

Molekulare Morphologie der 18S rRNA der Bivalvia

Dissertation der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

Eingereicht durch Dipl. Biol. Susanne Feistel

Am 28. November, 2011

Betreut durch Prof. Dr. Gerhard Haszprunar

1. Gutachter: Prof. Dr. Gerhard Haszprunar

2. Gutachter: Dr. Michael Schrödl

Mündliche Prüfung am: 12. Juni, 2012

Inhaltsverzeichnis

1. Zusammenfassung.....	5
2. Einleitung.....	7
2.1 Klassische Morphologie.....	7
2.2 Phylogenetische Systematik.....	8
2.3 Numerische Taxonomie.....	9
2.4 Genetische Analyse.....	9
2.5 Molekulare Morphologie.....	17
2.6 Bivalvia.....	19
2.6.1 Stand des Wissens zur Phylogenie.....	20
2.6.2 Protobranchia.....	24
2.6.3 Pteriomorphia.....	29
2.6.4 Palaeoheterodonta.....	29
2.6.5 Heterodonta.....	31
2.6.6 Anomalodesmata.....	31
2.7 Zielsetzung.....	33
2.7.1 Methodische Bewertung der Molekularen Morphologie.....	33
2.7.2 Phylogenie der Bivalvia mittels Molekularer Morphologie.....	34
3. Material & Methoden.....	37
3.1 Material.....	37
3.1.1 Sequenzen.....	37
3.1.2 Software.....	43
3.2 Faltung.....	59
3.2.1 Faltung nach Gibbs / Minimum Free Energy (MFE).....	59
3.2.2 Faltung mit festen Bindungen, Near Natural (NNC).....	64
3.3 Analyse.....	74
3.3.1 Sekundärstruktur.....	74
3.3.2 Merkmalsdefinition.....	79
3.3.3 Merkmalsmatrix.....	84
4. Ergebnisse.....	90
4.1 Methode „Molekulare Morphologie“.....	90
4.1.1 Minimum Free Energy (MFE).....	90
4.1.2 Near Natural mit Constraint (NNC).....	93
4.1.3 Vergleich von MFE und NNC.....	101
4.2 Phylogenie der Bivalvia.....	106
4.2.1 Protobranchia nach MFE.....	107
4.2.2 Protobranchia nach NNC.....	108
4.2.3 Pteriomorphia nach MFE.....	109
4.2.4 Pteriomorphia nach NNC.....	110
4.2.5 Palaeoheterodonta nach MFE.....	113
4.2.6 Palaeoheterodonta nach NNC.....	113

4.2.7	Heterodonta nach MFE	114
4.2.8	Heterodonta nach NNC.....	114
4.2.9	Bivalvia nach MFE	116
4.2.10	Bivalvia nach NNC	116
5.	Diskussion	119
5.1	Die Methode „Molekulare Morphologie“	119
5.2	Phylogenie der Bivalvia.....	125
5.3	Schlussfolgerungen.....	127
6.	Literaturverzeichnis	128
7.	Anhang	136
7.1	Abbildungen	136
7.1.1	Stammbäume Protobranchia.....	137
7.1.2	Stammbäume Pteriomorphia	144
7.1.3	Stammbäume Heterodonta	157
7.1.4	Stammbäume Palaeoheterodonta.....	164
7.1.5	Stammbäume Bivalvia.....	171
7.2	Glossar	182
7.3	Artenliste Bivalvia	188
7.4	Inhalt des digitalen Anhangs (CD)	189
7.5	Abbildungsverzeichnis.....	190
8.	Danksagung	193

1. Zusammenfassung

Mit der besseren und leichter verfügbaren Rechentechnik seit den 1970er Jahren und der Entwicklung einer schnelleren und billigeren Sequenzierungstechnik von DNA und RNA hat die Bedeutung der numerischen Sequenzanalysen gegenüber der klassischen, deskriptiven Morphologie für phylogenetische Untersuchungen stark zugenommen. Zu den jüngeren numerischen Methoden gehört die Analyse von genetischen Sekundärstrukturen, also fiktiver 2-dimensionaler Faltungen von RNA-Molekülen nach physikalisch-energetischen Kriterien. Für die numerische Durchführung dieser Faltungen existieren verschiedene physikalisch-energetische Modelle, die jeweils in diverse Algorithmen implementiert wurden, die ihrerseits jeweils eine Anzahl willkürlich wählbarer, freier Parameter aufweisen.

Die erste Teilaufgabe dieser Arbeit bestand darin, eine Auswahl vorhandener Methoden und Algorithmen der Sekundärstrukturanalyse an einem biologisch relevanten Beispiel zu testen und ihre Vor- und Nachteile gegenüber anderen Methoden der molekularen Systematik herauszuarbeiten. Als Ergebnis kann grundsätzlich festgestellt werden, dass der methodische Vorteil der Sekundärstrukturanalyse, relativ unempfindlich zu sein gegenüber chemisch irrelevanten Punktmutationen der Primärstruktur, den Nachteilen gegenübersteht, dass die neue Methode zeitaufwändig ist, sich zur automatisierten Bearbeitung wenig eignet, und stark von einer größeren Zahl willkürlicher Entscheidungen des Bearbeiters abhängt, also zu einem erheblichem Maß subjektiven Kriterien unterliegt.

Als biologisches Anwendungsbeispiel wurde die Phylogenie der Bivalvia ausgewählt. Die Verwandtschaft vieler Taxa, zu ermitteln über ihre paarweise Ähnlichkeit gemessen nach bestimmten Kriterien, ist ungeachtet zahlreicher Studien noch immer unsicher. Unterschiedliche taxonometrische Untersuchungen führten bisher häufig zu unbefriedigenden, inkonsistenten Resultaten.

Die zweite Teilaufgabe dieser Arbeit war die phylogenetische Analyse der Bivalvia mit der Methode der Sekundärstrukturanalyse. Als genetisches Ausgangsmaterial wurde die 18S Fraktion der ribosomalen RNA (18S rRNA) bevorzugt, weil (1) für rRNAs mehrere Modelle

bzw. Algorithmen verfügbar waren und (2) deren Mutationsraten dieser Fraktion hinreichend niedrig sind, um zumindest potenziell frühe Verzweigungen bis zum Kambrium zeitlich auflösen zu können. Die Ausführung der Analyse hat ergeben, dass die neue Methode grundsätzlich als Alternative bzw. als Ergänzung zu herkömmlichen Verfahren geeignet ist, um eine phylogenetische Evolutionsgeschichte aus den vorliegenden Primärsequenzen abzuleiten. Jedoch gestaltet sich das Ergebnis als empfindlich gegenüber relativ willkürlichen, methodischen Details und entbehrt der erwünschten Stabilität und Objektivität. Aus diesem Grunde kann die neue Methode im derzeitigen Stand nur wenig dazu beitragen, die bekannten Probleme der Phylogenie der Bivalvia maßgeblich zu verringern.

In dieser Arbeit vorkommende Fachbegriffe sind im Glossar im Anhang erklärt. Die in Abschnitt 4.2, „Phylogenie der Bivalvia“, vorgestellten Stammbäume befinden sich im Abschnitt 7.1, „Abbildungen“. Die beigefügte CD enthält die genutzten Primärsequenzen und die in dieser Arbeit bestimmten Sekundärstrukturen.

2. Einleitung

“Of all natural systems, living matter is the one which, in the face of great transformations, preserves inscribed in its organization the largest amount of its own past history.[...] We may ask the questions where in the now living systems the greatest amount of their past history has survived and how it can be extracted.”

*Emile Zuckerkandl und Linus Pauling
Molecules as Documents of Evolutionary History, 1965*

Vor mehr als 150 Jahren schrieb Charles Darwin sein Buch „Über die Entstehung der Arten durch natürliche Zuchtwahl“. Seitdem versucht die Wissenschaft, die Wege der realen Evolutionsgeschichte nachzuvollziehen und Verwandtschaften zwischen den Arten und Gruppen zu entschlüsseln.

Dazu dienten im Laufe der Zeit viele, sehr verschiedene Methoden, wie die klassische Morphologie, die phylogenetische Systematik, die numerische Taxonomie, kladistische Analyse, sowie eine Fülle an Analysen des Genotyps.

2.1 Klassische Morphologie

Die klassische Morphologie ist bei weitem die älteste Methode, um Organismen zu untersuchen und zu klassifizieren. Ähnlichkeiten zwischen den äußeren und inneren phänotypischen Merkmalen dienten schon *Aristoteles* in seiner *Historia animalium* zur Einteilung der Lebewesen in „verwandte“ Gruppen. Mit der Entwicklung von optischen Instrumenten wie Linsen und Mikroskopen wurden die Vergleiche des Körperbaus auf die Untersuchung von Organen und Geweben ausgedehnt. Heute kann man mit Hilfe der Elektronenmikroskopie Strukturen bis hinab zur Größe der Zelle und ihrer Bestandteile auflösen.

Für eine erste Ordnung der Arten innerhalb der Stämme haben morphologische Merkmale weitgehend ausgereicht. Für die übergeordnete Phylogenie hat sich jedoch gezeigt, dass

diese Merkmale allein nicht aussagekräftig genug sind und zusätzliche Indikatoren benötigt werden.

Arten, die wenig miteinander verwandt sind, können trotzdem große phänotypische Ähnlichkeit besitzen und deshalb eine nahe Verwandtschaft suggerieren. Stammbäume, die ausschließlich auf morphologischen Merkmalen basieren, können dementsprechend systematische Fehler enthalten, die durch konvergente phylogenetische Entwicklungen verursacht werden.

Viele Evolutionsmodelle, wie z.B. die morphologische Taxonomie, basieren auf der Smoothness-Hypothese (*Rechenberg, 1973; Tlustý, 2010*), dass genotypisch benachbarte Arten auch phänotypisch ähnlich sind. Im Gegensatz dazu ist konvergente Evolution ein Beispiel für große genotypische Abstände von phänotypisch ähnlichen Arten.

2.2 Phylogenetische Systematik

Den Grundstein für die heutige phylogenetische Methodik legte der Zoologe *Willi Hennig (1950, 1966)* in den 50er und 60er Jahren des 20. Jh. Die bis dahin übliche Analyse, basierend auf eher intuitiven Beurteilungen morphologischer Merkmale, wurde durch eine stringente Theorie der systematischen Methodik ersetzt, die die evolutionäre Verwandtschaft zu Grunde legt. Basierend auf den Homologie-Kriterien von *Adolf Remane (1952)* wurden insbesondere die heute gängigen Begriffe Plesiomorphie und Apomorphie eindeutig festgelegt. Sie bilden gemeinsam mit den Elementen der Numerical Taxonomy (*Sneath & Sokal, 1973*) die Eckpunkte für die neue Kladistik.

Für die phylogenetische Systematik werden die untersuchten morphologischen Merkmale in einer Matrix zusammengefasst und analysiert. Ein Stammbaum kann zum Beispiel mit dem Maximalen Sparsamkeitsprinzip („Maximum Parsimony“), d.h. einer Minimierung der notwendigen Schritte, durch eine Analyse von Merkmalsausprägungen generiert werden. An dieser Stelle greifen klassische und Molekulare Morphologie ineinander und können in einer Merkmalsmatrix zusammengefasst und ausgewertet werden.

2.3 Numerische Taxonomie

Mit der immer besseren Verfügbarkeit von Rechentechnik, insbesondere seit den 1970er Jahren, gewannen automatisierte Methoden zur systematischen und statistischen Verwandtschaftsanalyse zunehmend praktische Bedeutung. Die numerische Taxonomie setzt auf eine große Anzahl von morphologischen Merkmalen, die beobachtbar und quantifizierbar sein müssen. Mittels komplexer, computergestützter Verfahren soll so ein objektives Maß für Ähnlichkeit von Individuen gefunden werden. Die paarweise Bewertung von Ähnlichkeiten in Form einer Matrix kann dann als Ausgangspunkt für die mathematische Konstruktion eines taxonomischen Systems, in der Regel in Form eines graphentheoretischen Baums, dienen (*Fitch & Margoliash, 1967*). Phylogenetische Gesichtspunkte wurden bei dieser Methode zunächst nicht berücksichtigt (*Sokal & Rohlf, 1995*), später aber in Analysesoftware zur Phylogenetik implementiert (zum Beispiel HENNIG86 und PAUP (*Swofford, 2003*)).

Der reguläre Aufbau der DNA gestattet es, sie auf einfache Weise, als mathematisches Modell darzustellen. Die Numerische Taxonomie zur Analyse bietet sich an (*Jiménez-Montaña et al, 2004*).

2.4 Genetische Analyse

Die Erbsubstanz trägt nicht nur die genetische Information, die zum ontogenetischen Entfalten und zur Lebenserhaltung eines Organismus notwendig ist. Der Vergleich der DNA verschiedener Arten erlaubt Rückschlüsse auf ihre Verwandtschaft und phylogenetische Hypothesenbildung. Die Individuen und ihre Arten existieren nur für begrenzte Zeiträume, ihre Gene jedoch reichen weit darüber hinaus in die Vergangenheit zurück – bis zur Entstehung des Lebens. Da bestimmte Mutationen in der DNA keinen Einfluss auf die biochemische Stabilität oder Funktion des durch die Sequenz codierten Moleküls haben, bleiben alle Veränderungen, die nicht zu einem phänotypischen Nachteil des Organismus führen, in der DNA über längere Zeit erhalten. So stellt die DNA mit diesen zufälligen „Markern“ grundsätzlich ein „Tagebuch“ ihrer eigenen Entwicklungsgeschichte dar, das entschlüsselt werden kann (*Zuckerlandl & Pauling, 1965; Feistel, 1990; Wells, 2004*).

Nach der RNA-Welt-Hypothese (*Gilbert, 1986*) ist die RNA der älteste Baustein der heutigen Lebensformen. Sie arbeitet sowohl als universeller Informationsspeicher als auch als Katalysator chemischer Reaktionen. Damit vereint die RNA in sich alle Funktionen und Eigenschaften, die später in der Evolution durch DNA und Proteine übernommen wurden. Um „Deep Phylogeny“ zu untersuchen, ist es deshalb sinnvoll, die verschiedenen Formen, in denen RNA in Lebewesen vorkommt, zu analysieren.

Die rRNA (ribosomale RNA) ist am Aufbau und der enzymatischen Aktivität des Ribosoms und damit direkt an der Proteinsynthese beteiligt: Eukaryotische Ribosomen setzen sich aus zwei Untereinheiten, bestehend aus Proteinen und mehreren Abschnitten rRNA zusammen. Die rRNAs werden je nach ihrem Sedimentationskoeffizienten benannt: Die große Untereinheit der Ribosomen besteht aus drei Fraktionen, 28S, 5,8S und 5S rRNA; die kleine Untereinheit besteht aus der 18S rRNA.

Die rRNA als Element der Proteinsynthese ist essentiell für die Lebensfähigkeit eines Organismus. Zum einen sind Abschnitte der Primärsequenz der rRNA direkt an den Reaktionen beteiligt, zum anderen ist die auf der Primärsequenz aufbauende räumliche Struktur entscheidend für die Funktion.

Die RNA wird, wie alle anderen Bestandteile der Zelle, durch bestimmte Abschnitte der DNA codiert. Die 18S rDNA codiert die 18S rRNA. Im Folgenden wird immer dann von rRNA gesprochen, wenn die an der Proteinsynthese beteiligte funktionelle rRNA gemeint ist. Die rRNA ist auch die Form, die in einer 3D-Struktur vorkommt und analysiert wird. Wenn es aber um die Primärsequenz und ihre Mutation geht, wird hier der Begriff rDNA verwendet für den Abschnitt, der die rRNA codiert.

Die räumliche Struktur der eukaryotischen DNA ist bereits seit 1953 durch die Arbeit von Watson und Crick bekannt. Während bei der DNA die Sequenz (die Primärstruktur) mehr phylogenetische Informationen birgt, als ihre räumliche Struktur (Sekundär- und Tertiärstruktur), sind die Dinge bei der RNA wesentlich komplizierter.

RNA-Moleküle sind, im Gegensatz zur DNA, Einzelstränge. Sie können in weiten Abschnitten auf sich selbst zurückfalten und weisen dort eine Helix-Struktur auf. Die Doppelstrangabschnitte werden durch flexible Einzelstrangabschnitte, sogenannte Loops, unterbrochen. Die resultierende räumliche Struktur (Sekundär- und Tertiärstruktur) ist grundlegend für die Funktion der RNA (Zucker & Stiegler, 1981; Rijk & Wachter, 1997). Eine detaillierte Erklärung zur Sekundärstruktur von RNA findet sich in Abschnitt 3.3.1, „Sekundärstruktur“.

Entstehen Mutationen in der rDNA, die den „Phänotyp“ und die funktionellen Abschnitte der dadurch codierten rRNA maßgeblich verändern, wird die gesamte Funktion gestört – der Organismus ist mit hoher Wahrscheinlichkeit nicht überlebensfähig. Da Mutationen nur an wenigen Positionen toleriert und vererbt werden können, ist die effektive Mutationsrate der Gensequenz gering - die funktionellen sowie die strukturellen Abschnitte sind hoch konserviert. In der rDNA bleiben dementsprechend nur „stille“ oder durch Proteine kompensierbare Mutationen erhalten und werden an die Nachkommen weitergegeben. Erwartungsgemäß sind zahlreiche Abschnitte der 18S rDNA funktionsgebunden und deshalb fixiert. Für die Analyse in dieser Arbeit sind deshalb nur wenige Abschnitte der rDNA von Bedeutung.

Die rRNA kommt in fast unveränderter Form (mit leichten Unterschieden zwischen Pro- und Eukaryoten) in allen lebenden Organismen vor. Ihre Primärsequenz ist weitgehend hoch konserviert, um ihre Funktion in der Proteinsynthese zu gewährleisten. Bisherige Erfahrungen weisen darauf hin, dass die Sekundär- und Tertiärstrukturen von Molekülen noch einmal deutlich konservativer sind als die jeweiligen Primärstrukturen, d.h. die Nukleotid- oder Aminosäuresequenzen (Billoud *et al.*, 2000; Lydeard *et al.*, 2000, 2002; Wang & Lee, 2002; Ender & Schierwater, 2003). So ist etwa bei der Proteinsynthese aufgrund der Redundanz des genetischen Codes ($4^3 = 64$ Nukleotidtriplets: wenn man die „Stop-Codons“ abzieht, werden 20 Aminosäuren von 61 Triplets codiert) die Aminosäuresequenz konservativer als die Nukleotidsequenz. In analoger Weise wie bei Proteinen gestattet auch die Untersuchung der Primärsequenz und Sekundärstruktur von rRNA im Prinzip eine

Einleitung

Analyse der zeitlichen genetischen Variabilität und damit der phylogenetischen Verwandtschaft zwischen den Taxa über große Zeitspannen.

Bereits in den Siebziger Jahren des 20. Jh. wurde aufgrund einer Ähnlichkeitsanalyse der ribosomalen RNA (16S und 18S) die Einteilung aller lebenden Organismen in die drei Domänen der Bakterien, Archaeen und Eukaryoten vorgenommen (*Woese & Fox, 1977*).

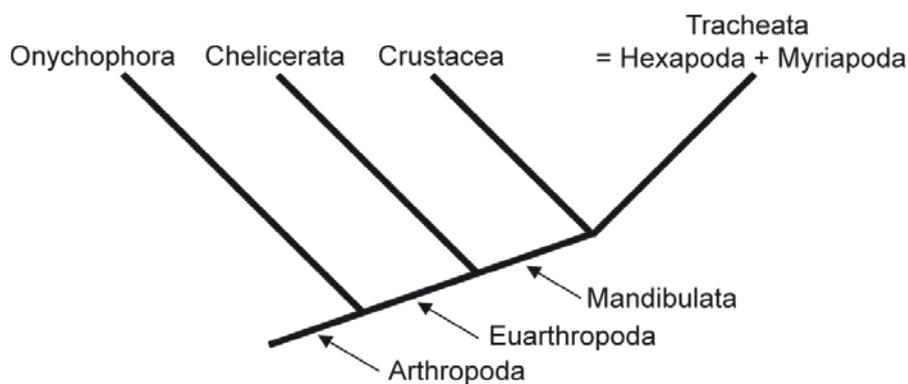
Anhand von phylogenetischen Untersuchungen an der 18S rRNA, der zur Analyse bevorzugt verwendeten Einheit der rRNA, wurde auch die derzeit diskutierte Einteilung der Protostomia in Ecdysozoa (Arthropoda, Tardigrada, Onychophora, Nematoda, Kinorhyncha, Gastrotricha und Priapulida) und Lophotrochozoa (u.a. Brachiopoda, Entoprocta, Phoronida, Mollusca, Annelida, Nemertea, Sipuncula, Echiura, Bryozoa) vorgenommen (*Aguinaldo et al., 1997; Giribet, 2003; Telford et al. 2008; Waagele et al., 2009*). Warum diese RNA für diese Fragestellung besonders geeignet ist, wird später im Detail diskutiert.

Die mitochondriale DNA und die DNA der pflanzlichen Plastiden bieten neben der 18S rRNA eine weitere Möglichkeit, die Phylogenie von Organismen zum Teil bis ins Präkambrium zurück zu verfolgen. Durch die vollständige Analyse der Genome von Rotalgen und Cyanobakterien konnte zum Beispiel belegt werden, dass alle Plastiden als Endosymbionten der Eukaryoten, einen gemeinsamen Ursprung haben (*McFadden & van Dooren, 2004*).

Ein besonders erfolgreiches Beispiel für die Verwandtschaftsanalyse mittels mtRNA, der RNA von Mitochondrien, ist die menschliche Genealogie und globale Migrationsgeschichte (*Wells, 2004; Hamel, 2007*).

Fast ebenso hoch konserviert wie rRNA und mtRNA sind die sogenannten Hox-Gene, welche die Zelldifferenzierung und Morphogenese eines Individuums steuern. Dementsprechend sind Veränderungen auch in dieser Genfamilie in den meisten Fällen letal. Eine Analyse ist z.B. im Fall der Phylogenie der Arthropoda interessant. Die Familie der Hox-Gene datiert vor die Radiation der Arthropoda im frühen Kambrium (*Grenier et al., 1997*). Dadurch bietet sich dieser Genabschnitt an, die Verwandtschaft der Großgruppen innerhalb der Gliederfüßler zu untersuchen.

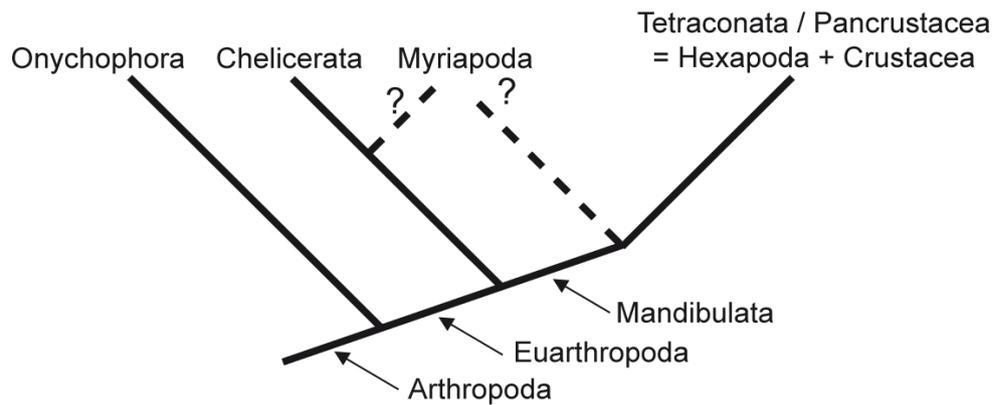
Die Ähnlichkeiten der Morphologie des Atmungssystems und die Reduktion eines Antennenpaares deutet auf eine nahe Verwandtschaft von Hexapoda und Myriapoda hin (= Tracheata / Antennata). Als Schwestergruppe von beiden wurden aufgrund ihrer komplexen Mandibeln die Crustacea eingeordnet (Mandibulata). Basal stehen Chelicerata und Onychophora (Abbildung 1) (Ax, 1999).



nach Ax, 1999

Abbildung 1: Klassische Phylogenie der Arthropoda nach morphologischen Merkmalen (Ax, 1999)

Im Widerspruch zur morphologischen Ähnlichkeit stehen die molekularen Daten zur genetischen Verwandtschaft. Die Untersuchung verschiedener Hox-Gene sowie verschiedener rRNA und neuronale Strukturen, sowie neuerdings auch sogenannte EST-Datensätze (Adams *et al.*, 1991) widerlegen den klassischen „morphologischen“ Stammbaum. Die genetische Analyse hat ergeben, dass sich die Hexapoda aus Crustaceen herleiten (Tetraconata oder Pancrustacea-Konzept) und mit den Myriapoda nicht näher verwandt sind. Letztere sind näher mit den Chelicerata verwandt, als mit dem Pancrustacea-Zweig (Cook *et al.*, 2001; Deutsch, 2011) (Abbildung 2).



nach Deutsch, 2011

Abbildung 2: Neue Phylogenie der Arthropoda unter Einbeziehung von molekularen Daten, insbesondere der Analyse von Hox-Genen und rRNA

Für die Untersuchung präkambrischer Evolution wurde durch *David & Alm (2010)* ein detailliertes Modell zur Analyse von fast 4000 Genfamilien von 100 Genomen von Arten aus allen drei Domänen entwickelt. Dabei wurden Informationen gewonnen, die eine „Archaische Expansion“ vor 3,3 bis 2,8 Milliarden Jahren vermuten lassen. Aus diesem Zeitraum stammen 27 Prozent aller heute existierenden Genfamilien. Die dabei neu entstandenen Gene für den biochemischen Elektronentransfer bedeuteten einen großen Fortschritt für die Entwicklung des Lebens auf der Erde: Durch sie wurden Photosynthese und Sauerstoffatmung möglich. Erstere führte vor 2,4 Milliarden Jahren zum ersten „Great Oxidation Event“. Der produzierte Sauerstoff veränderte die Lebensbedingungen auf der Erde entscheidend: Das in den Weltmeeren gelöste Eisen oxidierte und fiel aus, Sauerstoff gelangte in immer größeren Mengen in die Atmosphäre (*Lyons & Reinhard, 2009*). Ein großer Teil der damals existierenden anaeroben Mikroorganismen könnte deshalb ausgestorben sein. Die Gene, die damals von den überlebenden Mikroorganismen weitergegeben wurden, existieren bis heute.

Für eine der bisher umfangreichsten phylogentischen Analysen von *Goloboff et al. (2009)* wurden von mehr als 73 000 Arten mehr als 600 morphologischen Merkmale mit 13 verschiedenen Genen (unter anderem verschiedene rRNA, COX I-II und CytB) kombiniert und ausgewertet. Das Resultat war ein Stammbaum (Abbildung 3), der alle eukaryotischen Großgruppen umfasst und in Beziehung stellt. Auch die Mollusca wurden mit Hilfe großer

Datensätze untersucht (*Kocot et al., 2011; Smith et al., 2011*). Solche Analysen von mehreren Tausend Arten und/oder Genen sind mit Hilfe sogenannter ESTs (Expressed Sequence Tags) praktisch machbar geworden. ESTs sind kurze, für das spezielle Gen möglichst eindeutige Genabschnitte (bzw. deren Komplementärsequenzen) am Anfang und Ende der jeweiligen kompletten Primärsequenz. Durch ESTs können bestimmte gesuchte Gene in einer großen Anzahl nicht näher identifizierter Sequenzen (in der GenBank) numerisch lokalisiert und isoliert werden (*Adams et al., 1991*).

Die oberen Beispiele verdeutlichen die neue Qualität der Forschung durch die Analyse einer großen Anzahl von Sequenzen. Die Sequenzierung von DNA-Molekülen ist mit der Zeit preiswerter und schneller geworden. Während im Human Genome Project noch dreizehn Jahre notwendig waren, um die etwa drei Milliarden Basenpaare der menschlichen DNA zu sequenzieren, wurden sie wenige Jahre später schon in einem Monat entschlüsselt (*Pushkarev et al., 2009*).

Fortschritte in der Sequenzierungstechnik machen es möglich, dass heute eine große Anzahl an neuen Sequenzen jedes Jahr entschlüsselt wird. Datenbanken wie die GenBank stellen die Sequenzen der Allgemeinheit per Internet zur Verfügung. Während die Datenmenge exponentiell wächst, sind kritische Auswertungsmethoden trotz einiger bemerkenswerter Fortschritte (Parsimony Ratchet von *Nixon, 1999*; Cladistics TNT von *Goloboff et al., 2008*; RAxML-VI-HPC von *Stamatakis, 2006*) zurückgeblieben. Die Menge an bekannten Sequenzen bietet eine Vielzahl an Möglichkeiten, die darin verborgenen Informationen aufzuspüren, die weit über die numerische Ähnlichkeit der Primärsequenzen hinausgeht.

Für diese Arbeit wurde sequenzierte 18S rDNA genutzt, deren Basenfolgen im Internet frei verfügbar sind.

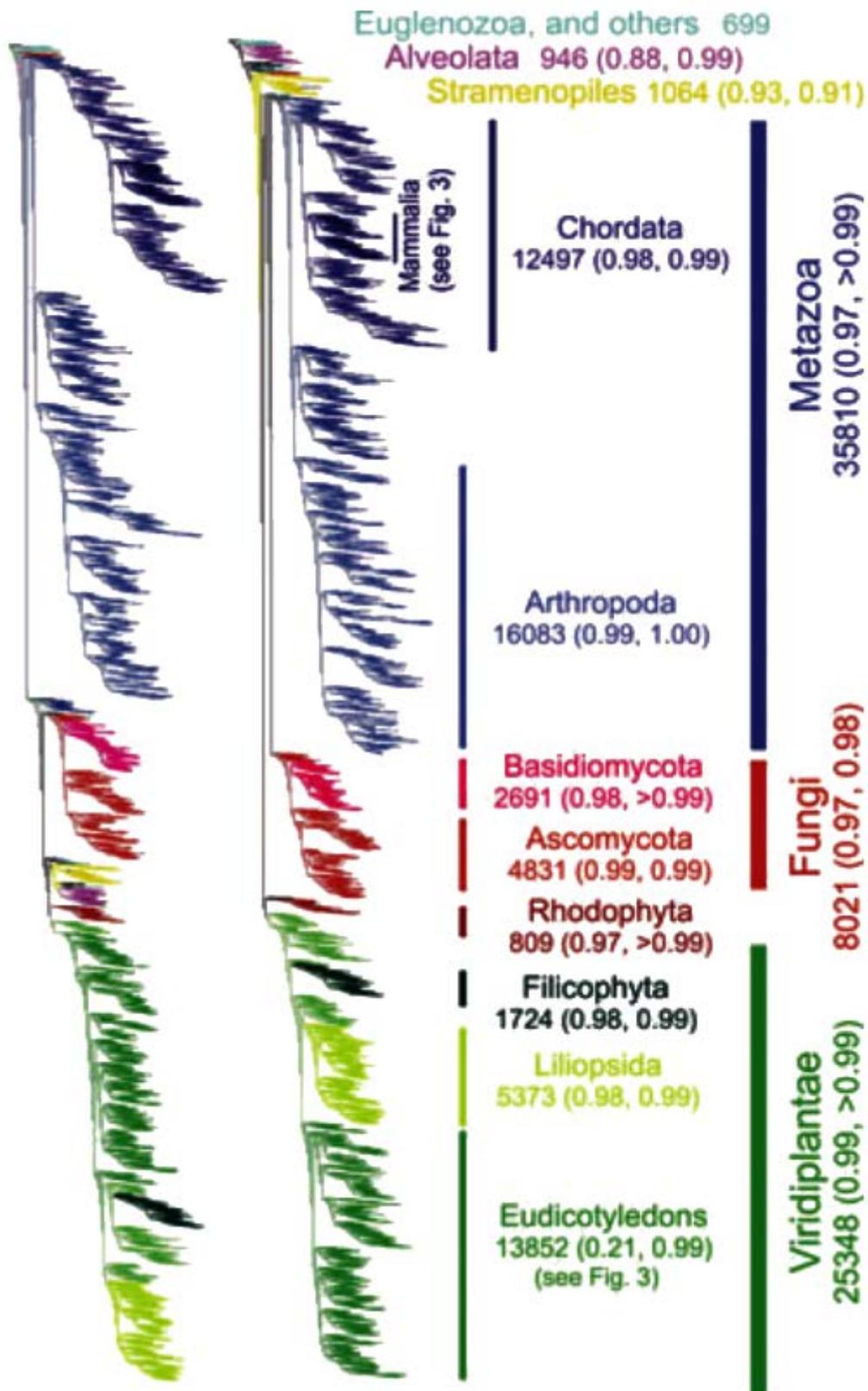


Abbildung 3: Stammbaum der eukaryotischen Großgruppen, ermittelt aus morphologischen und genetischen Informationen von 73 060 Arten. Stammbaum links: nur molekulare Daten; Stammbaum rechts: kombinierter Datensatz. Aus Goloboff (2009).

2.5 Molekulare Morphologie

Die große Bedeutung der rRNA und ihre vermutlich seit mehr als einer Milliarde Jahren innerhalb der Eukaryonta prinzipiell kaum veränderte Sekundärstruktur legen es nahe, letztere genauer zu untersuchen. Die ersten Algorithmen zur Ermittlung der Sekundärstruktur von RNA wurden Ende der 1970er Jahre entwickelt:

- ⇒ 1978 Nussinov-Algorithmus (*Nussinov et al., 1978*)
- ⇒ 1981 Zuker-Algorithmus (*Zuker & Stiegler, 1981*)
- ⇒ 1985 Sankoff-Algorithmus (*Sankoff, 1985*)

Grundlage der drei Algorithmen ist die Stärke der biochemischen und physikalischen Wechselwirkungen zwischen den Nukleotiden der Sequenz im gefalteten Zustand. Der Nussinov-Algorithmus versucht, das Molekül so zu falten, dass die Anzahl von entstehenden Basenpaaren maximal wird. Biologisch und physikalisch ist diese Herangehensweise für die vorliegende Arbeit jedoch nicht sinnvoll. Die Algorithmen von Zuker und Sankoff beruhen auf thermodynamischen Modellen und legen die Minimum Free Energy (MFE) im gefalteten Molekül zugrunde. Dies führt (im weitesten Sinne) zu biologisch sinnvolleren Strukturen. Der Sankoff-Algorithmus soll die Möglichkeit bieten, die Aufgaben des Alignments der Sequenzen, der Sekundärstruktur-Vorhersage und die daraus resultierende Berechnung phylogenetischer Stammbäume simultan zu lösen. Dieses Verfahren würde die hier gesuchte Molekulare Morphologie vollständig überspringen und damit einer kritischen Bewertung oder Modifikation entziehen. In der vorliegenden Arbeit wurden deshalb Programme verwendet, die auf dem Zuker-Algorithmus aufbauen.

Die relativ einfache Vorhersage der Sekundärstruktur, z.B. mit dem Zuker-Verfahren, ohne Laboraufwand oder experimentell-technische Hilfsmittel macht ihre Untersuchung um vieles leichter und kostengünstiger. Während der letzten Jahre wurde daraus eine zukunftsweisende Methode "Molekulare Morphologie" entwickelt, d.h., die morphologische Analyse von 2D- oder 3D-Strukturen der untersuchten Moleküle eingeführt (*Billoud et al., 2000; Wang & Lee, 2002; Ender & Schierwater, 2003*). Da sich die Strukturvorhersage und die Darstellung von 3D-Strukturen mit heutigen numerischen Mitteln noch immer als schwierig

erweist, stellt die Analyse der vereinfachten 2D-Strukturen einen hilfreichen Kompromiss zwischen der vergleichsweise einfachen Sequenzanalyse und der komplizierten Bewertung der biochemischen Aktivität des gefalteten Moleküls dar.

Die Molekulare Morphologie hat zwei grundlegende methodische Teilaufgaben, die von relativ willkürlichen Entscheidungen der Anwender abhängen und für die es alternative Ansätze gibt, die in dieser Arbeit kritisch bewertet werden. Das ist zum einen das Konstruktionsverfahren der 2D-Struktur aus der Primärsequenz, das verschiedene Algorithmen einsetzt, um automatisch ein Minimum der Bindungsenergie der gefalteten 2D-Struktur zu finden. Neben der unklaren Effektivität und Eindeutigkeit des Lösungsalgorithmus gibt es weitere Freiheitsgrade, die das Ergebnis wesentlich beeinflussen können, wie etwa die angenommene Temperatur oder der pH-Wert der Umgebung. Zum anderen besteht die Aufgabe darin, ein Maß für die Ähnlichkeit zwischen zwei berechneten 2D-Strukturen geeignet zu definieren und auf jedes Paar von untersuchten Molekülen anzuwenden. Bisher führen visuelle Vergleichsverfahren zu deutlich besseren Ergebnissen als automatisierte, sie sind jedoch mit einem hohen Zeit- und Arbeitsaufwand verbunden und unterliegen subjektiven und möglicherweise von Zeit zu Zeit oder Person zu Person wechselnden Kriterien.

Im Vergleich zu früheren Methoden bietet die Molekulare Morphologie einen qualitativ neuartigen Datensatz zur Untersuchung phylogenetischer Aufgabenstellungen. Molekulare Morphologie zeichnet sich durch einige charakteristische Eigenschaften aus:

- ⇒ Ein Vorteil des Verfahrens besteht darin, dass irrelevante („blinde“) Mutationen, die als Abweichungen in den 1D-Sequenzvergleichen auftreten, kaum eine Veränderung der 2D-Strukturen bewirken. Zusätzlich lassen genau die Regionen, die in den Primärsequenzen die größten Schwierigkeiten in der Sequenzanalyse bereiten – aufgrund von unklaren Alignments – und deshalb ausgeschlossen werden müssen, erkennbare phylogenetische Muster erwarten (*Billoud et al., 2000*). Das gleiche gilt für Insertionen und Deletionen, die sogenannten „InDels“, Mutationen im Genom, die durch das Einfügen oder Herausfallen einer oder mehrerer Basen entstehen.

-
- ⇒ Die fatalen Auswirkungen einer „falschen“ Mutation schränken die „Tagebuchfunktion“ der rDNA stark ein. Typischerweise liefern die Daten keinen voll aufgelösten phylogenetischen Baum. Sie bieten jedoch Anhaltspunkte für bestimmte Verzweigungen im Baum.
 - ⇒ Die Daten (Vorkommen oder Fehlen von molekular-morphologischen Merkmalen) können als potentielle Synapomorphien direkt in eine morphologische Datenmatrix integriert werden. Im Gegensatz zu Analysen der Primärsequenz können molekular-morphologische Merkmale direkt mit morphologischen kombiniert und ausgewertet werden.
 - ⇒ Bereits bekanntes Wissen zu den Variationen in Sekundär- und Tertiärstrukturen der untersuchten Moleküle verbessert die Modelle zur Nukleotid-Substitution, die Anwendung in der Sequenzanalyse der jeweiligen Moleküle finden (*Wang & Lee, 2002*). Die Analyse der Primärsequenz hängt vor allem von einem sorgfältigen Alignment der Sequenzen ab – sodass analoge Bereiche gleichgesetzt werden und Mutationen, Insertionen und Deletionen als solche richtig bei der Untersuchung erkannt werden. Durch eine bekannte Sekundärstruktur kann das Alignment maßgeblich verbessert werden.

Mit diesen Eigenschaften bietet die Molekulare Morphologie nicht notwendig die ultimativ beste Methode. Sie kann aber dazu beitragen, nicht gefestigte phylogenetische Annahmen - stets Hypothesen - zu bestätigen und neue Impulse für die Aufklärung noch unsicherer Verwandtschaftsverhältnisse zu liefern.

2.6 Bivalvia

Als eine Anwendung für die Methode der Molekularen Morphologie soll in dieser Arbeit die Phylogenie der Bivalvia untersucht werden.

Die Bivalvia sind ein gut untersuchtes Taxon des Tierreichs. Die verschiedenen Nutzungsmöglichkeiten machten ein grundlegendes Wissen um die Muscheln schon frühzeitig notwendig. Als Nahrungsmittel sind besonders Miesmuscheln, Austern oder Jakobsmuscheln bekannt, aber auch Herz- und Venusmuscheln werden kommerziell

genutzt. Bereits im Altertum waren Perlen und Perlmutter als Schmuck und Dekoration beliebt, aus dem Mittelalter sind Handschuhe aus Byssusfäden („Muschelseide“) erhalten (Ausstellungsstück, Zoologische Sammlung Rostock). Nicht nur die Schiffe von Kolumbus wurden von den Schiffsbohrmuscheln angegriffen. Die ökologische Wirkung der Muscheln als Filtrierer ist erst in der Neuzeit ausführlich untersucht worden. Trotz ihrer Bedeutung ist die Verwandtschaft innerhalb der Bivalvia unzureichend untersucht und wurde in den letzten 100 Jahren mehrfach revidiert.

Die Radiation der Großgruppen und Ordnungen innerhalb der Bivalvia fand vom späten Kambrium bis zum Silur statt; fast alle ökologischen Nischen der Muscheln waren ausgeprägt. Als geeignetes Gen, das diesen Zeitbereich auflösen kann, wurde die sehr konservative 18S rDNA ausgewählt (*Winnepenninckx et al., 1996*). Die Molekulare Morphologie ihrer Sekundärstruktur soll Aufschluss geben über die Verwandtschaftsverhältnisse innerhalb der bislang nur unsicher klassifizierbaren Gruppen der Bivalvia. Für die Aufschlüsselung der rezenten Gattungen und Familien eignet sich mitochondriale DNA, wie 12S oder 16S rRNA (*Puslednik & Serb, 2008*).

2.6.1 Stand des Wissens zur Phylogenie

Die klassische Morphologie ist bei weitem die älteste Methode, um Organismen zu untersuchen und zu klassifizieren. Auch die Bivalvia, durch ihr häufiges Vorkommen, ihre vielseitigen Nutzungsmöglichkeiten und nicht zuletzt ihre Schönheit, wurden seit alters her gern untersucht und beschrieben.

Die Schalen sind das bekannteste Merkmal der Bivalvia. In der Paläontologie ist das Vorkommen von Muschelschalen einer der wichtigsten Indikatoren für Erdzeitalter. Auch postglaziale Stadien der Entwicklung der Ostsee wurden beispielsweise nach der Salztoleranz der jeweils vorherrschenden Muschelpopulation klassifiziert, so z.B. das „Yoldia-Meer“ nach *Portlandia arctica*, früher als *Yoldia arctica* bezeichnet. Außer der Tatsache, dass Form und Größe der Schalen vom Alter und genetischen Eigenschaften der Muscheln abhängen, stellen sie auch einen Umweltindikator dar, der z.B. stark von der Kalzium-Übersättigung und damit vom Salzgehalt, von der Temperatur und dem jeweiligen pH-Wert der marinen

Umwelt abhängt (*Ivany et al., 2011; Marion et al., 2009*). Form, Ultrastruktur, Schichtzusammensetzung, Mantellinie, Muskelabdrücke und Schlosszähne der Schale bilden für Paläontologen und Zoologen seit langer Zeit wichtige Merkmale zur taxonomischen Einordnung der Bivalvia (Abbildung 4).

Die Untersuchung der Organsysteme rezenter Arten liefert eine Reihe weiterer Unterscheidungsmerkmale. Die verschiedenen Typen von Kiemen (Proto-, Fili-, Eulamelli- oder Septibranchien, siehe Abbildung 5), die Ausbildung und Faltung des Mantelrandes, Anzahl und Art der Schließmuskeln sowie Magenstrukturen bieten viele Anhaltspunkte zur Klassifizierung der Muscheln.

Bei der Entwicklungsgeschichte der Muscheln stellt sich jedoch ein besonderes Problem: Multiple konvergente Entwicklungen betreffen fast alle morphologischen Merkmale und erschweren die Klassifizierung. So ergeben Analysen der Schlosszähne (*Newell, 1965, 1969*) oder der Kiemen (*Franc, 1960*) sehr unterschiedliche Stammbäume – je nachdem auf welche Merkmale der Schwerpunkt der Analyse gesetzt wurde.

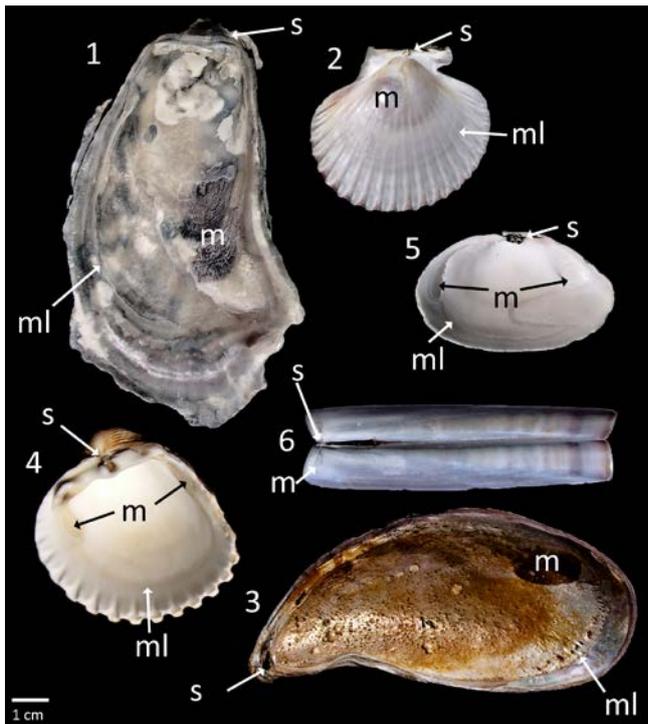


Abbildung 4: Verschiedene Muschelschalen: 1-3 Pteriomorpha (1 - Ostreoidea, 2 - Pectinoidea, 3 - Mytiloidea), 4-6 Heterodonta (4 - Veneroidea: Cardioidea, 5 - Myoidea: Myoidea, 6 - Veneroidea: Solenoidea). Markiert sind charakteristische Merkmale: m – Muskelansatzstellen, ml – Mantellinie, s – Schloss mit Schlosszähnen.

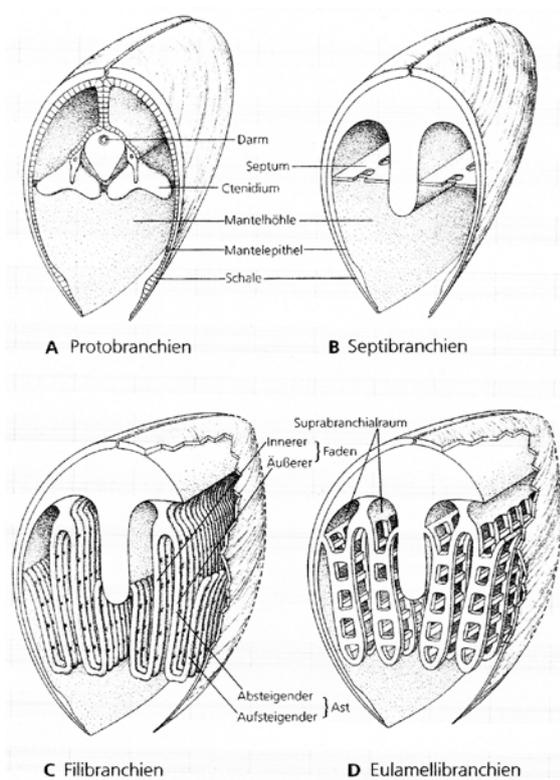


Abbildung 5: Verschiedene Kiementypen bei den Bivalvia: A - Protobranchien, B - Septibranchien, C - Filibranchien, D - Eulamellibranchien (aus Westheide & Rieger, 1996)

Der jüngste Ansatzpunkt der morphologischen Forschung ist deshalb die Ontogenese der Bivalvia. Nach der Grundregel von *Haeckel (1866)*, dass die Ontogenese die Phylogenese rekapituliert, könnte die Larvalentwicklung entscheidende Hinweise zu den oben genannten Parallelismen liefern (*Bändel, 1988; Cannuel & Benninger, 2006; Cannuel et al., 2009; Evseev et al., 2005*).

Eine der ersten wichtigen Arbeiten zur Systematik der Muscheln war die von *Thiele (1935)*. Er unterteilte die Bivalvia in drei Großgruppen, Taxodonta, Anisomyaria und Eulamellibranchia. Die Taxodonta enthielten die heutigen Protobranchia und die Arcoida; fünf Gruppen der heutigen Pteriomorphia bildeten die Anisomyaria. Die größte Ordnung der Eulamellibranchia teilte sich in vier Unterordnungen – Schizodonta (heute Teil der Palaeoheterodonta), Heterodonta, Adapedonta (polyphyletische Reste „Myida“ und „Venerida“) und Anomalodesmata.

In den 60er Jahren wurden durch *Cox (1960)* und *Newell (1965, 1969)* die morphologischen Erkenntnisse zusammengefasst und die fünf Unterklassen Protobranchiata, Pteriomorphia, Palaeoheterodonta, Heterodonta und Anomalodesmata eingeführt. Diese morphologisch begründeten Taxa blieben zwar bis heute erhalten, ihre Positionen im Stammbaum der Bivalvia haben sich jedoch inzwischen mehrfach gegeneinander verschoben (Abbildung 6).

Seit der Nutzung von DNA-Sequenzen zur Verwandtschaftsanalyse wurden von den Mollusca fast alle Taxa untersucht, lediglich die Verwandtschaft der Bivalvia blieb bis in die späten 1990er Jahre in Hinsicht auf Sequenz-Analyse unbeachtet. In den letzten 15 Jahren wurde das genetische Material zahlreicher Arten der Bivalvia sequenziert und in der GenBank zur Verfügung gestellt.

Während sich die morphologische Untersuchung durch Parallelismen in der Evolution der Bivalvia unsicher gestaltet, entstehen in der molekularen Analyse wiederum Probleme durch rapide Radiationsereignisse (*Adamkewicz et al. 1997, Lindgren et al. 2004; Plazzi & Passamonti, 2010*). Diese erschweren zum Beispiel die phylogenetische Auflösung der Pteriomorphia in ihre Untergruppen.

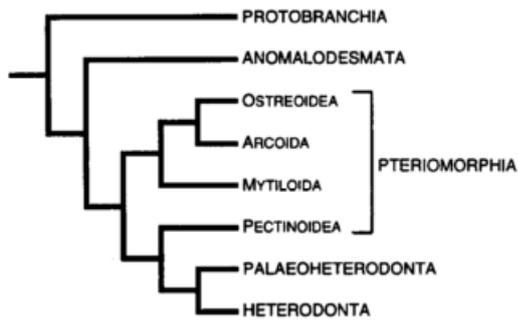
Inzwischen konnte eine Anzahl molekularbasierter Arbeiten zur Phylogenie der Bivalvia (Tabelle 1) einige offene morphologische Fragen klären. Die Position der Anomalodesmata innerhalb der Heterodonta wurde bestätigt. Die Monophylie vieler Taxa (Heterodonta, Palaeoheterodonta, Pteriomorphia, Arcoida, Mytiloida) wurde stärker belegt.

Dennoch bleibt die Entwicklungsgeschichte der Muscheln und dabei insbesondere die Heterodonta noch immer in vielen Punkten im Dunkeln. Die Zugehörigkeit der Nuculanoidea zu den Protobranchia wurde durch die molekulare Analyse in Frage gestellt. Einige Stammbäume auf Niveau von Familien und Ordnungen bleiben strittig und bedürfen weitergehender Untersuchungen.

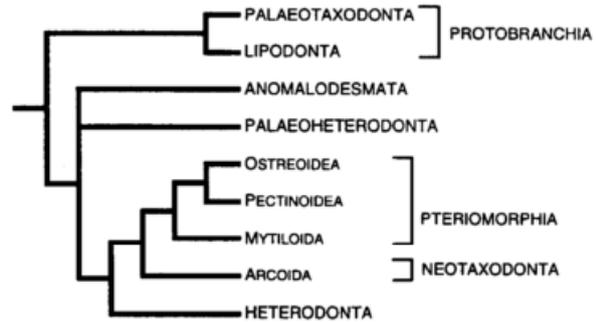
Im Folgenden soll ein kurzer Überblick über den gegenwärtigen Stand der Systematik der Bivalvia (Abbildung 7) gegeben werden. Die Abbildungen 6 bis 11 sollen die noch immer bestehenden Unsicherheiten in diesem Forschungsgebiet verdeutlichen.

2.6.2 Protobranchia

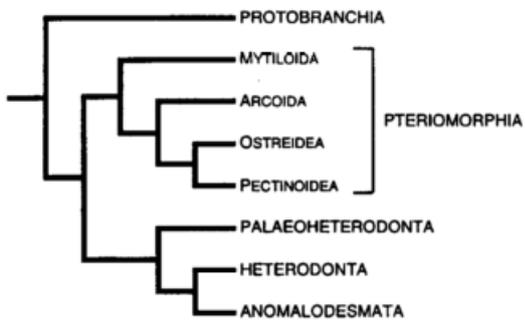
Protobranchia sind ursprüngliche, marine Muscheln mit Perlmutter (Solemyoidea, Nuculoidea) oder ohne (Nuculanoidea), deren Ursprung im Unterordovizium zu suchen ist. Ihr charakteristisches Merkmal sind die Ctenidien (Kiemen), die denen basaler Gastropoda ähneln (Reid, 1998). Morphologisch stehen sie basal als monophyletisches Adelphotaxon zu den restlichen Autobranchia. Molekulare Analysen stellen die Monophylie jedoch in Frage: die Nuculanoidea gehören demnach in die Pteriomorphia oder in deren Nähe (Giribet & Wheeler, 2002; Giribet & Distel, 2003; Plazzi & Passamonti, 2010), die Solemyoidea und Nuculoidea bilden die Protobranchia *sensu stricto*.



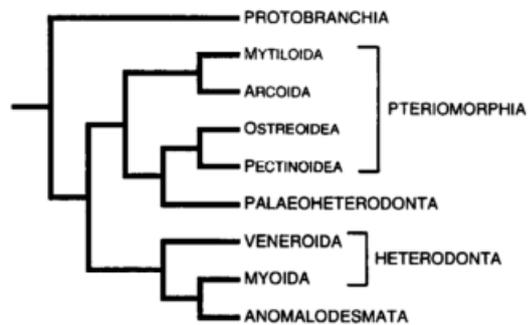
A STAROBOGATOV (1992)



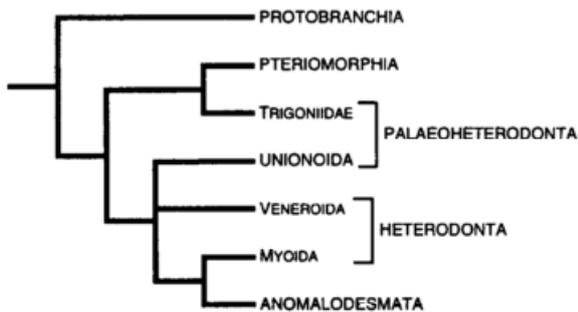
B COPE (1996)



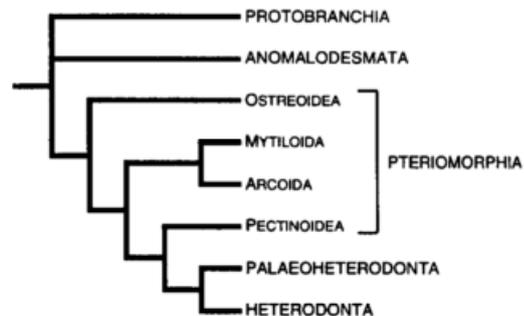
C WALLER (1997)



D MORTON (1996)



E SALVINI-PLAWEN & STEINER (1996)



F ADAMKEWICZ ET AL. (1997)

Abbildung 6: Ausgewählte Beispiele von hypothetischen Phylogenen der Bivalvia. Fünf Morphologie-basierte Stammbäume (A - E) und ein Stammbaum basierend auf der Analyse von 18S rDNA (F - Adamkewicz et al., 1997) (aus Adamkewicz et al., 1997).

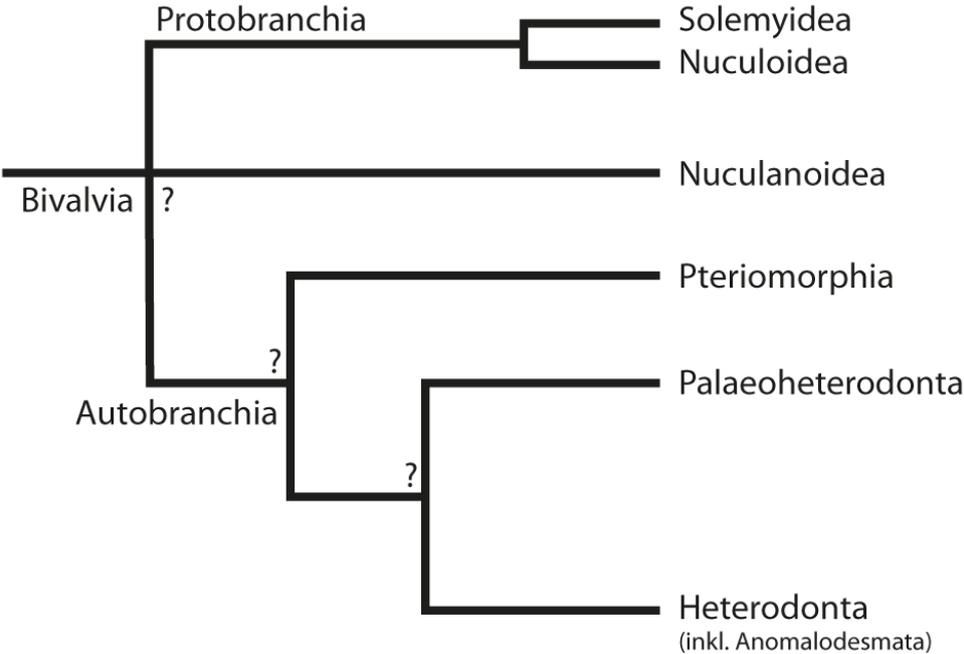


Abbildung 7: Hypothetische Phylogenie der Bivalvia, angepasst nach verschiedenen Autoren.

Tabelle 1: Auswahl von Publikationen mit molekularen Untersuchungen zur Phylogenie der Bivalvia.

