

Molekulare Evolution der Vertebraten snoRNA

Diplomarbeit

zur Erlangung des akademischen Grades
Diplom Ing. (FH)

durchgeführt unter der Leitung

von Prof. Dr. Bernhard Haubold
und
Prof. Dr. Peter F. Stadler

eingereicht von Bettina Müller

im Fachbereich Biotechnologie/Bioinformatik,
Fachhochschule Weihenstephan

Ochsenfurt, am 6. Oktober 2004

Eidesstattliche Erklärung

Gemäss §23 Abs. 6 der Prüfungsordnung

Ich erkläre hiermit an Eides statt, dass die von mir vorgelegte Arbeit selbst und ohne fremde Hilfe verfasst und noch nicht anderweitig für Prüfungszwecke vorgelegt wurde.

Es wurde keine als die angegebenen Quellen oder Hilfsmittel benutzt. Wörtliche und sinngemässe Zitate sind als solche gekennzeichnet.

Ochsenfurt, 6. Oktober 2004

Bettina Müller

Danksagung

Zuerst möchte ich mich besonders bei Prof. Dr. Peter Stadler für die exzellente Betreuung und tatkräftige Unterstützung bedanken. Des Weiteren gilt mein Dank Prof. Dr. Bernhard Haubold von der Fachhochschule Weihenstephan für die Unterstützung und Hilfe bei der Arbeit.

Ausserdem bedanke ich mich beim gesamten Leipziger Bioinformatiklehrstuhl für die Ratstellung bei biologischen oder softwaretechnischen Problemen. Insbesondere gilt mein Dank Andrea Tanzer für ihre Betreuung während des Praktikums und Unterstützung bei der Diplomarbeit.

Gleichfalls bedanke ich mich bei meinen Eltern und Geschwistern, die mich im letzten Jahr unterstützt haben.

Abstract

SnoRNAs sind nicht kodierende RNAs, die für die Modifikation und Prozessierung der pre-rRNA zuständig sind und dadurch zur Stabilisierung der ribosomalen RNA beitragen. Die ribosomale RNA wird durch zwei unterschiedliche snoRNA-Typen modifiziert. SnoRNA-Typ Box C/D ist für die 2-O-Methylierung der ribosomalen RNA zuständig; snoRNA-Typ Box H/ACA modifiziert die ribosomale RNA durch Pseudouridylierung.

Im Genom liegen beide snoRNA-Typen sowohl in Intronen, die ihre Aufgabe in der Translation haben, als auch auf Hostgenen, die funktionslos und als Trägergenen für die snoRNA vorhanden sind.

Ziel dieser Diplomarbeit war zum einen die detaillierte Suche nach neuen snoRNAs in den Vertebraten. Zu Beginn der Arbeit waren insgesamt 25 Teleosten-snoRNAs in öffentlichen Datenbanken publiziert. Bei der Suche nach neuen snoRNAs wurde besonderer Augenschein auf die ältesten Vertebratengenome geworfen. Zum anderen war ein weiteres Ziel der Diplomarbeit eine detaillierte Beschreibung der Evolution der snoRNA anhand ausgewählter Vertebraten.

Bevor evolutionäre Aussagen getroffen werden können, müssen neben den schon veröffentlichten snoRNAs neue snoRNAs an unterschiedlichen Genomen gefunden werden. Zu Beginn der Arbeit existierten zwei Programme, die für die Suche nach snoRNAs spezialisiert sind (Programm `Fisher` und Programm `SnoScan`) Mit dem Programm `Fisher` und anschliessend mit dem Programm `SnoScan` wurden neue snoRNAs am Genom des *Takifugu rubripes* gesucht. Die Homologiemethode `Blast` wurde des Weiteren zur Suche neuer snoRNAs an unterschiedlichen Genomen verwendet. Die mittels den Programmen `Fisher`, `SnoScan` und `Blast` gefundenen snoRNAs wurden durch Bestimmung der ribosomalen Bereiche, Lokalisation im Genom und Sequenzhomologie verifiziert.

Anhand der aus der Suche gewonnenen Daten wurden, basierend auf phylogenetischen Analysen, evolutionäre Szenarien für vier unterschiedliche snoRNA-Klassen aufgestellt (E1/U17, E2, E3 und U36). Dabei zeigte sich, dass es in der Evolution der snoRNA zu Genduplikationen, -mutationen und zu Sprüngen in der genomischen Lokalisation gekommen ist.

Inhaltsverzeichnis

1	Einführung	1
1.1	Nicht kodierende RNA	1
1.2	SnoRNA	2
1.2.1	Struktur der snoRNA Boxen	3
1.2.2	Genomische Organisation der snoRNAs	5
1.2.3	Wie funktionieren snoRNAs?	7
1.2.4	Modifikation unterschiedlicher nicht kodierender RNAs	9
1.2.5	Evolutionärer Ursprung der snoRNA	9
1.3	Ribosomale RNA	11
1.3.1	Struktur der Ribosomen	11
1.3.2	Genomische Lokalisation der rRNA	12
1.3.3	Reifung der rRNA	13
1.3.4	Warum wird die rRNA modifiziert?	15
1.3.5	Evolution der ribosomalen RNA	16
1.4	Evolution anderer nicht kodierender RNAs	19
1.4.1	Transfer RNA	19
1.4.2	MicroRNA	20
1.5	Ziel der Arbeit	24
1.6	Vorgehensweise	24
2	Bioinformatische Programme	25
2.1	Fisher: Suche nach H/ACA snoRNAs	25
2.2	SnoScan: Suche nach C/D snoRNAs	29
2.3	Blast: Homologie-Vergleiche	31
2.4	Das Vienna RNA Package	33

2.5	Methoden der phylogenetischen Analyse	34
2.5.1	Neighbor-Joining	34
2.5.2	Bootstrap	35
3	Neue snoRNAs in den Vertebraten	36
3.1	Ergebnisse der Fisher snoRNA Suche	37
3.1.1	Analyse der ψ -Bereiche und Distanzen	37
3.1.2	Resultate des Testlaufs	40
3.1.3	Probleme	41
3.2	Suche nach Box C/D snoRNAs mit SnoScan	43
3.2.1	Auswertung der Ergebnisse	43
3.2.2	Probleme	45
3.3	Resultate des NCBI-Blast	46
3.3.1	Die Geschichte der snoRNA-Familie U17/E1	46
3.3.2	Familie U24	51
4	Evolution der snoRNA	54
4.1	Stammbaum der Wirbeltiere	54
4.2	Szenarien der snoRNA-Evolution	57
4.2.1	Evolution der snoRNA U17/E1	57
4.2.2	Evolutionsszenarien für snoRNA E2 und E3	59
4.2.3	Evolution der snoRNA U36	64
5	Diskussion und Ausblick	71
5.1	Diskussion der Ergebnisse	71
5.2	Ausblick	73
A	Sequenzen und Strukturen	I
A.1	Sekundärstrukturen	I
A.2	Sequenzen	II
B	Bootstrapbäume	X
C	Verzeichnisse	XV

Kapitel 1

Einführung

1.1 Nicht kodierende RNA

Fast 98 % der menschlichen Transkripte bestehen aus nicht kodierender RNA. Unter nicht kodierender RNA werden alle RNA-Transkripte verstanden, die nicht die Fähigkeit haben, Proteine zu kodieren. Diese Definition schließt zu einem die Introns, als auch die Exons, die keine Proteinkapazität besitzen, ein [87].

Innerhalb der letzten 20 Jahre ist eine große Zahl von nicht kodierenden RNAs gefunden worden (Tab.: 1.1). Das zentrale Dogma der Molekularbiologie besagt, dass die genetische Information, die als DNA vorliegt, an einem Strang zu RNA-Molekülen transkribiert wird. Diese RNA gelangt als messenger RNA (mRNA) ins Cytoplasma. Dort erfolgt mittels transfer RNA-Molekülen (tRNA) dann die Bindung der mRNA an die Ribosomen und danach die Translation der mRNA in ein Protein.

Dieses Postulat, das 1956 von Francis Crick aufgestellt wurde, hat bis heute seine Gültigkeit bewahrt. Allerdings berücksichtigt es nicht die Möglichkeit der reversen Transkription und, dass eine Anzahl von bestimmten RNAs existiert, die keine kodierende Funktion haben. Sondern eine funktionelle Rolle haben - die nicht kodierenden RNAs.

Nicht kodierende RNAs haben wichtige Aufgaben: Zum einen in der Prozessierung und Modifikation der RNA, zum anderen auch in der Regulation der Transkription. Small nuclear RNA (snRNA) hat ihre Funktion in der Prozessierung der mRNA, indem sie im Spleißvorgang der pre-mRNA involviert ist. MicroRNA (miRNA)(Kap.: 1.4.2) und siRNA (small interfering RNA) sind

nicht kodierende RNAs, die für die Regulation der Transkription zuständig sind. Weiterhin existiert eine ganze Reihe von ncRNAs, denen noch keine Funktion zugeordnet werden kann. Ein Beispiel einer solchen nicht kodierenden RNA ist Y RNA. Diese RNA ist eine kleine cytoplasmatische RNA, deren Funktion noch nicht bekannt ist, obwohl viele Jahre daran geforscht wurde [86].

Gabriele Storz sagte im Jahre 2002 *“A big challenge for the future will be to identify the whole complement of ncRNAs and to elucidate their functions. This is an exciting time for investigations whose work focused on ncRNAs [86].“*

Nicht kodierende RNAs	
NcRNA	Funktion
tRNA	Translation von genetischer Information
rRNA	Bestandteil des Ribosoms
snRNA	Bestandteil des Spliceosoms, Splicing der pre-mRNA
snoRNA	RNA Modifikation der snRNA und der rRNA
RNase P RNA	Änderung der 5'-Enden der pre-tRNA
miRNA	Unterdrückung der Translation von pre-mRNA
tmRNA	Trans-Translation
siRNA	Unterdrückung der Translation von pre-mRNA
dsRNA	doppelsträngige RNA, involviert im RNA Silencing

Tabelle 1.1: Verzeichnis einiger nicht kodierender RNAs deren Funktion bekannt ist.

1.2 SnoRNA

Der Zellkern besitzt ein Organell, den Nucleolus, in dem die Prozessierung der ribosomalen RNA stattfindet. Neben der pre-mRNA befindet sich im Nucleolus noch die snoRNA, die so genannte small nucleolar RNA.

Die Aufgabe von ein paar wenigen snoRNAs ist die Spaltung der pre-rRNA. Der Hauptteil der snoRNAs hat jedoch die posttranslationale Modifikation der rRNA [5], snRNA [44] und mRNA als Aufgabe. SnoRNAs modifizieren

die genannten RNA-Arten durch die 2'-O-Methylierung oder durch die Pseudouridylierung. SnoRNAs können entsprechend der unterschiedlichen Modifizierungsarten in zwei Subklassen unterteilt werden, Box C/D und Box H/ACA. Beide Klassen zeichnen sich durch das Auftreten zweier konservierter Sequenzmotive aus.

1.2.1 Struktur der snoRNA Boxen

Box C/D

Box C/D ist für die 2'-O-Methylierung der rRNA, der snRNA und der mRNA zuständig [44]. Der Begriff Box C/D entwickelte sich aus der Benennung der konservierten Elemente der snoRNA. Als charakteristische Merkmale dienen bei der Box C/D folgende zwei Sequenzmotive: Box C (5'UGAUGA3') und Box D (5'CUGA3'). Diese Sequenzmotive befinden sich in der Nähe vom 5'- und 3'- Ende.

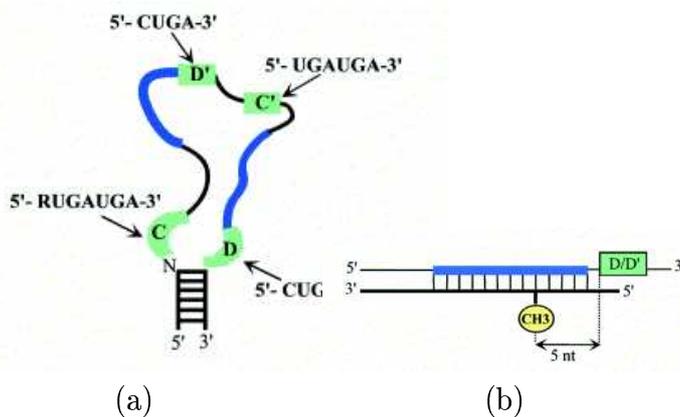


Abbildung 1.1: (a) Struktur der Box C/D mit ihren Erkennungsmotiven. Im blau markierten Bereich ist snoRNA komplementär zur rRNA. (b) Die Methylierungsstelle ist 5 Nukleotide von der Box D (D') entfernt (Abbildung Bachelierie[2002])

Neben diesen zwei Sequenzmotiven tritt zusätzlich eine Kopie der Box C (Box C') und der Box D (Box D') auf. Box C' ist in der Mitte der Struktur situiert,

gionen, liegen die Bereiche, die komplementär zur rRNA und für die Pseudouridylierung zuständig sind (Kap.: 1.3.4). Die Pseudouridylierungs-Bereiche sind 4 - 8 nt lang und auf zwei Stellen der Schleife verteilt. SnoRNA und Targetsequenz gehen, wie die Box C/D, eine direkte Basenpaarung ein und formen zwei kleine Duplexe, die 4 - 8 bp lang sind. Die Bereiche für Pseudouridylierung treten wie bei Box C/D in definierter Distanz zur Box H oder ACA auf. Sie liegen 15 bp vor der Box H oder ACA (Abb. 1.2.1) [44].

Die Ausbildung von unterschiedlichen H/ACA snoRNAs beruht auf der Variation der Anzahl von Nukleotiden innerhalb der Domänen [5].

1.2.2 Genomische Organisation der snoRNAs

Die genomische Organisation der snoRNAs ist sehr vielfältig. Zum einen können snoRNAs in Introns der pre-rRNA oder in Hostgenen vorkommen. Zum anderen sind snoRNAs auch in unabhängigen polycistronischen Genen vorhanden.

Die mono- bzw. polycistronische Organisation kommt speziell bei Pflanzen oder bei der Hefe vor. SnoRNAs der Vertebraten sind in Introns oder in Hostgenen zu finden [89, 10].

Während bei den polycistronischen Transkripten mehrere snoRNAs auftreten können, kann in den Introns nur eine snoRNA auftreten. In den Hostgenen kann allerdings auch mehr als nur eine snoRNA vorkommen [89, 10].

Die Mehrzahl der intronischen snoRNAs liegt auf Genen, die ihre Funktion bei der Translation haben. Diese Gene sind zum Beispiel ribosomale Proteine (RPS3, RPS8, RPL1a, RPL5, RPL7a und RPL13a), Proteinsynthese Faktoren (eIF4A, EF-2 und EF-1 β) und nukleolare Proteine (Nucleolin, hsc70) [5]. Eine Ausnahme dazu bilden die Hostgene der Vertebraten. Hostgene haben, im Gegensatz zu den Introns, keine Funktion im Translationsmechanismus. Sie stehen für snoRNAs als Trägergen zur Verfügung [72]. Die intronische Sequenz ist bei den Hostgenen snoRNAs besser konserviert als die exonische Struktur, da der alleinige funktionelle Teil des Hostgens die snoRNA ist [5]. Beiden genomischen Organisationen intronischer und nicht-intronischer snoRNA ist jedoch ein Schritt in der Prozessierung der snoRNA gemein. Beide Organisationen werden durch die 5' - 3' Exonukleasen Rat1p und Xrn1 prozessiert. Die Spaltung der snoRNA aus den unterschiedlichen Lokalisationen verläuft unterschiedlich (Abbildung: 1.3). Die intronischen snoRNAs oder die snoRNAs, die auf einem Hostgen sitzen, können durch Spleißvorgang oder durch Endonukleasen von der pre-rRNA gespaltet werden (Abbildung: 1.3).

Polycistronische Sequenzen, auftretend bei Pflanzen und bei der Hefe, werden durch die Endonuklease RNase III gespalten [5, 89].

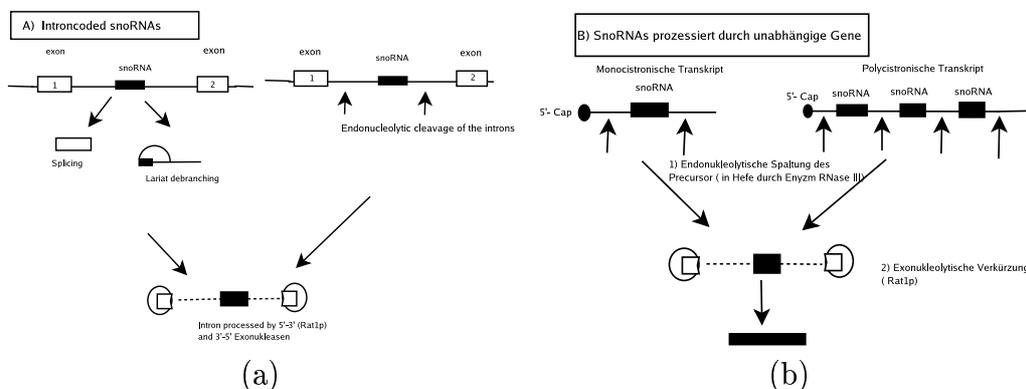


Abbildung 1.3: (a) Prozessierungsschritt der snoRNAs die auf Introns oder Hostgenen liegen. Prozessierung kann mittels einem kleineren und größeren Modifizierungsschritt vollzogen werden. (b) Prozessierungsschritt der polycistronischen snoRNA in den Pflanzen. (Abbildungen nach Bachellerie [2000])

Gehirnspezifische snoRNAs

Neben snoRNAs, die in Introns oder in Hostgenen kodiert sind und in allen Gewebearten auftreten, wurden auch snoRNAs gefunden, die nur oder vorherrschend im zentralen Nervensystem exprimiert werden [79, 12].

Besonders sechs menschliche geirnspezifische snoRNAs, alle auf einem Intron kodierend, fielen in Augenschein, da sie auf dem Bereich einer väterlich imprinted Transkriptionseinheit, IC-SNURF-SNRPN, liegen. Der Bereich der Transkriptionseinheit umfasst insgesamt 150 Exone und umspannt einen Bereich von 460 kp.

Es gilt der Grundsatz: SnoRNAs sitzen in Introns oder in Hostgenen, Introns beheimaten immer nur eine snoRNA (Kap. 1.2.2)[80]. Die Transkriptionseinheit dient als Hostgen für väterlich exprimierte snoRNAs, die auf den Introns der Transkriptionseinheit lokalisiert sind. Dem obigen Grundsatz nicht entsprechend sind allerdings bei der Transkriptionseinheit IC-SNURF-SNRPN 3 snoRNAs vorhanden, die auf einem Exon liegen. Des Weiteren beinhalten vier Introns jeweils zwei snoRNAs [80].

Die Transkriptionseinheit IC-SNURF-SNRPN ist involviert in neuronale Erkrankungen, wie z. B dem Prader-Willi Syndrom (PWS) oder dem Angelman-

Syndrom (AS).

Die väterlich vererbte Löschung von IC-SNURF-SNRPN führt bei der Maus dazu, dass Symptome auftreten, die für PWS charakteristisch sind. Durch die Lage der snoRNAs auf dem Bereich des IC-SNURF-SNRPN stellt sich die Frage, ob snoRNAs beim Prader-Willi-Syndrom eine Rolle spielen können. Als signifikantes Merkmal gegenüber nicht gewebespezifischen snoRNAs zeigen die gehirnspezifischen snoRNAs keine Komplementarität zur rRNA, snRNA auf. Die gehirnspezifische snoRNA MBII-52, weist allerdings eine 18 nt lange Sequenz auf, die komplementär zu einem menschlichen und mäuseichen Rezeptor 5HT 2C mRNA ist [12].

1.2.3 Wie funktionieren snoRNAs?

Die Modifizierung der ribosomalen RNA vollzieht sich in den ersten Spaltungsvorgängen der 35S pre-rRNA (Kap.: 1.3.3)[30]. Da die snoRNA die ribosomale RNA nicht alleine modifizieren kann, tritt sie immer in einem Komplex spezifischer Proteine auf. Box C/D hat vier konservierte Proteine: Fibrillarin (Nop1p), Nop56p, Nop58p und Snu13p. Die Proteine Cbf5p/Dyskerin, Gar1p, Nhp2p und Nop10p assoziieren mit der Box H/ACA snoRNA (Tab.: 1.2) [30, 25, 6].

Die Bindung der ribosomalen RNA findet zwischen komplementären Bereichen der snoRNA und der ribosomalen RNA statt. Die Modifizierung dieser komplementären ribosomalen RNA wird durch jeweils ein bestimmtes Protein katalysiert. Fibrillarin katalysiert als 2'-O-Methyltransferase die 2'-O-Methylierung, Cbf5 katalysiert als Pseudosynthase die Pseudouridylierung. Für die Box C/D snoRNA sind neben Fibrillarin die Proteine Nop56p und Nop58p wichtig. Beide Proteine bilden mit Fibrillarin einen Komplex. Das Protein Snu13p bindet direkt an Box C und Box D und formt dadurch eine Kink-Turn-Struktur (K-Turn) [44, 93, 45]. Unter einem Kink-Turn versteht man eine bestimmte Sekundärstruktur, die in der LSU rRNA, SSU rRNA und in der U4 snRNA zu finden ist. Die Consensusstruktur eines Kink-Turn besteht aus zwei Nukleotidstämmen, die durch einen drei nukleotidlangen einzelsträngigen Bereich getrennt sind (Abb.: 1.2.3) [38]. Der Stamm 1 zeichnet sich durch zwei G-C Basenpaare aus; Stamm 2 hat zwei aufeinander folgende G-A Basenpaare.

Der zweite Nukleotidstamm hat eine wichtige Funktion: Er erzeugt eine Basenpaarung zwischen der Box C und Box D. Nop58p bindet im Bereich der Box C; Fibrillarin bindet im Bereich der Box D [38]. Aufgrund dessen ist das

1.2.4 Modifikation unterschiedlicher nicht kodierender RNAs

SnoRNAs modifizieren entweder die durch RNA-Polymerase I transkribierte 18S, 5.8S, 28S rRNAs oder die durch RNA-Polymerase II und III spezifische snRNA. Die Modifikation der entstandenen Transkripte, rRNA und U6 snRNA, wird durch die Synthese von snoRNP (Tabelle 1.2) vollzogen.

Bei den snRNAs U1, U2, U4, U5 zeigte sich allerdings eine andere Modifikationsweise. Während die Pol III-spezifische U6 snRNAs im Nukleolus durch die snoRNA modifiziert wird, wandern die Pol II-spezifischen snRNA ins Cytoplasma. Im Cytoplasma assoziieren die snRNAs mit dem Sm Protein, wobei eine 3' - End Prozessierung durchlaufen wird. Das neu gebildete snRNP wird wieder in den Nukleolus importiert. Bevor es in den Nukleolus transportiert wird, assoziieren die snRNAs mit einem nukleoplasmatischen Organell, den Cajal Bodies [17, 83]. Cajal Bodies (auch coiled Bodies genannt) kommen in der Nachbarschaft zum Nukleolus vor und sind involviert in der snRNP Biogenese und der Wanderung der snoRNPs und snRNP durch die Cajal Bodies zum Nukleolus [31, 83]. Alle vier Pol II - spezifischen snRNAs assoziieren mit den Cajal Bodies. Es wird angenommen, dass die Modifikation dieser vier snRNA nicht im Nukleolus, sondern in den Cajal Bodies stattfindet [43]. Daher kann angenommen werden, dass sich snoRNAs auch außerhalb des Nukleolus befinden können. Sowie sie auch außerhalb des Nukleolus ihrer Funktion nachgehen können [44].

1.2.5 Evolutionärer Ursprung der snoRNA

Das heutige Leben ist höchstwahrscheinlich aus einer RNA-Welt entstanden, in der RNA-Moleküle die biochemischen Reaktionen katalysierten und die genetische Information inne hatten [69]. Am *E. coli*-Genom wurde festgestellt, dass vier Nukleotide der *E. coli* rRNA durch 2'-O-Methylierung modifiziert wurden sowie 11 Nukleotide durch Pseudouridylierung. Jedoch werden diese Nukleotide nicht durch eine modifizierende RNA, sondern durch spezifische Proteinenzyme modifiziert [18, 36, 66]. TruA und TruB sind Enzyme, die Nukleotide der *E. coli* transfer oder ribosomalen RNA modifizieren [48]

Das prokaryotische Genom *Archaea* besitzt keinen Zellkern. Jedoch finden auch hier DNA Replikation, Transkription und Translation in ähnlicher Weise zu den Eukaryoten wie zu den Bakterien statt. Am prokaryotischen *Archaea*-Genom existieren Homologien zu eukaryotischen Box C/D-Sequenzen

und zum eukaryotischen snoRNP Fibrillarin und Nop56 [68]. Keine Homologien zwischen den Box H/ACA snoRNAs bestehen zwischen den Eukaryoten und Archaea. Die bestehenden Homologien zu eukaryotischen snoRNAs am prokaryotischen *Archaea*-Genom zeigen, dass sich das auf snoRNA basierende Modifizierungssystem vor zwei bis drei Milliarden Jahren im gemeinsamem Vorfahren von *Archaea* und Eukaryoten entwickelt haben kann [47, 44].

Diese zwei unterschiedlichen Modifizierungssysteme in Bakterien und Eukaryoten ließen die Frage aufkommen, wie die Evolution vom proteinbasierenden Modifizierungssystem zum snoRNA basierenden Modifizierungssystem verlaufen ist. Homologievergleiche haben gezeigt, dass das bakterielle Enzym TruB mit Cbf5p verwandt ist. Es wird vermutet, dass ein Vorfahre vom Protein Cbf5p (Tabelle: 1.2) durch eine Duplikation des Genes, das für die Codierung des Enzymes TruB zuständig ist, entstanden ist. Dieses Protein war fähig, Bindungen zwischen einem Vorfahren der Box H/ACA und der pre-rRNA einzugehen. Vermutlich könnte ein Vorfahre der Box H/ACA snoRNA snR30 gewesen sein. Cbf5p könnte mittels dieser snoRNA mit der ribosomalen RNA verbunden worden sein. Diese snoRNA hat nicht die RNA-Modifikation als Aufgabe, sondern nur die Prozessierung der pre-rRNA. Durch Duplikation und Mutationen der Box H/ACA snoRNA kann es mit der Zeit zur Menge der existierenden verschiedenen Box H/ACA snoRNAs gekommen sein [48]. Die Entwicklung der Box C/D snoRNA könnte genauso durch Duplikation und Mutation einer möglichen Box C/D snoRNA, die nicht ihre Aufgabe in der Modifizierung besaß, verlaufen sein [48].

Lafontaine schliesst aus der Tatsache, da snoRNAs nicht in Bakterien vorkommen, dass snoRNAs nicht aus einer RNA-Welt hervorgegangen sind. Es wäre allerdings auch möglich, dass snoRNAs der ursprüngliche Zustand sind und das snoRNA-basierende Modifizierungssystem in den Bakterien verloren gegangen ist. Um eine Entscheidung zu treffen, ist noch zu wenig über die Evolution der nicht kodierenden RNA bekannt.

Ein Beispiel für die zweite Möglichkeit wäre tmRNA. Bakterien, Chloroplasten und die Mitochondrien der *Reclinomonas americana* [46] benutzen die tmRNA zur Termination von unvollständigen Translationsprodukten, wohingegen diese Aufgaben in Eukaryoten von Protein-Enzymen erfüllt werden [62, 42].

1.3 Ribosomale RNA

Viele Zellen enthalten Tausende von Ribosomen. Ribosomen bestehen aus zwei Untereinheiten: einer großen und einer kleinen (Kap.: 1.3.1). Diese Einheiten werden jeweils aus RNA-Molekülen und Proteinen gebildet; der Komplex wird als ribosomale RNA (rRNA) bezeichnet. Ungefähr zwei Drittel eines Ribosoms besteht aus rRNA. Aufgrund der Fülle an Ribosomen in einer Zelle und wegen des Massenanteils der rRNA am Ribosom ist die rRNA der häufigst vorkommende RNA-Typ [11].

Die Ribosomen haben die Aufgabe, während der Proteinbiosynthese die tRNA-Moleküle mit den Codons der mRNA zu verbinden und die Bildung der Peptidbindung zu katalysieren.

Die Synthese der ribosomalen Untereinheiten findet bei den Eukaryoten im Nukleolus statt (Kap.: 1.3.3). Die ribosomalen RNA-Gene werden zuerst aus dem Primärtranskript transkribiert, dann die RNA prozessiert und schließlich mit den aus dem Cytoplasma importierten Proteinen verbunden. Die gebildeten Untereinheiten werden dann wiederum ins Cytoplasma transportiert.

1.3.1 Struktur der Ribosomen

Ribosomen bestehen aus zwei Untereinheiten: einer großen Untereinheit und einer kleinen Untereinheit. Eukaryotische Ribosomen haben zwei Untereinheiten, die 60S und die 40S Untereinheit. Der S-Wert gibt den Sedimentationskoeffizienten an [9]. Die große Untereinheit des eukaryotischen Ribosoms gliedert sich in drei rRNAs (28S; 5,8S und 5S) und in 49 Proteine. Die kleine Untereinheit besitzt nur eine rRNA, die 18S rRNA und hat 33 Proteine.

Bei den Prokaryoten teilt sich die große Untereinheit in 23S und in 5S rRNA auf. Diese prokaryotische Einheit enthält 31 Proteine. Die kleine prokaryotische Untereinheit besitzt nur die 16S rRNA, die 21 Proteine enthält. Abbildung 1.6 gibt einen Überblick über die jeweiligen Untereinheiten bei Prokaryoten und Eukaryoten. Tabellarisch wird dargestellt, dass die Anzahl an RNA-Molekülen größer ist als die Anzahl an Proteinen und die RNA-Moleküle bei den Eukaryoten größer sind und mehr Proteine als die Prokaryoten besitzen (Abbildung: 1.6) [52].

Die Funktion des Ribosoms spiegelt sich in der Struktur wider. Das Ribosom verknüpft die aminosäurebeladene tRNA mit der mRNA. Es besitzt drei Bindungsstellen für die tRNA und eine Bindungsstelle für die mRNA. Die

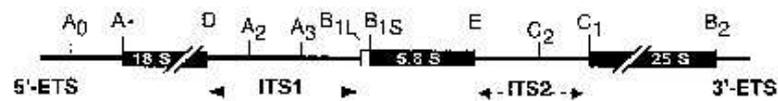


Abbildung 1.5: Genomische Organisation der eukaryotischen ribosomalen RNA. A_0 - E sind Syntheschnittstellen der pre-rRNA. (Abbildung: Fromont-Racine[2003])

Bindungsstellen der tRNA sind die so genannten P-, A- und E-Stellen. Die Bindungsstelle, die die tRNA mit der wachsenden Polypeptidkette verbindet, ist die P-Stelle (Peptidyl-tRNA-Stelle). Die Aminoacyl-Stelle (A-Stelle) nimmt die tRNA auf, die neue Aminosäuren liefert. An der E-Stelle (Austrittsstelle) verlassen ungebundene tRNAs das Ribosom [52].

Das Ribosom hält die tRNA- und mRNA-Moleküle fest beieinander und katalysiert die Bindung der Aminosäure an die freie Carboxylgruppe der Polypeptidkette.

1.3.2 Genomische Lokalisation der rRNA

Das Ribosom enthält je eine Kopie von unterschiedlichen rRNA-Molekülen. Am wirkungsvollsten ist es, wenn jeweils die gleiche Menge der unterschiedlichen rRNAs in der Zelle synthetisiert wird. Allerdings kann auch jeweils eine dem Verbrauch unterschiedliche Menge von rRNA in der Zelle synthetisiert werden. Nach dem Verbrauch der rRNA, die in kleinster Menge vorliegt, würden jedoch die anderen rRNA-Moleküle im Überfluß übrig bleiben.

Damit rRNA nicht verschwenderisch produziert wird, werden rRNAs in gleicher Menge synthetisiert. Dies wird durch eine rRNA-Transkriptionseinheit gewährleistet. Hier liegt je ein Satz von rRNA-Molekülen in einer Einheit vor. Das Produkt der Transkription, das Primärtranskript, ist eine RNA-Vorstufe, die alle rRNAs besitzt, die durch kurze Spacer getrennt sind (Abb: 1.5). Die Spacer werden durch Bearbeitungsgänge entfernt, so dass rRNA freigesetzt wird [9].

RIBOSOM		rRNA	Protein
Bakterien 70S Masse: 2.5×10^6 D 66% RNA 		23S = 2904 Basen 5S = 120 Basen	31
		16S = 1542 Basen	21
Säugetier 80S Masse: 4.2×10^6 D 60% RNA 		28S = 4718 Basen 5.8S = 160 Basen 5S = 120 Basen	49
		18S = 1874 Basen	33

Abbildung 1.6: Tabellarische Zusammenfassung der rRNA Untereinheiten bei den Bakterien und Säugetieren.

1.3.3 Reifung der rRNA

Unter rRNA Reifung versteht man die Modifikation der pre-rRNA zur fertig synthetisierten, getrennten rRNA.

Die Reifung der rRNA und die Assemblierung in die ribosomalen Untereinheiten benötigt insgesamt 170 Proteine. Diese Proteine unterteilen sich in Exon- und Endoribonukleasen, ATP-abhängige RNA-Helikasen und Assemblierungsfaktoren [30]. Außerdem spielen auch die snoRNP eine wichtige Rolle in der rRNA Reifung durch die Methylierung und die Pseudouridylierung der rRNA.

Der erste Schritt, um ein funktionstüchtiges Ribosom zu erhalten, ist die Synthese der pre-rRNA. Die 35S rRNA ist der erste erkennbare Zwischenzustand der rRNA in der Hefe. Diese pre-rRNA besteht aus 5' und 3' "external transcribed" Regionen (ETS) und den Bereichen der 18S rRNA, 5.8S rRNA und 25S rRNA, die durch intergenisch transkribierte Bereiche (ITS) getrennt sind (Abb.: 1.3.3) [30].

