50):18081–18086, Dec 2005.
- [2] Jana Hertel, Manuela Lindemeyer, Kristin Missal, Claudia Fried, Andrea Tanzer, Christoph Flamm, Ivo L. Hofacker, Peter F. Stadler, and The Students of Bioinformatics Computer Labs 2004 and 2005. The expansion of the metazoan microRNA repertoire. *BMC Genomics*, 7:25 [epub], 2006.
- [3] Y Nakatani, H Takeda, Y Kohara, and S Morishita. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res*, 17(9):1254–1265, Sep 2007.

- [4] K R Norman, R T Fazio, J E Mellem, M V Espelt, K Strange, M C Beckerle, and A V Maricq. The rho/rac-family guanine nucleotide exchange factor vav-1 regulates rhythmic behaviors in *c. elegans*. *Cell*, 123(1):119–132, Oct 2005.
- [5] D Steinke, S Hoegg, H Brinkmann, and A Meyer. Three rounds (1r/2r/3r) of genome duplications and the evolution of the glycolytic pathway in vertebrates. *BMC Biol*, 4:16–16, 2006.
- [6] A Tanzer and P F Stadler. Evolution of micrnas. *Methods Mol Biol*, 342:335–350, 2006.