Publikation	Anzahl untersuchter Arten		Focus der Untersuchung						Untersuchte Gene								
	Bivalvia	Outgroup	Bivalvia	Protobranchia	Pteriomorphia	Palaeoheterodonta	Heterodonta	Anomalodesmata	18S rDNA	28S rDNA	12S rDNA	16S rDNA	cox1	cytb	COI	Histone H3	Myosin
<i>Adamkewicz et al. (1997)</i>	28	5	x														
<i>Steiner et al. (2000)</i>	33	3	x	x													
<i>Campbell (2000)</i>	72	9	x														
<i>Canapa et al. (2001)</i>	29						x										
<i>Giribet & Wheeler (2002)</i>	62	14	x											x			
<i>Matsumoto (2003)</i>	75	11			x											x	
<i>Giribet & Distel (2003)</i>	133	9	x											x			
<i>Dreyer et al. (2003)</i>	26								x								
<i>Graf & Cummings (2006)</i>	50							x									
<i>Smith & Snyder (2007)</i>	1				x												
<i>Taylor et al. (2007)</i>	103									x							
<i>Puslednik & Serb (2008)</i>	46				x												
<i>Hoeh et al. (2009)</i>	50																
<i>Ren et al. (2010)</i>	2				x												
<i>Doucet-Beaupré et al. (2010)</i>	29	5			x												
<i>Plazzi & Passamonti (2010)</i>	35	4															
<i>Tsubaki et al. (2011)</i>	49	5															
<i>Plazzi et al. (2011)</i>	122	5															

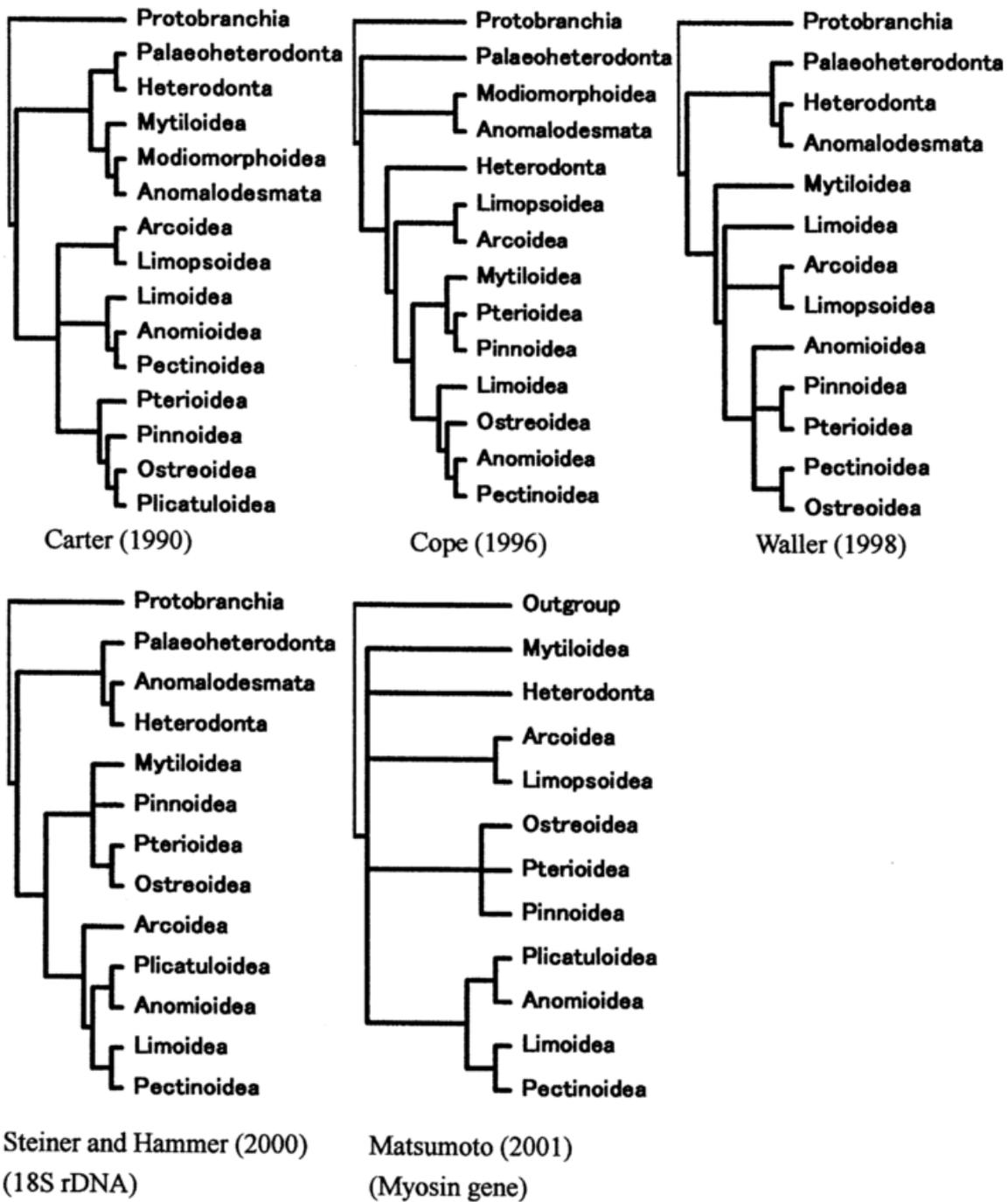


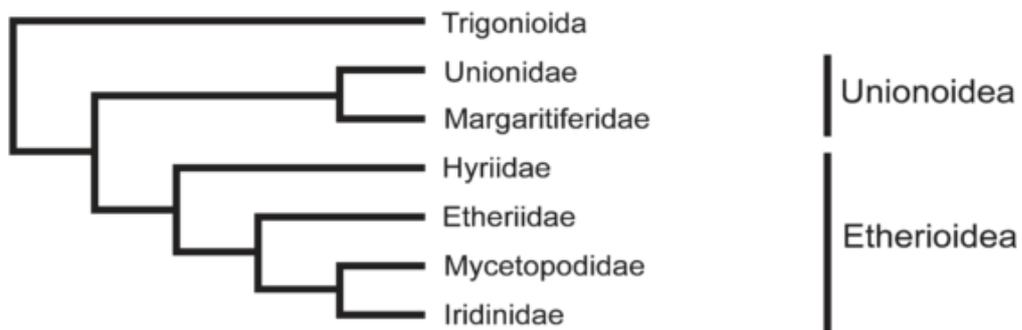
Abbildung 8: Vergleich von fünf hypothetischen Stammbäumen der Pteriomorpha (aus Matsumoto, 2003). Oben: basierend auf morphologischen Merkmalen – Unten: basierend auf molekularen Untersuchungen.

2.6.3 Pteriomorphia

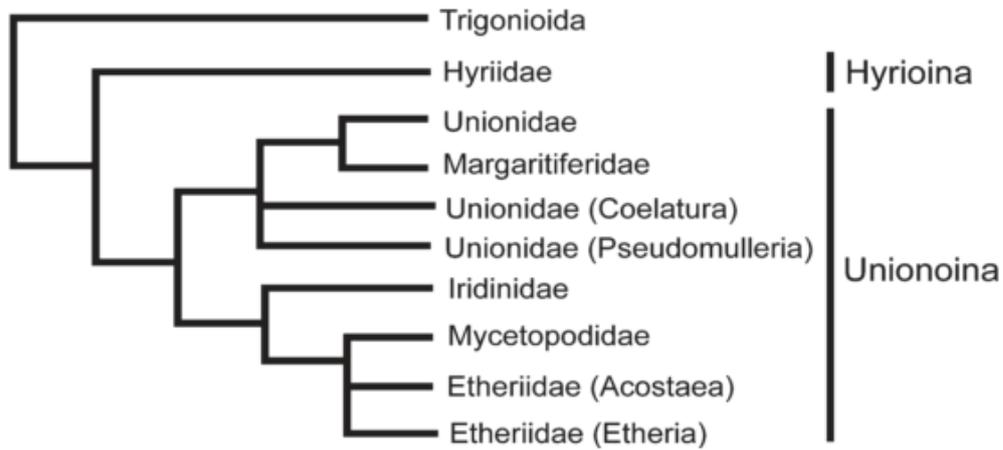
Pteriomorphia sind marine, meist sesshafte Tiere. Bekannte Vertreter sind Jacobsmuscheln, Miesmuscheln oder Austern. Sie zeichnen sich durch eine Asymmetrie der Schließmuskeln und der Schale aus. Bis auf die aufgeworfene Frage nach den Nuculanoidea, weisen molekulare Untersuchungen auf eine Monophylie der Pteriomorphia hin (*Giribet & Distel, 2003*). Innerhalb der Pteriomorphia ist die Analyse weit fortgeschritten, aber von einem Konsens noch weit entfernt. Während Arcoidea und Mytiloidea gut als monophyletische Zweige gesichert sind, werden die Positionen der Ostreoida, Pterioidea, Limoidea und Pectinoidea, sowie die Zuordnung ihrer Untergruppen diskutiert (Abbildung 8). Die Überfamilien (Arcoidea, Limopsoidea, Mytiloidea, Limoidea, Pectinoidea, Pinnoidea, Pterioidea, Anomioidea und Ostreoida) selbst sind dagegen weitgehend gefestigt (*Carter, 1990; Cope, 1996; Waller, 1998; Steiner & Hammer, 2000; Matsumoto, 2003*). Lediglich die Mytiloidea zeichnen sich in den verschiedenen Abstammungsbäumen durch Sprünge in die Nähe der Heterodonta und Anomalodesmata aus. Die Position der Pteriomorphia im Stammbaum der Bivalvia ist noch nicht geklärt. Möglich wäre zum Beispiel die Einordnung als Schwestergruppe zu den Heterodonta (*Cope, 1997; Plazzi & Passamonti, 2010*) oder als Schwestergruppe zu Heterodonta + Palaeoheterodonta (*Waller, 1990, 1998*). In der Phylogenie der Autobranchia wechseln die Palaeoheterodonta demnach von basal zu stark abgeleitet.

2.6.4 Palaeoheterodonta

Palaeoheterodonta setzen sich aus den marinen Trigonioidea und den Süßwassergroßmuscheln der Unionoidea zusammen (Abbildung 9). Trigonioidea waren im Mesozoikum weit verbreitet; die einzige rezente Gattung ist *Neotrigonia* mit sechs Arten. Die Palaeoheterodonta (früher Schizodonta + Modiomorpha) scheinen monophyletisch zu sein (*Giribet & Wheeler, 2002*) und stehen entweder basal zu den anderen Autobranchia oder als Schwestergruppe zu den Heterodonta (*Waller, 1990, 1998*). Die Phylogenie innerhalb der Unionoidea ist noch weitgehend unsicher und bedarf weiterer Untersuchungen (*Hoeh et al., 2009*).



GRAF & CUMMINGS 2006



BOGAN & HOEH 2000 + HOEH ET AL. 2001

Abbildung 9: Hypothetische Stammbäume der Palaeoheterodonta (aus Hoeh et al., 2009)

2.6.5 Heterodonta

Die Heterodonta sind die artenreichste Ordnung der Muscheln, die auch mehrfach das Süßwasser erobert haben. Die hoch entwickelten Vertreter der Heterodonta sind in Anatomie und Ökologie sehr variabel und schwierig gemeinsam zu definieren. Zu ihren bekannteren Vertretern gehören unter anderem die Riesenmuscheln, Herzmuscheln, Venusmuscheln, Dreikantmuscheln, Erbsenmuscheln oder die „Schiffsbohrwürmer“. Das Taxon Heterodonta wurde früher als paraphyletisch angesehen. Molekulare Untersuchungen legen jedoch eine Monophylie nahe (*Plazzi & Passamonti, 2010*) und inkludieren auch die Anomalodesmata in dieser Großgruppe (Abbildung 10). Morphologisch gibt es Argumente für und gegen eine Einordnung der Anomalodesmata in die Heterodonta.

Die auf Anatomie und Ökologie basierende Einordnung der Untergruppen „Veneroida“ und „Myoida“ werden nun als polyphyletische Taxa angesehen und gelten als weitgehend überholt. Sie werden auf der Basis der molekularen Analysen wahrscheinlich besser aufgelöst werden. Besonders in der internen Phylogenie der Heterodonta besteht weitergehender Forschungsbedarf.

2.6.6 Anomalodesmata

Anomalodesmata bestehen aus den teilweise ausgestorbenen Pholadomyoida, und den rezenten Septibranchia. Die Gruppe enthält seltsame und stark spezialisierte Arten, die im Salz- oder Brackwasser leben. Sie sind kleine, grabende oder eingebettete Arten, deren Kiemen dementsprechend angepasst oder reduziert sind. Die Anomalodesmata sind wahrscheinlich das am meisten „umhergeschobene“ Taxon innerhalb der Bivalvia. In vielen Arbeiten wurden sie als Taxon den Pteriomorphia oder Heterodonta gleichgestellt (*Thiele, 1935; Waller, 1998; Newell, 1965; Cope, 1996*). Die letzten Arbeiten weisen sie jedoch als relativ basalen Zweig innerhalb der Heterodonta aus (u.a. *Giribet & Wheeler, 2002; Bieler & Mikkelsen, 2006; Harper et al., 2006*). Aufgrund ihrer starken Spezialisierung ist auch in diesem Taxon die phylogenetische Klassifizierung schwierig (Abbildung 11).

Einleitung

A. Bieler & Mikkelsen 2006	B. Giribet & Distel 2003; Giribet, in press	C. Taylor et al. 2007
Heterodonta	Heterodonta	Heterodonta
Carditoida	Archiheterodonta	Archiheterodonta
Crassatelloidea	Carditoida	Crassatelloidea
Anomalodesmata	Crassatelloidea	Euheterodonta
13 families	Astartoidea	Thyasioidea
Veneroidea	Euheterodonta	Anomalodesmata (undivided at superfamily level see Harper <i>et al.</i> 2006)
Lucinoidea	Anomalodesmata	Lucinoidea (without Ungulinidae and Thyasiridae)
Chamoidea	Arcticoidea	Solenoidea
Galeommatoidea	Cardioidea	Hiatelloidea
Hiatellea	Chamoidea	Galeommatoidea (with ?Cyamioidea)
Hiatelloidea	Cyamioidea	Gastrochaenoidea
Gastrochaenoidea	Dreissenoida	Cardioidea (without Hemidonacidae)
Arcticoidea	Galeommatoidea	Tellinoidea
Glossoidea	Glossoidea	Neoheterodonte
Cyamioidea	Hiatelloidea	Sphaeroidea (without Corbiculidae)
Sphaeroidea	Lucinoidea	Myoidea (Myidae and Corbulidae)
Cardioidea	Mactroidea	Pholadoidea
Veneroidea	Myoidea	Dreissenoida
Tellinoidea	Solenoidea	Gaimardioidea
Solenoidea	Tellinoidea	Mactroidea
Mactroidea	Veneroidea	Ungulinoidea (elevated to superfamily status)
Dreissenoida		Corbiculoidea (without Sphaeridae)
Myoidea		Chamoidea
Myoidea		Veneroidea (Veneridae plus Petricolidae)
Pholadoidea		<i>Superfamilial status unresolved</i>
		Glossidae
		Hemidonacidae
		Glauconomidae
		Trapezidae
		Arctidae
		Vesicomidae
		Kelliellidae

Abbildung 10: Verschiedene Klassifizierungen der Heterodonta nach molekularen Untersuchungen (aus Taylor *et al.*, 2007)

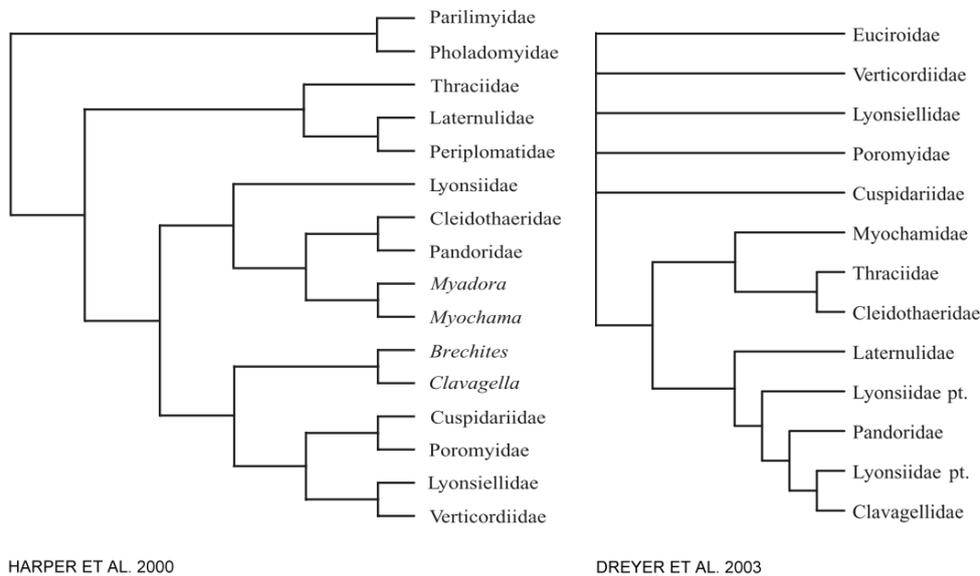


Abbildung 11: Hypothetische Phylogenien der Anomalodesmata. Links nach morphologischen Merkmalen, rechts nach 18S rDNA (aus Dreyer *et al.*, 2003).

2.7 Zielsetzung

2.7.1 Methodische Bewertung der Molekularen Morphologie

In dieser Arbeit sollen sowohl Potenzial als auch Restriktionen, also Vor- und Nachteile der Molekularen Morphologie gegenüber anderen Verfahren untersucht und anschließend die Methode an einem konkreten Beispiel, der Klasse der Bivalvia (Muscheln) getestet werden. Dementsprechend ist ein Ziel dieser Arbeit, durch die vergleichende Analyse der Sekundärstrukturen von sehr langsam evolvierenden Molekülen Rückschlüsse auf sehr alte stammesgeschichtliche Verzweigungen (Kambrium - Ordovicium) zu ermöglichen. Der Schwerpunkt liegt dabei auf den Sequenzen der 18S Fraktion der ribosomalen Gene, die bereits in großer Anzahl sequenziert in Online-Datenbanken vorhanden sind. Primär wird die 18S rRNA untersucht, da sowohl 16S als auch 28S höhere Mutationsraten aufweisen und folglich weniger aussagekräftig im Sinne der „Deep Phylogeny“ sind.

Seit 2003 sind mehrere Algorithmen zur Generierung von Sekundärstrukturen in Software-Paketen verfügbar. Sie erlauben, eine größere Anzahl von Sequenzen auf einfache Weise zu analysieren. Folgende Programme werden hier auf ihre Nutzbarkeit für Molekulare Morphologie getestet:

- ⇒ RNAfold (2003)
- ⇒ RNAsalsa (2007)
- ⇒ Mfold (2003)
- ⇒ RNAstructure (2006)
- ⇒ RNAViz (1997-2003)

Mit der jeweiligen Software werden die Sequenzen u.a. mit verschiedenen einstellbaren Parametern gefaltet und die entstehenden 2D-Strukturen untersucht. Dabei soll festgestellt werden, welcher Algorithmus mit welchen Parametern für die weitere Analyse geeignete Faltungen generiert. Die Unterschiede zwischen diesen Verfahren werden im Abschnitt 3.1.2, „Software“, genauer erläutert.

Ebenso sollen die untersuchten Verfahren in ihrem Aufwand und Ergebnis verglichen werden:

- ⇒ Wie erfolgt die Nutzung der Software (Installationsprogramm, Webanwendung)?
- ⇒ Welche Begleit- und Sekundärsoftware (Betriebssystem etc.) ist ggf. notwendig?
- ⇒ Wie groß ist der Zeitaufwand für eine gefaltete Sequenz?
- ⇒ Wie geeignet sind die 2D-Strukturen für die weitere Analyse?
- ⇒ Können die 2D-Strukturen automatisch auf Ähnlichkeiten untersucht werden?
- ⇒ Wie ähnlich sind die 2D-Strukturen desselben Moleküls bei anderer Faltungssoftware?

Im zweiten Teil dieser Arbeit wird die Methode an den Bivalvia getestet.

2.7.2 Phylogenie der Bivalvia mittels Molekularer Morphologie

Schwerpunkt im zweiten Teil dieser Arbeit soll die Untersuchung der phylogenetischen Entwicklung innerhalb der Bivalvia sein. Zum einen sollen die Ergebnisse der Molekularen Morphologie mit den bekannten phylogenetischen Stammbäumen innerhalb der Muscheln, zum Beispiel den Pteriomorphia, verglichen werden. Der Vergleich soll zeigen, ob die durch die neue Methode gewonnenen Resultate plausibel sind. Zum anderen sollen mit der getesteten Methode stammesgeschichtlich unklare Verwandtschaften innerhalb der Bivalvia (z.B. Heterodonta oder Anomalodesmata) untersucht werden.

Im Idealfall soll diese Arbeit zur Empfehlung einer Methode führen, die mit wenigen zusätzlichen Anpassungen weitgehend universell auf viele Organismengruppen anwendbar ist.

Die ersten (teilweise) Schalen tragenden Arten der Mollusca sind im Fossil-Record bereits für das frühe Kambrium (*Butterfield, 2008; Caron et al., 2006*) und möglicherweise sogar Präkambrium (*Fedonkin et al., 2007*) nachgewiesen. Während der „Kambrischen Explosion“ (vor 542 Mio. Jahren) erfolgte neben der Entstehung von vielen anderen Groß-Taxa auch die Aufspaltung der Bivalvia in die wichtigsten

Ordnungen. Diese sogenannten „Deep Nodes“ sollen mit Hilfe der Molekularen Morphologie untersucht werden.

Die Bivalvia bieten ein vielfältiges Spektrum aus:

- ⇒ gut untersuchten und phylogenetisch fest eingeordneten Taxa (Pteriomorphia)
- ⇒ gut untersuchten aber schwierig klassifizierbaren Taxa (Heterodonta, Veneroidea)
- ⇒ stark abgewandelten Taxa (Anomalodesmata, Teredinidae)
- ⇒ rezenten, aber phylogenetisch sehr alten Taxa (*Spondylus* 285 Mio., *Lima* 320 Mio. Jahre alt)

In Abschnitt 2.6, „Bivalvia“, unter „Stand des Wissens zur Phylogenie“ wurden die bisherigen Erkenntnisse zur Phylogenie der Muscheln diskutiert. Generell stellt sich gerade in dieser Gruppe das Problem von multiplen morphologischen Parallelentwicklungen. Davon betroffen sind insbesondere die höheren Bivalvia (Heterodonta) und nahezu alle Organsysteme.

So sind die Pteriomorphia (u.a. *Mytilus*, *Pecten*, *Ostrea*) gut untersucht und ihre Verwandtschaft weitgehend sicher. Das Taxon soll als Ausgangspunkt zur Untersuchung verschiedener Methoden der Molekularen Morphologie dienen. Die Verwendbarkeit dieser Methoden als Mittel zur phylogenetischen Analyse steht im Vordergrund. Erstes Ziel mit der jeweiligen Methode ist die Überprüfung der Konsistenz mit bekannten Verwandtschaften innerhalb der Pteriomorphia. Auf diese Weise soll die zuverlässigste Methode gefunden werden. Mit der so entwickelten Routine soll die Analyse auf andere Taxa der Bivalvia ausgeweitet werden.

Als Nächstes soll die basale Gruppe der Protobranchia mit der Molekularen Morphologie untersucht werden. Die namensgebenden protobranchen Kiemen deuten auf eine Monophylie der Gruppe hin, molekulare Analysen stellen das aber in Frage. Das Verwandtschaftsverhältnis der zwei, eventuell drei Untergruppen Nuculoidea, Solemyoidea und Nuculanoidea ist daher ungeklärt. Diese Untersuchung setzt eine hinreichend große Anzahl von sequenzierten Arten voraus.

Die höher evolvierten Heterodonta bilden einen weitaus artenreicheren, sehr umstrittenen Zweig der Bivalvia, deren klassische interne Systematik basierend auf Anatomie und Ökologie weitgehend als überholt gilt. Nach allen neueren Sequenzanalysen sind auch die „Anomalodesmata“ innerhalb der Heterodonta zu stellen (*Giribet, 2008*). Hier soll die Molekulare Morphologie neue Daten und damit einen Beitrag zur Aufklärung der Phylogenie liefern.

Falls sich die hier analysierte Methode bewährt, kann darauf aufbauend die Analyse auf die Taxa der Mollusca und weiter auf die Lophotrochozoa ausgeweitet werden.

3. Material & Methoden

3.1 Material

Im Zuge dieser Arbeit soll die Molekulare Morphologie als kostengünstige Methode zur Untersuchung der phylogenetischen Verwandtschaft von Taxa vorgestellt werden. Kostengünstig ist die Methode in dem Sinne, dass sie als solche keinerlei Feld- oder Laborarbeit benötigt. Zu ihrer Anwendung reicht es aus, nur mit geeigneter Software und den bereits vorhandenen Daten einer vorausgegangenen einmaligen experimentellen Sequenzierung zu arbeiten. Die Vorbereitung im Rahmen der vorliegenden Arbeit umfasste eine Recherche der verfügbaren Software sowie die Sammlung und kritische Sichtung der zu nutzenden vorhandenen Sequenzen.

Als Arbeitsmittel standen mehrere PCs mit den Betriebssystemen Windows XP, Windows 7 und Linux Ubuntu (Distribution Version 7.04 „Feisty Fawn“) zur Verfügung. Diese Betriebssysteme waren zur Nutzung der verschiedenen Programme ausreichend. Die Suche und Sammlung von Software und Sequenzen erforderte einen Internetzugang, der keine speziellen Anforderungen an Bandbreite oder Geschwindigkeit stellte. Alle hier verwendeten Materialien waren im Internet frei verfügbar.

3.1.1 Sequenzen

Abhängig davon, welches Taxon phylogenetisch untersucht werden soll, muss das zu verwendende Gen, entsprechend seiner geschätzten Mutationsgeschwindigkeit, geeignet ausgewählt werden. Je länger zwei ursprünglich identische Taxa getrennte Evolutionswege durchlaufen, desto größer werden die Unterschiede in ihren Nukleotidsequenzen. Diese Annahme steht im Einklang mit dem Zweiten Hauptsatz der Thermodynamik und bildet die Grundlage für das Konzept der Molekularen Uhr (Zuckerlandl & Pauling, 1965). Sie bezieht sich auf die Mutationsraten von Genen und/oder Proteinen, wobei angenommen wird, dass deren Taktfrequenz konstant ist

und z.B. durch Fossilienfunde annäherungsweise geeicht werden kann (mehr Informationen dazu sind im Glossar, Abschnitt „Molekulare Uhr“, beschrieben).

Für Aufspaltungsereignisse eines Taxons im frühen Paläozoikum (Kambrium bis Silur) eignet sich die sehr langsam evolvierende 18S rDNA als Untersuchungsmaterial, für Entwicklungen seit dem Mesozoikum die schneller veränderliche 16S rDNA (*Campbell, 2000; Lydeard et al., 2002*). In der vorliegenden Arbeit soll die stammesgeschichtliche Entwicklung der Bivalvia seit dem Kambrium untersucht werden. Die primäre Aufspaltung der Bivalvia erfolgte, paläontologischen Daten zufolge, bereits vom frühen Kambrium („Kambrische Explosion“) bis in das frühe Ordovizium hinein, vor ca. 550-470 Mio. Jahren (*Cope, 1997; Fedonkin, 2007; Butterfield, 2008; Plazzi et al. 2011*). In dieser Zeit entstanden alle großen Muscheltaxa (Protobranchia, Pteriomorpha, Heterodonta), die bis heute Bestand haben. Die 18S rDNA ist ausreichend hoch konserviert; geschätzt ergibt sich eine 10%-Divergenz für einen Zeitraum von 500 Mio. Jahren (*Brown et al., 1979*). Detaillierte Informationen darüber, welche Arten und Taxa in die Untersuchung einbezogen wurden, finden sich in Abschnitt 3.1, „Material & Methoden“, unter „Bivalvia“.

Voraussetzungen für die Verwendung eines bestimmten Gens für eine molekular-morphologische Untersuchung sind:

- ⇒ das Vorhandensein lückenloser Primärsequenzen des Gens für alle zu untersuchenden Arten
- ⇒ für die „Near-Natural“-Methode (NNC): das Vorhandensein eines „Constraint“ für eine oder mehrere der zu untersuchenden Arten, d.h., gewisse 2D- oder 3D-Strukturinformationen müssen bereits bekannt sein
- ⇒ eine begrenzte Sequenzlänge, um Faltungsfehler und Programmlaufzeiten einzuschränken. Mfold begrenzt die Sequenzlänge auf 9000 Basen; das Limit von RNAfold liegt bei 10 000 Basen für Faltungen nach Minimum Free Energy (MFE)

(Stand März, 2011)

Die verwendeten Sequenzen stammen aus der GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>), zusätzlich wurden einige Sequenzen aus der

European Ribosomal Database (inzwischen ersetzt durch SILVA rRNA database, <http://www.arb-silva.de/>) genutzt. Insgesamt standen zu Beginn dieser Arbeit 136 Sequenzen 18S rDNA von Bivalvia zur Verfügung, bei Abschluss waren es 159 Sequenzen. Ihre Länge liegt im Schnitt bei ca. 1800 Nukleotiden, sodass keine Limitierung bei der automatischen Faltung wirksam wurde.

3.1.1.1 Bivalvia

Für diese Arbeit wurden publizierte Primärstrukturen von bereits sequenzierter 18S rDNA genutzt. Die meisten der verwendeten Sequenzen stammen aus der GenBank, zusätzlich wurden einige Sequenzen aus der European Ribosomal Database genutzt.

Zum Zeitpunkt dieser Arbeit waren 160 nutzbare 18S rDNA-Sequenzen der Bivalvia für die Analyse verfügbar und wurden in den zu untersuchenden Datensatz aufgenommen. Drei der Sequenzen wurden von vornherein von der Analyse ausgeschlossen: *Cuspidaria cuspidata*, *Varicorbula dissimilis* und *Hiatella arctica*. Die sehr kurzen Sequenzen, die zu diesen Arten veröffentlicht wurden, lassen auf Teilstücke oder Fehler (bei Identifikation als 18S / Veröffentlichung als 18S) schließen. Der untere Grenzwert für die Anzahl von Nukleotiden wurde bei 1300 angesetzt. Die Sequenzen haben meist etwa 1800 (± 100) Nucleotide, wobei das Minimum bei 1370 (Heterodonta > Lucinidae > *Codakia orbiculata*) und das Maximum bei 2143 (Anomalodesmata > Pandoridae > *Pandora arenosa*) lag. Die berechneten Sekundärstrukturen für Sequenzen mit einer Länge unter 1700 (6 Sequenzen), sowie die mit einer Länge über 1900 (4 Sequenzen) sollten in dieser Hinsicht als vermutlich unvollständig oder aberrant angesehen werden.

Die in dieser Arbeit genutzten Sequenzen wurden in den folgenden Arbeiten publiziert: *Giribet & Wheeler (2002)*; *Giribet et al. (2000)*; *Steiner & Hammer (2000)*; *Neulinger et al. (2006)*; *Campbell (2000)*; *Kenchington et al. (1995)*; *Page (2001)*; *Lindgren et al. (2004)*; *Steiner & Mueller (1996)*; *Winnepenninckx et al. (1996)*; *Page & Linse (2002)*; *Itoh et al. (2001)* (direct submission); *Frischer et al. (1998)*; *Dufour et al. (2004)* (direct submission); *Canapa et al. (1999)* (direct submission); *Kenchington et al. (1993)*; *Rice (1990)*; *Masaoka & Kobayashi (2005)* (direct submission); *Zhang et al. (2001)* (direct

submission); Huff et al. (2004) (direct submission); Canapa et al. (1999); Rice et al. (1993); Maruyama et al. (1998); Distel et al. (2000); Distel (2000); Distel et al. (2011); Littlewood et al. (1991); Frischer et al. (1999); Bell & Grassle (1998).

Die hier untersuchten Arten sind durch die aktuelle Datenlage vorgegeben. Die Verteilung der Sequenzen auf die einzelnen Taxa ist in Tabelle 2 aufgelistet. Die Auswahl an Sequenzen ist weit gefächert und deckt die wichtigen Taxa der Bivalvia gut ab. Es wurde im Rahmen dieser Arbeit keine zusätzliche 18S rDNA sequenziert.

Für die Außengruppe wurden die folgenden Arten aus den Mollusca ausgewählt:

- ⇒ Cephalopoda > Nautilidae > *Nautilus macromphalus*
- ⇒ Cephalopoda > Octopodidae > *Bathypolypus arcticus*
- ⇒ Polyplacophora > Chitonidae > *Liolophura japonica*
- ⇒ Polyplacophora > Leptochitonidae > *Lepidopleurus cajetanus*
- ⇒ Polyplacophora > Leptochitonidae > *Lepidochitona corrugata*
- ⇒ Scaphopoda > Rhabdidae > *Rhabdus rectius*
- ⇒ Scaphopoda > Dentaliidae > *Antalis vulgaris*
- ⇒ Gastropoda > Helicidae > *Cepaea nemoralis*
- ⇒ Gastropoda > Pleurotomariidae > *Entemnotrochus adansonianus*

Tabelle 2: Anzahl der Sequenzen pro Großgruppe

		Familie	Anzahl Sequenzen		
Protobranchia	Nuculoidea	Nuculidae	3		
		Yoldiidae	3		
	Nuculanoida	Nuculanidae	3		
		Neilonellidae	1		
Solemyoidea	Solemyidae	4			
Autobranchia	Palaeoheterodonta	Trigonioida	Neotrigoniidae	2	
		Unionoida	Unionidae	4	
	Margaritiferidae		6		
	Pteriomorpha	Mytiloidea	Mytilidae	27	
			Arcoidea	Limopsidae	2
		Arcidae		6	
		Limoida	Limidae	4	
		Ostreoida	Ostreidae	5	
			Gryphaeidae	2	
		Pectinoidea	Plicatulidae	2	
			Spondylidae	3	
			Propeamussidae	2	
			Pectinidae	13	
			Anomiidae	2	
		Pterioidea	Pteriidae	12	
			Isognomonidae	2	
			Pinnidae	2	
			Malleidae	2	
		Heterodonta	Anomalodesmata	Poromyidae	1
				Cuspidariidae	2
				Pandoridae	1
	Lyonsiidae			1	
	"Myoidea"		Teredinidae	1	
			Pholadidae	1	
			Myidae	1	
			Corbulidae	1	
			Hiatellidae	1	
	"Veneroidea"		Veneridae	3	
			Semelidae	1	
			Donacidae	1	
			Pharidae	2	
			Mactridae	7	
			Ungulinidae	1	
			Lucinidae	3	
			Vesicomysidae	1	
			Gastrochaenidae	2	
Lasaeidae		1			
Galeommatidae		2			
Dreissenidae		1			
Basterotiidae		1			
Sphaeridae	1				
Corbiculidae	1				
Chamidae	1				
Carditidae	3				
Tridacnidae	2				
Cardiidae	2				
Arctidae	1				
Astartidae	1				

3.1.1.2 Beispiel: Sequenz der 18S rDNA von *Pecten maximus*

Die Sequenz der 18S rDNA von der Pilgermuschel (*Pecten maximus*) wurde von der GenBank unter der folgenden URL abgerufen:

<http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=1479992>

(Stand Oktober, 2011)

Außer der eigentlichen Sequenz sind eine Reihe weiterer Informationen verfügbar, unter anderem die Publikation, in der die Sequenz veröffentlicht wurde. Diese Daten werden hier als ein typisches Beispiel wiedergegeben:

LOCUS	PCTRRE	1815 bp	rRNA	linear	INV	08-NOV-1999
DEFINITION	Pecten maximus 18S ribosomal RNA, complete sequence.					
ACCESSION	L49053					
VERSION	L49053.1 GI:1479992					
KEYWORDS	18S ribosomal RNA; ribosomal RNA.					
SOURCE	Pecten maximus					
ORGANISM	Pecten maximus Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Pectinoida; Pectinoidea; Pectinidae; Pecten.					
REFERENCE	1 (bases 1 to 1815)					
AUTHORS	Frischer, M., Williams, J. and Kenchington, E.					
TITLE	A molecular phylogeny of some major groups of Pectinidae inferred from 18S rRNA gene sequences					
JOURNAL	(in) Johnston, P.A. and Haggart, J.W. (Eds.); BIVALVES: AN EON OF EVOLUTION - PALEOBIOLOGICAL STUDIES HONORING NORMAN D. NEWELL; University of Calgary Press, Calgary (1998)					
FEATURES	Location/Qualifiers					
source	1..1815 /organism="Pecten maximus" /mol_type="rRNA" /db_xref="taxon:6579" /haplotype="2n" /tissue_type="adductor muscle"					
rRNA	1..1815 /product="18S ribosomal RNA"					
ORIGIN	1 caacctggtt gatcctgcc gtagtcatat gcttgtctca aagattaagc catgcatgtc 61 taagtacata ctttttgatg gtgaaaccgc gaatggctca ttaaatcagt tatggttcct 121 tagatcgtag gatcctactt ggataactgt ggcaattcta gagctaatac atgcaaaaag 181 gcaccgactc acggaggtgc gattttatca gttcaaaacc aatcggtcgc aaggccgtca 241 ctttggtgaa tctggataac tttgtgctga tcgcacggcc ctctgcccgg cgacgtatct 301 ttcaaatgtc tgacctatca actttcgatg gtacgtgcta tgctaccat ggttghtaacg 361 ggtaacggag aatcagggtt cggttccgga gaggagcat gagaacggc taccacatcc 421 aaggaaggca gcaggcgcgc aaattaccca ctctggcac ggggaggtag tgacgaaaaa 481 taacaatacg ggactctttc gaggccccgt aattggaatg agtacacttt aaacccttta 541 acgagtatcc attggagggc aagtctggtg ccagcagccg cggtaattcc agctccaata 601 gcgtatatta aagttggtgc agttaaagag ctctgagttg gatctcagat ccaggcttgc 661 ggtccactcc ttggtggcta ctgctcgctc tgatctacct cccggtttta cccttgggtgc 721 tcttgattga gtgtctcggg tggccggaac atttactttg aaaaaattag agtgttcaaa 781 gcaggcagct cgcctgaata atggtgcatg gaataatgga ataggacctc ggttctatct 841 tgttggtttt cggaacttga ggtaatgatt aagagggaca gacgggggca ttcgtattac 901 ggtgtagtag gtgaaattct tggatcgccg taagacgaac tactgcgaaa gcatttgcca 961 agaattgttt cattaatcaa gatcgaagt cagaggttcg aagacgatca gataccgtcg					

```

1021 tagttctgac cataaacgat gccaaactagc gatccgcgg agttgcttca atgactcggc
1081 gggcagctta cgggaaacca aagtttttgg gttccggggg aagtatgggt gcaaagctga
1141 aacttaaagg aattgacgga agggcaccac caggagtgga gcctgcggct taatttgact
1201 caacacggga aaactcacc gccccggaca ctgtaaggat tgacagattg agagctcttt
1261 cttgattcgg tgggtggtgg tgcattggcg ttcttagttg gtggagcgat ttgtctgggt
1321 aattccgata acgaacgaga ctctagcctg ctaaatagtt caccgatccg ctctgcgctg
1381 gtgcaacttc ttagagggac aagtggcgtt tagccacacg agattgagca ataacaggtc
1441 tgtgatgccc ttagatgttc gggccgcac gcgcgtaca ctgaaggaat caacgtgctc
1501 ttatccttgc ccggaagggt tgggtaacc gttgaacctc ctctgtgcta gggattgggg
1561 cttgtaattc ttccccatga acgaggaatt ccagtaagc gcgagtcata agctcgcgtt
1621 gattacgtcc ctgccctttg tacacaccgc ccgtcgctgc taccgattga atggtttagt
1681 gagctcctcg gattggtccc gacacggggg gaaactctcg agtcggtgtg ccgaaaagac
1741 gagcaaactt gattatntag aggaagtaaa agtcgtaaca aggtttccgt aggtgaacct
1801 gcagaaggat ctagt

```

3.1.2 Software

Die dreidimensionale Struktur als Folge der jeweiligen Primärsequenz ist ausschlaggebend für die Funktion von realen kettenförmigen Biomolekülen wie Enzymen oder Nukleinsäuren. Während die wirkliche dreidimensionale Tertiärstruktur bisher kaum durch Algorithmen vorherzusagen ist, kann eine projektive, das heißt letztlich fiktive zweidimensionale Sekundärstruktur durch geeignete numerische Verfahren auf verschiedene Weisen annähernd dargestellt werden. Die grundlegenden Algorithmen wurden in Abschnitt 2.5, „Molekulare Morphologie“, genannt. Die Sekundärstrukturen der 18S rRNA wurden durch die Nutzung der folgenden Programme erstellt:

- ⇒ Mfold (*Zuker, 2003*)
- ⇒ RNAstructure (*Mathews, 2006*)
- ⇒ RNAfold (*Hofacker, 2003*)
- ⇒ RNAsalsa (*Stocsits, 2009*)

Es gibt weitere Programme zur Vorhersage von DNA-Sekundärstrukturen, die im Rahmen dieser Arbeit aber nicht oder nicht vollständig (zum Beispiel XRNA) untersucht wurden. Die grafische Darstellung der 2D-Struktur wird durch das folgende Programm ermöglicht bzw. verbessert:

- ⇒ RNA Viz (*Rijk & Wachter, 1997*)

Für das Alignment wurde ClustalX genutzt. Die Stammbäume wurden mit PAUP 4.0b 10 berechnet.

3.1.2.1 Mfold

Der mfold-Webserver wurde bereits 1995 durch die Washington University School of Medicine aufgesetzt. Seit 1996 können dort DNA-Faltungen von Sequenzen, die der Nutzer beliebig vorgeben kann, berechnet werden. Der aktuelle mfold-Server bietet die Möglichkeit, sowohl Einzelstrang-RNA, als auch -DNA nach dem Prinzip von MFE zu falten. Als Algorithmen stehen regelmäßig aktualisierte Versionen (derzeit Version 3.2) des Zuker-Algorithmus zur Verfügung.

Der Server ist über das WWW unter der URL

<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form>

(Stand Oktober, 2011)

verfügbar. Dort dient eine Weboberfläche als Zugriff auf die Software, sodass eine Nutzung auch ohne Installation auf einem lokalen PC möglich ist (Abbildung 12).

3.1.2.1.1 Webinterface

Die Eingabemaske erlaubt einen 40 Zeichen langen Namen für die einzugebende Sequenz. Die Sequenz selbst wird als Zeichenkette eingegeben und automatisch auf „RNA/DNA-Format“ getrimmt, d.h. Zeichen verschieden von A-Z, a-z werden entfernt. Für die DNA-Faltung werden U in T umgewandelt, für RNA-Sequenzen T in U. Nach mfold-Konvention werden bestimmte Buchstaben (außer A, C, G, T, U) als Platzhalter für bestimmte Eigenschaften genutzt. Genaue Angaben zu den Sequenzkonventionen finden sich bei *Zuker (2003)*. Des Weiteren können Constraint-Informationen im Formelformat eingegeben werden, um bestimmte Basenpaare zu erzwingen oder zu verhindern. Für die Steuerung der Faltung der Sequenzen stehen verschiedene frei wählbare Parameter zur Verfügung (Tabelle 3, Seite 57). Die Faltungstemperatur ist ab Version 2.4 der RNA-Parameter auf 37°C festgelegt. Die Ausgabe des Ergebnisses erfolgt als Bilddatei in den Formaten PNG und JPG. Zusätzlich zur einfachen Faltung gibt es die Möglichkeit, Informationen zu Wahrscheinlichkeit, Verlässlichkeit und Energie-Verteilung zu erhalten.



THE RNA INSTITUTE
COLLEGE OF ARTS AND SCIENCES
UNIVERSITY AT ALBANY State University of New York

The mfold Web Server

Home
DINAMelt Web Server
Mfold Web Server
Forum

Applications

- [RNA Folding Form](#)
- [DNA Folding Form](#)
- [Structure Display and Free Energy Determination](#)
- [RNA Folding Form \(version 2.3 energies\)](#)

View Folding Results

- [Folding Results](#)

Documentation

- [Mfold References](#)
- [FAQs](#)
- [Folding & output options](#)
- [Folding with constraints](#)

Software

- [Mfold](#)

About

- [About](#)

Contact

- [Contact](#)

RNA Folding Form

The folding temperature is fixed at 37°. You may still fold with the older version 2.3 RNA parameters, which allow the temperature to be varied.

[DNA mfold server](#), [Quikfold](#). Fold many short RNA or DNA sequences at once.

Enter a name for your sequence:

Enter the sequence to be folded in the box below. All non-alphabet characters will be removed. FASTA format may be used.

Enter [constraint information](#) in the box at the right. (optional) You may:

1. force bases $i, i+1, \dots, i+k-1$ to be double stranded by entering:
 $F _ i _ O _ k$ on 1 line in the constraint box.
2. force consecutive base pairs $i, j, i+1, j-1, \dots, i+k-1, j-k+1$ by entering:
 $F _ i _ j _ k$ on 1 line in the constraint box.
3. force bases $i, i+1, \dots, i+k-1$ to be single stranded by entering:
 $P _ i _ O _ k$ on 1 line in the constraint box.
4. prohibit the consecutive base pairs $i, j, i+1, j-1, \dots, i+k-1, j-k+1$ by entering:
 $P _ i _ j _ k$ on 1 line in the constraint box.
5. prohibit bases i to j from pairing with bases k to l by entering:
 $P _ i _ j _ k _ l$ on 1 line in the constraint box.

The RNA sequence is

Folding temperature is fixed at 37°.

Ionic conditions: 1M NaCl, no divalent ions.

Abbildung 12: Die Eingabemaske des mfold Web Servers (<http://mfold.rna.albany.edu/?q=mfold>)

3.1.2.1.2 Lokale Installation

Die mfold-Software für akademische Zwecke gibt es als kostenfreien Download unter

<http://mfold.rna.albany.edu/?q=mfold/download-mfold>

(Stand Oktober, 2011)

Die ausführbaren Dateien sind als Binaries für Linux/Unix-Systeme kompiliert (*Zuker et al., 1999*) und bestehen aus einzelnen Programmen in verschiedenen Programmiersprachen (C, Fortran, Perl). Die Unterprogramme können individuell oder gemeinsam über die Kommandozeile aufgerufen werden.

Einige Unterprogramme wurden für Windows-Systeme angepasst und sind unter der Bezeichnung RNAstructure verfügbar.

3.1.2.2 RNAstructure

RNAstructure ist der Name der Windows-Variante von mfold. Das Programm ist in C++ geschrieben und als Open Source-Bibliothek oder als Installationsdatei zum Download unter

<http://rna.urmc.rochester.edu/RNAstructure.html>

(Stand Oktober, 2011)

verfügbar. Ein mit JAVA programmiertes „Graphical User Interface“ (GUI) macht eine einfache Handhabung möglich. Sequenzen können direkt als Zeichenkette (nach bestimmten Regeln) eingegeben werden und als das mfold-eigene Format SEQ abgespeichert werden. Die Faltung erfolgt nach MFE (Turner) oder seit 2010 (also nach Abschluss der in dieser Arbeit beschriebenen Untersuchungen) nach MEA (Maximum Expected Pair Accuracy) (Abbildung 13). Die Faltung mit einem Constraint ist nicht möglich.

Eine detaillierte Beschreibung der Funktionen von RNAstructure findet sich unter

<http://rna.urmc.rochester.edu/GUI/html/Contents.html>

(Stand Oktober, 2011)

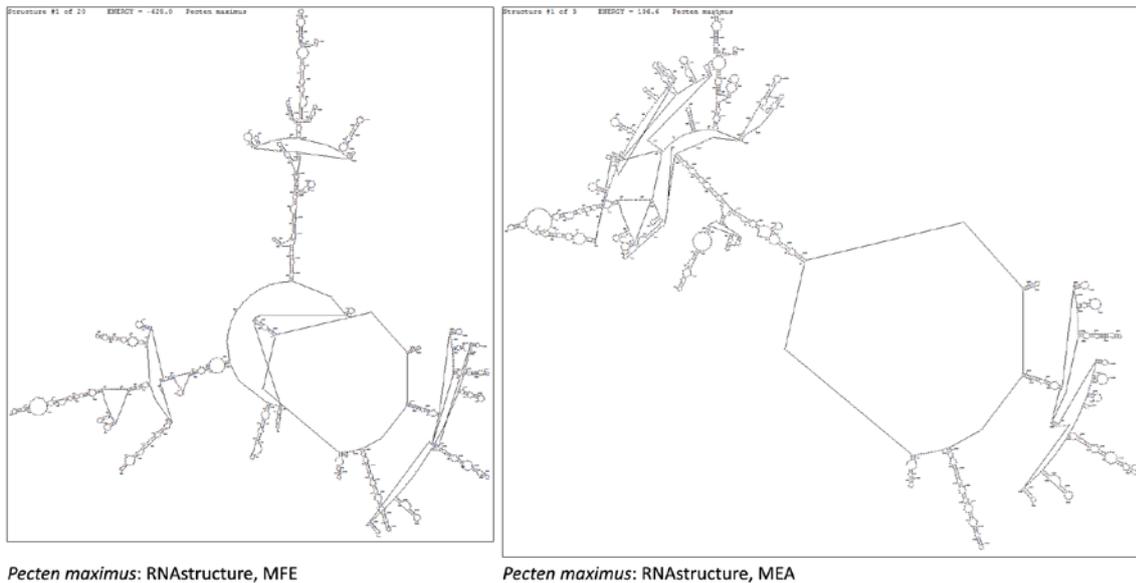


Abbildung 13: Sekundärstruktur von *Pecten maximus*, nach RNAstructure. Links nach Prinzip MFE, rechts nach Prinzip MEA.

3.1.2.3 RNAfold

Das Programm RNAfold ist Teil des Vienna RNA Software Package, das von Hofacker (2003) vorgestellt wurde. Des Weiteren basiert auch die Variante RNAsalsa auf der RNAfold-Software.

Es gibt zwei Möglichkeiten, RNAfold zu nutzen, nämlich das Software-Paket:

1. auf einem Linux-Rechner zu installieren oder
2. über den Vienna RNA secondary structure server zu nutzen.

Beide Optionen wurden in dieser Arbeit genutzt.

Die Variante (1) bietet den gesamten Funktionsumfang des Software-Pakets. Für die Nutzung wurde ein PC mit Ubuntu Linux aufgesetzt und die Software installiert. Das Paket ist nur für Unix-erfahrene Nutzer geeignet, da die meisten Funktionen nur über Kommandozeile ausgeführt werden können.

Das Webinterface (2) bietet nur die Teilmenge der am häufigsten genutzten Funktionen des Software-Pakets als Webservice an. Dazu gehört u. a. die Faltung von RNA-Sequenzen nach dem Minimum-Free-Energy-Prinzip (MFE) und die Faltung mit Constraint (Near Natural mit Constraint / NNC).

3.1.2.3.1 Webinterface

Die verschiedenen Funktionen des Vienna RNA Webservers sind unter der folgenden URL erreichbar:

<http://rna.tbi.univie.ac.at/>

(Stand Oktober, 2011)

Die Eingabemaske (Abbildung 14) des RNAfold WebServers bietet keine Namensgebung für die zu faltende Sequenz an. Die Sequenz selbst wird, wie beim mfold-Server, direkt eingegeben oder eingefügt. Für die Faltung unter Anwendung eines Constraint kann zusätzlich eine Zeichenkette mit entsprechenden Informationen (nach RNAfold-Konventionen) eingegeben oder im sogenannten FASTA-Format importiert werden.

Im einfachsten Fall wird die Faltung nach Minimum Free Energy durchgeführt. Diese wird nach dem Algorithmus von Zuker und Stiegler bestimmt (*Zuker & Stiegler, 1981*). Das Format RNA oder DNA wird durch die Faltungsparameter festgelegt; die Sequenz wird dementsprechend automatisch angepasst. Als Faltungs- oder Energieparameter stehen vier verschiedene Optionen zur Verfügung:

- ⇒ zwei Faltungsmöglichkeiten nach RNA-Parametern, jeweils nach Energieparametern, die durch die *Turner-Arbeitsgruppe* ermittelt wurden (*1999, 2004*)
- ⇒ die RNA-Faltung nach dem *Andronescu-Modell* (*2007*)
- ⇒ die Faltung mit DNA-Eigenschaften, nach dem *Mathews-Modell* (*2004*)

Die Faltungstemperatur ist standardmäßig auf 37°C gesetzt, kann aber bei Bedarf vom Nutzer anders angepasst werden. Die Ausgabe des Ergebnisses erfolgt als interaktive Vektorgrafik im SVG-Format. Zusätzlich können Informationen zur Zuverlässigkeit der ermittelten Sekundärstruktur ausgegeben werden (diese Option reduziert die Höchstgrenze der zulässigen Basen pro Sequenz auf 7500).

RNAfold WebServer
1 Enter Input Parameters 2 View Results

[\[Home\]](#) [\[New job\]](#) [\[Help\]](#)

The **RNAfold web server** will predict secondary structures of single stranded RNA or DNA sequences. Current limits are 7,500 nt for partition function calculations and 10,000 nt for minimum free energy only predictions.

Simply paste or upload your sequence below and click *Proceed*. To get more information on the meaning of the options click the symbols. You can test the server using [this sample sequence](#).

Paste or type your **sequence** here: [\[clear\]](#)

[Show constraint folding](#)

Or upload a file in FASTA format:

Fold algorithms and basic options

- minimum free energy (MFE) and partition function
- minimum free energy (MFE) only
- no GU pairs at the end of helices
- avoid isolated base pairs

[Hide advanced options](#)

Advanced folding options

Dangling end options

- no dangling end energies
- unpaired bases can participate in at most one dangling end (MFE folding only)
- dangling energies on both sides of a helix in any case
- allow coaxial stacking of adjacent helices in multi-loops (MFE folding only)

Energy Parameters

- RNA parameters (Turner model, 2004)
- RNA parameters (Turner model, 1999)
- RNA parameters (Andronescu model, 2007)
- DNA parameters (Matthews model, 2004)

Other Parameters

rescale energy parameters to given temperature (C)

assume RNA molecule to be circular

Output options

- interactive RNA secondary structure plot
- RNA secondary structure plots with reliability annotation (Partition function folding only)
- Mountain plot

Notification via e-mail upon completion of the job (optional):

Institute for Theoretical Chemistry | University of Vienna | rna@tbi.univie.ac.at

Abbildung 14: Die Eingabemaske des RNAfold WebServers (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>)

3.1.2.3.2 Lokale Installation

Der Vienna RNA Webserver bietet verschiedene Versionen der Programme als Download an. Zum Beispiel wurde der komplette Quelltext der angebotenen Software in C verfügbar gemacht. Damit ist es Nutzern mit C-Kenntnissen möglich, das Programm für ihr jeweiliges Betriebssystem zu kompilieren, oder auch beliebige Änderungen/Erweiterungen der Funktionen vorzunehmen.

Unter UNIX / Linux

Für ein UNIX / Linux Betriebssystem kann das Vienna RNA Package heruntergeladen und per „configure“ und „make“ installiert werden. Eine detaillierte Anweisung dazu gibt es auf der Homepage des Vienna Webservers:

<http://www.tbi.univie.ac.at/~ivo/RNA/INSTALL.html>

(Stand Oktober, 2011)

Die lokale Installation bietet, im Gegensatz zur Webanwendung, die Möglichkeit, durch verfügbare höhere Rechnerkapazitäten die Obergrenze der Sequenzlänge herauf zu setzen. Für diese Grenze spielt der Arbeitsspeicher die entscheidende Rolle: Für eine zu faltende Sequenz der Länge n ist ein Arbeitsspeicher von $4 \times n^2$ Bytes notwendig. Da die Länge der in dieser Arbeit genutzten Sequenzen bei ca. 1800 Basen liegt, ist jeder moderne PC (Stand 2006: typischerweise 1-3GB Arbeitsspeicher) für die Berechnungen mit dem Vienna RNA Package geeignet.

Zur Faltung wird eine entsprechende Sequenz zunächst im spezifisch vorgeschriebenen FASTA-Format eingelesen:

```
>Pecten_maximus  
CAACCUGGUUGAUCCUGCCAGUAGUCA ...
```

Der Befehl zur nachfolgenden Berechnung setzt sich syntaktisch folgendermaßen zusammen:

```
RNAfold [-p[0/2]] [-C] [-T temp] [-4] [-d[0|1|2|3]] [-noLP] [-noGU] [-noCloseGU] [-e 1/2] [-P paramfile] [-nsp pairs] [-S scale] [-circ] [-MEA [gamma]]
```

Ohne Vorgabe von Optionen wird die gegebene Sequenz nach Minimum Free Energy (MFE) gefaltet und das Ergebnis in Dot-Bracket-Format ausgegeben. Zusätzlich erzeugt RNAfold PostScript-Dateien für die grafische Darstellung der MFE-Struktur und den sogenannten Dot-Plot für die Basenpaarungen.

Die für diese Arbeit wichtigsten Optionen für RNAfold sind „-p[0/2]“ und „-C“. Erstere berechnet zusätzlich zur Struktur die Zustandssumme („Partition Function“) und die Wahrscheinlichkeiten der berechneten Basenpaare. Die Option „-C“ ermöglicht die Verwendung eines Constraints zur Faltung der Sequenz. Der gewünschte Constraint muss im Dot-Bracket-Format in der Eingabe enthalten sein:

```
>Pecten_maximus
CAACCUUGGUUGAUCCUGCCAGUAGUCA ...
...((((((((.....))))))(((.( ...
```

Eine installierte Version des Vienna RNA package bietet außerdem die Möglichkeit, Sekundärstrukturen nach „Maximum Expected Pair Accuracy“ (MEA) zu generieren.

Eine detaillierte Beschreibung der weiteren Optionen ist unter der folgenden Adresse verfügbar:

<http://www.tbi.univie.ac.at/~ivo/RNA/man/RNAfold.html>

(Stand Oktober, 2011)

Unter Windows

Das Vienna RNA Package ist grundsätzlich für UNIX / Linux Systeme geschrieben, sodass die Nutzung unter Windows mit größerem Aufwand verbunden ist. Die einfachste und zuverlässigste Variante ist der Einsatz der Linux-Umgebung „Cygwin“ (<http://www.cygwin.com/>). Damit ist es unter Windows möglich, die gesamten Funktionen von RNAfold analog zu der Handhabung unter Linux zu nutzen.

3.1.2.4 RNAsalsa

Der Ableger von RNAfold war zum Zeitpunkt der Nutzung (2007) nur als Installation auf einem Linux-System zu betreiben. Inzwischen gibt es die Software auch zum Installieren

unter Windows-Systemen. Das Programm und der Source Code sind als Download unter der folgenden Adresse kostenfrei erhältlich:

[http://zfmk.de/web/Forschung/Abteilungen/AG_Wgele/
Software/RNAsalsa/index.en.html](http://zfmk.de/web/Forschung/Abteilungen/AG_Wgele/Software/RNAsalsa/index.en.html)

(Stand Oktober, 2011)

Die Steuerung erfolgt in allen Versionen der Software über die Kommandozeile:

```
RNAsalsa [-h] [-v] [-X] [-p] [-s1] [-s2] [-s3] -c FILE -i FILE
```

Die zu faltenden Sequenzen müssen als eine Textdatei im FASTA-Format (-i FILE) vorliegen. Dazu ist eine zweite Datei mit den Constraint-Informationen notwendig (ebenfalls im FASTA-Format) (-c FILE). Die Ausgabe erfolgt als PostScript-Dateien (-p) aller gefalteten Sekundärstrukturen.

RNAsalsa ermöglicht die Erstellung von Sekundärstrukturen nach dem Prinzip „Near Natural mit Constraint“ (NNC). Das heißt, ein Alignment von Sequenzen, deren Sekundärstrukturen nicht bekannt sind, wird mit der Sequenz einer nah verwandten Art kombiniert, deren Sekundärstruktur bekannt ist. Diese Strukturinformation dient als Vorlage zur Faltung der anderen Sequenzen.

Die Faltung der Sequenzen in RNAsalsa hängt in starkem Maße von der konkreten Wahl des Alignments ab (Abbildung 15). Die wichtigsten Parameter, die das Programm zusätzlich bietet, betreffen die Anwendung des Constraints (-s1, -s2, -s3). Bei den Parametern von Alignment und Constraint ist meist ausführliches Testen notwendig, um alle Parameter sinnvoll festzulegen. (Detaillierte Erklärungen zu den Constraint-Parametern enthält das RNAsalsa Handbuch) (*Stocsits, 2009*).

Wie in Abbildung 15 zu sehen ist, kann das Sequenzalignment trotz festem Constraint das Ergebnis der Faltung beeinflussen.

Pecten maximus: Faltung mit RNAsalsa (NN) mit gleichem Constraint, gleichen Faltungsparametern aber unterschiedlichen Sequenzalignments.

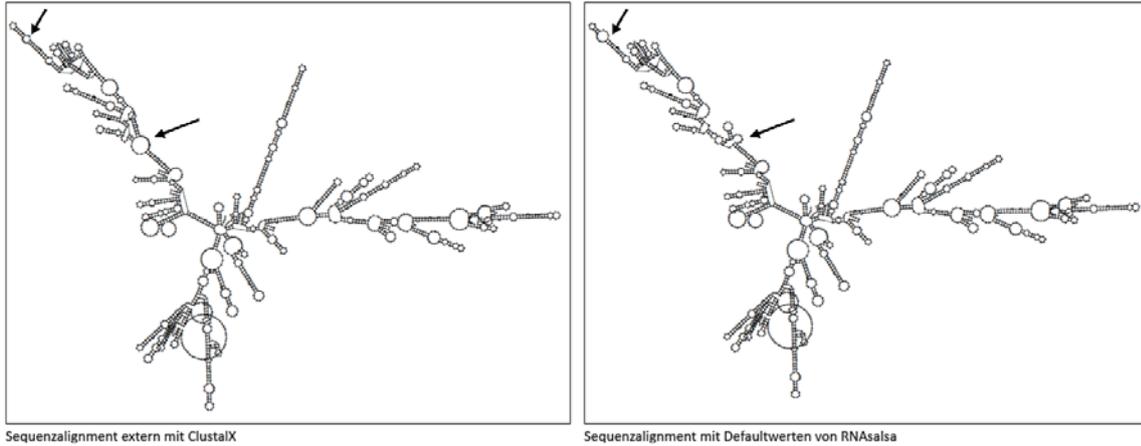


Abbildung 15: Faltung der 18S rRNA von *Pecten maximus* mit RNAsalsa unter Nutzung verschiedener Sequenzalignments.