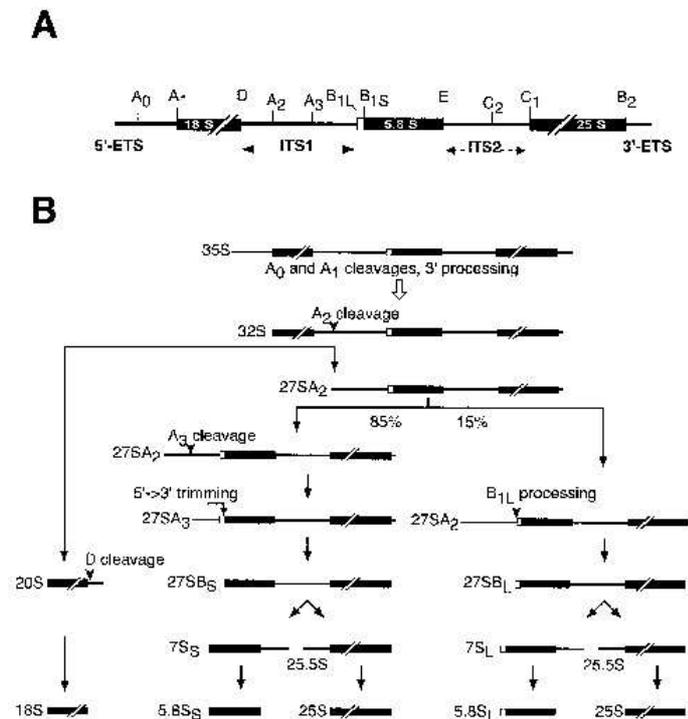


Abbildung 1.7: Synthese der rRNA von der pre-rRNA. (A) Genomische Organisation der pre-rRNA und Angabe der Schnittstellen. (B) Darstellung der Prozessierung von der 35S pre-rRNA zur 5S, 18S und 28S rRNA (Abbildung: Fromont-Racine [2003])

Die rRNA Komponente der kleinen Untereinheit 40S ist die 18S rRNA. Die große Untereinheit (60S) hat als Komponente die 5.8S und 25S rRNA (Abb: 1.6) (Kap. 1.3.1).

Die Spaltung der 35S pre-rRNA an Stelle A_0 im 5'-ETS Bereich führt zur Erstellung der 32S pre-rRNA. Sofern die 35S pre-rRNA am 5'-Ende der 18S rRNA, Position A_1 , gespaltet wird, kommt es zur Ausbildung der 32S pre-rRNA (Abb.: 1.3.3). Nach der Generierung der 32S pre-rRNA wird durch einen Spaltungsvorgang an Stelle A_2 die pre-rRNA 20S und 27SA₂ hergestellt. Die 20S rRNA ist der Vorläufer der fertig synthetisierten 18S rRNA. Die 18S rRNA wird im Cytoplasma durch eine Spaltung an Stelle D prozes-

siert. Die $27SA_2$ kann auf zwei verschiedene Weisen zur 5.S und 25S rRNA gespalten werden. Eine Spaltung an Stelle A_3 durch RNase MRP und eine Kürzung der 5', bzw. 3' Enden zur $27B_{1S}$ kommt zu 85 % vor. Die restlichen 15 % der $27SA_2$ pre-rRNA werden an Stelle B_{1L} gespalten.

Der Prozessierungsweg der $27SB_S$ und $27SB_L$ verläuft daraufhin wieder völlig identisch. Durch eine letzte Spaltung an Stelle C_1 und C_2 sowie Verkürzung der 5', bzw. 3'-Enden kommt es zur Ausbildung der 5.8S rRNA und der 28S rRNA [30, 3]

1.3.4 Warum wird die rRNA modifiziert?

Es stellen sich die Fragen: Warum wird die rRNA modifiziert? Was für einen Nutzen hat die Modifikation der rRNA?

Die Aufgabe der snoRNA ist die Modifizierung von rRNA, snRNA und mRNA durch Pseudouridylierung und Methylierung (Kap.: 1.2). Die zu modifizierenden Nukleotide der ribosomalen RNA werden am häufigsten durch die Methylierung und Isomerisierung des Uridins verändert. Insgesamt je 100 Nukleotide der rRNA des Menschen werden durch die zwei Methoden modifiziert [30].

Es stellt sich daher als erste Frage: Wie funktionieren die Pseudouridylierung und die Methylierung? Die Umwandlung des Uridins in ein Pseudouridin erfolgt durch eine enzymatische Spaltung und anschließend durch eine 180°-Rotation um die Achse des Uracil-Rings (Abbildung: 1.3.4). Durch die Rotation wechselt das Kohlenstoffatom (C_5) seinen Platz mit dem Stickstoffatom (N_1). Das C_5 -Atom geht eine Bindung mit dem Kohlenstoffatom des Glycosyl ein. Um die Bindung eingehen zu können, gibt das Kohlenstoffatom C_5 sein Wasserstoffatom unter der Freisetzung von Wasserstoff ab. Als zweiter Modifizierungsschritt erhält das Stickstoffatom (N_1) noch ein Wasserstoffatom [67]. Bei der Methylierungsreaktion wird eine zusätzliche Methylgruppe durch die S-Adenosylmethion-Methyltransferasen angeheftet (Abbildung: 1.3.4) [67, 48].

Nachdem die erste Frage geklärt ist, wie die rRNA chemisch modifiziert wird, stellt sich als zweite Frage: Warum wird die rRNA pseudouridyliert bzw. methyliert?

Ein wichtiger Aspekt der Pseudouridylierung ist, dass am Stickstoffatom N_1 ein neuer Wasserstoffdonor gebildet wird. Dieser zusätzliche Wasserstoffdonor führt zur erhöhten Stabilisation der rRNA. Durch das Wasserstoffatom ist die rRNA stabiler geworden, um dann intra- bzw. intermolekulare Basen-

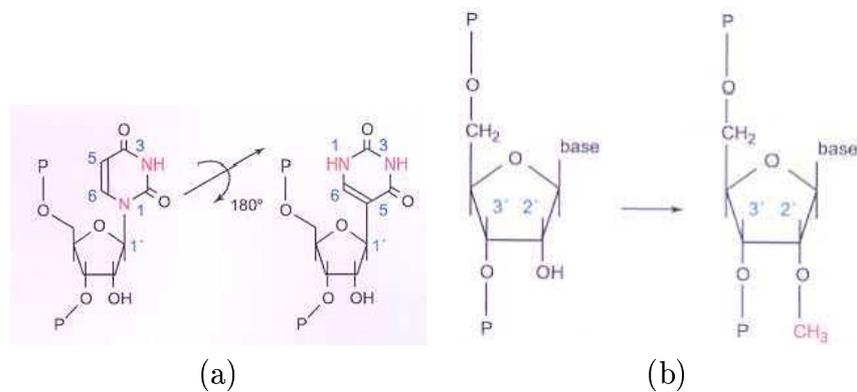


Abbildung 1.8: (a) Strukturformeldarstellung der Transformation des Uridins in Pseudouridin (b) Prozess der Methylierung als Strukturformel dargestellt (Abbildung: Lafontaine [1998])

paarungen eingehen zu können [67].

Die zwei Gründe, warum rRNA Nucleotide modifiziert werden, sind: Zum einen wirkt es sich auf eine erhöhte Stabilisation der rRNA aus und zum anderen wird durch die Modifikation die Ribosomenaktivität bei der Translation verfeinert [30]. Abbildung 1.9 stellt die 16S rRNA der *P.abysssi* mit ihren zu modifizierenden Nucleotiden dar.

1.3.5 Evolution der ribosomalen RNA

Ribosomale RNA Gene kommen in einer hohen Kopienanzahl (100 - 1000 Genkopien) in den Genomen vor. Die jeweiligen rRNA-Moleküle (23S, 5S, 16S) liegen im Genom auf einem einzigen Operon (Kap.: 1.3.2). Liao verglich die Genkopien der 16S rRNA innerhalb einer Organismenart und zwischen 19 verschiedenen bakteriellen und archaeischen Spezies [54]. Durch eine phylogenetische Analyse konnte ermittelt werden, dass die Genkopien innerhalb einer Spezies völlig identisch sind und die Genkopien zwischen verschiedenen Organismen sich sehr gering unterscheiden [54]. Die Homogenität zwischen den einzelnen Spezies ist so groß, dass die Genkopien verschiedener Organismen phylogenetisch näher verwandt sind als intrachromosomale Genkopien [90].

In den intergenischen Bereichen, die die einzelnen rRNA-Moleküle umgeben, kommt es zu Insertionen, Deletionen und Mutationen. Die Distanz zu den

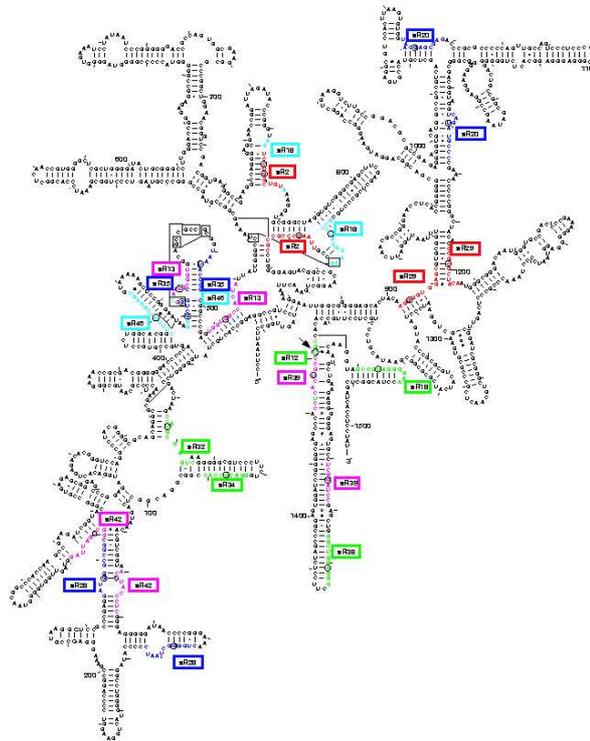


Abbildung 1.9: Sekundärstruktur der *P.abysssi* 16S rRNA mit den zu modifizierenden Nukleotiden. Zur rRNA komplementäre Bereiche wurden farblich markiert, Kreise zeigen das zu modifizierende Nukleotid (Abbildung: Gaspin [2000])

16S rRNA-Molekülen ist ein Indiz für die Heterogenität des Bereiches. Je größer die Distanz zum rRNA-Molekül ist, umso höher ist die Mutations-, Insertions- und Deletionsrate. Der Bereich um die rRNA-Moleküle liegend, ungefähr 150 Basen, zeigt ebenso hohe Homogenität zueinander wie die einzelnen rRNA Kopien [53].

Diese Ergebnisse zeigen, dass rRNA Gene dem Mechanismus der konzentrierten Evolution unterliegen. Von konzentrierter Evolution wird gesprochen, wenn paraloge Sequenzen in tandem repeats angeordnet sind und gemeinsam evolvieren. Ein Kennzeichen der konzentrierten Evolution ist, dass mehrere Duplikate des RNA-Moleküles innerhalb einer Spezies und zwischen verschiedenen Spezies eine sehr geringe Divergenz haben (Abbildung 1.3.5) [90]. Die

Hypothese der “molecular clock “ [96] besagt, dass sich homologe Proteine bzw. DNA-Sequenzen über die Zeit mit einer etwa konstanten Evolutionsrate entwickelt haben. Die Mechanismen der konzentrierten Evolution und Grund für die Homogenität sind *Gene Conversion* (Abb.: 1.3.5) und ungleiches *Crossing Over* [53]. Sich dispergierende Gene werden über *Gene Conversion* homogenisiert. In Abbildung 1.3.5 wird der Mechanismus der *Gene Conversion* detailliert dargestellt. Die Kopien sind im Chromosom in tandem repeats angeordnet. Sofern eine Genkopie mutiert, werden die intraspezifischen Genkopien über intrachromosomale Homogenisierung angeglichen. Die Homogenität zwischen den einzelnen Spezies auf homologen Chromosomen wird über die *Gene Conversion* erreicht. Eine paraloge Genkopie wird homogenisiert; diese Kopie führt dann wieder über intrachromosomale Homogenisierung dazu, dass alle Genkopien gleich sind [53]. Ein weiteres Kenn-

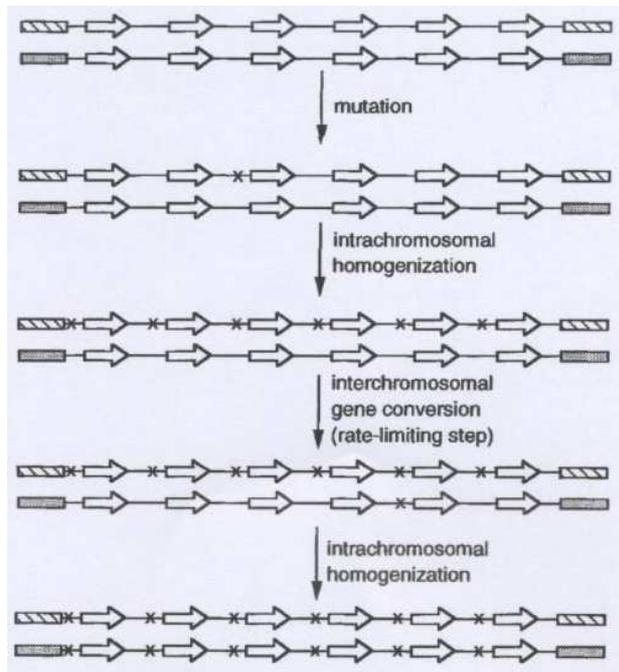


Abbildung 1.10: Ablauf der konzentrierten Evolution eines RNA-Moleküls an zwei unterschiedlichen Organismen. Kodierende Bereiche werden durch Pfeil, nicht kodierende Bereiche durch Linie dargestellt. Mutation wird durch X angegeben. (Abbildung: Liao [1997])

zeichen der konzentrierten Evolution ist die Heterogenität der intergenischen Bereichen. Dies zeigt, dass im Laufe der Evolution nicht ein ganzes Operon, sondern nur bestimmte Bereiche des Operons, die rRNA-Moleküle, kopiert wurden.

1.4 Evolution anderer nicht kodierender RNAs

Dieses Kapitel befasst sich mit zwei weiteren evolutionären Geschichten von RNA-Molekülen, transfer RNA und microRNA. Die Darstellungen der Evolutionen werden zeigen, dass beide RNA-Moleküle auf unterschiedliche Arten evolvieren. Transfer RNA evolviert genauso wie die ribosomale RNA durch konzentrierte Evolution (Kap.: 1.4.1, Kap.: 1.3.5). Die paralogen Kopien der microRNA evolvieren unabhängig voneinander (Kap.: 1.4.2).

1.4.1 Transfer RNA

Der genetische Code hat sich aus einem einfacheren, weniger spezifischen Code entwickelt [15, 23]. Durch die Analyse aller bekannter tRNA-Moleküle wurde festgestellt, dass bestimmte tRNAs, die zum Beispiel als Startzeichen für die Translation dienen, in allen Spezien verwandt sind und sich selbst durch einen phylogenetischen Baum darstellen lassen. Der Baum zeigt, dass selbst die am weitesten entfernten tRNA-Moleküle Ähnlichkeiten aufweisen. Daraus resultiert, dass die frühere tRNA bis heute noch gut erhalten sein muss [24]. Der frühere genetische Code enthielt weniger als 20 Aminosäuren. Er besaß die Eigenschaft, dass die Identität eines Triplet-Nukleotides, später die Identität zweier Nukleotide, zur Spezifizierung der Aminosäuren ausreichte. Die ersten vier Codons waren vermutlich Glycin, Asparginsäure, Alanin und Valin [23, 24]. Alle vier Codons zeichnen sich durch einen hohen GC-Gehalt aus.

Der heutige genetische Code identifiziert die spezifische Aminosäure über drei Nukleotide. Unser heutiger genetischer Code entwickelt sich aus vier Codons, die dem Muster GNC entsprechen. Im Jahre 1981 wurde von Eigen eine evolutionäre Entwicklungsgeschichte des genetischen Codes aufgestellt. Eigen sagt, dass sich aus dem ersten genetischen Code (GNC), ein genetischer Code entwickelte der an Stelle des Cytosins ein purines Nukleotid (Y) hat (GNY). In der weiteren Entwicklung bildete sich aus dem Guanin (erste Nukleotid) eine pyrimidines Nukleotid (R). Damit alle drei Nukleotide ami-

nosäurenspezifisch sind kann an der Position des purinen Nukleotides auch ein pyrimidines Nukleotid stehen und das pyrimidine Nukleotid durch ein purines Nukleotid ersetzt werden.

$GNC \rightarrow GNY \rightarrow RNY \rightarrow RNN \rightarrow NNN$

Entwicklungslinie des genetischen Codes.

Der Wandel von einem zu drei spezifischen Nukleotiden kann nur durch die Entwicklung von verschiedenen anderen transfer RNA (tRNA) erklärt werden [15]. Eigen und Hopfield trafen Aussagen über das erste tRNA-Molekül. Das erste tRNA-Molekül entspricht der 3'-Hälfte der heutigen tRNA (Abb.: 1.4.1). Über eine komplementäre Reproduktion der 3'-Hälfte konnte die 5'-Hälfte ermittelt werden. Die beiden Hälften bilden gemeinsam die Sekundärstruktur der heutigen tRNA, das Kleeblatt. Im Laufe der Evolution der tRNA muss es zu einer Sequenzduplikation der 3'-Hälfte gekommen sein.

Die Evolution der tRNA ist, genauso wie die Evolution der rRNA (Kap.: 1.3.5), ein Beispiel für konzentrierte Evolution. Transfer RNA-Moleküle treten in verschiedenen eukaryotischen Genomen auf zwei verschiedenen Arten auf: Zum einen treten tRNA-Moleküle, die je einen Abstand von 100 - 500 Basen haben, in Clustern auf, zum anderen können tRNA-Moleküle auftreten, die als einzige auf dem Gen lokalisiert sind [55].

Die Ähnlichkeit von tRNA-Kopien zueinander ist auch hier wieder identisch. Die intergenischen Bereiche, die zwischen den tRNA-Kopien liegen, haben eine schlechtere Konservierung zueinander als die tRNA-Kopien.

Es ist ersichtlich, dass auch bei der Evolution der tRNA nicht das ganze Operon auf einmal kopiert wurde, sondern nur einzelne Bereiche. Auch in der Evolution der tRNA-Kopien wurde als Methode die Gene Conversion verwendet (Abb.: 1.3.5).

Des Weiteren konnten einzelne tRNA-Moleküle in nicht verwandte Bereiche über RNA intermediates eingebracht werden. Diese tRNA-Moleküle liegen als Pseudogene, als funktionslose Moleküle, in den einzelnen Genen.

1.4.2 MicroRNA

MicroRNAs (miRNA) haben genauso wie die siRNA die Aufgabe, die Translation zu unterbrechen; sie vollziehen somit eine regulatorische Funktion. Sie

haben gleichfalls wie die snoRNA eine haarnadelähnliche Sekundärstruktur. Jede miRNA ist aus einer 75 nt langen pre-miRNA entstanden. Diese liegt als einzige Kopie im Genom. Ein paar wenige pre-miRNA bestimmter miRNAs können allerdings auch in zwei- oder dreifacher Ausfertigung im Chromosom vorkommen. Des Weiteren können bestimmte miRNA aus zwei oder mehr unterschiedlichen pre-miRNA gebildet werden. MiRNA liegen üblicherweise auf Introns von bekannten Genen oder in intergenischen Bereichen [65].

MiRNA werden in einer Biogenese, die aus zwei Schritten besteht, gebildet [51]. Zuerst wird aus einer *pri*-miRNA, die 75 nt lang ist, eine *pre*-miRNA hergestellt. Die primary miRNA (*pri*-miRNA) wird so gefaltet, dass eine stabile Stamm-Schleifen-Struktur entsteht. Diese Struktur wird vom Enzym Droscha erkannt [26], welches die pre-miRNA aus der *pri*-miRNA spaltet. *Pri*-miRNA liegt im Nukleus vor, *pre*-miRNA wird über ein Exportin-5, einem tRNA Transportfaktor für den weiteren Prozessierungsschritt ins Cytoplasma transportiert.

Im Cytoplasma wird die miRNA aus der *pre*-miRNA mittels DICER, einem RNase III-ähnlichen Enzym gespalten. Dieser Vorgang gleicht dem bei der siRNA. Bei der siRNA wird aus einer dsRNA mittels DICER die doppelsträngige RNA in zwei einzelsträngige RNAs gespalten [65]. Zwei Stränge der siRNA sind stabil und können mit dem RISC-Komplex eine Bindung eingehen. Bei der microRNA geht jedoch bei zwei *pre*-miRNA Strängen nur einer eine Bindung mit dem mRNA-Target ein. Wie bei der siRNA muss auch die miRNA ebenfalls einen Komplex mit dem RISC (RNA induced silencing complex) eingehen, um sich an das mRNA Target binden zu können. Der RISC-Komplex führt an der bindenden Stelle dazu, dass die Translation unterdrückt oder eine neue mRNA hergestellt wird.

Die Evolution der miRNA wurde von Andrea Tanzer anhand des mir17 Cluster untersucht [88]. Dabei wurde erkannt, dass microRNA in Clustern auftreten. Mir17 besteht aus den microRNA mir17/mir18/19a/20/19b [61, 49]. Auf dem *Homo sapiens*-Genom ist das mir17 Cluster auf dem Chromosom 13 [49]. Paraloge von diesem mir17 Cluster konnten auf dem menschlichen X Chromosom (mir106a/18X/20X/19b/92) und dem Chromosom 7 (mir106b/93/25) festgestellt werden. Mir17 ist homolog zu mir106 [61]; mir19b-1 ist homolog zu mir19b-2. Tanzer untersuchte in ihrer Arbeit neun verschiedene Wirbeltiere auf Paralogie zum mir17 Cluster, um ein denkbares Evolutionsszenario der microRNA aufstellen zu können.

Es wurde festgestellt, dass insgesamt drei Genome (Ratte, Maus und Schimpanse) die gleiche genomische Organisation wie das menschliche mir17 Clu-

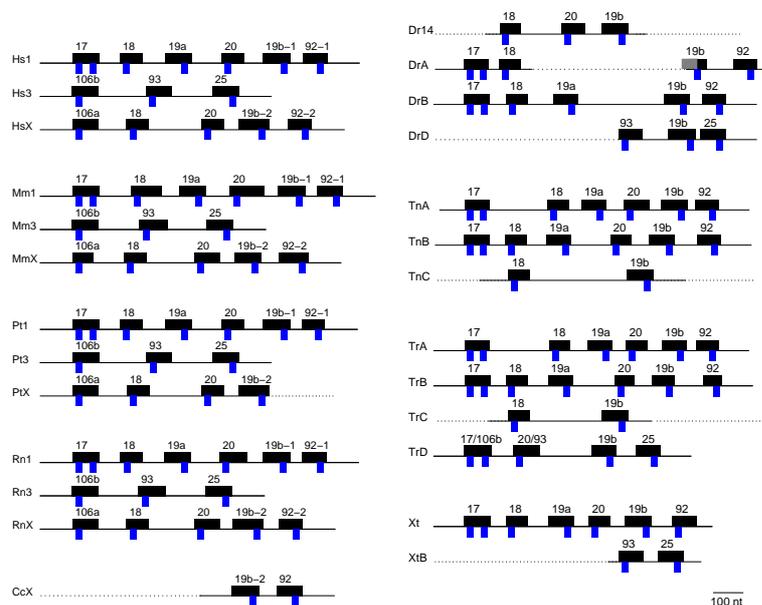


Abbildung 1.12: Mir17 Cluster an verschiedenen Organismen (Abbildung: Tanzer [2004])

ster besitzen. Bei den Fischen (*D. rerio*, *T. rubripes*, *T. nigroviridis*) konnte keine völlige Konservierung erkannt werden, da nur vier Kopien des mir17 Clusters wieder gefunden wurden [88]. Des Weiteren konnten zwei mir17 Cluster im *X. tropicalis*-Genom und ein mir17 Cluster im *C. familiaris* entdeckt werden (Abb.: 1.4.2) [88].

Aufbauend auf diese Ergebnisse wurde von Andrea Tanzer eine phylogenetische Analyse der drei paralogen Cluster an den unterschiedlichen Genomen durchgeführt. Es zeigte sich, dass zwei Typen von Clustern unter den mir17 Clustern existieren. Die beiden mir17 Cluster, die auf den Chromosomen 1 und X lokalisiert sind, waren homolog zueinander (TYP I). Das mir17 Cluster lokalisiert auf Chromosom 7, konnte mit beiden Clustern nicht gut aligniert werden (TYP II). TYP I und TYP II Cluster sind höchstwahrscheinlich durch Cluster- und Tandemduplikationen sowie durch den Verlust von microRNA entstanden [88].

Ein mögliches Evolutionsszenario für das mir17 Cluster ist, dass das Cluster durch eine Anzahl von Duplikationen und Verlusten von anfänglichen Duplikationen entstanden ist (Abbildung: 1.13) [88].

Es wird angenommen, dass mir17/19/92 am Ursprung der Clusterevolution

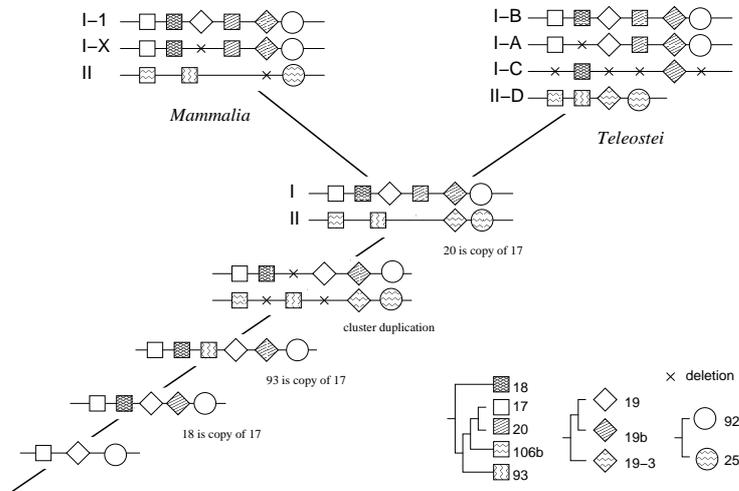


Abbildung 1.13: Evolutionäres Szenario des mir17 Cluster (Abbildung: Tanzer [2004])

steht. Mir17/19 wurden nur in den Vertebraten gefunden, wobei mir92 nur im *C. elegans*, *D. melangoster* auftritt. Der erste wichtige Evolutionsschritt war die Aufspaltung von mir18 aus mir17. Ein weiterer wichtiger Schritt in der Evolution des microRNA Clusters war die Duplikation vom gemeinsamen Vorfahren mir17/18/93/19/19b/92 in die TYP I (mir17/18/19a/20/19b/92) und TYP II (mir106b/93/25) Cluster [88].

1.5 Ziel der Arbeit

Ziel dieser Diplomarbeit ist zum einen die detaillierte Suche nach neuen snoRNAs in den Vertebraten. Zu Beginn der Arbeit waren insgesamt 25 Box C/D und Box H/ACA Teleosten-snoRNAs in öffentlichen Datenbanken publiziert. Aufgrund dessen wurde bei der Suche nach neuen snoRNAs besonderer Augenschein auf die ältesten Vertebratengenome geworfen. Zum anderen ist ein Ziel der Diplomarbeit eine detaillierte Beschreibung der snoRNA-Evolution anhand ausgewählter Vertebraten snoRNAs.

1.6 Vorgehensweise

Um evolutionäre Aussagen treffen zu können, müssen homologe Kopien der snoRNAs an den unterschiedlichen Genomen gefunden werden. Zwei Programme, die für die Suche nach snoRNAs spezialisiert sind, existierten zu Beginn der Diplomarbeit: `Fisher` und `SnoScan`. Ein weiteres Programm `snoGPS` wurde am Ende der Diplomarbeit veröffentlicht. In der Arbeit wurde versucht über das Programm `Fisher` und anschließend über das Programm `SnoScan` neue snoRNAs am Genom des *Takifugu rubripes* zu finden. Als weitere Methode wurde versucht über einen Homologie-Vergleich, Programm `Blast`, neue snoRNAs an unterschiedlichen Genomen zu finden. Die durch die Programme `Fisher`, `SnoScan` und `Blast` gefundenen snoRNAs sind durch Bestimmung der ribosomalen Bereiche, Lokalisation im Genom und Sequenz-homologie verifiziert wurden. Mittels den neu gewonnenen snoRNAs wurden evolutionäre Szenarien für vier unterschiedliche snoRNA-Klassen, basierend auf phylogenetischen Analysen, aufgestellt. Die evolutionäre Analyse zeigt, dass es in der Evolution der snoRNA zu Genduplikationen, -mutationen und zu Sprüngen in der genomischen Lokalisation gekommen ist.

Die Arbeit besteht aus vier Kapiteln. Im ersten Kapitel, das bereits abgehandelt ist, wird in das Thema snoRNA eingeführt und die Evolutionen verschiedener nicht kodierender RNAs vorgestellt. Das zweite Kapitel gibt einen Einblick in die Vorgehensweise der bioinformatischen Programme, mittels denen nach neuen snoRNAs gesucht wurde. Im darauf folgenden Kapitel werden die Ergebnisse der bioinformatischen Programme ausgewertet und die Resultate bzw. Probleme der einzelnen Programme dargestellt. Die Evolution der snoRNA wird im letzten Kapitel anhand vier ausgewählter snoRNAs beschrieben.

Kapitel 2

Bioinformatische Programme

Bei der Suche nach neuen RNAs kann die Standardsuchmethode **BLAST** [2] verwendet werden. Bei einer **BLAST**-Suche wird das Genom komparativ mit bekannten snoRNAs auf Homologie verglichen (Kap. 2.3).

Des Weiteren existiert eine Anzahl von Programmen (**tRNAscan-SE** [56], **QRNA** [75], **SnoScan** (Kap.: 2.2), **Fisher** (Kap.: 2.1), **snoGPS** [82]), die speziell für die Suche von nicht kodierenden RNAs entwickelt worden sind.

Da **SnoScan** und **Fisher** spezielle Programme für die snoRNA-Suche sind, wird in den nächsten Kapiteln näher auf das Suchverfahren der beiden Programme eingegangen.

2.1 Fisher: Suche nach H/ACA snoRNAs

Das von Paul Gardner entwickelte Programm kann über bekannte H/ACA snoRNAs nach neuen H/ACA snoRNAs suchen. Die Suche basiert auf der Primärstruktur. Für die Primärstruktursuche sind die Erkennungsmotive der Box H/ACA sowie bestimmte definierte Distanzen zwischen den jeweiligen Boxen und Pseudouridylationsstellen notwendig [21]. Diese durch die Primärstruktursuche ermittelten Sequenzen werden durch verschiedene Sekundärstrukturfilter verifiziert.

Die Sequenzen, die Ergebnisse bei der Primär- und Sekundärstruktur liefern, werden im letzten Schritt anhand der Daten aus der Primär- und Sekundärstruktur bewertet [21].

Gardner durchsuchte anhand dem Programm **Fisher** das Hefegenom und konnte drei H/ACA snoRNA, sowie weitere 47 Kandidaten ermitteln.

Primärstruktursuche

Zuerst wird das Genom auf die Erkennungsmotive der Box H und der Box ACA untersucht. Die Boxen und Pseudouridylationsstellen treten in definierten Abständen zueinander auf. Diese Distanzen sind ein Indiz für die Primärstruktur der snoRNA. Im Programm werden folgende Distanzen benötigt: Länge von Anfang zur Box H, Länge von Box H zu ψ_3 , Länge von ψ_3 zu ψ_4 , Länge von Box H zu Box ACA (Abb.: 2.1).

Die Sequenz wird zuerst nach konservierten Box H-Elementen durchsucht.

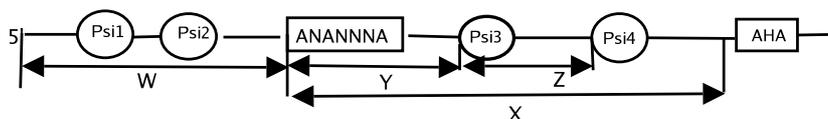


Abbildung 2.1: Distanzen einer *H. sapiens* H/ACA snoRNA. Länge von Anfang zu Box H (W), Länge von Box H zu ψ_3 (Y), Länge von ψ_3 zu ψ_4 (Z), Länge von Box H zu Box ACA (X).

Von diesen vermutlich gefundenen H-Boxen wird ein Score mit Wahrscheinlichkeitsfrequenzen aus den variablen Stellen der Box H ($AN_1AN_2N_3N_4N_5$, wobei N_4 oder $N_5 = A$) errechnet. Die Frequenzen der variablen Positionen wurden am Hefegenom ermittelt. Sofern die Basen der H-Boxen miteinander übereinstimmen, erhalten die Matches einen Score von 200; wenn die Basen nicht übereinstimmen, erhalten sie den aus der Nukleotidmatrix entsprechenden Wert (Tab.: 2.1).

Bei einem Score über 800 wird die H-Box als glaubwürdig eingestuft. Die Se-

%	N_1	N_2	N_3	N_4	N_5
A	12.50	25.00	18.75	81.25	87.50
U	6.25	62.50	43.75	6.25	12.50
C	0	6.25	37.50	0	0
G	81.25	6.25	0	12.50	0

Tabelle 2.1: Häufigkeiten der Nukleotide, die benötigt werden, um die H-Box zu scoren

quenzen, die eine Box mit einem Score über 800 besitzen, werden als nächstes ab der H-Box auf ihre Distanzen zu den ψ -Stellen (ψ_3 und ψ_4) untersucht. Von den ψ -Stellen muss sich dann in bestimmter Distanz die ACA-Box befinden.

Durch diese Vorgehensweise wird die Primärstruktur einer Box H/ACA snoRNA gefunden.

Sekundärstrukturfilter

Die durch die Primärstruktursuche gefundenen SnoRNA-Sequenzen werden im zweiten Schritt auf ihre Richtigkeit untersucht. Die Sequenzen werden mit 100 - 200 Nukleotiden abwärts von der H-Box durch das Subprogramm `RNAfold` des `Vienna RNA Package` [39] gefaltet.

Dies bietet die Möglichkeit, die Gemeinsamkeiten von Primär- und Sekundärstruktur im Mountain Plot darzustellen (Abbildung: 2.2) [41, 60]. Dieser Typ an Sekundärstrukturdarstellung bildet auf der X-Achse die n-Nukleotide der Sequenz ab und auf der Y-Achse die Anzahl der Basenpaarungen, die dieses Nukleotid umschließen. Ein weiterer Vorteil für diese Darstellung ist, um einen Score auszurechnen, dass bei unterschiedlicher Sequenzlänge die Mountain Plots normalisiert werden können.

Der erste Sekundärstrukturfilter beruht auf den Eigenschaften des Mountain Plots. Der entstandene Mountain Plot wird erst dann akzeptiert, wenn sich ein lokales Minimum in der Nähe der H-Box befindet. In Abbildung 2.2 kann man erkennen, dass die Sekundärstruktur aus zwei separaten Mountains besteht. Der erste Mountain stellt den Bereich vor der H-Box dar, der zweite Mountain ist der Bereich zwischen H-Box und ACA-Box.

Das Programm `Fisher` versucht, nicht nur den Bereich der snoRNA zwischen Box H und Box ACA zu finden, sondern auch den Bereich vor der Box H. Der Bereich vor der Box H bildet gleichermaßen wie der Bereich zwischen Box H und ACA eine Schleifenstruktur aus. Im Bereich vor Box H treten auch ribosomale Bindungsstellen auf.