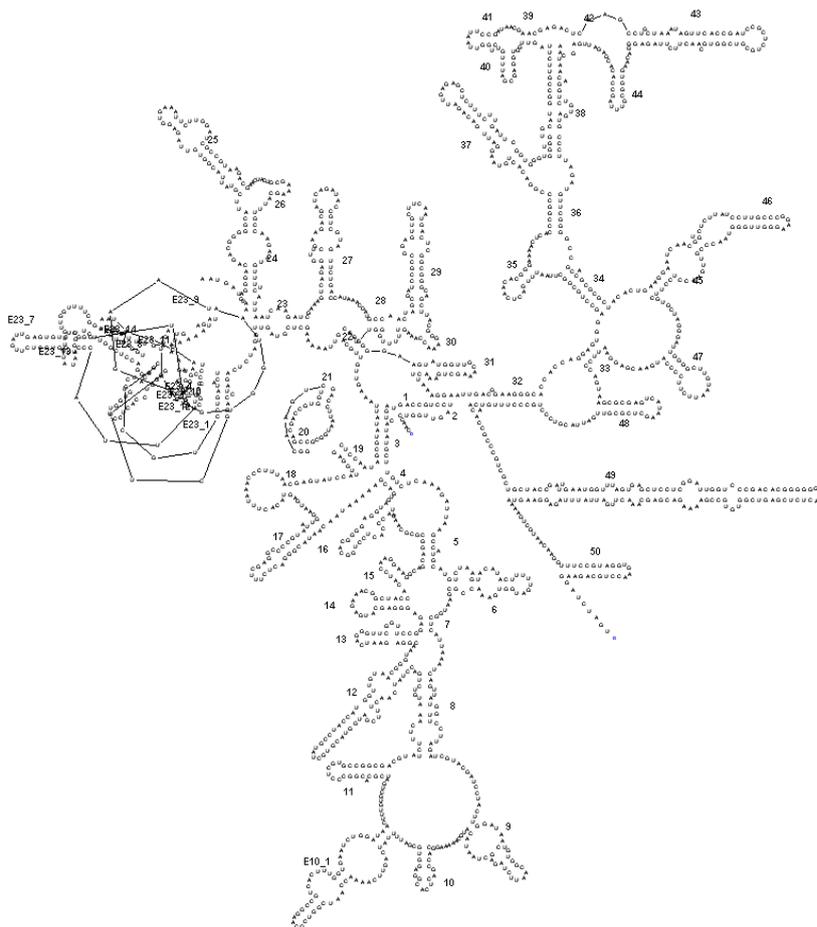


Abbildung 16: Darstellung der Sekundärstruktur von *Pecten maximus* mit RNAViz.

3.1.2.5 RNAViz

RNAViz ist eine lokal zu installierende Software, welche die grafische Darstellung von RNA-Sekundärstrukturen ermöglicht. Als Eingabe dienen die ct-Dateien, die mfold liefert oder DCSE-Alignments, welche zusätzliche Strukturinformationen enthalten.

Das Ziel von RNAViz ist es, Sekundärstrukturen übersichtlich und ohne überlappende Bereiche darzustellen. Beim Laden einer neuen Sekundärstruktur erweist sich die erste Darstellung meist als die schwierigste, da es in der Regel noch viele überlappende Bereiche gibt und die eigentliche Struktur kaum zu erkennen ist. Die GUI von RNAViz ist so gestaltet, dass der Benutzer per Drag&Drop einzelne Sequenzen oder Loops bewegen und so die grafische Darstellung leicht verändern kann. Es bleibt dem Geschick des Nutzers überlassen, wie übersichtlich die fertige Anzeige ist (Abbildung 16). Aus der „geordneten“ Sekundärstruktur kann RNAViz automatisch Strukturinformationen ableiten und in einer sogenannten Skeleton-Datei speichern.

Die Darstellung von Pseudoknoten und unformatierten Abschnitten, sowie die automatische Nummerierung oder Markierung von Abschnitten ist ebenfalls möglich. Mit Hilfe von Skeleton-Dateien kann eine größere Anzahl ähnlicher RNA-Strukturen mit geringem Aufwand in ähnlicher Weise dargestellt werden.

Die Arbeit mit der Software zeigte, dass RNAViz für lange Sequenzen wie die 18S rRNA einen größeren Aufwand als Nutzen bedeutet. RNAViz ist für die anschauliche Darstellung kurzer Sequenzabschnitte und einzelner Merkmale gut geeignet, weniger aber für die professionelle Analyse von einer großen Anzahl Sequenzen mit größeren Längen, wie sie bei dieser praktischen Genanalyse häufiger vorkommen.

3.1.2.6 Zusammenfassung

Die gegenwärtig verfügbare Faltungssoftware ist primär zu dem Zweck entwickelt worden, die Funktionsweise von RNA zu verstehen und die Sequenzalignments für die Analyse der Primärsequenzen zu verbessern. Das Alignment der Sequenzen spielt in der Analyse der Primärstruktur eine grundlegende Rolle. Die Methode der Molekularen

Morphologie entstand als ein interessantes und aussichtsreiches „Nebenprodukt“ der Verfügbarkeit hochwertiger Faltungsalgorithmen.

Im Prinzip ist die „Korrektheit“ der erzeugten Strukturen für die Molekulare Morphologie zwar erwünscht, aber nicht entscheidend, da die Ergebnisse in jedem Fall künstlich (3D-Struktur wird auf 2D abgebildet) sind, und weil eine vergleichende phylogenetische statt einer biochemischen Analyse durchgeführt wird. Maßgebliches Kriterium für den Einsatz eines bestimmten Verfahrens war in jedem Falle, dass die generierte 2D-Struktur eine hinreichend große Anzahl klar erkennbarer und voneinander separierte Strukturmerkmale aufweisen musste, um eine möglichst eindeutige morphologische Analyse zu gestatten.

Eine grundsätzliche praktische Schwierigkeit der Methode besteht darin, dass MFE die Erzeugung einer Anzahl alternativ möglicher, in gewissem Sinne gleichberechtigter Sekundärstrukturen bedeutet. Die Ursachen dafür werden unter dem Begriff „Frustration“ im Abschnitt 3.2.1 erläutert. Nur ein Teil dieser Menge der formal möglichen Sekundärstrukturen ist für die Analyse mittels Molekularer Morphologie geeignet. Die Genauigkeit (d.h. Annäherung an die „natürliche“/„wahre“ Sekundärstruktur) der Vorhersage kann mittels Verwendung von Constraints bei Wechsel von MFE zu NNC deutlich verbessert werden.

Außer MFE und NNC gibt es weitere Möglichkeiten, Sekundärstrukturen vorherzusagen. Eine davon ist die Methode nach dem Prinzip der „Maximum Expected Pair Accuracy“ (MEA) (*Do et al., 2006; Lu et al., 2009; Reuter & Mathews, 2010*), bei der die Wahrscheinlichkeiten der Basenpaarungen analysiert werden. Die sich daraus ergebende Freie Enthalpie spielt eine untergeordnete Rolle. Diese Methode war zum Zeitpunkt dieser Arbeit noch nicht als implementierte Funktion verfügbar, ist es aber inzwischen bei RNAstructure und RNAfold.

Die meisten Algorithmen sind auf kurze Sequenzen ausgelegt (<700 Nukleotide), die Genauigkeit der Vorhersage der Sekundärstruktur sinkt mit der Länge der Sequenz. Die Übereinstimmung zwischen berechneten und im Labor untersuchten RNA liegt bei

kurzen Sequenzen bei etwa 70-80%, dabei verbessert sich die Vorhersage durch die vergleichende Analyse einer oder mehrerer homologer Sequenzen (zum Beispiel mit Dyn- und Multialign von RNAstructure) (Reuter & Mathews, 2010; Andronescu et al., 2010).

In überschaubarer Zeit soll eine größere Anzahl Sequenzen sinnvoll gefaltet und analysiert werden. Die Berechnung der Faltung einer Sequenz dauert im Schnitt drei bis zehn Minuten, je nach Serverauslastung kann die notwendige Zeit variieren. Die Faltung auf einem normalen PC (Win7, 32bit, 3GB) dauert in etwa die gleiche Zeit. Die Faltung des gesamten Alignments (mit ca. 120 Sequenzen) mit RNAsalsa dauert je nach Parametern mehrere Stunden bis zu einem Tag.

Die Entscheidung für oder gegen eine bestimmte Software erfolgte nach den Kriterien der bequemen Handhabbarkeit, Import-Export-Möglichkeiten, Nachvollziehbarkeit der Ergebnisse, konsistente Verfügbarkeit von MFE und NNC aus der gleichen Programmfamilie, Rechenzeit / Zeitaufwand, Aktualität der Software, und nach morphologischen Kriterien, das heißt wieviele Merkmale gibt es und wie stark sie ausgeprägt sind.

Keines der getesteten Faltungs-Programme kann automatisch Pseudoknoten vorhersagen, wie sie in der Realität vorkommen, siehe Abschnitt 3.3.1.1, „Pseudoknoten“.

In der Praxis haben sich RNAfold und RNAsalsa als die günstigsten Programme für die vorliegende Arbeit herausgestellt, RNAfold für einfaches MFE-Falten, RNAsalsa für die Faltung nach NNC.

Die von RNAfold und RNAsalsa gelieferten PostScript-Dateien waren für die visuelle Analyse ausreichend, sodass die Nutzung von RNAViz und XRNA nicht über das Teststadium hinausging. RNAViz wurde aber für die Darstellung einzelner Merkmale genutzt.

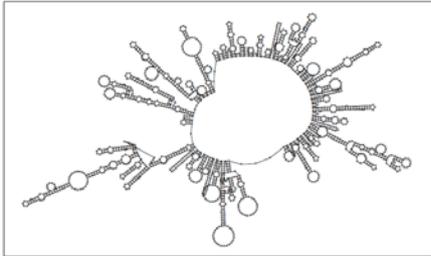
Tabelle 3: Ein- und Ausgabemöglichkeiten der Software-Varianten zur Vorhersage von RNA-Sekundärstruktur

Software	Mfold 3.2	RNAstructure	RNAfold 2.0	RNAalsa
Algorithmus	Zuker	Zuker	Zuker	Zuker
Installationssoftware für	UNIX/Linux	Win	UNIX/Linux	UNIX/Linux, Win
Webinterface	x	-	x	-
Inputformat	FASTA	SEQ	FASTA	FASTA
Basenlimit	9000		10000	
Constraint	proprietäres Format		Dot-bracket	Dot-bracket

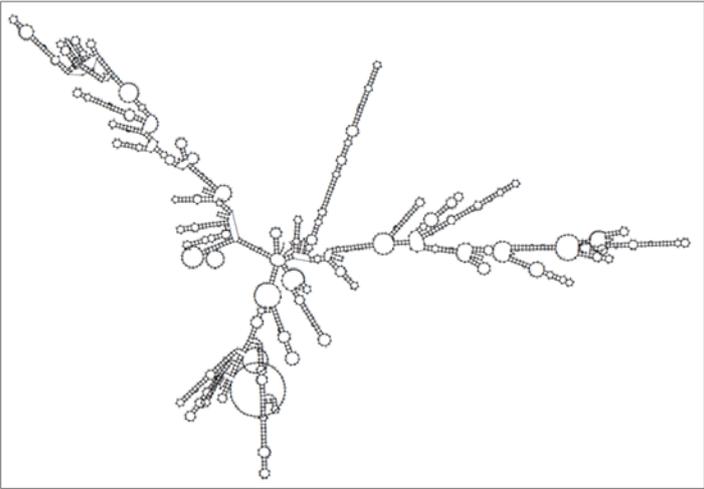
Parameter	Mfold 3.2	RNAstructure	RNAfold 2.0	RNAalsa
Temperatur	37°C		Variabel	fest
Turner model (2004)	X	X	X	X
Turner model (1999)	X	X	X	X
Andronescu model (2007)			X	
Matthews model (2004)		X	X	
MFE	X		X	-
NNC	X	-	X	X
MEA	-	X	X	-

Output	Mfold 3.2	RNAstructure	RNAfold 2.0	RNAalsa
Format	ps, png, jpg, svg	ct	ps	ps
Reliability Notation	X		X	
Base Pair Probability		X		
Energy Dot Plot	X			
Partition Function			X	X

Material & Methoden



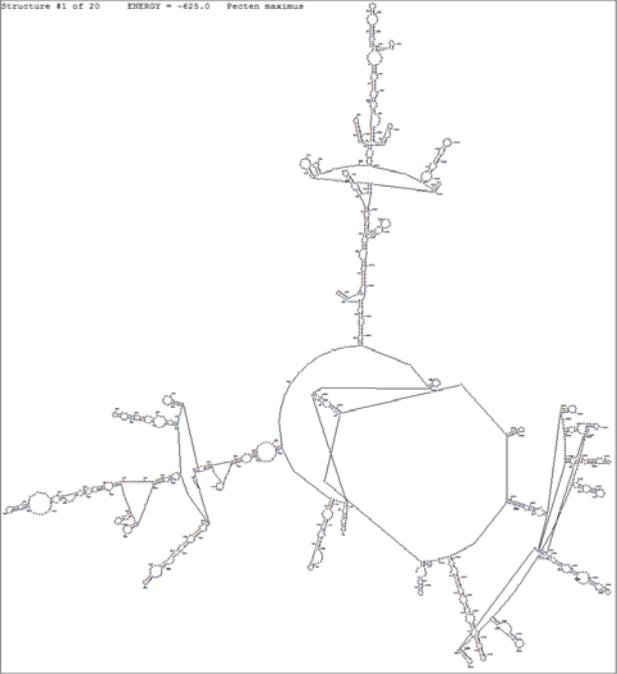
Pecten maximus: RNAfold, MFE, DNA Parameter



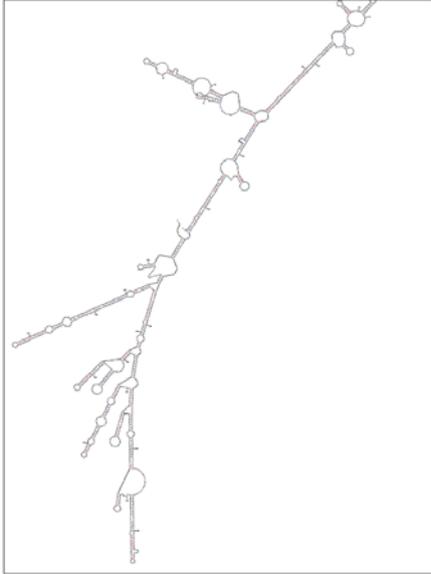
Pecten maximus: RNAsalsa, NNC



Pecten maximus: RNAfold, MFE, RNA Parameter



Pecten maximus: RNAstructure, MFE



Pecten maximus: mfold, MFE

Abbildung 17: Unterschiedliche Ergebnisse für die Sekundärstruktur der 18S rRNA von *Pecten maximus* nach Anwendung der verschiedenen Faltungsalgorithmen RNAfold, RNAsalsa, mfold, RNAstructure.

3.2 Faltung

Für die Faltung der Sequenzen wurden verschiedene Methoden angewandt. Dieses Verfahren wurde mit dem identischen Satz von Sequenzen für verschiedene Faltungsalgorithmen und -parameter wiederholt.

3.2.1 Faltung nach Gibbs / Minimum Free Energy (MFE)

Wie bereits im Abschnitt 3.1, „Material“, beschrieben, gibt es verschiedene Algorithmen und Parameter, mit denen Primärsequenzen zu 2D-Strukturen gefaltet werden können. Die einfachste Methode ist die nach dem thermodynamischen Prinzip der „Minimum Free Energy“ (MFE).

Besonders bei Sekundärstrukturen nach MFE stellt sich das Problem der sogenannten Frustration (*Ebeling et al., 1990*). Eine Aufgabe ist frustriert, wenn mehrere gleichwertige, im Detail aber sehr unterschiedliche Lösungen (d.h., Kompromisse zwischen den widersprüchlichen Anforderungen) für ein Problem möglich sind, und keine Lösung definitiv mit Abstand die „beste“ ist. Dieses Problem ist z.B. auch von der Faltung und Evolution von Proteinen bekannt (*Oliveira et al., 2006*), aber auch von verwandten mathematischen Aufgaben wie dem „travelling salesman problem“, bei dem der kürzeste Weg gesucht wird, um eine gegebene Menge von Orten auf einer Karte zu besuchen (*Bohr & Brunak, 1989*). So lassen sich für eine Sequenz oft mehrere unterschiedliche Sekundärstrukturen finden, die eine ähnlich niedrige Freie Enthalpie aufweisen. Kleine Modifikationen der Faltungsparameter führen dann dazu, dass die jeweils „besten“ Lösungen sehr unterschiedliche Strukturen besitzen können. Das Faltungsprogramm mfold bietet deshalb als Ausgabe die 30 Sekundärstrukturen mit den niedrigsten Enthalpiewerten an. Aus diesen Möglichkeiten ist dann eine auszuwählen, deren Freie Enthalpie möglichst wenig vom besten Wert abweicht und die 2D-Struktur denen der anderen betrachteten RNA möglichst ähnlich ist. Solche Entscheidungen sind nur bedingt zu automatisieren und lassen stets einen gewissen subjektiven Spielraum.

3.2.1.1 Faltung mit RNAfold

Nach Austesten der verschiedenen Programme wurde die Faltung nach MFE mit RNAfold vom Vienna-Server durchgeführt.

Zur Vorbereitung brauchten die Sequenzen nur geringfügig modifiziert zu werden: Zu kurze Sequenzen und „falsche“ Buchstaben und Leerzeichen wurden entfernt.

Der erste Schritt vor dem seriellen Falten war das Testen der Parameter, die RNAfold anbietet. Die generierten Sekundärstrukturen sollten eine möglichst geringe Variabilität zwischen den Arten aufweisen, um zwischen konservativen und variablen Bereichen unterscheiden zu können.

Die intuitiv logische Variante der „RNA-Parameter“ erzeugt eine Struktur, die sehr wenige und wenig variable Merkmale zur Analyse aufweist (Abbildung 17). Während dies im Sinne der „konservativen“ RNA vielleicht „richtig“ erscheint, muss beachtet werden, dass sich die „Konservierung“ der RNA auf funktionelle Abschnitte bezieht. Die Faltung nach MFE nimmt jedoch auf funktionelle Regionen keine Rücksicht. Es entsteht eine „nur“ scheinbar „konservierte“ Sekundärstruktur.

Um für die visuelle Analyse im Sinne der Molekularen Morphologie eine geeignetere Sekundärstruktur zu erhalten, wurde deshalb die Faltung auch mit „DNA-Parametern“ durchgeführt. Die daraus resultierenden Strukturen wiesen nun eine Vielzahl sehr variabler Merkmale (zwischen 30 und 50) auf (Abbildung 17).

Im Vorfeld bestand die Vermutung, dass die zu erwartenden Sekundärstrukturen im Grunde sehr ähnlich sind, und die Unterschiede (d.h. nicht funktionelle, „frei“ mutierende Bereiche) zwischen ihnen Rückschlüsse auf die Verwandtschaft der untersuchten Arten zulassen. Diese Annahme erwies sich jedoch für die Faltung nach MFE mit DNA-Parametern als zu optimistisch. Dementsprechend wurde der Schwerpunkt der visuellen Vergleiche auf die gemeinsamen Merkmale (d.h. auf funktionelle, konservierte Bereiche) verlegt und den variablen Bereichen geringere Bedeutung zugemessen.

Bei der anschließenden Analyse wurden wegen ihrer noch unbekanntem Relevanz zunächst alle Merkmale in die Merkmalsmatrix aufgenommen. Erst am Ende konnte festgelegt werden, welche Merkmale tatsächlich gleichzeitig bei einer größeren Anzahl Arten zu finden waren.

3.2.1.1.1 RNAfold: Parameter

Die Faltung der Sequenzen nach MFE wurde Anfang 2006 über das Webinterface von RNAfold (siehe Abschnitt 3.1.2.3, „RNAfold“) des Vienna Servers durchgeführt. Die Parameter wurden wie folgt gesetzt:

- ⇒ minimum free energy only
- ⇒ use DNA parameters
- ⇒ Rescale energy parameters to temperature 37°C

- ⇒ special tetraloops
- ⇒ dangling end energies
- ⇒ GU pairs at the end of helices
- ⇒ avoid isolated base pairs

Zum jetzigen Zeitpunkt (2011) sind auch noch andere Parameter und Faltungen möglich.

3.2.1.2 Beispiel *Pecten maximus* mit RNAfold

Am ausgewählten Einzelbeispiel der Pilgermuschel (*Pecten maximus*) kann die Methode in umfassender Weise exemplarisch dargestellt und deutlich gemacht werden. Die in Abschnitt 3.1.1 „abgerufene“ Information wurde zunächst in das FASTA-Format umgewandelt; diese Option bietet GenBank inzwischen automatisch an. Die resultierende Zeichenkette enthielt keine Leerzeichen oder Buchstaben außer den Nukleinsäuren a, c, g und t.

```
caacctggttgatcctgccagtagtcatatgcttgtctcaaagattaagccatgcatgtctaagtacataactttttgatg
gtgaaaccgcgaatggctcattaaatcagttatggttccttagatcgtagatcctacttggataactgtggcaattcta
gagctaatacatgcaaaaaggcaccgactcacggaggtgcgatttttatcagttcaaaaccaatcggtcgcaaggccgtca
ctttggtgaatctggataactttgtgctgatcgcacggccctcgtgccggcgacgtatctttcaaatgtctgacctatca
```

Material & Methoden

```
actttcgatgggtacgtgctatgcctaccatgggttgtaacgggtaacggagaatcagggttcgggtccggagagggagcat
gagaaacggctaccacatccaaggaaggcagcagggcgcgcaaattaccactcctggcacggggaggtagtgcgaaaaa
taacaatacgggactctttcgaggccccgtaattggaatgagtacactttaaacctttaacgagatccattggagggc
aagtctgggtgccagcagccgcggttaattccagctccaatagcgtatattaaagttggtgcagttaaaaagctcgtagttg
gatctcagatccaggcttgccgtccactccttggtggctactgctcctgatctacctcccgggtttacccttggtgc
tcttgattgagtgctcgggtggccggaacatttactttgaaaaaattagagtggtcaaagcaggcagctcgcctgaata
atggtgcatggaataatggaataggacctcgggttctattttggtggttttcggaacttgaggtaatgattaagagggaca
gacgggggcatctcgtattacgggtgtagagggtgaaattcttggtatcgccgtaagacgaaactactgcgaaagcatttgcca
agaatgttttcattaatcaagatcgaaagtcagagggtcgaagacgatcagataccgtcgtagttctgaccataaacgat
gccaactagcgcacccggagttgcttcaatgactcggcgggcagcttacgggaaaccaaagttttgggtccggggg
aagtatggttgcaaagctgaaacttaaaggaattgacggaagggcaccaccaggagtggagcctgcggcttaatttgact
caacacgggaaaactcaccggcccgacactgtaaggattgacagattgagagctcttcttgattcgggtgggtggtgg
tgcatggccggttcttagttggtggagcgatttgctggttaattccgataacgaacgagactctagcctgctaaatagtt
caccgatccgctctgcgtcgggtgcaacttcttagagggacaagtggcgttttagccacacgagattgagcaataacaggtc
tgtgatgcccttagatggtcggggccgcacgcgcgctacactgaaggaatcaacgtgctcttatccttgcccggaaggtt
tgggtaacccttgaaacctccttcgtgctagggattggggcttgtaattcttccccatgaacgaggaattcccagtaagc
gcgagtcataagctcgcgttgattactgcctccttgacacaccgcccgtcgtgctaccgatgaaatgggttagt
gagctcctcggattggtcccgcacgggggaaactctcgagtcggtgtgccgaaaagacgagcaaaacttgattatttag
aggaagtaaaagtcgtaacaaggtttccgtaggtgaacctgcagaaggatctagt
```

Diese Zeichenkette konnte per Copy & Paste in das Eingabefeld vom RNAfold Webserver eingesetzt oder als Datei im FASTA-Format hochgeladen werden. Das Resultat hängt nicht von der gewählten Weise der Eingabe ab. Als Parameter für die Faltung wurden die im Abschnitt 3.2.1.1 angegebenen Werte ausgewählt.

Die 18S rDNA von *Pecten maximus* liegt mit 1815 Basen unter der Längenbegrenzung, bis zu der RNAfold falten kann (siehe Abschnitt 3.1.1, „Sequenzen“). Die eingegebene Sequenz wurde als „Batch-Job“ in eine Warteschlange eingereiht, das heißt, sie wurde an das Ende aller vorher online in Auftrag gegebenen Faltungen gestellt. Der Nutzer erhielt eine URL, unter der die generierte Sekundärstruktur im Anschluss abgerufen werden konnte.

Mit den genannten Einstellungen hat RNAfold die Sekundärstruktur in Abbildung 18 generiert.

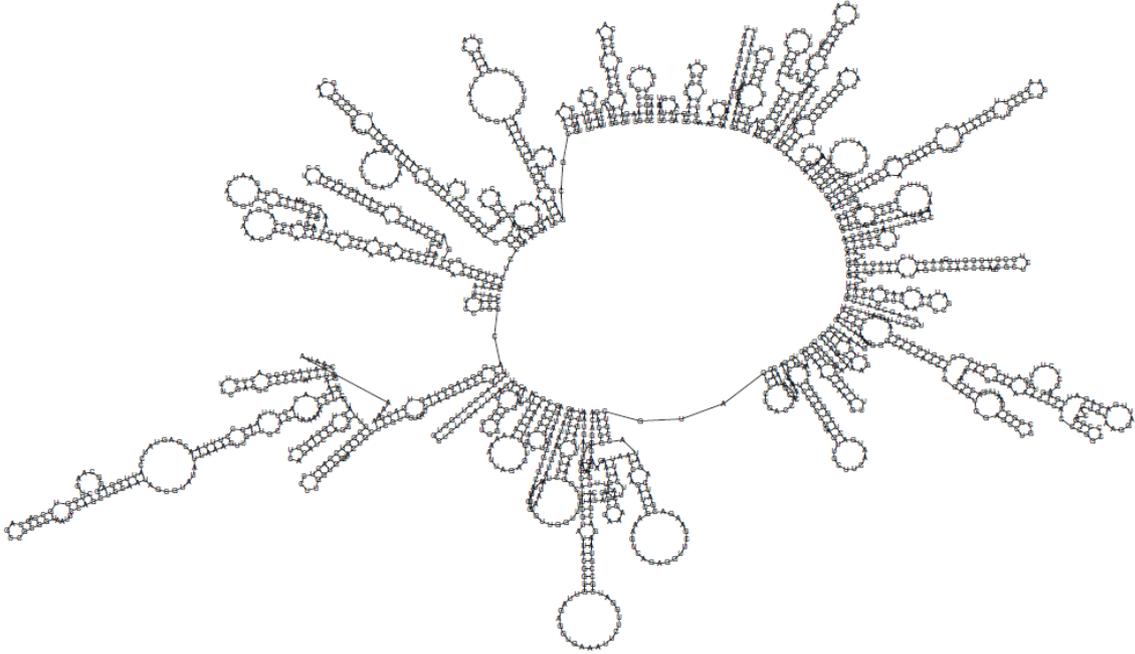


Abbildung 18: MFE-Sekundärstruktur der 18S rDNA von *Pecten maximus* nach dem RNAfold-Algorithmus

3.2.2 Faltung mit festen Bindungen, Near Natural (NNC)

Um die Ergebnisse für die nach MFE gefalteten Sekundärstrukturen in Bezug auf die Problemstellung zu optimieren und das Frustrationsproblem der möglichen 2D-Strukturen zu reduzieren, gibt es die Möglichkeit, bei der Faltung der Sequenz einen sogenannten Constraint anzuwenden. Solch ein Constraint wurde meist experimentell im Labor ermittelt und gibt für bestimmte Gruppen von Basen gewisse Regeln für die Paarbildung vor. Dabei wird jeder einzelnen Base je eine der Alternativen vorgeschrieben, dass

- ⇒ sie nicht gebunden werden darf
- ⇒ mit welcher anderen Base sie gebunden werden muss
- ⇒ sie optional zur Paarbildung nach dem MFE-Algorithmus verwendet werden soll

Ein Constraint ist eine Einschränkung in der MFE-Faltung und soll Sekundärstrukturen liefern, die näher an der natürlichen Form der rRNA sind als das bei völliger Freigabe aller Basen in der Regel der Fall ist. Auf diese Weise sollen weniger und dafür aussagekräftigere Merkmale erzeugt werden.

Im Gegensatz zur uneingeschränkten Faltung nach "Minimum Free Energy" erfordert das Falten mit Constraint zusätzliche Vorbereitungen. Für die Formulierung der ergänzenden Vorschriften ist ein Alignment der Sequenzen notwendig und die Zuweisung eines Consensus Constraint für alle einbezogenen Sequenzen.

Für die Faltung standen die Programme mfold, RNAfold und RNAsalsa zur Auswahl. Als am besten geeignete Software erwies sich RNAsalsa, siehe Abschnitt 3.1.2.4. Dieses Programm bietet die Möglichkeit, mehrere Sequenzen vergleichend zu analysieren und unter Vorgabe des sogenannten Consensus Constraint zu falten.

Notwendig für die Faltung mit RNAsalsa ist also ein bekanntes Constraint einer (möglichst nah verwandten) Art, die im Alignment der Sequenzen enthalten sein muss. Experimentell sind Constraints aufgrund der aufwendigen Untersuchung bisher nur für eine begrenzte Anzahl Arten ermittelt worden.

Bei der Abarbeitung des Algorithmus wird die Sequenz mit Constraint zum „Ausgangspunkt“ für das Alignment aller Sequenzen. Im Programm ClustalX wird das Alignment vorgenommen. Ein Alignment von mehreren Sequenzen verschiedener Arten hilft dabei, vergleichbare bzw. verwandte funktionelle Abschnitte innerhalb der Gensequenzen zu identifizieren.

Im Rahmen des Alignments werden „Leerstellen“ in die Sequenz „eingefügt“, um InDels auszugleichen. Die vom Nutzer gesetzten Parameter legen die Höhe der Strafen für eingefügte Leerstellen und die Zahl der Pluspunkte für übereinstimmende Basen fest. Entsprechend wird ein „bestes“ Alignment durch maximale Punktzahl gemäß den gesetzten Parametern ermittelt. Bereits bekannte funktionelle/analoge Abschnitte können ein Alignment verbessern (\Leftrightarrow Constraint-Informationen). RNAsalsa bietet optional an, intern ein automatisches Alignment mit Constraint vorzunehmen, ohne explizites Zutun des Anwenders.

Ähnlich wie bei der Faltung nach MFE, bei der gleichwertige Sekundärstrukturen ein frustriertes Problem darstellen, besteht ein analoges Problem auch für das Alignment der Primärsequenzen. In der Natur gibt es nur ein „richtiges“ Alignment, dieses kann aber nicht leicht bzw. in eindeutiger Weise gefunden werden. Je nach Einstellung der Parameter in der Alignment-Software können gleichwertige Alignments gefunden werden, die durch den Benutzer nicht a priori als „falsch“ oder „richtig“ klassifiziert werden können. Dadurch entsteht auch an dieser Stelle Raum für relativ willkürliche, subjektive Entscheidungen des jeweiligen Bearbeiters.

Nach Festlegung des Alignment steht für jede Sequenz ein (Consensus) Constraint zur Verfügung. Dieser ist in der Regel nicht exakt so wie es ein im Labor untersuchter wäre, nähert sich diesem erfahrungsgemäß aber zumeist hinreichend gut an und ist mit deutlich weniger Aufwand zu ermitteln (*Stocsits, 2009*). Außerdem ergibt sich die Möglichkeit, NNC-Faltungen in größerer Anzahl automatisiert vorzunehmen, ohne vorher experimentell Strukturinformationen ermitteln zu müssen.

Nach Vorliegen eines Consensus Constraint können alle Sequenzen einer Gruppe von Spezies mit RNAsalsa nach der NNC-Methode gefaltet werden.

3.2.2.1 Vorbereitung

Ähnlich wie bei der Faltung mit MFE (siehe Abschnitt 3.2.1) müssen die Sequenzen in das einheitliche FASTA-Format gebracht werden. Alle Sequenzen müssen in einer Datei zusammengefasst werden:

```
>Acila_castrensis
tcatatgcttgtctcaaagattaagccatgcatgtctaagtacaaactttcacatagtgaaaccg...
>Nucula_proxima
tcatatgcttgtctcaaagattaagccatgcatgtctaagtacagactttcacatagtgaaaccg...
>Nucula_sulcata
tcatatgcttgtctcaaagattaagccatgcatgtctaagtacaaactttcacatagtgaaaccg...
```

Die erste Zeile enthält einen eindeutigen Namen für die Sequenz (ASCII, ohne Leerzeichen), die zweite Zeile enthält die Sequenz (ohne Leerzeichen oder Zeilenumbrüche). In diesem Format können sie in ClustalX aligniert oder durch RNAsalsa gefaltet werden.

3.2.2.2 Alignment mit ClustalX

Die in einer Datei im FASTA-Format vorliegenden Sequenzen können in ClustalX eingeladen werden. Es folgt eine Ansicht aller Sequenzen untereinander, die Reihenfolge entspricht der in der Datei vorgegebenen (siehe Abbildung 19). Die farbliche Markierung der Basen im Hauptfenster des Programms lassen bereits gewisse Versetzungen von Abschnitten in den einzelnen Sequenzen vermuten. Diese gilt es, durch sinnvolles Einfügen von Leerstellen auszugleichen.

ClustalX berechnet in mehreren Schritten das Alignment der Sequenzen für die jeweils Parameter gesetzt werden müssen. Die wichtigsten sind die Einstellungen für das „Pairwise Alignment“ und das „Multiple Alignment“, siehe die in Abbildung 20 gezeigten Einstellungsmasken des Programms. Der Nutzer legt fest, welche Strafen ClustalX für das Einfügen eines „Gaps“ (Leerstelle) und dessen Ausweitung einrechnen soll.

Eine genaue Beschreibung der Parameter kann im Handbuch von Jarno Tuimala unter

http://www.csc.fi/english/research/sciences/bioscience/programs/clustalw/clustalx_pdf
(Stand Oktober, 2011)

abgerufen werden.

Das endgültige Festlegen der Parameter ist jedoch weiterhin mit Testen verbunden und muss vom Benutzer visuell kontrolliert und ggf. korrigiert werden. Das heißt, an bestimmten Stellen werden Leerstellen manuell entfernt oder zusätzlich zum automatischen Alignment eingefügt. Daraus resultiert eine aneinander ausgerichtete Sortierung der Basen aller Sequenzen (Abbildung 21). Zusammen mit einem Constraint einer nah verwandten Art kann dieses Alignment von RNAsalsa zur Faltung aller Sequenzen genutzt werden, die im betrachteten Genpool enthalten sind.

Als Parameter-Sets stellten sich die in Abbildung 20 dargestellten Einstellungen als günstig heraus. Anschließend wurde das Alignment auf Fehler überprüft und gegebenenfalls manuell ausgebessert: Die einzelnen Sequenzen mussten in ihrer Ausrichtung visuell miteinander verglichen werden. Einzelne Basen, bzw. größere Abschnitte mussten zum Teil manuell verschoben werden. Auf diese Weise wurden analoge (und gegebenenfalls homologe) Sequenzabschnitte einander zugeordnet. Es wurden zusätzlich die Faltungsinformationen des Constraints mit ausgerichtet.

3.2.2.3 Faltung mit RNAsalsa

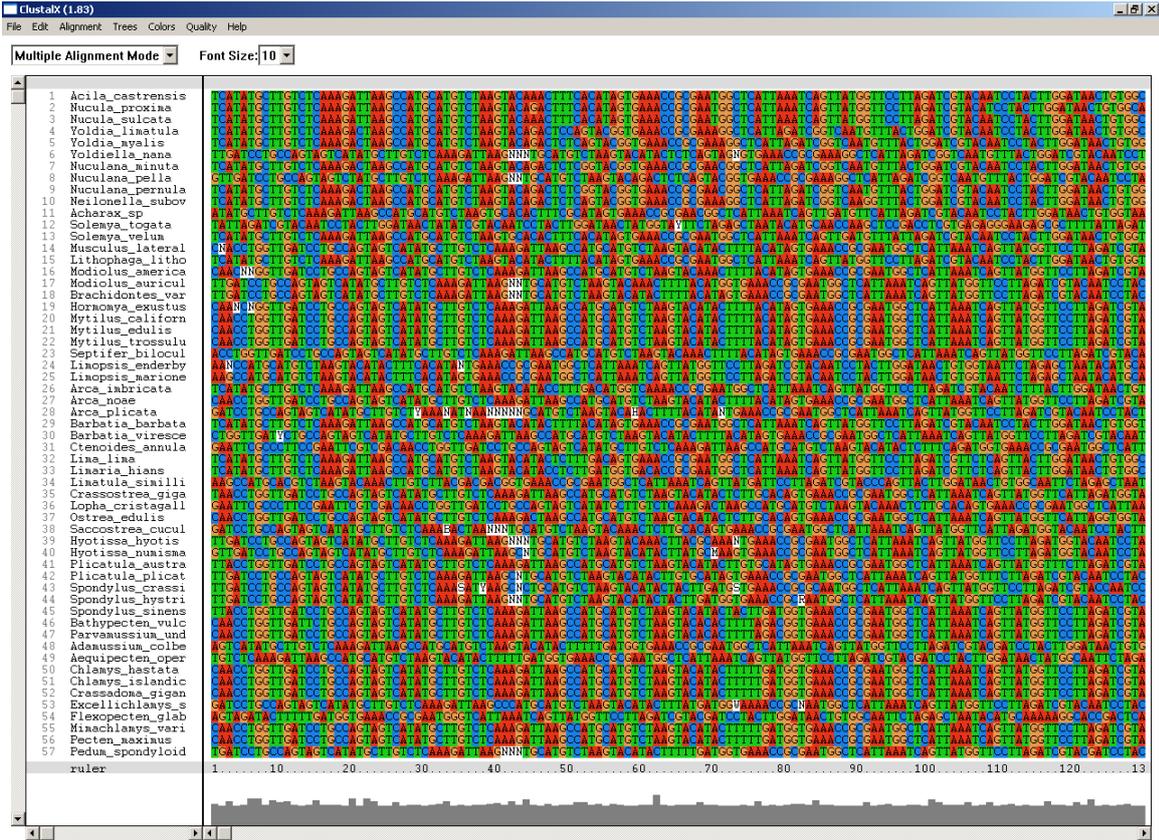
Die Voraussetzung für die Herstellung einer oder mehrerer Sekundärstrukturen durch Faltung mittels RNAsalsa ist das geeignete Alignment aus zwei oder mehr Sequenzen (siehe Abschnitt 3.2.2.2) und die Constraint-Informationen von einer nah verwandten Art. Alignment und Constraint müssen als separate Dateien im FASTA-Format an das Programm übergeben werden. Das Sequenzalignment muss so formatiert sein, wie es in Abschnitt 3.2.2.1 beschrieben ist. ClustalX bietet die Ausgabe des Alignments in der richtigen Form an (*.fasta).

Der Constraint muss ebenfalls in einer bestimmten Form sein, z.B.:

Material & Methoden

```
>Mytilus_californianus
-----CAACCUGGUUGAUCUGCCAGUAGUCAUAUGCUUG...
>structure
.....((((((...((((())))).(((.((((((((((...
```

Im Fall der Bivalvia wurde die Constraint-Information von der Kalifornischen Miesmuschel (*Mytilus californianus*) genutzt. Diese stammt aus der European Ribosomal Database. Die Sequenz in der Constraint-Datei muss mit dem gleichen Namen im Alignment aller Sequenzen angegeben sein. Der Programmaufruf zur Faltung aller Sequenzen erfolgt dann wie in Abschnitt 3.1.2.4 beschrieben.



File D:\User\Feiste\Daten\# Alignments\Parameterset 1\Bivalvia_185_r

Abbildung 19: Ansicht der unalignierten Sequenzen in ClustalX.

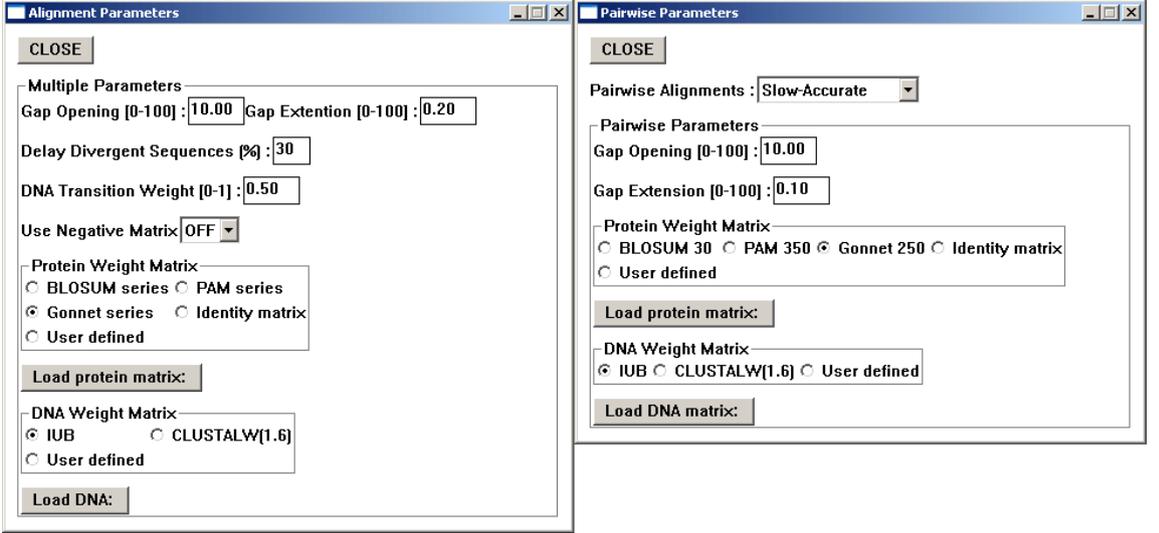


Abbildung 20: Eingabemasken für Alignmentparameter bei ClustalX.

Der Zeitaufwand für die Faltung mit RNAsalsa ist natürlich von der Rechenleistung des ausführenden Systems und von der Größe des Alignments abhängig. Mit einem typischen PC (Dual Core 2GHz, 4 GB RAM) dauert eine Beispielrechnung mit sechs Sequenzen und einer Sequenz mit Constraint etwa zehn Stunden. Für größere Alignments (wie in dem Fall der Bivalvia 130 bis 160 Sequenzen) sind leistungsstärkere Workstations zu empfehlen.

3.2.2.3.1 RNAsalsa Parameter

Die Faltung der Sequenzen wurde mit verschiedenen Parametern durchgeführt. RNAsalsas Standardparameter für die Anwendung des Constraints (siehe Abschnitt 3.1.2.4) wurden für die Bivalvia nicht verändert, da sie sich für das Alignment der verwendeten Sequenzen als sinnvoll erwiesen hatten. In anderen Fällen können sich veränderte Constraint-Parameter aber durchaus als notwendig erweisen.

Der Vergleich der von RNAsalsa generierten Sekundärstrukturen, jeweils einerseits mit externem Alignment (ClustalX) und andererseits mit RNAsalsa-internem Alignment, erbrachte zwar kleine Unterschiede (siehe Abbildung 15), diese gingen aber kaum über die Dimension eines Merkmals (siehe Abschnitt 3.3.2) hinaus. Die grundlegende Form der Sekundärstruktur blieb jedoch erhalten. Die veränderten Merkmale können aber in jedem Einzelfall wiederum weder als „richtig“ noch als „falsch“ eingestuft werden, solange nicht die vollständige Analyse der Merkmale mit bekannten Verwandtschaftsdaten verglichen wurde.

3.2.2.4 Beispiel *Pecten maximus* mit RNAsalsa

Die Faltung nach den oben genannten Parametern ergibt für das Beispiel Pilgermuschel (*Pecten maximus*) die in Abbildung 22 dargestellte Sekundärstruktur. Im Vergleich zur Faltung nach MFE (Abbildung 18) sind deutliche Unterschiede zu erkennen, der auffälligste darunter ist das Fehlen eines zentralen Multi-branch Loops bei NNC. Während bei MFE alle Merkmale von diesem zentralen Loop ausgehen, gibt es bei der

NNC-Struktur fünf größere Merkmalskomplexe (siehe Abschnitt 3.3.2, „Merkmalsdefinition“).

3.2.2.5 Faltung mit RNAfold im Vergleich mit RNAsalsa

Die Sekundärstruktur nach RNAfold unter Nutzung des gleichen Constraints und der gleichen Parameter, wie bei RNAsalsa, erweist sich als sehr ähnlich aber nicht identisch. Abbildung 23 zeigt das Ergebnis der Faltungen mit beiden Programmen am Beispiel von *Abra prismatica*.

Material & Methoden



File D:\User\SFeste\Daten\# Alignments\Parameterset 1\Bivalvia_18S_r

Abbildung 21: Ergebnis der Bearbeitung der Sequenzen aus Abbildung 19.

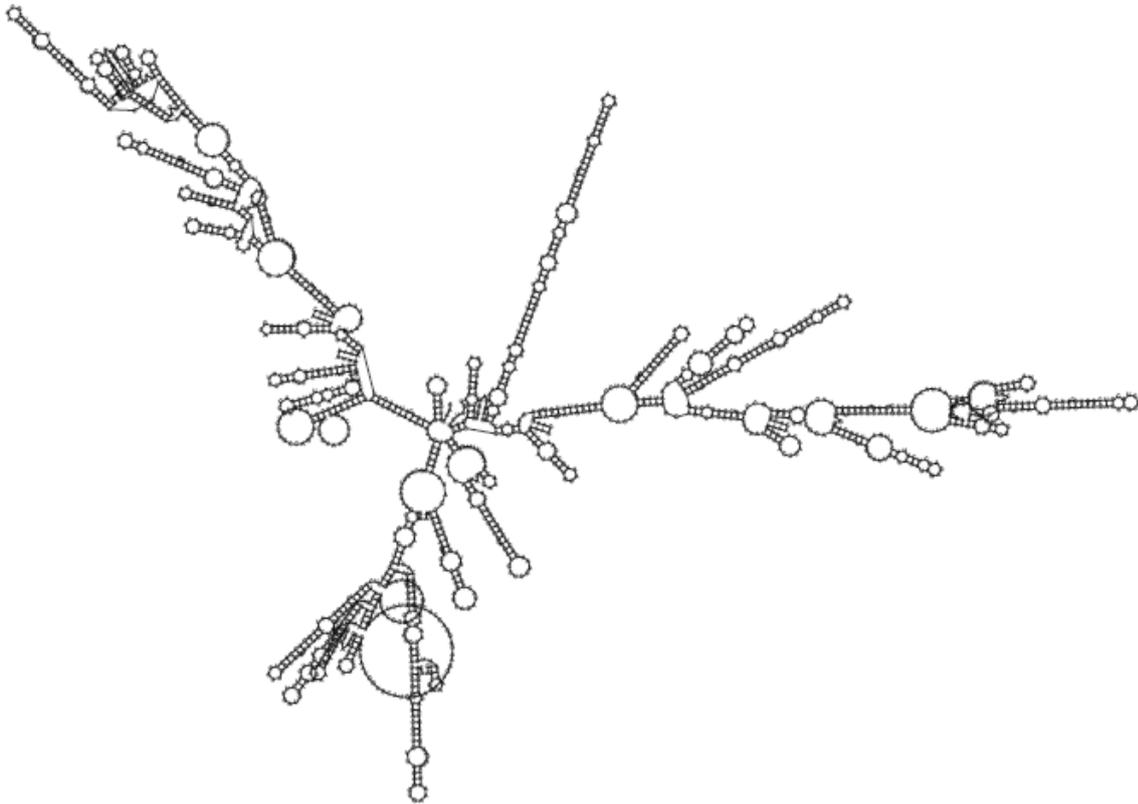
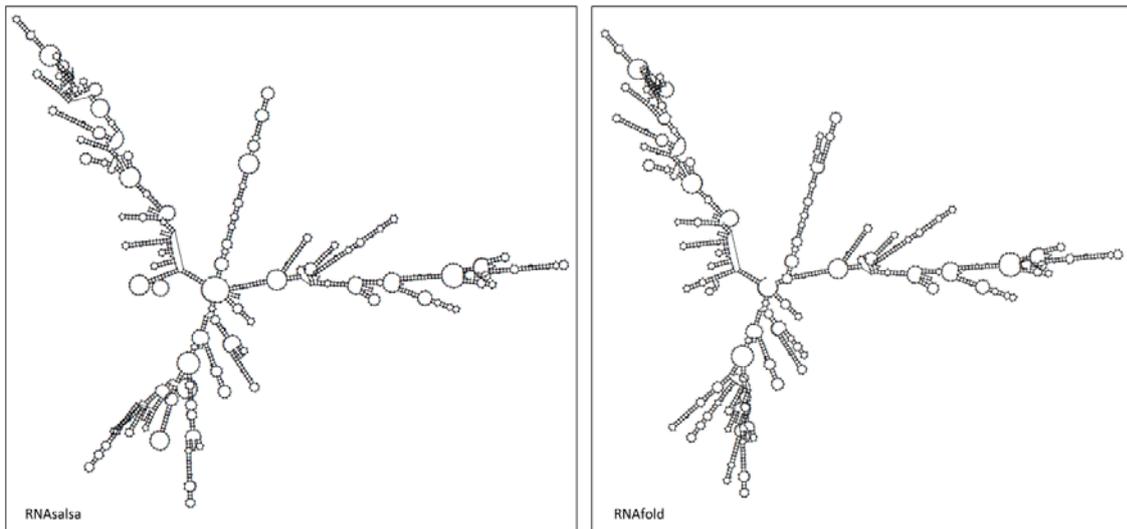


Abbildung 22: NNC-Sekundärstruktur von *Pecten maximus* mit RNAsalsa.



Sekundärstruktur von *Abra prismatica*, nach NNC mit RNAsalsa und RNAfold generiert.

Abbildung 23: Vergleich der NNC-Sekundärstruktur von *Abra prismatica*, gefaltet mit RNAsalsa (links) und RNAfold (rechts).

3.3 Analyse

3.3.1 Sekundärstruktur

Von der DNA ist bekannt, dass sie aus Nucleinsäuren als den vier Grundbausteinen besteht. Zwei dieser Nucleinsäuren sind Purine, Adenin (A) und Guanin (G), und zwei sind Pyrimidine, Thymin (T) und Cytosin (C). Im Zellkern liegen zwei komplementäre DNA-Einzelstränge nebeneinander. Aus den sich gegenüberliegenden Bausteinen entstehen feste Basenpaare zwischen jeweils einem Purin und einem Pyrimidin. Es wurde chemisch nachgewiesen, dass das Verhältnis der Anzahl von Adenin- zu Thymin-Basen in einem DNA-Doppelstrang immer 1 : 1 ist, ebenso das Verhältnis der Anzahl von Guanin- zu Cytosin-Basen. Das Mengenverhältnis von Adenin zu Guanin, beziehungsweise Cytosin zu Thymin ist dagegen von Art zu Art verschieden (*Chargaff, 1950*). Daraus ließ sich schlussfolgern, dass Adenin und Thymin ein Basenpaar bilden (Ausbildung von zwei Wasserstoffbrücken). Das andere Basenpaar, Guanin und Cytosin, ist durch drei Wasserstoffbrücken verbunden. Die daraus resultierende Struktur ist die chemisch stabile Doppelhelix (Abbildung 24).

Die rRNA liegt, wie andere Arten von RNA, zumeist als Einzelstrang vor. Dadurch sind, im Gegensatz zur doppelsträngigen DNA, die Chargaff-Regeln nicht erfüllt. Die Sekundärstruktur der RNA entsteht dadurch, dass der Einzelstrang unter bestimmten physiko-chemischen Bedingungen und unter Bildung intramolekularer Basenpaare auf sich selbst zurückfaltet. Unter der Sekundärstruktur einer RNA-Sequenz versteht man im Wesentlichen eine Liste der gepaarten Basen (*Fontana, 1996*). Des Weiteren ist in der RNA die Base Thymin durch die chemisch ähnliche Base Uracil (U) ersetzt. Die Watson-Crick-Paarungen in der RNA sind dementsprechend A-U und G-C. Diese kommen in der Sekundärstruktur am häufigsten vor.

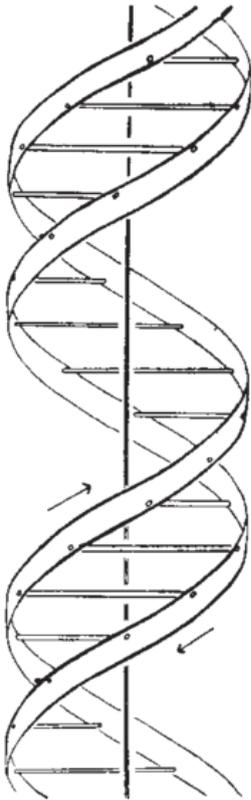


Abbildung 24: Doppelhelixstruktur der DNA aus *Watson & Crick, 1953*. Es wird schematisch dargestellt, wie die zwei Phosphat-Zucker-Ketten durch die horizontalen Bindungen der Basenpaare zusammengehalten werden.

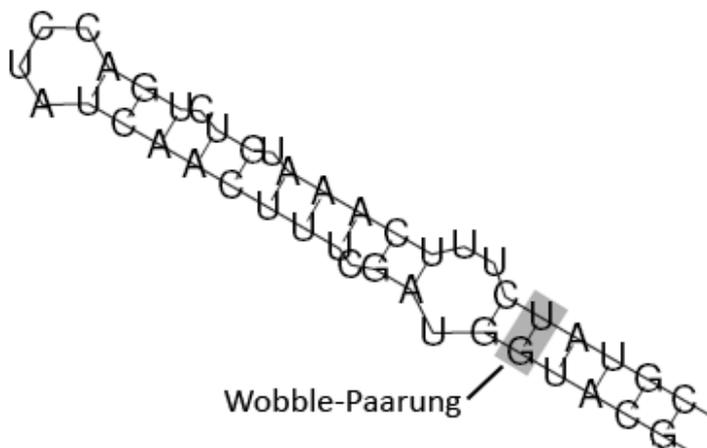


Abbildung 25: Ausschnitt einer Sekundärstruktur nach RNAfold. Grau markiert ist die "Wobble"-Bindung G-U.

Durch die chemischen Wechselwirkungen in der Sekundär- und Tertiärstruktur der RNA sind auch andere Basenpaare als A-U und G-C möglich. Diese bezeichnet man als non-Watson-Crick- oder Wobble-Paarungen (*Crick, 1966*). Dazu zählen unter anderem A-C, A-G oder G-U, wobei G-U als Wobble-Paarung in der RNA am häufigsten erscheint (Abbildung 25). Aus energetischen Gründen ist die Zahl der Wobble-Paare in einer gefalteten RNA möglichst niedrig. Häufig ist die genaue Position von Wobble-Paaren im Strang nicht eindeutig festgelegt, sie können als „Fehlstellen“ des Doppelstrangs „wandern“ und so zu unterschiedlichen Sekundärstrukturen bei gleicher Primärstruktur führen.

Da die entstehenden Sekundärstrukturen für die katalytische Aktivität einer RNA entscheidend sind, sollen diese hier anstelle der Primärstruktur zur Bewertung der Ähnlichkeit zweier RNA-Moleküle herangezogen werden. Die Vermutung, dass 40% bis 60% einer realen RNA-Sequenz eine partielle Helix-Struktur ausweist (*Fontana, 1996*) und die Doppelhelix-Abschnitte durch Einzelstrang-Abschnitte, sogenannte Loops, unterbrochen werden, bedeutet, dass RNA eine morphologisch untersuchbare Molekülstruktur mit unterschiedlichen Merkmalen einnimmt.

Die in dieser Arbeit untersuchten morphologischen Merkmale der Sekundärstruktur, im Folgenden 2D-Merkmale genannt, entstehen durch die abwechselnde Aneinanderreihung von Abschnitten mit Helixstruktur und unverpaarten Einzelsträngen; es liegt somit eine Kombination aus charakteristischen, sich wiederholenden Grundstrukturen zur Analyse vor. Die Elemente dieser übergeordneten Strukturen werden in Helices und Loops (Schleifen) unterteilt, die dann ihrerseits aus Nukleotiden aufgebaut sind.

Nach *Fontana (1996)* reicht es in der Regel aus, Loops durch nur zwei Zahlen zu charakterisieren. Der „Grad“ eines Loops ist die Anzahl der Helices, die von ihm ausgehen und die dann auch als „Stämme“ bezeichnet werden. Die Anzahl n der im Loop enthaltenen Basen sind die „Größe“ des Loops.

Abhängig vom Grad werden Loops in drei qualitativ verschiedene Typen eingeteilt, Hairpin, Bulge und Multi-branched Loops. Die vier grundlegenden topologischen Strukturelemente sind dementsprechend

- ⇒ Helix / Stamm: eine ununterbrochene Helixstruktur aus $1+n$ Basenpaaren
- ⇒ Hairpin Loop (Form einer „Sackgasse“): Die Rückfaltung des RNA-Strangs am Ende eines Stammes / Helix, d.h., ein Loop mit Grad 1 und Größe n .
- ⇒ Bulge Loop / Interior Loop: Ein Abschnitt ungebundener Basen zwischen zwei Helixstrukturen, d.h., ein Loop mit Grad 2 und Größe n . Beim Interior Loop (Form einer „Linse“) sind ungepaarte Basen auf beiden Seiten des Strangs vorhanden, beim Bulge Loop (Form einer „Blase“) sind ungepaarte Basen nur einseitig.
- ⇒ Multi-branched Loop (Form eines „Kreisverkehrs“): Ein Ring ungebundener Basen, von dem mehr als 2 Helices / Stämme ausgehen, d.h., ein Loop mit Grad > 2 und Größe n .

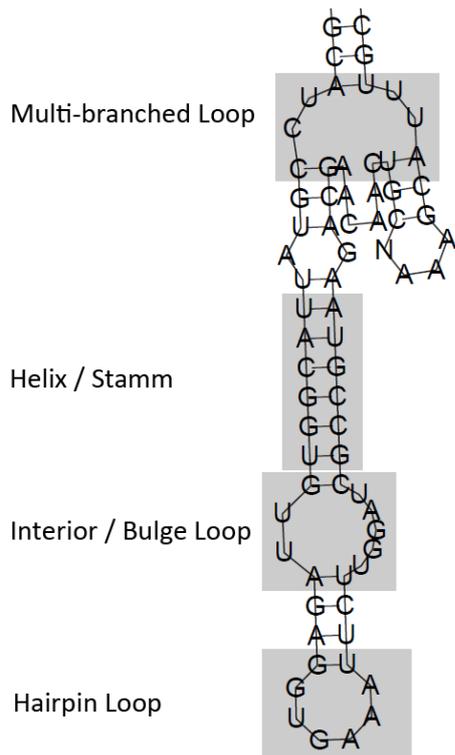


Abbildung 26: Ausschnitt einer Sekundärstruktur nach RNAfold. Grau markiert sind die vier grundlegenden Strukturelemente der Sekundärstruktur.

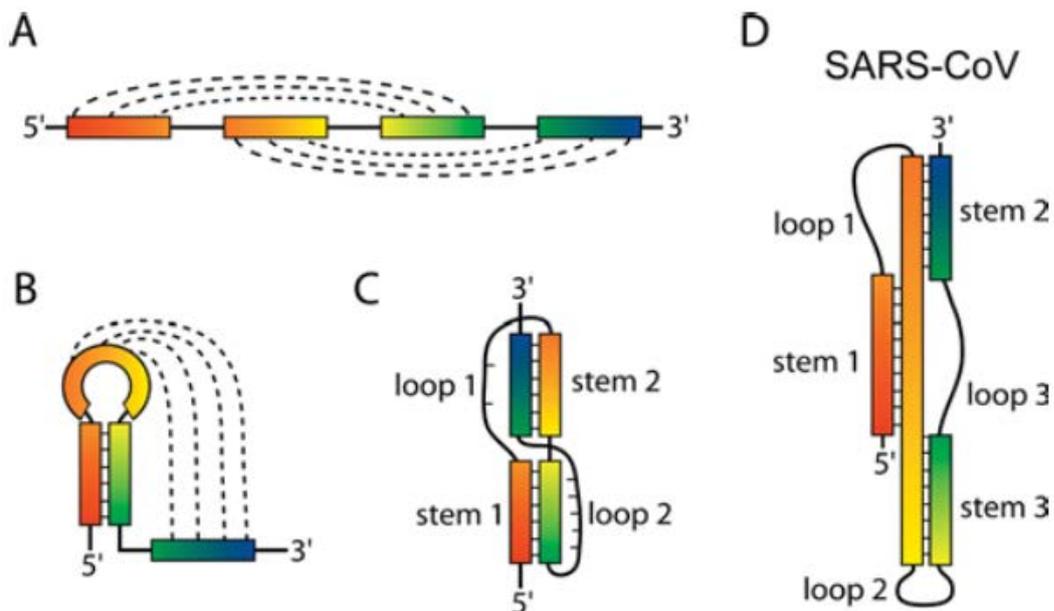


Abbildung 27: Verschiedene Typen von Pseudoknoten aus Staple & Butcher (2005). Gestrichelte Linien stellen Basenpaarungen dar.

Diese vier Strukturelemente sind in Abbildung 26 illustriert. Gegebene Sekundärstrukturen wurden nach Kombinationen aus diesen vier Elementen visuell durchsucht und die diesbezüglichen Unterschiede für die Verwandtschaftsanalyse protokolliert. Die oben genannten 2D-Merkmale mussten aber noch präziser spezifiziert werden, um eine geeignete Vorschrift zu gewinnen, die auf die Fülle der möglichen und komplizierten 2D-Strukturen bei der eigentlichen Auswertung angewendet werden konnte und zu quantitativen Ergebnissen führte, die von Fall zu Fall nicht (oder möglichst wenig) von subjektiven Einzelentscheidungen des jeweiligen Bearbeiters abhingen. Das Herausfinden und Festlegen geeigneter solcher Regeln stellte eine der wesentlichen Herausforderungen dieser Arbeit dar.

3.3.1.1 Pseudoknoten

1982 wurde an einem Mosaikvirus nachgewiesen, dass RNA auch in andere als die in Abschnitt 3.3.1, „Sekundärstruktur“ beschriebenen Formen falten kann (*Rietveld et al., 1982*). Als Pseudoknoten bezeichnet man eine RNA-Sekundärstruktur, bei der die Schleife eines Loops einen Teil des Stamms eines anderen Loops bildet (Abbildung 27).

Derartige Formen kommen in der Natur vor, können aber durch automatische Faltungsalgorithmen kaum vorhergesagt werden. Aus diesem Grund und weil die generierten Sekundärstrukturen eine vereinfachte Darstellung und nicht unbedingt eine naturgetreue Abbildung sind, wurden Pseudoknoten aus der Betrachtung herausgelassen. Im unwahrscheinlichen Fall, dass ein Pseudoknoten doch als Ergebnis des Faltungsalgorithmus auftritt, sollte dieser bei der Auswertung ignoriert werden. Praktisch kam dieser Fall bei den untersuchten Strukturen aber nicht vor.

3.3.2 Merkmalsdefinition

Um eine sinnvolle und eindeutige Klassifikation und Bewertung der Sekundärstrukturen mittels der hier verwendeten 2D-Merkmale zu erreichen, mussten letztere zunächst möglichst streng definiert werden. Damit ein 2D-Merkmal bei der vergleichenden Analyse berücksichtigt wird, sollte es aus mindestens einem Stamm und einem „signifikanten“ Loop bestehen. Aufgrund der Variabilität der Loops wurde zusätzlich

festgelegt, dass sich Loops aus mindestens sechs Nukleotiden zusammensetzen, d.h. mindestens die Größe 6 besitzen müssen, um als signifikant zu gelten.

Beim Vergleich der Sekundärstrukturen stellen sich Merkmalskomplexe heraus, die aus einer Kombination von 2D-Merkmalen bestehen. Ein solcher Merkmalskomplex kann beim Vergleich zweier Sekundärstrukturen einmal vorhanden sein und einmal nicht, er kann identisch oder mit mehr oder weniger großen Unterschieden vorkommen. Solche Merkmale und Merkmalskomplexe galt es zu identifizieren, zu definieren, zu vergleichen und zu bewerten (Abbildung 28).

Es wurden für den Algorithmus folgende Regeln festgelegt:

1. Ein Merkmal und Merkmalskomplex beginnt immer als Stamm / Helix an einem Interior oder Multi-branched Loop. Im Falle der untersuchten MFE-Faltungen beginnen sie meist am Multi-branched Loop mit dem höchsten Grad.
2. Ein Merkmalskomplex K besteht aus M Merkmalen, je ein Merkmal besteht aus $N > 1$ Grundelementen. Ein Merkmal kann nicht nur aus einem Stamm bestehen.
3. Merkmale gelten als gleich, wenn sie die gleiche Anzahl Basen n enthalten und die relativen Positionen ihrer Grundelemente gleich sind.

Da nur 2D-Merkmale zur Klassifikation der Spezies und ihrer Verwandtschaft verwendet werden sollten, sind Punktmutationen der Primärstruktur, die keine Auswirkung auf die Sekundärstruktur haben (d.h. keine Veränderung der Anzahl von Stämmen und/oder signifikanten Loops zur Folge haben), für die Bewertung irrelevant.

In der Merkmalsmatrix wurde jedem Merkmal und jedem Merkmalskomplex eine eindeutige Kennung zugewiesen (siehe Abschnitt 3.3.2.1, „MFE: Merkmalsbeispiel“).

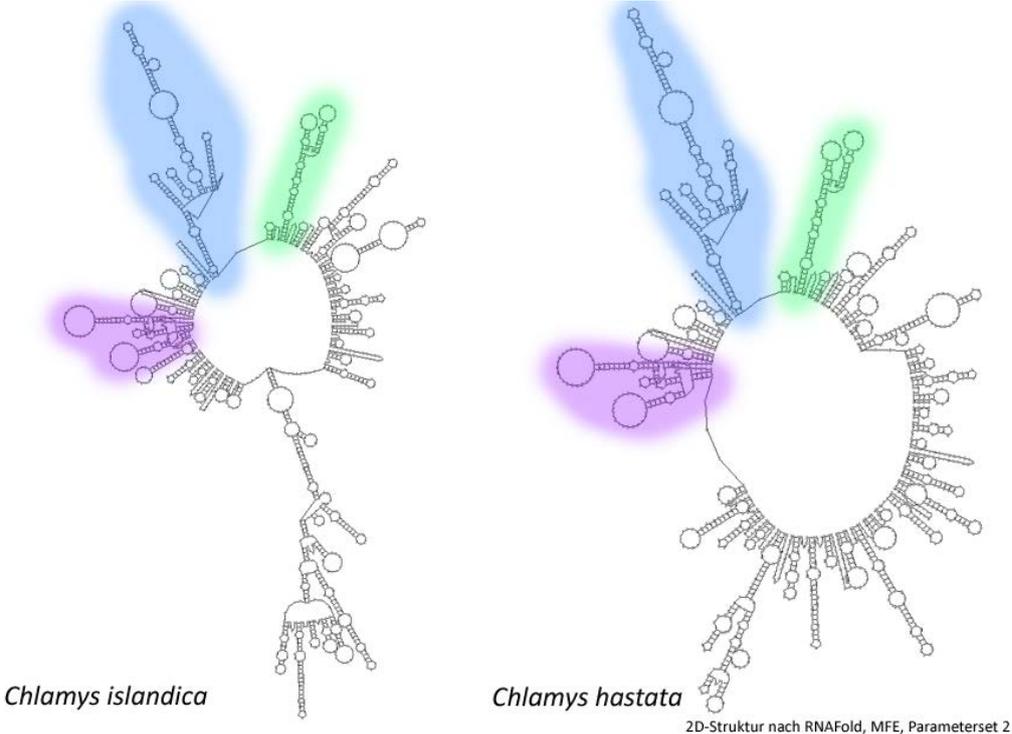


Abbildung 28: Berechnete Sekundärstrukturen von *Chlamys islandica* und *Ch. hastata* nach RNAfold, MFE. Farblich markiert sind drei übereinstimmende Merkmalskomplexe.

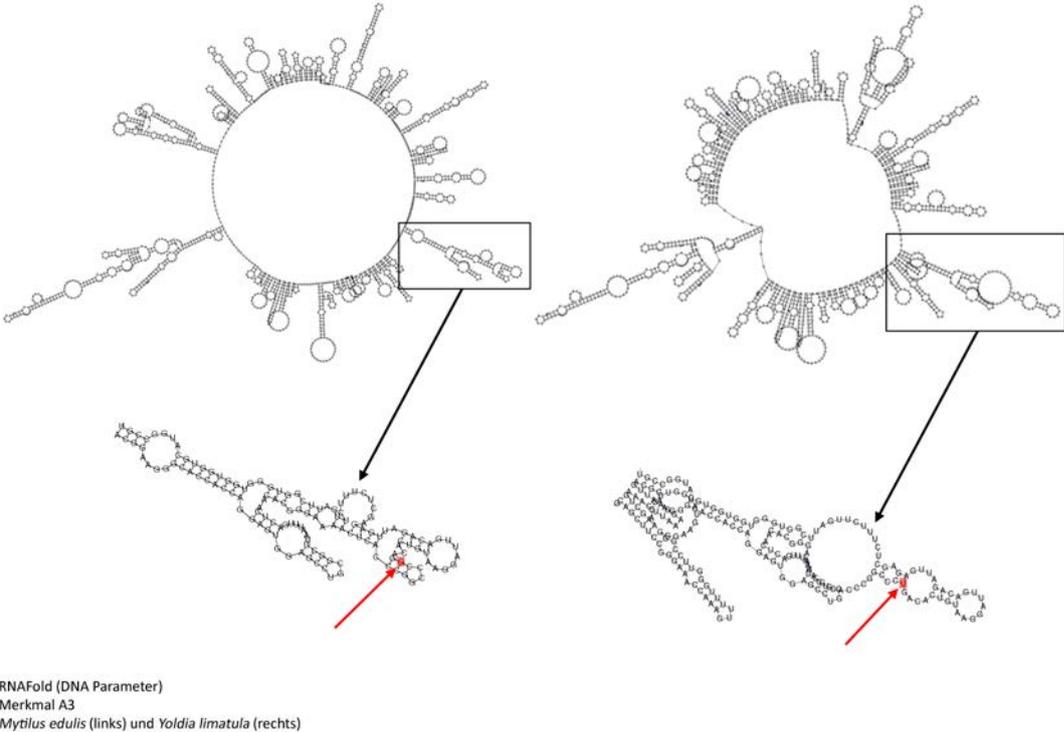


Abbildung 29: Herausgestellt ist jeweils ein Merkmalskomplex mit unterschiedlich ausgeprägten Merkmalen (Sekundärstruktur von *Mytilus edulis* und *Yoldia limatula* nach RNAfold), rot markiert ist die ursächliche Punktmutation.

3.3.2.1 MFE: Merkmalsbeispiel

Die geeignete Merkmalsdefinition bei MFE gestaltet sich grundsätzlich schwierig, wie an den folgenden Beispielen gezeigt werden soll. In beiden Fällen handelt es sich um (in dieser Arbeit so genannte) Merkmalskomplexe, die wiederum aus mehreren Merkmalen bestehen. Allgemein können sowohl Merkmale als auch Merkmalskomplexe unterschiedliche Anzahlen von Basen haben.

Merkmale und Merkmalskomplexe können identisch, in unterschiedlichen Ausprägungen oder gar nicht vorkommen. In Abbildung 29 ist ein Merkmalskomplex herausgestellt, der bis zum ersten Multibranch-Loop identisch ist, danach aber in einem Merkmal abweicht. Die Anzahl der beteiligten Basen ist gleich (59). Wenn man die einzelnen Basen in der Struktur vergleicht, kann man erkennen, dass eine lokale Punktmutation ($G \leftrightarrow U$) die ausgedehnte Strukturveränderung ausgelöst hat.

Es ist zu diskutieren, inwiefern diese Merkmalsveränderungen signifikant für phylogenetische Stammbäume sind.

3.3.2.2 NNC: Merkmalsbeispiel

Die Merkmalsdefinition bei NNC ist im Gegensatz zu MFE relativ einfach. In Abbildung 30 ist ein Merkmal (6) aus einem Merkmalskomplex (B) herausgestellt. Obwohl drei Punktmutationen und die Insertion einer Base zu erkennen sind, bleibt das Merkmal (und der Merkmalskomplex) größtenteils erhalten.

Die Sekundärstrukturen von NNC zeichnen sich durch eine gewisse Unempfindlichkeit gegenüber Punktmutationen und kleineren InDels aus.

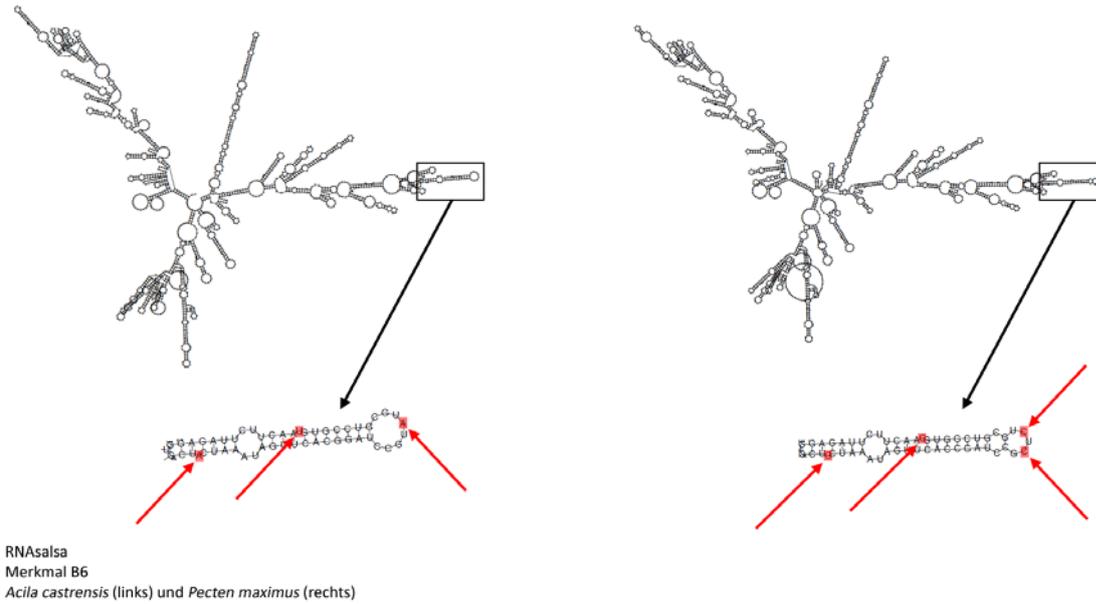


Abbildung 30: Herausgestellt ist jeweils ein Merkmal (B6) mit unterschiedlichen Grundelementen (Sekundärstruktur von *Acila castrensis* und *Pecten maximus* nach RNAalsa), rot markiert sind Punktmutationen, bzw. eine Insertion von einer Base.

3.3.3 Merkmalsmatrix

Nach erfolgreicher Ausführung der Faltung aller gegebenen Sequenzen wurde eine Gesamtliste aller relevanten Merkmale angelegt, typischerweise zwischen $M = 100$ und $M = 500$, herausgesucht aus der Gesamtmenge von $S = 160$ der zu vergleichenden Bivalvia-Sequenzen. Hinzu kamen Sequenzen von anderen Mollusca als sogenannte „Outgroup“, um festzustellen, welche der gefundenen Merkmale ggf. plesiomorph für Bivalvia sind. Während für die Verwandtschaft der Bivalvia untereinander die apomorphen Merkmale ausschlaggebend sind, dienen die plesiomorphen Merkmale zur Einschätzung der Verwandtschaft zwischen Bivalvia und anderen Mollusken. Insbesondere kann die Wurzel des Stammbaums der Bivalvia aus den plesiomorphen Merkmalen bestimmt werden. Der Unterschied zwischen plesiomorphen und apomorphen Merkmalen ergibt sich erst im Ergebnis der Merkmalsanalyse.

Von den verfügbaren Bivalvia-Sequenzen wurden 13 ausgeschlossen. Davon betroffen waren drei bereits vor der Analyse, weil ihre Sequenzen zu kurz waren. Die restlichen wurden nach der Analyse ausgeschlossen, weil sie vermutlich falsch bestimmt oder zugeordnet waren, oder aber zu stark abweichende, aberrante Strukturen aufwiesen, wie z.B. bei *Tridacna*.

Für jede 2D-Struktur aus der Menge wurde mit wahr oder falsch (1 oder 0) in einer ($S \times M$) - großen Merkmalsmatrix notiert, ob das jeweilige Merkmal der Gesamtliste bei der jeweiligen Struktur vorhanden war oder nicht. Im Ergebnis konnte für jedes Paar von Sequenzen die Zahl der Übereinstimmungen oder Nichtübereinstimmungen abgezählt und daraus eine ($S \times S$) - große Ähnlichkeits- oder Abstandsmatrix gebildet werden (Abbildung 31).

Dieses Verfahren wurde mit dem identischen Satz von Sequenzen für verschiedene Faltungsalgorithmen und -parameter wiederholt. Insgesamt wurden so mehrere Zehntausend einzelne Merkmale an ausgedruckten 2D-Strukturen identifiziert, manuell markiert, ausgezählt und in Listen vermerkt. Die optische Analyse am Bildschirm erwies sich bei dieser Aufgabe als wenig praktikabel.

3.3.3.1 Merkmalsanalyse

Die Analyse erfolgte für MFE und NNC ähnlich: Es wurden zunächst alle Merkmalskomplexe identifiziert, benannt und jeweils in einer Matrix erfasst. Anschließend wurden die Merkmale, aus denen sich die Komplexe zusammensetzen, in der Matrix erfasst.

Es gibt bei dem Vergleich und der Analyse unterschiedliche, aber eng miteinander verknüpfte Probleme zu beachten: die Homologisierung der Merkmale, die Ähnlichkeit homologer Merkmale und die Gewichtung der Merkmale und ggf. ihre Veränderung.

Das erste Problem ist das der Homologisierung, d.h. der Suche nach der wahrscheinlichsten gemeinsamen Wurzel von später differenzierten Merkmalskomplexen. In der klassischen Morphologie gibt es die sogenannten homologen Merkmale, die sich von einem gemeinsamen, ursprünglichen Zustand ableiten. Ein bekanntes Beispiel sind die Kiemenbögen der Wirbeltiere. Im Embryo entwickeln sie sich als Ausstülpungen des Mesoderms, ursprünglich erscheinen sie als der Kiemendarm der Chordatiere. Die Anlage kann durch die Ontogenese und die Jungstadien bis zum erwachsenen Organismus verfolgt werden. In Haien entwickeln sich die namensgebenden Kiemenbögen aus der Anlage, in Knochenfischen entwickeln sich Teile von Schädel und Kiefer daraus. Im Menschen wiederum sind diese Knochen nach innen „gewandert“ und bilden dort u.a. die Gehörknöchelchen. Ein anderes bekanntes, relativ einfaches Beispiel für eine Homologisierung ist die Wobble-Hypothese von Crick, bei der nach bestimmten Regeln die wahrscheinlichste ursprüngliche Form des genetischen Codes gesucht wird, aus der die moderne Tabelle durch sukzessive Differenzierung entstehen konnte.

In Analogie zur klassischen Morphologie stellte sich hier die Frage, auf welche Weise die Merkmale der Molekularen Morphologie homologisiert werden können. Dafür mussten geeignete formale Regeln aufgestellt und später systematisch angewendet werden. Im Verlauf der Arbeit erwies es sich als sinnvoll, für MFE und NNC verschiedene Regeln zur Homologisierung zu formulieren.

Homologisierung bei MFE: Falls bei den Sekundärstrukturen zweier Arten jeweils ein Merkmalskomplex den gleichen „primären“ Stamm, bestehend aus mindestens fünf gleichen zusammenhängenden Basenpaaren, hat, wird dieser Merkmalskomplex als homolog angesehen. Es wird unter dieser Bedingung davon ausgegangen, dass ein gemeinsamer Ursprung der beiden Komplexe vorliegt. Eine hundertprozentige manuell-visuelle Zuordnung homologer Strukturen ist jedoch vom zeitlichen Aufwand her nicht praktikabel. Falls kein „ähnlicher“ Primärstamm ausgemacht werden kann, ist es nicht möglich, eine Homologisierung vorzunehmen.

Homologisierung bei NNC: Die Zuordnung homologer Strukturen ist bei NNC deutlich einfacher als bei MFE. Durch die relative Position eines Merkmals innerhalb eines Merkmalskomplexes sind verschiedene Ausprägungen / Versionen dieses Merkmals leicht zu identifizieren. In vielen Fällen ist auch die ursächliche Mutation zu erkennen. Lediglich bei stark abweichenden, durch große InDels ausgelösten Strukturen ist eine Homologisierung nicht mehr möglich.

Im Folgenden werden Merkmale, die als homolog definiert wurden, auch als Merkmalsversionen oder –ausprägungen bezeichnet.

Nach der Identifizierung der homologen Merkmale und Merkmalskomplexe wurde anschließend deren Ähnlichkeit untersucht unter Anwendung einer zweiten Regel, die drei Situationen unterscheidet:

- a) Der einfachste Fall liegt vor, wenn die Merkmalskomplexe in allen Merkmalen übereinstimmen; (Punkt-) Mutationen ohne Auswirkungen auf die Sekundärstruktur werden nicht berücksichtigt. In diesem Fall werden Merkmalskomplex und alle Merkmale als „vorhanden“ eingestuft.
- b) Vollständig unterschiedliche Merkmalskomplexe, das heißt solche, die in den jeweils anderen Sekundärstrukturen überhaupt nicht vorhanden sind, werden zusammen mit allen dazugehörigen Merkmalen als „nicht vorhanden“ gezählt. Es konnte keine homologe Struktur gefunden werden.
- c) Homologe Merkmalskomplexe, bei denen sich ein oder mehrere Merkmale

unterscheiden, werden als „vorhanden“ gezählt und auf ihre einzelnen Merkmalsübereinstimmungen untersucht.

Für die Bewertung der Merkmalsvarianz wurde, soweit möglich, untersucht, wodurch Änderungen bei homologen Merkmalen verursacht wurden. Es wurde der am häufigsten auftretende Zustand als Referenzpunkt genommen und die Veränderung relativ zu dieser Ausprägung festgestellt. Die Ursache einer Veränderung wurde anhand der Anzahl der veränderten/zusätzlichen/fehlenden Basen bewertet, die eigentliche Veränderung über die veränderte Komplexität. Als Veränderungen gelten nur Mutationen, die eine andere Ausprägung von Merkmalen zur Folge haben. Diese Analyse ist aufgrund der Anzahl der homologen Merkmale nur bei NNC sinnvoll.

Die in dieser Arbeit verwendete formale, eindeutige Bezeichnung eines Merkmals(-komplexes) setzt sich aus Buchstaben und Zahlen zusammen. Merkmalskomplexe werden mit einem oder mehreren Buchstaben gekennzeichnet (ein Buchstabe bei NNC, da weniger Merkmalskomplexe vorliegen; mehrere Buchstaben bei MFE, da deutlich mehr als 26 Merkmalskomplexe auftreten). An den Komplexbuchstaben schließt sich eine zweistellige Positions-Nummer für die Merkmale innerhalb eines Komplexes an. Merkmale werden im Idealfall durchlaufend im Uhrzeigersinn nummeriert.

3.3.3.2 Stammbaumberechnung

Die hypothetischen Phylogenien der Bivalvia und ihrer Großgruppen nach MFE und NNC wurden mit verschiedenen Ansätzen untersucht. Zur Berechnung der möglichen Stammbäume diente das Programm PAUP 4.0 beta 10 Win. Alle Analysen wurden nach dem Prinzip von Maximum Parsimony durchgeführt.

Die einfachste Herangehensweise war die Analyse aller Merkmale und Merkmalsversionen als sogenannte single state characters, mit gleicher Gewichtung. Diese setzt alle identifizierten Merkmale ungeachtet ihrer Homologie oder Größe gleich.

Darauf folgte die Analyse der gleichen single state characters mit differenzierter Gewichtung, basierend auf der Merkmalsgröße oder Komplexität. Dieses Vorgehen lässt

wiederum die Homologie außer Acht, bezieht aber die Größe der gemeinsamen Merkmale ein. Wenige große gemeinsame Merkmale beeinflussen den Stammbaum stärker als viele kleine Merkmale.

Die Analyse der Matrix mit Merkmalen, die als homolog definiert wurden, erfolgte mit sogenannten multi state characters. Diese Methode bezieht zwar die Homologie der Merkmale mit in die Untersuchung ein, da aber die jeweilige „Richtung“ der Entwicklung von „ursprünglich“ zu „neu entstanden“ nicht festgestellt werden kann, erfolgte die Berechnung mit sogenannten ungeordneten Merkmalszuständen. Auch die Analyse mit multi state characters wurde mit undifferenzierter und differenzierter Gewichtung durchgeführt, wobei als Gewichtung die Größe der vermutlich ursprünglichsten Merkmalsversion gewählt wurde.

Bei jeder Berechnung wurde soweit möglich mittels Bootstrap-Analyse (*Felsenstein, 1985*) deren Verlässlichkeit untersucht.

3.3.3.3 Gewichtung

Die Gewichtung wurde mit zwei verschiedenen Maßen getestet - über die Anzahl der Basen und über die Komplexität des Merkmals. Die „Größe“ eines Merkmals, d.h. die Anzahl seiner Basen, dient der Vergleichbarkeit nicht homologer Merkmalskomplexe. „Große“ Merkmalskomplexe mit vielen Basen und Merkmalen sollen höher gewichtet werden, als Merkmalskomplexe, die lediglich aus einem Merkmal bestehen. Die Größenbestimmung für single-state Merkmale erfolgte direkt aus der Anzahl der Basen. Für Merkmale mit verschiedenen Zuständen („multi-state characters“) (siehe Abbildung 30) wurde das am häufigsten vorkommende Merkmal als Referenz festgelegt.

Die Komplexität eines Merkmals wurde wie folgt festgelegt:

- ⇒ Merkmalskomplex K mit M Merkmalen => M ist das Maß der Komplexität von K
- ⇒ Merkmal M mit N Grundelementen => N ist das Maß der Komplexität von M
- ⇒ Grundelement N mit n Basen => n ist das Maß der Komplexität von N

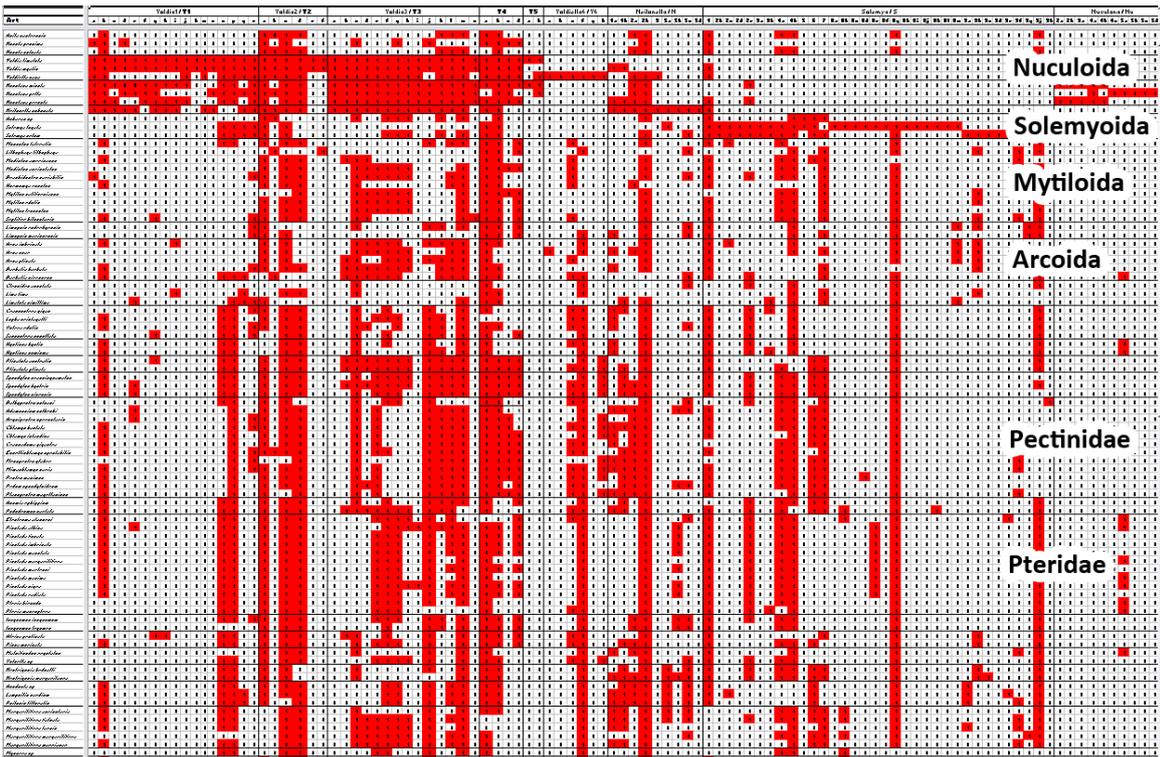


Abbildung 31: Ausschnitt aus der Rohfassung der Merkmalsmatrix von der MFE-Faltung mit RNAfold. Untersuchte Arten sind in den Zeilen, Merkmale in den Spalten. Rot markiert sind "vorhandene" Merkmale.

4. Ergebnisse

4.1 Methode „Molekulare Morphologie“

Während die Analyse, wie sie im Abschnitt 3.3 beschrieben wurde, bei der Faltung nach MFE aufgrund der extrem großen Anzahl von unterschiedlichen Merkmalen nur mit erheblichem Zeitaufwand möglich war, ist NNC in dieser Hinsicht wesentlich effektiver.

Für beide Methoden muss angemerkt werden, dass immer eine geringe Anzahl (ca. 5-10) „Ausreißer“ entstand, deren 2D-Strukturen entweder keine oder nur sehr wenige Gemeinsamkeiten mit denen ihren nah verwandten Arten aufweisen. Da die Einbeziehung der Ausreißer die Analyse deutlich erschwert, wurden diese Arten in den meisten Fällen von der Analyse ausgenommen. Die Gründe für diese starken Abweichungen wurden nicht explizit erforscht, deuten aber wiederum auf eine hohe Empfindlichkeit der Verfahren gegen kleine Veränderungen der Eingangsdaten hin, wie etwa zu kurze oder zu lange Sequenzen oder einzelne kritische Sequenzierungsfehler.

4.1.1 Minimum Free Energy (MFE)

Mit Hilfe der RNAfold-Software des Vienna Servers wurden 2D-Strukturen generiert, die dem Prinzip der „Minimum Free Energy“ gehorchen. Insgesamt wurden die 18S rDNA-Sequenzen von 133 Arten der Bivalvia, Schwerpunkt Pteriomorpha, analysiert. Die Anzahl der Basenpaare pro Sequenz variierte zwischen 1600 und 2000.

Nach dem „Falten“ erfolgte die Untersuchung der Strukturen auf Ähnlichkeiten bzw. Unterschiede. Die berechneten Strukturen sind jeweils ringförmig, mit einzelnen nach außen weisenden, unterschiedlich komplex ausgeprägten Doppelsträngen nach außen. Es ergaben sich im Schnitt 30 Merkmale pro Art (siehe Abbildung 28 und Abbildung 29).

Es wurden mehr als 500 verschiedene Merkmale definiert und für jede der 133 Strukturen individuell ausgewertet. Jedes Merkmal ist ein visuell markantes, qualitativ

abgegrenztes Strukturelement, das von einem zusammenhängenden Sequenzabschnitt von 6 bis 150 Basen gebildet wird (siehe Abschnitt 3.3.2, „Merkmalsdefinition“).

Merkmalskomplexe, soweit vorhanden, enthielten meist drei bis vier Merkmale. Sie waren jedoch nie als konservativ, das heißt als mit einem Vorkommen bei in mindestens 50% der Arten, anzusehen.

Einen Ausschnitt aus der ermittelten Merkmalsmatrix zeigt Abbildung 31. Einerseits sind signifikant erscheinende Merkmalsblöcke zu erkennen, andererseits ist ein gewisses „Rauschen“ nicht zu übersehen.

4.1.1.1 Statistische Auswertung

Weil gerade bei der Faltung nach MFE eine sehr große Anzahl unterschiedlicher Merkmale auszuwerten ist, kann eine statistische Zusammenstellung hilfreich sein, um einen Überblick über die Verteilung und Ausprägung der Merkmale zu bekommen. Dabei wurden die Größen der Merkmale und ihre Häufigkeiten erfasst.

Bei der Faltung nach MFE ist die Mehrzahl der Basen, circa 80%, in Doppelbindungen gebunden. „Freie“ Abschnitte mit ungebundenen Basen gibt es fast nur in den Loops, die die klassifizierten Merkmale ausmachen. Trotz dieser „fest“ erscheinenden Strukturen sind die Faltungen sehr instabil und von Art zu Art sehr variabel.

Die Größen der Merkmale überschreiten kaum die Anzahl von 100 Basen, 70% der definierten Merkmale haben weniger als 30 Basen; am häufigsten treten Merkmale mit 20 bis 50 Basen auf. Bei MFE sind die Merkmale selten so groß und stabil, dass man sie als Merkmalskomplex bestehend aus mehreren elementaren Merkmalen definieren kann. Meist liegt der Fall so, dass der gesamte Merkmalskomplex nicht mehr vorhanden ist, sobald sich auch nur ein darin enthaltenes Merkmal ändert. Der Großteil der identifizierten Strukturen sind Merkmale, die aus wenigen Grundelementen bestehen.

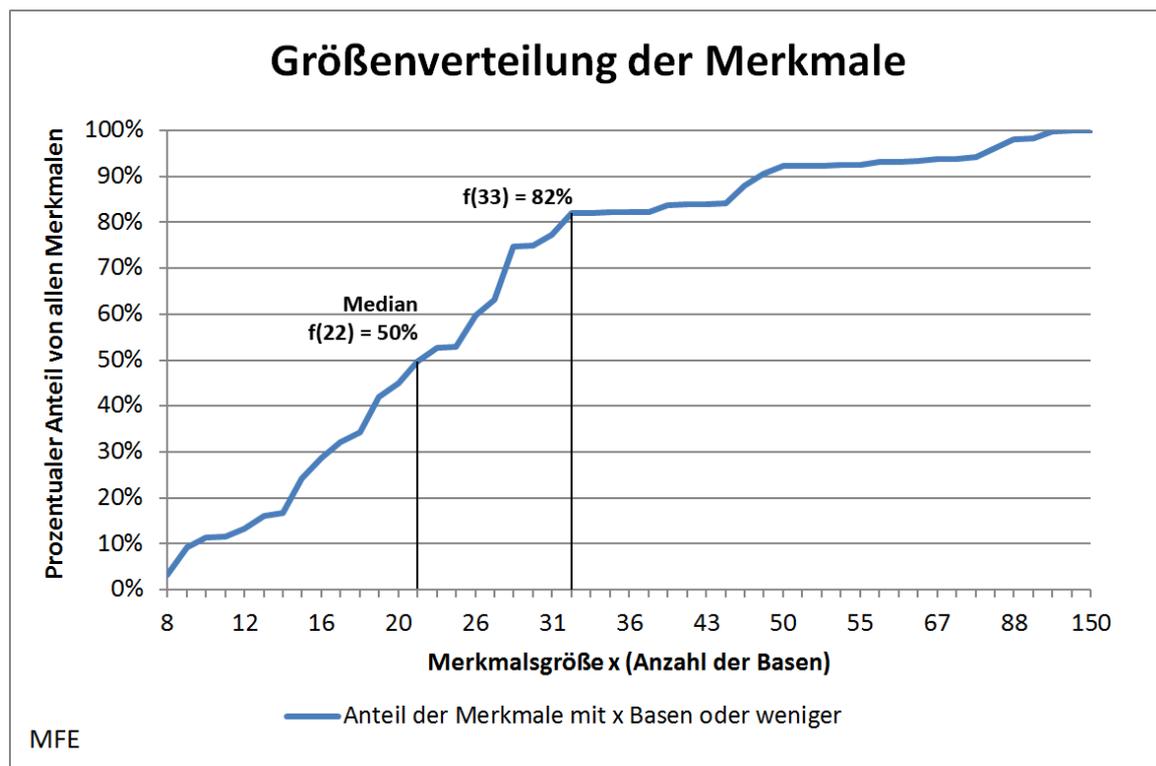


Abbildung 32: Größenverteilung der Merkmale in einem Satz MFE-Faltungen.

Auf der x-Achse ist die Merkmalsgröße x aufgetragen, als Verteilungsfunktion auf der y-Achse der Prozentsatz derjenigen Merkmale, die kleiner oder gleich der Größe x sind.

Kleine und mittelgroße Merkmale kommen am häufigsten vor. Große oder sehr große Merkmale, die aus mehr als 33 Basen bestehen, stellen nur etwa 20%.

Die Größenverteilung, dargestellt in Abbildung 32, stellt sich zunächst als näherungsweise linearer Anstieg dar. Der Median liegt bei Merkmalen mit 22 Basen. Ab einem Wert von 33 Basen knickt die Gerade ab – große oder sehr große Merkmale sind selten (weniger als 20%) und können als „Ausreißer“ angesehen werden. Gerade diese seltenen, großen und sehr großen „Merkmale“ sind die einzigen, die man als Merkmalskomplexe einordnen kann.

Es ist deutlich zu erkennen, dass viele „kleine“, variable Merkmale einen großen Teil des Datensatzes ausmachen. Dies weist auf eine erhebliche Instabilität beim Falten der Sequenzen hin. Merkmale gelten hier als „variabel“ wenn an der gleichen relativen Position in der Sekundärstruktur bei mehr als 50% der untersuchten Arten an Stelle des

betrachteten Merkmals ein oder mehrere andere Merkmale ausgeprägt sind, das heißt, es handelt sich um Merkmale, die bei weniger als 50% der Arten vorkommen.

Von 25 bis 46 Merkmalen pro Art sind weniger als 10% der Merkmale konservativ. Wenn man im Gegensatz zu „variabel“ hier definiert, dass „konservativ“ heißen möge, dass ein Merkmal bei mindestens 50% der Arten vorkommt, sollten bei untersuchten 130 Arten für ein entsprechendes „konservatives“ Merkmal 65 Merkmalsträger vorhanden sein. Es wurden aber nur 19 Merkmale gefunden, die bei 65 oder mehr Arten vorkommen.

Die vorher vermutete Stabilität aufgrund der hohen Anzahl an Doppelbindungen, hat sich im Ergebnis der Analyse nicht bestätigt.

4.1.2 Near Natural mit Constraint (NNC)

Im Gegensatz zur MFE-Methode, bei der sich bei den Sekundärstrukturen fast nur Merkmale definieren ließen, gibt es bei NNC durchgehende Merkmalskomplexe.

Die Analyse der Merkmale der NNC-Sekundärstrukturen ergab in dem hier untersuchten Fall fünf Merkmalskomplexe, die sich aus jeweils bis zu 31 definierten Merkmalen zusammensetzen. Alle fünf Komplexe traten in jeweils 98% der Sekundärstrukturen auf. In den meisten Fällen unterschieden sie sich um bis zu fünf Merkmale. Tabelle 4 fasst die weiteren numerischen Ergebnisse zusammen.

Pro Merkmalskomplex treten im Schnitt 10 der definierten Merkmale auf. Abbildung 33 zeigt die fünf Merkmalskomplexe wie sie bei *Acila castrensis* (Nuculidae) ausgeprägt sind. Die Abbildungen 34 bis 38 zeigen verschiedene Ausprägungen der Merkmalskomplexe A-E.

Tabelle 4: Numerische Zusammenfassung der Merkmalsanalyse.

Merkmalskomplex	Größe (Anzahl Basen)	Durchschnittliche Anzahl Merkmale	Anzahl Basen pro Merkmal	Anzahl homologer Merkmale
A	144	10	6-18	31
B	311	9	13-70	9
C	73	4	6-24	7
D	494	10	10-70	12
E	431	14	8-71	15

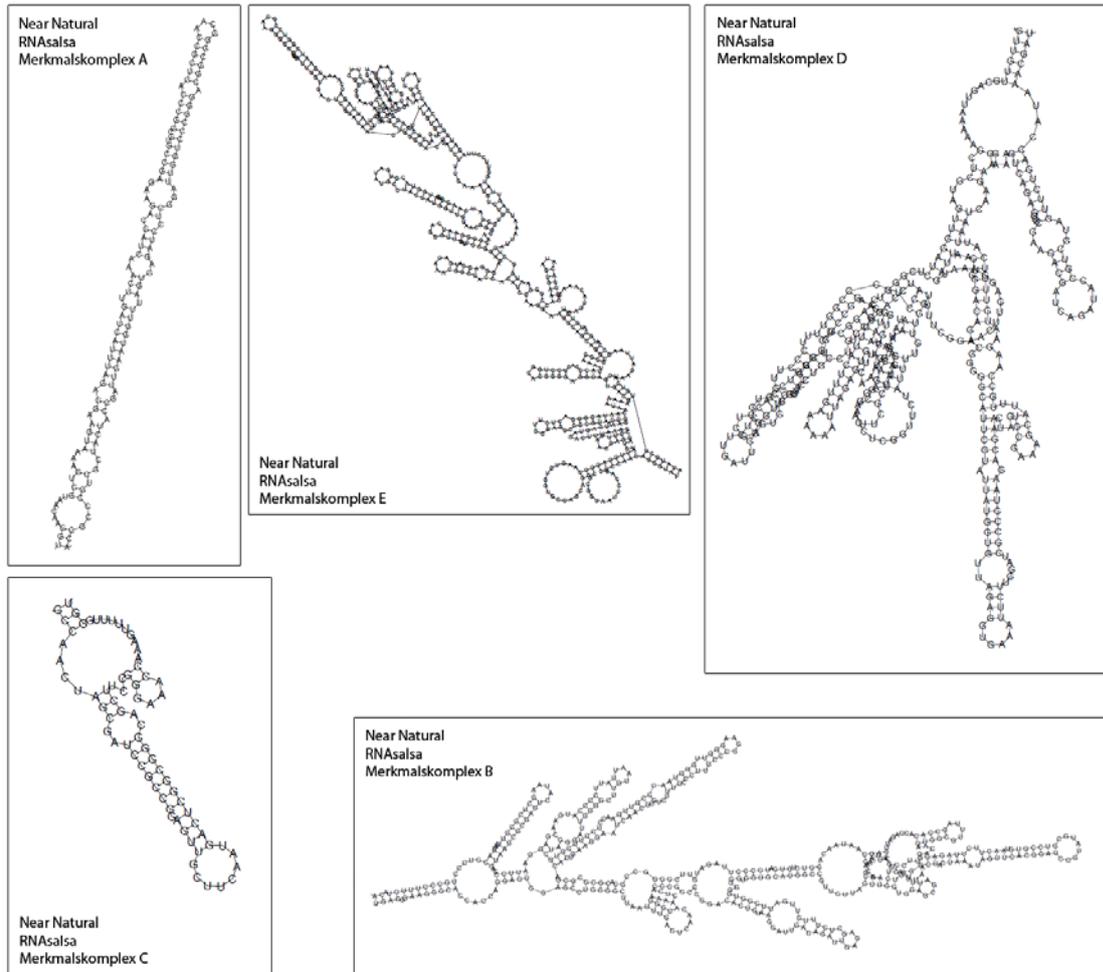


Abbildung 33: Die fünf häufigsten gefundenen Merkmalskomplexe in den Sekundärstrukturen nach NNC, mit RNAsalsa gefaltet. Merkmalskomplexe im Uhrzeigersinn: A, E, D, C, B.

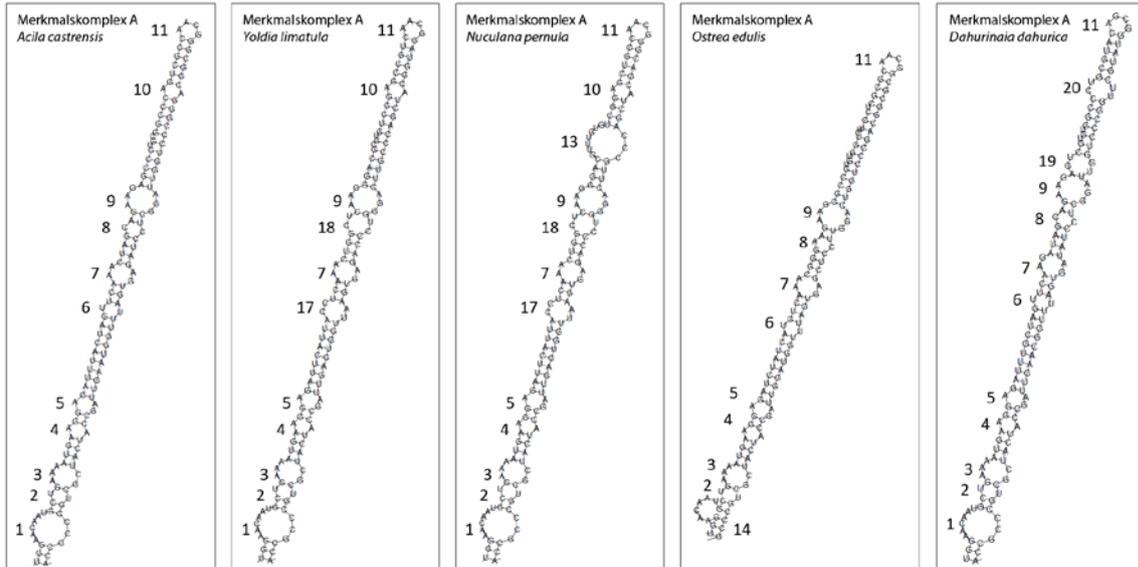


Abbildung 34: Near Natural (NNC) nach RNAsalsa. Verschiedene Ausprägungen von Merkmalskomplexen am Beispielkomplex A mit hier insgesamt 17 definierten Merkmalen in: v.l.n.r. *Acila castrensis*, *Yoldia limatula*, *Nuculana pernula*, *Ostrea edulis*, *Dahurinaia dahurica*.

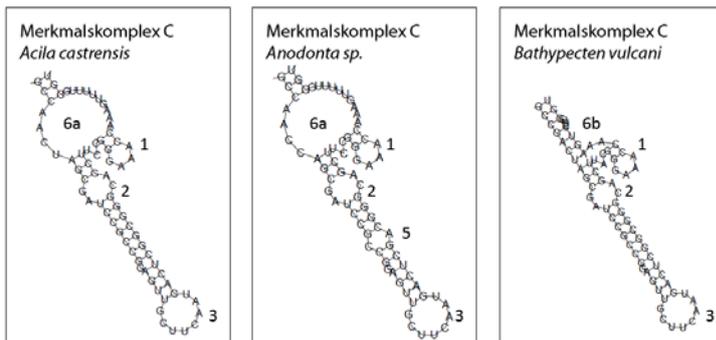


Abbildung 35: Near Natural (NNC) RNAsalsa. Verschiedene Ausprägungen von Merkmalskomplexen am Beispielkomplex C mit fünf Merkmalen, z.T. mit verschiedenen Ausprägungen in: v.l.n.r. *Acila castrensis*, *Anodonta sp.* und *Bathypecten vulcani*.

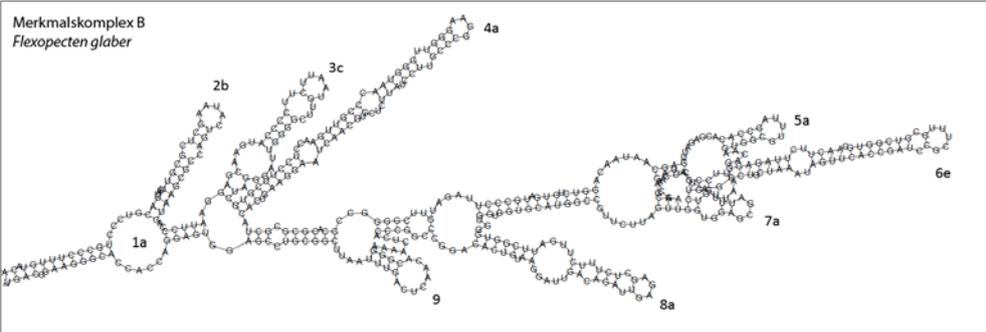
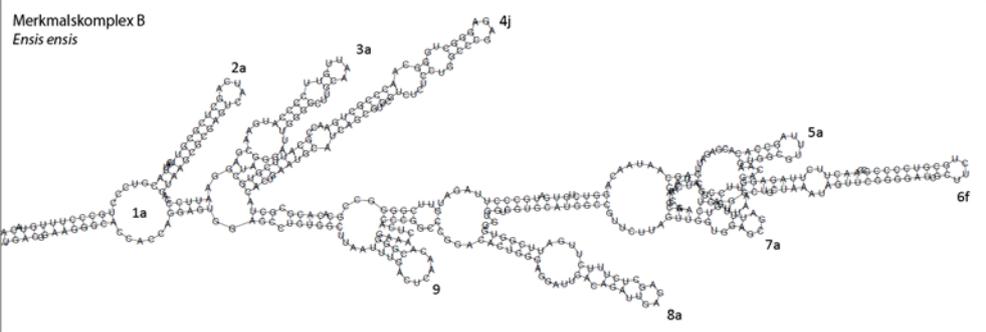
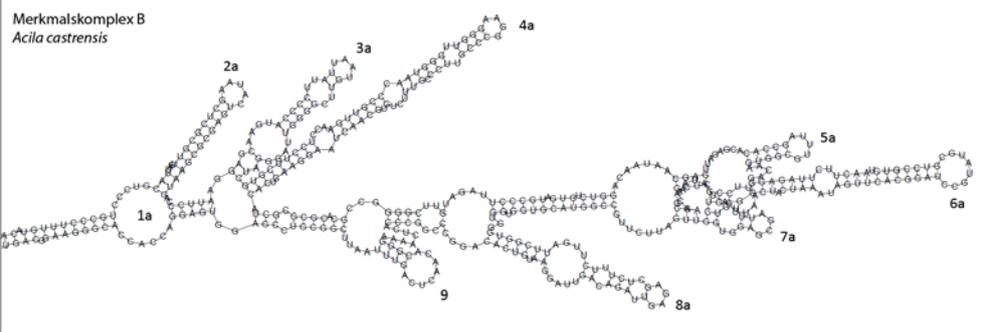


Abbildung 36: Near Natural (NNC) RNAsalsa. Verschiedene Ausprägungen von Merkmalskomplexen am Beispielkomplex B mit neun Merkmalen, jeweils mit verschiedenen Ausprägungen in: v.l.n.r. *Acila castrensis*, *Ensis ensis* und *Flexopecten glaber*.

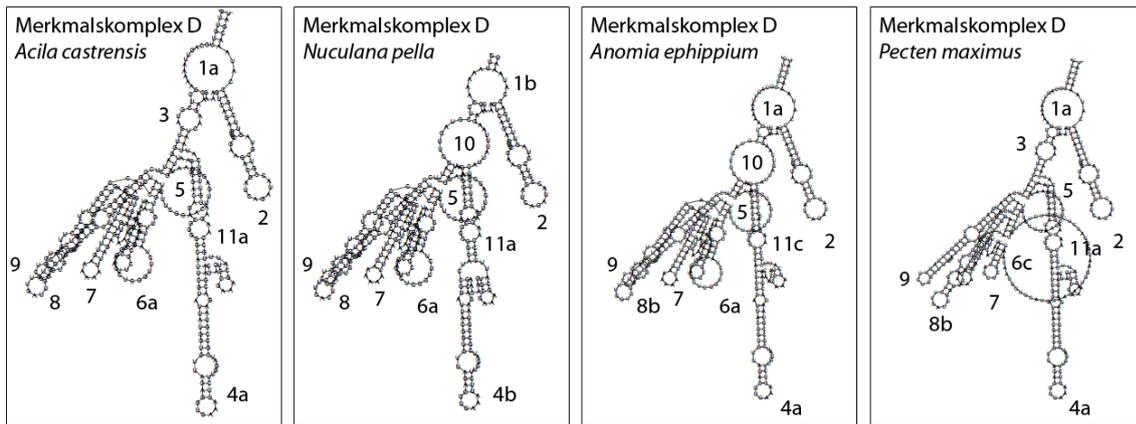


Abbildung 37: Near Natural (NNC) RNAsala. Verschiedene Ausprägungen von Merkmalenkomplexen am Beispielkomplex D mit 11 Merkmalen, z.T. mit verschiedenen Ausprägungen in: v.l.n.r. *Acila castrensis*, *Nuculana pella*, *Anomia ephippium* und *Pecten maximus*.

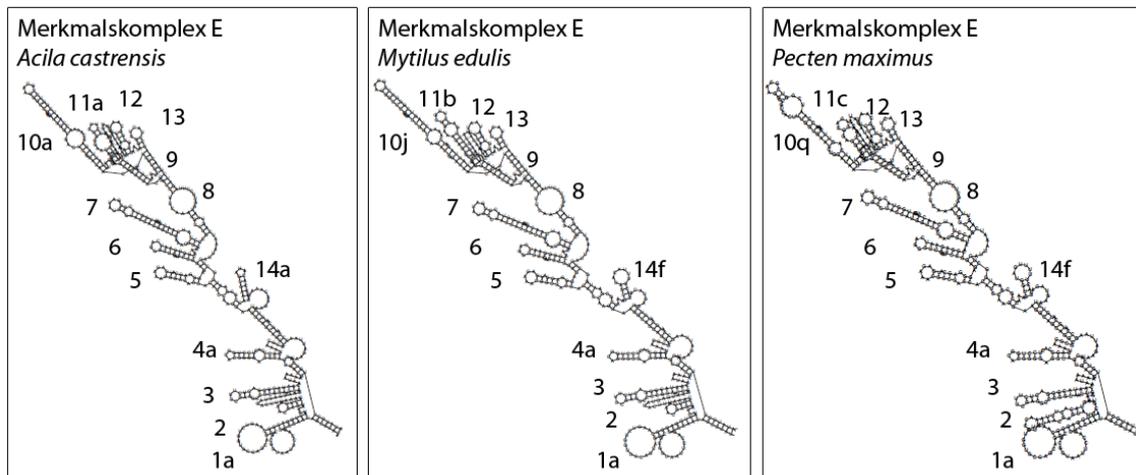


Abbildung 38: Near Natural (NNC) RNAsala. Verschiedene Ausprägungen von Merkmalenkomplexen am Beispielkomplex E mit 14 Merkmalen, z.T. mit verschiedenen Ausprägungen in: v.l.n.r. *Acila castrensis*, *Mytilus edulis* und *Pecten maximus*.

Abbildung 39 zeigt die Merkmalsmatrix für die Faltung nach NNC mit RNAsalsa. Für die Bivalvia wurden 74 Merkmale mit insgesamt 194 verschiedenen Merkmalsausprägungen definiert. Dabei ist anzumerken, dass die Unterscheidung von Merkmal und Merkmalsausprägung sowie deren Anfangs- und Endpunkte sehr stark von relativ willkürlichen Entscheidungen des Bearbeiters abhängen. Auch die Bestimmung der Ursache für unterschiedliche Merkmalsausprägungen erwies sich als nicht trivial, denn feste Regeln konnten kaum angewendet werden. Wie in der klassischen Morphologie tritt hier das Problem von multiplen Mutationsereignissen auf, was die numerische Einordnung erheblich erschwert.

Die Analyse von NNC zeigt ein weites Spektrum an Merkmalen und Merkmalsversionen

- ⇒ Bereiche, die zu variabel sind, um sinnvoll Merkmale zuzuordnen
- ⇒ Merkmale, die sowohl in Größe als in Ausprägung stark variieren, aber bei 95% der Arten vorhanden sind (Merkmal „E10“, Abbildung 38)
- ⇒ Merkmale, deren Definition als Merkmal oder Merkmalsversion schwierig ist (Merkmale von Komplex „A“, Abbildung 34)
- ⇒ Merkmale, die zwar in der Ausprägung variieren, aber deren Größe weitgehend konstant ist (Merkmal „B4“ oder „B6“, Abbildung 36)
- ⇒ Merkmale, die bis auf offensichtliche einzelne „Ausreißer“ nicht variieren (Merkmale „E5“ oder „E6“, Abbildung 38)

In den definierten Merkmalen traten Punktmutationen und kleinere InDels (bis zu 30 Basen) als Ursachen für verschiedene Ausprägungen auf. Lediglich in der Außengruppe traten signifikant größere InDels auf (*Bathypolypus arcticus* mit InDels > 100 Basen). Eine Punktmutation zieht jedoch nicht zwangsweise eine veränderte Merkmalsausprägung nach sich. Abgesehen von Merkmalskomplex A hatten die meisten Merkmale zwei bis fünf verschiedene Ausprägungen. Die maximale Anzahl von Merkmalsversionen fand sich bei Merkmal E10 mit 21 verschiedenen Ausprägungen. Diese wurden durch Punktmutationen und verschiedene InDels ausgelöst.

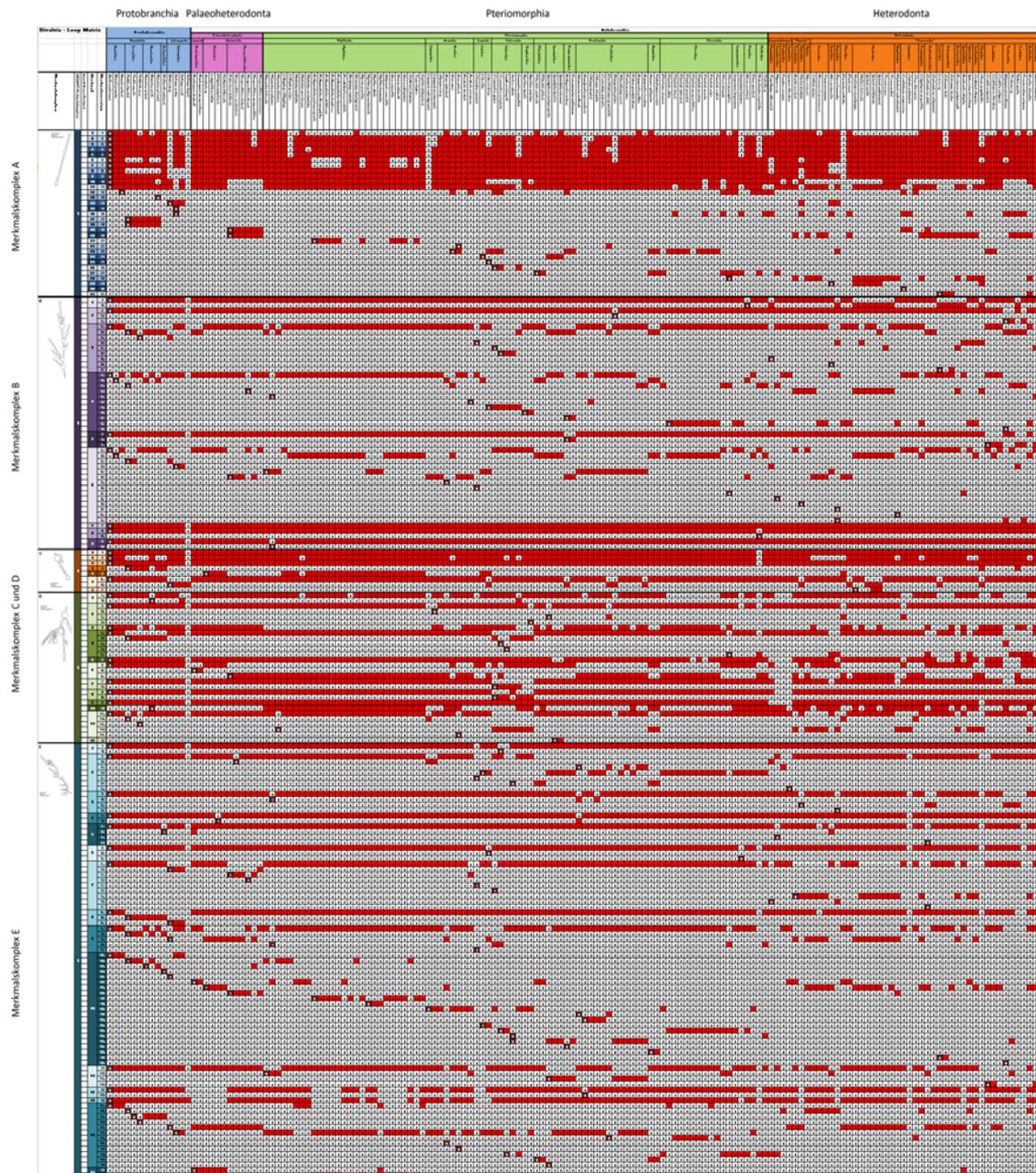


Abbildung 39: Merkmalsmatrix für 159 Sekundärstrukturen gefaltet nach NNC mit RNAsalsa. V.l.n.r.: Protobranchia, Palaeoheterodonta, Pteriomorpha, Heterodonta. V.o.n.u.: Merkmalskomplexe A - E mit 74 Merkmalen in 194 Merkmalsversionen.

4.1.2.1 *Statistische Auswertung*

Um einen Vergleich zwischen den Methoden quantitativ zu belegen, wurde auch bei der Faltung nach NNC eine statistische Auswertung vorgenommen. Diese fiel aufgrund der stark abweichenden Merkmalsstrukturen anders aus als bei MFE. Während bei den Sekundärstrukturen von MFE die Merkmale durch ihre Anordnung am zentralen Multi-branch Loop vorgegeben sind, sind die Merkmale bei NNC anders angeordnet. Bei der Definition von Merkmalen mussten diese willkürlich abgegrenzt werden. In manchen Fällen war eine klare Unterscheidung zwischen Merkmalen und Merkmalsversionen (homologen Merkmalen) nicht möglich, so zum Beispiel bei dem in Abbildung 34 dargestellten Merkmalskomplex A. Deshalb ist auch die Bestimmung der Größe der Merkmale relativ willkürlich. Sie wurde in den meisten Fällen durch die maximale Anzahl gemeinsamer Basen in der häufigsten Merkmalsversion bestimmt.