Die Kandidaten, die im ersten Filter akzeptiert wurden, passieren einen weiteren Filter. Der zweite Filter vergleicht den resultierenden normalisierten Mountain Plot mit den alten (v1.3) und neuen (v1.4) Faltparameter [58] des `Vienna Packages`. Die Parameter können als zweiter Filter herangezogen werden, da sich die Sekundärstrukturen der snoRNA bei beiden Parametern sehr ähneln. Durch die zwei unterschiedlichen Faltparameter werden zwei verschiedene Distanzen aus dem normalisierten Mountain Plot errechnet. Die

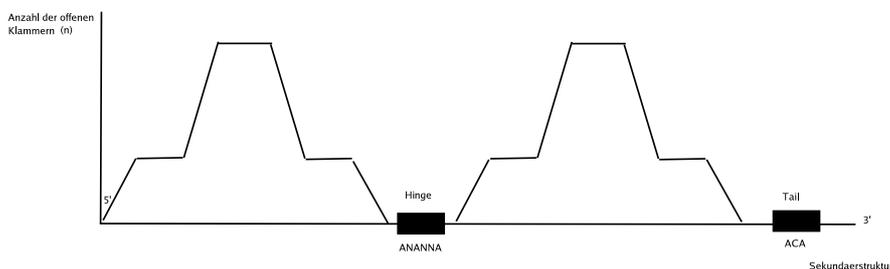


Abbildung 2.2: Sekundärstrukturdarstellung einer H/ACA snoRNA mittels Mountain Plot

Sequenzen werden beim zweiten Filter akzeptiert, wenn sie der Gleichung $d_{old} - d_{new} < 300$ entsprechen.

Scoring

Im letzten Schritt werden die aus der Primär- und Sekundärstruktur erhaltenen Informationen gescort.

Als zweiter Score wird der Score der ACA-Box berechnet. Die ACA-Box kann anstelle des Cytosin auch die nicht motivtypischen Basen Adenine oder Uridin einnehmen. Dieser Score wird zu dem, bei der Primärstrukturfiltersuche errechneten Wert dem H-Box Score, addiert. Die beiden addierten Scores werden normalisiert auf einen Wert zwischen Null und Eins. Zuzüglich werden drei weitere Quantitäten P_2 , P_3 , P_4 dazu addiert.

P_2 wird gleichgesetzt 1, wenn $X < 40$ (Abbildung: 2.1), ansonsten $P_2 = 0,5$. Wenn $66 < X+Y < 100$, dann $P_3 = 1$, sonst $P_3 = 0,5$. Quantität P_4 stellt die durchschnittliche Distanz dar. Hier gilt: Wenn $P_4 = 0$ ansonsten

$$P_4 = \frac{\bar{d} - |d - \bar{d}|}{\bar{d}}$$

Alle vier Quantitäten werden durch folgende Gleichung zu einem Score zusammengerechnet:

$$P_{tot} = 100 \times \left[\frac{w_1 P_1 + w_2 P_2 + w_3 P_3 + w_4 P_4}{w_1 + w_2 + w_3 + w_4} \right]$$

W sind konstante Gewichte, die mittels der Nelder-Mead Methode [64] errechnet wurden.

Sofern der Score größer 70 ist kann die gefundene Sequenz als glaubwürdig angesehen werden. Der Score von 70 ist ein Schwellenwert, der von den Entwicklern gewählt wurde, um die Anzahl an Sequenzen einzugrenzen (persönl. Kommunikation).

2.2 SnoScan: Suche nach C/D snoRNAs

Neben **Fisher** existiert ein weiteres SnoRNA-Suchprogramm, welches für die Suche nach Box C/D snoRNAs spezialisiert wurde. **Fisher**, Kap.: 2.1, und **SnoScan** haben gemeinsam, dass zuerst nach den snoRNA Sequenzmerkmalen gesucht wird. **SnoScan** verwendet den Greedy Such-Algorithmus, um Box C/D snoRNAs zu finden [57]. Der Greedy-Algorithmus analysiert den aktuellen Zustand anhand einer Bewertungsfunktion. Alle möglichen Folgezustände werden errechnet und dann die Möglichkeit ausgewählt, welche den größten Gewinn verspricht.

Für **SnoScan** sind bei der Suche nach Box C/D snoRNAs folgende Merkmale wichtig: Box D, Box C, komplementärer rRNA Sequenzbereich, Box D', Stellen der Methylation (Guide Sequenz) und Terminal Stem [57]. **SnoScan** untersucht die Sequenz in folgender Weise (Abb.: 2.3). Zuerst wird die Box D (A/CUGA) auf der Sequenz identifiziert. Sofern das Pattern der Box D gefunden wurde, wird in 35 - 200 bp aufwärts des Box D Musters nach der höchst gescorten Box C gesucht. Alle möglichen Box C Muster im Abstand von 35 - 200 bp werden durch einen logarithmischen odds Score [7] eingestuft. Da zwischen dem Bereich von Box C und Box D der Bereich liegt, der komplementär zur ribosomalen RNA ist, wird nach komplementären rRNA Sequenzen gesucht. Erst wird nach komplementären Sequenzen gesucht, die 9 bp oder länger sind sowie weniger als vier Mismatches besitzen. In der Nähe des 3'-Endes des rRNA komplementären Bereichs liegt die Box D'. Zuletzt wird nach rRNA-Methylierungsstellen fünf Nukleotide abwärts von Box D oder Box D' gesucht [57].

Um alle falsch Positiven zu eliminieren, werden die Strukturmerkmale nach fest definierten Kennzahlen durchsucht. Alle SnoRNA Kandidaten werden gegen ein SnoRNA Modell aligniert (Tab.: 2.2).

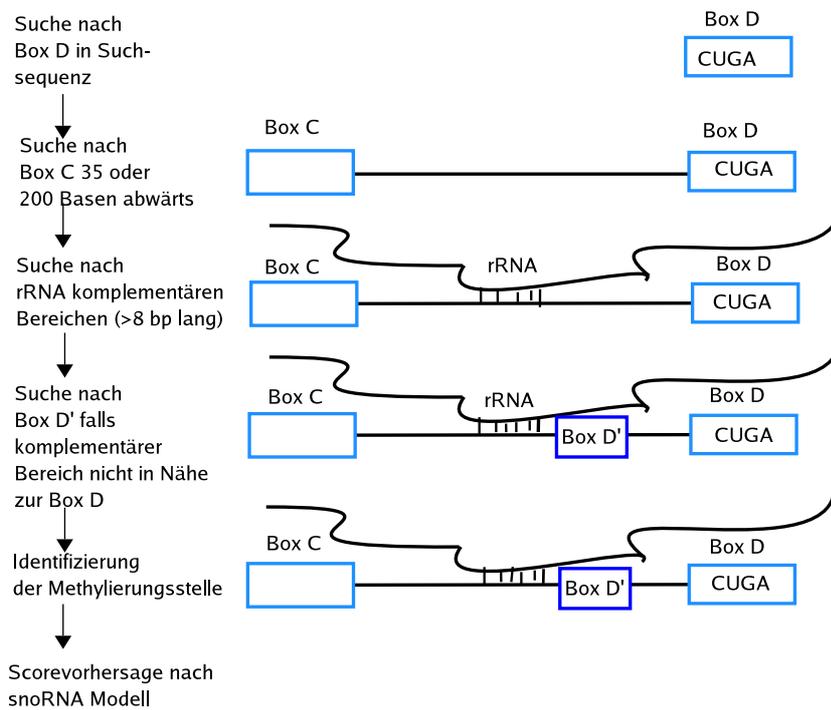


Abbildung 2.3: Suchvorgehensweise des Programm SnoScan.

SnoRNA Modell			
Merkmal	Modell	Consensus	Score (Hoch Mittel Niedrig)
Terminal Stem	SCFG, 4 bis 8bp	6bp	7.60/ 3.09 /0.35
Box C	7 bp	AUGAUGA	12.73/ 11.63/ 5.84
Gap		Länge 6-10bp	-1.59/ -2.09/ -4.76
Guide Sequenz	HMM	12 bp Duplex	15.67/ 11.11/ 2.54
Box D'	4 bp	CUGA	7.34/ 4.85/ -3.74
Gap		Länge 36-45 bp	-1.59/ -2.43/ -5.36
Box D	4bp	CUGA	8.05/ 7.92/ 5.43
Gap		Länge 56-75bp	-1.5/ -2.1/ -4.17
Guide Sequenz	HMM	14bp Duplex	18.96/ 13.98/ 9.95

Tabelle 2.2: Zusammenfassung des snoRNA Modells verwendet bei SnoScan [57].

Alle durch oben beschriebene Primärstruktursuche ermittelten snoRNAs werden zum Schluss anhand einem log odds-Score [7], der alle Informationen der SnoRNA-Merkmale vereinigt, gescort.

2.3 Blast: Homologie-Vergleiche

Als dritte Methode zur snoRNA-Suche soll die Standardmethode zur Suche von nicht kodierenden RNAs dargestellt werden [2]. Mittels dieser Methode wurden kleine regulatorische RNAs an Vertebraten und Nichtvertebraten gefunden [71].

Zur Suche von neuen snoRNAs wurden bekannte publizierte menschliche snoRNA gegen die Nukleotid-Datenbank von NCBI geblastet.

Gefundene Blast-Hits, die in Homologie zu *Danio rerio*, *Takifugu rubripes*, *Cyprinus carpio*, *Gallus gallus*, *Xenopus tropicalis* standen wurden in einem zweiten BLAST gegen das Fugugenom geblastet. Diese gefundenen homologen Fugusequenzen wurden wiederum gegen die NCBI Nukleotiddatenbank [2] auf Ähnlichkeiten zu bekannten snoRNA, snoRNP und ribosomalen Proteinen untersucht.

Um die Sequenzen zu verifizieren, wurden folgende Schritte angewendet: Als erstes wurden die Sequenzen, die im Homologievergleich gegen die Nukleotid-datenbanken Ähnlichkeiten zu snoRNAs hatten, im Fugu Ensembl Browser [84] auf ihre Lokalisation im Genom untersucht. Alle Sequenzen, die im Fu-

gengenom auf einem ribosomalen Protein oder auf einem Gen liegen, das für die rRNA- oder mRNA-Prozessierung zuständig ist, wurden durch weitere Schritte untersucht.

Durch ein Alignment aller snoRNA Familienmitglieder samt der Fugu snoRNA wurde die Konservierung der snoRNA ermittelt. In diesem Schritt ist besonders wichtig, dass die spezifischen snoRNA-Motive (Box C/D oder Box H/ACA) auf gleicher Nukleotidposition konserviert sind. Neben den spezifischen Motiven müssen die Sequenzen Bindungstargets für die jeweilige rRNA-Untereinheit an gleicher Nukleotidposition haben. Diese rRNA Bindungstargets müssen im Alignment konserviert sein.

Sofern die Sequenz die Kriterien erfüllte, wurde sie als snoRNA eingestuft. Eine Zusammenfassung aller Schritte dieser Methode wird in Tabelle 2.3 dargestellt.

Als zweiter Homologievergleich wurden anstelle der *Homo Sapiens* Sequenzen bekannte Fisch snoRNA Sequenzen gegen die *Orzias latipes*, *Takifugu rubripes*, *Tetradon nirogordis* und *Danio rerio* Genome geblastet. Hier wurden die Schritte 1 - 3 aus Tabelle 2.3 zu einem Schritt zusammengefasst und danach sofort mit der Verifikation der Sequenzen begonnen.

Homologieverfahren	
Schrittnr.	Methode
1	Blast bekannter <i>H. sapiens</i> -Sequenzen gegen Nukleotiddatenbank
2	2. Blast von Sequenzen aus fünf bestimmten Vertebraten gegen <i>T. rubripes</i>
3	3. Blast Fugu-Sequenzen werden mittels Nukleotiddatenbank verglichen
4	Lokalisation der Fugu-Sequenz mittels dem Ensembl Fugu Genome Browser
5	Vergleich der Consensusstruktur bekannter snoRNA des Types mit gefundener snoRNA
6	Alignen der Sequenz gegen bekannte snoRNAs des Types mittels clustalw

Tabelle 2.3: Zusammenfassung der Schritte, die bei der Homologiemethode angewendet werden

2.4 Das Vienna RNA Package

Das Vienna RNA Package bietet die Möglichkeit, RNA Sekundärstrukturen über die kleinste Energie vorherzusagen. Das Paket beinhaltet drei Arten von dynamischen Strukturvorhersage-Algorithmen:

1. Algorithmus der kleinsten frei werdenden Energie von Zuker und Stiegler, der eine einzige optimale Struktur vorhersagt [97]
2. Partition Function Algorithmus von McCaskill, der die Basenpaarwahrscheinlichkeiten errechnet [59].
3. Suboptimal Folding Algorithmus von Wuchty, der alle suboptimalen Strukturen in einem bestimmten Energierahmen vorhersagt [94].

RNAfold

RNAfold ist ein Subprogramm aus dem Vienna RNA Package [39]. Dieses Programm liest RNA Sequenzen aus dem `stdin` ein und berechnet ihre mfe Struktur, Partition Function und die Basenpaarwahrscheinlichkeitsmatrix [59]. Die mfe Struktur wird in der Bracket Notation in den Standardoutput zurückgeschrieben. Des Weiteren wird im Standardoutput die Energie, die freie Energie des thermodynamischen Ensembles und die Häufigkeit der mfe Struktur im Ensemble erhalten. Zudem wird ein Ausgabefile im Postscript-Format erzeugt, das zum einen die resultierende Sekundärstruktur und zum anderen einen Dotplot der Basenpaarwahrscheinlichkeiten skizziert. Der Dotplot wird als quadratische Matrix dargestellt (Tab.: 2.4). In der unteren Hälfte stehen die proportionalen Paarungswahrscheinlichkeiten und in der oberen Hälfte die Paarungen in der mfe Struktur. Das Output von RNAfold wird im Programm `alidot` als Input benötigt [39, 40].

RNAalifold

RNAalifold kann die Consensus Sekundärstruktur einer Menge alignierter RNA oder DNA Sequenzen vorhersagen. Diese Methode gelingt durch die Kombination eines Standard-Sekundärstruktur-Energiemodelles mit dem Kovarianzterm.

Jeder möglichen Sekundärstruktur wird ein Mean-Energiewert zugewiesen, gemessen am Durchschnitt aller Alignmentsequenzen. Zuzüglich wird zur

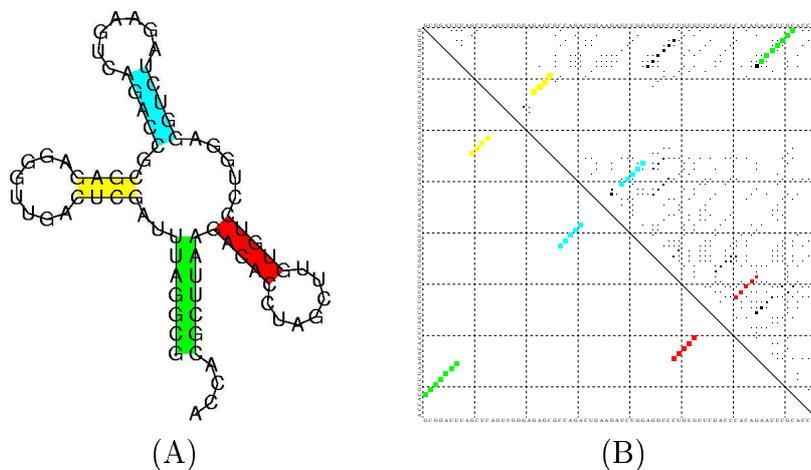


Abbildung 2.4: Strukturdarstellung einer tRNA, (A) tRNA als Sekundärstruktur dargestellt, (B) tRNA als Dotplot dargestellt.

Mean-Energie der Sequenz Bonuswerte für kompensatorische und konsistente Mutationen dazugerechnet und Bonuspunkte für Basenpaarungen, die bei allen anderen Sequenzen nicht auftreten, abgezogen.

2.5 Methoden der phylogenetischen Analyse

2.5.1 Neighbor-Joining

Neighbor Joining wendet als algorithmische Approximation das Minimum-Evolution-Prinzip an [81]. Die Ausgangsdaten der Analyse liegen als Distanzmatrix, der Anfangsbaum liegt als Sternbaum vor. Zuerst wird die Gesamtlänge für jedes mögliche Paar S_{ij} ausgerechnet. Das Paar S_{ij} mit dem minimalsten Wert wird als erste Verzweigung in den phylogenetischen Baum übernommen. Nach der Bildung der ersten Gruppe wird die neue Distanzmatrix, mit einem Taxon weniger, neu durch Mittelwertbildung errechnet. In jedem weiteren Schritt werden zwei Verzweigungen zu einer Verzweigung zusammengeführt, solange, bis zwei Verzweigungen übrigbleiben, die durch einen gemeinsamen Ast getrennt werden. Als Ausgangsbaum erhält man im Phylogenie-Programm `Phylip` [27] bei der Anwendung des Algorithmus einen ungerichteten Baum.

2.5.2 Bootstrap

Das Bootstrapverfahren analysiert die Stabilität der Stammbaumverzweigungen. Die Bootstrapwerte sind ein Indiz für die Vertrauenswürdigkeit der Verzweigungen [22]. Teile der Datenmatrix werden zufällig herausgeschnitten oder kopiert und wieder eingefügt, um einen veränderten Datensatz gleichen Umfangs zu erhalten. Einige Stichproben aus der Anfangsdatenmenge werden nicht mehr vorkommen, andere dafür aber mehrfach [27].

Felsenstein integrierte das Bootstrapping-Verfahren in das Programm `Phylip` [27, 28]. Über das Programm `seqboot` wird eine neue Datenmenge durch das zufällige Ziehen aus der Originalmenge gewonnen. Das Programm `consense` ermittelt aus allen errechneten Bäumen, die über den Neighbor Joining-Algorithmus ermittelt wurden, einen Consensusbaum. `Consense` errechnet den Baum nach dem Majoritätsprinzip und gibt an jeder Verzweigung den entsprechenden Bootstrapwert an. Dieser Wert ist proportional zum Vorkommen des Teilbaumes in allen ermittelten Bäumen [27]. Bei hohen Bootstrapwerten über 75% kann die Verzweigung als ausreichend gesichert gesehen werden. Verzweigung mit einem niedrigerem Bootstrapwert müssen jedoch nicht unbedingt falsch sein.

Kapitel 3

Neue snoRNAs in den Vertebraten

Bis zum heutigen Stand sind in der Rfam-Datenbank [35] insgesamt 25 snoRNAs auf unterschiedlichen Teleostengenomen publiziert. Diese snoRNAs wurden am *Danio rerio*, *Takifugu rubripes*, *Oncorhynchus mykiss* und *Cyprinus carpio* gefunden (Tab.: 3.1).

Durch Homologievergleiche der publizierten Fischsequenzen gegen die Geno-

Publizierte snoRNAs			Gefundene snoRNAs			
Familie	Vorkommen	Anzahl	Familie	Vorkommen	Anzahl	
U14	<i>Danio rerio</i>	4	U14	<i>Tetradon nirgovirdis</i>	8	
	<i>Onchorhynchus mykiss</i>	1		<i>Takifugu rubripes</i>	2	
	U15	<i>Takifugu rubripes</i>	1	U17	<i>Orzyias latipes</i>	1
U17	<i>Onchorhynchus mykiss</i>	1	<i>Tetradon nirgovirdis</i>		4	
	<i>Takifugu rubripes</i>	4	<i>Danio rerio</i>		4	
	<i>Danio rerio</i>	4	E2		<i>Takifugu rubripes</i>	1
U23	<i>Onchorhynchus mykiss</i>	1	E3	<i>Danio rerio</i>	2	
				<i>Cyprinus carpio</i>	2	<i>Takifugu rubripes</i>
	U24	<i>Takifugu rubripes</i>	1	U23	<i>Tetradon nirgovirdis</i>	6
					<i>Danio rerio</i>	3
U36a	<i>Takifugu rubripes</i>	1	U24	<i>Tetradon nirgovirdis</i>	1	
U36b	<i>Takifugu rubripes</i>	1	U36a	<i>Tetradon nirgovirdis</i>	1	
			U2	<i>Orzyias latipes</i>	1	
U4	<i>Danio rerio</i>	1	U36b	<i>Tetradon nirgovirdis</i>	1	
U70	<i>Danio rerio</i>	1		<i>Orzyias latipes</i>	1	
U8	<i>Danio rerio</i>	2		<i>Tetradon nirgovirdis</i>	4	
				<i>Takifugu rubripes</i>	6	

Tabelle 3.1: Auflistung aller publizierten und gefundenen snoRNAs an unterschiedlichen Fischen.

me *Tetradon nirgovirdis*, *Orzyias latipes*, *Takifugu rubripes* und *Danio rerio* konnten insgesamt 52 weitere snoRNAs gefunden werden. Die gefundenen snoRNAs verteilen sich wie folgt auf die Familien U17, E2, E3, U23, U24, U36a/b, U2 und U14 (Tab: 3.1).

3.1 Ergebnisse der Fisher snoRNA Suche

Im Programm `Fisher` [21] (Kap.: 2.1) werden snoRNA des Types H/ACA mittels bekannten Mustern gesucht. Die Muster, auf denen die Suche nach H/ACA snoRNA basiert, sind bekannte Box H-Muster sowie bekannte Sequenzmuster der ψ -Bereiche. Die bestimmten Erkennungsmerkmale einer H/ACA snoRNA treten innerhalb definierter Distanzen auf. Um snoRNAs zu finden, müssen Pattern von ribosomalen Targets und die organismenspezifischen Distanzen bekannt sein. Da nicht genügend *Takifugu rubripes* H/ACA Sequenzen bekannt sind, um aus den Sequenzen die Distanzen als auch die ribosomalen Bindungsstellen herauszufinden, wurde das Programm anhand einer Trainingsmenge bekannter menschlicher H/ACA-Sequenzen gegen das *X. tropicalis*-Genom getestet. Durch die resultierenden *X. tropicalis* H/ACA-Sequenzen sollte das Fugu-Genom untersucht werden.

3.1.1 Analyse der ψ -Bereiche und Distanzen

Die Erkennung von H/ACA snoRNAs am *X. tropicalis*-Genom wurde mit Targetsequenzen folgender bekannter *H. sapiens* snoRNA-Sequenzen vollzogen. Die ψ -Bereiche wurden durch einen Homologievergleich der reverse komplementären humanen rRNA-Sequenzen (3'-5') gegen die humanen snoRNA-Sequenzen (5'-3') (Tab.: 3.2) ermittelt. Die gemeinsamen Homologien zwischen den Sequenzen müssen sechs bis zehn Nukleotide groß sein, um annehmen zu können, dass es sich um eine Bindungsstelle handelt. Von drei humanen Sequenzen konnten alle vier Pseudouridylationsstellen gefunden werden (Tab.: 3.3). Diese drei Sequenzen dienen als Pattern für die Suche nach *X. tropicalis* Sequenzen.

Nachdem die ribosomalen Targets feststanden, wurden die Sequenzen auf die Distanz zur Box H, die Distanz zwischen Box H und ψ_3 , Distanz zwischen ψ_3 zu ψ_4 und der Gesamtlänge von Box H zu Box ACA untersucht. Durch einen Primärstrukturvergleich aller Sequenzen (Tab.: 3.2) wurden je die minimale

Accession Nummer	Beschreibung
Y11164	H.sapiens U70 small nucleolar RNA gene
Y11163	H.sapiens U69 small nucleolar RNA gene
Y11162	H.sapiens U68 small nucleolar RNA gene
Y11161	H.sapiens U67 small nucleolar RNA gene
Y11160	H.sapiens U66 small nucleolar RNA gene
AY349599	Homo Sapiens HACA box snoRNA U98b, complete sequence
Y11167	H.sapiens U72 small nucleolar RNA gene
AF492209	Homo Sapiens U93 scaRNA gene, complete sequence
AY349598	Homo Sapiens HACA box snoRNA U98a, complete sequence
AB070559	Homo sapiens RPL30 and U72 genes for ribosomal protein L30 and U72 snoRNA, complete cds and sequence
AB061842	Homo sapiens RPS20, U54 genes for ribosomal protein S20 and U54 snoRNA, complete cds and sequence
AB061835	Homo sapiens RPL39, U69 genes for ribosomal protein L39 and U69 snoRNA, complete cds and sequence
AF375472	Homo sapiens small nucleolar RNA M2, complete sequence and LAMR1 gene, intron2
M14061	Human U3 small nuclear RNA gene
L07384	Homo sapiens E3 small nucleolar RNA
L07383	Homo sapiens E2 small nucleolar RNA
L16791	Human complete U17A small nuclear RNA (U17A) gene
X94290	H.sapiens small nuclear RNA
S63405	E1 small nucleolar RNA
AY349600	Homo sapiens HACA box snoRNA U99 gene, complete sequence
AJ007015	Human sapiens U23 gene
L16792	Human complete U17B small nuclear RNA (U17B) gene

Tabelle 3.2: *H. sapiens* HACA snoRNA

Sequenzname	Box H	ψ_1	ψ_2	ψ_3	ψ_4
Y11164	ACAGCA	CUGUGUG	CAGCCAU	AACCCAU	GUCCC
Y11162	ACAGCA	CCUAA	GAGGAGGA	CAGCCCCC	AGGGCC
Y11161	AUAGAA	CCUCU	UAGCA	GCUUAUU	UUCCUU

Tabelle 3.3: Sequenzen, deren ψ_{1-4} Bereiche Pattern für Programm Fisher sind.

als auch die maximale Distanz ermittelt. Durch den Primärstrukturvergleich ergab sich, dass die minimale Distanz von Anfang der Sequenz zu Box H 47 Basen beträgt, die maximale Distanz lag bei 168 Basen. Die Distanz von Box H zu ψ_3 bei den menschlichen Sequenzen hat als maximale Distanz 63 Basen. Die Länge des Abstandes von Pseudouridylationsstelle ψ_3 zu ψ_4 hatte als größten Abstand 12 Basen. Die Gesamtlänge der Sequenzen von Box H zu Box ACA lag in einem Intervall von 24 bis zu 125 Basen.

Beim Vergleich der minimalen und maximalen Distanz vom *H. sapiens* und vom *S. cerevisiae*-Genom ist ersichtlich, dass die Distanzen sich stark ähneln. Die Distanzen zwischen ψ_3 und ψ_4 variieren innerhalb der Genome. Die minimale Distanz von ψ_3 zu ψ_4 am Hefegenom ist größer als beim Menschen-genom.

Sequenzname	W	X	Y	Z
Y11164	62	60	34	2
Y11162	60	60	21	8
Y11161	61	60	8	-
Y11160	60	60	18	7
AY349599	47	74	14	-
Y11167	58	62	34	2
AY349598	58	58	0	3
M14061	84	-	63	-
L07384	59	63	0	3
L07383	73	66	46	-
22L16791	109	82	-	-
X94290	60	125	-	-
S63450	110	80	52	-
AY349600	58	78	0	12
AY007015	61	63	14	-
L16792	168	24	-	-
L07382	-	86	-	-
Minimale Distanz	47	24	0	2
Minimale Distanz (Hefe)	-	-	0	30
Maximale Distanz	168	125	63	12
Maximale Distanz (Hefe)	-	135	53	-

Tabelle 3.4: Distanzen aller *H. sapiens* H/ACA snoRNAs sowie maximale und minimale Distanzen der *H. sapiens* und *S. cerevisiae* H/ACA snoRNA. W: Länge vom Anfang zur Box H, X: Distanz von Box H zu ACA, Y: Länge von Box H zu ψ_3 , Z: Länge von ψ_3 zu ψ_4 . Minimale Distanzen von W und Z werden im Programm nicht benötigt.

Vergleich der Nukleotidwahrscheinlichkeiten

Die Nukleotidwahrscheinlichkeiten der einzelnen variablen H-Box Positionen wurden mittels den im Homologie-Blast gefundenen Vertebraten-Sequenzen (siehe Anhang A) errechnet.

Beim Vergleich der Nukleotidwahrscheinlichkeiten der *S. cerevisiae* (Tabel-

%	N ₁	N ₂	N ₃	N ₄	N ₅
A	34.00	23.00	26.00	79.00	87.10
U	8.00	39.00	21.00	7.00	4.30
C	13.00	13.00	24.00	0	2.90
G	45.00	25.00	29.00	14.00	5.70

Tabelle 3.5: Ermittelte Nukleotidhäufigkeiten in den variablen Positionen der Vertebraten Box H

le: 3.6), benutzt im Programm **Fisher**, und den Wahrscheinlichkeiten der

Vertebraten (Tabelle: 3.5) fällt auf, dass besonders die variablen Positionen N_1 , N_2 und N_3 sich prozentual unterscheiden. Es ist zu erkennen, dass jede variable Position der *S. cerevisiae* ein Nukleotid besitzt, das hauptsächlich konserviert ist. Die variablen Positionen der Vertebratensequenzen (N_2, N_3) liegen fast gleichverteilt vor. Position N_4 und N_5 ähneln sich prozentual an den Genomen der Hefe und der Vertebraten.

%	N_1	N_2	N_3	N_4	N_5
A	12.50	25.00	18.75	81.25	87.50
U	6.25	62.50	43.75	6.25	12.50
C	0	6.25	37.50	0	0
G	81.25	6.25	0	12.50	0

Tabelle 3.6: Häufigkeiten der Nukleotide in den variablen Positionen der *S. cerevisiae* Box H.

3.1.2 Resultate des Testlaufs

Das Programm `Fisher` wurde mit den ermittelten minimalen und maximalen humanen Distanzen (Tab.: 3.4) und humanen ψ_{1-4} Targets (Tab.: 3.3) implementiert und gegen das *X. tropicalis*-Genom getestet.

Aus dem Testlauf resultierten für die drei Mustersequenzen ungefähr 4500 Übereinstimmungen.

Für die Datenauswertung sind Übereinstimmungen mit einem totalen Score bis zu 75 % verwendet worden. Aus der ersten Mustersequenz ergaben sich 30 Übereinstimmungen, aus der zweiten 90 und aus der dritten weitere 90 Übereinstimmungen. Diese 210 möglichen snoRNAs wurden mittels NCBI-Blast [2] auf Homologie zu ribosomalen Proteinen überprüft.

Durch diesen Homologievergleich ergaben sich aus den drei Mustersequenzen 20 snoRNAs, die Ähnlichkeiten zu einem ribosomalen Protein im Bereich von Box H zu Box ACA haben. Die Strukturen dieser 20 snoRNA (siehe Anhang A) zeigen alle Ähnlichkeiten zu einer Haarnadel, wobei Box H und Box ACA weitgehend im einzelsträngigen Bereich liegen (Tab.: 3.7). Die gefundenen snoRNAs haben Ähnlichkeiten zu folgenden ribosomalen Proteinen oder snRNAs: RPS1, RPS8, RPL14, U5 snRNP specific protein, U7 snRNA, ELaB (ribonucleoprotein) (Anhang A).

Die 20 möglichen snoRNAs wurden des Weiteren gegen bekannte menschliche reverse komplementäre rRNAs geblastet, um die ribosomalen Bindungsstellen in der Sequenz zu finden. Hier konnten bei allen Sequenzen die ribosomalen Bindungsstellen zwischen Box H und Box ACA (ψ_3 und ψ_4) gefunden werden. Die ribosomalen Targets vor der Box H (ψ_1 und ψ_2) konnten nur vereinzelt entdeckt werden (Tab.: 3.7).

Box H	Box ACA	ψ_{3-4}	ψ_{1-2}
12/20	17/20	20/20	14/20

Tabelle 3.7: Statistik über die Anzahl der einzelsträngigen Box H und ACA-Motive und die Anzahl der gefundenen ψ_{1-4} -Bereiche an den 20 möglich gefundenen snoRNA-Sequenzen

3.1.3 Probleme

Die Auswertung der Ergebnisse mittels des **Vienna RNA-Package** [39] zeigte, dass die spezifischen Boxen H und ACA nur vereinzelt auf einzelsträngigen Bereichen liegen. Die ribosomalen Bindungsstellen wurden weitgehend im Abstand von 15 Basen zur Box H bzw. Box ACA gefunden. Jedoch konnte mittels eines Homologie-Blastes gegen die reverse komplementären rRNAs nicht fest definiert werden, an welcher Position die ribosomalen Bindungsstellen genau liegen. Durch den Blast wurden 10 nt lange Bereiche gefunden die komplementär zur reversen rRNA waren. Nur durch experimentelle Versuche könnte genau bestimmt werden, an welche 4 - 8 Basen die rRNA binden kann.

Auch die Entwickler des Programmes berichteten über ihre Probleme bei der Verifikation der gefundenen *S. cerevisiae* snoRNAs. Die von ihnen gefundenen snoRNA konnten nur durch ein weiteres snoRNA Suchprogramm **snoGPS** von Peter Schattner genau verifiziert werden (persön. Mitteilung). **SnoGPS** sucht im ersten Schritt nach neuen snoRNAs über einen deterministischen Suchalgorithmus. Die gewonnenen Sequenzen werden im zweiten Schritt durch einen Vergleich der snoRNA-Merkmale gegenüber bekannten snoRNAs bewertet. Durch die Analyse von Schattner konnte gesichert werden, dass nur drei von den 47 Testkandidaten des **Fisher** mögliche snoRNAs sind [82].

Des Weiteren ist die lange Suchlaufzeit (Tabelle: 3.8) eine Schwachstelle des

Programmes **Fishers**. Das Programm brauchte zwei Tage um das *Xenopus tropicalis*-Genom auf snoRNAs zu untersuchen. Die lange Laufzeit vom Programm **Fisher** liegt daran, dass es zuerst die Boxen H und ACA sucht und alle gefundenen Sequenzen, die diese Targetsequenzen enthalten, mittels dem **Vienna RNA Package** faltet. Erst in späteren Filterschritten werden die Haarnadelstrukturen auf mögliche Pseudouridylierungsstellen untersucht und die Anzahl an positiven snoRNA Sequenzen minimiert. Durch das Falten aller Sequenzen benötigt das Programm viel Zeit. In Tabelle 3.8 wurde eine Targetsequenz (Tabelle: 3.3) unter dem Aspekt der Suchlaufzeit gegen das *X. tropicalis*-Genom getestet. Insgesamt 102232 mögliche Sequenzen wurden gefunden, die alle gefaltet wurden. 95% der Programmlaufzeit wird alleine durch das Falten in Anspruch genommen. `Fill_array` und `LoopEnergy` sind Funktionen, die für die Faltung der Sequenzen zuständig sind.

Da zum einen die Verifikation der von mir gefundenen Daten nur durch ein

% Zeit	Sekunden	Aufrufe	Funktion
57.23	2069.62	-	<code>fill_arrays</code>
38.63	1387.32	-	<code>LoopEnergy</code>
1.31	47.53	-	<code>get_line</code>
0.84	30.44	-	<code>get_array1</code>
0.59	21.51	-	<code>check_symmetry</code>
0.55	19.78	-	<code>HairpinE</code>
0.33	12.02	-	<code>make_ptypes</code>
0.18	6.59	102232	<code>CheckHtoAHA</code>
0.15	5.45	-	<code>space</code>
0.11	3.89	43470336	<code>wobbler4</code>
0.10	3.75	-	<code>ignore_comment</code>
0.07	2.70	-	<code>rd_int22</code>
0.04	1.47	6440616	<code>wobbler3</code>
0.04	1.37	11148	<code>local_min_low</code>
0.03	1.08	-	<code>backtrack</code>
0.02	0.90	-	<code>gettype</code>
0.00	0.14	-	<code>main</code>

Tabelle 3.8: Tabellarische Zusammenfassung der prozentual benötigten Zeit, Anzahl von Sekunden und Aufrufen der einzelnen Funktionen im Programm **Fisher**. Funktionen wurden sortiert nach der prozentualen Zeit. Daten wurden mit dem Programm **Gprof** ermittelt.

weiteres H/ACA snoRNA Programm vollzogen werden kann, kann nicht mit

den gefundenen *X. tropicalis* Sequenzen das *T. rubripes*-Genom untersucht werden.

Es sind noch weitere Trainingsläufe des Programmes **Fishers** mit bekannten snoRNAs gegen reduzierte Genome nötig, um die Sensitivität des Programmes zu testen.

Allerdings muss auch ein Fehler in der von mir durchgeführten Analyse eingeräumt werden. Der Vergleich der Nukleotidwahrscheinlichkeiten der variablen Positionen (Kap.: 3.1.1) zeigte, dass sich die Wahrscheinlichkeiten der einzelnen variablen Positionen zwischen der Hefe und den Vertebraten unterscheiden. Daher können nicht die von mir verwendeten Nukleotidwahrscheinlichkeiten der *S. cerevisiae* benutzt werden, um Sequenzen an den Vertebratengenomen zu finden.

3.2 Suche nach Box C/D snoRNAs mit SnoScan

3.2.1 Auswertung der Ergebnisse

Das Programm **SnoScan** (Kap. 2.2) wurde am Tetradon und am Fugu-Genom angewendet. Alle Fugu Sequenzen, die einen 'Strong Terminal' besitzen, wurden durch den **Ensembl Fugu Genome Browser** [84] auf ihre Lokalisation im Genom untersucht. Durch den Genome Browser können snoRNAs ermittelt werden, die auf Genen liegen, die in der Translation involviert sind. 12 Sequenzen konnten gefunden werden, die entweder auf einem ribosomalen Protein oder auf auf einem Gen für mRNA/rRNA-Prozessierung sitzen.

Diese Sequenzen wurden gegen alle bekannten Box C/D snoRNA Familien aus der **RFAM-Datenbank** [35] aligniert, um die Sequenz einer etwaigen Familie zuzuordnen. Eine Sequenz, lokalisiert auf scaffold_757 Position 94612 - 94688, zeigte gute Konservierungen zu den spezifischen Boxmotiven. Diese Sequenz liegt im Fugu-Genom auf dem ribosomalen Protein RPL0. Durch das Alignieren gegen alle bekannte Box C/D snoRNA-Familien zeigt die Sequenz die beste Konservierung zu den snoRNA-Familien U27, U33, U50, U73. Durch die Analyse mit dem Programm **HMMER** [20] wurden die Erwartungswerte der Alignments ermittelt, die sich im Bereich von $E = 1.0 - 2.0$ bewegen. **HMMER** ist eine Möglichkeit statistische Modelle von multiplen Sequenzalignments zu erhalten. **HMMER** erstellt Profile HMMs, die Aussagen über die Konservierung einer Alignmentzeile über den Erwartungswert geben.

Neben der Konservierung zu einer snoRNA-Familie ist auch die Konservie-

zung zu komplementär ribosomalen Bindungsstellen entscheidend. Alle vier Alignments sind gegen reverse komplementäre ribosomale RNAs aligniert worden. Beim Vergleich der ribosomalen Stellen zwischen den bekannten Familiensequenzen und der Sequenz zeigte snoRNA-Familie U50, dass sie die gleiche ribosomale Bindungsstelle 28S rRNA hat wie die Sequenz (Abb. 3.1). Die ribosomale Bindungsstelle weist bei der Fugusequenz an Position 59 - 69 fünf Matches zur bekannten ribosomalen Bindungsstelle der humanen U50-Sequenz auf. Eine Sequenz aus der snoRNA Familie U50 (Acc.Numb.: X96662) hat an der Stelle 14 - 24 das 28S Target. Die Box C ist völlig konserviert, Box D ist teilweise konserviert. Die teilweise Konservierung von Box D rührt von einer 2 nt langen Deletion her. Eine weitere Familie U33 zeigte

```

CLUSTAL W (1.8) multiple sequence alignment

xxxxx  -----CCCCC-RRRRRRRRRR-----
HsU50   TATCTGTGATGATCTTATCCCGAACCT--GAACTTCTGTTGAAAAAAAAA
TrU50   -ACCTGTGATGAAGCCGTCTCTCTCCAGGAGCGTGTGGGAAACACCA
          * **** *      * *   * *   * * * * * * * * * *
xxxxx  -----rrrrrrrrrrDDddd
HsU50   ACTTTTACGGATCTGG-CTTCTGAGAT-
TrU50   GTAAACCGGGAGCAGCTCCCTCTGATG
          *** * * * * ** ***
E-value: 2.0

```

Abbildung 3.1: Alignment der snoRNA Familie U50 gegen Fugu-Sequenz Scaffold_757 (Tr_U50). C: Box C, D: Box D, R: ribosomale Bindungsstelle (*Homo sapiens*), r: ribosomale Bindungsstelle (*Takifugu rubripes*/*Homo sapiens*)

auch sehr gute Konservierungen an den Box C und Box D Stellen auf (Abb.: 3.2). Jedoch sind die ribosomalen Bereiche, die bei Position 20 - 33, 36 - 46 der humanen snoRNA U33 Sequenz (Acc.Numb.: AB028893) und bei der *T. rubripes*-Sequenz auf Position 59 - 69 liegen, nicht so gut konserviert, wie bei Familie U50. Auch die snoRNA Familie U33 hat wie die *Takifugu rubripes*-Sequenz als Bindungsstelle 28S rRNA.

Die einzelnen Sequenzen der snoRNA (U33, U50 und scaffold_757) liegen alle auf unterschiedlichen Proteinen, so dass ausgeschlossen werden kann, dass es sich um Tr_U33 oder Tr_U50 handelt. Auch aus den Erwartungswerten kann geschlossen werden, dass es sich nicht um eine bekannte Box C/D snoRNA handelt.

```

CLUSTAL W (1.8) multiple sequence alignment

xxxxxx  -----CCCCC-----RRRRRRRRRRRRRR--RRRRRRRRRR--
Hs_U33   GGCCGGTGATGAGAACTTCTCCCACTCACATTTCGAGTTCCCGACCATGA
Mm_U33   AGCTTGTGATGAGA-CATCTCCCACTCATGTTTCGAGTTGCTCGACTATGA
Tr_U33   -ACCTGTGATGAAGCCGTCTCTCTCCAGGAGCGTGTGTGGGAAACACCA
          *  *****  *  **** * * **  ** **  *  *  *
xxxxxx  -----rrrrrrrrrrrr--DDDD-----
Hs_U33   GATGACTCCACATGCACTACCATCTGAGGCCA
Mm_U33   GATGACTCTACATGCACTACCATCTGAGGCTG
Tr_U33   G-TAAACCGGGAGCAGCTCCCCTCTGATG---
          * * * * *  ** ** ***** *
E-value: 1.0

```

Abbildung 3.2: Alignment der snoRNA Familie U33 gegen Fugu-Sequenz Scaffold_757 (Tr_U33). C: Box C, D: Box D, R: ribosomale Bindungsstelle (Säugetiere), r: ribosomale Bindungsstelle (*Takifugu rubripes*).

3.2.2 Probleme

Das Programm *SnoScan* bietet die Möglichkeit, sehr schnell mögliche snoRNA-Sequenzen über das Subprogramm *snoScan* zu finden. Dieses Programm untersucht das Genom über die im Kap. 2.2 beschriebene Vorgehensweise. Eine genauere und sichere Methode wäre allerdings das Subprogramm *scan-yeast* gewesen. Dieses Programm untersucht die Genomsequenz nach der beschriebenen Vorgehensweise und mittels bekannten Methylierungsstellen der rRNA. Die Methylierungsstellen der *Takifugu rubripes* rRNAs sind nicht bekannt und hätten ebenfalls durch experimentelle Versuche ermittelt werden müssen.

Neben der experimentellen Suche nach Methylierungsstellen gibt es als zweite Möglichkeit das *Blast2*-Sequenzprogramm [2], um Methylierungsstellen zu finden. Maria Carmela Accardo verwendete diese Möglichkeit bei der Untersuchung des *Drosophila melanogaster*-Genoms [1]. Sie alignierte die bekannten *S. cerevisiae*-Methylierungsstellen gegen das *S. cerevisiae*-Genom und markierte diese gefundenen Stellen auf den *D. melanogaster* rRNAs. Durch einen Test der gefundenen *D. melanogaster*-Methylierungsstellen gegen das *S. cerevisiae*-Genom konnte sie wiederum feststellen, ob die *D. melanogaster*-Methylierungsstellen die richtige *S. cerevisiae* snoRNA finden.

Diese Methode kann auch an den Vertebraten angewendet werden. Sie wäre

in dieser Diplomarbeit zum Einsatz gekommen, sofern die Publikation von Maria Carmela Accardo nicht erst kurz vor Ende der Diplomarbeit erschienen wäre.

3.3 Resultate des NCBI-Blast

3.3.1 Die Geschichte der snoRNA-Familie U17/E1

Die snoRNA Familie U17 ist die am häufigsten publizierte snoRNA Familie mit ihren 89 Publikationen (Stand: September 2004). Die Forschungsgruppe von *Cecconi* stellte einen evolutionären Vergleich dieser Familie auf sowie auch *Cecconi* die einzige bis heute publizierte H/ACA snoRNA am *Takifugu rubripes*-Genom fand [14, 16]. Durch einen Homologievergleich konnten wir weitere U17 snoRNAs an den Teleostengenomen *Tetradon nigrovirdis*, *Danio rerio*, *Orzyias latipes* und an den Säugetieren *Sus scrofa*, *Canis familiaris*, *Pan troglodytes*, *Rattus norvegicus* und *Bos taurus* finden (Abbildung: 4.3).

Eigenschaften und Aufgabe der U17

SnoRNA U17/E1 wurde bis heute an verschiedenen Organismen (Säugetieren, Amphibien, Reptilien, Teleosten und an der Hefe (snR30)) gefunden. U17 besitzt bei den Vertebraten eine Länge von 205 - 225 nt, snR30 bei der Hefe ist um 100 bp länger [3].

Die Funktion der Mehrzahl von H/ACA snoRNA ist Pseudouridylierung von rRNA. H/ACA snoRNA U17 hat allerdings die Spaltung der pre-rRNA aus der 5' ETS (external transcribed spacer), die zur Entstehung der 18S rRNA führt als Aufgabe [16].

In der humanen U17 Sequenz ist die snoRNA im RCC1 Gen lokalisiert. Bei den Reptilien und Teleosten sind die snoRNAs in den Introns des ribosomalen Protein RPS7 verteilt [14]. Am Säugetier (Mensch, Maus, Ratte) wurde keine Sequenz gefunden, die sich auf dem Gen RPS7 befindet. Bei einem evolutionären Vergleich zwischen Vertebraten, Amphibien und Teleosten stellte Cervelli fest, dass sich der evolutionäre Sprung zwischen dem ribosomalen Protein RPS7 und dem Hostgen RCC1 in der Entwicklung vom Vogel zum Säugetier ergeben hat. Diese Ergebnisse von Cervelli beruhen auf der Tatsache, dass beim Sekundärstrukturvergleich aller Organismen die Sekundärstruktur der *Gallus gallus* U17 Sequenz eine geringe Verkürzung der Sequenz zeigt und das rRCSI-Motiv fehlt.

Die *Takifugu rubripes* Sequenzen sind auf dem ribosomalen Protein RPS7 lokalisiert. Das Protein besteht aus sieben Exonen und sechs Intronen. In jedem Intron ist eine Kopie der U17 zu finden, wobei nur bei den letzten vier Intronen eine funktionstüchtige U17 Kopie existiert. Bei den Kopien lokalisiert auf Intron 1 und 2 wurde festgestellt, dass es sich um Pseudogene handelt [14], da den Kopien U17a und U17b im Gegensatz zu Kopie U17c-f die rRCS-Motive fehlen.

Spezielle Merkmale der U17

U17 snoRNA haben spezielle Motivsequenzen rRCSI, rRCSIII sowie zwei weitere definierte konservierte Sequenzen (Tab.: 3.9).

sequenzen rRCSI, rRCSIII sowie zwei weitere definierte konservierte Sequenzen (Tab.: 3.9).

rRCSI und rRCSIII wurden nach ihrer Funktion als rRNA komplementäre Sequenz benannt. Alle vier Motive haben als Funktion mit der 18S rRNA zu interagieren. Im Vorkommen unterscheiden sich die Motive. Fünfzig Nukleotide vor Box H tritt in jedem Vertebraten das Motiv rRCSIII auf. rRCSI tritt ca. 5 nt nach der Box H auf. Dieses Motiv kommt allerdings nur bei Amphibien, Reptilien und Teleosten vor. Das Fehlen des Motives ist ein Merkmal dafür, dass sich der evolutionäre Wandel zwischen Vögeln und Säugetieren vollzogen haben kann (Kap.: 3.3.1) [16].

Motive m1 und m2 haben die 18S-Produktion als Aufgabe. Auch sind sie für die Zellebensfähigkeit wichtig. Die Produktion von rRNA wirkt sich auf die Zellebensfähigkeit aus. Wenn weniger rRNA produziert wird können weniger Proteine produziert werden. Beide Motive kommen neben den Vertebraten auch in der Hefe vor [3].

Suche nach U17 snoRNA an den Teleosten

Von der bekannten Fugu snoRNA U17 (Acc. Numb: X94942) ausgehend wurde ein Homologievergleich gegen die *Tetradon nigrovirdis*, *Danio rerio* und *Orzyias latipes*-Genome durchgeführt. Nukleotidstränge mit einer Länge von 214 Basen (*Tetradon*) wurden von den vorhandenen 217 Basen gefunden. Prozentual bestand zwischen den *T. rubripes* und *T. nigrovirdis*-Sequenzen eine Homologie von 90 %. Die Sequenzhomologie bei den *D. rerio*- und *T. rubripes*-Sequenzen betrug 86 % bei einer gemeinsamen Länge von 186 Nu-

Spezielle U17 Features			
Benennung	Sequenz	Merkmale	Vorkommen
rRCSI	TGTCCCCGGCC	18S Target	Amphibien, Teleosten, humaner Kopie f
rRCSIII	TCTGTCCTGG	18S Target	Vertebraten
m1	ATATTCCTA	18S Produktion	Hefe, Vertebraten
m2	AAACCAT	18S Produktion	Hefe, Vertebraten

Tabelle 3.9: Sequenzmotive, die in bestimmten Distanzen zu Box H bzw. Box ACA bei der snoRNA U17 auftreten.

kleotiden. Der Homologievergleich der Fugu-Sequenzen gegen das *Orzyias latipes*-Genom wurde durch das *NIG DNA Sequencing Center* vollzogen. Der Vergleich der Sequenzen gegen das Genom erbrachte hier eine Homologie von 191 Nukleotiden, prozentual beträgt die Homologie 90 % gegenüber dem *Ta-kifugu rubripes*.

Bei näherer Analyse der Sequenzen zeigte sich, dass die Sequenzen gleiche Konservierungen an den für die U17 snoRNA spezifischen Motiven sowie an Box H und Box ACA haben (Abb.: 3.3). Die spezifischen Motive rRCSIII, rRCSI, m1 und m2 kommen, wie von Atzorn [3] und Cervelli [16] beschrieben, an definierten Positionen in der U17 vor. Die Motive m1 und m2 bilden den ungepaarten Nukleotidbereich der 3' Haarnadel. Die ersten beiden Nukleotide von m1 und die letzten zwei Nukleotide von m2 gehen als einzige Nukleotide der beiden Motive eine Basenpaarung ein. Motiv m1 kommt im Alignment (Abb.: 3.3) an Position 147 - 155 der *Tetradon nirgovirdis* Sequenz (Tn_E1_1) vor. Das Motiv m2 tritt bei allen Vertebraten sieben Nukleotide vor der Box ACA auf.

Motive rRCSI und rRCSIII kommen an den Positionen 110 - 120 bzw. im Positionsbereich 32 - 40 der *Tetradon nirgovirdis* Sequenz (Tn_E1_1) vor. Bei beiden Motiven sind zwei Nukleotide nicht konserviert. Am Motiv rRCSI tritt eine Mutation beim dritten Nukleotid der *Orzyias latipes* Sequenz auf. Anstelle der Motivbasensequenz TGTCCCCGGCC tritt die Basensequenz TGCCCCCGGCC auf. Die Mutation an Stelle 3 hat allerdings keine Auswirkung auf die Komplementarität der Target Region. Motiv rRCSI der *Orzyias latipes* Sequenzen kann noch an die 18S rRNA binden. Die Mutation des Motives rRCSIII tritt an Position 8 und 9 bei den Sequenzen von *Orzyias latipes* und *Danio rerio* auf. Die *Orzyias latipes*-Sequenzen haben an

Position 8 anstelle eines T ein G, die *Danio rerio*-Sequenzen haben anstelle der Basenkombination TG die Basenkombination GT.

Box H tritt bei allen Sequenzen an gleicher Position auf. Das Sequenzmuster ANANNA entspricht bei den Sequenzen von *O. latipes*, *T. rubripes*, *D. rerio* der Box H. Die Sequenzen vom *T. nigrivirdis* Genom entsprechen dem Box H-Muster ANANNNA. Atzorn schreibt in seiner Veröffentlichung über die U17 snoRNA, dass nur die Konservierung der Nukleotide A_1 und A_3 unabdingbar ist für die Anheftung der 18S rRNA [3].

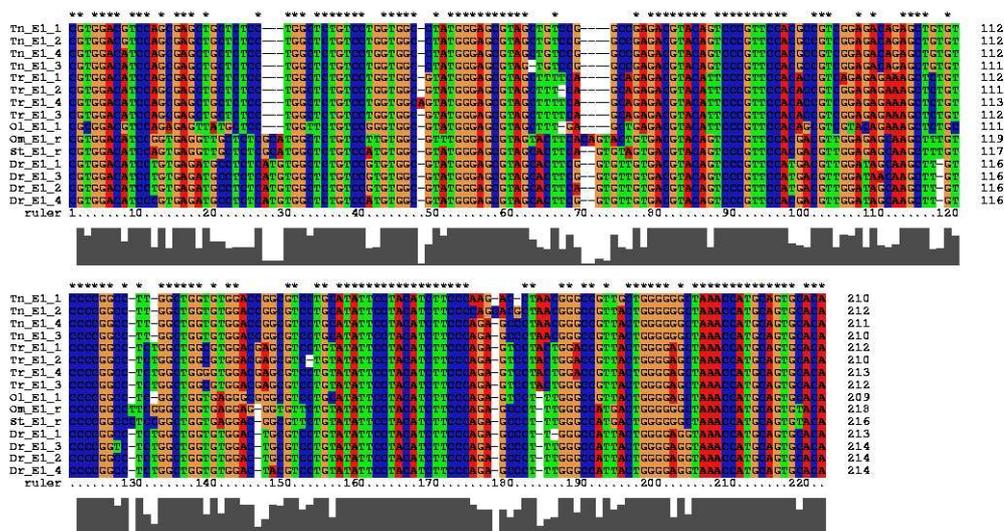


Abbildung 3.3: Alignment der *Orzyias latipes*, *Takifugu rubripes*, *Tetradon nigrivirdis*, *Danio rerio*, *Onchorhynchus mykiss* und *Salmo trutta* U17 snoRNA.

Genomische Organisation der U17 snoRNA

U17 snoRNA kommen entweder auf einem Hostgen bei den Säugetieren oder auf einem ribosomalen Protein bei den Amphibien, Reptilien und Teleosten vor (Kap.: 3.3.1). Die Sequenzen der U17 snoRNA liegen auf dem ribosomalen Protein RPS7. Insgesamt existieren vier Kopien der U17 snoRNA, die jeweils auf einem eigenen Intron des Gens liegen. Auch durch die Publikation von 'He' (unveröffentlichtes Paper) über die U17 snoRNA Sequenzen des

Zebrafisches ist bekannt, dass auf dem ribosomalen Protein RPS7 snoRNA-Sequenzen liegen. Ausgehend von dieser Tatsache fing die Suche nach den U17 snoRNA-Sequenzen an den Teleostengenomen an. Anhand des Ensembl Genome Browsers wurde die genomische Organisation der Zebrafisch-Sequenzen festgestellt. Alle vier Sequenzen liegen auf dem RPS7 Gen (Abb.: 3.4) und sind verteilt zwischen einzelnen Introns. Die Se-

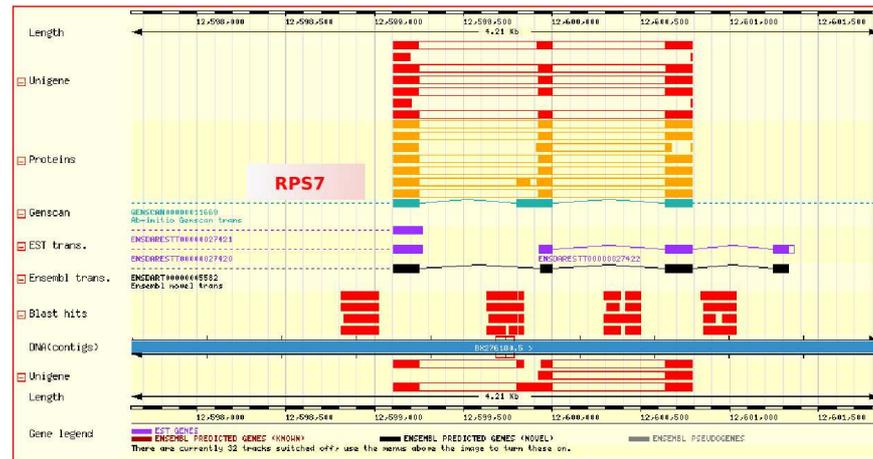


Abbildung 3.4: BlastHit der *Danio rerio* U17 snoRNA Sequenzen. Alle vier U17 snoRNA Kopien (rote Rechtecke) liegen auf dem ribosomalen Protein RPS7. Exonstruktur des Genes RPS7 wird im Genome Browser grün dargestellt.

quenzen vom *Tetradon nirogovirdis* liegen gleichfalls auf verschiedenen Introns des ribosomalen Proteins RPS7. Von diesen Tatsachen ausgehend, dass alle bisherigen Teleostensequenzen auf verschiedenen Introns eines Genes lagen, wurde die genomische Organisation der *Orzyias latipes* Sequenzen über GeneID [70] untersucht. GeneID ist ein Programm zur Vorhersage von Genen in anonymen Vertebraten DNA-Sequenzen. Das Programm sucht im ersten Schritt nach Gensignalen, wie z.B Start- und Stopcodon, die durch eine Positionsgewichtsmatrix (PWM) gescort werden. Danach wird aus den gefundenen Signalen mögliche Exone konstruiert, die aus der Summe der einzelnen Gensignalscores bewertet werden. Aus den vorhergesagten Exonen wird im letzten Schritt eine Genstruktur assembliert.

Durch das Programm konnte festgestellt werden, dass der Bereich, auf dem

die snoRNA liegen, aus sieben Exonen besteht.

Die *Orzyias latipes* U17 Sequenzen liegen auf folgenden Bereichen (Abbildung 3.5):

- OlU17c Bereich 214869 - 215099
- OlU17d Bereich 215567 - 215783
- OlU17e Bereich 216226 - 216439
- OlU17f Bereich 216787 - 217002

Die Exone, die diese snoRNA umspannen liegen in den Bereichen: Exon 1 (214200-214276), Exon 2 (214738- 214811), Exon 3 (215133- 215229), Exon 4 (215970- 216034), Exon 5 (216501- 216644), Exon 6 (217071- 217142), Exon 7 (217226- 217310).

Aus den obigen Daten kann man erkennen, dass OlU17c zwischen den Exonen 2 - 3 liegt, OlU17d zwischen Exon 3 - 4, OlU17e zwischen Exon 4 - 5, OlU17f zwischen Exon 5 - 6. Aus der Exonstruktur und Lage der snoRNAs kann festgestellt werden, dass die genomische Organisation der *Orzyias latipes*-Sequenzen der genomischen Organisation der *Takifugu rubripes* U17 entspricht.

3.3.2 Familie U24

Familie U24 ist ein Mitglied der Box C/D snoRNA. Diese Familie enthält zwei komplementäre Bereiche zur 2'-O-Methylierung der 28S rRNA, die 12 bis 21 nt lang sind. Beide komplementäre Bereiche sind verwandt mit den ribosomalen Bindungstargets der snoRNA Familien U14, U20, U21 [73].

Eine weitere Funktion der snoRNA U24 ist die Anheftung der snoRNA U14 an die Terminalstruktur der snoRNA U24. Des Weiteren enthält snoRNA U24 Signale für die Prozessierung der U14 aus dem Hostgen hsc70. Beim Vergleich der U24 snoRNA in Vertebraten und Hefe zeigte sich, dass während der Evolution dieser Spezien Mutationen aufgetreten sind [73]. Jedoch hatten diese Mutationen keinen Einfluss auf die Komplementarität der snoRNA zur ribosomalen RNA.

Die U24 snoRNA ist im Genom auf einem für sie spezifischen Gen zu finden. U24 ist beim *Gallus gallus* und *Homo sapiens* auf dem zweiten Intron des ribosomalen Proteins RPL7 lokalisiert.

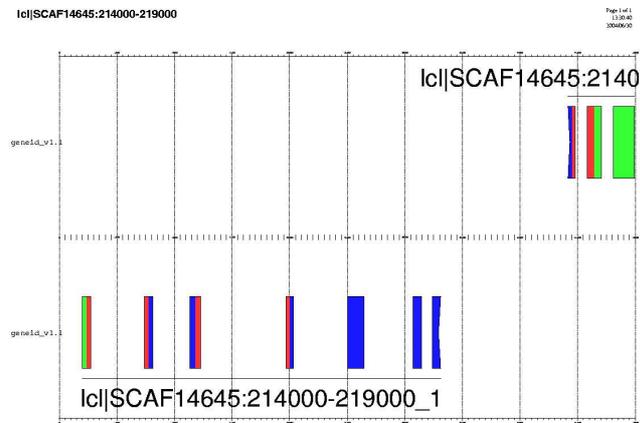


Abbildung 3.5: Genomische Organisation der *Orzyias latipes* U17 snoRNA Sequenzen. Der obere Bereich stellt den 5'-3' dar, der im Bereich von 214000-219000 3 Exone hat. Im unteren Teil wird der 3'-5' Bereich visualisiert. Die U17 Sequenzen liegen im 3'-5' Bereich, der insgesamt 7 Exone (farbige Rechtecke) hat. Zwischen den Exonen sind die U17 Sequenzen lokalisiert

Lediglich eine Sequenz dieser Familie wurde am *Tetradon nirgovirdis* gefunden. Die Sequenz wurde durch Homologie zur *Takifugu rubripes* U24 Sequenz entdeckt (Abbildung: 3.6).

Interessanterweise hat U24 an den Fischsequenzen nicht die übliche H-Box Konservierung. Das sechste Nukleotid der Box H ist sowohl an der Fugu-Sequenz als auch an der Tetradon-Sequenz mutiert [73]. Interessant wären Homologe von *Danio rerio* und *Orzyias latipes* gewesen, um betrachten zu können, ob auch hier die Mutation stattgefunden hat.

'Qu' publizierte, dass snoRNA U24 aus zwei 12 Nukleotid langen Sequenzen besteht, die komplementär zur 28S rRNA sind.

Die Tetradonsequenz hat an zwei Stellen Komplementarität zur 28S rRNA gezeigt. Die Fugu-Sequenz U24 hat allerdings im ersten komplementären Bereich eine Nukleotid-Deletion. Dieser Bereich zeigt keine Komplementarität zur 28S rRNA.

```

CLUSTAL W (1.8) multiple sequence alignment

#####
TrU24      CCCCC-----XXXXX-XXXXX-----XXXXXXXXXX-D
TnU24      TGAATGTAAC--GATATT-GCTATCTTAATGACGATGACAAATCCACCAAGATCGC
SecStr     .(((.....(((((((.....((((.....)))))).....)))))).....))....

#####
TrU24      DDD
TnU24      TGA
SecStr     ...

```

Abbildung 3.6: Zusammenfassung aller U24 Sequenzen. Alignment zeigt Consensusstruktur aller Sequenzen. Erste Zeile gibt komplementäre ribosomale Stellen an (X), Box C (C) und Box D (D) an.

Kapitel 4

Evolution der snoRNA

Zwei evolutionäre Szenarien der snoRNA sind bis zum heutigen Zeitpunkt bekannt. Jonathan Gilley und Mike Fried untersuchten die Evolution der snoRNA U36 (Kap.: 4.2.3). Die Evolution der snoRNA U17/E1 entschlüsselte Martina Cervelli (Kap.: 4.2.1).

Beide Szenarien zeigen, dass die Evolution der snoRNA interessant abgelaufen ist. Zum einen duplizieren snoRNA innerhalb der Spezies und zwischen unterschiedlichen Organismen rezent, zum anderen evolvierten snoRNAs im Laufe der Entwicklung auf unterschiedliche Hostgene.

Weitere zwei evolutionäre Szenarien für die snoRNA E2 und die snoRNA E3 wurden an den Wirbeltieren (Kap.: 4.2.2) aufgestellt.

4.1 Stammbaum der Wirbeltiere

Die Wirbeltiere lassen sich in fünf Klassen unterteilen: Osteichthyes, Amphibien, Reptilien, Vögel und Säugetiere. In der zweiten Unterteilung werden die Vertebraten unterschieden nach kiefertragendem Tier und nach kieferlosem Tier.

Alle Tiere, die zur Klasse Amphioxus und Agnatha gehören, gruppieren sich zur Klasse der kieferlosen Tiere (Abb.: 4.1). Die Klassen der kiefertragenden Tiere, in die über 95 % der heute lebenden Wirbeltiere eingeordnet werden, sind Chondrichthyes, Osteichthyes, Coelacanthimorpha (Latimeria), Dipnoi (Lungenfische) Amphibien, Reptilien, Vögel und die Säugetiere.

Die kieferlosen Fische waren die ersten Tiere, die auf der Erde auftauchten. Die Agnatha sind kieferlose Fische und der Ursprung der Vertebraten.

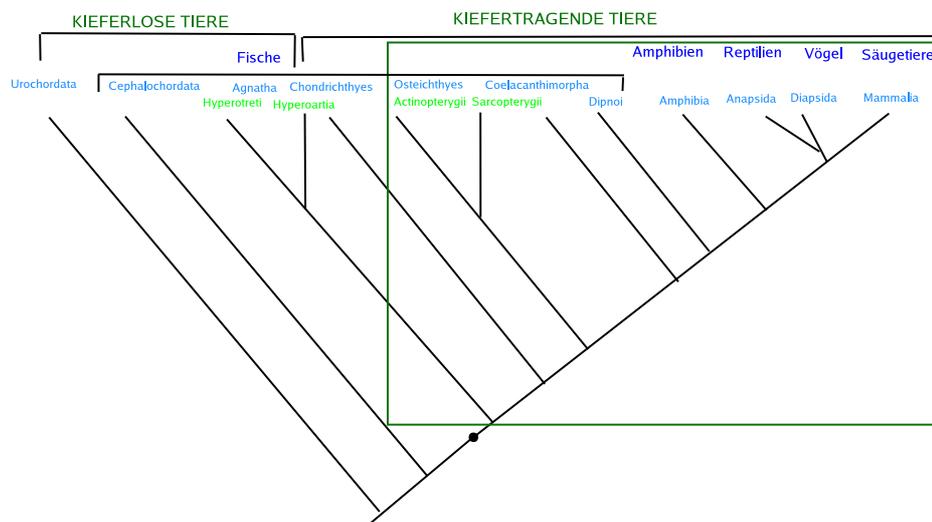


Abbildung 4.1: Stammbaum der Vertebraten, fett markierter Punkt ist der Ursprung der Vertebraten (Abbildung nach Fritzsich [2004]).

Erst 160 Millionen Jahre später sind wahrscheinlich die kiefertragenden Fische entstanden. Diese Gruppierung bildete im Gegensatz zu den kieferlosen Fischen einen Kiefer und paarige Flossen aus [92]. Eine rasante Entwicklung durchliefen die kiefertragenden Fische in ihrer Schwimmeigenschaft und Nahrungsaufnahme. Nach der Entstehung der Agnatha kam es zur Ausbildung der Quasten- und Strahlenflosser (Sarco- und Actinopterygii). Fast alle heutigen Fische stammen von den Strahlenflossern ab. Aus der Gattung der Strahlenflosser entwickelten sich auch die Teleosteen. Zu dieser Gruppierung gehören die vier Modellfische *Takifugu rubripes*, *Tetradon nigrovirdis*, *Orzias latipes* und *Danio rerio* (Abbildung: 4.1).

Die Entwicklung der Amphibien beruht auf den im Wasser schlecht überlebenden Quastenflosser. Der Übergang vom Fisch zur Amphibie war der wichtigste evolutionäre Schritt in der Entwicklung der Vertebraten. Anatomisch bildete sich bei den Amphibien neben den ehemaligen Kiemen Lungen aus, was sie überlebensfähig auf Land und zu Wasser machte [85]. Aus der Gruppierung der Amphibien kann für phylogenetische Analysen der *Xenopus laevis* und *Xenopus tropicalis* verwendet werden. Beide Amphibien-Genome stehen sich in ihrer Phylogenie sehr nahe.

Die Reptilien waren die erste Tiergruppe, die vollständig auf dem Festland

lebte. Sie entwickelten sich aus primitiven Amphibien. Zwei verschiedene Stammbaumzweige sind aus dem Reptilienstamm hervorgegangen: Vögel und Reptilien. Da zum einen *Archosauria* Beziehungen zu den Vögeln haben und zum anderen die Synapsiden ähnlich zu den Säugetieren sind, kam es zur Zweiteilung des Reptilienstammbaumes (Abb.: 4.1). Ancestrale Formen der Reptilien differenzieren nur wenig zu den Amphibien.