In dem hier untersuchten Fall waren von 74 Merkmalen (in 194 Ausprägungen) 42 Merkmale in mehr als 50% der Arten vorhanden. Dazu kommen die in 98% der Sekundärstrukturen auftretenden Merkmalskomplexe. Mehr als 50% der Merkmale sind konservativ und ändern sich zwischen den Arten kaum. Es konnte nicht eindeutig festgestellt werden, ob dieses Ergebnis durch den extrapolierten, gemeinsamen Constraint verursacht wurde, oder ob eine „echte“ gemeinsame, genetisch bedingte Struktur vorliegt. Die gewählte numerische Strenge, mit der der Constraint angewendet wird, spielt für dieses Ergebnis sicherlich eine entscheidende Rolle.

Für NNC lässt sich sagen, dass wenige Merkmalskomplexe auftreten. In dem hier ausgewerteten Datensatz konnten fünf signifikante Merkmalskomplexe festgelegt werden, jeweils mit 5-14 variablen Merkmalen. Im Gegensatz zu MFE sind diese Merkmalskomplexe so stabil, dass Veränderungen einzelner Merkmale auch durch größere InDels kaum Auswirkungen auf den ganzen Komplex haben. Diese Eigenschaft erhöht die Aussagekraft der NNC-Resultate für die Verwandtschaftsanalyse.

Die NNC-Sekundärstrukturen weisen eine hohe Stabilität auf (siehe Abbildung 30). Merkmale ändern durch kleine Mutationen entweder gar nicht ihre Ausprägung, bzw.

nur geringfügig. Diese Merkmale wurden, wie in Abschnitt 3.3.3.1, „Merkmalsanalyse“ beschrieben, als homolog eingestuft. In Abbildung 40 ist dargestellt, wie variabel die Merkmale sind, das heißt wieviele homologe Ausprägungen pro Merkmal vorkommen. Die Hälfte aller Merkmale kommt in höchstens zwei Ausprägungen vor. Mehr als fünf homologe Merkmalsversionen kommen bei weniger als 20% der Merkmale vor. Die Merkmale können als sehr stabil angesehen werden.

Für den Fall, dass es sich um ein Merkmal mit verschiedenen Ausprägungen handelt, wurde untersucht, wie stark die Basenzahl (die Größe des Merkmals) schwankt. Das Ergebnis ist in Abbildung 41 dargestellt. Wiederum ist die Variabilität sehr gering: 50% der homologen Merkmale haben weniger als zwei Basen Größenunterschied. Größenunterschiede von mehr als 11 Basen machen weniger als 20% der Merkmale mit verschiedenen Versionen aus. 37% der Merkmalsversionen entstehen nicht durch InDels sondern durch Punktmutationen.

In Abbildung 42 ist schließlich die Korrelation der beiden Faktoren, Variabilität und Größenänderung homologer Merkmale, dargestellt. Deutlich zu erkennen ist, dass sich ein Großteil der Merkmale im linken unteren Viertel befindet und eine geringe Variabilität aufweist. Es gibt zwei „Ausreißer“ mit 40 Basen Größenunterschied, bzw. 21 homologen Merkmalsausprägungen.

4.1.3 Vergleich von MFE und NNC

Wie an den oben vorgestellten Beispielen zu erkennen ist, weichen sowohl die Faltungen in ihren Merkmalen als auch die resultierenden Stammbäume (unter Abschnitt 7.1, „Abbildungen“) von MFE und NNC stark voneinander ab.

Die Unterschiede im Aussehen der Strukturen wurden bereits in der Abbildung 17 dargestellt.

Die absolute Anzahl von Merkmalen ist bei MFE um ein vielfaches höher als bei NNC. Merkmale bei MFE sind sehr variabel, nur ein geringer Teil ist als konservativ anzusehen. Es sind selten Merkmalskomplexe zu finden, diese sind nicht stabil. Punktmutationen

können zur Auflösung von ganzen Merkmalen und Merkmalskomplexen führen. Subjektiv scheint die Anzahl der Merkmale nahezu unbegrenzt. Gerade bei MFE wird deutlich, dass viele auf diese Weise generierte Strukturen artifiziell sind.

Die Merkmalsvarianz bei NNC ist durch die Anwendung eines Constraints stark eingeschränkt und bei weitem nicht so ausgeprägt wie bei MFE. Merkmalskomplexe sind über den gesamten Datensatz konstant vorhanden, mit einer nur beschränkten Varianz der Merkmale. Punktmutationen führen in den seltensten Fällen zur Auflösung eines Merkmals, sondern zu einer anderen Ausprägung. Merkmalskomplexe sind durch Punktmutationen nicht beeinflusst. Erst große InDels verursachen die Entstehung neuer Merkmale. Diese Eigenschaft der NNC-Faltungen verbessert das Ergebnis deutlich.

Bereits in den Merkmalsmatrizen ist ein deutliches „Rauschen“ bei MFE nicht zu übersehen. Bei NNC ist die Merkmalsverteilung deutlich stetiger. Abbildung 43 zeigt den direkten Vergleich der Merkmalsmatrizen.

Der Vergleich der resultierenden Stammbäume am Beispiel der Pteriomorphia (Abschnitte 4.2.3 und 4.2.4) zeigt eine Fortsetzung dieses „Rauschens“. In diesem Sinne muss angenommen werden, dass sich die Molekulare Morphologie nach NNC für die phylogenetische Analyse zumindest besser als MFE eignet. Die Analyse der Stammbäume soll klar stellen, ob die Methode nach NNC sinnvolle Ergebnisse im Sinne der „Deep Phylogeny“ liefern kann.

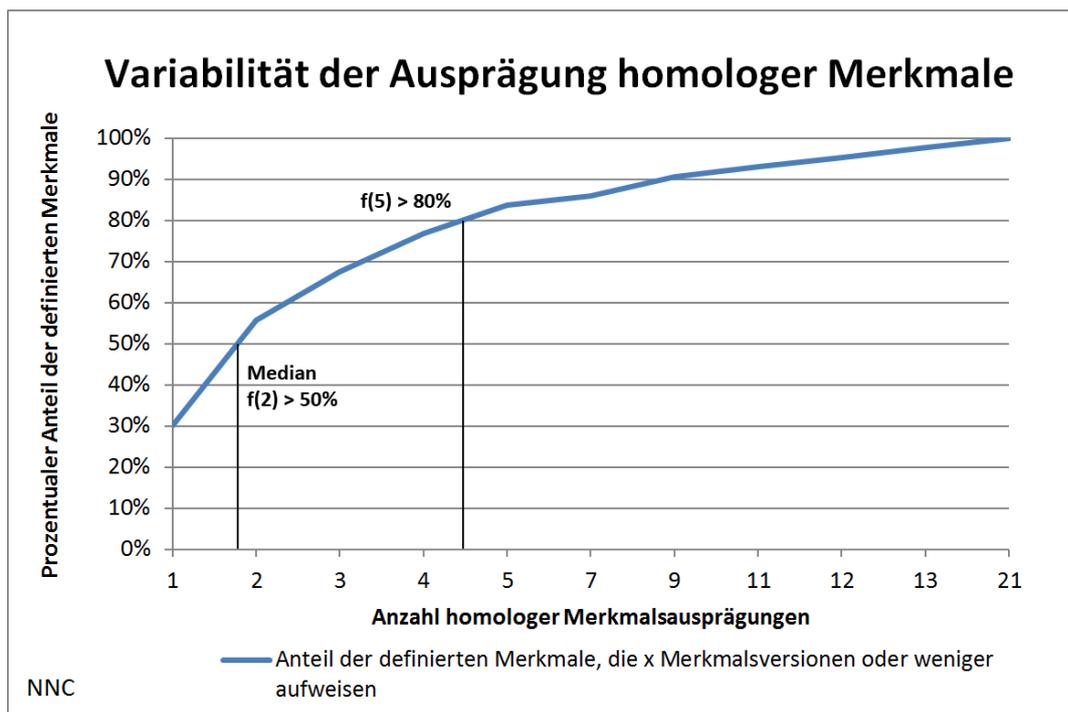


Abbildung 40: Variabilität der Ausprägung homologer Merkmale. 50% der definierten Merkmale haben zwei oder weniger homologe Ausprägungen. 80% der Merkmale haben weniger als fünf verschiedene Ausprägungen.

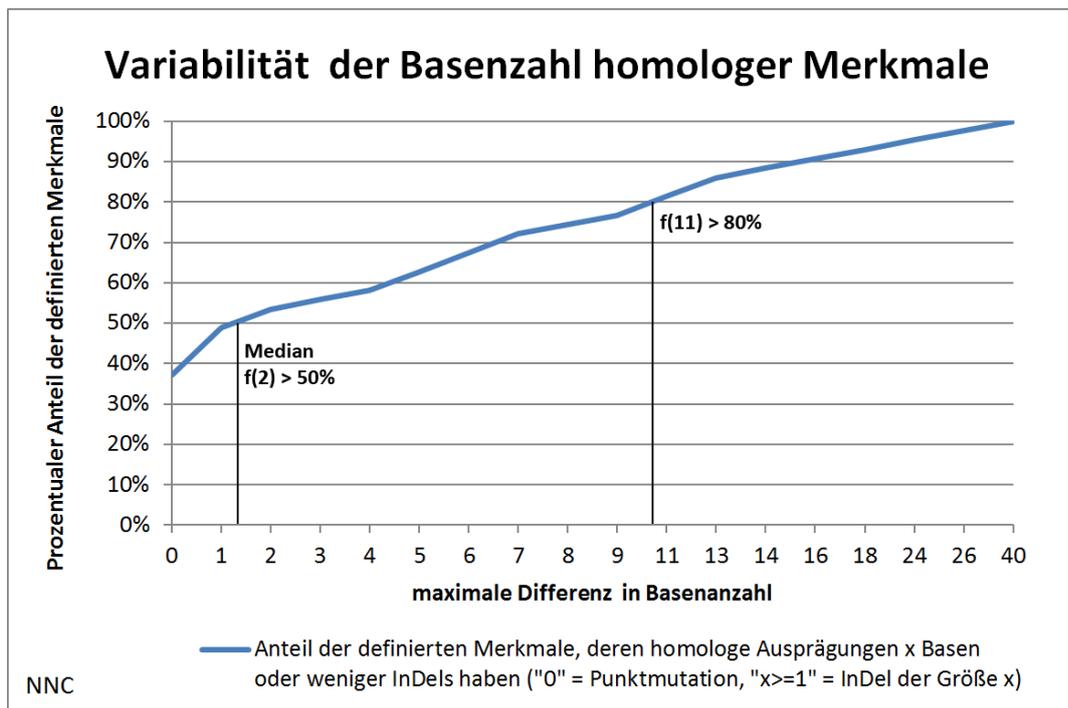


Abbildung 41: Maximaler Größenunterschied zwischen verschiedenen Ausprägungen eines Merkmals. 50% der Merkmalsausprägungen unterscheiden sich durch weniger als zwei Basen. Maximale Varianz sind 40 Basen.

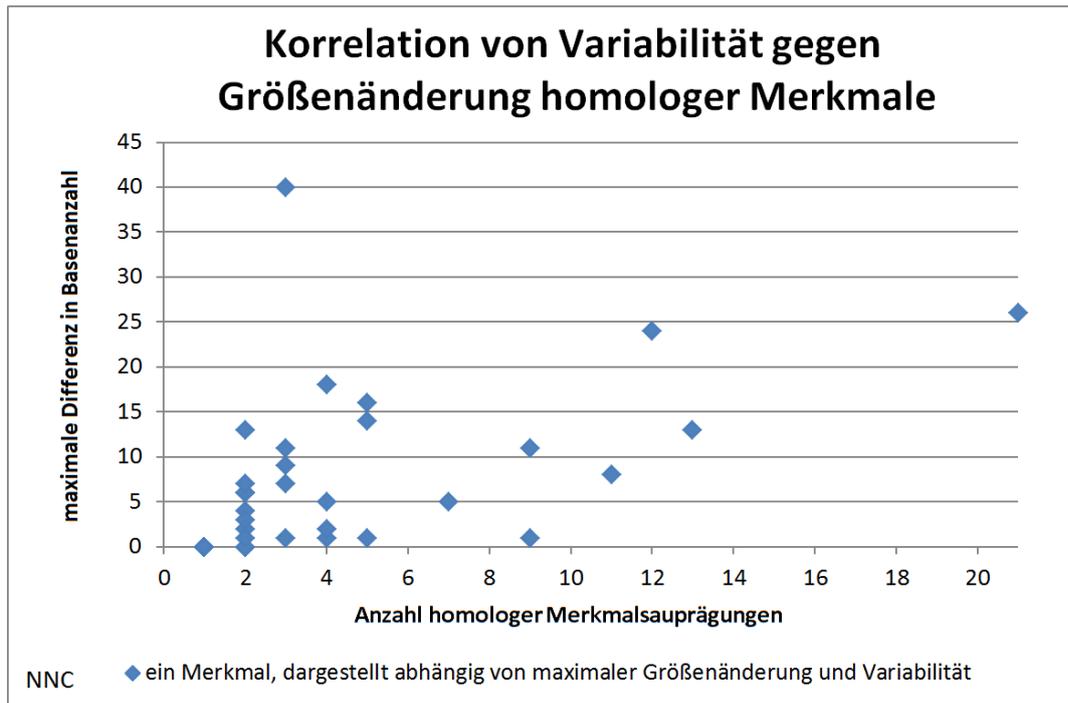


Abbildung 42: Korrelation zwischen Größenunterschied und Anzahl homologer Ausprägungen eines Merkmals. Der Großteil der Merkmale liegt im linken unteren Viertel und hat nur eine geringe Variabilität.

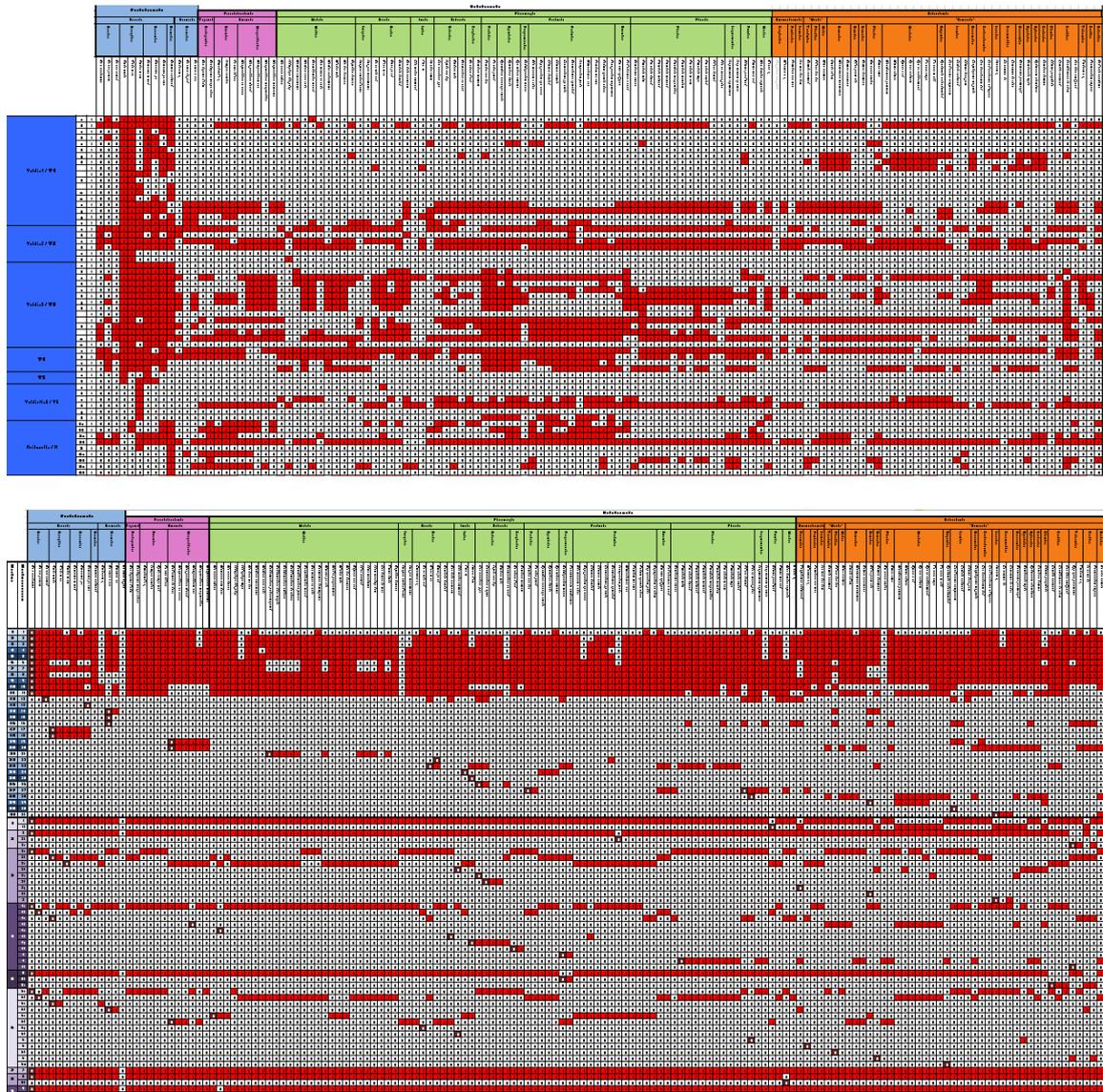


Abbildung 43: Ausschnitte aus den Merkmalsmatrizen von MFE (oben) und NNC (unten). Horizontal sind die Arten aufgetragen (v.l.n.r.: Protobranchia, Palaeoheterodonta, Pteriomorphia, Heterodonta), vertikal stehen die definierten Merkmale. Rot markiert sind vorhandene Merkmale. Bei MFE ist ein deutliches Rauschen zu erkennen, die Merkmale bei NNC zeigen eine größere Stetigkeit.

Tabelle 5: Übersicht über die erstellten Stammbäume.

			Protobranchia	Palaeo- heterodonta	Heterodonta	Pteriomorphia	Bivalvia	
MFE	single state	gleichgewichtet	x	x	x	x		
		gewichtet	Größe	x	x	x	x	x
NNC	single state	gleichgewichtet	x	x	x	x		
		gewichtet	Größe	x	x	x	x	x
			Komplexität	x	x	x	x	
	multi state	gleichgewichtet	x	x	x	x	x	
gewichtet		Größe	x	x	x	x	x	

4.2 Phylogenie der Bivalvia

Eine Übersicht zu allen erstellten Stammbäumen findet sich in Tabelle 5. In den detaillierten Abschnitten zu den Großgruppen der Bivalvia befindet sich jeweils eine Übersicht zu den verwendeten Methoden, der Außengruppe und dem Abschnitt im Anhang, wo die jeweilige Grafik zu finden ist.

Zusammenfassend lässt sich sagen, dass die Stammbäume nach MFE nur wenige sichere Aussagen zu der Verwandtschaft innerhalb der Bivalvia zulassen. Es besteht eine generelle Fragmentierung auf Familien-Niveau. Die Arten der Außengruppe sind nur in wenigen Fällen klar abgesetzt. Nach Bootstrap-Analysen weisen die meisten errechneten Phylogenien eine Vielzahl nicht aufgelöster Polytomien auf.

Die Stammbäume nach NNC sind deutlich besser aufgelöst. Eine richtige Zuordnung der untersuchten Arten auf Familien-Niveau ist oft gegeben. In vielen Fällen sind Überfamilien als zusammenhängende Blöcke zu finden. Es gibt bei NNC aber auch viele „Ausreißer“, bei denen einzelne Arten in diesen Blöcken fehlen, bzw. kein zusammenhängendes, übergeordnetes Taxon zugeordnet werden kann.

Ein bemerkenswertes Ergebnis der berechneten Stammbäume ist, dass die Nuculanoidea in jedem Fall monophyletisch sind. Die Ostreoidea, Pectinidae, Veneroidea, Mytiloidea, Mactroidea und Archiheterodonta sind in vielen Fällen monophyletisch, selten paraphyletisch oder mit „Ausreißern“. Die Anomiidae sind in vielen Stammbäumen mit den Plicatulidae gruppiert.

Als besonders instabil erscheinen die Protobranchia (in den Bivalvia-Stammbäumen), Pterioidea, Pinnoidea (obwohl nur durch zwei Arten präsentiert, wurden die in den seltensten Fällen gemeinsam eingeordnet) und Lucinoidea. Die Arten dieser Taxa sind selten zusammen gruppiert und oft auf alle Zweige des jeweiligen Stammbaums verteilt.

4.2.1 Protobranchia nach MFE

	Stammbaum	Methoden	Außengruppe
Protobranchia	Abschnitt 7.1.1.1	MFE, single-state, gewichtet nach Größe, gleichgewichtet	1 Palaeoheterodonta 2 Heterodonta 2 Pteriomorphia

Die Stammbäume der Protobranchia nach MFE weisen eine schlechte Auflösung auf. Nach einer Bootstrap-Analyse sind nur wenige Verzweigungen erhalten. Die Arten der Außengruppe sind nicht klar von den Nuculoidea oder Solemyidea getrennt. Eine 100% Zuordnung dieser Arten zu ihren übergeordneten Taxa ist in keinem der errechneten Stammbäume gegeben. Einzig die Nuculanoidea sind stabil monophyletisch und werden durch Bootstrapwerte gestützt.

4.2.2 Protobranchia nach NNC

	Stammbaum	Methoden	Außengruppe
Protobranchia	Abschnitt 7.1.2.2 Abschnitt 7.1.2.3	NNC, single-state, multi-state, gewichtet nach Größe oder Komplexität, gleichgewichtet	1 Palaeoheterodonta 2 Heterodonta 3 Pteriomorphia

Die Stammbäume nach NNC sind deutlich besser aufgelöst, als die mit MFE erstellten. Die Ergebnisse der meisten Analysen, mit unterschiedlich gewichteten oder gleichgewichteten, single-state oder multi-state Merkmalen unterstützen die Positionen der Arten in den bekannten, etablierten Überfamilien. Diese Einordnung ist im Falle der gleichgewichteten Merkmale auch durch Bootstrap-Werte unterstützt. Bei differenziert gewichteten Merkmalen ist diese Bestätigung nicht zu finden, hier sind fast keine sicheren Aussagen mehr zu machen.

Die Arten der Außengruppe sind bis auf Ausnahmen (*Ensis ensis*, nach gleichgewichteten, single-state Merkmalen oder Pteriomorphia und Heterodonta in multi-state Analysen) abgesetzt. Eine Zuordnung der Nuculanoidea in die Nähe der Pteriomorphia (vertreten durch drei Arten in der Außengruppe) findet hier nicht statt.

Die Nuculoidea und Solemyidea sind in den meisten Fällen monophyletisch, halten aber kaum einer Bootstrap-Analyse stand. Es besteht außerdem eine monophyletische Gruppe aus Yoldiidae, Nuculanidae und Neilonellidae. Diese Gruppe entspricht den Nuculanoidea und ist als einzige in allen Analysen stabil und durch mehr oder weniger hohe Bootstrap-Werte belegt. Die richtige Zuordnung der Arten zu ihren Familien ist zwar angedeutet, aber nicht stabil.

4.2.3 Pteriomorphia nach MFE

	Stammbaum	Methoden	Außengruppe
Pteriomorphia	Abschnitt 7.1.2.1	MFE, single-state, gewichtet nach Größe, gleichgewichtet	1 Palaeoheterodonta 2 Heterodonta 1 Protobranchia

Auf den ersten Blick ist bei beiden Stammbäumen eine Fragmentierung der größeren Taxa zu erkennen. Auch bei der Analyse der Familienzuordnung ist dieser Trend deutlich. Die Arten der Außengruppe sind auf die jeweiligen Stammbäume verteilt. Es sind große zusammenhängende Blöcke von Familien oder Überfamilien zu finden, zum Beispiel Mytilidae, Pteridae oder Pectinoidea, diese enthalten jedoch in keinem Fall alle dazugehörenden Arten. Allein die Ostreoidea sind vollständig und monophyletisch.

4.2.3.1 Gleichgewichtete, single-state Merkmale

Bei der Analyse mit gleichgewichteten Merkmalen sind von 15 Familien nur fünf eng zusammenhängend: Plicatulidae, Gryphaeidae, Isognomidae, Limopsidae und Pectinidae – nur die letzte besteht aus mehr als zwei Arten. Die Ostreidae bilden nur zusammen mit den Gryphaeidae ein Monophylum. Die Propeamussidae sind nur mit einer Art vertreten. Alle anderen Familien erscheinen separiert und über den Stammbaum verteilt.

Wenn man von den vielen „Ausreißern“ absieht, kann man eine nähere (zusammen monophyletische) Verwandtschaft von Pectinoidea, Pterioidea und Ostreoidea interpretieren. Mytiloidea, Limoidea und Arcoidea stehen paraphyletisch an der Basis des Baums. Diese Phylogenie ist am ehesten mit der von *Waller (1998)* zu vergleichen. Mit den neueren Arbeiten von *Steiner & Hammer (2000)* oder *Matsumoto (2001)* bestehen fast keine Gemeinsamkeiten.

4.2.3.2 Differenziert gewichtete, single-state Merkmale

Die zweite Analyse mit unterschiedlich gewichteten Merkmalen bietet ebenfalls nur wenige Ergebnisse die mit bestehenden Phylogenien übereinstimmen. Hier sind nur die Anomiidae, Plicatulidae und Isognomonidae, jeweils bestehend aus zwei Arten, monophyletisch. Alle anderen Familien sind zersplittert. Von den höheren Taxa sind nur die Ostreoidea monophyletisch.

Wie bei den gleichgewichteten Merkmalen ist die Fragmentierung so stark, dass jede Interpretation der Verwandtschaft der höheren Taxa sehr unsicher ist. Nach differenziert gewichteten, single-state Merkmalen sind Pectinoidea und Arcoidea nah verwandt, ihnen nächststehend sind die Ostreoidea. Die Mytiloidea sind als Adelphotaxon zu den drei Taxa eingeordnet. Die Pterioidea sind in kleineren Blöcken über den gesamten Stammbaum verteilt. Die größte Übereinstimmung besteht mit der von Steiner & Hammer (2000) aufgestellten Phylogenie.

4.2.4 Pteriomorphia nach NNC

	Stammbaum	Methoden	Außengruppe
Pteriomorphia	Abschnitt 7.1.2.2 Abschnitt 7.1.2.3	NNC, single-state, multi-state, gewichtete nach Größe oder Komplexität, gleichgewichtet	1 Palaeoheterodonta 2 Heterodonta 1 Protobranchia

4.2.4.1 Gleichgewichtete, single-state Merkmale

Die Außengruppe ist in dieser Analyse sehr gut abgesetzt. Viele Familien sind monophyletisch abgebildet: Mytiloidea, Spondylidae, Propeamussidae, Plicatulidae, Anomiidae, Ostreidae und Gryphaeidae. Die Mytiloidea bilden das Schwestertaxon zu einem polyphyletischen Komplex aus Pectinoidea und Limidae (ohne *Ctena divergens*). Ohne *Arca noae* und *A. imbricata* bilden die Arcoidea (Limopsoidea und Arcoidea) ein Monophylum. Die Pteriidae sind zweigeteilt und zusammen mit den Plicatulidae und

Anomiidae eingeordnet. Die Ostreoidea sind als Monophylum aus Ostreidae und Gryphaeidae dargestellt.

Die in dieser Analyse vorkommenden Gruppierungen von Pectinoidea + Limoidea, Plicatulidae + Anomiidae oder Arcoidea + Limopsoidea kommen in mehreren der in Abbildung 8 vorgestellten Phylogenien vor. Andererseits ist die nahe Verwandtschaft von Mytiloidea und Pectinoidea+Limidae in keiner der in Abbildung 8 vorgestellten Phylogenien zu finden.

4.2.4.2 Differenziert gewichtete, single-state Merkmale

Der hypothetische Stammbaum mit gewichteten, single-state-Merkmalen zeigt trotz „Ausreißen“ eine gute Zuordnung der Arten zu den etablierten Familien: Limopsidae, Ostreidae, Gryphaeidae, Malleidae, Plicatulidae, Anomiidae, Spondylidae, Pectinidae und Propeamussidae sind demnach monophyletisch. Pectinoidea und Limidae (ohne *Ctenoides annulata*) bilden zusammen innerhalb der Mytilidae ein Monophylum. Diese Zuordnung der Pectinidae in den Mytilidae erscheint zweifelhaft. Die Arcoidea bilden (ohne *Arca noae* und *A. imbricata*) das Schwestertaxon zu ebendiesen Mytilidae. Weitere Monophyla bilden die Ostreoidea und die Pterioidea (ohne *Pinna muricata*).

Die Außengruppe ist relativ gut abgesetzt, obwohl ein Zweig der Pteriomorphia, bestehend aus Propeamussidae, Spondylidae und *Ctenoides annulata*, ebenso abgesetzt ist.

Im Vergleich mit den Arbeiten von *Steiner & Hammer (2000)* und *Matsumoto (2001)* (Abbildung 8), lassen sich folgende Aussagen treffen: Die Gruppierung von Limoidea und Pectinoidea stimmt überein, ebenso die von Plicatuloidea und Anomioidea. Die bei Matsumoto dargestellte Verwandtschaft von Arcoidea und Limopsoidea kommt in dieser Analyse auch zum Vorschein. Auf der anderen Seite kann zum Beispiel die Anordnung der Mytilidae zusammen mit den Pectinoidea in keiner der vorangegangenen Arbeiten gefunden werden.

Die wahllos erscheinende Verteilung der Arcidae, Limidae, Pinnidae oder Pectinoidea im gesamten Stammbaum lassen aber Zweifel an der völligen Korrektheit dieser berechneten Phylogenie aufkommen.

4.2.4.3 Gleichgewichtete, multi-state Merkmale

Aufgrund von Programm-Limitationen von PAUP musste die Anzahl Taxa für diesen Stammbaum stark reduziert werden. Da in vorangegangenen Analysen nach NNC, die Zuordnung von Familien relativ sicher war, wurden bewusst „stabile“ Familien auf eine geringere Anzahl Arten reduziert.

Die Außengruppe ist gut abgetrennt. Mytiloidea und Ostreoidea, sowie Anomiidae, Spondylidae und Propeamussidae sind monophyletisch. Die Limidae, Pinnidae und Pectinidae sind paraphyletisch. Die Arcoida (ohne *Arca imbricata*) sind unaufgelöst in der Nähe der Außengruppe eingeordnet.

Einige Gruppierungen, wie Pectinoidea+Limoidea, Plicatulidae+Anomiidae oder Mytiloidea+Pterioidea+Pinnoidea, sind in den in Abbildung 8 dargestellten Stammbäumen zu finden. Gemeinsamkeiten in den Verzweigungen sind nicht zu finden.

4.2.4.4 Differenziert gewichtete, multi-state Merkmale

Abgesehen von *Ensis ensis* sind die Arten der Außengruppe gut abgesetzt. Die Mytiloidea, Arcoida (ohne *Arca noae* und *A. imbricata*) und Ostreoidea, sowie Spondylidae, Propeamussidae und Gryphaeidae sind monophyletisch. Die Pectinoidea (Plicatulidae) und Limidae (ohne *Ctena divergens*) bilden ein Monophylum.

Die Pterioidea sind über den gesamten Stammbaum verteilt. Anomiidae, Isognomonidae, Pinnidae und Malleidae, bestehend aus jeweils zwei Arten, sind nicht gemeinsam gruppiert.

Es gibt nur wenige Gemeinsamkeiten mit den in Abbildung 8 dargestellten Phylogenien.

4.2.5 Palaeoheterodonta nach MFE

	Stammbaum	Methoden	Außengruppe
Palaeo-heterodonta	Abschnitt 7.1.4.1	MFE, single-state, gewichtet nach Größe, gleichgewichtet	1 Protobranchia 2 Heterodonta 2 Pteriomorphia

Die Stammbäume nach MFE weisen nach einer Bootstrap-Untersuchung eine große Anzahl Polytomien auf. Die Unionidae und Trigonioidea sind als einzige monophyletisch und stabil. Die Zuordnung der Margaritiferidae ist unsicher, ebenso wie ihre Einordnung in die Unionoidea zusammen mit den Unionidae. Die Außengruppe ist nicht sinnvoll abgesetzt.

4.2.6 Palaeoheterodonta nach NNC

	Stammbaum	Methoden	Außengruppe
Palaeo-heterodonta	Abschnitt 7.1.4.2 Abschnitt 7.1.4.3	NNC, single-state, multi-state, gewichtet nach Größe oder Komplexität, gleichgewichtet	1 Protobranchia 2 Heterodonta 2 Pteriomorphia

Die Analyse von unterschiedlich gewichteten und gleichgewichteten Merkmalen stimmt darin überein, dass die bestehenden Familien der Neotrigoniidae, Unionidae und Margaritiferidae jeweils monophyletisch und klar voneinander abgegrenzt sind. Auch die Bootstrap-Werte unterstützen die Monophylien der Familien. Es werden in jeder Analyse Neotrigoniidae als nah verwandt mit den Unionidae gezeigt. Eine gemeinsame Linie der Unionoidea ist in keinem Stammbaum zu finden.

Eine Auflösung der unter- oder übergeordneten Taxa ist mit oder ohne Bootstrap nicht mit Sicherheit möglich.

4.2.7 Heterodonta nach MFE

	Stammbaum	Methoden	Außengruppe
Heterodonta	Abschnitt 7.1.3.1	MFE, single-state, gewichtet nach Größe, gleichgewichtet	1 Protobranchia 1 Palaeoheterodonta 2 Pteriomorphia

Die Analyse der Heterodonta nach MFE ergibt Stammbäume, die nur wenige zusammenhängende Überfamilien aufweisen: Nur die Tellinioidea und Pholadoidea sind mit unterschiedlich gewichteten und gleichgewichteten Merkmalen monophyletisch. Solenoidea sind bei differenziert gewichteten Merkmalen monophyletisch, Mactroidae sind (ohne *Mactromeris polynyma*) bei gleichgewichteten Merkmalen monophyletisch. Alle anderen höheren Taxa sind auf den Stammbaum verteilt. Die von *Giribet & Distel (2003)* und *Taylor et al. (2007)* postulierten Archiheterodonta sind angedeutet, aber nicht monophyletisch. Die Anomalodesmata sind kein Monophylum. Es gibt sowohl bei differenziert gewichteten als auch bei gleichgewichteten Merkmalen einen zusammenhängenden Block der von *Taylor et al. (2007)* eingeführten Neoheterodonte, es fehlt jedoch eine große Anzahl Taxa die nach Abbildung 10 dazugehören sollten, zum Beispiel Pholadoidea, *Dreissena polymorpha* oder *Chama gryphoides*.

4.2.8 Heterodonta nach NNC

	Stammbaum	Methoden	Außengruppe
Heterodonta	Abschnitt 7.1.3.2 Abschnitt 7.1.3.3	NNC, single-state, multi-state, gewichtet nach Größe oder Komplexität, gleichgewichtet	1 Protobranchia 2 Heterodonta 2 Pteriomorphia

Wie schon bei den oben genannten Protobranchia, Pteriomorphia und Palaeoheterodonta ist sowohl bei unterschiedlich gewichteten als auch bei gleichgewichteten Merkmalen die Zuordnung der Arten zu den bestehenden Familien und Überfamilien relativ sicher: Tellinoidea, Galeommatoidea, Solenoidea, Cardioidea und Mactroidea sind jeweils monophyletisch. Die Pholadoidea, Gastrochaenoidea, Tridacnoidea oder Veneroidea sind mono- bzw. paraphyletisch.

Die Carditoidea bilden zusammen mit *Astarte castanea* ein Monophylum (welches die Archiheterodonta von *Giribet & Distel (2003)* und *Taylor et al. (2007)* belegen würde).

Die Neoheterodonteia zeichnen sich zwar in allen Stammbäumen ab, die Zuordnung der Subtaxa ist jedoch nicht sicher. Die Positionen der Tellinoidea, Lucinoidea und Solenoidea inner- oder außerhalb dieses Taxons sind nicht aufzulösen.

Bis auf die Phylogenie mit gleichgewichteten Merkmalen, bilden die Pteriomorphia-Arten (Außengruppe) zusammen mit den Anomalodesmata (ohne *Tropidomya*) jeweils ein Monophylum innerhalb der Heterodonta.

Auch bei den Heterodonta gibt es Arten, bei denen keine sichere Zuordnung möglich ist: so sind die Arten der Lucinoidea in keinem der Stammbäume zusammen eingeordnet.

Gemeinsamkeiten auf Ordnungs-Niveau, die bei allen Stammbäumen vorkommen und so bei einer Auflösung der Heterodonta Belege liefern könnten, lassen sich nicht finden. Die Familien und Überfamilien sind immer wieder anders angeordnet.

4.2.9 Bivalvia nach MFE

	Stammbaum	Methoden	Außengruppe
Bivalvia	Abschnitt 7.1.5.1	MFE, single-state, gewichtet nach Größe	1 Scaphopoda 2 Polyplacophora 1 Gastropoda

4.2.9.1 Differenziert gewichtete, single-state Merkmale

Die Außengruppe ist schlecht abgesetzt, die Arten sind über alle Zweige verteilt. Grob vereinfacht stellen die Palaeoheterodonta zusammen mit den Nuculanoidea die Schwestergruppe zu Heterodonta, Pteriomorphia und den anderen Protobranchia. Obwohl es große zusammenhängende Blöcke von Pteriomorphia und Heterodonta gibt, sind fast die Hälfte der Arten (aus den Taxa Solemyidae, Anomalodesmata, Pteriidae, Solenoidea, Ostreoidea, Galeommatidea, Arcoidea, Limidae, Archiheterodonta) „bunt gemischt“.

4.2.10 Bivalvia nach NNC

	Stammbaum	Methoden	Außengruppe
Bivalvia	Abschnitt 7.1.5.2	NNC, single-state, multi-state, gewichtet nach Größe oder Komplexität, gleichgewichtet	1 Gastropoda 1 Scaphopoda 1 Polyplacophora 2 Cephalopoda

4.2.10.1 Differenziert gewichtete, single-state Merkmale

Die Außengruppe ist vermischt mit den Anomalodesmata, Tridacnidae, Cardiidae, sowie *Solemya togata* und *Chama gryphoides*.

Die Gruppierung der Arten auf Niveau von Familien und Überfamilien ist sehr gut erfolgt: Nuculanoidea, Cardiidae, Ostreoidea, Tellinoidea, Gastrochaenidae, Trigonioidea, Unionidae, Mactroidea oder Veneroidea sind monophyletisch. Viele andere Taxa sind paraphyletisch: Tridacnidae, Galeommatidea, Spondylidae oder Mytiloidea.

Die sinnvolle Zuordnung zu den Großgruppen ist jedoch nicht erfolgt. Die Palaeoheterodonta sind zweigeteilt (Neotrigoniidae+Unionidae, sowie Margaritiferidae). Pteriomorphia und Heterodonta bilden viele miteinander abwechselnde Blöcke. Eine höhere Verwandtschaft kann nicht sinnvoll abgeleitet werden.

4.2.10.2 Differenziert gewichtete, multi-state Merkmale

Aufgrund der Programm-Limitationen von PAUP musste die Anzahl Taxa für diesen Stammbaum stark reduziert werden. Da in vorangegangenen Analysen nach NNC, die Zuordnung von Familien relativ sicher war, wurden bewusst „stabile“ Familien auf eine geringere Anzahl Arten reduziert.

Wie schon in den vorangegangenen Stammbäumen nach NNC, erfolgte auch in diesem Stammbaum die Zuordnung der Familien und Überfamilien relativ sicher. Die Nuculanoidea, Ostreoidea, Anomioidea, Trigonioidea, Unionidae, Archiheterodonta, Mactroidea, Tellinoidea, Tridacnidae, Margaritiferidae, Spondylidae und Limidae sind monophyletisch. Dagegen sind die Pectinidae, Cardiidae, Plicatulidae und Veneroidea paraphyletisch.

Es gibt einen monophyletischen Zweig aus Nuculanoidea und Ostreoidea, dieser ist zusammen mit den Anomalodesmata, zwei Arten Mytilidae und zwei Arten der Außengruppe angeordnet. Die Pteriomorphia bilden zwei größere paraphyletische Blöcke. Die Palaeoheterodonta sind wiederum zweigeteilt (Neotrigoniidae+Unionidae, sowie Margaritiferidae). Die Heterodonta bilden ein Monophylum mit Teilen der Palaeoheterodonta und einigen Protobranchia.

4.2.10.3 Gleichgewichtete, multi-state Merkmale

Aufgrund der Programm-Limitationen von PAUP musste die Anzahl Taxa für diesen Stammbaum stark reduziert werden. Da in vorangegangenen Analysen nach NNC, die Zuordnung von Familien relativ sicher war, wurden bewusst „stabile“ Familien auf eine geringere Anzahl Arten reduziert. Dadurch wird auch die Anzahl der „Ausreißer“ verringert.

Der Stammbaum weist eine Reihe Polytomien auf – die wohl wichtigste besteht zwischen dem Großteil der Heterodonta, den Palaeoheterodonta, den Nuculanoidea und den Pteriomorphia. Die Arten der Außengruppe (außer *Liolophura*) sind zusammen mit den Anomalodesmata abgesetzt.

Die Pteriomorphia bilden zusammen mit den Arten der Protobranchia ein Monophylum (ohne *Lima lima*). Darin sind die Mytilidae, Pectinoidea (ohne *Plicatula*), Anomioidea und Ostreoidea jeweils monophyletisch. Die Arcoidea sind nicht aufgelöst. Die Pterioidea sind nicht vollständig zusammen gruppiert – *Electroma* und *Pinctada* sind bei den Pectinoidea eingeordnet.

Von den Palaeoheterodonta ist *Margaritifera* bei den Heterodonta eingeordnet, sonst bilden sie ein Monophylum parallel zu den Pteriomorphia, Nuculanoidea und Heterodonta.

Die Heterodonta sind zum Teil paraphyletisch außerhalb der oben genannten Polytomie, zum Teil monophyletisch (zusammen mit *Margaritifera*) innerhalb. Im äußeren Teil sind die Archiheterodonta monophyletisch angesiedelt, allerdings auch die Tellinoidea und die „Ausreißer“ *Lima lima*, *Ctenoides annulata* und *Pholax dactylus*. Im inneren Teil der Heterodonta sind die Mactroidea und die Galeommatoidea monophyletisch. Die Veneroidea sind schlecht aufgelöst und paraphyletisch.

5. Diskussion

5.1 Die Methode „Molekulare Morphologie“

Die Molekulare Morphologie, wie sie in dieser Arbeit auf das Beispiel der 18S rRNA der Bivalvia angewendet wurde, ist nach ihrem grundlegenden Prinzip eine logische und viel versprechende Methode, um sehr alte Verwandtschaftsbeziehungen innerhalb eines Taxons aufzulösen. Auf der anderen Seite zeigt die Methode auf dem Stand der Ausarbeitung, wie sie zum Zeitpunkt dieser Arbeit verfügbar war, noch eine Anzahl systematischer Schwierigkeiten, technischer Probleme und offener logischer Fragen. Es war das Hauptziel dieser Arbeit, durch praktische Anwendung mehrerer verfügbarer Algorithmen die wichtigsten Vor- und Nachteile dieser neuen Methode herauszufinden.

Nach dem Ergebnis dieser Arbeit wird das schwerwiegendste Problem der Molekularen Morphologie dadurch verursacht, dass methodisch bedingt die Analyse mit Sekundärstrukturen durchgeführt wird, die mittels künstlicher, formal-mathematischer Vorschriften generiert werden. Diese mit numerischen Algorithmen nach relativ einfachen physikalischen und chemischen Prinzipien errechneten Strukturen zeigen bestimmte qualitative (topologische) Merkmale wie Verzweigungen oder Schlaufen, deren Auftreten, Anzahl und relative Position sehr empfindlich von verhältnismäßig willkürlichen Details der unterschiedlichen Algorithmen und der darin frei wählbaren Parametern abhängt. Ein typischer Effekt ist der der sogenannten Frustration; zu den jeweils gegebenen Bedingungen existieren oft mehrere, sehr unterschiedlich aussehende Kompromisslösungen, die aber ein vorgegebenes Gütekriterium (wie z.B. das Minimum der Gesamtenergie) nahezu gleich gut erfüllen. Kleine Änderungen im Verfahren oder in den eingestellten Parametern können dann leicht dazu führen, dass im Ergebnis ein völlig anderer Kompromiss als „beste“ Lösung ermittelt wird, wenn auch mit nur geringem Güte-Vorteil gegenüber den nächstbesten Alternativen.

Dem gegenüber wirken in der Natur jedoch weitere Faktoren auf die Tertiärstrukturen ein, die die Funktionalität und Stabilität des Moleküls gewährleisten und eine extreme

Empfindlichkeit gegen Feinheiten der herrschenden Bedingungen, wie etwa winzige Temperaturschwankungen, weitgehend verhindern. Solche Faktoren können, wenn auch nur mit begrenztem Erfolg, in der Molekularen Morphologie im Prinzip mittels sogenannter Constraints imitiert werden, doch entsprechend geeignete Constraints sind derzeit nur für die 18S rRNA einer kleinen Anzahl von Bivalvia-Arten bekannt, im Gegensatz etwa zur großen Zahl der inzwischen bekannten genetischen Primärsequenzen. Im Falle der Bivalvia lag das Verhältnis bei nur einem bekannten Constraint zu ca. 160 bekannten Sequenzen. Unter bestimmten Voraussetzungen ist es sinnvoll und auch möglich, einen bekannten Constraint mit mathematischen Methoden auf weitere, nahe verwandte Arten zu extrapolieren.

Sowohl die Berechnung der Sekundärstruktur durch einen gegebenen Algorithmus, die Reduktion der natürlichen 3D-Strukturen auf zwei Dimensionen, als auch die Nutzung von extrapolierten Constraints unterliegen relativ willkürlichen, undifferenzierten mathematischen Festlegungen. Viele vom Nutzer unabhängige Entscheidungen werden notwendigerweise von den Algorithmen automatisch und ohne Rücksicht auf die spezifischen natürlichen, zumeist auch unbekanntem Selektionsfaktoren getroffen. So können bereits Punktmutationen erhebliche qualitative Auswirkungen auf umfangreiche berechnete Merkmalskomplexe haben.

Ein weiterer Mangel der getesteten gegenwärtigen Verfahren besteht darin, dass komplexere Merkmale wie etwa Pseudoknoten, wie sie von realen RNA-Strukturen bekannt sind (*Rietveld et al., 1982*), durch die Algorithmen noch nicht vorhergesagt werden können.

Nicht nur die Entscheidungen, die durch die Algorithmen getroffen werden, beeinflussen die zu analysierende Sekundärstruktur. Auch der Benutzer muss bei der Einrichtung und Anwendung der Programme eine Reihe von Entscheidungen treffen, die das Ergebnis der Berechnungen wesentlich beeinflussen. Das beginnt mit der Auswahl des geeigneten Software-Pakets, im Falle von Near-Natural (NNC) folgt dann die Ermittlung eines vermeintlich optimalen Sequenz-Alignments und endet mit der Festlegung der numerischen Parameter, wie z.B. der angenommenen

Umgebungstemperatur, auf bestimmte, relativ plausible, zumeist aber nur schwach begründete Werte für die anschließende eigentliche Berechnung der Sekundärstruktur.

Das vorbereitende Alignment der Sequenzen unterliegt sehr stark den subjektiven Entscheidungen des Bearbeiters. Automatische Alignments mit fest gesetzten Parametern liefern einen ersten Vorschlag, müssen aber ggf. noch manuell korrigiert bzw. optimiert werden. Diese Korrektur erfordert eine gewisse Erfahrung und Wissen um Mutationswahrscheinlichkeiten.

Obwohl alle Programme zur Vorhersage der Sekundärstrukturen grundsätzlich auf dem Zuker-Algorithmus aufbauen, sind die jeweiligen Implementierungen und die damit erzielten Ergebnisse sehr unterschiedlich. Auch der Satz der von außen justierbaren Parameter ist von Programm zu Programm verschieden. Meistens resultieren aus alternativ gewählten Variablen letztlich sehr unterschiedliche Sekundärstrukturen bei identischen gegebenen Primärsequenzen.

Die Faltung des Kettenmoleküls wird durch den Algorithmus bei angenommenen hypothetisch optimalen Bedingungen simuliert, die nicht notwendig realistisch sind. Allein die Vorgabe der Software, für die Faltung des Moleküls eine Umgebungstemperatur von 37°C festzulegen, ist fragwürdig. Im Falle der poikilothermen Bivalvia kann diese Umgebungstemperatur mit Blick auf die Wassertemperaturen, die gegenwärtig und in den letzten 500 Myr in Ozeanen und Seen als typisch angenommen werden können, weitgehend ausgeschlossen werden. Auch die restriktive Verallgemeinerung, dass dieselbe Faltungstemperatur für alle Taxa und für alle weit zurückliegenden Zeiträume angenommen wird, entspricht sicherlich nicht der Wirklichkeit. Allein die Bivalvia besiedeln ökologische Nischen mit einem sehr weiten Temperaturspektrum – von *Bathymodiolus thermophilus*, die bei hydrothermalen Quellen zu finden ist (aus denen bis zu 400°C heißes Wasser strömt, Reed (2006)), bis zur protobranchen *Deminucula atacellana*, die rezent in der Tiefsee bei wenigen Graden über 0°C lebt. Man muss solche stark vereinfachten Parameterwerte, ähnlich wie die Methode insgesamt, im Ergebnis der Untersuchung so einschätzen, dass es sich

prinzipiell um künstliche stereochemische RNA-Strukturen mit künstlichen physikochemischen Merkmalen handelt.

Es ist nachgewiesen worden, dass Proteine, wenn es der jeweilige Prozess erfordert, dynamisch ihre räumliche Struktur verändern können (*Dunker & Kriwacki, 2011*). Erst durch diese Flexibilität sind sie in der Lage, ihre verschiedenen Funktionen erfolgreich wahrzunehmen. Falls man diese Dynamik mit einer gewissen Berechtigung auch RNA-Molekülen zuschreibt, und außerdem deren chemische Metastabilität (*Lehman, 2010*) berücksichtigt, dann ist Molekulare Morphologie als Methode der Verwandtschaftsanalyse vielleicht nur mit künstlichen, „stabilen“ und vergleichbaren Strukturen aussichtsreich, mit "natürlichen" Sekundärstrukturen dagegen kaum.

Je länger die untersuchten Sequenzen sind, desto unsicherer und empfindlicher erwies sich die Berechnung der Sekundärstruktur. Um dieses Problem zu verringern, könnte man die Analyse auf kürzere Sequenzabschnitte konzentrieren, bzw. die Analyse aus den einzelnen Ergebnissen für kürzere Teilstücke zusammensetzen. Andererseits gibt es aber im gefalteten Molekül auch Wechselwirkungen und Bindungen zwischen Abschnitten, die in der Primärsequenz weit auseinander liegen und nicht nur lokal beschränkt sind auf Nachbarbasen in kurzen Sequenzabschnitten. Solche Wechselwirkungen größerer Reichweite sind aber für die globale Molekülstruktur essentiell und würden bei derartiger Aufspaltung nicht mehr berücksichtigt werden.

Die beschriebenen Fehlerquellen, die das Ergebnis der Molekularen Morphologie maßgeblich beeinflussen, potenzieren sich mit jedem Arbeitsschritt und sind dabei in den seltensten Fällen als Fehler erkennbar.

Wie die bisher dargestellten Einschätzungen zeigen, bedarf die sinnvolle Anwendung der Algorithmen der Molekularen Morphologie der häufigen Kontrolle und der Korrekturen durch einen mit der Methode technisch erfahrenen und mit dem Untersuchungsgegenstand wissenschaftlich vertrauten Nutzer. Diese Anforderung schließt eine routinemäßige Anwendung auf Massendaten, wie z.B. kürzlich erfolgreich

bei der genetischen Analyse der Phylogenie von Korallen geschehen (*Shinzato et al., 2011*), weitgehend aus.

Bezüglich des notwendigen humanen Arbeitsaufwands und der Einflüsse subjektiver Entscheidungen des Bearbeiters im Einzelfall ist neben der Berechnung der Sekundärstrukturen deren Auswertung noch kritischer zu sehen. Es gibt (noch) keine verlässlichen numerischen Verfahren, um automatisch ein sinnvolles, objektives Abstandsmaß für ein gegebenes Paar von Molekülen zu ermitteln. Stattdessen müssen Merkmale, deren Ähnlichkeit und Gewichtung individuell und oft subjektiv visuell beurteilt werden. Da die Zahl der notwendigen Vergleiche mit der Zahl der Sequenzen quadratisch wächst, wird hier erst recht sehr schnell eine Grenze der praktischen Machbarkeit erreicht.

Anschließend verbleiben grundsätzliche Fragen:

- ⇒ Wie erkennt man qualitative Unterschiede beim Vergleich einer künstlichen Struktur?
- ⇒ Wie kann die Homologisierung von Merkmalen systematisch erreicht werden, ggf. über das Alignment der Primärstruktur?
- ⇒ Wie stark / systematisch beeinflusst der extrapolierte Constraint das Ergebnis?
- ⇒ Inwiefern beeinflusst die Auswahl der Arten die Analyse der Merkmale?

Vorteile gegenüber anderen Verfahren sind:

- ⇒ Neue Merkmale, die vorhandene Hypothesen belegen, aber eigenständig nicht zuverlässig sind, können einbezogen werden
- ⇒ Es sind keine Labormittel notwendig
Auswertung erfolgt mittels bereits vorhandener Daten

Der entscheidende Vorteil der Molekularen Morphologie gegenüber der Analyse der Primärsequenz entsteht aus der zusätzlichen Berücksichtigung der paarweisen chemischen und physikalischen räumlichen Wechselwirkung der Basenmoleküle. Dadurch erhält ein einfaches Modell-Element (Nukleotid: A, C, G, T/U) in einer linearen Abfolge weitere Eigenschaften zugewiesen (Paarbindung und räumliche Anordnung).

Die Untersuchung der neuartigen topologischen Merkmale, die von der Primärsequenz ausgehend durch diese Verbesserung entstehen, stellt eine neue Qualität für die Beschreibung der Struktur und Eigenschaften des berechneten Makromoleküls dar.

Bisher gibt es nur wenige Arbeiten, in der die Methode der Molekularen Morphologie angewandt wurde. Dazu gehört zum Beispiel die Arbeit von Wang & Lee (2002), welche die Sekundärstruktur der 12S rRNA von 32 Fischarten untersucht. Dabei wurde aus einem experimentell abgeleiteten Strukturmodell ein Constraint entwickelt. Lydeard et al. (2000) haben die Sekundärstruktur der 23S rRNA von zehn Mollusca mit Hilfe eines Strukturmodells und vergleichender Sequenzanalyse untersucht.

Im Gegensatz dazu wurden in der vorliegenden, vergleichsweise umfassenden Arbeit verschiedene mögliche Sekundärstrukturen der 18S rRNA von bis zu 168 Arten der Bivalvia untersucht, zum Einen die Sekundärstrukturen ohne vorgegebenes Strukturmodell (nach MFE), zum Anderen wurde ein neues numerisches Verfahren (RNAsalsa) eingesetzt, um mittels Alignment naturnahe Sekundärstrukturen für die untersuchten Sequenzen zu konstruieren.

5.2 Phylogenie der Bivalvia

Die Phylogenie der Bivalvia ist kompliziert und von Grund auf schwierig nachzuvollziehen. Es gibt eine deutliche Lücke in den frühen Fossilienfunden der Bivalvia, die vom mittleren Kambrium bis zum frühen Ordovizium reicht. In dieser Phase fand eine weitreichende Radiation statt, die die Diversität deutlich erhöhte und den Muscheln neue ökologische Nischen erschloss (*Harper et al., 2000*).

Viele morphologische Merkmale der Bivalvia sind konvergent oder mehrfach parallel oder sukzessiv entstanden (*Campbell, 2000*). Immer wieder wurden nach Ereignissen des Massensterbens von den überlebenden Taxa dieselben Lebensräume und ökologischen Nischen neu besiedelt; morphologisch konvergente, ähnliche Anpassungen setzten sich immer wieder durch.

Genetisch wird die Analyse durch Radiationsereignisse erschwert, die in einem sehr kurzen Zeitraum ablaufen. Unter diesen Umständen ist die Annahme einer gleichförmigen, langsamen Mutationsrate („Molekulare Uhr“), wie sie bei der molekularen Evolutionsanalyse vorausgesetzt wird, nicht mehr wahrscheinlich (*Plazzi & Passamonti, 2010*).

Alle bisherigen Versuche, die Systematik der Bivalvia genauer zu klären, haben ihre spezifischen Schwierigkeiten, so auch die Molekulare Morphologie. Dabei ist bisher allerdings weniger der Forschungsgegenstand das Problem. Die Methode selbst ist unsicher und liefert bisher keine eindeutigen, robusten Ergebnisse.

Die im Anhang dargestellten Stammbäume zu den Bivalvia und ihren Großgruppen, sowie deren Analyse in Abschnitt 4.2 zeigen das beschriebene Problem recht deutlich.

In allen Großgruppen sind ähnliche Ergebnisse aufgetreten:

Die MFE-Sekundärstrukturen der 18S rRNA enthalten gewisse phylogenetische Informationen. Das „phylogenetische Rauschen“, welches ggf. durch die Algorithmen und die oben beschriebenen Freiheitsgrade entsteht, macht eine eindeutige und zuverlässige

Auflösung der Taxa in einem Stammbaum unmöglich. Bekannte und bestätigte phylogenetische Taxa sind zwar in allen getesteten Methoden erkennbar, aber immer durch „Ausreißer“ durchsetzt.

Bei der Analyse mit NNC erfolgt die Zuordnung der untersuchten Arten zu den etablierten Familien meist sicher (Ausnahmen sind Heterodonta: Lucinidae oder Protobranchia: Nuculanidae). Jede weiterführende Auflösung der über- und untergeordneten Taxa ist selten zuverlässig, wie die Bootstrap-Werte belegen (siehe Palaeoheterodonta), bzw. nicht unbedingt im Einklang mit bereits als gesichert angesehenen Erkenntnissen (siehe Pteriomorphia).

Besonders die Stammbäume, die alle untersuchten Bivalvia umfassen, sind je nach numerischen Verfahren untereinander sehr unterschiedlich und nicht im Einklang mit den bisherigen Erkenntnissen zur Phylogenie.

Die Methode sollte, so der ursprüngliche Plan, zunächst an der bereits etablierten und anerkannten Phylogenie der Pteriomorphia „geeicht“ werden. Keine der getesteten Verfahrensweisen der Molekularen Morphologie konnte jedoch zu 100% in Übereinstimmung mit den bisherigen Erkenntnissen gebracht werden, etwa durch die Festlegung von willkürlichen Parametern der Algorithmen. Die Ergebnisse, die in Abschnitt 4.2 beschrieben wurden, stellen den jeweils besten Kompromiss dar.

Durch die vielen Freiheitsgrade dieser Methode (Sequenzauswahl, Algorithmen und Parameter, 3D=>2D, Merkmalsdefinition und Gewichtung) können die Ergebnisse kaum zur zuverlässigen Bestätigung bisheriger Erkenntnisse oder zur Aufstellung neuer Hypothesen genutzt werden. Die Methode kann ggf. aber auf Schwächen in bekannten Phylogenien hinweisen.

Mit der Molekularen Morphologie, wie sie im Moment möglich ist, konnte kein eindeutiges Ergebnis zur Phylogenie der Bivalvia bestimmt werden. Eine Automatisierung der Methode, wie sie am Anfang geplant war, steht nicht in Aussicht.

5.3 Schlussfolgerungen

Die in dieser Arbeit untersuchte Methode liefert eindeutig ein Ergebnis, das phylogenetische Informationen enthält. Die Zuordnung von Familien und Überfamilien erfolgt relativ sicher. Einige Taxa sind über alle Stammbäume stabil dargestellt. Dieses Ergebnis deutet darauf hin, dass Molekulare Morphologie zumindest im Prinzip funktionieren „kann“. Künftige Fortschritte sind eine Frage der verbesserten Bewertung der Sekundärstrukturen und der Vorschriften, wie sie ermittelt werden. Falls in Zukunft die Möglichkeit besteht, die Molekulare Morphologie mit natürlichen, experimentell ermittelten oder „naturnäheren“ Strukturen durchzuführen, könnte sich das Ergebnis vermutlich deutlich gegenüber dem jetzigen Stand verbessern.