Die Säugetiere haben sich aus den Reptilien entwickelt. Es existieren zahlreiche Ordnungsklassen verschiedener Säugetiere. Säugetiere, die von uns in der phylogenetischen Analyse verwendet wurden, sind in folgenden Säugetierordnungen zu finden: Primaten, Rodentia (Nagetiere), Carnivora (Raubtiere) und Paraxonia (Paarhufer) [85]. Schaf (*Ovis aries*), Rind (*Bos taurus*) und Pferd (*Equus caballus*) gehören zur Gattung der Paarhufer. Die beiden erst genannten lassen sich des Weiteren in die Klasse der Wiederkäuer einteilen. *Felis catus* und *Canis familiaris* gruppieren sich in der Klasse der Raubtiere. *Mus musculus* und *Rattus norvegicus* sind Nagetiere. *Pan troglodytes*, *Homo Sapiens* sind Primaten. Die phylogenetischen Verhältnisse der einzelnen Vertebraten, deren Genom weitgehend fertig sequenziert ist, sind in Abbildung 4.1 zusammengefasst. Tabelle 4.1 gibt ein Überblick über die momenta-

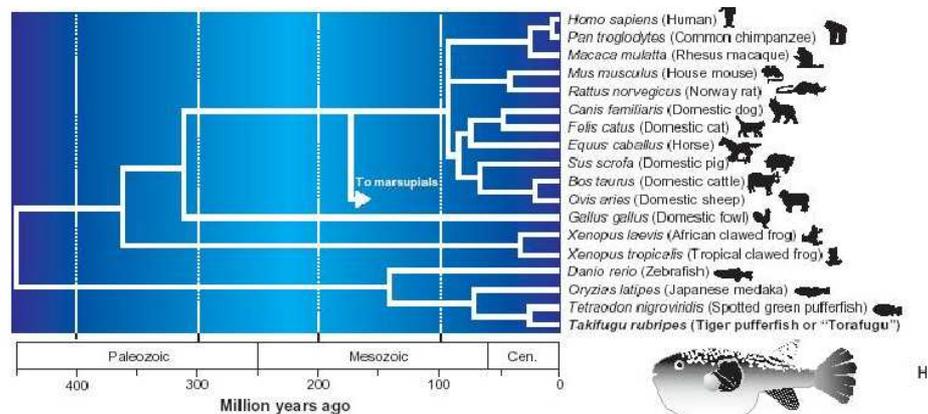


Abbildung 4.2: Phylogenetischer Baum der die Verwandtschaftsverhältnisse der Vertebratengenome darstellt, die sequenziert wurden (Abbildung: Hedges [2002])

nen Assemblierungszustände, der in den Analysen verwendeten Organismen.

Organismus	Art	Gattung	Genomdatenbank	Status
H. sapiens	Genome	Primaten	NCBI	vollständig Nature 15.02.2001
P. troglodytes	Genome	Primaten	NCBI	unvollständig
M. musculus	Genome	Säugetier	NCBI	vollständig Nature 05.12.2002
R. norvegicus	Genome	Säugetier	NCBI	vollständig Nature 01.04.2004
C. familiaris	Genome	Säugetier	NCBI	unvollständig
S. scrofa	EST	Säugetier	Roslin Institute	unvollständig
B. taurus	Genome	Säugetier	Cattle Genome Database	unvollständig
G. gallus	Genome	Vogel	NCBI	unvollständig
X. laevis	Genome	Amphibien	TIGR	unvollständig
X. tropicalis	Genome	Amphibien	JGI	unvollständig
D. rerio	Genome	Fisch	NCBI	unvollständig
O. latipes	EST	Fisch	TIGR	unvollständig
T. nirogovirdis	Genome	Fisch	Genoscope	unvollständig
T. rubripes	Genome	Fisch	JGI	vollständig teilweise öffentlich Science 23.08.2002

Tabelle 4.1: Auflistung aller verwendeten Organismen mit Nennung des Herkunftortes und Genomzustand.

4.2 Szenarien der snoRNA-Evolution

4.2.1 Evolution der snoRNA U17/E1

Die Evolution der snoRNA U17 ist sehr vielseitig (Abb.: 4.3). Zwischen den Teleosteen und Säugetieren kam es im Laufe der Entwicklung zu Duplikationen, Mutationen. Auch wurde das Hostgen der snoRNA U17 gewechselt.

In den Teleosteen liegen jeweils vier funktionstüchtige Kopien der snoRNA U17 im Gen RPS7 vor [14]. Im Gen RPS7 kommen zusätzlich noch zwei weitere Kopien der U17 snoRNA des *Takifugu rubripes* vor, die allerdings keine Funktion haben. Diese Kopien konnten an den Genomen des Zebrafisches (*D. rerio*), Medaka (*Orzias latipes*) und am grünen Pufferfisch (*T. nirogovirdis*) nicht mehr gefunden werden (Kap.: 3.3.1).

Im Laufe der Entwicklung kam es jedoch zwischen dem *Danio rerio* und dem *Xenopus laevis* wieder zur Ausbildung von zwei weiteren Kopien. *X. laevis* enthält insgesamt sechs Kopien der snoRNA U17, die sich auf die Introns 1 bis 6 des ribosomalen Proteins RPS8 (Synonym von RPS7) verteilen. Im phylogenetischen Baum ist die Verzweigung und die Ausbildung von sechs *X. laevis*-Kopien mit einem Bootstrapwert von 963 gesichert.

Zwischen *X. laevis* und *G. gallus* kam es zum Wechsel des Hostgenes [16]. Die

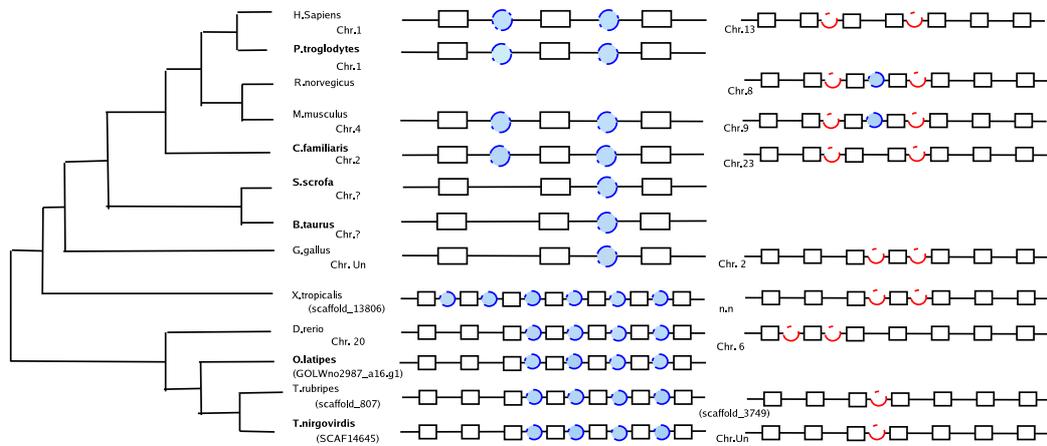


Abbildung 4.3: Evolutionäre Geschichte der snoRNA E1 (Mitte) und snoRNA E2 (rechte Seite).

ursprüngliche Organisation der snoRNA U17 wurde zwischen *Xenopus tropicalis* und *Gallus gallus* geändert. In den Vögeln und Säugetieren liegen die Kopien der snoRNA U17 auf dem Gen RCC1 (zellenregulatorisches Protein). Eine homologe Kopie der snoRNA U17 ist auf dem Gen RCC1 lokalisiert. Eine Kopie der snoRNA U17 des *X. tropicalis* duplizierte wahrscheinlich als einzige Kopie auf das Huhn.

Es ist evolutionär bekannt, dass es zu einer erneuten Duplikation der snoRNA U17 zwischen *X. laevis* und *M. musculus* gekommen ist. Um konkreter sagen zu können, zwischen welchen Organismen die Duplikation von einem Homolog zu zwei homologen Kopien der U17 stattgefunden hat, wurden weitere Säugetier-Genome (*C. familiaris*, *S. scrofa*, *P. troglodytes*, *B. taurus*) nach Paralogen der snoRNA U17 untersucht. Es konnte festgestellt werden, dass jeweils eine Kopie der snoRNA U17 am Genom des *Sus scrofa* und des *Bos taurus* vorkommt. Weiterhin existieren zwei Kopien am *Canis familiaris* und am *Pan troglodytes*.

Es ist daher möglich, dass sich die Duplikation der snoRNA U17 von einem Duplikat auf zwei Duplikate zwischen *S. scrofa* und *C. familiaris* vollzogen hat.

Dies sollte durch eine phylogenetische Analyse (Abbildung: 4.4) bestätigt werden. Die Verzweigung der *S. scrofa*, *B. taurus* und *C. familiaris* im phylogenetischen Baum zeigt, dass die Sequenzen des *B. taurus*, *C. familiaris* und

S. scrofa verwandt sind. Die Verzweigungen zwischen den Organismen wird im phylogenetischen Baum nicht korrekt dargestellt, da die Sequenz des *S. scrofa* zwischen den Sequenzen des *B. taurus* und *C. familiaris* stehen müsste. Da die Bootstrapwerte zwischen *B. taurus* und *C. familiaris* sowie zwischen der Gruppierung *B. taurus*, *C. familiaris* und *S. scrofa* sehr gering sind, kann jedoch trotzdem angenommen werden, dass der Sprung von einem Duplikat auf zwei Duplikate zwischen *S. scrofa* und *C. familiaris* stattgefunden hat und die Verzweigungen zwischen den Organismen im phylogenetischen Baum falsch dargestellt wurden.

Die zwei Duplikate des *C. familiaris* wurden auch an den Genomen der Primaten und *M. musculus* wieder gefunden. Im phylogenetischen Baum (Abb.: 4.4) ist erkennbar, dass drei *M. musculus* Kopien der snoRNA U17/E1 und nur eine *R. norvegicus* Kopie der snoRNA U17/E1 existieren. Auf dem Gen RCC1, Trägergen der snoRNA U17/E1 ab dem *G. gallus*, konnte nur auf dem *M. musculus* Genom zwei Kopien der snoRNA U17/E1 gefunden werden. Zusätzlich dazu konnte allerdings eine Kopie auf dem Gen LAMR1, Trägergen der snoRNA E2, am Genom des *M. musculus* und *R. norvegicus* gefunden werden. Der phylogenetische Baum zeigt, dass die Nagetiersequenzen der snoRNA U17 miteinander verwandt sind. Der Bootstrapwert aller Nagetiersequenzen hat einen Wert von 628 bei 1000 Replikationen.

4.2.2 Evolutionsszenarien für snoRNA E2 und E3

Neben der U17/E1 (Kap.: 3.3.1) wurden im Jahre 1993 zwei weitere H/ACA snoRNA, E2 und E3, gefunden. Da zu diesem Zeitpunkt die Existenz der zwei verschiedenen snoRNA-Boxen noch nicht bekannt war, erhielten die ersten drei H/ACA snoRNAs eine andere Benennung [78, 74, 63]. Alle drei snoRNAs unterscheiden sich hinsichtlich der Länge, Aufgabe und Vorkommen und haben ausgenommen von den zwei spezifischen H/ACA-Boxen keine Gemeinsamkeiten (Tab.: 4.2).

SnoRNA E2 ist insgesamt 154 Nukleotide lang und hat als Aufgabe die Modifikation der 28S rRNA. Sie kommt in den Intronen des laminbindenden Proteins (LAMR1, Synonym: RPSA und RPS4) vor.

In den Intronen des Transkriptionsfaktors EIF4A2 kommt die snoRNA E3 vor. Diese snoRNA hat eine Länge von 135 nt und modifiziert, genauso wie snoRNA U17/E1, die 18S rRNA.

Neben den beschriebenen Unterschieden (Tab.: 4.2) der snoRNA Typen weisen alle drei snoRNAs Unterschiede in ihrer Evolution auf. SnoRNA E2,

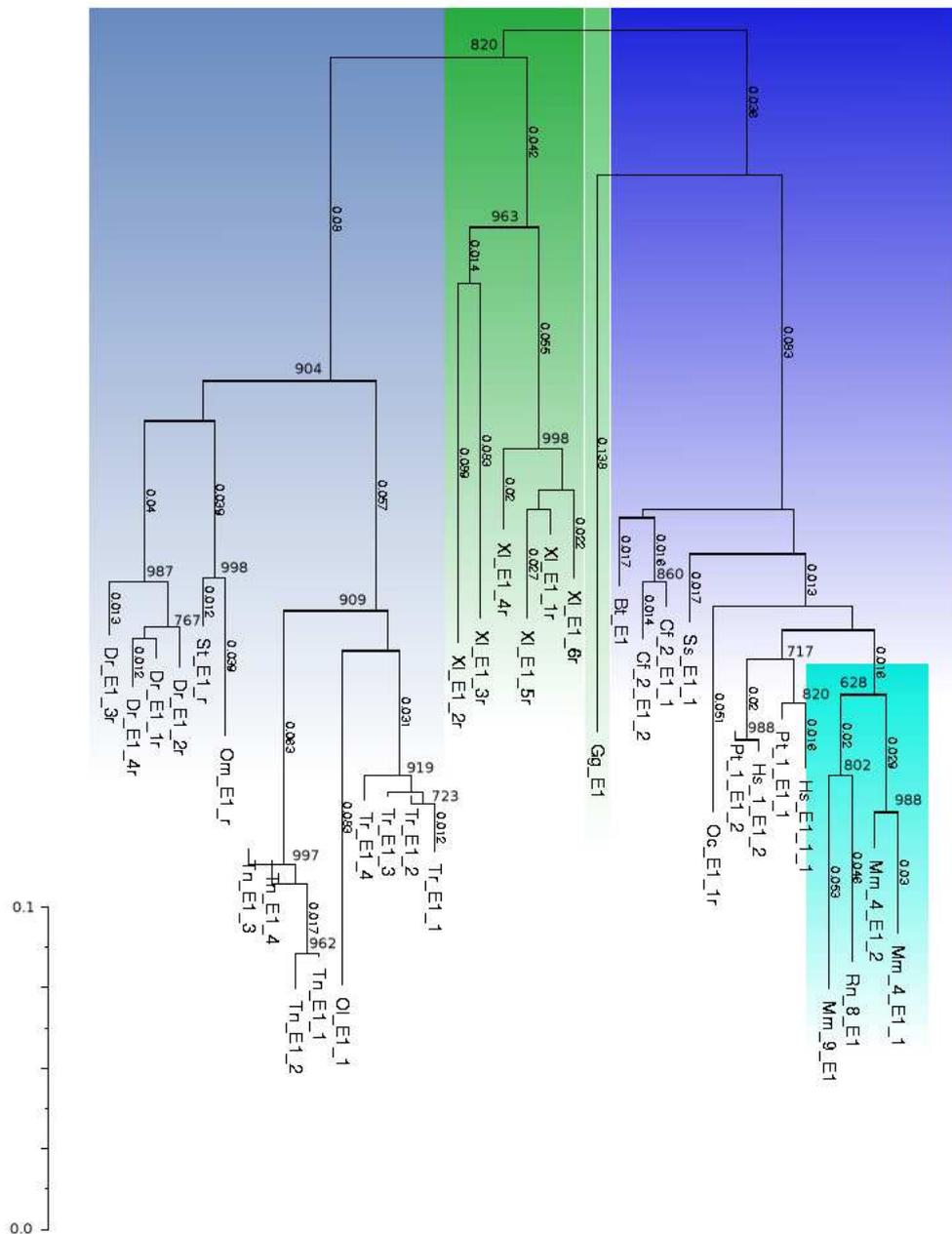


Abbildung 4.4: Stammbaumrekonstruktion der snoRNA U17/E1 erstellt mit der Neighbor-Joining Methode im Programm Phylip. Die Astlängen entsprechen der evolutionären Distanz. Distanzen wurden im Programm dnadist ermittelt. Die Werte an den Verzweigungen stellen Bootstrapwerte dar, die nach 1000 Replikationen erhalten wurden. Die unterschiedlich farblichen Boxen zeigen die unterschiedlichen Vertebratengattungen auf. Besonders hervorgehoben wurde die Gruppierung der Nagetiersequenzen, die auf das Gen der snoRNA E2 gesprungen sind (hellblaues Rechteck) Hs: Homo sapiens, Pt: Pan troglodytes, Mm: Mus musculus, Rn: Rattus norvegicus, Oc: Oryctolagus cuniculus, Cf: Canis familiaris, Bt: Bos taurus, Ss: Sus scrofa, Gg: Gallus gallus, Xl: Xenopus laevis, Dr: Danio rerio, Tr: Takifugu rubripes, Tn: Tetradon nirgovirdis Ol: Orzyias latipes, Om: Onchorhynchus mykiss, St: Salmo trutta.

SnoRNA	Aufgabe	Vorkommen	Länge
E1	Modifikation von 18S rRNA	RCC1, bzw. RPS7	207 nt
E2	Modifikation von 28S rRNA	LAMR1	154 nt
E3	Modifikation von 18S rRNA	EIF4A2, bzw. RPLP0	135 nt

Tabelle 4.2: Auflistung der Unterschiede der snoRNA E1, E2 und E3 in Bezug auf Vorkommen, Länge und Aufgabe.

snoRNA E3 und snoRNA U17/E1 (Kap.: 4.2.1) wurden alle in ihrer Evolution von den Teleosten bis zu den Primaten verglichen (Abb. 4.5).

Wie bei der Evolution der snoRNA U17/E1 beinhaltet auch die Evolution der snoRNAs E2 und E3 neben einer Serie von Duplikationen und Verlusten von Duplikaten ebenfalls genetische Sprünge zwischen verschiedenen Hostgenen.

Evolution der snoRNA E2

Die evolutionäre Geschichte der snoRNA E2 enthält eine Reihe von Duplikationen und Sprüngen der snoRNA-Duplikate zwischen den einzelnen Intronen des Hostgenes LAMR1 (Abb.: 4.5).

Der Pufferfisch *Takifugu rubripes* besitzt eine Kopie der snoRNA E2 im Intron 3 des Hostgenes. Zwischen dem Zebrafisch und dem Pufferfisch kommt es zur ersten Duplikation und zu Sprüngen zwischen den Intronen des Hostgenes. Das Gen LAMR1 des *Danio rerio* hat zwei Duplikate der snoRNA E2 in den Intronen 1 und 2 (Abb.: 4.6). Diese zwei Kopien wechselten wiederum ihre Position beim *Xenopus tropicalis*. Im Gen LAMR1 sind beim *Xenopus tropicalis* und beim *Gallus gallus* die snoRNA E2 in den Intronen 3 und 4 wieder zu finden.

Ein erneuter Positionswechsel der *Gallus gallus* snoRNA E2, lokalisiert auf Intron 3 des Hostgenes, fand zwischen *Gallus gallus* und *Canis familiaris* statt. Der Hund besitzt zwei Kopien der snoRNA E2 in den Intronen 2 und 4 des Hostgenes LAMR1.

Nach der Evolution des Hundes kam es zu einer Aufspaltung zwischen den Rodentia und den Primaten. An den Genomen von *Rattus norvegicus* und *Mus musculus* sind jeweils zwei Duplikate der snoRNA E2 und je eine Kopie der snoRNA U17/E1 zu finden. Bei der phylogenetischen Analyse der U17/E1 snoRNA (Abb. 4.4) ist ersichtlich, dass die *Mus musculus*-Sequenzen, lokalisiert auf dem Trägergen der snoRNA E2, verwandt mit den U17 snoRNA

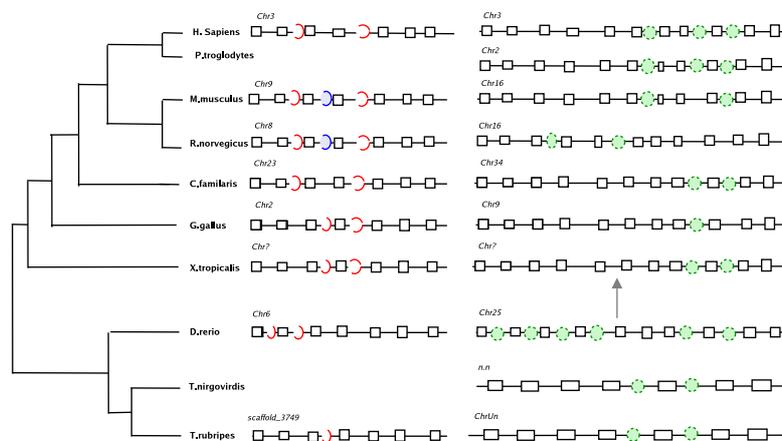


Abbildung 4.5: Evolution der snoRNA E2 und snoRNA E3 an den Organismen *H. sapiens*, *M. musculus*, *R. norvegicus*, *C. familiaris*, *G. gallus*, *X. tropicalis*, *T. rubripes*, *T. nigrovirdis* und am *D. rerio*. In der ersten Spalte wird die Evolution der snoRNA E2 dargestellt, in der zweiten die Evolution der snoRNA E3. Exone sind durch Rechtecke dargestellt, grüne Kreise stellen snoRNA E3 dar. Blaue Kreise snoRNA U17/E1, rote Kreise snoRNA E2.

Kopien, lokalisiert auf Trägern RCC1, sind (Bootstrapwert: 628). Die gesprungenen Sequenzen der snoRNA (*Ms_9_X1*, *Rn_8_X1*) sind homolog zu den *M. musculus* Duplikaten auf dem Hostgen *RCC1*. Zwei Kopien der snoRNA E2 in den Intronen 2 und 4 und eine Kopie der snoRNA U17/E1 im Intron 3 des Hostgenes existieren in den Nagetieren (*M. musculus* und *Rattus norvegicus*).

Die Kopie der snoRNA U17/E1 fehlt bei den Primaten (*H. Sapiens*). Die Primaten besitzen je zwei Kopien der snoRNA E2 in den Intronen 2 und 4 des Hostgenes *LAMR1*. Die Verzweigungen der Säugetieren gelten mit ihren Bootstrapwerten von 923 (Kopie 2) und 750 (Kopie 1) als statisch sicher. Am *Pan troglodytes* konnte keine Aussage über das Vorhandensein der snoRNA getroffen werden, da die zweite Kopie der E2 snoRNA am Anfang eines scaffolds (Tabelle: 4.1) sitzt.

Evolutionäres Szenario der snoRNA E3

Die Evolution der snoRNA E3 verlief im Gegensatz zur snoRNA E2 völlig anders. Zwei Kopien der snoRNA E3 befinden sich in den Pufferfischen

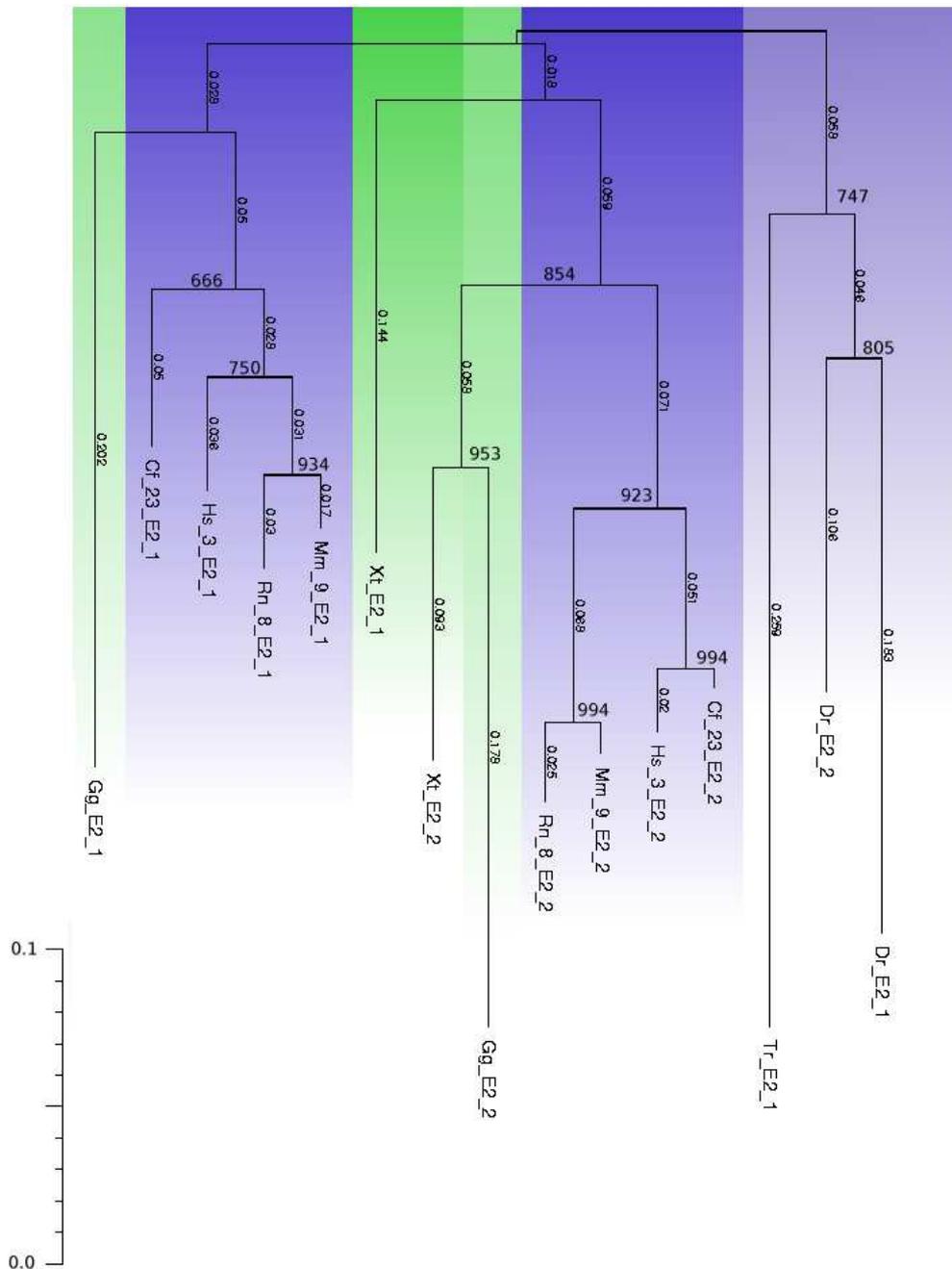


Abbildung 4.6: Stammbaumrekonstruktion der snoRNA E2 erstellt mit der Neighbor-Joining Methode im Programm Phy1ip. Die Astlängen entsprechen der evolutionären Distanz. Distanzen wurden im Programm dnadist ermittelt. Die Werte an den Verzweigungen stellen Bootstrapwerte dar, die nach 1000 Replikationen erhalten wurden. Die unterschiedlich farblichen Boxen zeigen die unterschiedlichen Vertebratengattungen auf. Hs: Homo sapiens, Pt: Pan troglodytes, Mm: Mus musculus, Rn: Rattus norvegicus, Cf: Canis familiaris, Gg: Gallus gallus, Xt: Xenopus laevis, Dr: Danio rerio, Tr: Takifugu rubripes, Tn: Tetradon nirgovirdis.

auf dem Hostgen RPLP0 (Abb.: 4.5). Am *Takifugu rubripes* und *Tetradon nigrovirdis* liegen je zwei Kopien der E3 vor. Bei beiden Duplikaten des *T. rubripes* ist ersichtlich, dass beide auf das Genom des *Tetradon nigrovirdis* dupliziert sind (Abb.: 4.7). Beide Duplikate zeichnen sich durch einen hohen Bootstrapwert von 885 bzw. 966 aus sowie der Homologisierungsgrad sehr hoch ist. Sechs Duplikate wurden auf dem *Danio rerio* gefunden. Beim phylogenetischen Vergleich der Kopien zeigte sich, dass Kopie E3_1, E3_3 und E3_5 Duplikate (Bootstrapwert: 793) voneinander sind, ebenso Kopie E3_4 und Kopie E3_6 (Bootstrapwert: 997).

Zwischen der Evolution des Zebrafisches und den Fröschen kam es zum Wechsel des Hostgenes, auch in der Evolution der snoRNA E3 wurde der ursprüngliche Zustand gewechselt. Insgesamt zwei Kopien der snoRNA E3 werden am Genom des *Xenopus tropicalis* auf dem Gen EIF4A2 gefunden. Die snoRNAs sind auf dem Gen EIF4A2 in den Intronen 8 und 9 lokalisiert.

Auf dem Hostgen EIF4A2 des *Gallus gallus*-Genom kann lediglich nur im Intron 8 eine Kopie der snoRNA E3 gefunden werden. Beim *C. familiaris* kommen jedoch wieder zwei Kopien der snoRNA E3 in den Intronen 8 und 9 des Hostgenes EIF4A2 vor. Diese zwei Kopien haben phylogenetisch keine Gemeinsamkeiten. Cf_34_E3_1 ist eine Kopie der *G.gallus* Sequenz, die auf dem Intron 8 lokalisiert ist. Die zweite Kopie muss ein Paralog der *Gallus gallus*-Sequenz sein. Ab dem *Canis familiaris* liegt eine klare Trennung zwischen Duplikat 1 und 2 im phylogenetischen Baum vor (Abb.: 4.7).

In der Entwicklung von *Canis familiaris* und *Rattus norvegicus* kommt es des Weiteren zu einem Sprung der snoRNA von Intron 8 und 9 in Intron 3 und 5 des Hostgenes EIF4A2. Die *M. musculus*-Kopien kommen allerdings wieder in den Intronen 6 und 9 des Genes EIF4A2 vor. Daher kann angenommen werden, dass die Kopien des *R. norvegicus* ebenfalls auf den Intronen 6 und 9 liegen. Meiner Vermutung nach handelt es sich hierbei um einen Annotationsfehler des Genome Browser.

In der Evolution von Maus und Schimpanse kommt es nochmals zu einer weiteren Duplikation. Die Primaten besitzen drei Kopien in den Intronen 6, 8 und 9.

4.2.3 Evolution der snoRNA U36

Familie U36, eine Box C/D snoRNA, kommt auf den Intronen des ribosomalen Proteins RPL7a in mehreren Kopien vor. Die Kopien der snoRNA U36 liegen auf den Introns 4 (U36b) und 5 (U36a) des ribosomalen Proteins. Des

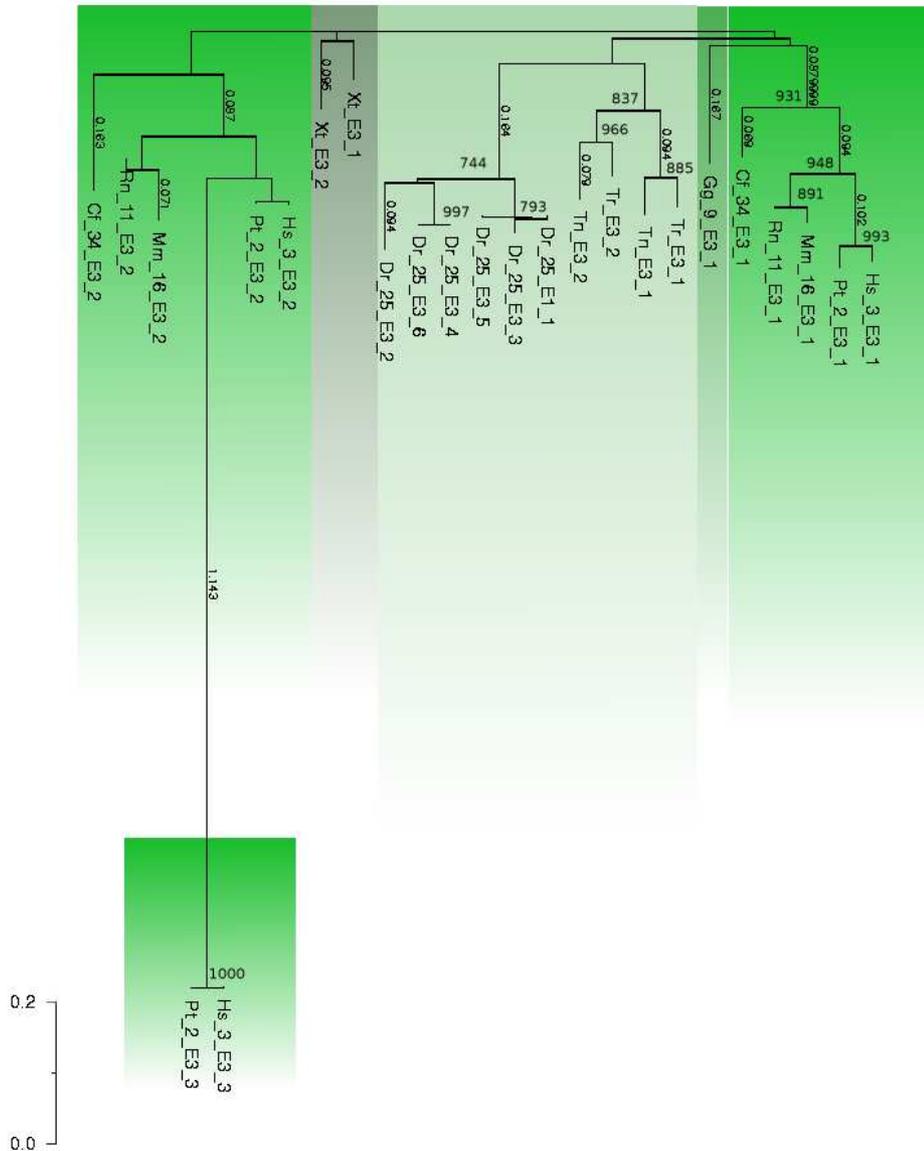


Abbildung 4.7: Stammbaumrekonstruktion der snoRNA E3 erstellt mit der Neighbor-Joining Methode im Programm PhyIip. Die Astlängen entsprechen der evolutionären Distanz. Distanzen wurden im Programm dnadist ermittelt. Die Werte an den Verzweigungen stellen Bootstrapwerte dar, die nach 1000 Replikationen erhalten wurden. Die unterschiedlich farblichen Boxen zeigen die unterschiedlichen Vertebratengattungen auf. Hs: Homo sapiens, Pt: Pan troglodytes, Mm: Mus musculus, Rn: Rattus norvegicus, Cf: Canis familiaris, Gg: Gallus gallus, Xt: Xenopus tropicalis, Dr: Danio rerio, Tr: Takifugu rubripes, Tn: Tetradon nirgovidis.

Weiteren kommt eine dritte Kopie der snoRNA U36 beim Menschen und bei der Maus vor. Diese Kopie (U36c) liegt auf dem sechsten Intron des ribosomalen Proteins, RPL7a.

Das Auftauchen von mehreren snoRNA-Kopien auf ein und demselben Hostgen bzw. auf verschiedenen Introns eines Gens ist nicht selten. Dies zeigten schon die Evolutionen der snoRNA E1, E2 und E3.

SnoRNAs der Familie U36 wurden bis heute an folgenden Genomen gefunden: *Homo sapiens*, *Mus musculus*, *Gallus gallus* und *Takifugu rubripes*. Zudem ist die snoRNA U36 homolog zur snoRNA R47. SnoRNA R47 konnte an den Genomen *Oryza sativa* und *S. cerevisiae* entdeckt werden.

Evolutionsszenario von Gilley und Fried

Jonathan Gilley und Mike Fried konnten bei ihrer Suche nach U36 snoRNA die Homologen der *Takifugu rubripes* U36 snoRNA ausfindig machen [34]. Am *T. rubripes* Genom existieren zwei Kopien der U36 snoRNA in den Introns des *T. rubripes* RPL7a Gens. Die Kopienanzahl der snoRNA U36 kann auf dem Gen RPL7a unterschiedlich sein. *H. sapiens* und *M. musculus* erhalten je drei U36 snoRNA Kopien auf den Introns 4, 5 und 6. Am Genom des *G. gallus* und *T. rubripes* kommen allerdings nur zwei Kopien auf den Introns 4 und 5 vor.

Entsprechend der Eigenschaft der Box C/D snoRNAs enthalten alle Kopien der U36 Bereiche, die komplementär zur rRNA sind. SnoRNA U36 hat als Besonderheit, dass der erste ribosomal komplementäre Bereich 28S rRNA binden kann sowie der zweite Bereich komplementär zur 18S rRNA ist. Diese Eigenschaft tritt bei den Kopien des *T. rubripes* und bei den Kopien des *G. gallus* auf. Am Genom der Rodentia und Säugetiere tritt nur an einer Kopie (U36b) Komplementarität zur 18S rRNA und 28S rRNA auf. Kopien U36a und U36c haben jeweils nur zu einer rRNA Komplementarität. SnoRNA U36a kann nur an die 18S rRNA binden, U36c kann nur 28S rRNA binden.

Während der Evolution der snoRNA U36 kam es bei einem gemeinsamen Vorfahren von Ratte und Mensch zu einer Genduplikation. Dieser gemeinsame Vorfahre, der je eine Kopie der U36a und U36b snoRNA besitzt, duplizierte die Kopie der U36a snoRNA [34]. Die Duplikation der snoRNA U36 lokalisierte sich auf dem nächsten Intron des Gens RPL7A (Abb.: 4.8).

Die Duplikation erhöhte die Menge an 18S und 28S rRNA die modifiziert werden konnte. Um das ursprüngliche Gleichgewicht an modifizierter rRNA zu erhalten, mutierte je eine rRNA-Bindungsstelle der Kopie U36a und U36c

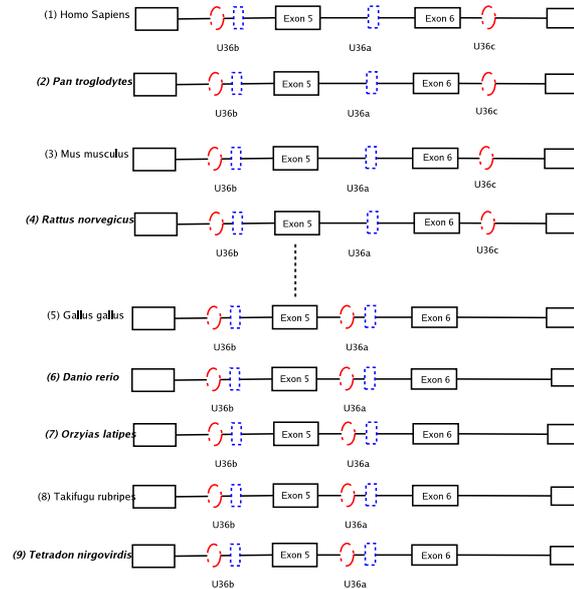


Abbildung 4.8: Szenario, dass die Evolution der snoRNA U36 vom den Teleosten bis zu den Primaten beschreibt. Die unterschiedlichen reverse komplementären Bereichen werden bei 18S rRNA rot, bei 28S rRNA blau dargestellt.

snoRNA, so dass nur eine Kopie in der Lage war, entweder 18S rRNA oder 28S rRNA zu binden. Die Mutation führte dazu, dass die Kopie U36a keine 28S rRNA modifizieren kann und U36c keine 18S rRNA [34].

Erweitertes Evolutionsszenario

Durch den von uns durchgeführten Homologievergleich mit der *T. rubripes*-Sequenz konnten jeweils eine *T. nigrovirdis*, *D. rerio*, *O. latipes* Sequenz für die Kopien der snoRNA U36a und snoRNA U36b gefunden werden. Die Konservierung dieser Sequenzen wird in Abb.: 4.9 und in Abb.: 4.10 dargestellt. Auch hier zeigen die Sequenzen die nicht übliche H-Box Konservierung. Bei beiden Kopien der Familie U36 (U36a und U36b) ist das sechste Nukleotid gleichfalls zur snoRNA U24 (Kapitel: 3.3.2) mutiert.

Bei den Sequenzen aller drei Genome kommt es ebenfalls zu einer Differenzierung der ribosomalen Targets. Die beiden Kopien U36a und U36b der Organismen können ebenso wie die *T. rubripes*-Sequenz 28S rRNA und 18S

```
#####          CCCCC  XXXXXXXXXXXXX          XXXXXXXXXXXXD
01U36b_1        TGATGTTTCAGAATTTCTTCACCTGAAACTTCTATGATGTTTTGAGCTTTTAAACC
01U36b_2        TGATGTTTCAGAATTTCTTCACCTGAAACTTCTATGATGTTTTGAGCTTTTAAACC
01U36b_3        TGATGTTTCAGAATTTCTTCACCTGAAACTTCTATGATGTTTTGAGCTTTTAAACC
Dr_U36b        TGATGTTCT-GAATTTCTTCACCTGAAACAACCATGTGGTT-TAAATGAGCTTTTAAACC
Tr_U36b        TGATGACTTGAATTTCTTCACCTGAAACCTTCAGTGATGTTTTGAGCTTTTAAACC
TnU36b         TGATGACTTGAATTTCTTCACCTGATGGCTTCAGTGATGCGTATTGAGCTTTTAAACC
*****
#SecStr        .....(((.....))).....
#####          DDD
01U36b_1        TGA
01U36b_2        TGA
01U36b_3        TGA
Dr_U36b        TGA
Tr_U36b        TGA
TnU36b         TGA
***
#SecStr        ...
```

Abbildung 4.9: Zusammenfassung aller U36b-Sequenzen. Alignment zeigt Consensusstruktur aller Sequenzen. Erste Zeile gibt komplementäre ribosomale Stellen an (X), Box C (C) und Box D (D) an.

rRNA binden.

Neben den Homologen am Fisch wurde noch nach weiteren Homologen bei

```
#####          CCCCC  XXXXXXXXXXXXX
TrU36a         TGATGT-TTGAATTTCTTCACCTGAAAGTCTCTATGTAGATCAAAACAGACGCGCTTTTA
TnU36a         TGATGT-TTGAATTTCTTCACCTGAAAGTCACTGTGTAGATCAAAACAGACGAGCTTTTA
01U36a         TGATGT-CTGAATTTCTTCACCTGAAATCCTCCGTGGATCCAAACATACGAGCTTTTA
DrU36a         TGATGTCTTGAATTTCTTCACCTGAAATCAAACTGAAGATCAAAAT--ACGAGCTTTTA
*****
#SecStr        ...(((.....(((.....))).....))).....
#####          DDDD
TrU36a         ACCCTGA
TnU36a         ACCCTGA
01U36a         ACCCTGA
DrU36a         ACCCTGA
*****
#SecStr        .....
```

Abbildung 4.10: Zusammenfassung aller U36a-Sequenzen. Alignment zeigt Consensusstruktur aller Sequenzen. Erste Zeile gibt komplementäre ribosomale Stellen an (X), Box C (C) und Box D (D) an.

den Säugetieren gesucht, um den Sprung von zwei Duplikaten beim *G. gallus* zu drei Duplikaten bei der Maus näher eingrenzen zu können.

Über den **Ensembl Genome Browser** [84] wurden Homologe am *Canis familiaris*, *Rattus norvegicus*, *Pan troglodytes* gesucht; über NCBI Blast [2] wurden die Genome von *Ovis aries*, *Felis catus* und *Sus scrofa* untersucht. Jeweils drei homologe Kopien der snoRNA U36 (U36a - c) konnten lediglich am Schimpansen-Genom und an der Ratte gefunden werden. Die Homologen an den anderen Genomen (*C. familiaris*, *S. scrofa*, *F. catus* und *O. aries*) konn-

ten, aufgrund der noch nicht vollständig sequenzierten Genome (Tabelle: 4.1) nicht gefunden werden.

Auch diese Sequenzen wurden phylogenetisch untersucht. Bei der Analyse des Baumes konnten Entwicklungslinien erkannt werden (Abb.: 4.11). Laut dem phylogenetischen Baum hat sich aus der Kopie A die dritte Kopie an den Nagetieren und Primaten (Kopie C) entwickelt. Die *G. gallus* Kopie B, die nahe verwandt mit der Kopie B des *Danio rerio* ist, führte zur Ausbildung der Kopie B bei den Säugetieren; Kopie A des *G. gallus* führte zur Entwicklung der Kopie A bei den Säugern.

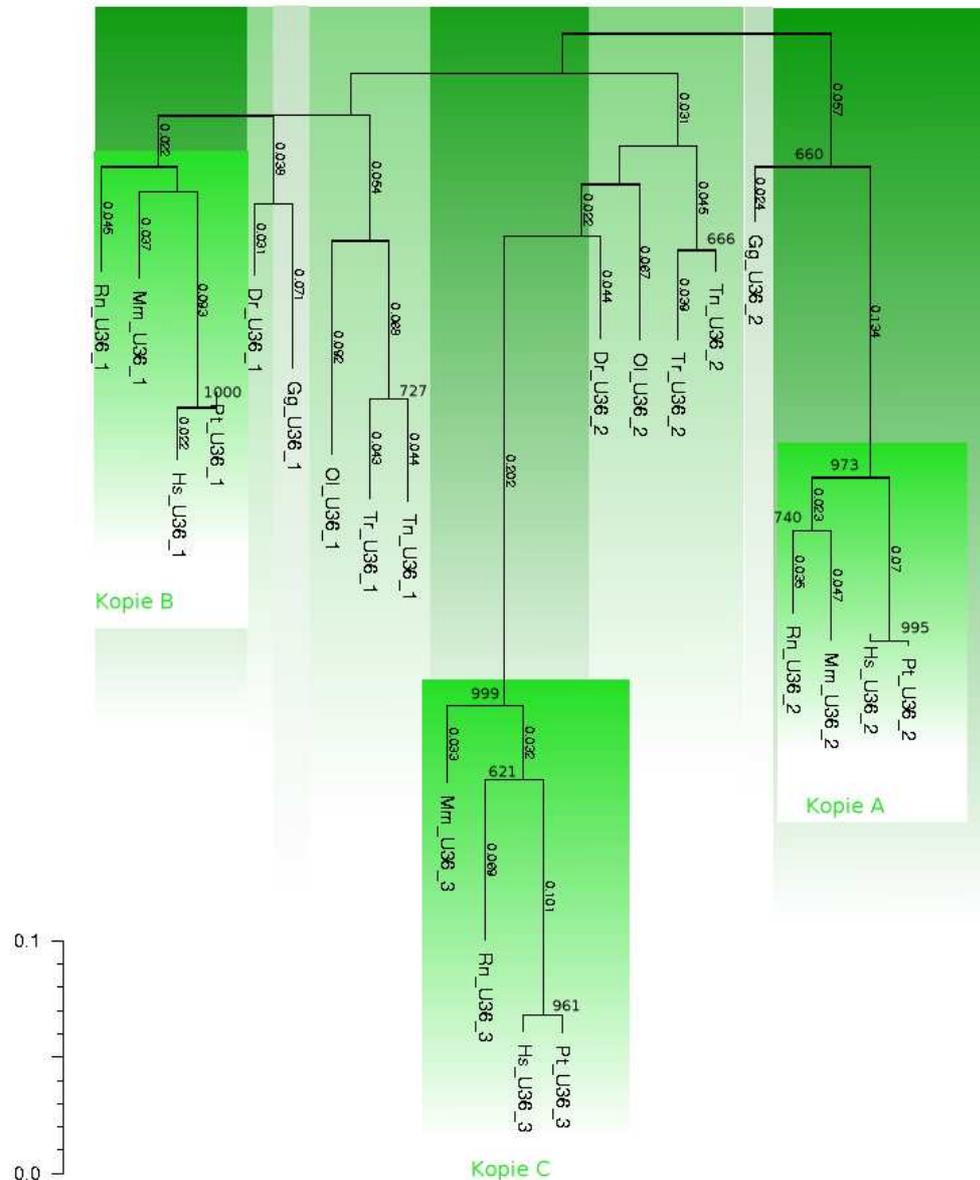


Abbildung 4.11: Stammbaumrekonstruktion der snoRNA U36 erstellt mit der Neighbor-Joining Methode im Programm Phylip. Die Astlängen entsprechen der evolutionären Distanz. Distanzen wurden im Programm dnadist ermittelt. Die Werte an den Verzweigungen stellen Bootstrapwerte dar, die nach 1000 Replikationen erhalten wurden. Die unterschiedlich farblichen Boxen zeigen die unterschiedlichen Vertebratengattungen auf. Hs: Homo sapiens, Pt: Pan troglodytes, Mm: Mus musculus, Rn: Rattus norvegicus, Gg: Gallus gallus, Ol: Orzyias latipes, Tr: Takifugu rubripes, Tn: Tetradon nirgovirdis.

Kapitel 5

Diskussion und Ausblick

5.1 Diskussion der Ergebnisse

Ein Ziel dieser Diplomarbeit war neue snoRNAs in den Vertebraten zu finden. Insgesamt 80 snoRNAs wurden an unterschiedlichen Genomen gefunden. Bei der Suche wurde insbesondere Augenmerk auf die Teleosten-Genome geworfen. Hier konnten allein 55 snoRNAs ermittelt werden.

Die 80 snoRNAs wurden durch einen Homologievergleich bekannter snoRNAs gegen die jeweiligen Genome gefunden.

Zur Bestimmung der Richtigkeit wurden drei Kriterien angewendet:

1. Sequenzähnlichkeit
2. Genomischer Kontext
3. Bestimmung der ribosomalen Targets.

Zu Beginn der Arbeit war über die Evolution der snoRNA bekannt, dass sich snoRNAs im Laufe der Zeit dupliziert haben, mutiert und zwischen einzelnen Hostgenen gesprungen sind.

Da bis heute die Evolution der snoRNAs in einzelnen Gattungen der Vertebraten, wie z. B. den Schildkröten näher untersucht wurden, war ein weiteres Ziel der Diplomarbeit die Evolution der Vertebraten von den Säugern bis zum Ursprung zurück zu verfolgen. Vier unterschiedliche snoRNAs wurden für die Suche ausgewählt: U17/E1, E2, E3, U36.

Mit Hilfe von phylogenetischen Bäumen wurde die Evolution der vier snoRNAs anhand ihrer Homologen rekonstruiert. Die Evolution der U17/E1 war zu Teilen am Anfang der Arbeit bekannt [16]. Im Laufe der Evolution der

U17/E1 kam es zum Wechsel des Hostgenes zwischen *X. laevis* und *G. gallus*. Nach Änderung der genomischen Organisation lag ein Homolog der U17 im Genom des *G. gallus* vor. Da es zwischen dem *G. gallus* und *M. musculus* zu einer Duplikation gekommen ist, wurde versucht den Sprung von einem Homolog auf zwei Homologe näher einzugrenzen. Je eine Kopie der snoRNA U17 kommt am *B. taurus* und beim *S. scrofa* vor, zwei Kopien sind allerdings schon beim *C. familiaris* vorhanden. Durch phylogenetische Analysen lässt sich der Zeitraum der Duplikation näher eingrenzen. Sie muss zwischen *Sus scrofa* und *Canis familiaris* stattgefunden haben.

Ein zweites Evolutionsszenario wurde mit der snoRNA E2 aufgestellt. Dieses Szenario stellte mich am Anfang vor ein großes Rätsel. Ausgehend von den Pufferfischen besitzt das Gen LAMR1 eine homologe Kopie der snoRNA bis es zur Duplikation der snoRNA im *Danio rerio* kommt. Diese zwei Homologen verteilen sich bis zum *Canis familiaris* auf zwei Introns des Genes. Bei der Suche nach Homologen der Nagetiere konnte anfänglich nur eine Kopie der snoRNA gefunden werden. Dies gab Rätsel auf, was nun im Laufe der Evolution passiert sei, da bei den Primaten wieder zwei homologe Kopien gefunden wurden. Es stellte sich heraus, dass die homologe Kopie an den Nagetieren eine Kopie der U17/E1 snoRNA ist und zwei Kopien der snoRNA E2 in den Intronen des Gen LAMR1 zu finden sind. Dieser Fund zeigt, dass zum einen die snoRNA genetisch mobil sein kann. Zum anderen muss es in der Evolution der E2 zwischen den Primaten und den Nagetieren zu einer Spaltung gekommen sein.

Das Szenario der snoRNA E3 Evolution ist ein Szenario, das bestimmt ist von Duplikationen, Verlusten und vom Wechsel des Hostgenes. Während der Evolution kam es zwischen den Teleosten und Reptilien zum Wechsel des Hostgenes und zur Änderung der genomischen Organisation. Ab dem *X. tropicalis* kommen die homologen Kopien der snoRNA E3 auf dem Transkriptionsfaktor EIF4A2 vor, bei den Teleosten liegen sie auf dem ribosomalen Protein RPL0.

Die Evolution zeigte auch, dass nicht unbedingt beide Kopien auf ein anderes Genom duplizieren müssen. Von den Sequenzen des *X. tropicalis* duplizierte nur eine auf das Genom des *G. gallus*.

Im letzten Evolutionsszenario wurde ein Blick auf die snoRNA U36 geworfen. Die Evolution der snoRNA verlief, wie von Gilley und Fried festgestellt, besonders interessant, da ab den Nagetieren drei Kopien der snoRNA existierten. Diese drei Kopien veränderten ihre ribosomalen Targets so durch Mutation, dass das Gleichgewicht der ribosomalen Bindungstarget zwischen

Teleosten, Reptilien und Säugetieren wieder hergestellt wurde. Auch hier wurde versucht, den Sprung der Duplikate von zwei Homologen auf drei Homologe näher einzugrenzen. Es konnten allerdings keine Kopien an *C. familiaris*, *S. scrofa* und am *B. taurus* gefunden werden. Die phylogenetische Analyse zeigte die Entwicklungslinien der homologen Kopien. SnoRNA U36a, führte wahrscheinlich zur Ausbildung der Kopien U36c in den Säugetieren, die *Gallus gallus* Kopie U36a führte, zur Ausbildung der Kopien U36a und die *Gallus gallus* Kopie U36b führte zur Ausbildung der Kopie U36b in den Säugetieren.

Durch die vier Stammbäume der unterschiedlichen snoRNA konnte festgestellt werden, dass es im Laufe der Evolution der snoRNA in den einzelnen snoRNA-Familien zu keiner zeitgleichen Evolution gekommen ist. Alle vier Evolutionen zeichnen sich jedoch durch die gleichen Mechanismen aus: die Duplikation, Mutation und den Wechsel des Hostgenes.

Des Weiteren zeigte sich, dass im Laufe der Evolution zwei verschiedene snoRNAs, die unterschiedlichste Funktionen haben, auf dem gleichen Hostgen vorkommen können.

5.2 Ausblick

Während dieser Arbeit entstanden immer neue Fragen, die in ferner Zukunft gelöst werden sollten. Unterliegt die Evolution der snoRNA auch dem Mechanismus der konzentrierten Evolution? Wie verhalten sich die umliegenden Bereiche der Hostgene? Heterogen oder Homogen?

Wie verlief nun die Evolution der snoRNA? Entwickelten sich die snoRNAs aus Proteinen oder entwickelten sie sich auch aus einer RNA-Welt, die bei den Bakterien durch Proteine ersetzt wurde? Diese Frage ist sicherlich, eine der spannendsten Fragen im Bereich der snoRNA!

Eine weitere Frage, die sich während des Schreibens stellte: Wie sind die Verwandtschaftsverhältnisse der einzelnen snoRNA-Familien zueinander? Haben E1, E2, E3 aufgrund ihrer Benennung doch etwas gemein?

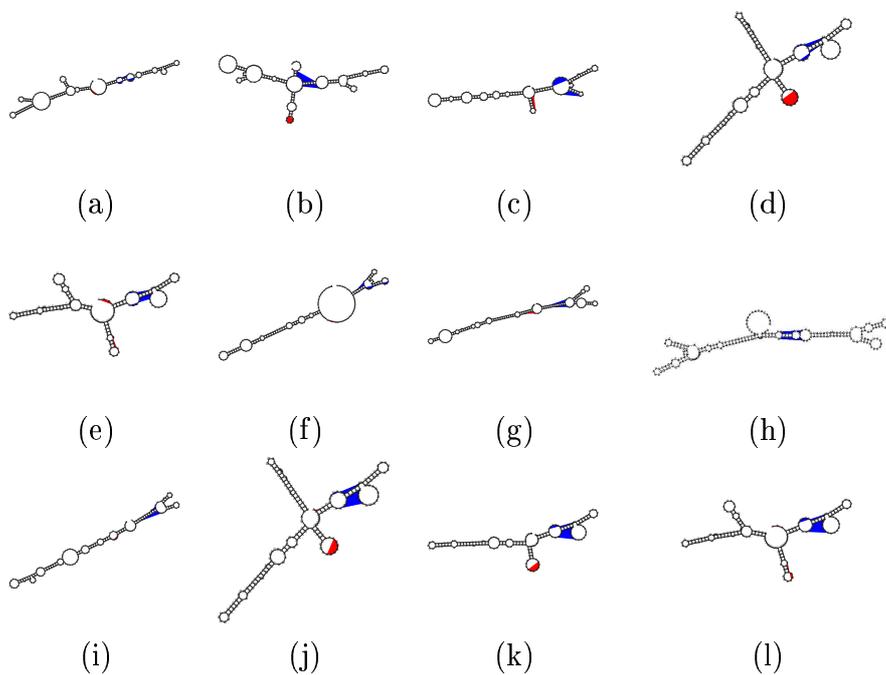
Ersichtlich ist, dass im Bereich der snoRNA noch einige Fragen zu lösen sind.

Anhang A

Sequenzen und Strukturen

A.1 Sekundärstrukturen

Abbildung A.1: 20 möglich gefundene *X. tropicalis* snoRNAs mittels Programm *fisher*. Rot markierter Bereich stellt die Boxen H bzw. ACA dar, blau markierter Bereich ψ 3/4-Bereiche.



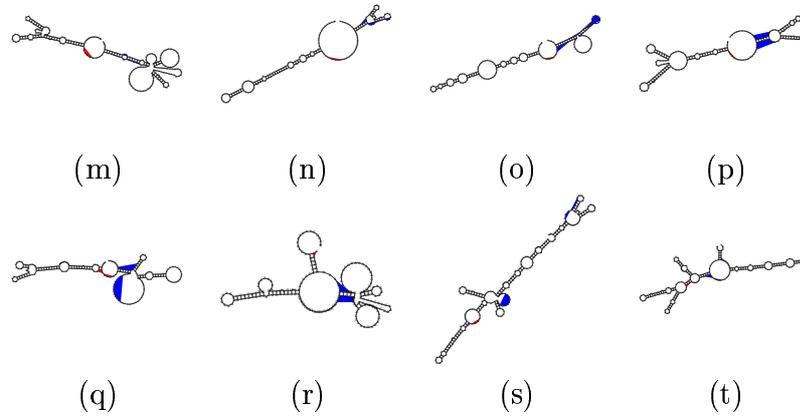


Abbildung A.2: Homologien der snoRNAs zu folgenden ribosomalen Proteinen

Ribosomales Protein, snRNA	snoRNA
RPS1	(a) (f) (n) (r) (t)
RPS8	(d) (g) (k) (l) (q)
RPL14	(e) (i) (j) (k) (l) (o) (s)
U5 snRNP	(b)
U7 snRNA	(o)
ELrB	(q)
A2B1	(h)

A.2 Sequenzen

Tabelle A.1: Sequenzen der U17/E1 snoRNA

```
>Tr_E1_1|X94942|1504-1721
CTGACGTGGACATCCAGCGAGCTGCTCTCCTGGCTCTGTCCTGGTGGCGTATGGGAGCGTAG
CTTTTCAGCAGAGACGTACATTCCCGTTCACACCGTCAGAGAGAAAGCTCTGTCCCCGGCC
TCTGGCTGGCGTGGACGAGCGTCTGTATATTCCTACATCTTCCCAGAGTCCTACTGGACCG
TTACTGGGGAGCTAAACCATGCAGTGCACAT
>Tr_E1_2|X94942.1|2086-2301
CCGACGTGGACATCCAGCGAGCTGCTCTCCTGGCTCTGTCCTGGTGGCGTATGGGAGCGTAG
CTTTTCAGCAGAGACGTACATTCCCGTTCACACCGTCGGAGAGAAAGCTCTGTCCCCGGCCT
CTGGCTGGCGTGGACGAGCGTCTGTATATTCCTACATCTTCCCAGAGTCCTACTGGAACCGTT
ACTGGGGAGCTAAACCATGCAGTGCACAT
>Tr_E1_3|X94942.1|2747-2965
CCGACGTGGACATCCAGCGAGCTGCTCTCCTGGCTCTGTCCTGGTGGCGTATGGGAGCGTAG
CTTTTCAGCAGAGACGTACATTCCCGTTCACACCGTCGGAGAGAAAGCTCTGTCCCCGGCC
TCTGGCTGGCGTGGACGAGCGTCTGTATATTCCTACATCTTCCCAGAGTCCTACTGGGCCG
TTACTGGGGAGCTAAACCATGCAGTGCACAT
>Tr_E1_4|X94942.1|3702-3921
CCGACGTGGACATCCAGCGAGCTGCTCTCCTGGCTCTGTCCTGGTGGCAGTATGGGAGCGTA
GCTTTTCAGCAGAGACGTACATTCCCGTTCACACCGTCGGAGAGAAAGCTCTGTCCCCGGC
CTCTGGCTGGGGTGGACGAGCGTCTGTATATTCCTACATCTTCCCAGAGTCCTACTGGACC
GTTACTGGGGAGCTAAACCATGCAGTGCACAA
>Bt_E1|BM030695.1|271-471
CCAACGTGGATAACCCGGGAGGTCACTCTCCCCGGGCTCTGTCCAGGTGGCGAACGGGAGCG
TAGGGCTCTGCTCAGTGACGTACAGTCCCTTTCCACAACGTTGGAGATGAAGCTGGGCCTTG
AGTCTGCGCCTGCATATTCCTACAGCTTCTCAGAGTCTGTGGACAATGACTGAGGAGACAA
ACCATGCAGGAAACA
>Cf_2_E1_1|Chr2|72632675-72632877
TCCAACGTGGATACTCCCGGGAGGTCACTCTCCCCGGGCTCTGTCCAGGTGGCGTATGGGAG
CTCAGGGCTCTGCCCCGTGACGTACAGTCCCTTTCCACAACGTTGGAGATGAAGCCGGGCTT
GAGTCTGCGCCTGCATATTCCTACAGCTTCTCAGAGTCTGTGGACCATGACTGAGGAGACA
AACCATGCAGGAAACAT
>Cf_2_E1_2|Chr2|72633598-72633793
CCAACGTGGATACTCCCGGGAGGTCACTCTCCCCGGGCTCTGTCCAGGTGGCGTGGGGAGC
TCAGGGCTCTGCCCCGTGACGTACAGTCCCTTTCCACAACGTTGGAGATGAAGCCGGGCTT
GAGTCTGCGCCTGCATATTCCTACAGCTTCTCAGAGTCTGTGGACCATGACTGAGGAGACA
AACCATGCAGGAAACA
>Hs_E1_1_1|Chr1|28521389-28521595
TCCAACGTGGATAACCCGGGAGGTCACTCTCCCCGGGCTCTGTCCAAGTGGCGTAGGGGAG
CATAGGGCTCTGCCCCATGATGTACAAGTCCCTTTCCACAACGTTGAAATAAAGCTGGGCC
TCGTGTCTGCGCCTGCATATTCCTACAGCTTCCCAGAGTCTGTGCAACAATTACTGGGGAGA
CAAACCATGCAGGAAACAGCC
>Hs_1_E1_2|Chr1|28522579-28522786
TCCAACGTGGATAACCTGGGAGGTCACTCTCCCAGGCTCTGTCCAAGTGGCATAGGGGAGC
TTAGGGCTCTGCCCCATGATGTACAGTCCCTTTCCACAACGTTGAAGATGAAGCTGGGCCTC
GTGTCTGCGCCTGCATATTCCTACAGCTTCCCAGAGTCTGTGGACAATGACTGGGGAGACA
AACCATGCAGGAAACATAT
>Ol_E1_1
AAACCCGACGCGGACGTCCAGAGAGTTATGCTCCTGGTTCTGTCCGGTGGCGTATGGGAGC
GTAGCTTTGAGCTGAGACGTACATTCCCGTTCACAGCGTCGTACAGAAAGCTCTGCCCCCG
GCCTCGGCTGGTAGGGGCGGGCGTCTGTGATATTCCTACATCTTCCCAGAGTCTTTGGGCC
GTTACTGGGGAGCTAAACCATGCAGTGCACAT
>Mm_4_E1_1|Chr4|130937113-130937303
TCCGACGTGGATAACCCGGGCTCTGTCTAGTGGCGAGCGGGAGCTTAGGCATTGCCCGGTG
ATGTACAGTCCCTTTCCACAAAACGTTGGACATAAAGCTGGGCTTTGAGTCTGCGCCTGCAT
ATTCCTACGACTTCTCCAGAGTCTGTGGACCATGACTGAGAGACAAACCATGCAGGAAACA
GTT
```

>Mm_4_E1_2|Chr4|130937549-130937754
TCCAACGTGGATATCCCGGGAGGTCACCTCTCCCCGGGCTCTGTCCCTAGTGGCGAACGGGAGC
TTAGGGCATTGCCCGGTGATGTACAGTCCCTTTCCACAACGTTGGAGATGAAGCTGGGCTTT
CGAGTCTGCGCCTGCATATTCCTACGACTTCTCAGAGTCCCGTGGACCATGACTGAGGAGAC
AAACCATGCAGGAAACAAA

>Pt_1_E1_1|Chr1|29224911-29225116
TCCAACGTGGATACACCCGGGAGGTCACCTCTCCCCGGGCTCTGTCCAAGTGGCGTAGGGGAG
CATAGGGCTCTGCCCCATGATGTACAGTCCCTTTCCACAACGTTGAAGATGAAGCTGGGCCT
CGTGTCTGCGCCTGCATATTCCTACAGCTTCCCAGAGTCTGTGGACAACTTACTGGGGAGAC
AAACCATGCAGGAAACAGC

>Pt_1_E1_2|Chr1|29226091-29226298
TCCAACGTGGATACCCTGGGAGGTCACCTCTCCCCAGGCTCTGTCCAAGTGGCATAGGGGAGC
TTAGGGCTCTGCCCCATGATGTACAGTCCCTTTCCACAACGTTGAAGATGAAGCTGGGCCTC
GTGTCTGCGCCTGCATATTCCTACAGCTTCCCAGAGTCTGTGGACAATGACTGGGGAGACA
AACCATGCAGGAAACA

>Rn_8_E1|Chr8|125121509-125121716
TTTCCAACGTGGATAACCCGGGAGGTCACCTCTCCCCGGGCTCTGTCCCTAGTGGCGTCAGGGA
GCAGAGGCATTGCCAGTGTACAGTCCCTTTCCACAACGTTGGAGATGAAGCTGGGCTT
CGTGCTCGCACCTGCATATTCCTACAACCTTCTCAGAGTCTGTGGACCATGACTGAGGAGGC
AAACCATGCAGGAAACAGTCC

>Om_E1_r|AJ310189.1|1-225
CCAACGTGGACATCCCGGTGAGGTTGCTCTCgcATGGCTCTGTCCCTTGTGGCGTTTGGGAGCG
TAGTACTTCacaGTACTGTGACGTACAGTCCCGTTCCACGACGTTGGAGAGCAAGCtTTGTC
CCCGGCCTTCGGGCTGGTGAAGGAGGTTCTGTATATTCCTACATCTTCCCAGAGCCCTTT
GGGCCATGACTGGGGGGCTAAACCATGCAGTGTACAACC

>St_E1_r|AJ310189.1|1-223
CCAACGTGGACATCCAGTGAGGTTGCTCTCgcATGGCTCTGTCCATGTGGCGTATGGGAGCG
TAGCACTTCaGTGTAGTGAAGTACAGTCCCGTTCCACGACGTTGGAGAGCAAGCtTTGTCCC
CGGCCCTCCGGCTGGTGAAGACGGCGTCTGTATATTCCTACATCTTCCCAGAGCCCTTTGG
GCCATGACTGGGGGGCTAAACCATGCAGTGTACAAC

>Dr_E1_3r|Chr20|327261758-327261979
CCAACGTGGACATCCCTGTGAGATGCCTCTCatGTGGCTCTGTCCGTGTGGCGTATGGGAGCG
TAGCACTTCaGTGTTGTGACGTACAGTCCCGTTCCATGACGTTGGATAGCAAGCTTGTCCCC
GGCCTCTGGCTGGTGTGGAAGTGCCTGTATATTCCTACATCTTCCCAGAGCCCTTTGGGC
CATTACTGGGGAGGTAACCATGCAGTGCACATAA

>Dr_E1_4r|Chr20|327262302-327262523
CCAACGTGGACATCCCGTGAGATGCCTCTCatGTGGCTCTGTCCATGTGGCGTATGGGAGCG
TAGCACTTCgGTGTTGTGACGTACAGTCCCGTTCCACGACGTTGGATAGCAAGCTTGTCCCC
GGCCTCTGGCTGGTGTGGAAGTGCCTGTATATTCCTACATCTTCCCAGAGCCCTTTGGGC
CATTACTGGGGAGGTAACCATGCAGTGCACACT

>Dr_E1_1r|Chr20|327260279-327260500
CCAACGTGGACATCCCTGTGAGATGCCTCTCatGTGGCTCTGTCCGTGTGGCGTATGGGAGCG
TAGCACTTCgGTGTTGTGACGTACAGTCCCGTTCCATGACGTTGGATAACAAGCTTGTCCCC
GGTCTCTGGCTGGTGTGGAAGTGCCTGTATATTCCTACATCTTCCCAGAGCCCTTTGGGC
CATTACTGGGGAGGTAACCATGCAGTGCACAACC

>Dr_E1_2r|Chr20|327261101-327261321
CCAACGTGGACATCCCTGTGAGATGCCTCTCatGTGGCTCTGTCCGTGTGGCGTATGGGAGCG
TAGCACTTCgGTGTTGTGACGTACAGTCCCGTTCCATGACGTTGGATAGCAAGCTTGTCCCC
GGCCTCTGGCTGGTGTGGAAGTGCCTGTATATTCCTACATCTTCCCAGAGCCCTTTGGGC
ATTACTGGGGAGGTAACCATGCAGTGCACAGCC