6. Literaturverzeichnis

- Adamkewicz SL, Harasewych MG, Blake J, et al. (1997) A Molecular Phylogeny of the Bivalve Mollusks. *Molecular Biology and Evolution* 14:619-629.
- Adams MD, Kelley JM, Gocayne JD, et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252:1651-1656.
- Aguinaldo AMA, Turbeville JM, Linford LS, et al. (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387:489-493. doi: 10.1038/387489a0
- Andronescu M, Condon A, Hoos HH, et al. (2007) Efficient parameter estimation for RNA secondary structure prediction. *Bioinformatics* 23:i19-i28.
- Andronescu MS, Pop C, Condon AE (2010) Improved free energy parameters for RNA pseudoknotted secondary structure prediction. *Rna New York Ny* 16:26-42.
- Ax P (1999) *Das System der Metazoa II - Ein Lehrbuch der phylogenetischen Systematik*. 1-383.
- Bell JL, Grassle JP (1998) A DNA probe for identification of larvae of the commercial surfclam (*Spisula solidissima*). *Molecular Marine Biology and Biotechnology* 7:127-137.
- Bieler R, Mikkelsen PM (2006) *Bivalvia - a look at the Branches*. *Zoological Journal of the Linnean Society* 148:223-235. doi: 10.1111/j.1096-3642.2006.00255.x
- Billoud B, Guerrucci MA, Masselot M, Deutsch JS (2000) Cirripede phylogeny using a novel approach: molecular morphometrics. *Molecular Biology and Evolution* 17:1435-1445.
- Bogan AE, Hoeh WR (2000) On becoming cemented: evolutionary relationships among the genera in the freshwater bivalva family Etheriidae (Bivalvia: Unionoidea). In: Harper EM, Taylor JD, Crane JA (eds) *The Evolutionary Biology of the Bivalvia* 2. Geological Society, Special Publication, London, pp 177: 159-168
- Bohr H, Brunak S (1989) A Travelling Salesman Approach to Protein Conformation. *Complex Systems* 3:9-28.
- Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Genetics* 76:1967-1971.
- Butterfield NJ (2008) An Early Cambrian Radula. *Journal of Paleontology* 82:543-554. doi: 10.1666/07-066.1
- Bändel K (1988) Stages in the ontogeny and a model of the evolution of bivalves (Mollusca). *Paläontologische Zeitschrift* 62:217-254. doi: 10.1007/BF02989494
- Campbell DC (2000) Molecular evidence on the evolution of the Bivalvia. In: Harper EM, Taylor JD, Crane JA (eds) *The Geological Society*, pp 31-46
- Canapa A, Barucca M, Marinelli A, Olmo E (2001) SHORT COMMUNICATION A Molecular Phylogeny of Heterodonta (Bivalvia) Based on Small Ribosomal Subunit RNA Sequences. *Molecular Phylogenetics and Evolution* 21:156 -161. doi: 10.1006/mpev.2001.0992
- Canapa A, Marota I, Rollo F, Olmo E (1999) The small-subunit rRNA gene sequences of venerids and the phylogeny of bivalvia. *Journal of Molecular Evolution* 48:463-468.
- Cannuel R, Beninger PG (2006) Gill development, functional and evolutionary implications in the Pacific oyster *Crassostrea gigas* (Bivalvia: Ostreidae). *Marine Biology* 149:547-563. doi: 10.1007/s00227-005-0228-6
- Cannuel R, Beninger PG, McCombie H, Boudry P (2009) Gill Development and its functional and evolutionary implications in the blue mussel *Mytilus edulis* (Bivalvia: Mytilidae). *The Biological bulletin* 217:173-88.

- Caron J-B, Scheltema A, Schander C, Rudkin D (2006) A soft-bodied mollusc with radula from the Middle Cambrian Burgess Shale. *Nature* 442:159-163.
- Carter JG (1990) Evolutionary significance of shell microstructure in the Palaeotaxodonta, Pteriomorpha and Isofilibranchia. In: Carter JG (ed) *Skeletal Biomineralization: Patterns, Process and Evolutionary Trends* volume I. Van Nostrand Reinhold, New York, New York, pp 135-411
- Chargaff E (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Cellular and Molecular Life Sciences* 6:201-209. doi: 10.1007/BF02173653
- Cook CE, Smith ML, Telford MJ, et al. (2001) Hox genes and the phylogeny of the arthropods. *Current biology* : CB 11:759-63.
- Cope JCW (1997) Early phylogeny of the class Bivalvia. *Palaeontology* 40:713-746.
- Cope JCW (1996) The early evolution of the Bivalvia. *The Malacological Society Of London* 30:361-370.
- Cox LR (1960) Thoughts on the classification of the Bivalvia. *Proceeding of the Malacological Society* 34:60-88.
- Crick FH (1966) Codon--anticodon pairing: the wobble hypothesis. *Journal of molecular biology* 19:548-55.
- David LA, Alm EJ (2010) Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* 1-37. doi: 10.1038/nature09649
- Deutsch JS (2011) Pending Issues in Development and Phylogeny of Arthropods. In: DeSalle R, Schierwater B (eds) *Key Transitions in Animal Evolution*. Science Publishers, Enfield, New Hampshire, pp 54-81
- Distel DL (2000) Phylogenetic relationships among mytilidae (Bivalvia): 18S rRNA data suggest convergence in mytilid body plans. *Molecular Phylogenetics and Evolution* 15:25-33.
- Distel DL, Amin M, Burgoyne A, et al. (2011) Molecular phylogeny of Pholadoidea Lamarck, 1809 supports a single origin for xylophagy (wood feeding) and xylophagous bacterial endosymbiosis in Bivalvia. *Molecular Phylogenetics and Evolution*. doi: 10.1016/j.ympev.2011.05.019
- Distel DL, Baco AR, Chuang E, et al. (2000) Do mussels take wooden steps to deep-sea vents? *Nature* 403:725-726.
- Do CB, Woods D a, Batzoglou S (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics (Oxford, England)* 22:e90-8. doi: 10.1093/bioinformatics/btl246
- Doucet-Beaupré H, Breton S, Chapman EG, et al. (2010) Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. *BMC Evolutionary Biology* 10:50.
- Dreyer H, Steiner G, Harper EM (2003) Molecular phylogeny of Anomalodesmata (Mollusca: Bivalvia) inferred from 18S rRNA sequences. *Zoological Journal of the Linnean Society* 139:229-246. doi: 10.1046/j.1096-3642.2003.00065.x
- Dunker AK, Kriwacki RW (2011) The orderly chaos of proteins. *Scientific American* 304:68-73.
- Ebeling W, Engel A, Feistel R (1990) *Physik der Evolutionsprozesse*. Akademie-Verlag
- Efron B (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7:1-26. doi: 10.1214/aos/1176344552
- Ender A, Schierwater B (2003) Placozoa are not derived cnidarians: evidence from molecular morphology. *Molecular Biology and Evolution* 20:130-134. doi: 10.1093/molbev/msg018
- Evseev GA, Semenikhina OY, Kolotukhina NK (2005) Early stage morphogenesis: an approach to problems of taxonomy, phylogeny and evolution in the Mytilidae (Mollusca: Bivalvia). *Current Zoology(formerly Acta Zoologica Sinica)* 51:1130-1140.

- Fedonkin MA, Simonetta A, Ivantsov AY (2007) New data on Kimberella, the Vendian mollusc-like organism (White Sea region, Russia): palaeoecological and evolutionary implications. *Geological Society, London, Special Publications* 286:157-179. doi: 10.1144/SP286.12
- Feistel R (1990) Ritualisation und die Selbstorganisation der Information. *Selbstorganisation*. Duncker & Humblot Berlin, Berlin, pp 83-98
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783-791. doi: 10.2307/2408678
- Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. *Science* 155:279-284.
- Fontana W (1996) Struktur einzelsträngiger Nukleinsäuren. 1-35.
- Franc A (1960) Classe de Bivalves. *Traite de Zoologie* 5/II
- Frischer M, Danforth J, Tyner L, et al. (1999) Development of an Argopecten-Specific 18S rRNA Targeted Genetic Probe. *Marine biotechnology* New York NY 2:11-20. doi: 10.1007/s101269900003
- Frischer M, Williams J, Kenchington E (1998) A molecular phylogeny of some major groups of Pectinidae inferred from 18S rRNA gene sequences. *Bivalves: An eon of evolution - Palaeobiological studies honoring Norman D. Newell*
- Gilbert W (1986) The RNA World. *Nature* 319:618.
- Giribet G (2008) Assembling the lophotrochozoan (= spiralian) tree of life. *Most* 1513-1522. doi: 10.1098/rstb.2007.2241
- Giribet G (2003) Molecules, development and fossils in the study of metazoan evolution; Articulata versus Ecdysozoa revisited. *Zoology Jena Germany* 106:303-326.
- Giribet G, Distel DL (2003) Bivalve phylogeny and molecular data. In: Lydeard C, Lindberg DR (eds) *Bivalve phylogeny and molecular data. Molecular Systematics and Phylogeography of Molluscs*. Washington and London: Smithsonian Books, pp 45–90
- Giribet G, Distel DL, Polz M, et al. (2000) Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelminthes, and Chaetognatha: a combined approach of 18S rDNA sequences and morphology. *Systematic Biology* 49:539-562.
- Giribet G, Wheeler W (2002) On bivalve phylogeny: a high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. *Invertebrate Biology* 121:271–324.
- Goloboff PA, Catalano SA, Mirande JM, et al. (2009) Phylogenetic analysis of 73 060 taxa corroborates major eukaryotic groups. *Cladistics* 25:211-230. doi: 10.1111/j.1096-0031.2009.00255.x
- Goloboff PA, Farris JS, Nixon KC (2008) Cladistics TNT , a free program for phylogenetic analysis. *Cladistics* 24:774-786. doi: 10.1111/j.1096-0031.2008.00217.x
- Graf DL, Cummings KS (2006) Palaeoheterodont diversity (Mollusca : Trigonioidea + Unionoidea): what we know and what we wish we knew about freshwater mussel evolution. *Zoological Journal of the Linnean Society* 343-394.
- Grenier JK, Garber TL, Warren R, et al. (1997) Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. *Current biology* : CB 7:547-53.
- Haeckel E (1866) *Generelle Morphologie. I: Allgemeine Anatomie der Organismen. II: Allgemeine Entwicklungsgeschichte der Organismen*. Berlin
- Hamel E (2007) *Das Werden der Völker in Europa : Forschungen aus Archäologie, Sprachwissenschaft und Genetik*. Tenea Verl.
- Harper EM, Dreyer H, Steiner G (2006) Reconstructing the Anomalodesmata (Mollusca: Bivalvia): morphology and molecules. *Zoological Journal of the Linnean Society* 148:395-420. doi: 10.1111/j.1096-3642.2006.00260.x

- Harper EM, Taylor JD, Crame JA (2000) Unravelling the evolutionary biology of the Bivalvia: a multidisciplinary approach. Geological Society Special Publication 177:1-9. doi: 10.1144/GSL.SP.2000.177.01.01
- Hennig W (1950) Grundzüge einer Theorie der Phylogenetischen Systematik. Deutscher Zentralverlag 1:370.
- Hennig W (1966) Phylogenetic Systematics. Univ. Illinois Press, Urbana
- Hoeh WR, Bogan AE, Heard WH (2001) A phylogenetic perspective on the evolution of morphological and reproductive characteristics in the Unionoida. In: Bauer G, Wächtler K (eds) Ecology and evolution of the freshwater mussels Unionoida. Ecological Studies, Volume 145. Springer-Verlag, Berlin, p xxii + 394 pp.
- Hoeh WR, Bogan AE, Heard WH, Chapman EG (2009) Palaeoheterodont phylogeny, character evolution, diversity and phylogenetic classification: A reflection on methods of analysis. *Dna Sequenc* 51:
- Hofacker IL (2003) Vienna RNA secondary structure server. *Most* 31:3429-3431. doi: 10.1093/nar/gkg599
- Ivany LC, Brey T, Huber M, et al. (2011) El Niño in the Eocene greenhouse recorded by fossil bivalves and wood from Antarctica. *Geophysical Research Letters* 38:L16709. doi: 10.1029/2011GL048635
- Jiménez-Montaño MA, Feistel R, Díez-Martínez O (2004) Information hidden in signals and macromolecules I. Symbolic time-series analysis. *Nonlinear dynamics, psychology, and life sciences* 8:445-478.
- Kenchington E, Landry D, Bird CJ (1995) Comparison of taxa of the mussel *Mytilus* (Bivalvia) by analysis of the nuclear small-subunit rRNA gene sequence. *Canadian Journal of Fisheries Aquatic Sciences* 52:2613-2620.
- Kenchington E, Naidu KS, Roddick DL, et al. (1993) Use of biochemical genetic markers to discriminate between adductor muscle of the sea scallop (*Placopecten magellanicus*) and the Iceland scallop (*Chlamys islandica*). *Canadian Journal of Fisheries Aquatic Sciences* 50:1222-1228.
- Kocot KM, Cannon JT, Todt C, et al. (2011) Phylogenomics reveals deep molluscan relationships. *Nature* 1-6. doi: 10.1038/nature10382
- Lehman N (2010) RNA in evolution. *Wiley interdisciplinary reviews RNA* 1:202-13. doi: 10.1002/wrna.37
- Linderstrøm-Lang KU (1952) Proteins and Enzymes. Lane Medical Lectures. Stanford University Publications, University Series, Medical Sciences, Stanford University Press., pp Bd. 6, S. 1-115
- Lindgren AR, Giribet G, Nishiguchi MK (2004) Cladistics A combined approach to the phylogeny of Cephalopoda (Mollusca). *Cladistics* 20:454-486.
- Littlewood DTJ, Ford SE, Fong D (1991) Small subunit rRNA gene sequence of *Crassostrea virginica* (Gmelin) and a comparison with similar sequences from other bivalve molluscs. *Nucl Acids Res* 19:6048- . doi: 10.1093/nar/19.21.6048
- Lu ZJ, Gloor JW, Mathews DH (2009) Improved RNA secondary structure prediction by maximizing expected pair accuracy. *Rna New York Ny* 15:1805-1813.
- Lydeard C, Holznagel WE, Schnare MN, Gutell RR (2000) Phylogenetic analysis of molluscan mitochondrial LSU rDNA sequences and secondary structures. *Molecular Phylogenetics and Evolution* 15:83-102.
- Lydeard C, Holznagel WE, Ueshima R, Kurabayashi A (2002) Systematic implications of extreme loss or reduction of mitochondrial LSU rRNA helical-loop structure in gastropods. *Malacologia* 44:349-352.
- Lyons TW, Reinhard CT (2009) Oxygen for heavy-metal fans. *Nature* 461:179-181.
- Maddison DR, Swofford DL, Maddison WP (1997) Nexus: An Extensible File Format for Systematic Information. *Systematic Biology* 46:590-621. doi: 10.1093/sysbio/46.4.590

- Marion GM, Millero FJ, Feistel R (2009) Precipitation of solid phase calcium carbonates and their effect on application of seawater SA–T–P models. *Ocean Science* 5:285-291. doi: 10.5194/os-5-285-2009
- Maruyama T, Ishikura M, Yamazaki S, Kanai S (1998) Molecular phylogeny of zooxanthellate bivalves. *The Biological Bulletin* 195:70-77.
- Mathews DH (2006) RNA secondary structure analysis using RNAstructure. *Current protocols in bioinformatics* editorial board Andreas D Baxevanis et al Chapter 12:Unit 12.6. doi: 10.1002/0471250953.bi1206s13
- Mathews DH (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *Rna New York Ny* 10:1178-1190.
- Matsumoto M (2001) Molecular phylogenetic analysis of pteriomorph bivalves. The Graduate School, Kanagawa University
- Matsumoto M (2003) Phylogenetic analysis of the subclass Pteriomorphia (Bivalvia) from mtDNA COI sequences. *Molecular Phylogenetics and Evolution* 27:429-440.
- McFadden GI, van Dooren GG (2004) Evolution: red algal genome affirms a common origin of all plastids. *Current biology* : CB 14:R514-6. doi: 10.1016/j.cub.2004.06.041
- Morton B (1996) The evolutionary history of the Bivalvia. In: Taylor J (ed) *Origin and evolutionary radiation of the Mollusca*. Oxford University Press, Oxford, pp 337-359
- Neulinger SC, Sahling H, Süling J, Imhoff JF (2006) Presence of two phylogenetically distinct groups in the deep-sea mussel *Acharax* (Mollusca : Bivalvia : Solemyidae). *Marine Ecology Progress Series* 312:161-168.
- Newell ND (1965) Classification of the Bivalvia. *American Museum Novitates* 2206:1-25.
- Newell ND (1969) Classification of Bivalvia. In: Moore RC (ed) *Treatise on Invertebrate Paleontology, Part N, Mollusca 6, Bivalvia*. Geological Society of America and University Of Kansas Press, Lawrence, p N205-N218
- Nixon KC (1999) The Parsimony Ratchet, a New Method for Rapid Parsimony Analysis. *Cladistics* 15:407-414. doi: 10.1111/j.1096-0031.1999.tb00277.x
- Nussinov R, Pieczenik G, Griggs JR, Kleitman DJ (1978) Algorithms for Loop Matchings. *SIAM Journal on Applied Mathematics* 35:68-82. doi: 10.1137/0135006
- Oliveira LC, Silva RTH, Leite VBP, Chahine J (2006) Frustration and hydrophobicity interplay in protein folding and protein evolution. *The Journal of chemical physics* 125:084904.
- Page TJ (2001) A Molecular Systematic Investigation of the *Limatula ovalis/pygmaea* species complex (Bivalvia: Limidae) in the Southern Ocean. University of London, London, United Kingdom
- Page TJ, Linse ÆK (2002) More evidence of speciation and dispersal across the Antarctic Polar Front through molecular systematics of Southern Ocean *Limatula* (Bivalvia : Limidae). *Polar Biology* 25:818-826. doi: 10.1007/s00300-002-0414-7
- Plazzi F, Ceregato A, Taviani M, Passamonti M (2011) A Molecular Phylogeny of Bivalve Mollusks: Ancient Radiations and Divergences as Revealed by Mitochondrial Genes. *PLoS ONE* 6:e27147. doi: 10.1371/journal.pone.0027147
- Plazzi F, Passamonti M (2010) Towards a molecular phylogeny of Mollusks: bivalves' early evolution as revealed by mitochondrial genes. *Molecular phylogenetics and evolution* 57:641-57. doi: 10.1016/j.ympev.2010.08.032
- Pushkarev D, Neff NF, Quake SR (2009) Single-molecule sequencing of an individual human genome. *Nature Biotechnology* 27:847-850.

- Puslednik L, Serb JM (2008) Molecular phylogenetics of the Pectinidae (Mollusca: Bivalvia) and effect of increased taxon sampling and outgroup selection on tree topology. *Molecular phylogenetics and evolution* 48:1178-88. doi: 10.1016/j.ympev.2008.05.006
- Rechenberg I (1973) *Evolutionsstrategien: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Friedrich Frommann Verlag
- Reed C (2006) Marine science: boiling points. *Nature* 439:905-907.
- Reid DG (1998) Class: Bivalvia. In: Beesley PL, Ross GJB, Wells A (eds) *Mollusca: The Southern Synthesis. Fauna of Australia Volume 5*. CSIRO PUBLISHING / Australian Biological Resources Study (ABRS), pp 195-234
- Remane A (1952) *Die Grundriegen des natürlichen Systems, der vergleichenden Anatomie und der Phylogenetik*. Leipzig: Akademische Verlagsgesellschaft Geest. English 364.
- Ren J, Shen X, Jiang F, Liu B (2010) The mitochondrial genomes of two scallops, *Argopecten irradians* and *Chlamys farreri* (Mollusca: Bivalvia): the most highly rearranged gene order in the family Pectinidae. *Journal of Molecular Evolution* 70:57-68.
- Reuter JS, Mathews DH (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11:129.
- Rice EL (1990) Nucleotide sequence of the 18S ribosomal RNA gene from the Atlantic sea scallop *Placopecten magellanicus* (Gmelin, 1791). *Nucleic Acids Research* 18:5551.
- Rice EL, Roddick D, Singh RK (1993) A comparison of molluscan (Bivalvia) phylogenies based on palaeontological and molecular data. *Molecular Marine Biology and Biotechnology* 2:137-146.
- Rietveld K, Van Poelgeest R, Pleij CW, et al. (1982) The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. *Nucleic Acids Research* 10:1929-1946.
- De Rijk P, De Wachter R (1993) DCSE, an interactive tool for sequence alignment and secondary structure research. *Computer applications in the biosciences : CABIOS* 9:735-40.
- Rijk PD, Wachter RD (1997) RnaViz , a program for the visualisation of RNA secondary structure. *Nucleic Acids Research* 25:4679-4684.
- Salvini-Plawen L, Steiner G (1996) Synapomorphies and plesiomorphies in higher classification of Mollusca. In: Taylor JD (ed) *Origin and evolutionary radiation of the Mollusca*. Oxford: Oxford University Press, pp 29-51
- Sankoff D (1985) Simultaneous Solution of the RNA Folding, Alignment and Protosequence Problems. *SIAM Journal on Applied Mathematics* 45:810-825. doi: 10.1137/0145048
- Shinzato C, Shoguchi E, Kawashima T, et al. (2011) Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476:320-323. doi: 10.1038/nature10249
- Smith DR, Snyder M (2007) Complete mitochondrial DNA sequence of the scallop *Placopecten magellanicus*: evidence of transposition leading to an uncharacteristically large mitochondrial genome. *Journal of Molecular Evolution* 65:380-391.
- Smith SA, Wilson NG, Goetz FE, et al. (2011) Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*. doi: 10.1038/nature10526
- Sneath PHA, Sokal RR (1973) *Numerical Taxonomy*. Freeman, San Francisco
- Sokal RR, Rohlf FJ (1995) *Biometry: the principles and practice of statistics in biological research*. WH Freeman 3:887.
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688-2690.

- Staple DW, Butcher SE (2005) Pseudoknots: RNA Structures with Diverse Functions. *PLoS Biology* 3:e213.
- Starobogatov YI (1992) Morphological basis for phylogeny and classification of Bivalvia. *Ruthenica* 2:1-25.
- Steiner G, Hammer S (2000) Molecular phylogeny of the bivalvia inferred from 18S rDNA sequences with particular reference to the Pteriomorphia. In: Harper EM, Taylor JD, Crame JA (eds) *The Evolutionary Biology of the Bivalvia*. The Geological Society of London, London, pp 11-29
- Steiner G, Hammer S, Harper EM, et al. (2000) Molecular phylogeny of the Bivalvia inferred from 18S rDNA sequences with particular reference to the Pteriomorphia. *Geological Society*, pp 11-29
- Steiner G, Müller M (1996) What can 18S rDNA do for bivalve phylogeny? *Journal of Molecular Evolution* 43:58-70.
- Stocsits RR, Letsch H, Hertel J, et al. (2009) Accurate and efficient reconstruction of deep phylogenies from structured RNAs. *Nucleic Acids Research* 37:6184-6193.
- Swofford DL (2003) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4 Sinauer As:
- Taylor J, Williams ST, Glover EA, Dyal P (2007) A molecular phylogeny of heterodont bivalves (Mollusca : Bivalvia : Heterodonta) : new analyses of 18S and 28S rRNA genes. *Zoologica Scripta* 36:587- 606. doi: 10.1111/j.1463-6409.2007.00299.x
- Telford MJ, Bourlat SJ, Economou A, et al. (2008) The evolution of the Ecdysozoa. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 363:1529-1537.
- Thiele J (1935) *Handbuch der Systematischen Weichtierkunde*. Dto *Zoologia of Biology* 41:1.
- Trusty T (2010) A colorful origin for the genetic code: information theory, statistical mechanics and the emergence of molecular codes. *Physics of life reviews* 7:362-76. doi: 10.1016/j.plrev.2010.06.002
- Tsubaki R, Kameda Y, Kato M (2011) Pattern and process of diversification in an ecologically diverse epifaunal bivalve group Pterioidea (Pteriomorphia, Bivalvia). *Molecular phylogenetics and evolution* 58:97-104. doi: 10.1016/j.ympev.2010.11.014
- Waegele JW, Erikson T, Lockhart P, Misof B (2009) The Ecdysozoa: Artifact or monophylum? *Journal of Zoological Systematics and Evolutionary Research* 37:211-223. doi: 10.1111/j.1439-0469.1999.tb00985.x
- Waller TR (1990) The evolution of ligament systems in the Bivalvia. In: Morton B (ed) *The Bivalvia. Proceedings of a memorial symposium in honour of Sir Charles Maurice Yonge (1899-1986)*, Edinburgh, 1986. Hong Kong University Press, Hong Kong. viii + 355 pp., Hong Kong, pp 49-71
- Waller TR (1998) Origin of the molluscan class Bivalvia and a phylogeny of major groups. *Bivalves an eon of evolutionpaleobiological studies honoring Norman D Newell* 1–45.
- Waller TR (1997) Origin of the molluscan class Bivalvia and a phylogeny of major groups. *The Bivalvia: half a billion years of evolution - essays in honor of Norman D. Newell*
- Wang H-Y, Lee S-C (2002) Secondary Structure of Mitochondrial 12S rRNA Among Fish and Its Phylogenetic Applications. *Molecular Biology* 19:138-148.
- Watson JD, Crick FHC (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171:737-738.
- Wells S (2004) *The Journey of Man: A Genetic Odyssey*. xvi, 224 p. [16] p. of plates.
- Westheide W, Rieger R (1996) *Spezielle Zoologie, Teil 1: Einzeller und Wirbellose Tiere*. Gustav Fischer Stuttgart 89:179–180.
- Winnepenninckx B, Backeljau T, De Wachter R (1996) Investigation of molluscan phylogeny on the basis of 18S rRNA sequences. *Molecular Biology and Evolution* 13:1306-1317.

- Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* 74:5088-90.
- Zuckerandl E, Pauling L (1965) Molecules as documents of evolutionary history. *Journal of Theoretical Biology* 8:357-366.
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Bioinformatics* 31:3406-3415. doi: 10.1093/nar/gkg595
- Zuker M, Mathews DH, Turner DH (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide.
- Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research* 9:133-148.

7. Anhang

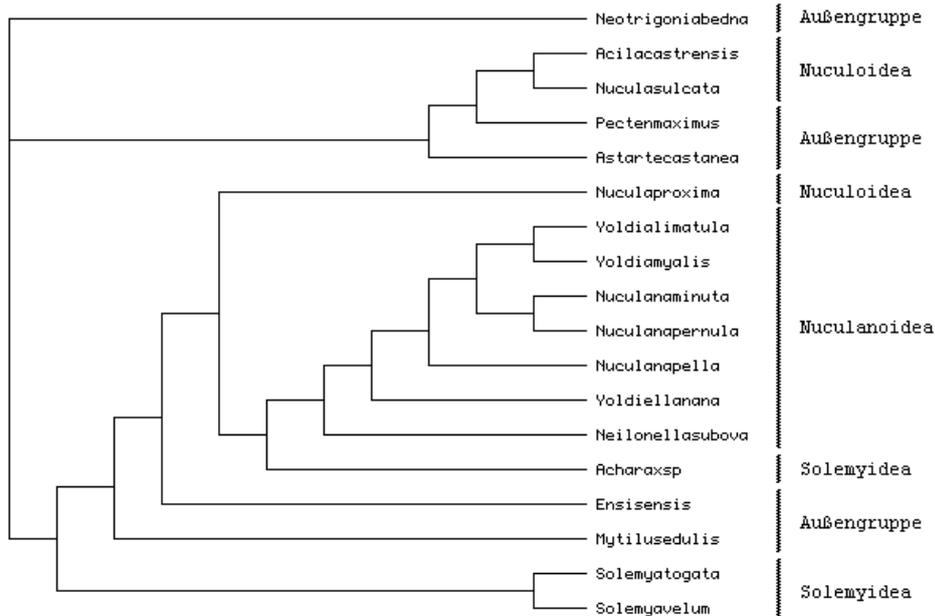
7.1 Abbildungen

7.1.1 Stammbäume Protobranchia

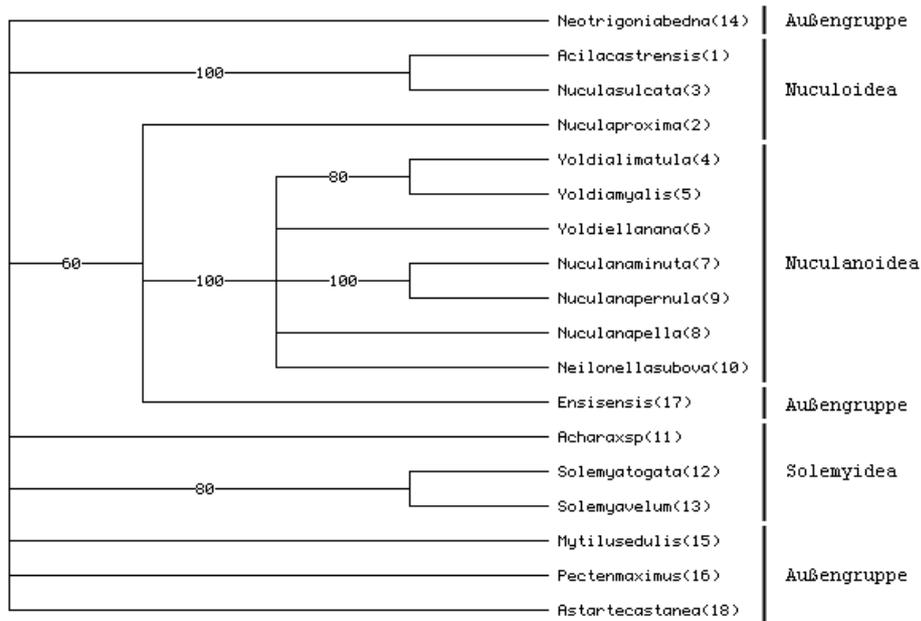
7.1.1.1 Protobranchia MFE Single State

7.1.1.1.1 Differenziert gewichtet nach Größe

Parsimony Tree für Protobranchia nach MFE
differenziert gewichtet nach Größe, single-state Merkmale



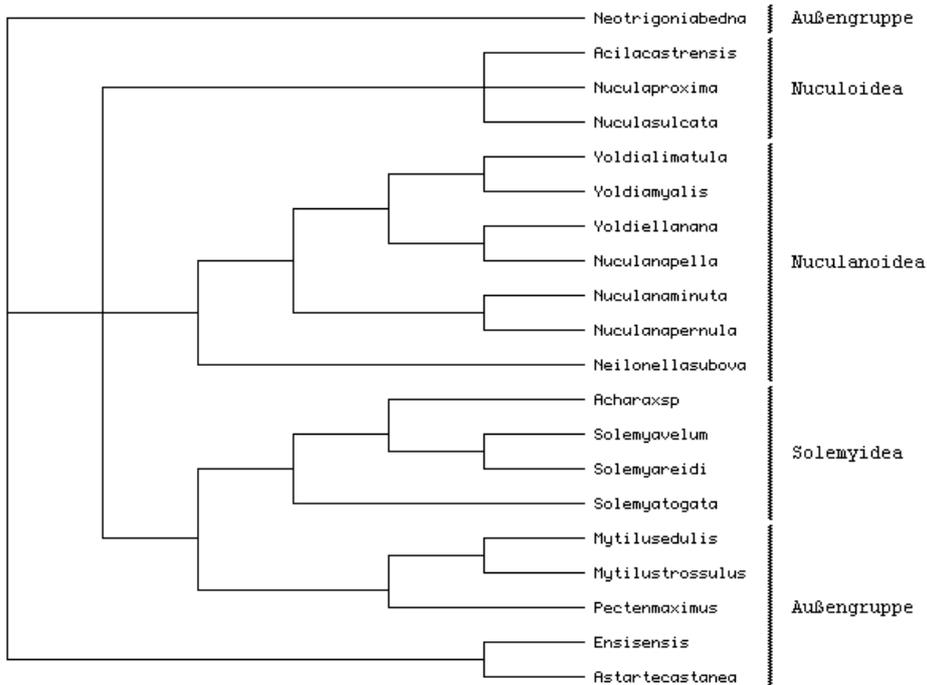
Parsimony Tree mit Bootstrap für Protobranchia nach MFE
differenziert gewichtet nach Größe, single-state Merkmale



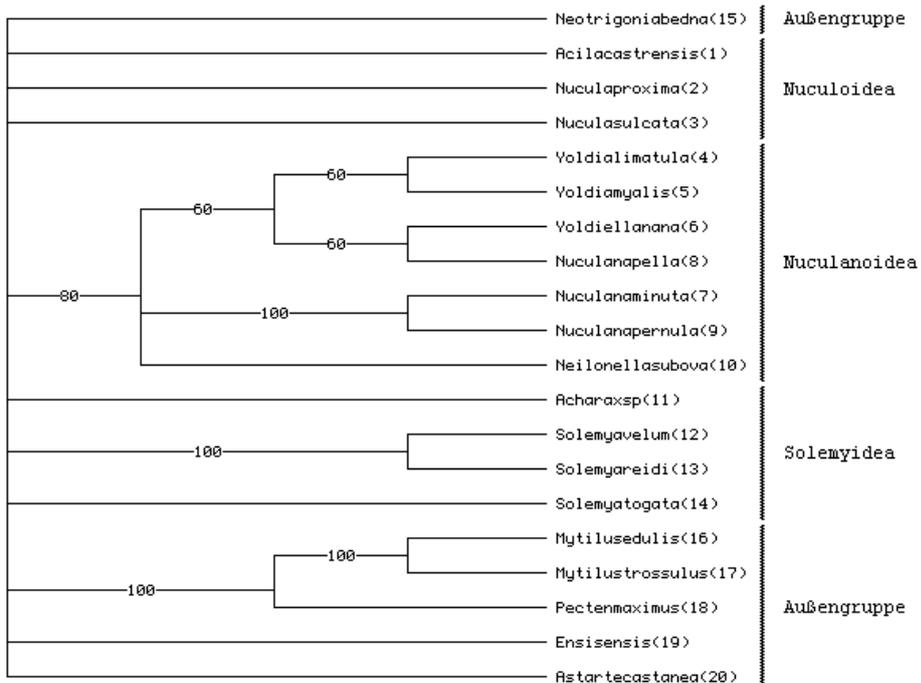
7.1.1.2 Protobranchia NNC Single State

7.1.1.2.1 Differenziert gewichtet nach Größe

Parsimony Tree für Protobranchia nach NNC
differenziert gewichtet nach Größe, single-state Merkmale

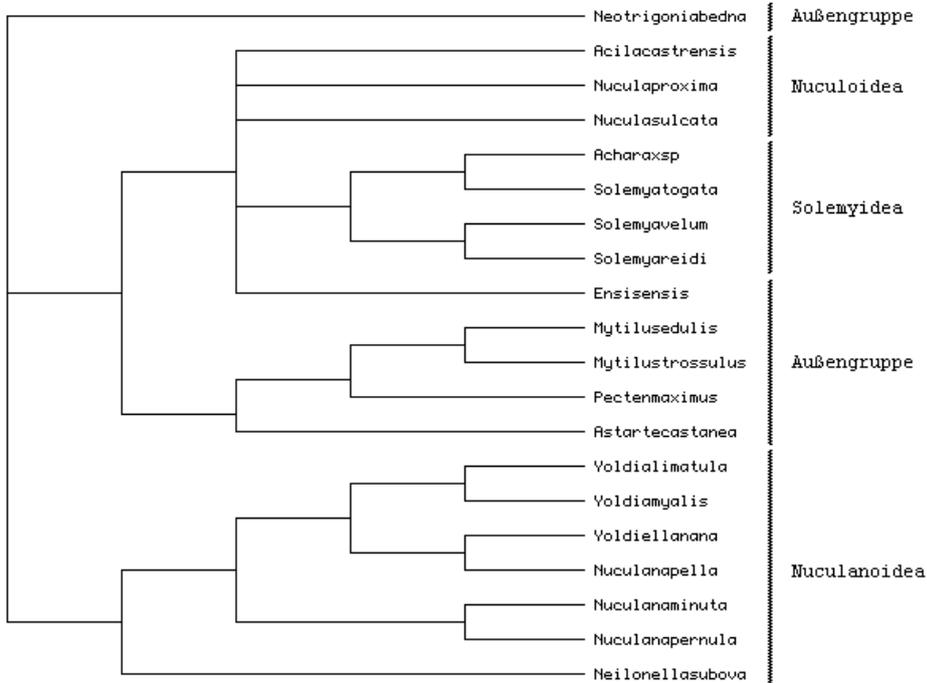


Parsimony Tree mit Bootstrap für Protobranchia nach NNC
differenziert gewichtet nach Größe, single-state Merkmale

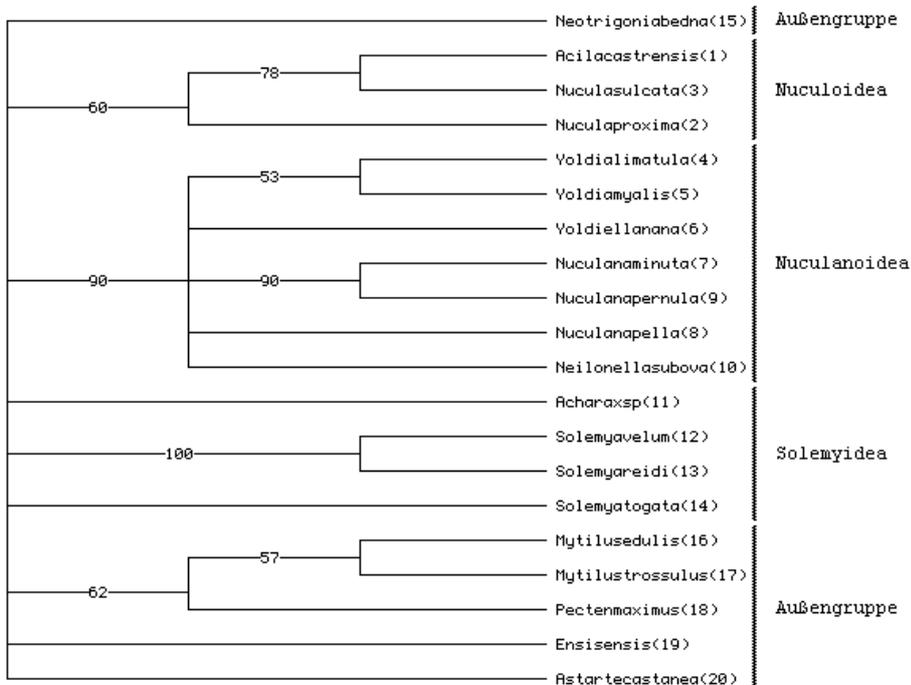


7.1.1.2.2 Differenziert gewichtet nach Komplexität

Parsimony Tree für Protobranchia nach NNC
differenziert gewichtet nach Komplexität, single-state Merkmale

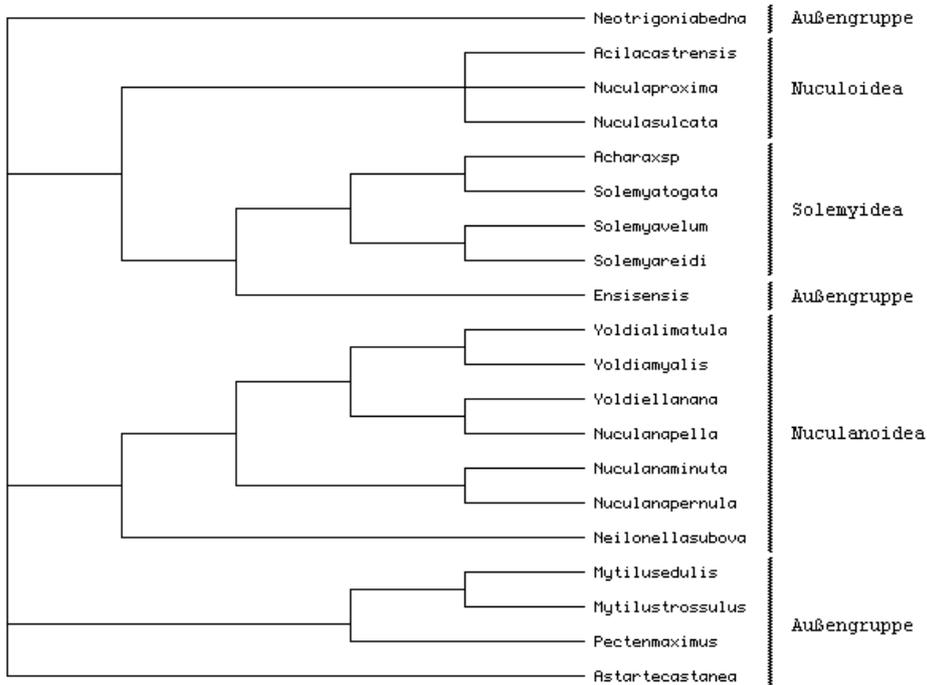


Parsimony Tree mit Bootstrap für Protobranchia nach NNC
differenziert gewichtet nach Komplexität, single-state Merkmale

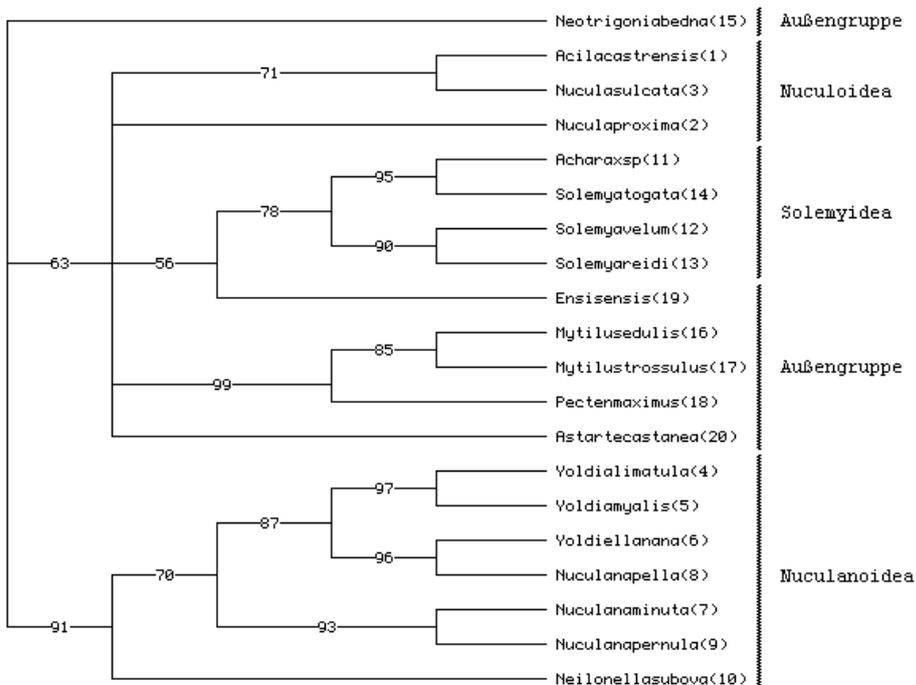


7.1.1.2.3 Gleichgewichtet

Parsimony Tree für Protobranchia nach NNC
gleichgewichtete, single-state Merkmale



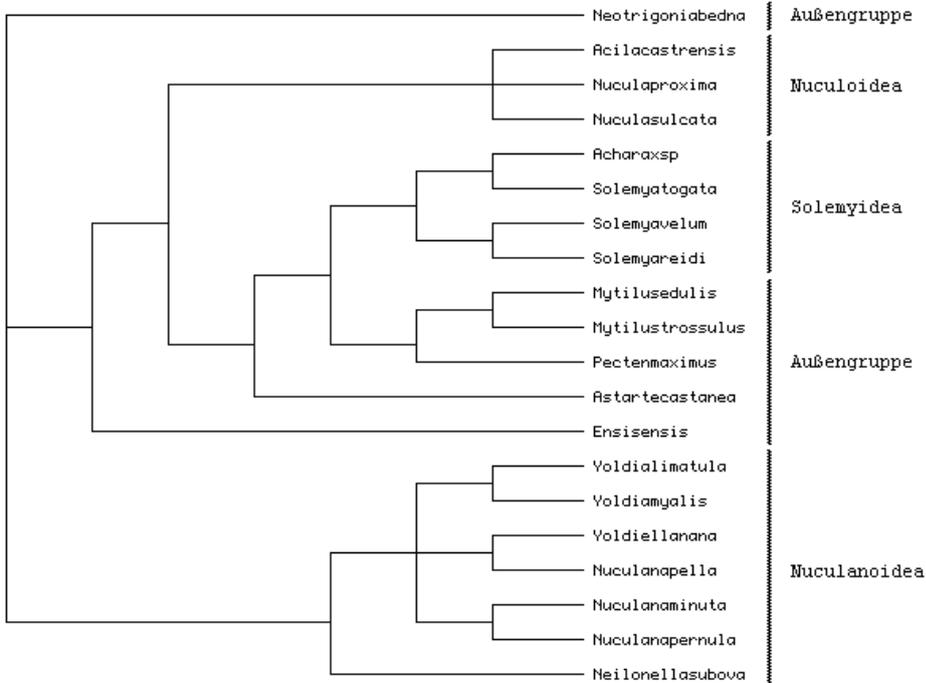
Parsimony Tree mit Bootstrap für Protobranchia nach NNC
gleichgewichtete, single-state Merkmale



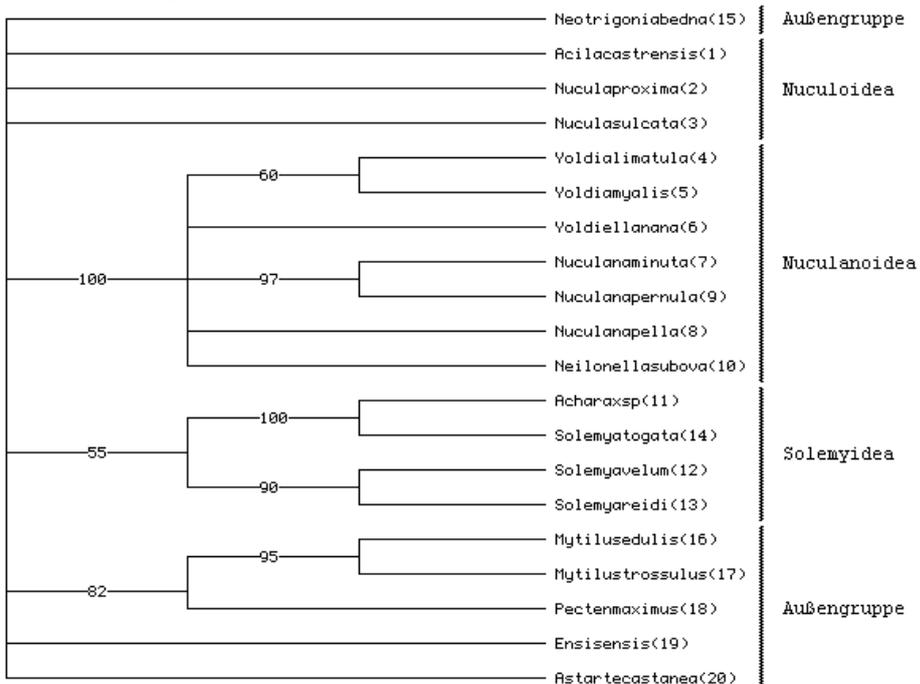
7.1.1.3 Protobranchia NNC Multi State

7.1.1.3.1 Differenziert gewichtet nach Größe

Parsimony Tree für Protobranchia nach NNC
differenziert gewichtet nach Größe, multi-state Merkmale

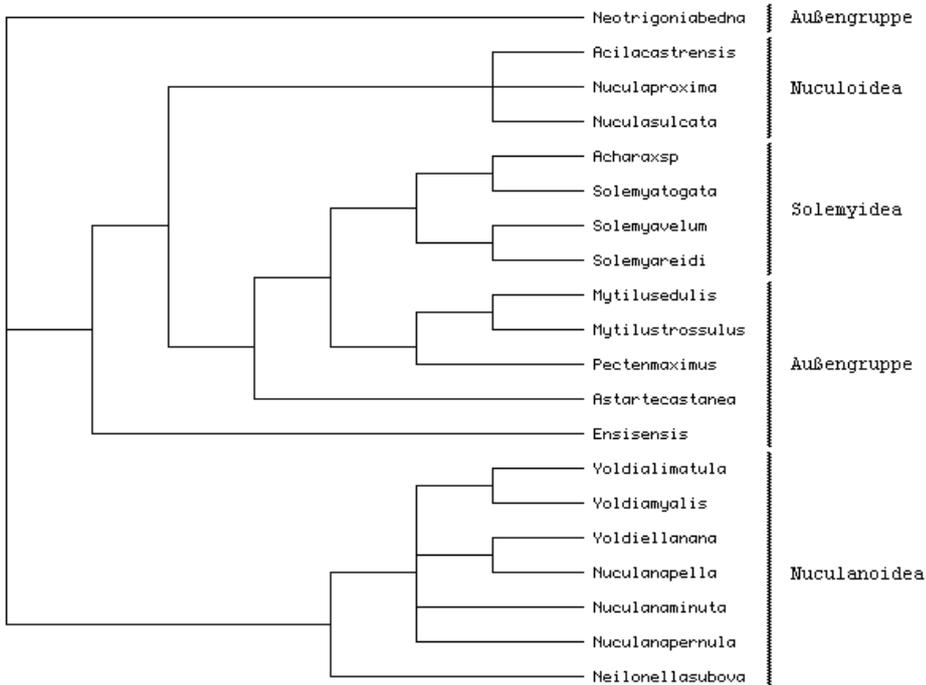


Parsimony Tree mit Bootstrap für Protobranchia nach NNC
differenziert gewichtet nach Größe, multi-state Merkmale

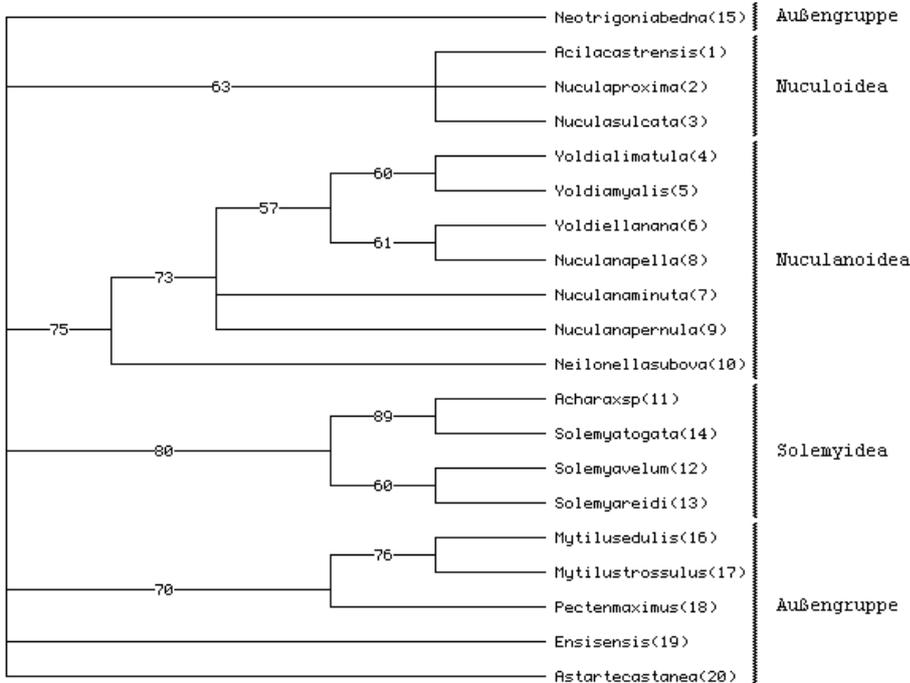


7.1.1.3.2 Gleichgewichtet

Parsimony Tree für Protobranchia nach NNC
gleichgewichtete, multi-state Merkmale



Parsimony Tree mit Bootstrap für Protobranchia nach NNC
gleichgewichtete, multi-state Merkmale

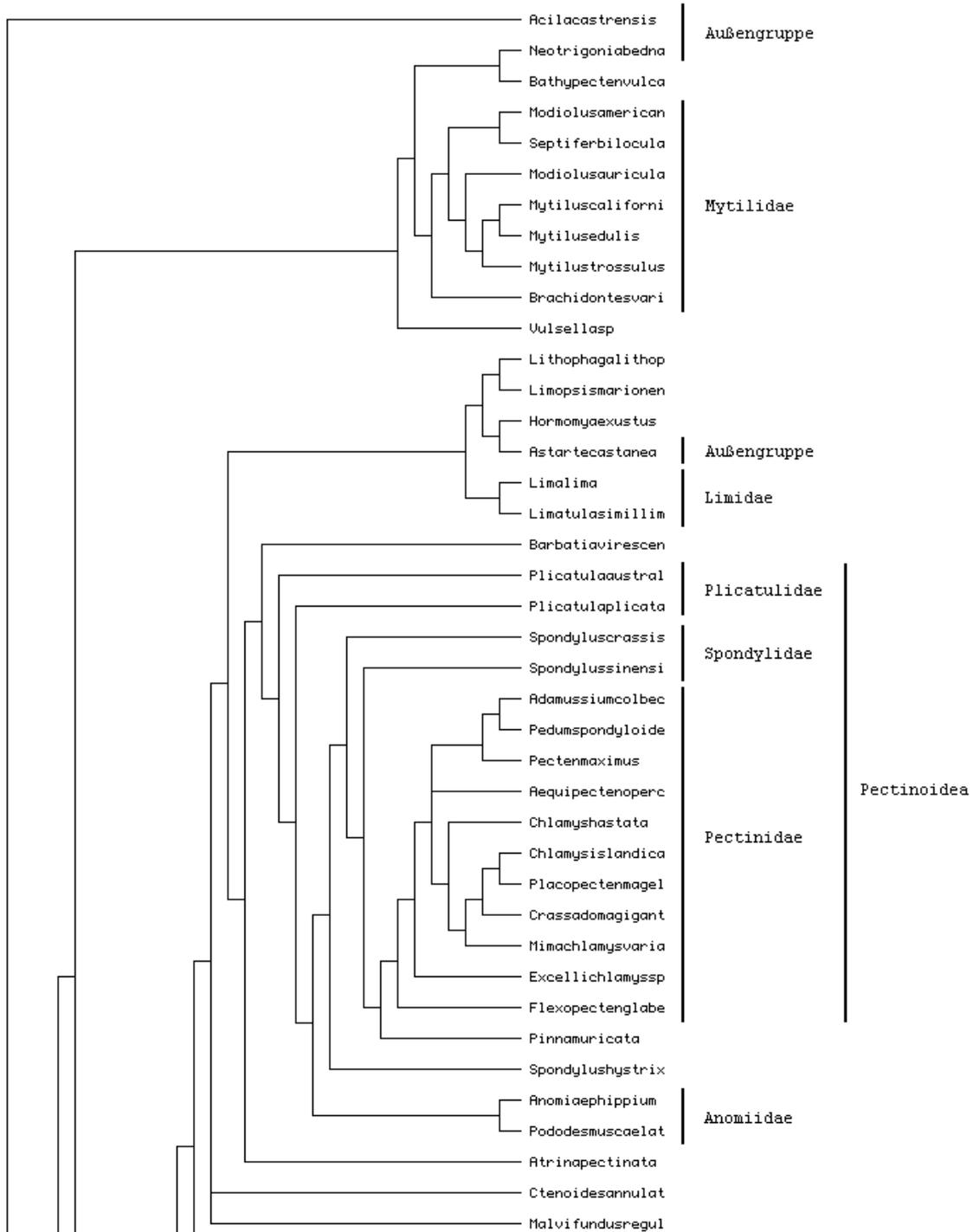


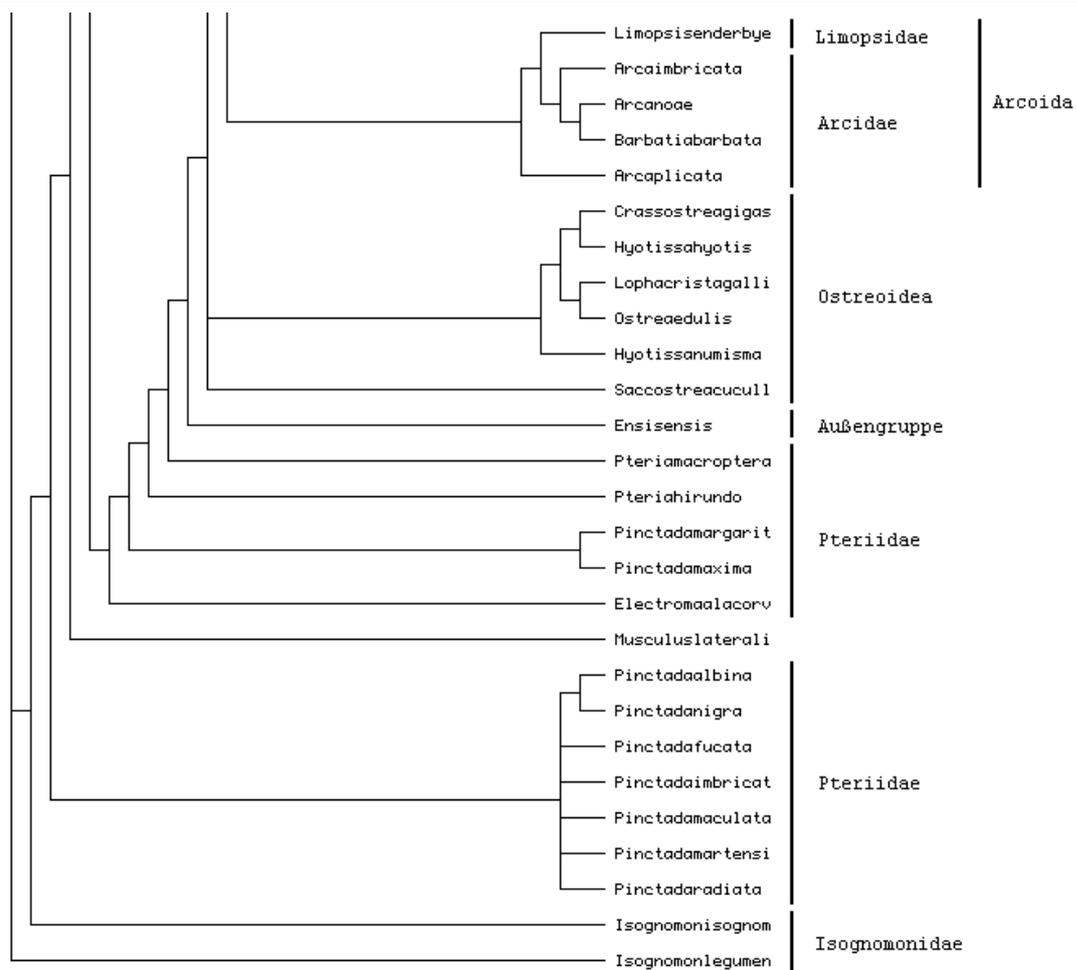
7.1.2 Stammbäume Pteriomorphia

7.1.2.1 Pteriomorphia MFE Single State

7.1.2.1.1 Differenziert gewichtet nach Größe

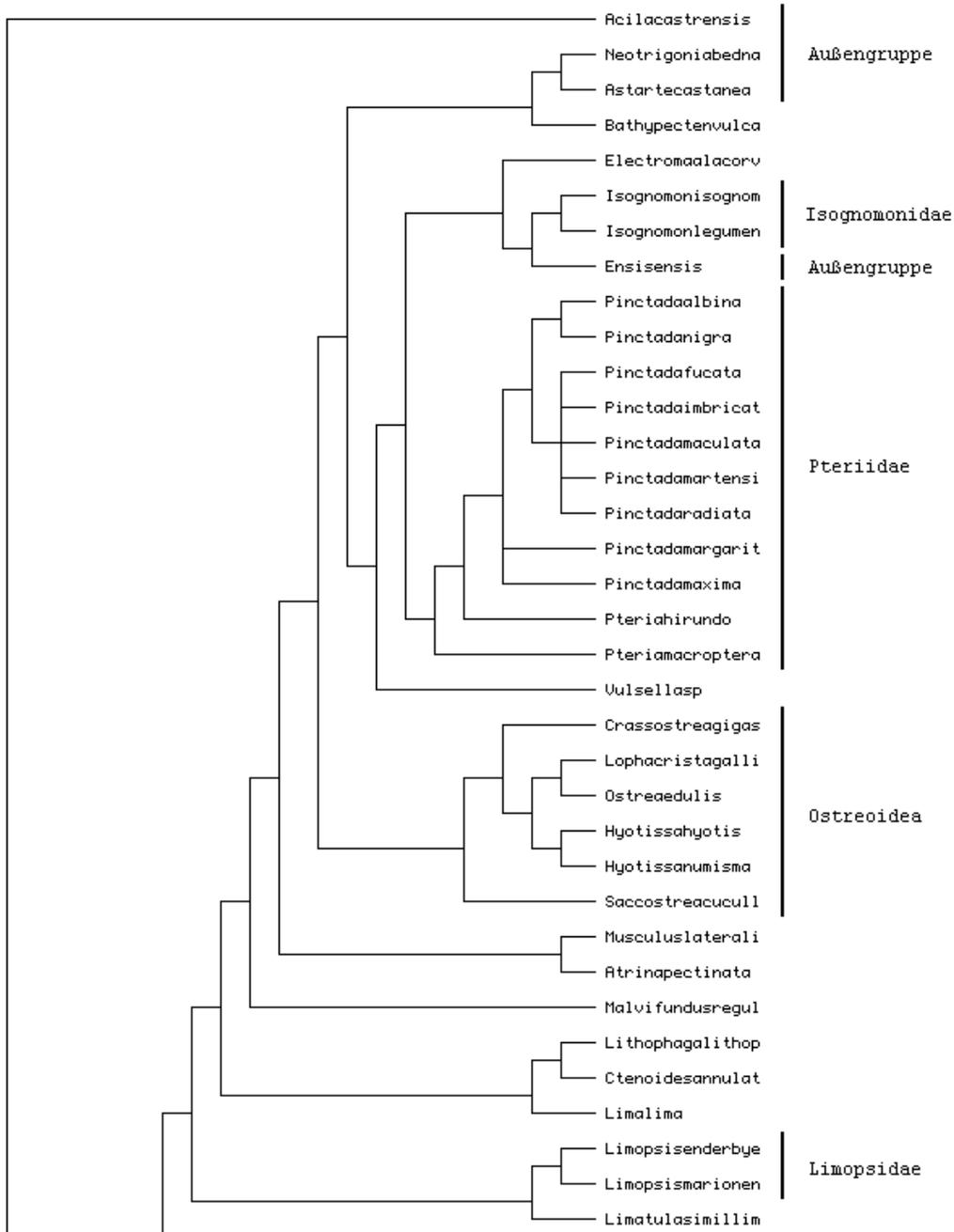
Parsimony Tree für Pteriomorphia nach MFE
differenziert gewichtet nach Größe, single-state Merkmale

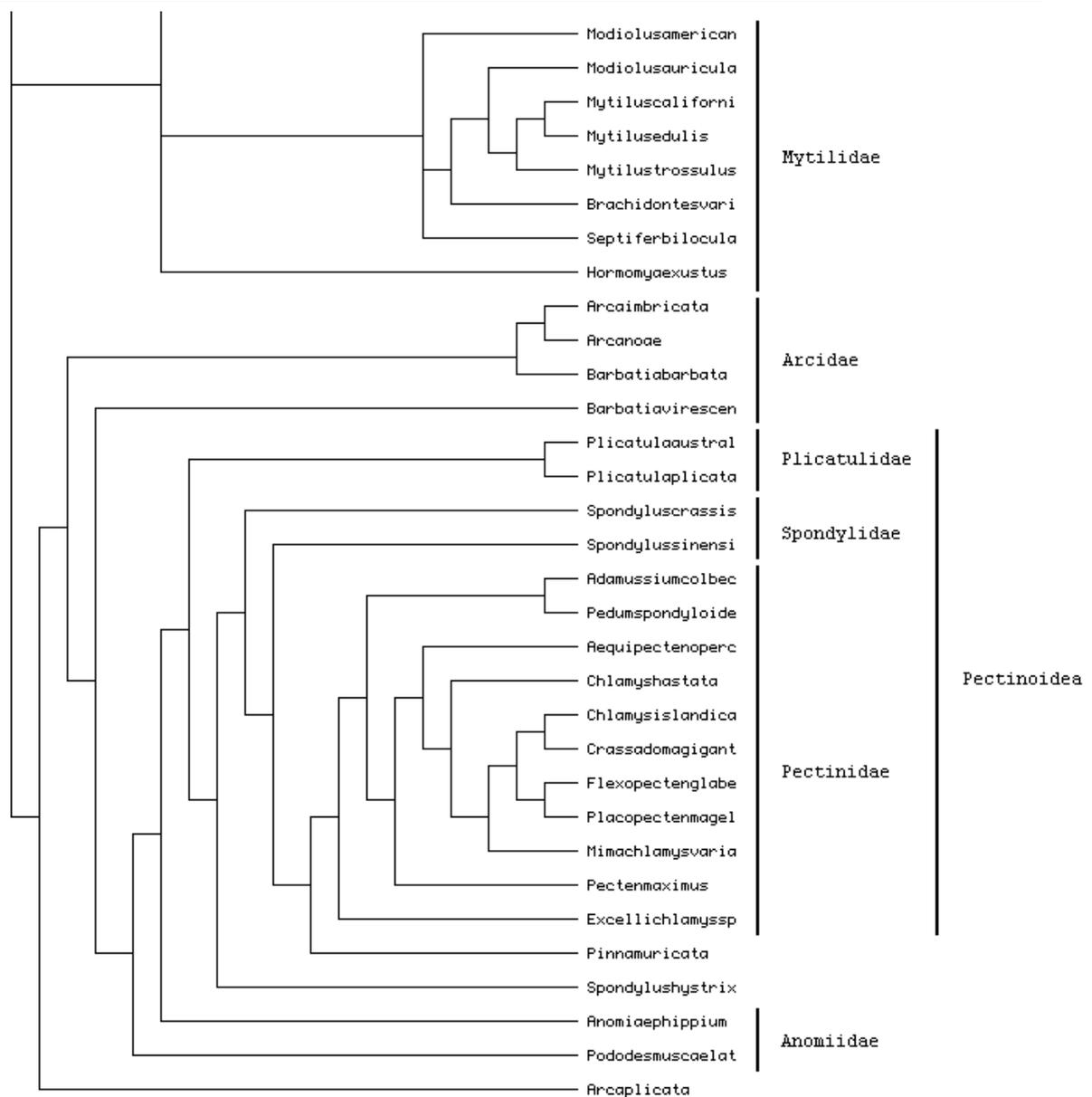




7.1.2.1.2 Gleichgewichtet

Parsimony Tree für Pteriomorpha nach MFE
gleichgewichtete, single-state Merkmale

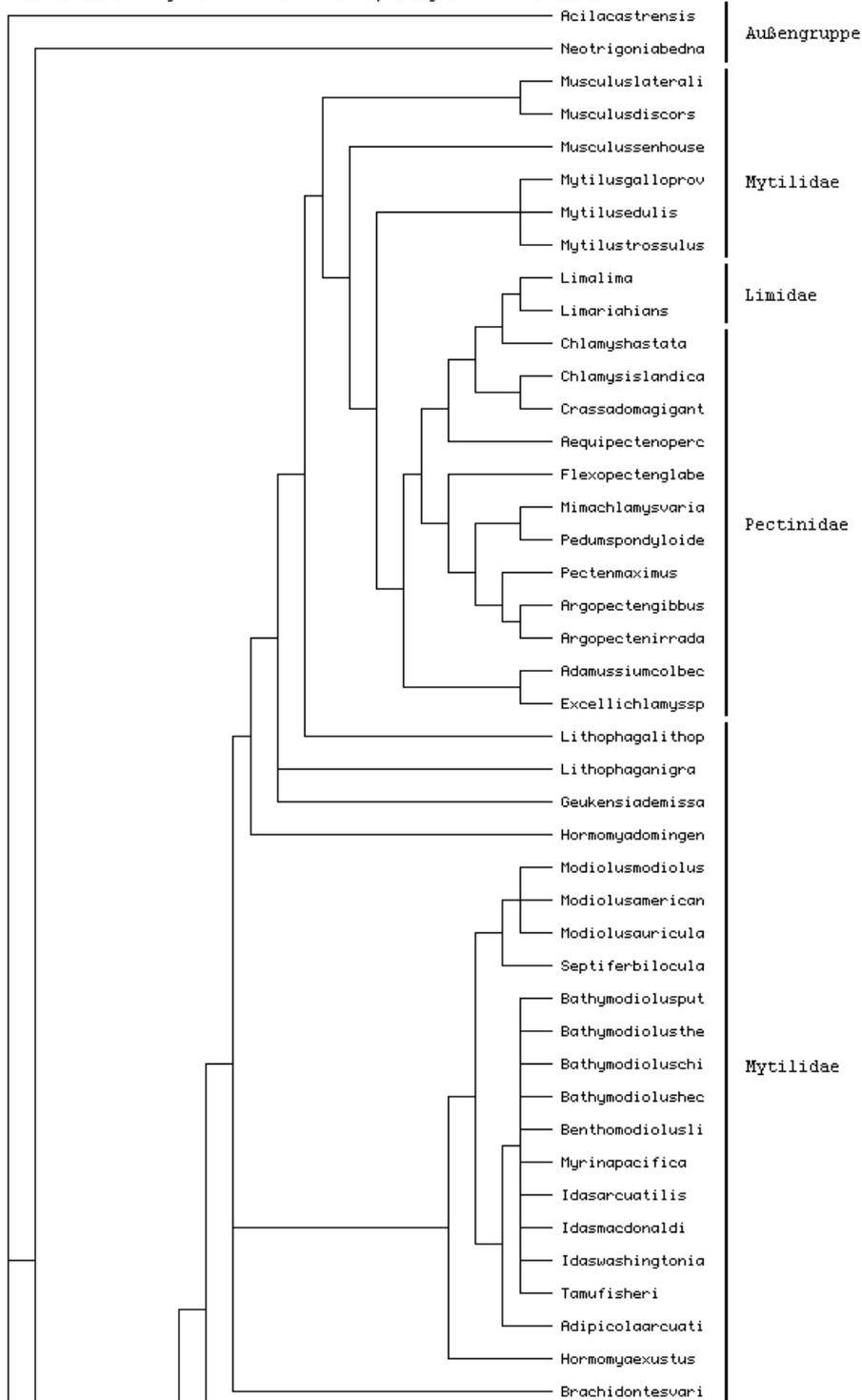


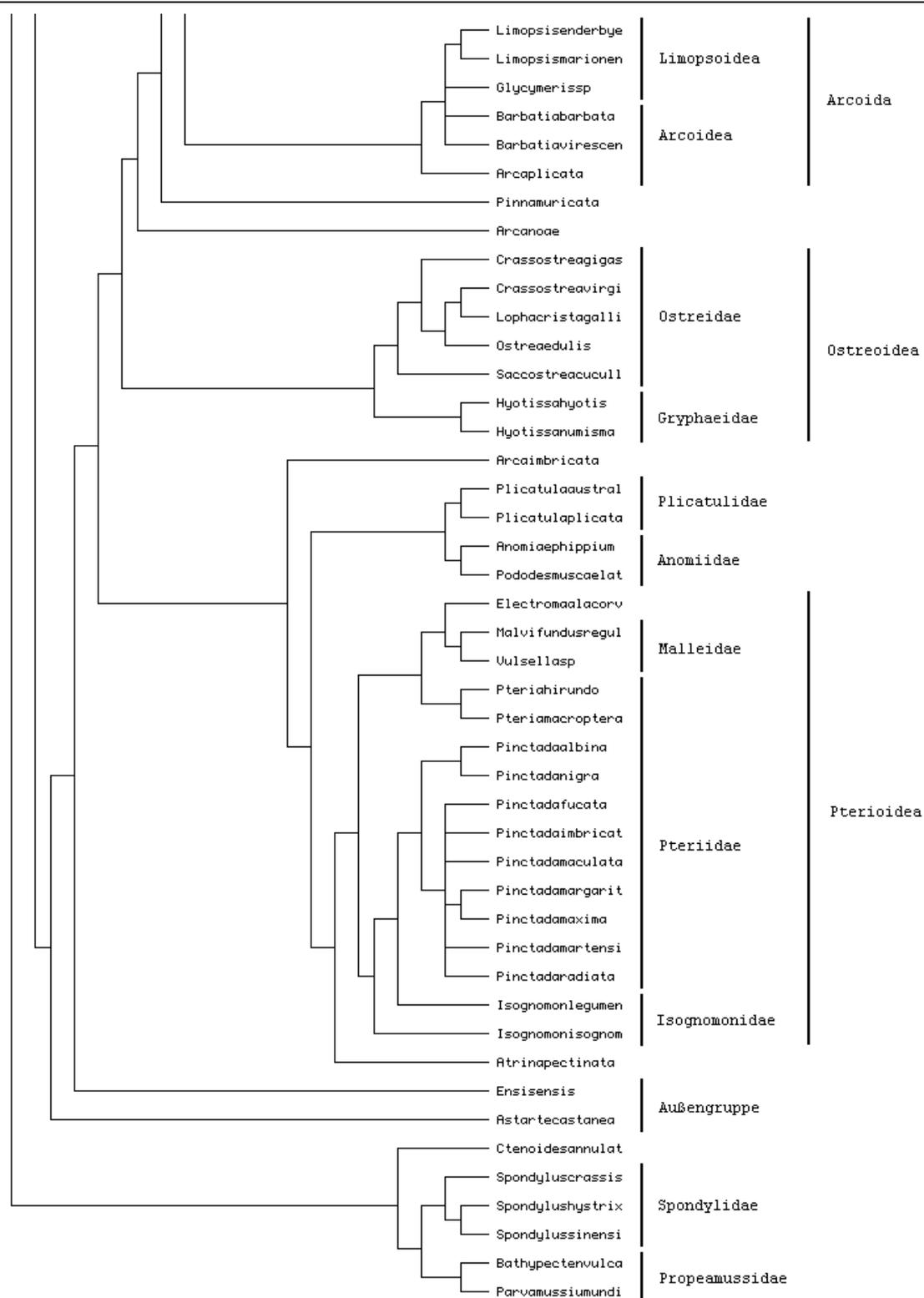


7.1.2.2 Pteriomorphia NNC Single State

7.1.2.2.1 Differenziert gewichtet nach Größe

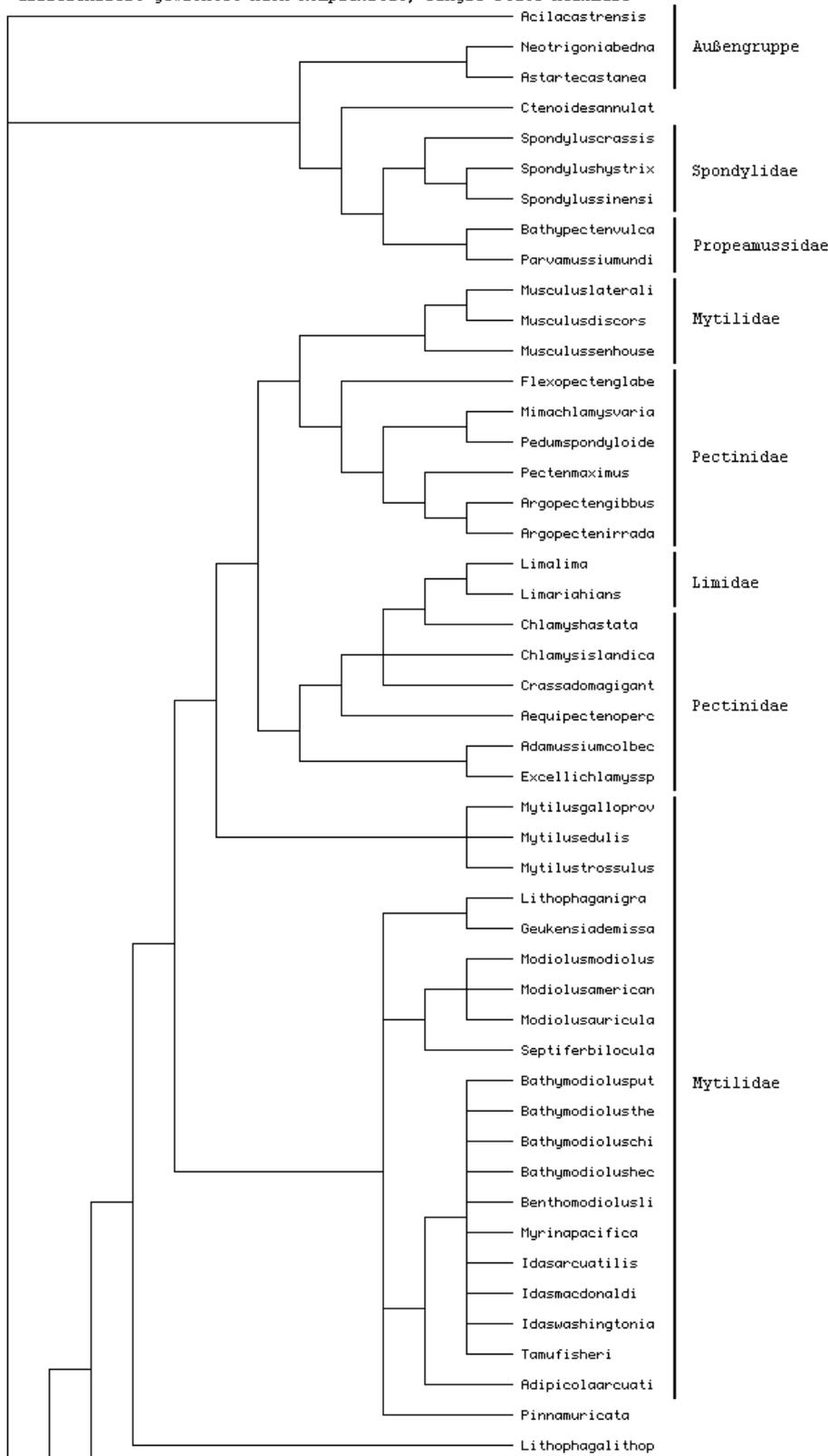
Parsimony Tree für Pteriomorphia nach NNC
differenziert gewichtet nach Größe, single-state Merkmale

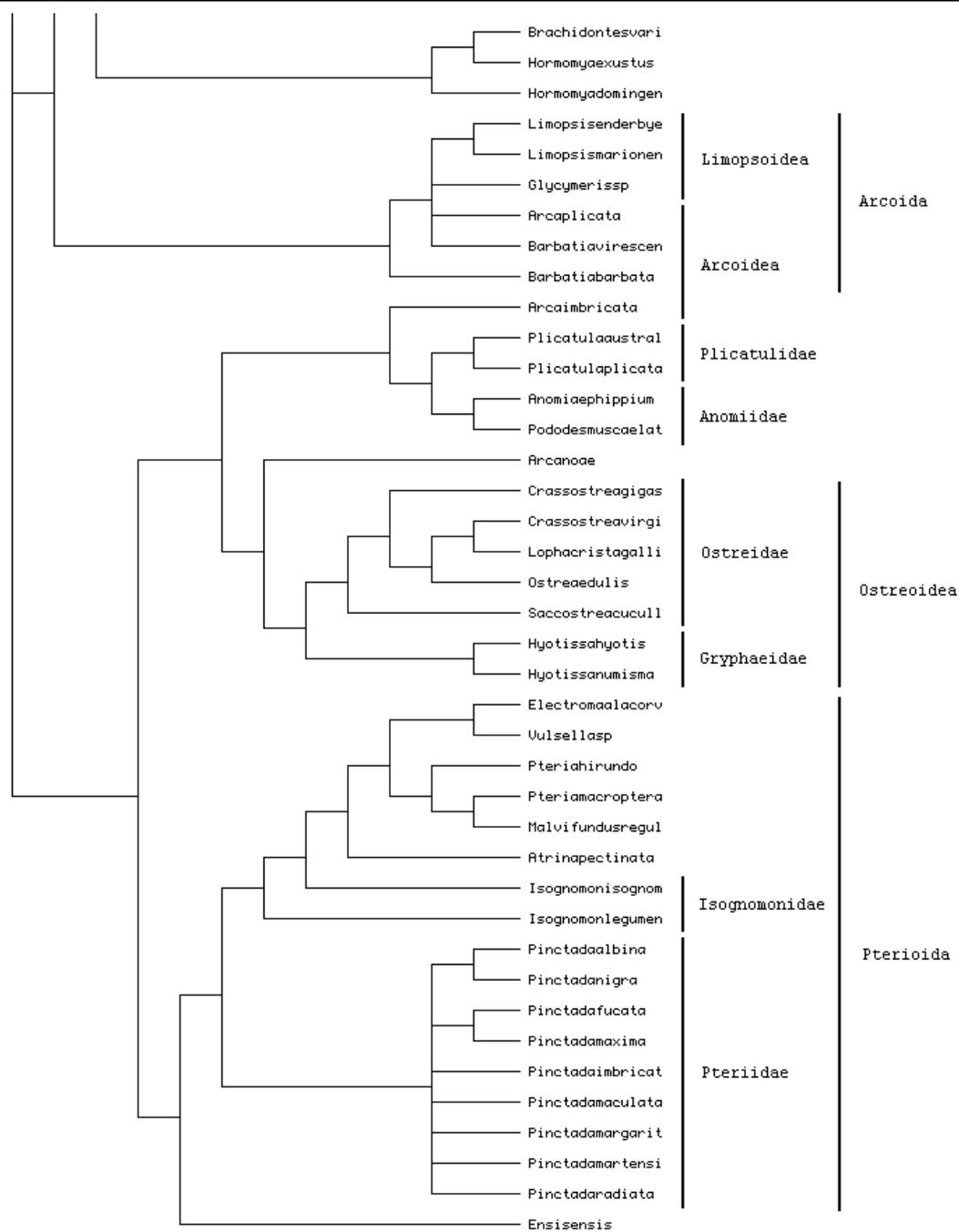




7.1.2.2.2 Differenziert gewichtet nach Komplexität

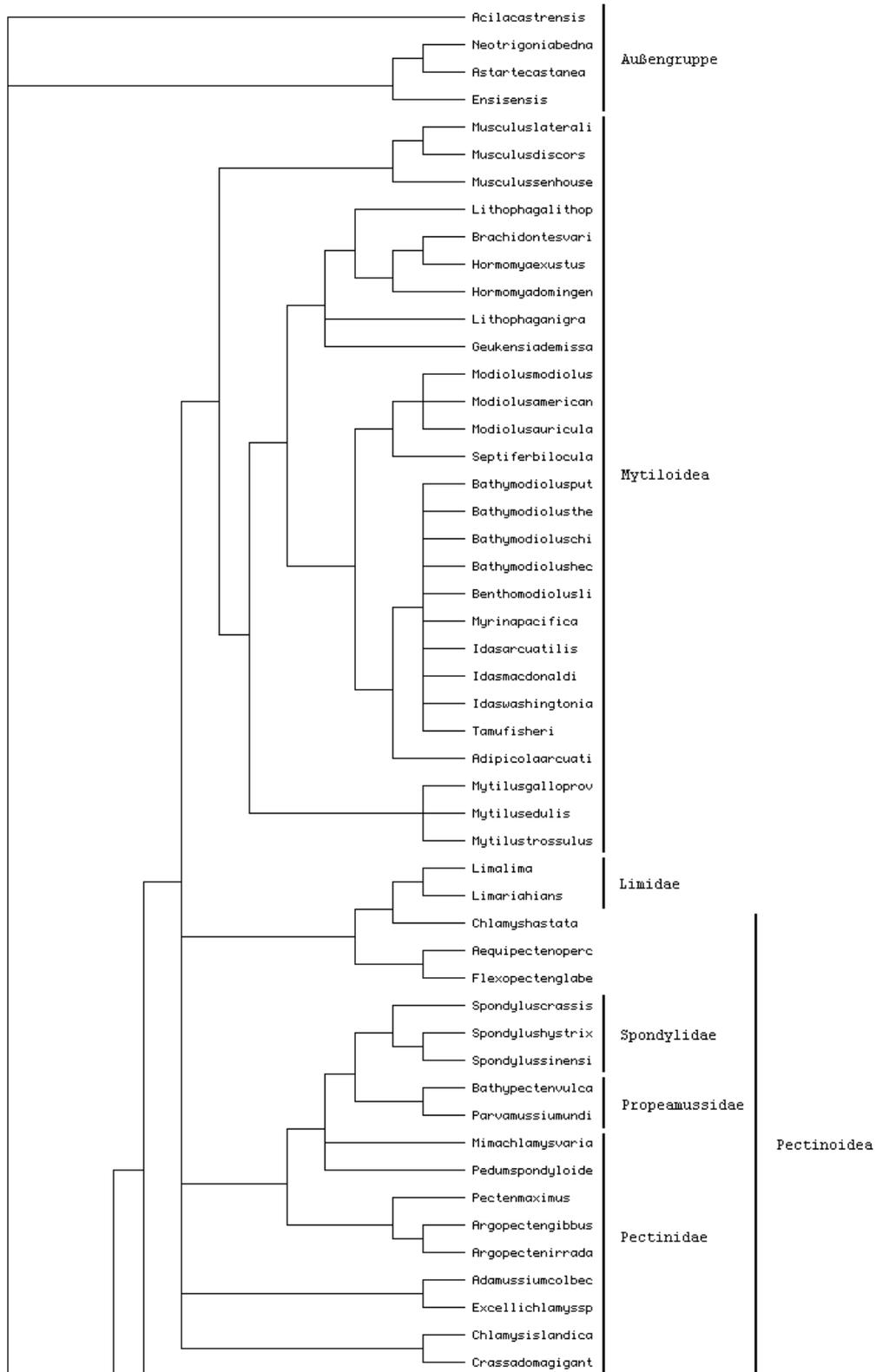
Parsimony Tree für Pteriomorpha nach NMC
differenziert gewichtet nach Komplexität, single-state Merkmale

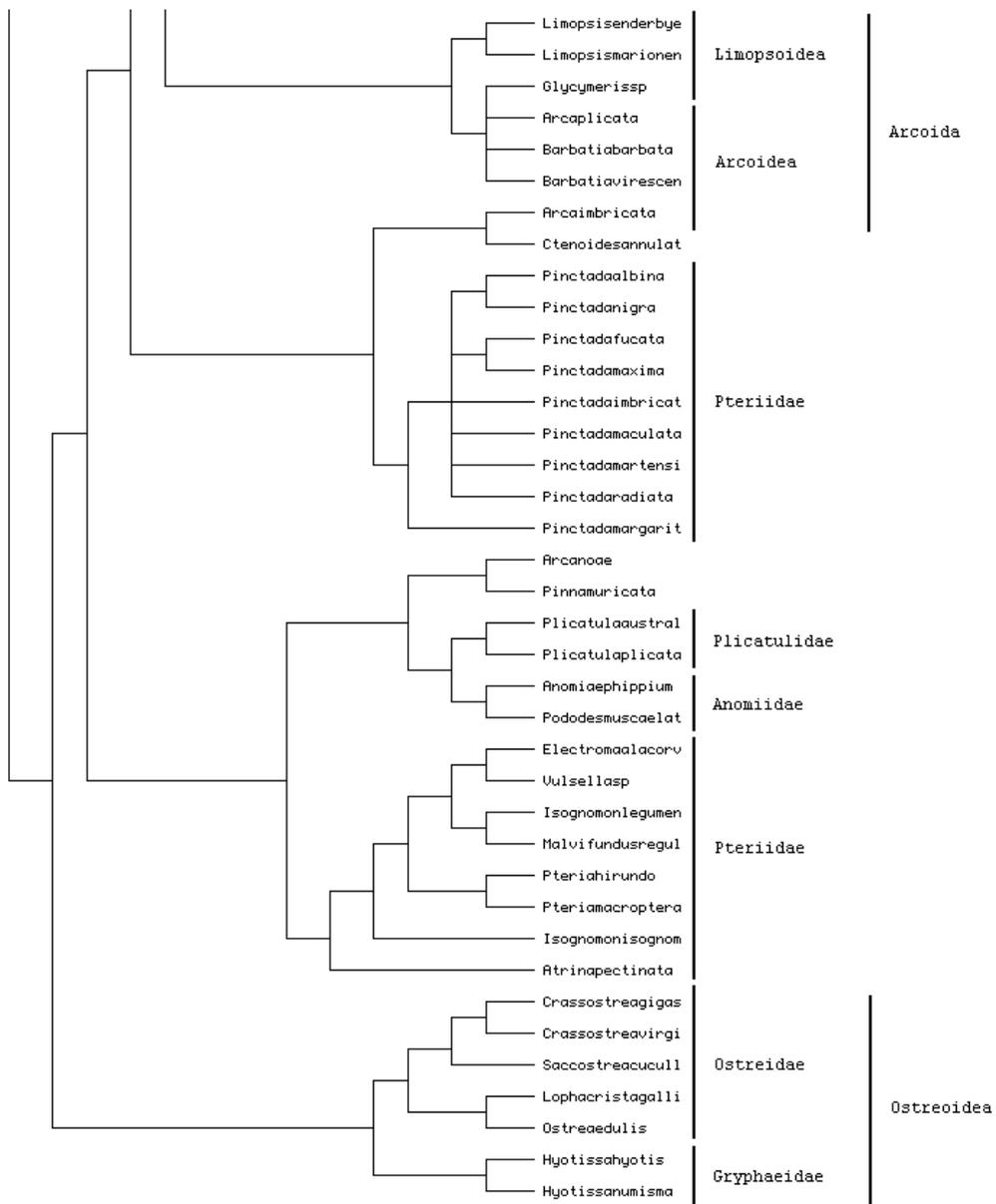




7.1.2.2.3 Gleichgewichtet

Parsimony Tree für Pteriomorpha nach MMC
gleichgewichtete, single-state Merkmale

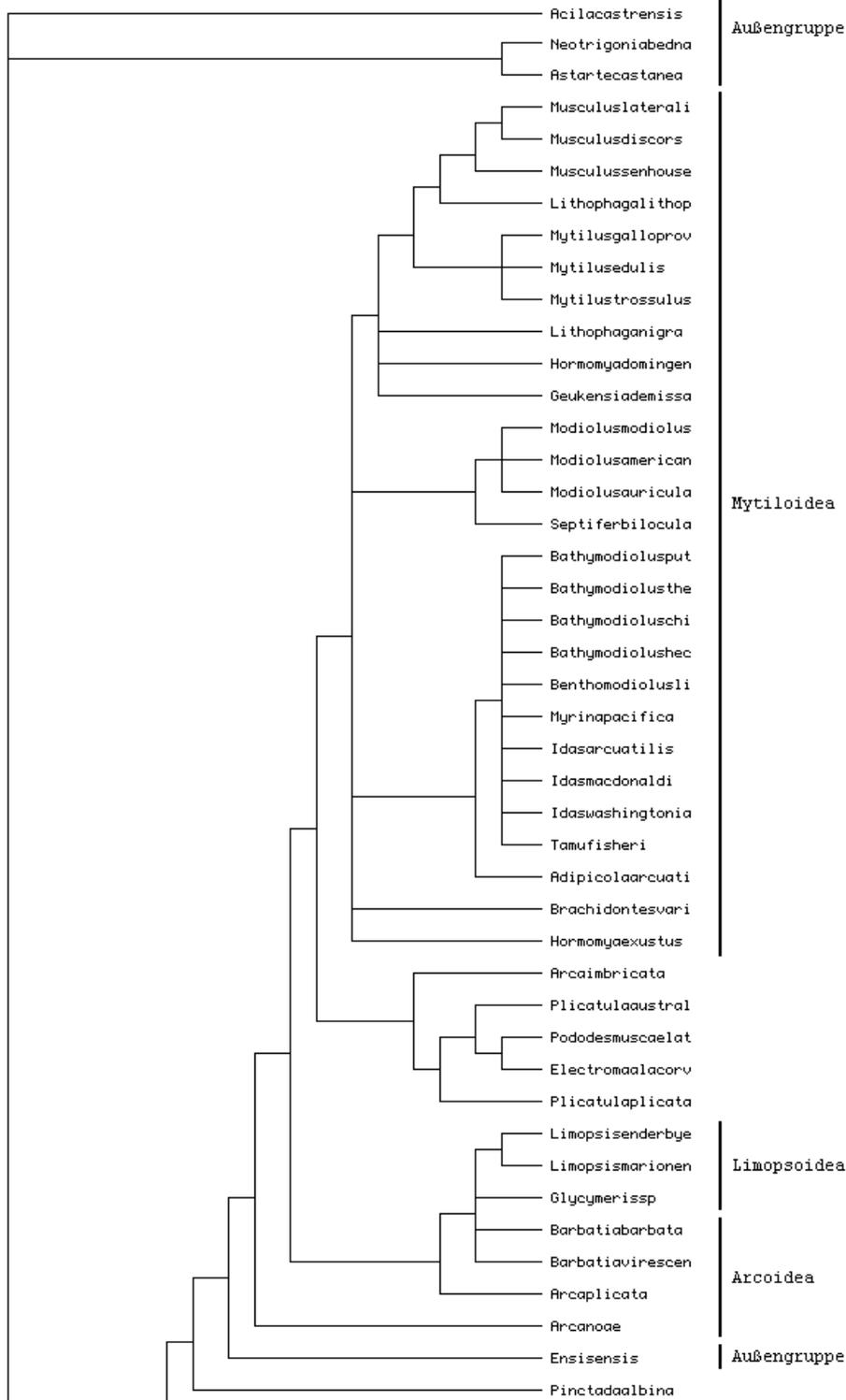


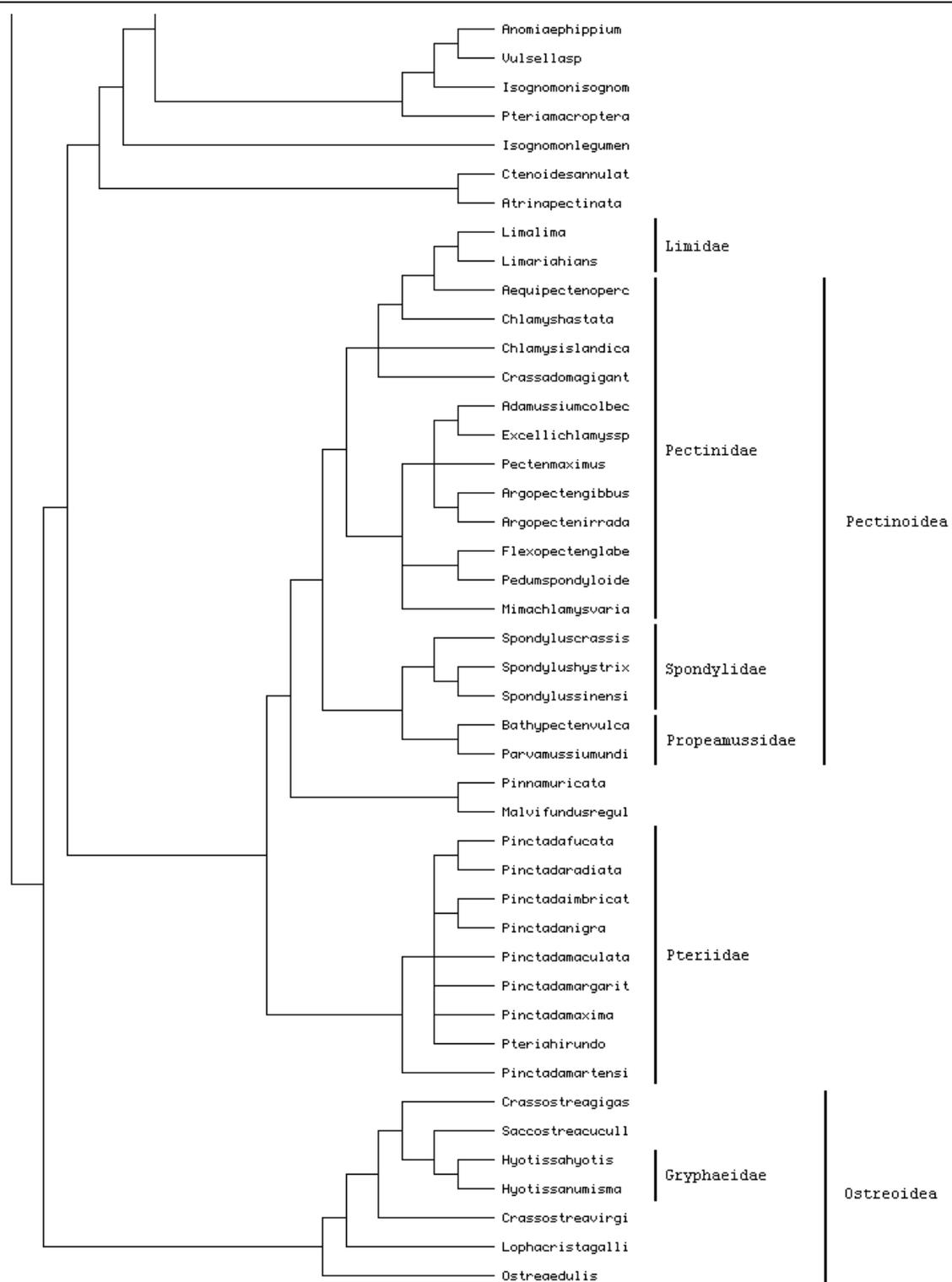


7.1.2.3 Pteriomorphia NNC Multi State

7.1.2.3.1 Differenziert gewichtet nach Größe

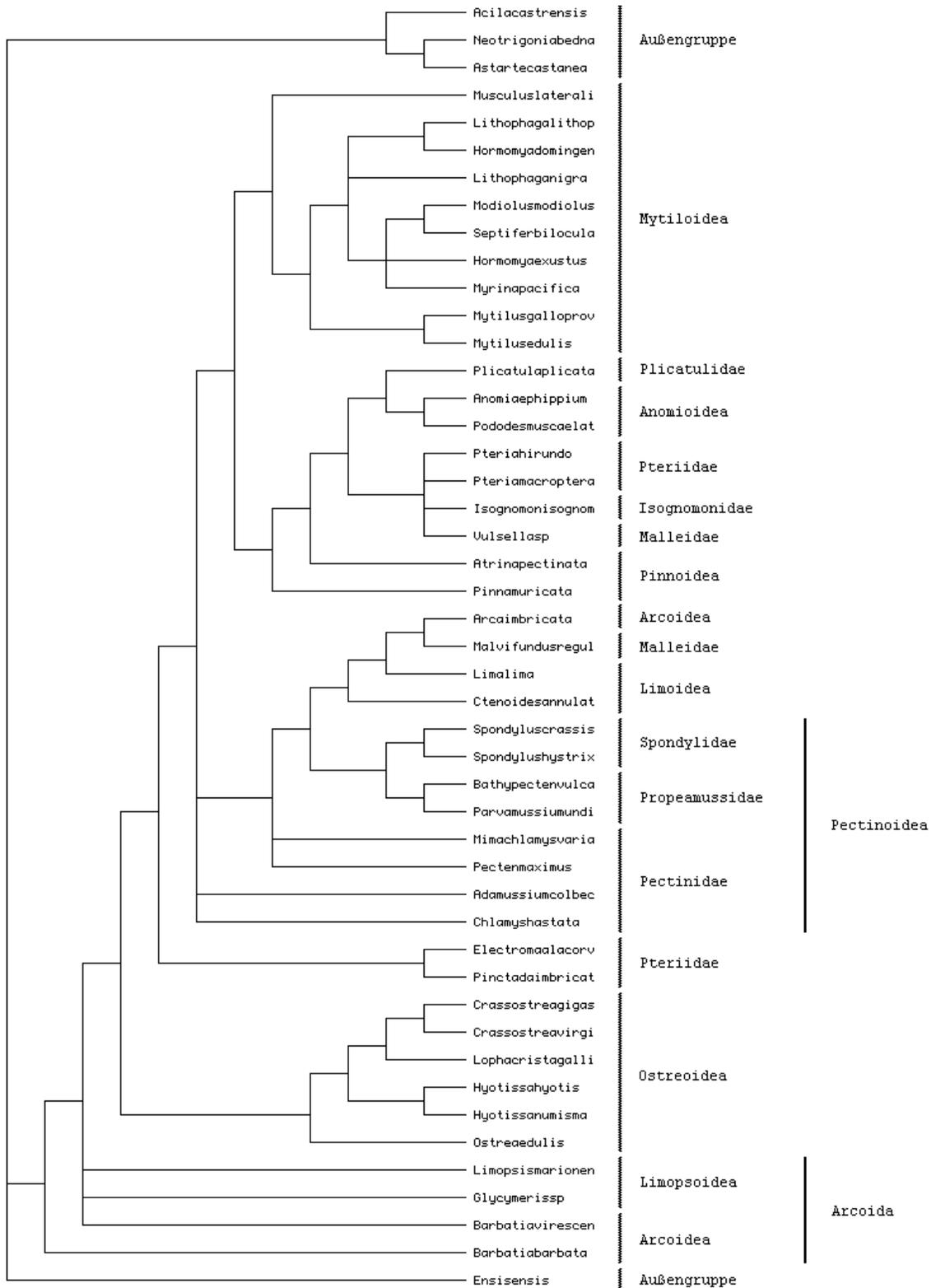
Parsimony Tree für Pteriomorphia nach NNC
differenziert gewichtet nach Größe, multi-state Merkmale





7.1.2.3.2 Gleichgewichtet

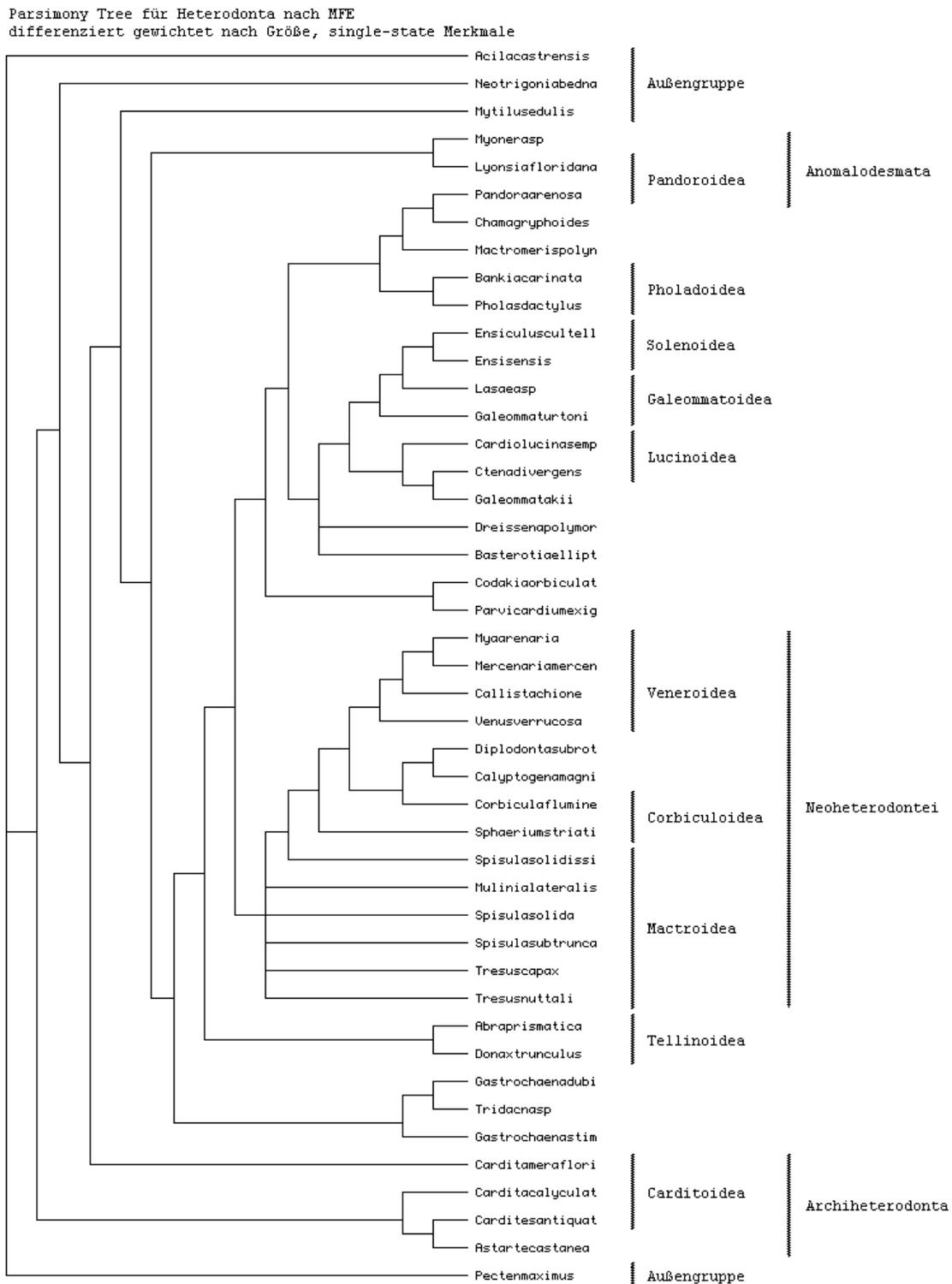
Parsimony Tree für Pteriomorpha nach NMC
gleichgewichtete, multi-state Merkmale



7.1.3 Stammbäume Heterodonta

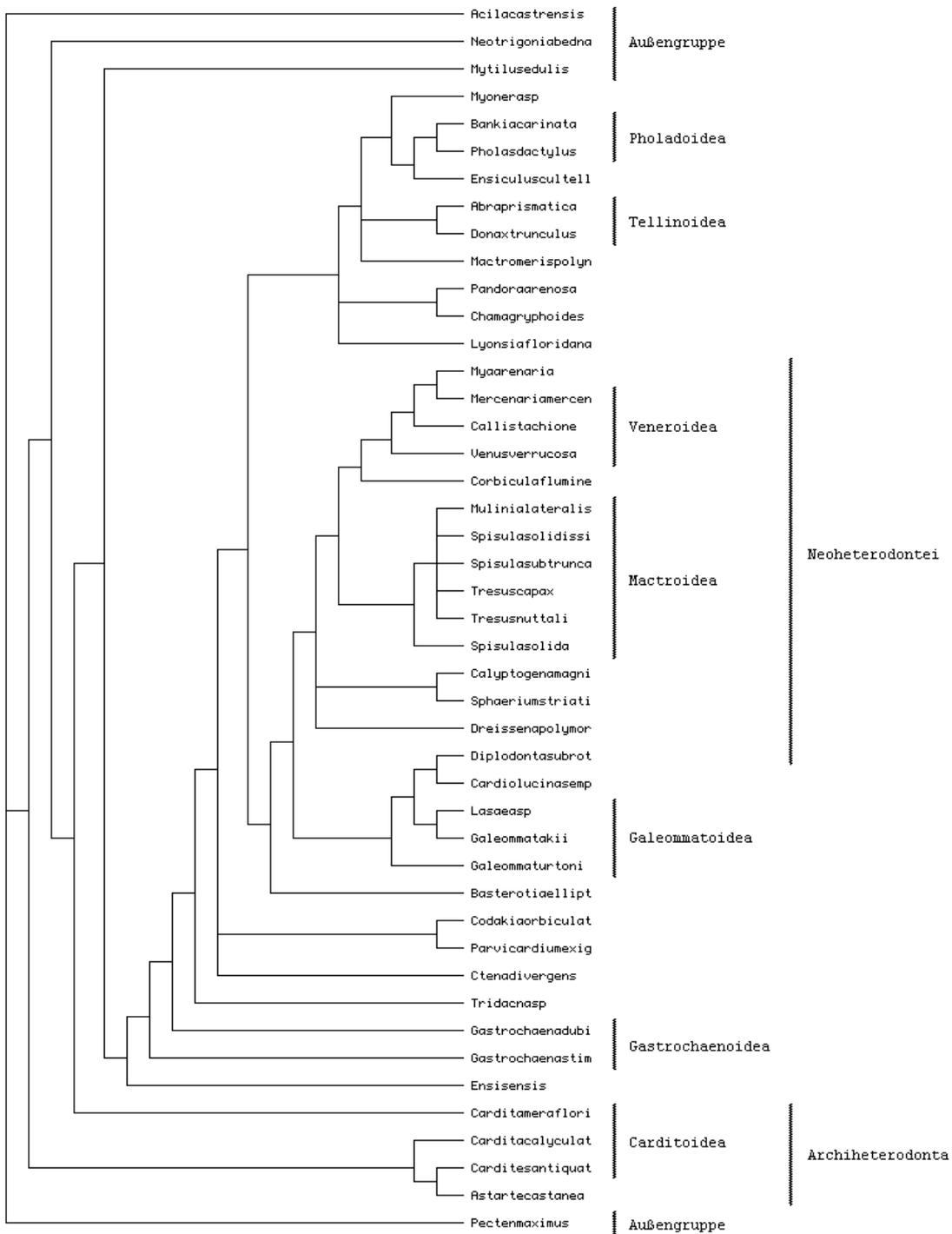
7.1.3.1 Heterodonta MFE Single State

7.1.3.1.1 Differenziert gewichtet nach Größe



7.1.3.1.2 Gleichgewichtet

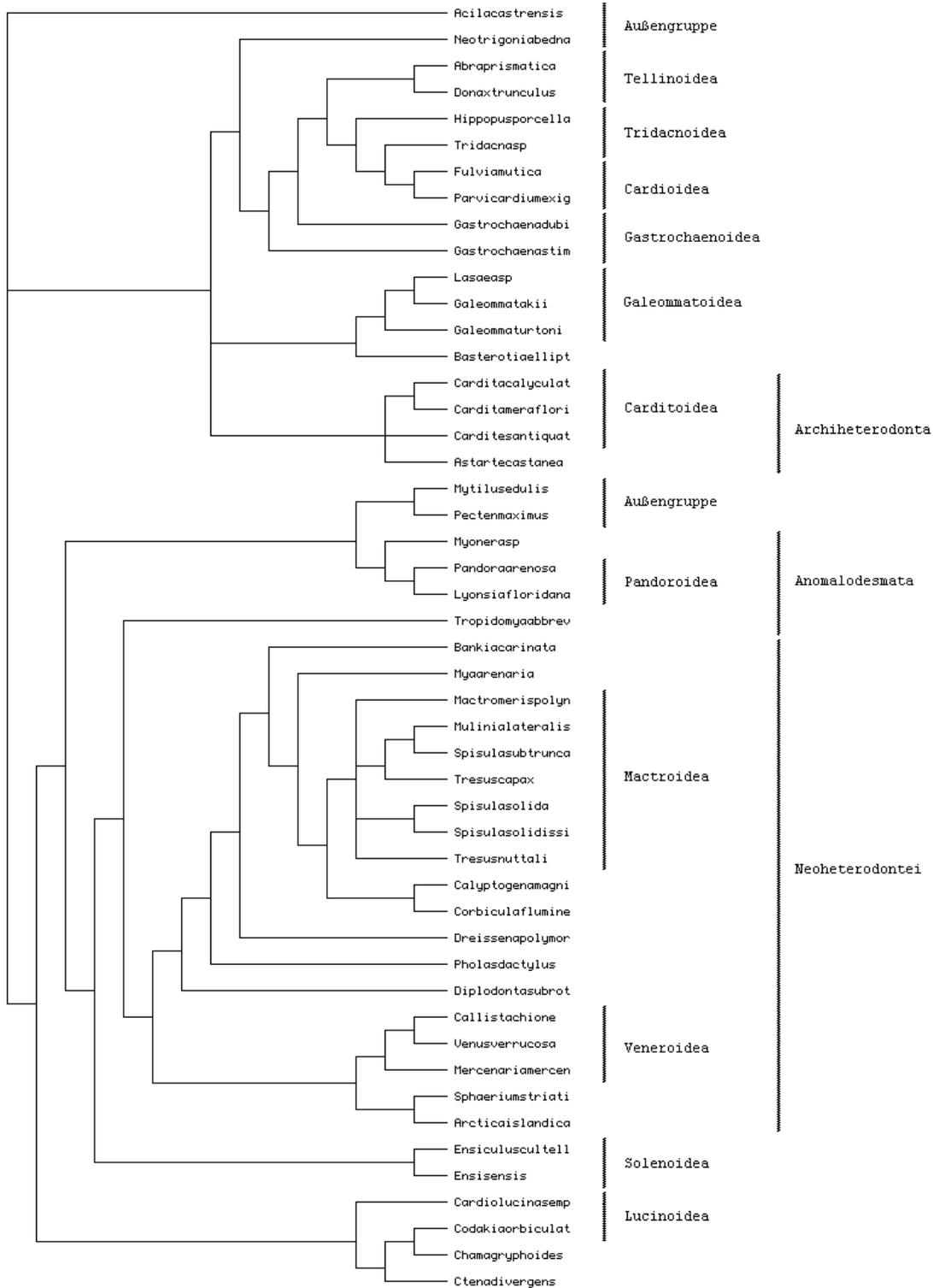
Parsimony Tree für Heterodonta nach MFE
gleichgewichtete, single-state Merkmale



7.1.3.2 Heterodonta NNC Single State

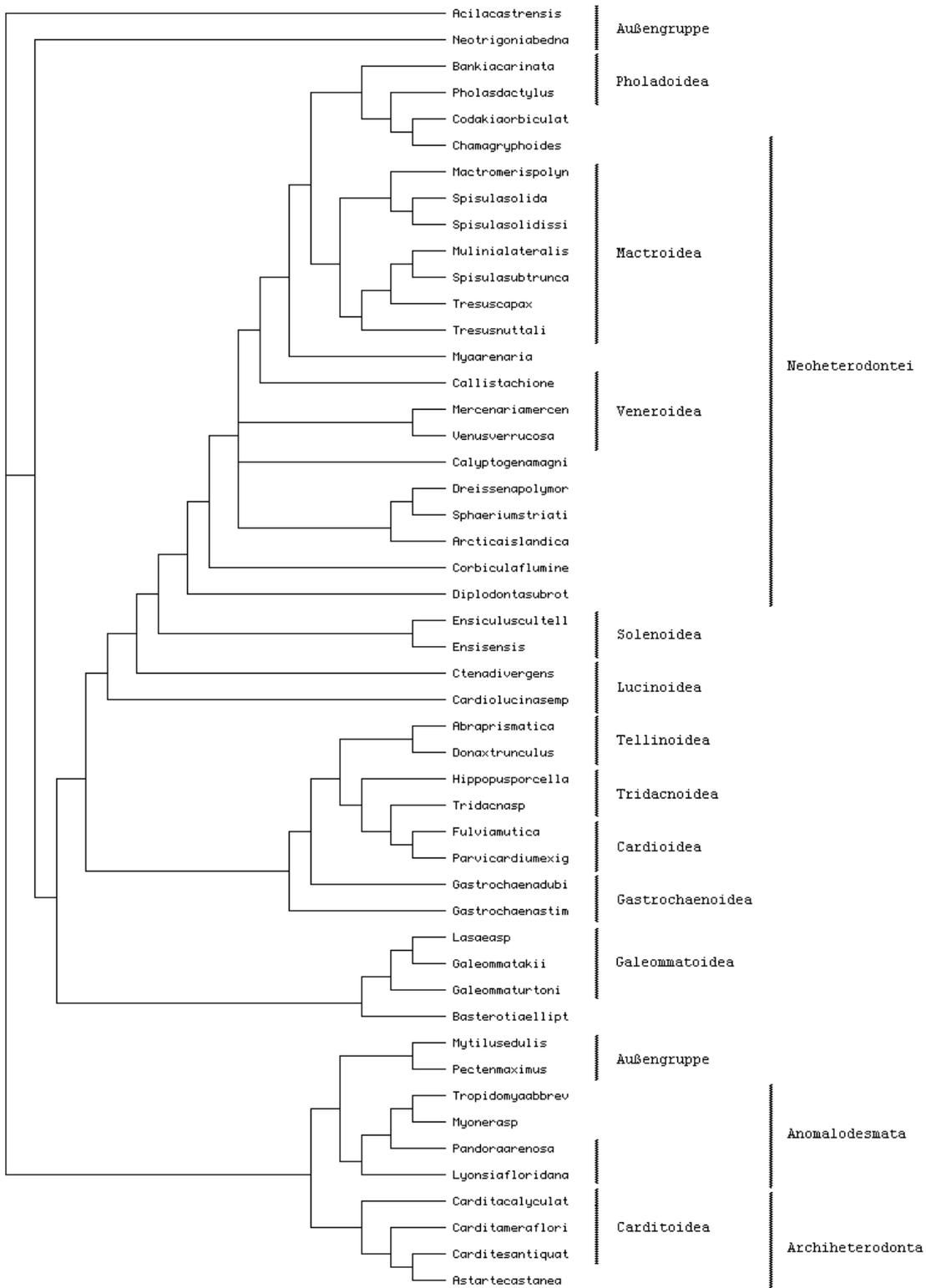
7.1.3.2.1 Differenziert gewichtet nach Größe

Parsimony Tree für Heterodonta nach NNC
differenziert gewichtet nach Größe, single-state Merkmale



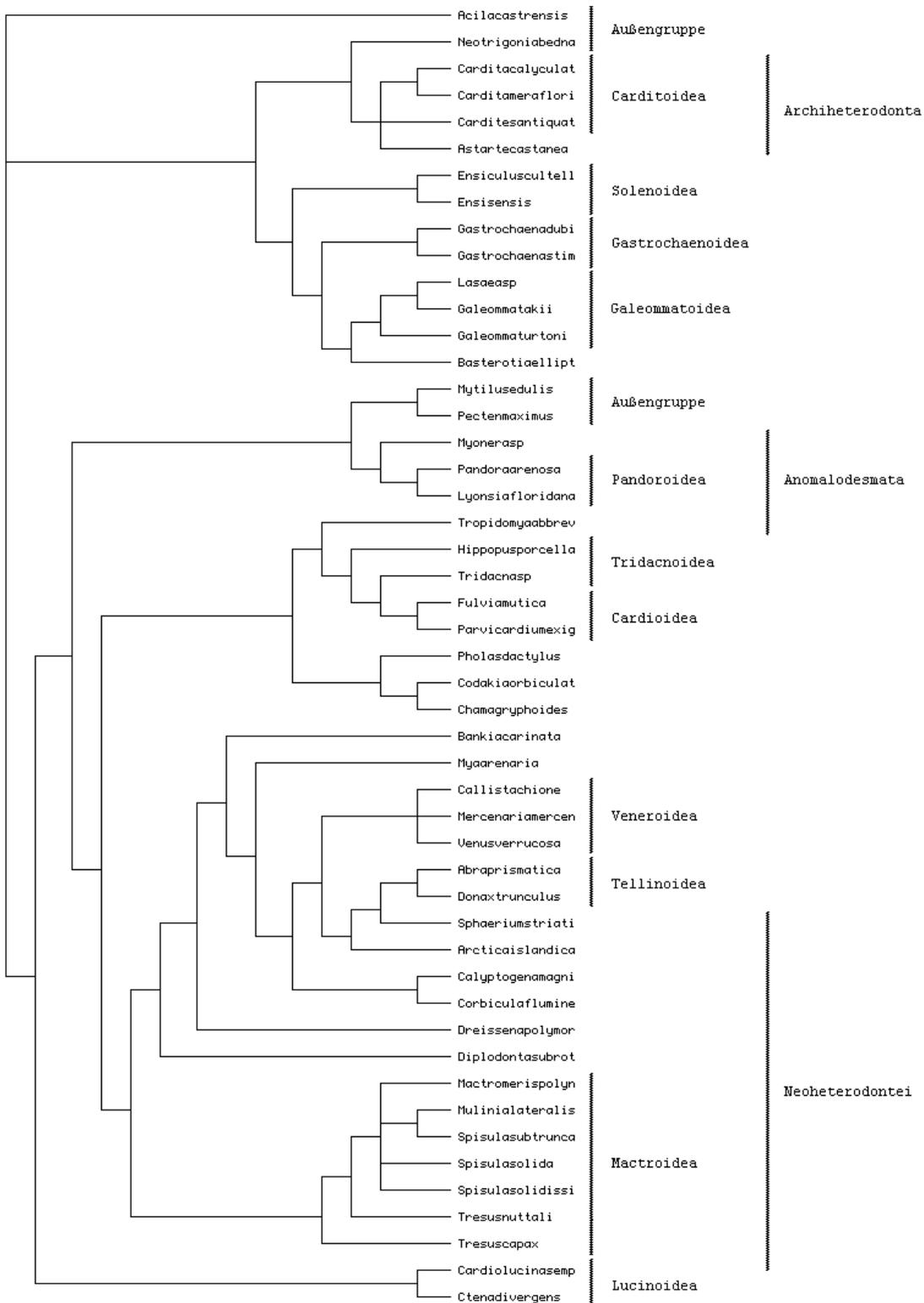
7.1.3.2.2 Differenziert gewichtet nach Komplexität

Parsimony Tree für Heterodonta nach NNC
differenziert gewichtet nach Komplexität, single-state Merkmale