>Xl_E1_3r|X71801.1|4425-4646
CCAATGTGGATATCCCTAGAGGTCTCTTTGGGCTCTGTCCAGGTGGCGTATGGGAGCATAat
aGCCCTTaaagGTGCAAGTACGTATAATCCCCTTCCACAACATTGGAGACAAGCATGTCCC
CAGCCATTAGGCGGGTGTGGCAGCATCCTGTATATTCCTAAAACCTTCCCAGAGCCCTTTGGC
CAATATCTGGGGAGACAAAACCATGCGGGATACATTT

>Xl_E1_6r|X71801.1|12459-12678
CCAACGTGGATATCTCATGAGGTTACTCTCaTAGGCTCTGTCTGGTGGCGTATGGGAGCGTCGCCCTGTgagGCACAGT
GACGTATAACCCCTTCCACAACGTTGGAGAAACAAGCATGTCCCCGGCCATTCCGCTGGTGTGGGCATCACCCCTGTATATC
CTACAACCTTCCCAGAGCCCTCGGGCAGTGACCGGGGAGACAAAACCATGCAGGATACAGAC

>Xl_E1_4r|X71801.1|6865-7083
CCAACGTGGATATCTCATGAGGTTACTCATGGGCTCTGTCTTGGTGGCGTATGGGAGCGTCGCCCTGTgagGCGCAGTGA
CGTATAATCCCCTTCCACAACGTTGGAGAAACAAGCATGTCCCTAGCCTTCCGCTGGTGTGGGCAGCACCCCTGTATATTC
TACAACCTTCCCAGAGCCCTCGGGCAATGACTGAGGAGACAAAACCATGCAGGACACACAA

>Xl_E1_2r|X71801.1|3489-3711
TCAATGTGGATATCTGTGAGGTTACTCTCaTGGGCTCTGTCCAAGTGGCGTATGGGAGCGTAgTgCTTTATgGCGCAGT
GACGTACAATCCCCTTCCACAACATTGAAGAACAAGCATGTCCCTCGGCCATTAGGCTGTTGTGGGCAGCGTCCTGTATA
TCCTACAACCTTCCCAAAGCCAAAATGGCTCTTACTGGGGAGACAAAACCATGCAGGTTACATCT

>Xl_E1_1r|X71801.1|1462-1682
CCAACGTGGATATCTCATGAGGTTACTCTCaTGGGCTCTGTCTTGGTGGCGTATGGGAGCGTCGCCCTGTgagGCGCAGTG
ACGTATAATCCCCTTCCACAACGTTGGAGAAACAAGCATGTCCCCGGCCATTCCGCTGGTGTGGGCATCACCCCTGTATATTC
CTACAACCTTCCCAGAGCCCTCGGGCAGTGACTGAGGAGACAAAACCATGCAGGATACATT

>Xl_E1_5r|X71801.1|9191-9412
CCAACGTGGATATCTCATGAGGTTACTCTCaTGGGCTCTGTCTTGGTGGCGTATGGGAGCGTCGCCCTGTgagGCGCAGTG
ATGTATAATCCCCTTCCACAACGTTGGAGAAACAAGCATGTCCCCGGCCATTCCGTTGGTGTGGGCAGCACACTGTATATTC
CTACAACCTTCCCAGAGCCGTCGGGCAGTGACCGGGGAGACAAAACCATGCAGGATACATTA

>Oc_E1_1r|U64696.1|3-159
TCTCCCCGGGCTCTGTCCAGGTGGCATAGGGGAGCACAGGGCTCTGCCCGATGACGTACAGTCCCTTCCCACGCCGTTGGA
GAGGAAGCGGCCCTCGAGTCTGCGCCTGCATATTCCTACAGCTTCCCAGAGTCTGTGGACAATGACTGGGGAGAC

>Ss_E1_1|BI342555.1|248-448
CCAACGTGGATACCCCGGAGGTCCTCTCCCCGGGCTCTGTCTGGTGGCGTAGAGGAGCATAGGGCTCTGCCCCATGATG
TACAGTCCCTTCCACAACGTTGGAGATGAAGCTGGGCCTTGAGTCTGCGCCTGCATATTCCTACAGCTTCTCAGAGTCTGT
TGGACAATGACTGAGGAGACAAAACCATGCAGGAAACA

>Tn_E1_1|Chr14|9588563-9588781
CCGACGTGGACGTCCAGCGAGCTGCTCTCCTGGCTCTGTCTGGTGGCCTATGGGAGCGTAGCTGTCCGGCCGAGACGTACA
GTCCCGTCCACGCCGTCGGAGACAGAGCTGTGTCCCCGGCCTTGGCTGGTGTGGACCGGCGTCTGCATATTCCTACATCT
TCCCAAGACCTAACGGGCCGTTGCTGGGGGGCTAAACCATGCAGTGCAC

>Tn_E1_2|Chr14|9589252-9589464
CGTGGACGTCCAGCGAGCTGCTCTCCTGGCTCTGTCTGGTGGCCTATGGGAGCGTAGCTGTCCGGCCGAGACGTACAGTCC
CGTTCCACGCCGTCGGAGACAGAGCTGTGTCCCCGGCCTTGGCTGGTGTGGACCGGCGTCTGCATATTCCTACATCTTCCC
CAGCAGCTAACGGGCCGTTACTGGGGGGCTAAACCATGCAGTGCAC

>Tn_E1_3|Chr14|9589909-9590120
CGTGGACATCCAGCGAGCTGCTCTCCTGGCTCTGTCTGGTGGCCTATGGGAGCGTAGCTGTCCGGCCGAGACGTACAGTCCC
GTTCCACGCCGTCGGAGACAGAGCTGTGTCCCCGGCCTTGGCTGGTGTGGACCGGCGTCTGCATATTCCTACATCTTCCCA
GAGCCCTAACGGGCCGTTACTGGGGGGCTAAACCATGCAGTGCACAT

>Tn_E1_4|Chr14|9590478-9590689
CGTGGACATCCAGCGAGCTGCTCTCCTGGCTCTGTCTGGTGGCCTATGGGAGCGTAGCTGTCCGGCCGAGACGTACAGTCC
CGTTCCACGCCGTCGGAGACAGAGCTGTGTCCCCGGCCTTGGCTGGTGTGGACCGGCGTCTGCATATTCCTACATCTTCCC
AGAGCCCTAACGGGCCGTTACTGGGGGGCTAAACCATGCAGTGCAC

>Mm_9_E1|Chr9|120114860-120115066
TATCCAACGTGGACAACCCAGGAGGTCACTCTCCCTGGGCTCTGTCTAGTGGCATAAGGGAGCATAGGCCTTGCCCAAGTGA
CGTACAGTCCCTTCCACGCCGTTGGAGATGAAGCTGGGCTTTGTGCCCGCCTGCATATTCCTACGACTTCTCAGAGTCC
TGTGGACAGTACTGAGGAGGCAAACCATGCAGGAAACAGT

>Gg_E1|ChrUn|9714-9874
GCTCTCTCTCGGCTCTGTCCGTGTGGCGACGGGAGCGTAGGCCTTCCGGCCCGAGACGTATATTCCTTCCACATCGTTG
GAGAAACAAGCTGGCCTCGAGCCTGCAGCCTGCATATTCCTACTGCTTCTCTGAGCCCTCGGGCAGTGACAGAGGAGAT

Tabelle A.2: Sequenzen der E2 snoRNA

```

>Cf_23_E2_1|Chr23|12080172-12080321
GCTGCACACTATTAAGCTCAAGGTTGAGGCCACCGTGGCCAACTGAACTTCTGAGTGTCCG
AAGTGTGCTATATAAACAGCAGGGTTTTCGCTAACACCAGTAGAGCTTGTCTCTATGACTG
GAGTTTGATAGTACTCACTGCTACATTG
>Cf_23_E2_2|Chr23|12082742-12082897
CTCTGCACATTGTTAGAGCTTGGAGTTGAGGCCACTGGCTGGCCGATGAACTCAAAAGTGT
AGGTAGTGTGCTACATGAGGGGCAAGTCTTTTCGCTAACACCACAAGGGTCTCTGGCCCAA
TGAGTGGAGTTTGATAGTAATTCTTGCTACAAG
>Dr_E2_1|Zv4_NA8930.1|68513-68669
ACGCACACTGTTAAAACCTCACACTAATGTTGGTGTAAAGGTGTGGAAAGTGTGCTAGAGTG
ATTCAGCCGGGATTTTCGCTAACACCAGTAGGGCTGTTACGCCCTTGACTGGAGTTTGATAG
CAACCTGAGCTACATTTATCCCACTCAACCT
>Dr_E2_2|Zv4_NA8930.1|69001-69157
ATGCACACGGTTAAAGCCTGGCACTGATGTTGGTGTGGTGTGGTGTGGAAAGTGTGCTAAATAA
ACATGGCTAGGTATTCGCTAACACCAGTAGGGCTGAAGCGCCTCATGACTGGAGTTTGATAG
CAACCTTTGCTACATTTGATCTAAATTCATAT
>Tr_E2_1|scaffold_3749|675-830
TGCACACTGTAAAGCTAGGCAACTGTCAGTGTCTAGGTGTGGAAAGTGTGCTAAAGCAA
GTGGACAAGGTTTTCACTAATGCCATGAGGACTTCTGTCTTGTGACTGGCAATTGATAGTGA
CCTTAGCCACACTTCTGTAACAAAATCAAAA
>Gg_E2_2|Chr2|43447109-43447267
TGCACACTGTAAAGCTTGTGCTGAGGTCAGTCTGACCAATGAGCTCTCAAGTGTGGAAAG
TGTGCTACACAAAGCTGTTGGCAGGTGCTTGCTAATGCCAGGAGAGGAGCTGCCTCTATGAT
TGGAGTTTGATAGCAGCCCTCCGACATGCT
>Gg_E2_1|Chr2|43445870-43446024
CACACAGTTGGAACCTGGGAGCTGAGAGCCAACCTGTGGCCAAGGAACTCTCAGTGTAGGATGT
GTGCTACAGTAACATGGCAGGACTTACTGACAGCCTGAGGGCCTTTGTGGCCAATGAAAGGAG
TTTGGTAGTCAATCTTGCCACAGTT
>Hs_3_E2_1|Chr3|39410473-39410622
TGCACACTATTAAGCTCAGGGTGGAGGCCAGTCTTGGCTCATGAACTTCTGAGTGTGGAAAGT
GTGCTATATCAATGGCAGGATTTTCGCTAACACCAGTAGAGCTTGCCTCTATGACTGGAGTTG
GTAGTACTCGCTGCCACATAG
>Hs_3_E2_2|Chr3|39413136-39413290
TGTGCACATTGTTAGAGCTTGGAGTTGAGGCTACTGACTGGCCGATGAACTCGCAAGTGTAGGT
AGTGTGCTACATGAGGGGCAAGTTTTCGCTAACACCACAAGGGTCTCTGGCCCAATGAGTGGAG
TTTGATAGTAATTCTTGCTACAAGTA
>Mm_9_E2_1|Chr9|120114293-120114448
CTGCACACTATTAAGCTCAGAGTGGAGGCCAGTCTTGGCCAGTGAACCTCTGAGTGTGGAAAG
TGTGCTACATTGAATTTATGGCAGGATTTTCGCTAACACCAGTAGAGCTTGCCTCTATGACTGG
AGTTTGGTAGTGGTCACTGCCACATGC
>Mm_9_E2_2|Chr9|120115926-120116071
CCCTGCACATTGTTAGAGCTTGGAGTTGAGGTCACACGCTGGTCCGATGAACTCAAAAGTGTAGGT
AGTGTGCTAAAAGAGTGGCAAGGTTTTTCGCTAACCCCCACTCTGGTCCAGTGAAGTGGGTTTGA
TAGTAACCTTTGCTA

```

```
>Rn_8_E2_1|Chr8|125120943-125121091
TGCACACTATTAAGCTCAGGGTGGAGGCCAGATTTGGCCAGTGAACTTCTGAGTGTGGAAAGTG
TGCTACATTGAACATACGGCAGGATTTTCGCTAACACCAGTAGAGCTTGCCCTCTATGACTGGAGT
TTGGTAGTGGTCACTGCCACATAC
>Rn_8_E2_2|Chr8|125122686-125122841
CCCTGCACATTGTTAGAGCTTGAGGTTGAGGTCACACGCTGGTCGATGAACTCCTAAGTGTAGGT
AGTGTGCTAAACGAGCGGCAAGGTTTTTCGCTAACACCCCACTCTGGTCCAGTGAGTGGGGTCTGA
TAGTAACCTTTGCTACAAGTTTGAT
>Xt_E2_1|scaffold_22171|633-785
TGCACACTGTTAAAGCTCAGATGAAGCCATGCTGGCTAATGAACTCTTGAGTGTGGAAAGTGTGC
CATAGCATTATTGTGGCAAGATTTTCGCTAACACCGTCAGAGTCCTTGCTCAATGATCGGCGTT
TGATAGTGATCTCTGCTACAGTT
>Xt_E2_2|scaffold_22171|1969-2120
TTGCACACAGTTAGAAGCTTGAGGCTGAGGCTTTTCAGCCAATGAACTCTGAGTGTAGGATGTGT
GCTACATTAACATGGCAGACTTTTCGCTAACGCCACAAGGAGCTTTAGAGCTCCAGTGACTGGAG
TTTGATAGTAATTCTTGCCACAGCT
```

Tabelle A.3: Sequenzen der E3 snoRNA

```
>Cf_34_E3_1|Chr34|19473959-19474073
TACAATATAGCTGTGAAGTGCTATATTGCCTTACCCCTGCTGAAAGCAACTGTTTCTACAATCT
CTGTCTAGCTCAATTCTATAGTAAGACAGGGACGCTTGGTCTCAAAACAT
>Cf_34_E3_2|Chr34|19475059-19475175
TACAATGTAGCTATAAAGTGTGCCTTGTCTTTACCCCTGCTCAAAATAAACTGTTATCTTTCT
AGCCCTGTCCGCCATGGTGCTTAGCAGCCAGGGATGCAGGGTCTCAGACAT
>Gg_9_E3_1|Chr9|15488664-15488795
AAAGCAGGATAAGTTATGATATAGCTGTGCAGTGTGTATTATCGTTCCTGCTTAAAGAAAAA
TGTTTCCTTCTATCCCTACCTGCTTAGTTCACAGCAGGATAGGACGCTGGGTTTCATACATCC
>Hs_3_E3_1|Chr3|187825023-187825149
AATGCAGGTACTGTTACAATACTGATGTGTTTTGTTGTCGTTCCCTGCTTAAAGCACTTGA
TGCATAACTCTGTCTACCTTCAATCCGTAGTAAGACAGAGACGCTTGGCTTCAGACATTT
>Hs_3_E3_2|Chr3|187826000-187826135
AAAGCAGGATTCAGACTACAATATAGCTGCTAAGTGTGTGTTGTCGTTCCCTGCTTAAATAA
AGTTGTTTCTTAACATACTGTCTGCTATTCTCCTGTAGCAGCCAGGGACGCTTGGTCTCATACA
TGT
```

>Mm_16_E3_1|Chr16|22954373-22954502
TCATGCAGGTACAGTTACAATAAACTGAAGTGTCTATTGTCGTTCCCCCTGCTTAAAGCATTG
ATACATTACTATCTCTGCCTAGCTCTACCATACTAAGGCAGAGACGCTTGGTTTCAAACATTT
>Mm_16_E3_2|Chr16|22955292-22955425
TTAAGCAGGATTTAACTACAATATAGCTGCTCAGTGCTGTGTTGTCGTTCCCCCTGCTCAGAATAA
TTGTTTCTTAACTATACTGTCTGCCATCTGCCTGTAGCAGCCAGGGACGCTTGGTCTCAAACATGA
>Pt_2_E3_1|Chr2|192078762-192078885
AATGCAGGTACTGTTACAATACTGACGTGTTTTATTGTCGTTCCCCCTGCTTAAAGCACTTGA
TGCATAACTCTCTCTAACTCCATTCCGTAGTAAGACAGAGACGCTTGGCTTCAAACAT
>Pt_2_E3_2|Chr2|192079738-192079872
AAAGCAGGATTCAGACTACAGTATAGCTGCTAAGTGCTGTGTTGTCGTTCCCCCTGCTTAAATAA
AGTTGTTTCTTAACTATACTGTCTGCTATTCTCTGTAGCAGCCAGGGACGCTTGGTCTCATACA
TGT
>Rn_11_E3_1|Chr11|80010709-80010823
TACAATATGGCTGCTCAGTGCTGTGTTGTCGTTCCCCCTGCTCAGAACAATTGTTTCTTAACTATA
CCTGTCTGTCTACCTGTAGCAGCCAGGGACGCTTGGTCTCATACAT
>Rn_11_E3_2|Chr11|80012052-80012181
TCATGCAGGTACAGTTACAATAAACTGAAGTGTCTATTGTCGTTCCCCCTGCTTAAAGTATCTG
ATACATTACTATCTCTGCCTAGCTCTACCATAGTAAGGCAGAGACGCTTGGTTTCAAACATTT
>Xt_E3_1|scaffold_11263|6840-6962
ATGCAGGATAAGTTACAATATAGCTGGCAGTGCTGTGTTGTCGTTCCCCCTGCTGAAAGTAAATG
TTTCTTCTATTCTGTCTGTTTCTATAGCAGACAGGAACGCTTGGTCTCAAACAA
>Xt_E3_2|scaffold_11263|5880-6010
CTGCAGGTTAAGTTACAATACATCTGTGCAGTGTTGTTGTCGTTCCCCCTGCAAAAAATAAACT
GCTTCCCTACTATCCCTGTCTGTTCTGACACTATCCAGCAGACAGGGACGCTTGGTCTCAGACAT
>Dr_25_E1_3|Chr25|557916-558049
GTGCAGGTTTAGTTACGGA CTGGCTGTTGCAGTGCTTGTTCGTCGTTCCCCCTGCAGAAATAAACA
TGTTTCCCTACTATCCCTGTCTGTCTACTACTTAGATCTGCAGGGACGCTTGGTCTCAGACATTTCAAT
>Dr_25_E3_2|Chr25|556141-556274
GTGCAGGTTCAAGTTATGGA CAGGCTGTTGCAGTGCTTGTTCGTCGTTCCCCCTGCACAAATGAACAT
GATTCCTTCTATCCCTGTGTCTGCTTCTCATGACTTGCAGGGACGCTTGGTCTCAGACAGTTTTAT
>Dr_25_E3_1|Chr25|552173-552306
GTGCAGGTTTAGTTACGGA CTGGCTGTTGCAGTGCTTGTTCGTCGTTCCCCCTGCAGACATAAACAT
GTTTCCCTACTATCCCTGTTTGTCTACTACTTAGATCTGCAGGGACGCTTGGTCTCAGACATTTCAAT
>Dr_25_E3_4|Chr25|560532-560666
GTGCAGGTTTAGTTACGGA CTGGCTGTTGCAGTGCTTGTTCGTCGTTCCCCCTGCAGAAATAAACAT
GATTCCTTCTATCCCTGTCTGTTTGCATCCTACAACTAACAGGGACGCTTGGTCTCAGACAAATTTAATC
>Dr_25_E3_5|Chr25|564135-564268
GTGCAGGTTTAGTTACGGA CTGGCTGTTGCAGTGCTTGTTCGTCGTTCCCCCTGCAGAAATAAACAT
GTTTCCCTACTATCCCTGTCTGTCTACTACTTAGATCTGCAGGGACGCTTGGTCTCAGACATTTCAAT
>Dr_25_E3_6|Chr11|567675-567809
GTGCAGGTTTAGTTACGGA CTGGCTGTTGCAGTGCTTGTTCGTCGTTCCCCCTGCAGAAATAAACAT
GATTCCTTCTATCCCTGTCTGTTTGCATCCTACAACTAACAGGGACGCTTGGTCTCAGACAAATTTAATC
>Tr_E3_1|scaffold_149|217573-217707
CATGCAGGATAAGTTACAGTGCAGCTGTGTAGTGTGTGTTGTCGTTCCCCCTGCAAAAAATGACA
TGTTTCCCTCGCTATACCTGTCAAGTCAACTTCCCTAGATTGACAGGGACGTAAGGTTTTCAGACAAGTA
>Tr_E3_2|scaffold_149|217937-218065
ACAGCAGGATAAGTTACAGTGCAGCTGTGGAAGTGTGTGTTGTCGTTCCCCCTGCACAACTAACT
TGTTTCCAGCTATCCCTGTTTGTCTCTTTTATTGGACAACCCAGGGACGCAAGGTCTCGGACA
>Tn_E3_2|SCAF14663|197666-197801
CGTAGCAGGATAAGTTACAGTGCAGCTGTGGAAGTGTGTGTTGTCGTTCCCCCTGCAAAAACTAAC
CTGTTTCCAGCTATCCCTGCTTGTCTTCTTCCCTCGTAGACACCCAGGGACGCAAGGTCTCAAACAGTTGCC
>Tn_E3_1|SCAF14663|197267-197401
CATGCAGGTTAGGTTACAACGCAGCTGTGGAAGTGTGTGTTTGTGTCGTTCCCCCTGCAAAAAGTGACA
TGTTTCCCTCGCTATACCTGTCAAGTACCCTCTCTTGA CTGACAGGGACGTAAGGTTTTCAGACAAGTA

Tabelle A.4: Sequenzen der U36 snoRNA

```
>Pt_U36_3|Chr11|118286841-118286909
TGAAC TTTGCCAGT GATGGTTAAGAAT TCTTCACCTGAATAAACCATGTGGTCAGCATTGCATCTGA
>Pt_U36_2|Chr11|118286455-118286524
CTTGCAATGATGTGAATCTCTCACTGAATCAACCTTGAAGTGCGAATCCATGAGCTTTTTAACCTGA
>Pt_U36_1|Chr11|118286094-118286163
GTTGCAGT GATGTAAAAT TCTTCGCCTGAAAT TACTGTGAAGAGTAAAACCGAGCTTTTTAACACTGA
>Mm_U36_3|Chr2|26872870-26872936
TGCCAGT GATGCTTAGGAAT TCTTCACCTGAATCAACTATGTGGTCACACTGTTGTCTGAGGCAA
>Mm_U36_2|Chr2|26872240-26872309
GTGCAGT GATGTGAAT TCTTCACCTGAAAATATTGTGAAGAGTAAATCGAGCTTTTTAACCTGAGT
>Mm_U36_1|Chr2|26872551-26872621
TTGCAATGATGTCAATCTTTGACTGAAGTGACCTTGAAGTGCAAT TACTGAGCTTTTTAACCTGAGCAA
>Hs_U36_1|X52138.1|3930-4011
TTGCAATGATGTGAATCTCTCACTGAATCAACCTTGAAGTGCGAATCCATGAGCTTTTTAACCTGAGCAA
>Hs_U36_2|X52138.1|4300-4372
GTTGCAGT GATGTAAAAT TCTTGGCCTGAAAT TACTGTGAAGAGTAAAACCGAGCTTTTTAACACTGAGT
>Hs_U36_3|X52138.1|4693-4760
TTGCCAATGATGGTTAAGAAT TCTTCACCTGAATAAACCATGTGGTCAGCATTGCATCTGAGGCAA
>Gg_U36_1|Chr17|3534505-3534579
TTGCAATGATGTGAAT TCTTCACCTGAATCAACAATGAAGAGCAAAATGAGCTTTTTAACACTGAGCAA
>Gg_U36_2|Chr17|3534958-3535030
ATGGCAGT GATGTATGAAT TCTTCACCTGAGCTCAAAGTGAAGAGCGAAATAGACGAGCTTTTTAACCTGA
>O1_U36_1
TGATGTCTGAAT TCTTCACCTGAATCCTCCGTGGATCCAAACATACGAGCTTTTTAACCTGA
>O1_U36_2
ATGATGTT CAGAAT TCTTCACCTGAAACTTTCTATGATGGTTTTT GAGCTTTTTAACCTGA
>Tn_U36_1|Chr4|3068454-3068522
TGATGACTTGAAT TCTTCACCTGATGGCTTCAGT GATGCGTATTGAGCTTTTTAACCTGA
>Tn_U36_2|Chr4|3068731-3068797
TGATGTTTGAAT TCTTCACCTGAAGTCACTGTGTAGATCAAAACAGACGAGCTTTTTAACCTGA
>Tr_U36_2|scaffold_416|116217-116282
TGATGTTTGAAT TCTTCACCTGAAGTCTCTATGTAGATCAAAACAGACGCGCTTTTTAACCTGA
>Tr_U36_1|scaffold_416|115938-116000
TGATGACTTGAAT TCTTCACCTGAAACCTTCAGT GATGCTTTTTT GAGCTTTTTAACCTGA
>Rn_U36_3|Chr3|5810326-5810391
TTGCCAATGATGGTTAAGAAT TCTTCACCTGAATCAACTATGTGGTCATACTCTTCTGTCTGA
>Rn_U36_2|Chr3|5809987-5810052
TTGCAATGATGTGAATCTTTTGCTGAAGTGACCTTGAAGTGTAATCCTGAGCTTTTTAACCTGA
>Rn_U36_1|Chr3|5809666-5809743
TGTGCAATGATGTGTGAAT TCTTCACCTGAAATCACCGTGAAGAGTAATTTAAGTAATCGAGCTTTTTAACCTGA
>Dr_U36_1|Zv4_NA6576.1|74728-74793
TGATGTCTTGAAT TCTTCACCTGAATCAAACTGAAGATCAAAATACGAGCTTTTTAACCTGA
>Dr_U36_2|Zv4_NA6576.1|75085-75145
TGATGTCTGAAT TCTTCACCTGAACAACCATGTGGTTAAATGAGCTTTTTAACCTGA
```

Anhang B

Bootstrapped

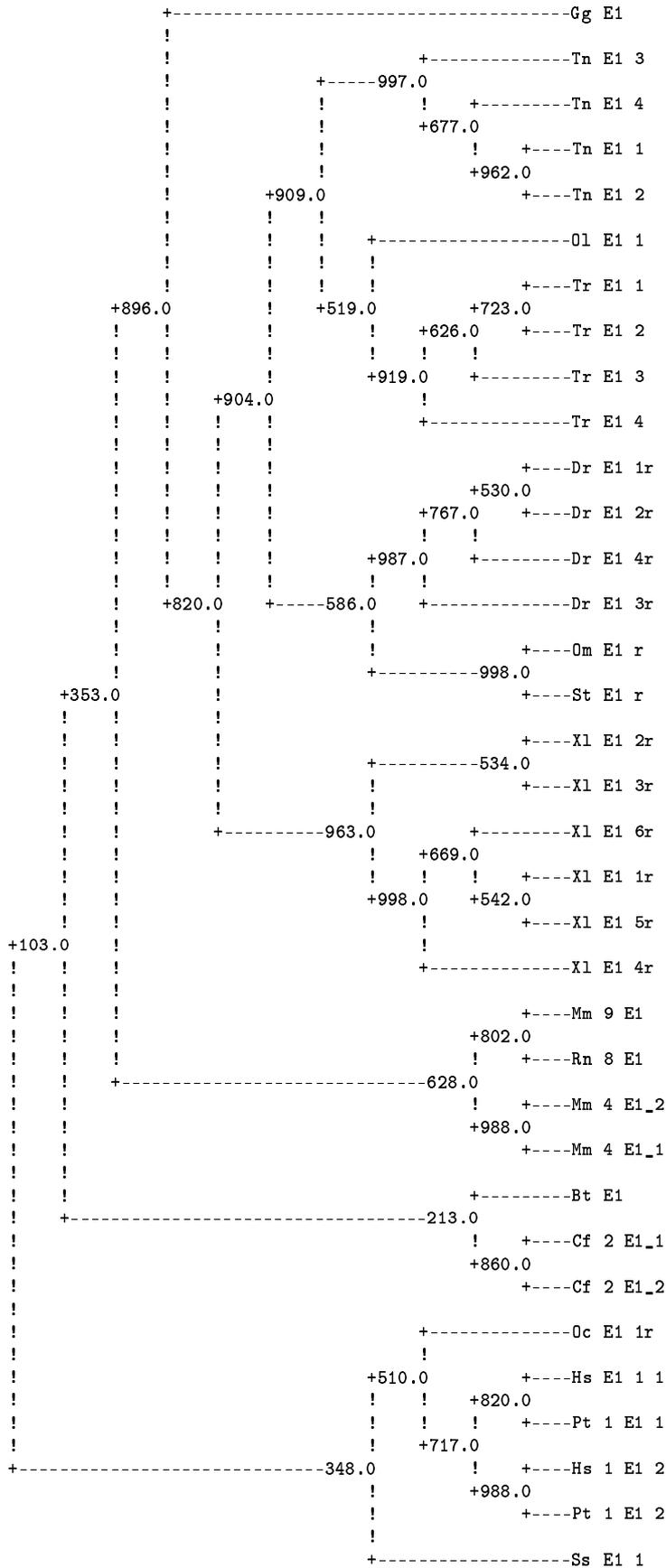


Abbildung B.1: Bootstraphbaum der U17/E1 snoRNA

```

          +----Mm 9 E2_1
        +934.0
      +750.0 +----Rn 8 E2_1
        !   !
    +666.0 +-----Hs 3 E2_1
        !   !
+-----286.0 +-----Cf 23 E2_1
!           !
!           +-----Gg E2 1
!
!           +----Hs 3 E2 2
!           +994.0
!           !   +----Cf 23 E2 2
!           +923.0
!           !   !   +----Mm 9 E2 2
!           !   +994.0
!           +854.0 +----Rn 8 E2 2
!           !   !
!           !   !   +----Xt E2 2
!           +349.0 +-----953.0
!           !   !   +----Gg E2 2
!           !   !
+231.0 +-----Xt E2 1
!
!           +----Dr_E2 2
!           +805.0
+-----747.0 +----Dr_E2 1
!
!           +-----Tr E2_1

```

Abbildung B.2: Bootstrapbaum der E2 snoRNA.

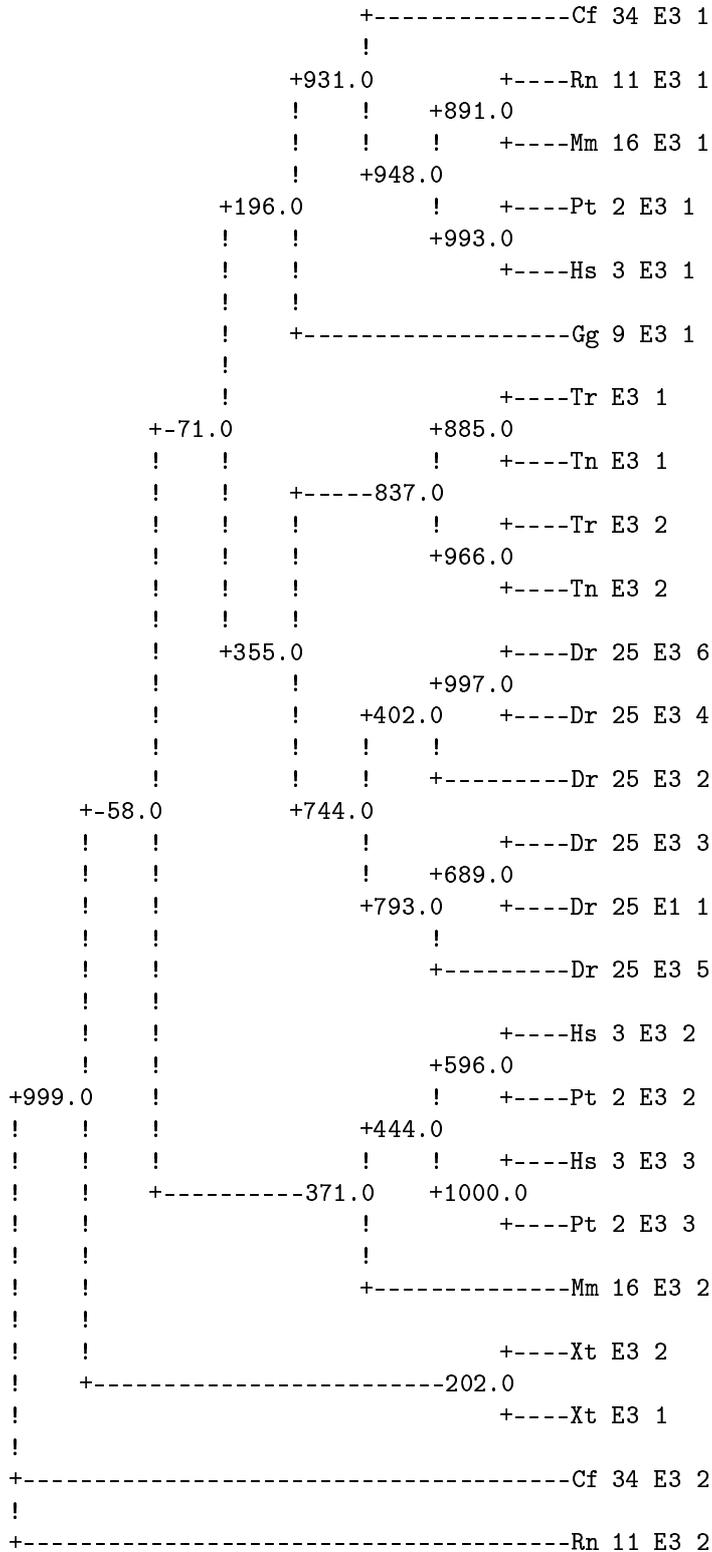


Abbildung B.3: Bootstrapbaum der snoRNA E3

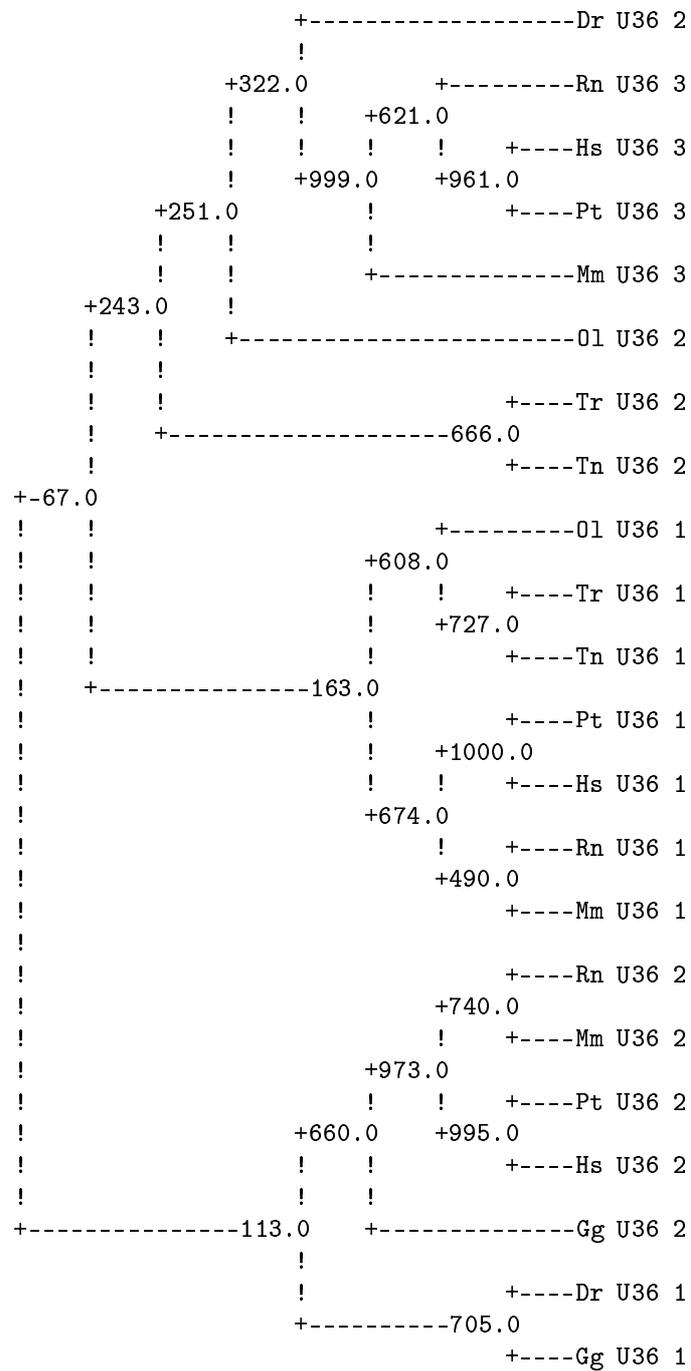


Abbildung B.4: Bootstrapbaum der snoRNA U36

Literaturverzeichnis

- [1] Accardo, MC., Giordano, E., Riccardo, S., Digilio, FA., Iazzetti, G., Calogero, RA., Furia, M., A computational search for box C/D snoRNA genes in the *D.melangoster* genome. *Bioinformatics*, **0**:3942-3950,2004.
- [2] Altschul, SF., Gish, W., Miller, W., Myers, EW., Lipman, DJ., Basic local alignment search tool. *J. Mol. Biol.*, **215**(3):403-410, 1990.
- [3] Atzorn, V., Fragapane, P., Kiss, T., U17/snR30 Is a Ubiquitous snoRNA with Two Conserved Sequence Motifs Essential for 18S rRNA Production. *Mol Cell Biol*, **24**(4):1769-1778, 2004.
- [4] Bachellerie, JP., Michot, B., Nicoloso, M., Balakin, A., Ni, J., Fournier, MJ., Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem. Sci.*, **20**(7):261-264, 1995.
- [5] Bachellerie, JP., Cavaille, J., Qu, LH Nucleotide modifications of eukaryotic rRNAs: the world of small nucleolar RNA guides revisited. In Garret, R., Douthwaite, R., Liljas, A., Matheson, A., Moore, P., Noller, H. (eds) *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions* ASM Press, Washington DC, pp.191-203, 2000.
- [6] Bachellerie, JP., Cavaille, J., Huttenhofer, A., The expanding snoRNA world. *Biochimie*, **84**(8):775-790, 2002.
- [7] Barrett, C., Hughey, R., Karplus, K., Scoring hidden Markov models. *Comput Appl Biosci.*, **13**(2):191-199, 1997.
- [8] Benson, D., Lipman, DJ., Ostell, J., GenBank. *Nucl. Acids. Res.*, **21**(13):2963-2965, 1993.
- [9] Brown, T.A, Moderne Genetik: eine Einführung. *Spektrum Akademischer Verlag* Heidelberg; Berlin; Oxford , pp.83-103, 1993.

-
- [10] Brown, JW; Echeverria, M., Qu, LH., Plant snoRNAs: functional evolution and new modes of gene expression *Trends Plant Sci*, **8**(1):42-49, 2003.
- [11] Campbell, NA, Reece, JB, Biologie. *Spektrum Akademischer Verlag* 6. Auflage, p.:371-374, 2003.
- [12] Cavaille, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, CI., Horsthemke, B., Bachellerie, JP., Brosius, J., Huttenhofer, A., Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci USA*, **97**(26):14311-14316, 2000.
- [13] Cavaille, J., Seitz, H., Paulsen, M., Ferguson-Smith, AC., Bachellerie, JP., Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum Mol Genet.*, **11**(13):1527-1538, 2002.
- [14] Cecconi, F., Crosio, C., Mariottini, P., Cesareni, G., Giorgi, M., Brenner, S., Amaldi, F., A functional role for some Fugu introns larger than the typical short ones: the example of the gene coding for ribosomal protein S7 and snoRNA U17. *Nucl. Acids Res.*, **24**(16):3167-3172, 1996.
- [15] Cedergren, R.J., LaRue, B., Sankoff, D., Lapalme, G., Grosjean, H., Convergence and minimal mutation criteria for evaluating early events in tRNA evolution. *Proc. Natl. Acad. Sci. USA*, **77**(5):2791-2795, 1980.
- [16] Cervelli, M., Cecconi, F., Giorgi, M., Annesi F., Oliverio, M., Mariottini, P., Comparative Structure Analysis of Vertebrate U17 Small Nucleolar RNA (snoRNA.) *J. Mol. Evol.*, **54**(2):166-179, 2002.
- [17] Darzacq, X., Jady, BE., Verheggen, C., Kiss, AM., Bertrand, E., Kiss, T., Cajal body-specific small nuclear RNAs: a novel class of 2'-O-Methylation and pseudouridylation quide RNAs. *EMBO J.*, **21**(11):2746-2756, 2002.
- [18] Dennis, PP., Omer, A., Lowe, T., A guided tour: small RNA function in Archaea. *Mol. Microbiol.*, **40**(3):509-519, 2001.

-
- [19] Dunbar, DA., Baserga, SJ., The U14 snoRNA is required for 2'-O-Methylation of the pre-18S rRNA in *Xenopus* oocytes. *RNA*, **4**(2):195-204, 1998.
- [20] Eddy, SR., Profile hidden Markov models. *Bioinformatics*, **14**(9):755-763, 1998.
- [21] Edvardsson, S., Gardner, PP., Poole, AM., Hendy, MD., Penny, D., Moulton, V., A search for H/ACA snoRNAs in yeast using MFE secondary structure prediction. *Bioinformatics*, **19**(7):865-873, 2003.
- [22] Efron, B., Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7**:1-26, 1979.
- [23] Eigen, M., Winkler-Oswatitsch, R., Transfer-RNA, an Early Gene? *Naturwissenschaften*, **68**(6):282-292, 1981.
- [24] Eigen, M., Gardiner, W., Schuster, P., Winkler-Oswatitsch, R., Ursprung der genetischen Information. In Mayr, E.(eds) *Evolution*, Spektrum der Wissenschaften, 7.Auflage, pp.: 60-82, Heidelberg, 1988.
- [25] Fatica, A., Tollervey, D., Insights into the structure and function of a guide RNP. *Nat Struct Biol.*, **10**(4): 237-239, 2003.
- [26] Filippov, V., Solovyev, V., Filippova, M., Gill, S.S., A novel type of RNase III family proteins in eukaryotes. *Gene*, **245**(1):213-221, 2000.
- [27] Felsenstein, J., PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistic*, **5**:164-166, 1989.
- [28] Felsenstein, J., Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**:783-791, 1985.
- [29] Fritzsich, G., Fachgespräch an der Universität Leipzig am 29.09.2004.
- [30] Fromont-Racine, M., Senger, B., Saveanu, C., Fasiolo, F., Ribosome assembly in eukaryotes. *Gene*, **313**:17-42, 2003.
- [31] Gall, JG., Cajal bodies: the first 100 years. *Annu. Rev. Cell Dev. Biol.*, **16**:273-300, 2000.

-
- [32] Ganot, P., Caizergues-Ferrer, M., Kiss, T., The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.*, **11**(7):941-956, 1997.
- [33] Gaspin, C., Cavaille, J., Erauso, G., Bachellerie, JP., Archaeal Homologs of Eukaryotic Methylation Guide Small Nucleolar RNAs: Lesson from the Pyrococcus Genomes. *J. Mol. Biol.*, **297**:895-906, 2000.
- [34] Gilley, J., Fried, M., Evolution of U24 and U36 snoRNAs Encoded within Introns of Vertebrate rpl7a Gene Homologs: Unique Features of Mammalian U36 Variants. *DNA Cell Biol.*, **17**(7):591-602, 1998.
- [35] Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., Eddy, SR., Rfam: an RNA family database. *Nucl. Acids Res.*, **31**(1):439-441, 2003.
- [36] Gustafsson, C., Reid, R., Greene, PJ., Santi, DV., Identification of new RNA modifying enzymes by iterative genome search using known modifying enzymes as probes. *Nucl. Acids Res.*, **24**(19):3756-3762, 1996.
- [37] Hedges, SB., Kumar, S., Genomics. Vertebrate Genomes Compared. *Science*, **297**(5585), 1283-1285, 2002.
- [38] Henras, AK., Dez, C., Henry, Y., RNA structure and function in C/D and H/ACA s(no)RNPs. *Curr. Opin. Struct. Biol.*, **14**(3):335-343, 2004.
- [39] Hofacker, IL., Fontana, W., Stadler, PF., Bonhoeffer, LS., Tacker, M., Schuster P., Fast Folding and Comparison of RNA Secondary Structures. *Monatshefte für Chemie*, **125**:167-188, 1994.
- [40] Hofacker, I.L., Fekete, M., Flamm, C., Huynen, M.A., Rauscher, S., Stolorz, P.E., Stadler, P.F., Automatic detection of conserved RNA structure elements in complete RNA virus genome. *Nucl. Acids Res.*, **26**(16):3825-3836, 1998.
- [41] Hogeweg, P., Hesper, B., Energy directed folding of RNA sequences. *Nucl. Acids Res.*, **12**(1 Pt 1):67-74, 1984.
- [42] Ivanov, PV., Zvereva, MI., Shpanchenko, OV., Dontsova, OA., Bogdanov, AA., Aglyamova, GV., Lim, VI., Teraoka, Y., Nierhaus, KH., How does tmRNA move through the ribosome? *FEBS Letters*, **514** (1):55-59, 2002.

-
- [43] Kiss, T., Small Nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.*, **20**(14):3617-3622, 2001.
- [44] Kiss, T. Small Nucleolar RNAs: An Abundant Group of Noncoding RNAs with Diverse Cellular Functions. *Cell*, **109**(2):145-148, 2002.
- [45] Klein, DJ., Schmeing, TM., Moore, PB., Steitz, TA., The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**(15):4214-4221, 2001.
- [46] Knudsen, B., Wower, J., Zwieb, C., Gorodkin, J., tmRDB (tmRNA database). *Nucl Acids Res.* **29**(1):171-172, 2001.
- [47] Kuhn, JF, Tran, EJ, Maxwell, ES, Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucl. Acids Res.*, **30**(4):931-941, 2002.
- [48] Lafontaine, DL., Tollervey, D., Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem Sci*, **23**(10):383-388, 1998.
- [49] Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T., Identification of novel genes coding for small expressed RNAs. *Science*, **294**(5543): 853-858, 2001.
- [50] Lange, TS, Borovjagin, A., Maxwell, ES, Gerbi, SA, Conserved Boxes C and D are essential nucleolar localization elements of U14 and U8 snoRNAs. *EMBO J.*, **17**(11):3176-3187, 1998.
- [51] Lee, Y., Jeon, K., Lee, JT., Kim, S., Kim, VN., MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.*, **21**(17):4663-4670, 2002.
- [52] Lewin, B., Genes VII *Oxford University Press*, 1. Auflage, p.:139-143, 2000.
- [53] Liao, D., Pavelitz, T., Kidd, JR., Kidd, KK., Weiner, AM., Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *EMBO J.*, **16**(3):588-598, 1997.

-
- [54] Liao, D., Gene Conversion Drives Within Genic Sequences: Concerted Evolution of Ribosomal RNA Genes in Bacteria and Archaea. *J. Mol. Evol.*, **51**(4):305-317, 2000.
- [55] Looney, JE., Harding, JD., Structure and evolution of a mouse tRNA gene cluster encoding *tRNA_{Asp}*, *tRNA_{Gly}* and *tRNA_{Glu}* and as unlinked, solitary gene encoding *tRNA_{Asp}*. *Nucl. Acids Res.*, **11**(24):8761-8775, 1983.
- [56] Lowe, TM., Eddy, SR., tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl. Acids Res.*, **25**(5):955-964, 1997.
- [57] Lowe, TM., Eddy, SR., A Computational Screen for Methylation Guide snoRNAs in Yeast. *Science*, **283**(5405):1168-1171, 1999.
- [58] Mathews, DH., Sabina, J., Zuker, M., Turner, DH., Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**(5), 911-940, 1999.
- [59] McCaskill, JS., The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**(6-7):1105-1119, 1990.
- [60] Moulton, V., Zuker, M., Steel, M., Pointon, R., Penny, D., Metrics on RNA secondary structures. *J. Comput. Biol.*, **7**(1-2), 277-292, 2000.
- [61] Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., Dreyfuss, G., miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.*, **16**(6):720-728, 2002.
- [62] Muto, A., Ushida, C., Himeno, H., A bacterial RNA that functions as both a tRNA and an mRNA. *Trends Biochem Sci.*, **23**(1): 25-29, 1998.
- [63] Nag, MK., Thai, TT., Ruff, EA., Selvamurugan, N., Kunnimalaiyaan, M., Eliceiri, GL., Genes for E1, E2 and E3 small nucleolar RNAs. *Proc. Natl. Acad. Sci. USA*, **90**(19):9001-9005, 1993.
- [64] Nelder, JA., Mead, R., A simplex method for function minimization. *The Computer Journal*, **7**:308-313, 1965.

- [65] Nelson, P., Kiriakidou, M., Sharma, A., Maniataki, E., Mourelatos, Z., The microRNA world: small is mighty. *Trends Biochem Sci.*, **28**(10):534-540, 2003.
- [66] Ofengand, J., Fournier, M.J., The pseudouridine residues in rRNA: number, location, biosynthesis, and function. In Grosjean, H., Benne, R.(eds) *Modification and Editing of RNA*. American Society for Microbiology, Washington DC, pp.229-253.
- [67] Ofengand, J., Rudd, KE, Bacterial, Archaeal, and Organellar rRNA Pseudouridines and Methylated Nucleosides and Their Enzymes. In Garret, R., Douthwaite, R., Liljas, A., Matheson, A., Moore,P., Noller, H.(eds) *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions* ASM Press, Washington DC, pp.175-189.
- [68] Omer, AD., Lowe, TM., Russell, AG., Ebhardt, H., Eddy, SR., Dennis, PP. Homologs of Small Nucleolar RNAs in Archaea. *Science*, **288**(5465): 517-522, 2000.
- [69] Orgel, LE., The Origins of Life- a review of facts and speculations. *Trends Biochem Sci.*,**23**:491-495, 1998.
- [70] Parra, G., Blanco, E., Guigo, R., GeneID in Drosophila. *Genome Res.*, **10**(4):511-515, 2000.
- [71] Pasquinelli, AE., Reinhart, BJ., Slack, F., Martindale, MQ., Kuroda, MI., Maller, B., Hayward, DC., Ball, EE., Degan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., Ruvkun, G., Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, **408**(6808):86-89, 2000.
- [72] Pelczar, P., Filipowicz, W., The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is member of the 5'-terminal oligopyrimidine gene family. *Mol. Cell. Biol.*, **18**(8):4509-4518, 1998.
- [73] Qu, LH., Henry, Y., Nicoloso, M., Michot, B., Azum, MC., Renalier, MH., Caizergues-Ferrer, M., Bachellerie, JP., U24, a novel intron-encoded small nucleolar RNA with two 12 nt long, phylogenetically con-

- served complementarities to 28S rRNA. *Nucl. Acids Res.*, **23**(14):2669-2676, 1995.
- [74] Rimoldi, OJ., Raghu, B., Nag, MK., Eliceiri, GL, Three New Small Nucleolar RNAs That are Psoralen Cross-Linked In Vivo to Unique Regions of Pre-rRNA. *Mol. Cell Biol.*, **13**(7):4382-4390, 1993.
- [75] Rivas E., Eddy SR., Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*, **2**(1):8, 2001.
- [76] Rogelj, B., Hartmann, CE, Yeo, CH., Hunt, SP., Giese, KP., Contextual fear conditioning regulates the expression of brain-specific small nucleolar RNAs in hippocampus. *Eur J Neurosci*, **18**(11):3089-3096, 2003.
- [77] Rozhdestvensky, TS., Tang, TH., Tchirkova, IV., Brosius, J., Bachelierie, JP., Huttenhofer, A., Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in archaea. *Nucl. Acids Res.*, **31**(3):869-877, 2003.
- [78] Ruff, EA., Rimoldi, OJ., Raghu, B., Eliceiri, GL., Three small nucleolar RNAs of unique nucleotide sequences. *Proc. Natl. Acad. Sci. USA*, **90**(2):635-638, 1993.
- [79] Runte, M., Huttenhofer, A., Gross, S., Kiefmann, M., Horsthemke, B., Buiting, K. The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum. Mol. Genet.*, **10**(23):2687-2700, 2001.
- [80] Runte, M., Kroisel, PM, Gillessen-Kaesbach, G., Varon, R., Horn, D., Cohen, MY., Wagstaff, J., Horsthemke, B., Buiting, K., SNURF-SNRPN and UBE3A transcript levels in patients with Angelman syndrome. *Hum. Genet.*, **114**(6):553-561, 2004.
- [81] Saitou, N., Nei, M., The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.*, **4**(4):406-425, 1987.
- [82] Schattner, P., Decatur, WA., Davis, CA., Ares, M JR., Fournier, MJ., Lowe, TM., Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the *Saccharomyces cerevisiae* genome. *Nucl. Acids Res.*, **32**(14):4281-4296, 2004.

- [83] Sleeman, JE., Lamond, AI., Newly Assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.*, **9**(19):1065-1074, 1999.
- [84] Stalker, J., Gibbins, B., Meidl, P., Smith, J., Spooner, W., Hotz, HR., Cox, AV., The Ensembl Web site: mechanics of a genome browser. *Genome Res.*, **14**(5):951-955, 2004.
- [85] Starck, D., Vergleichende Anatomie der Wirbeltiere auf evolutionsbiologischer Grundlage. Bd. 1.: Theoretische Grundlagen; Stammesgeschichte und Systematik unter Berücksichtigung der niederen Chordata. *Springer*, Berlin, Heidelberg, New York, 1978.
- [86] Storz G. An expanding universe of noncoding RNAs. *Science*, **296**(5571):1260-1263, 2002.
- [87] Szymanski, M., Barciszewski, J., Beyond the proteome: non-coding regulatory RNAs. *Genome Biol.*, **3**(5):reviews0005.1-0005.8, 2002
- [88] Tanzer, A., Stadler, PF., Molecular Evolution of a MicroRNA Cluster. *J. Mol. Biol.*, **339**(2):327-335, 2004.
- [89] Terns, MP., Terns, RM., Small Nucleolar RNAs: Versatile Trans-Acting Molecules of Ancient Evolutionary Origin *Gene Expr.*, **10**(1-2):17-39, 2002.
- [90] Teshima, KM., Innan, H., The Effect of Gene Conversion on the Divergence Between Duplicated Genes. *Genetics*, **166**(3): 1553-1560, 2004.
- [91] Thompson, JD., Higgins, DG., Gibson, TJ., ClustalW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.*, **22**(22):4673-4680, 1994.
- [92] Valentine, JW., Die Entwicklung vielzelliger Pflanzen und Tiere In Mayr, E.(eds) *Evolution*, Spektrum der Wissenschaften, 7.Auflage, p: 60-82, Heidelberg, 1988.
- [93] Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R., Ficner, R., Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell.*, **6**(6):1331-1342, 2000.

- [94] Wuchty, S., Fontana, W., Hofacker, IL., Schuster, P., Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**(2): 145-165, 1999.
- [95] Xia, L., Watkins, NJ., Maxwell, ES., Identification of specific nucleotide sequences and structural elements required for intronic U14 snoRNA processing. *RNA*, **3**(1):17-26, 1997.
- [96] Zuckerkandl, E., Pauling, L., Molecular disease, evolution and genetic heterogeneity. In Kasha, M., Pullmann, B., (eds) *Horizons in Biochemistry*, Academic Press, New York, pp:189-225, 1962
- [97] Zuker, M., Stiegler P., Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucl. Acids Res.*, **9**(1):133-148, 1981.

Abbildungsverzeichnis

1.1	(a) Struktur der Box C/D mit ihren Erkennungsmotiven. Im blau markierten Bereich ist snoRNA komplementär zur rRNA. (b) Die Methylierungsstelle ist 5 Nukleotide von der Box D (D') entfernt (Abbildung Bachelierie[2002])	3
1.2	(a) Konservierte Sekundärstruktur der Box H/ACA. Blau markierte Bereiche stellen die zur rRNA komplementären Bereiche dar. (b) Modifizierungsstelle liegt 14 - 15 nt von Box H und Box ACA entfernt (Abbildung: Bachelierie [2002])	4
1.3	(a) Prozessierungsschritt der snoRNAs die auf Introns oder Hostgenen liegen. Prozessierung kann mittels einem kleineren und größeren Modifizierungsschritt vollzogen werden. (b) Prozessierungsschritt der polycistronischen snoRNA in den Pflanzen. (Abbildungen nach Bachelierie [2000])	6
1.4	(a)Konsensstruktur des Kink-Turn (b)Fibrillarin und NOP58 können im Stamm 2 an die Struktur binden (Abbildung: Henras [2004])	8
1.5	Genomische Organisation der eukaryotischen ribosomalen RNA. A_0 - E sind Syntheschnittstellen der pre-rRNA. (Abbildung: Fromont-Racine[2003])	12
1.6	Tabellarische Zusammenfassung der rRNA Untereinheiten bei den Bakterien und Säugetieren.	13
1.7	Synthese der rRNA von der pre-rRNA. (A) Genomische Organisation der pre-rRNA und Angabe der Schnittstellen. (B) Darstellung der Prozessierung von der 35S pre-rRNA zur 5S, 18S und 28S rRNA (Abbildung: Fromont-Racine [2003]) . . .	14
1.8	(a) Strukturformeldarstellung der Transformation des Uridins in Pseudouridin (b) Prozess der Methylation als Strukturformel dargestellt (Abbildung: Lafontaine [1998])	16

1.9	Sekundärstruktur der <i>P.abyssi</i> 16S rRNA mit den zu modifizierenden Nukleotiden. Zur rRNA komplementäre Bereiche wurden farblich markiert, Kreise zeigen das zu modifizierende Nukleotid (Abbildung: Gaspin [2000])	17
1.10	Ablauf der konzertrierten Evolution eines RNA-Moleküls an zwei unterschiedlichen Organismen. Kodierende Bereiche werden durch Pfeil, nicht kodierende Bereiche durch Linie dargestellt. Mutation wird durch X angegeben. (Abbildung: Liao [1997])	18
1.12	Mir17 Cluster an verschiedenen Organismen (Abbildung: Tanzer [2004])	22
1.13	Evolutionäres Szenario des mir17 Cluster (Abbildung: Tanzer [2004])	23
2.1	Distanzen einer <i>H. sapiens</i> H/ACA snoRNA. Länge von Anfang zu Box H (W), Länge von Box H zu ψ_3 (Y), Länge von ψ_3 zu ψ_4 (Z), Länge von Box H zu Box ACA (X).	26
2.2	Sekundärstrukturdarstellung einer H/ACA snoRNA mittels Mountain Plot	28
2.3	Suchvorgehensweise des Programm SnoScan	30
2.4	Strukturdarstellung einer tRNA, (A) tRNA als Sekundärstruktur dargestellt, (B) tRNA als Dotplot dargestellt.	34
3.1	Alignment der snoRNA Familie U50 gegen Fugu-Sequenz Scaffold_757 (Tr_U50). C: Box C, D: Box D, R: ribosomale Bindungsstelle (<i>Homo sapiens</i>), r: ribosomale Bindungsstelle (<i>Takifugu rubripes</i> / <i>Homo sapiens</i>)	44
3.2	Alignment der snoRNA Familie U33 gegen Fugu-Sequenz Scaffold_757 (Tr_U33). C: Box C, D: Box D, R: ribosomale Bindungsstelle (Säugetiere), r: ribosomale Bindungsstelle (<i>Takifugu rubripes</i>).	45
3.3	Alignment der <i>Orzyias latipes</i> , <i>Takifugu rubripes</i> , <i>Tetradon nirovirdis</i> , <i>Danio rerio</i> , <i>Onchorhynchus mykiss</i> und <i>Salmo trutta</i> U17 snoRNA.	49
3.4	BlastHit der <i>Danio rerio</i> U17 snoRNA Sequenzen. Alle vier U17 snoRNA Kopien (rote Rechtecke) liegen auf dem ribosomalen Protein RPS7. Exonstruktur des Genes RPS7 wird im Genome Browser grün dargestellt.	50

3.5	Genomische Organisation der <i>Orzyias latipes</i> U17 snoRNA Sequenzen. Der obere Bereich stellt den 5'-3' dar, der im Bereich von 214000-219000 3 Exone hat. Im unteren Teil wird der 3'-5' Bereich visualisiert. Die U17 Sequenzen liegen im 3'-5' Bereich, der insgesamt 7 Exone (farbige Rechtecke) hat. Zwischen den Exonen sind die U17 Sequenzen lokalisiert . . .	52
3.6	Zusammenfassung aller U24 Sequenzen. Alignment zeigt Consensusstruktur aller Sequenzen. Erste Zeile gibt komplementäre ribosomale Stellen an (X), Box C (C) und Box D (D) an.	53
4.1	Stammbaum der Vertebraten, fett markierter Punkt ist der Ursprung der Vertebraten (Abbildung nach Fritzsich [2004]). .	55
4.2	Phylogenetischer Baum der die Verwandtschaftsverhältnisse der Vertebratengenome darstellt, die sequenziert wurden (Abbildung: Hedges [2002])	56
4.3	Evolutionäre Geschichte der snoRNA E1 (Mitte) und snoRNA E2 (rechte Seite).	58
4.4	Stammbaumrekonstruktion der snoRNA U17/E1 erstellt mit der Neighbor-Joining Methode im Programm Phylip. Die Astlängen entsprechen der evolutionären Distanz. Distanzen wurden im Programm dnadist ermittelt. Die Werte an den Verzweigungen stellen Bootstrapwerte dar, die nach 1000 Replikationen erhalten wurden. Die unterschiedlich farblichen Boxen zeigen die unterschiedlichen Vertebratengattungen auf. Besonders hervorgehoben wurde die Gruppierung der Nagetiersequenzen, die auf das Gen der snoRNA E2 gesprungen sind (hellblaues Rechteck) Hs: Homo sapiens, Pt: Pan troglodytes, Mm: Mus musculus, Rn: Rattus norvegicus, Oc: Oryctolagus cuniculus , Cf: Canis familiaris, Bt: Bos taurus, Ss: Sus scrofa, Gg: Gallus gallus, Xl: Xenopus laevis, Dr: Danio rerio, Tr: Takifugu rubripes, Tn: Tetradon nigrovirdis Ol: Orzyias latipes, Om: Onchorhynchus mykiss, St: Salmo trutta.	60

- 4.5 Evolution der snoRNA E2 und snoRNA E3 an den Organismen *H. sapiens*, *M. musculus*, *R. norvegicus*, *C. familiaris*, *G. gallus*, *X. tropicalis*, *T. rubripes*, *T. nirgovirdis* und am *D. rerio*. In der ersten Spalte wird die Evolution der snoRNA E2 dargestellt, in der zweiten die Evolution der snoRNA E3. Exone sind durch Rechtecke dargestellt, grüne Kreise stellen snoRNA E3 dar. Blaue Kreise snoRNA U17/E1, rote Kreise snoRNA E2. 62
- 4.6 Stammbaumrekonstruktion der snoRNA E2 erstellt mit der Neighbor-Joining Methode im Programm Phylip. Die Astlängen entsprechen der evolutionären Distanz. Distanzen wurden im Programm *dnadist* ermittelt. Die Werte an den Verzweigungen stellen Bootstrapwerte dar, die nach 1000 Replikationen erhalten wurden. Die unterschiedlich farblichen Boxen zeigen die unterschiedlichen Vertebratengattungen auf. Hs: Homo sapiens, Pt: Pan troglodytes, Mm: Mus musculus, Rn: Rattus norvegicus, Cf: Canis familiaris, Gg: Gallus gallus, Xt: Xenopus laevis, Dr: Danio rerio, Tr: Takifugu rubripes, Tn: Tetradon nirgovirdis. 63
- 4.7 Stammbaumrekonstruktion der snoRNA E3 erstellt mit der Neighbor-Joining Methode im Programm Phylip. Die Astlängen entsprechen der evolutionären Distanz. Distanzen wurden im Programm *dnadist* ermittelt. Die Werte an den Verzweigungen stellen Bootstrapwerte dar, die nach 1000 Replikationen erhalten wurden. Die unterschiedlich farblichen Boxen zeigen die unterschiedlichen Vertebratengattungen auf. Hs: Homo sapiens, Pt: Pan troglodytes, Mm: Mus musculus, Rn: Rattus norvegicus, Cf: Canis familiaris, Gg: Gallus gallus, Xt: Xenopus tropicalis, Dr: Danio rerio, Tr: Takifugu rubripes, Tn: Tetradon nirgovirdis. 65
- 4.8 Szenario, dass die Evolution der snoRNA U36 vom den Teleosteen bis zu den Primaten beschreibt. Die unterschiedlichen reverse komplementären Bereichen werden bei 18S rRNA rot, bei 28S rRNA blau dargestellt. 67
- 4.9 Zusammenfassung aller U36b-Sequenzen. Alignment zeigt Consensusstruktur aller Sequenzen. Erste Zeile gibt komplementäre ribosomale Stellen an (X), Box C (C) und Box D (D) an. 68

4.10	Zusammenfassung aller U36a-Sequenzen. Alignment zeigt Consensusstruktur aller Sequenzen. Erste Zeile gibt komplementäre ribosomale Stellen an (X), Box C (C) und Box D (D) an.	68
4.11	Stammbaumrekonstruktion der snoRNA U36 erstellt mit der Neighbor-Joining Methode im Programm Phylip. Die Astlängen entsprechen der evolutionären Distanz. Distanzen wurden im Programm dnadist ermittelt. Die Werte an den Verzweigungen stellen Bootstrapwerte dar, die nach 1000 Replikationen erhalten wurden. Die unterschiedlich farblichen Boxen zeigen die unterschiedlichen Vertebratengattungen auf. Hs: Homo sapiens, Pt: Pan troglodytes, Mm: Mus musculus, Rn: Rattus norvegicus, Gg: Gallus gallus, Ol: Orzyias latipes, Tr: Takifugu rubripes, Tn: Tetradon nirgovirdis.	70
A.1	20 möglich gefundene <i>X. tropicalis</i> snoRNAs mittels Programm fisher. Rot markierter Bereich stellt die Boxen H bzw. ACA dar, blau markierter Bereich ψ 3/4-Bereiche.	I
A.2	Homologien der snoRNAs zu folgenden ribosomalen Proteinen	II
B.1	Bootstrapbaum der U17/E1 snoRNA	XI
B.2	Bootstrapbaum der E2 snoRNA.	XII
B.3	Bootstrapbaum der snoRNA E3	XIII
B.4	Bootstrapbaum der snoRNA U36	XIV

Tabellenverzeichnis

1.1	Verzeichnis einiger nicht kodierender RNAs deren Funktion bekannt ist.	2
1.2	Proteine, die mit den snoRNA einen Komplex bilden.	8
2.1	Häufigkeiten der Nukleotide, die benötigt werden, um die H-Box zu scoren	26
2.2	Zusammenfassung des snoRNA Modells verwendet bei SnoScan [57]	31
2.3	Zusammenfassung der Schritte, die bei der Homologiemethode angewendet werden	32
3.1	Auflistung aller publizierten und gefundenen snoRNAs an unterschiedlichen Fischen.	36
3.2	<i>H. sapiens</i> HACA snoRNA	38
3.3	Sequenzen, deren ψ_{1-4} Bereiche Pattern für Programm Fisher sind.	38
3.4	Distanzen aller <i>H. sapiens</i> H/ACA snoRNAs sowie maximale und minimale Distanzen der <i>H. sapiens</i> und <i>S. cerevisiae</i> H/ACA snoRNA. W: Länge vom Anfang zur Box H, X: Distanz von Box H zu ACA, Y: Länge von Box H zu ψ_3 , Z: Länge von ψ_3 zu ψ_4 . Minimale Distanzen von W und Z werden im Programm nicht benötigt.	39
3.5	Ermittelte Nukleotidhäufigkeiten in den variablen Positionen der Vertebraten Box H	39
3.6	Häufigkeiten der Nukleotide in den variablen Positionen der <i>S. cerevisiae</i> Box H.	40

3.7	Statistik über die Anzahl der einzelsträngigen Box H und ACA- Motive und die Anzahl der gefundenen ψ_{1-4} -Bereiche an den 20 möglich gefundenen snoRNA-Sequenzen	41
3.8	Tabellarische Zusammenfassung der prozentual benötigten Zeit, Anzahl von Sekunden und Aufrufen der einzelnen Funktionen im Programm <code>Fisher</code> . Funktionen wurden sortiert nach der prozentualen Zeit. Daten wurden mit dem Programm <code>Gprof</code> ermittelt.	42
3.9	Sequenzmotive, die in bestimmten Distanzen zu Box H bzw. Box ACA bei der snoRNA U17 auftreten.	48
4.1	Auflistung aller verwendeten Organismen mit Nennung des Herkunftortes und Genomzustand.	57
4.2	Auflistung der Unterschiede der snoRNA E1, E2 und E3 in Bezug auf Vorkommen, Länge und Aufgabe.	61
A.1	Sequenzen der U17/E1 snoRNA	III
A.2	Sequenzen der E2 snoRNA	VI
A.3	Sequenzen der E3 snoRNA	VII
A.4	Sequenzen der U36 snoRNA	IX

Curriculum vitae

Bettina Müller

geboren am 06.04.1980 in Ostfildern/Ruit

Heimatadresse:

Wengertsleite 2
97199 Ochsenfurt

Schulische Laufbahn:

1991 - 1993	Mozart-Gymnasium, Würzburg
1993 - 2000	Egbert-Gymnasium, Münsterschwarzach Abschluss: <i>Allgemeine Hochschulreife</i>
2000 - 2004	Fachhochschule Weihenstephan, Freising
2002	<i>Vordiplom</i>

Praktika:

März 2002 - Okt.2002	Lynkeus Biotech GmbH, Würzburg
Sept. 2003 - Feb.2004	Lehrstuhl für Bioinformatik, Universität Leipzig, Leipzig
div. weitere Praktika	Bodengesundheitsdienst, Ochsenfurt Südzucker AG Mannheim/Ochsenfurt Hoffmann Menü, Boxberg