7.1.3.2.3 Gleichgewichtet

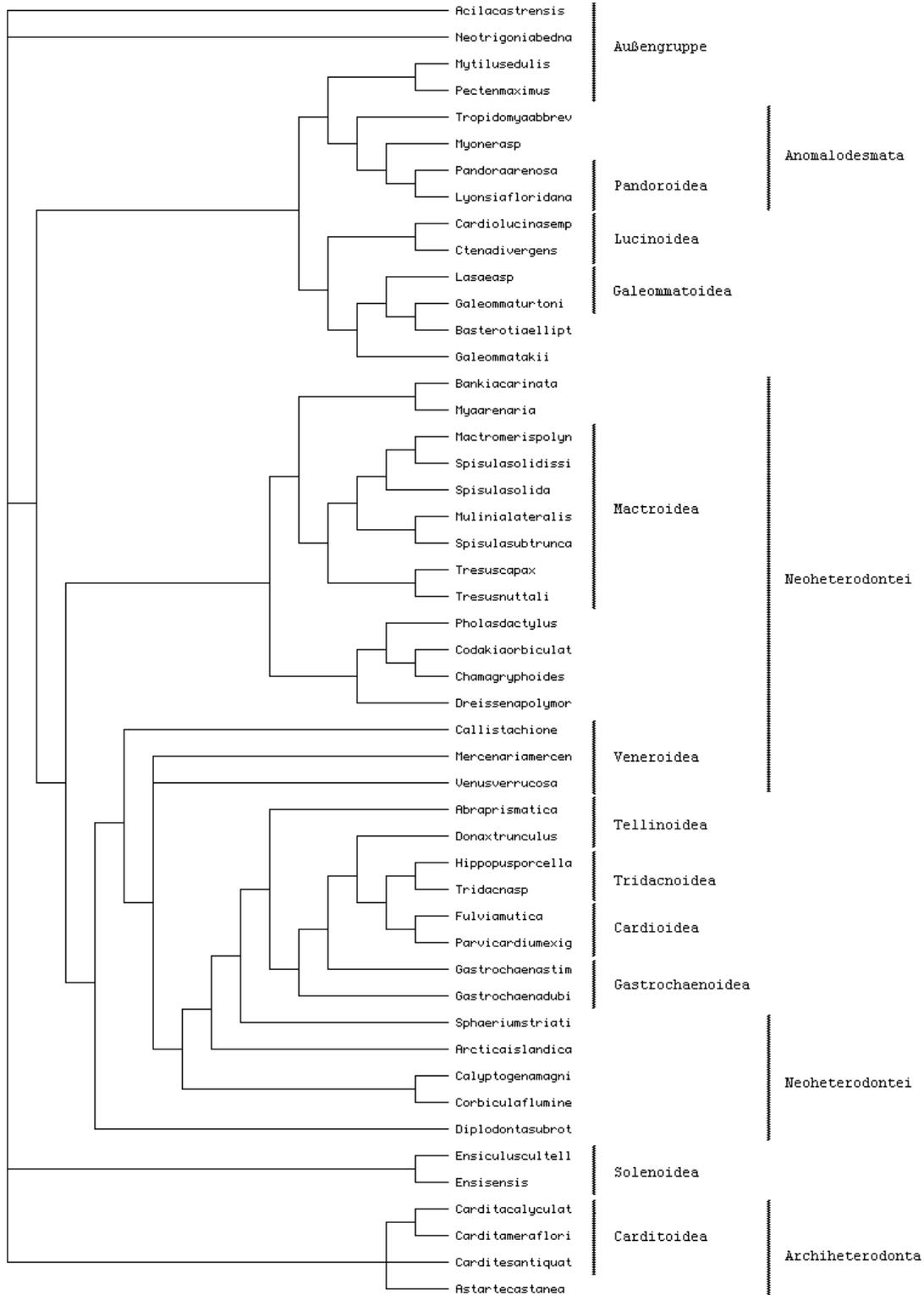
Parsimony Tree für Heterodonta nach NNC
gleichgewichtete, single-state Merkmale



7.1.3.3 Heterodonta NNC Multi State

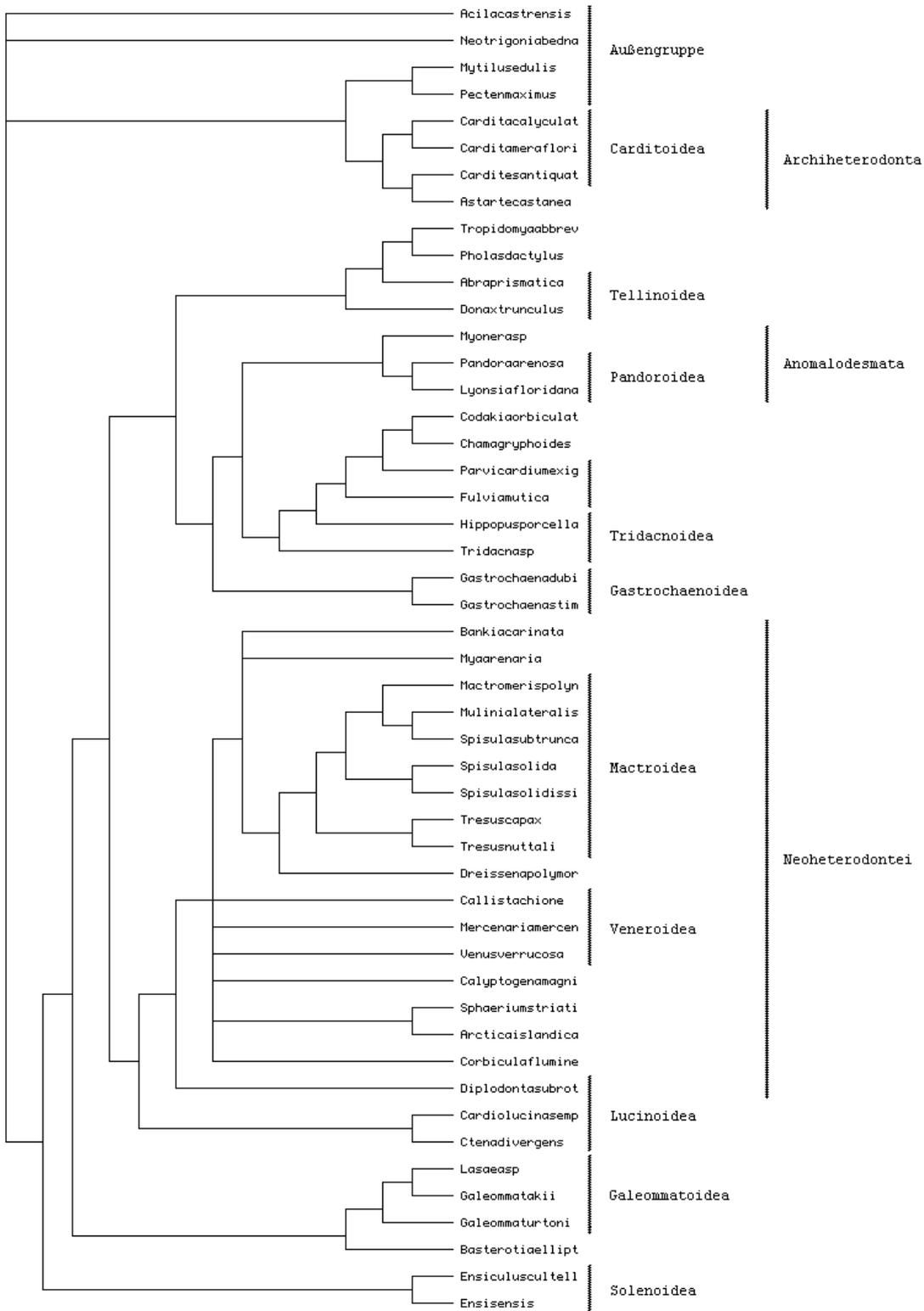
7.1.3.3.1 Differenziert gewichtet nach Größe

Parsimony Tree für Heterodonta nach NNC
differenziert gewichtet nach Größe, multi-state Merkmale



7.1.3.3.2 Gleichgewichtet

Parsimony Tree für Heterodonta nach NNC
gleichgewichtete, multi-state Merkmale

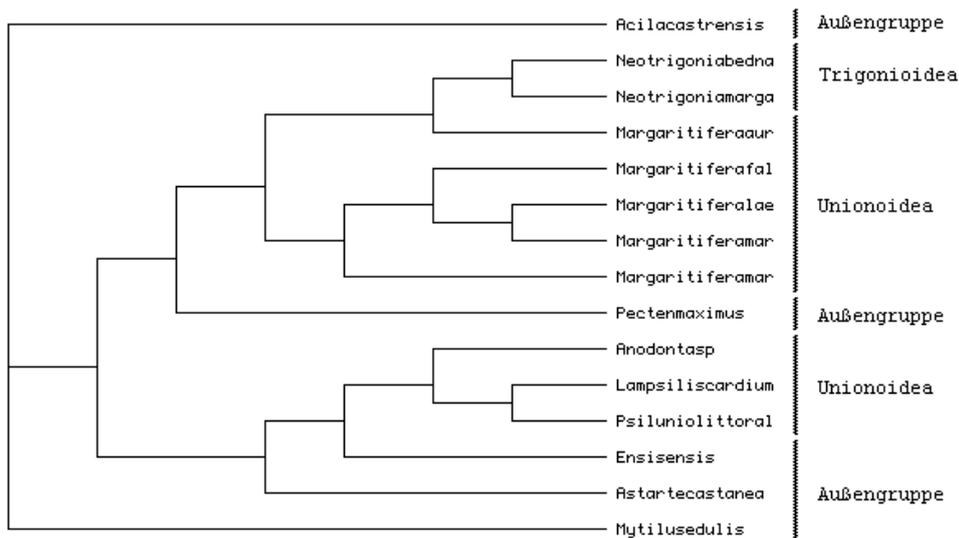


7.1.4 Stammbäume Palaeoheterodonta

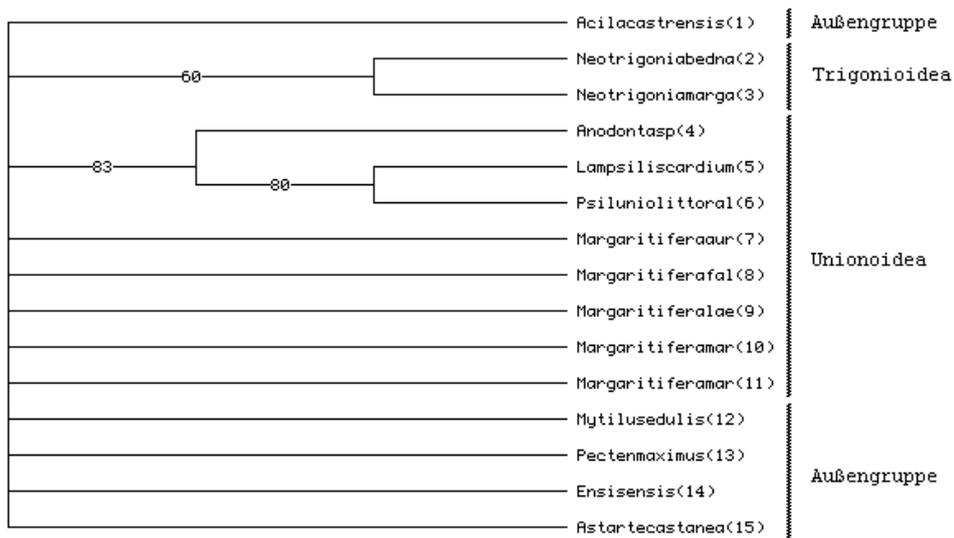
7.1.4.1 Palaeoheterodonta MFE Single State

7.1.4.1.1 Differenziert gewichtet nach Größe

Parsimony Tree für Palaeoheterodonta nach MFE
differenziert gewichtet nach Größe, single-state Merkmale

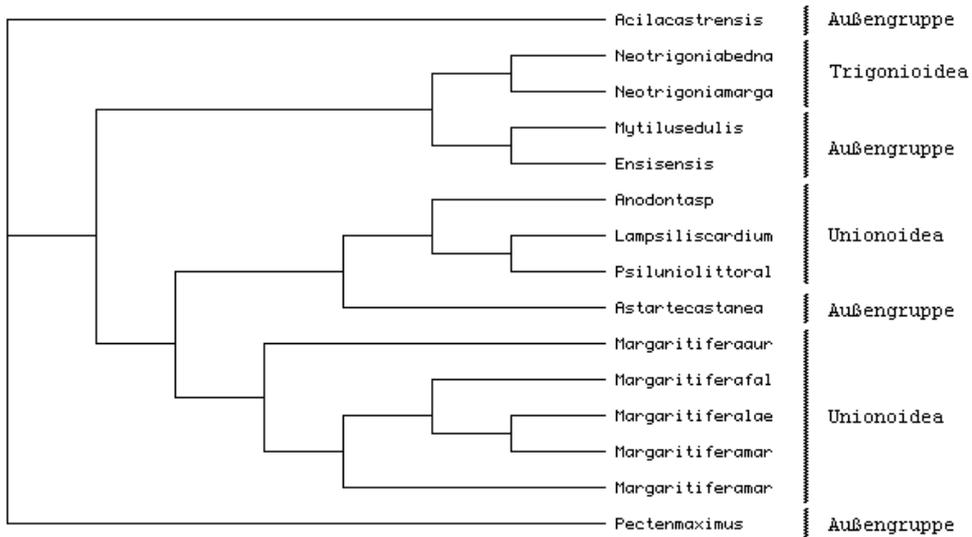


Parsimony Tree mit Bootstrap für Palaeoheterodonta nach MFE
differenziert gewichtet nach Größe, single-state Merkmale

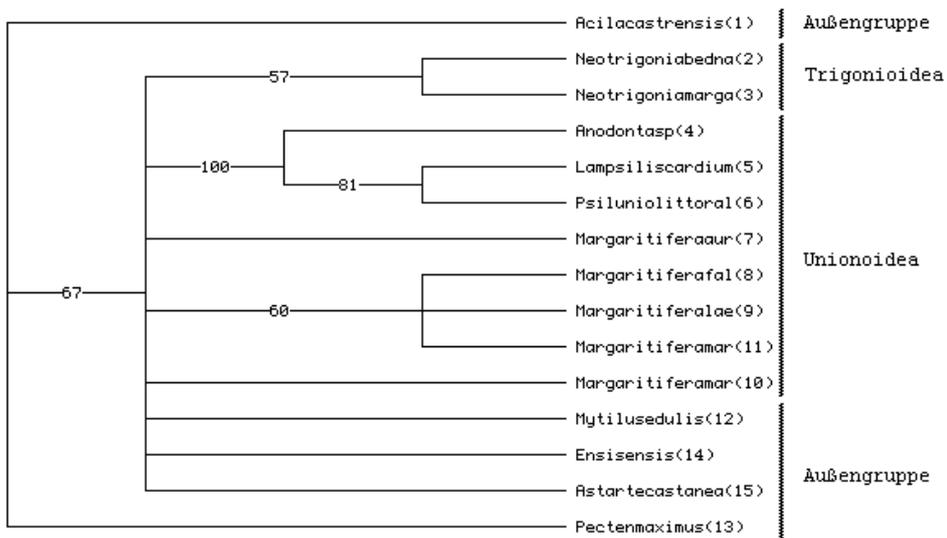


7.1.4.1.2 Gleichgewichtet

Parsimony Tree für Palaeoheterodonta nach MFE
gleichgewichtete, single-state Merkmale



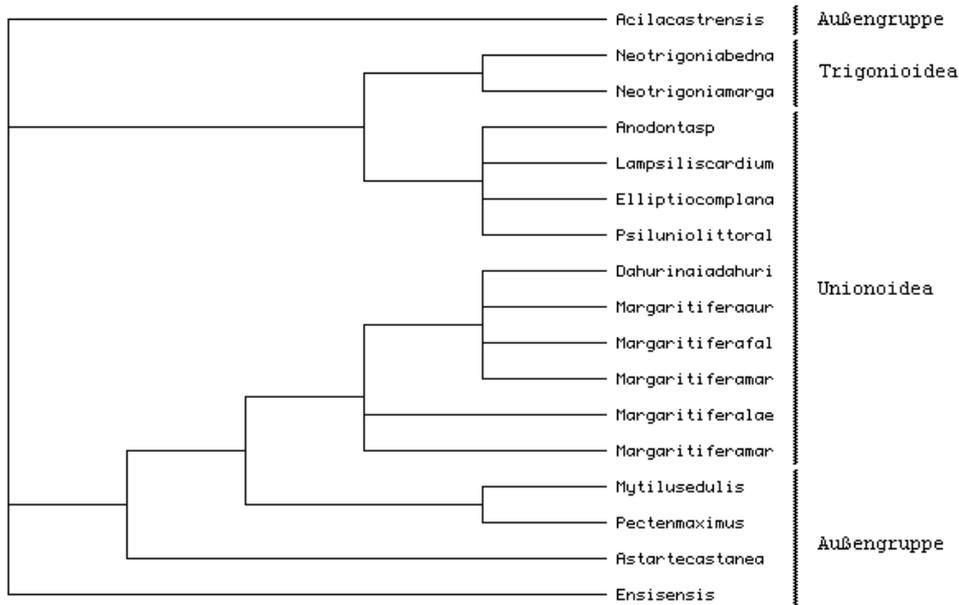
Parsimony Tree mit Bootstrap für Palaeoheterodonta nach MFE
gleichgewichtete, single-state Merkmale



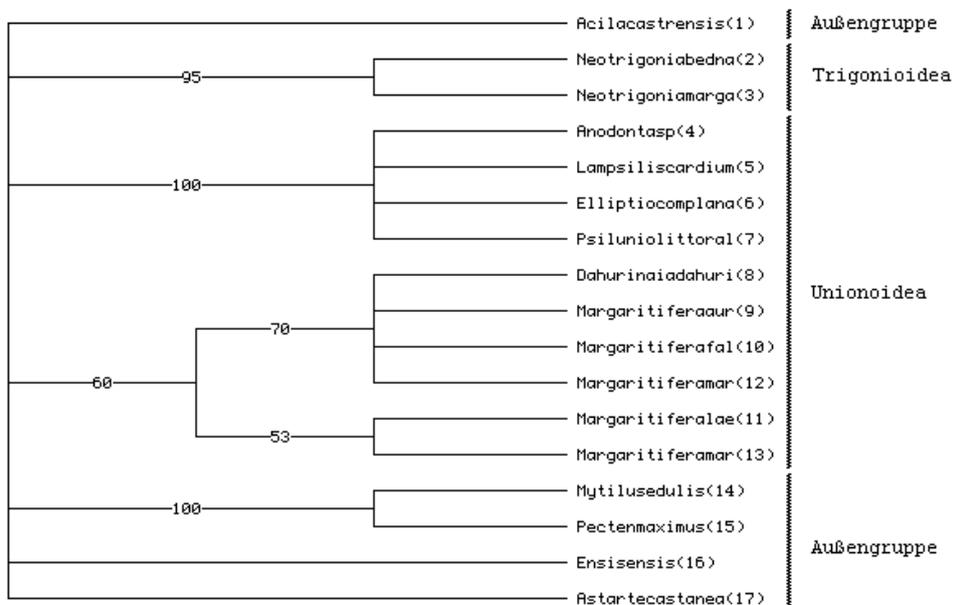
7.1.4.2 Palaeoheterodonta NNC Single State

7.1.4.2.1 Differenziert gewichtet nach Größe

Parsimony Tree für Palaeoheterodonta nach NNC
differenziert gewichtet nach Größe, single-state Merkmale

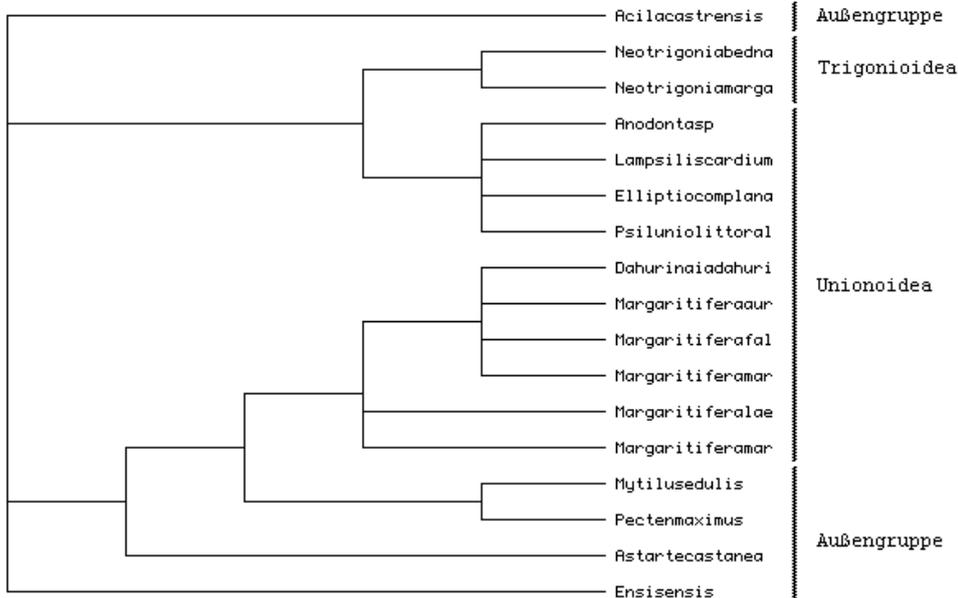


Parsimony Tree mit Bootstrap für Palaeoheterodonta nach NNC
differenziert gewichtet nach Größe, single-state Merkmale

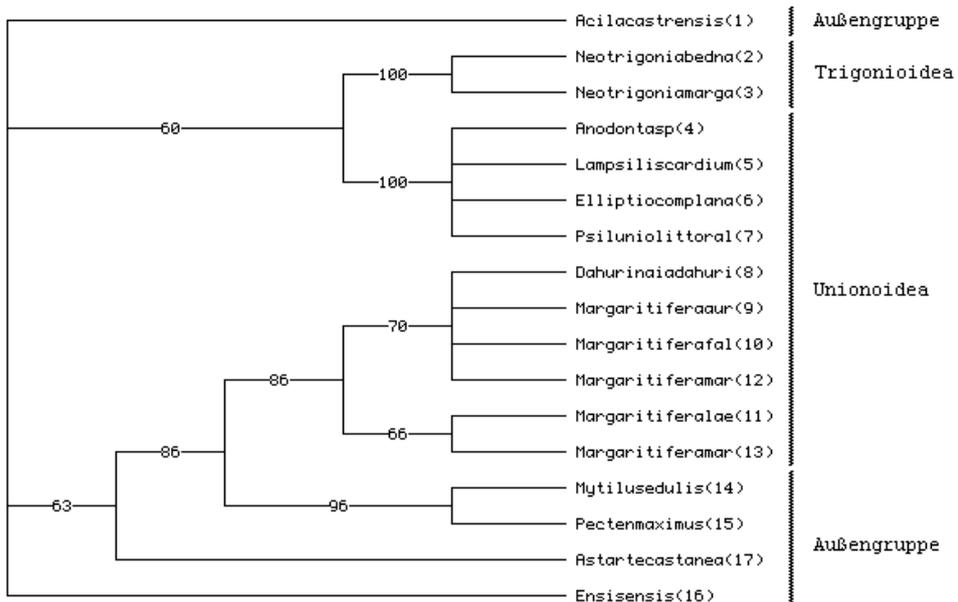


7.1.4.2.2 Differenziert gewichtet nach Komplexität

Parsimony Tree für Palaeoheterodonta nach NNC
differenziert gewichtet nach Komplexität, single-state Merkmale

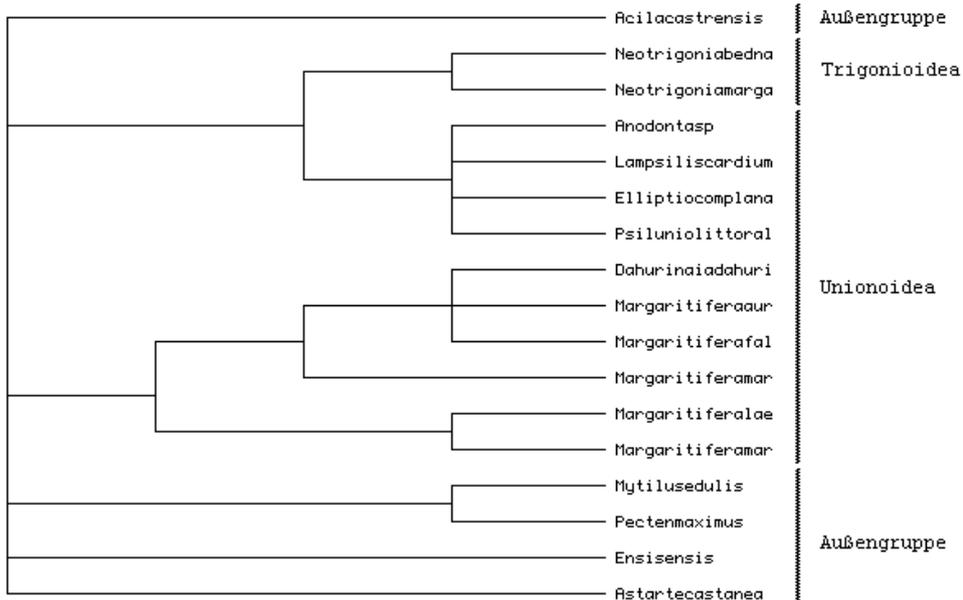


Parsimony Tree mit Bootstrap für Palaeoheterodonta nach NNC
differenziert gewichtet nach Komplexität, single-state Merkmale

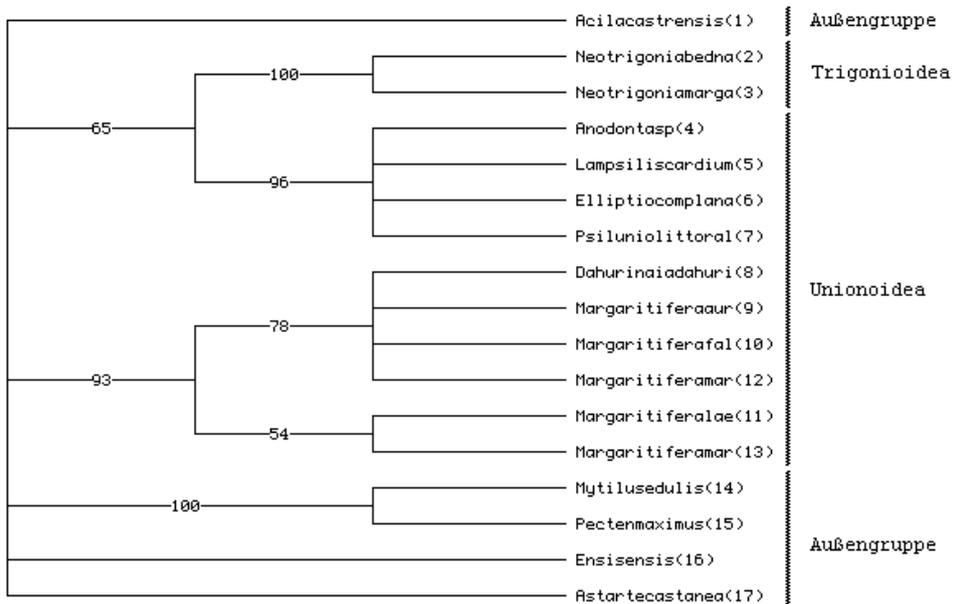


7.1.4.2.3 Gleichgewichtet

Parsimony Tree für Palaeoheterodonta nach NNC
gleichgewichtete, single-state Merkmale



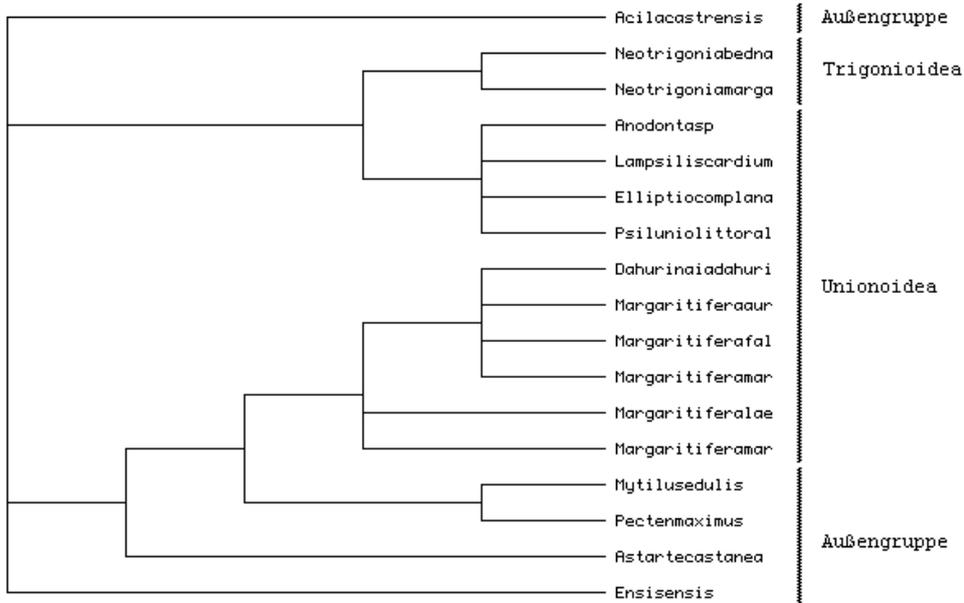
Parsimony Tree mit Bootstrap für Palaeoheterodonta nach NNC
gleichgewichtete, single-state Merkmale



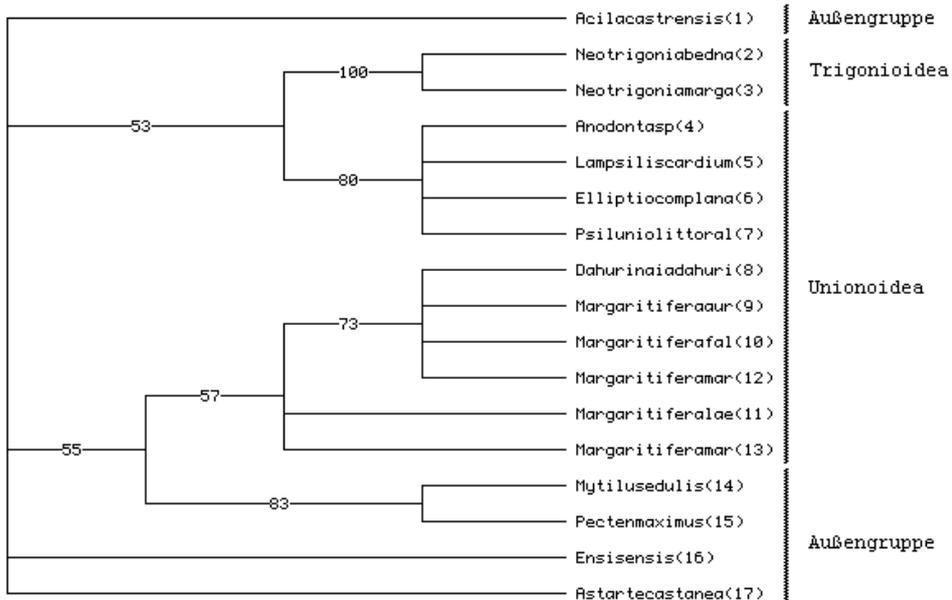
7.1.4.3 Palaeoheterodonta NNC Multi State

7.1.4.3.1 Differenziert gewichtet nach Größe

Parsimony Tree für Palaeoheterodonta nach NNC
differenziert gewichtet nach Größe, multi-state Merkmale

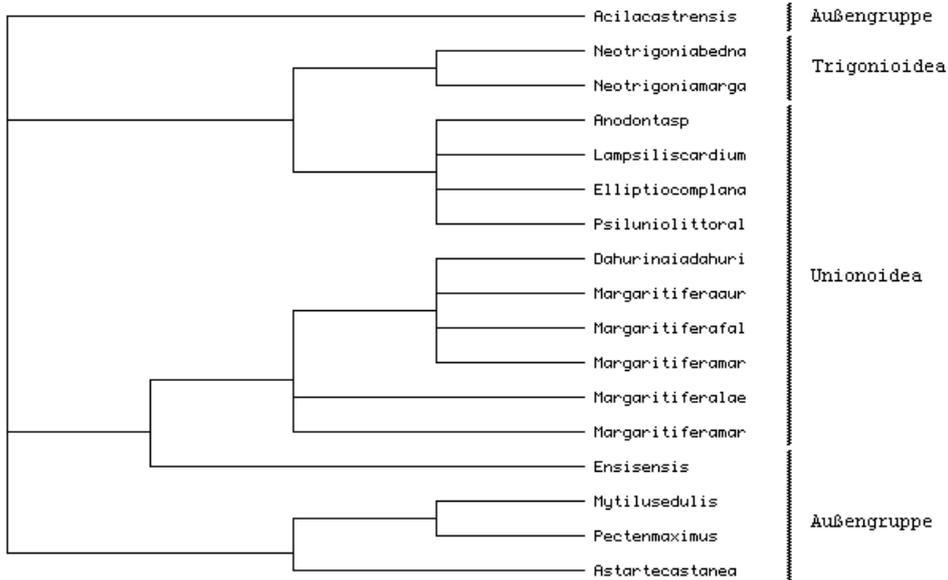


Parsimony Tree mit Bootstrap für Palaeoheterodonta nach NNC
differenziert gewichtet nach Größe, multi-state Merkmale

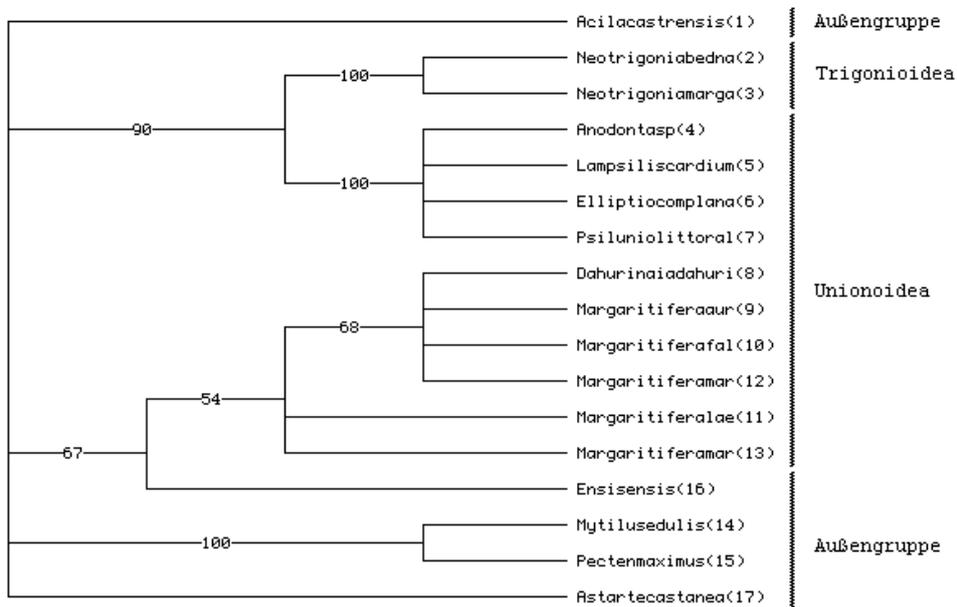


7.1.4.3.2 Gleichgewichtet

Parsimony Tree für Palaeoheterodonta nach NNC
gleichgewichtete, multi-state Merkmale



Parsimony Tree mit Bootstrap für Palaeoheterodonta nach NNC
gleichgewichtete, multi-state Merkmale

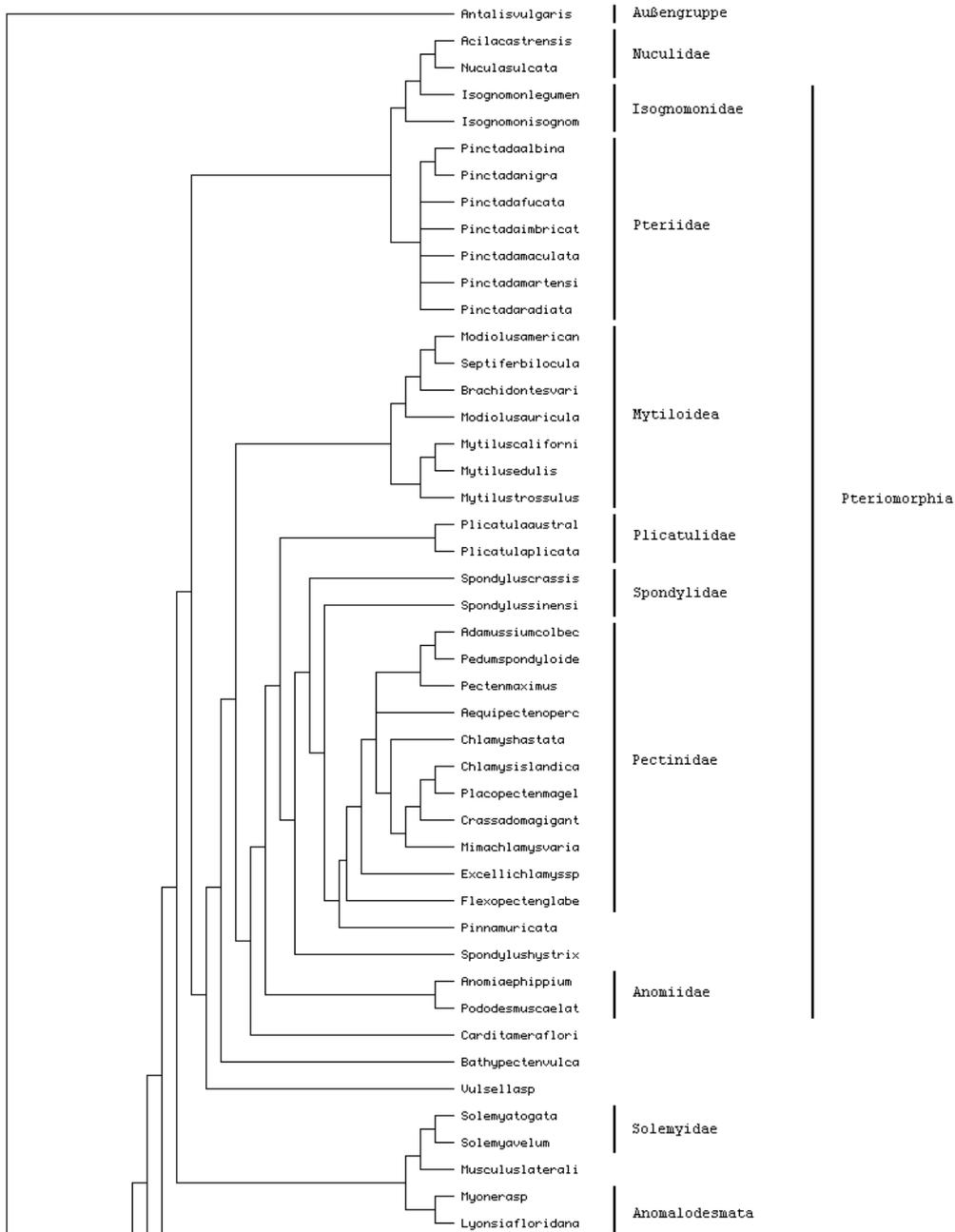


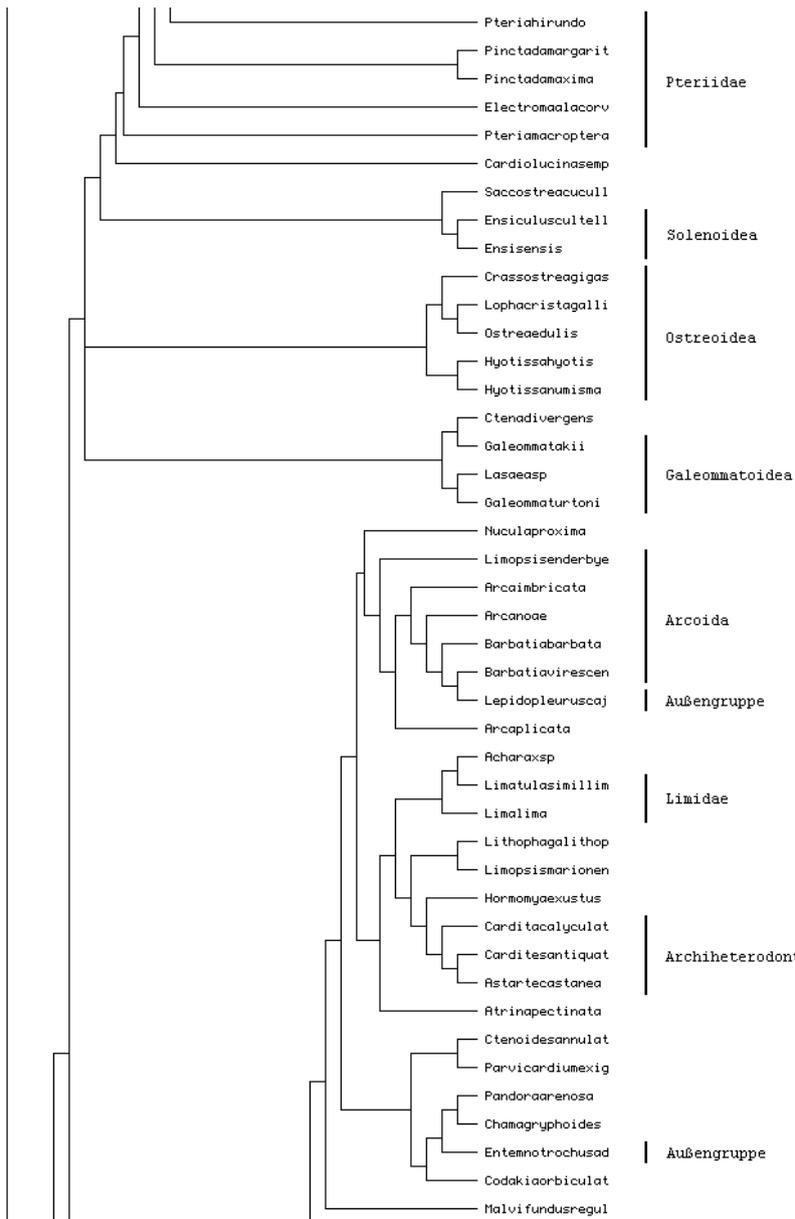
7.1.5 Stammbäume Bivalvia

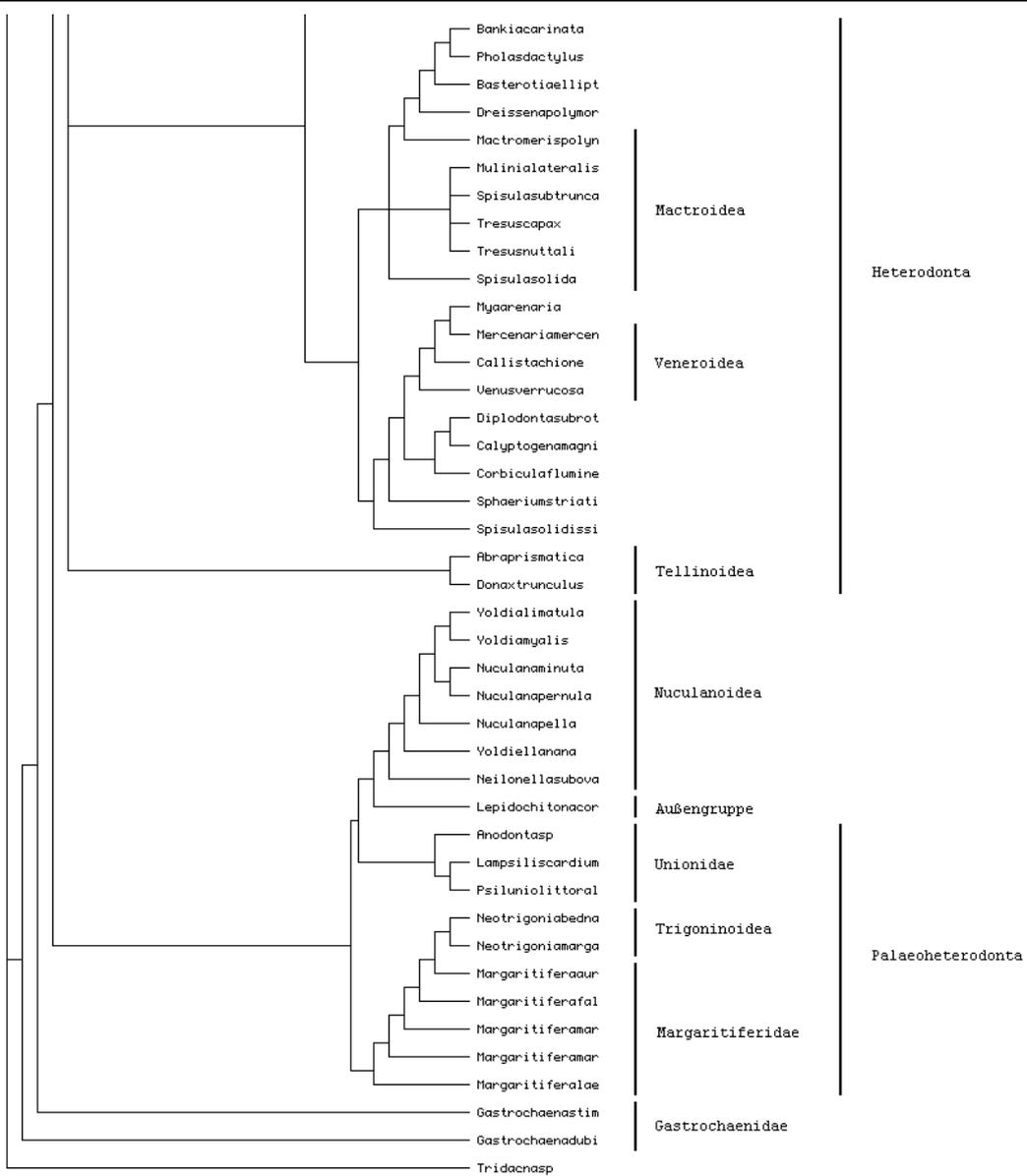
7.1.5.1 Bivalvia MFE Single State

7.1.5.1.1 Differenziert gewichtet nach Größe

Parsimony Tree für Bivalvia nach MFE
differenziert gewichtet nach Größe, single-state Merkmale



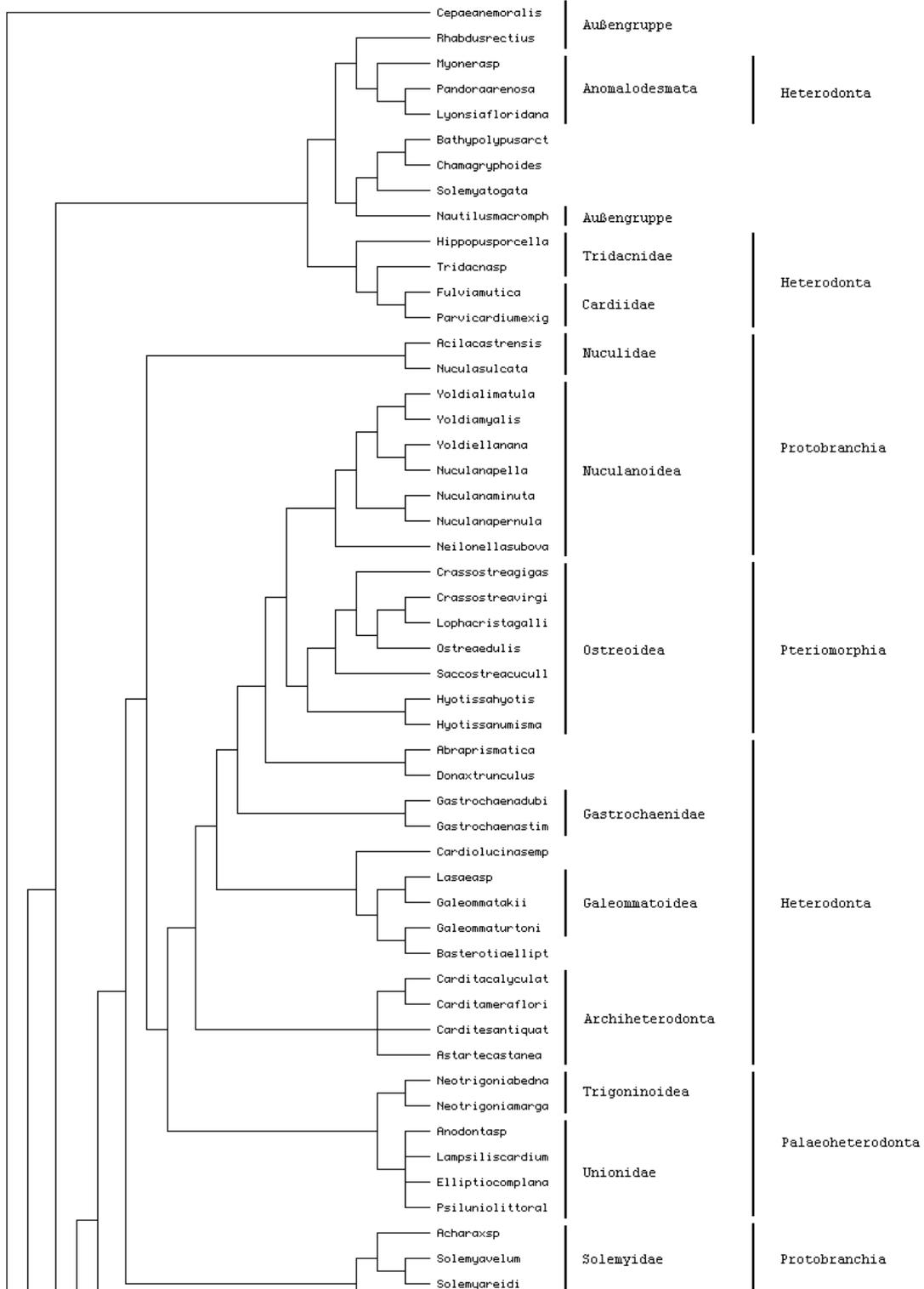


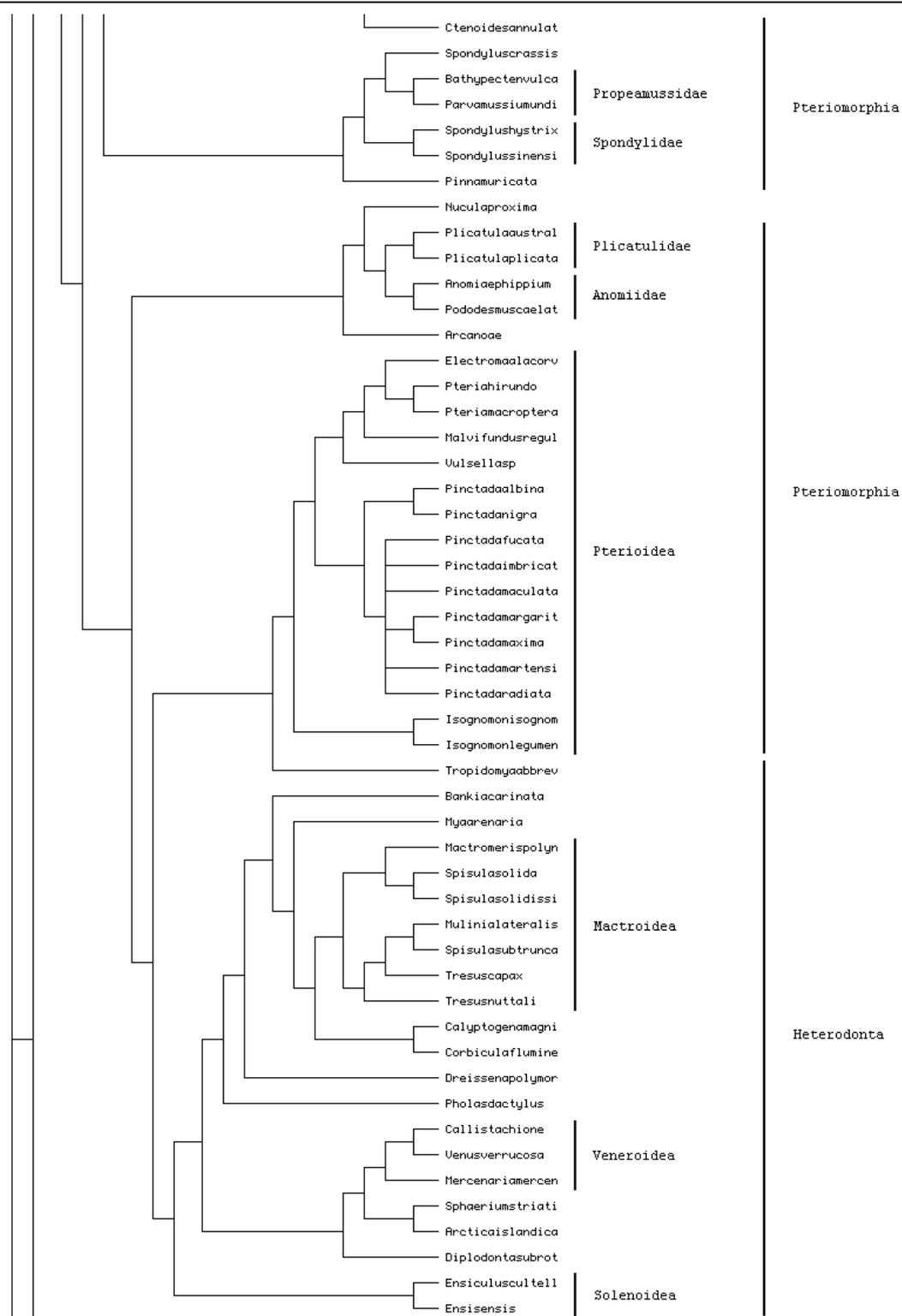


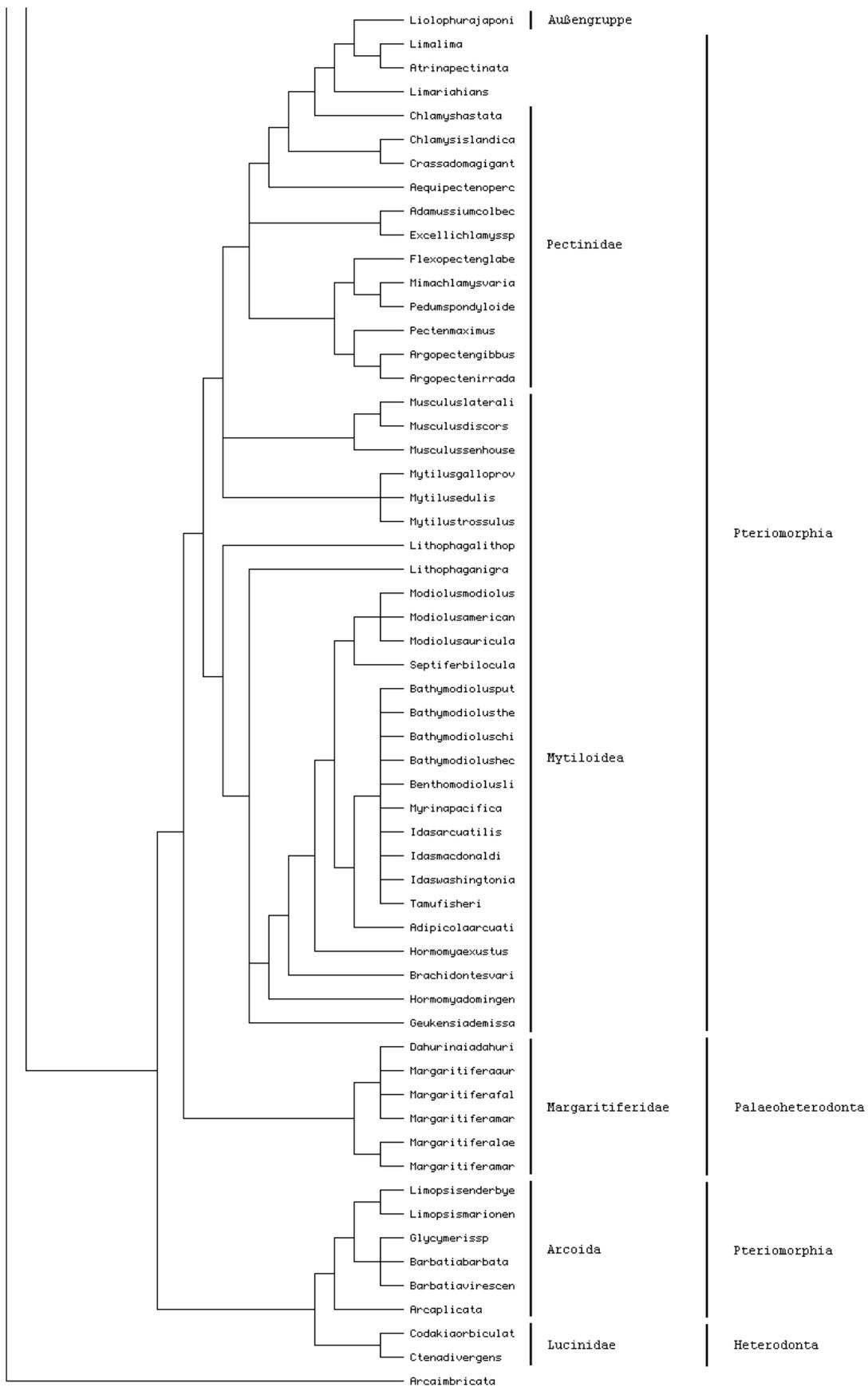
7.1.5.2 *Bivalvia NNC Single State*

7.1.5.2.1 Differenziert gewichtet nach Größe

Parsimony Tree für Bivalvia nach NNC
differenziert gewichtet nach Größe, single-state Merkmale





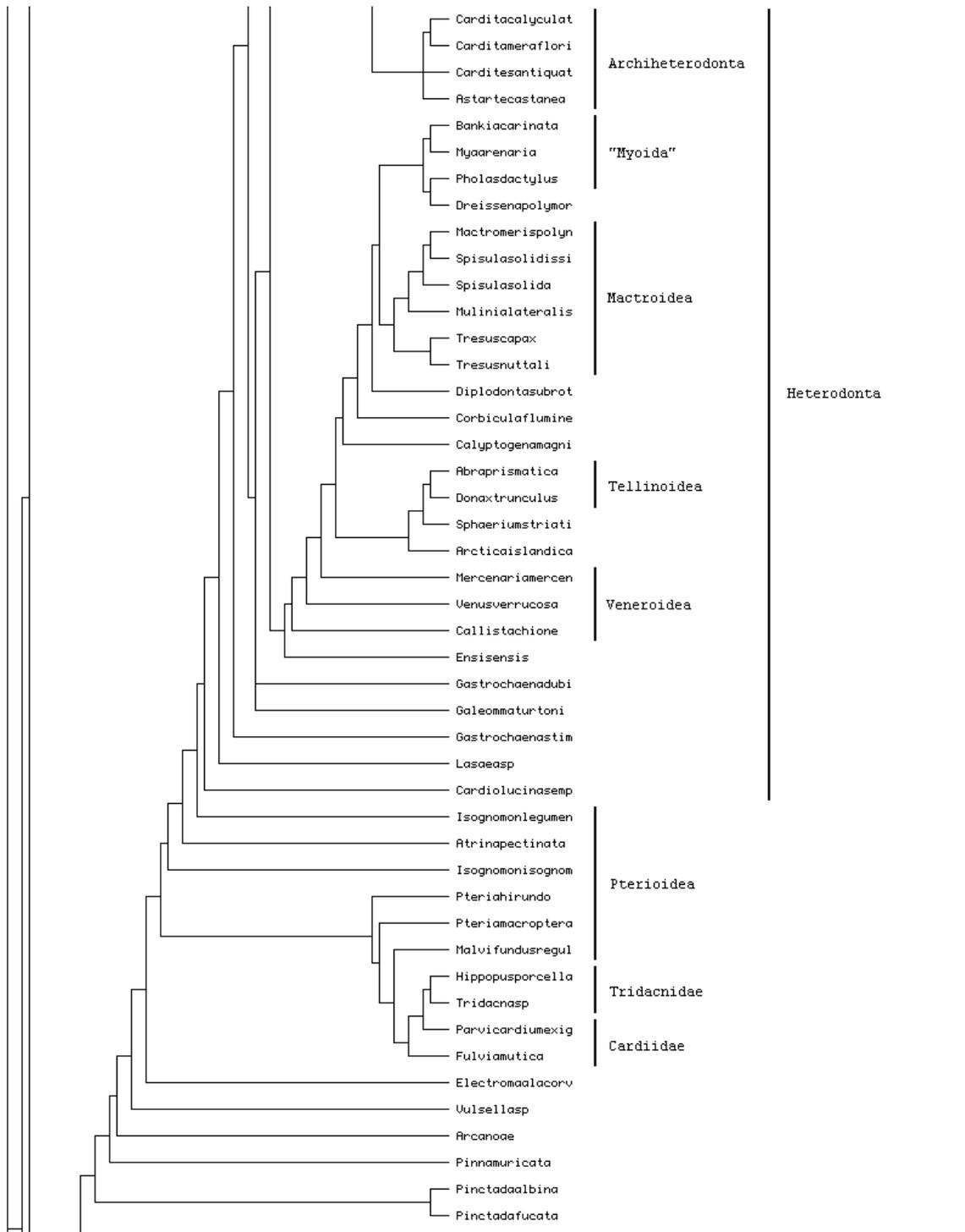


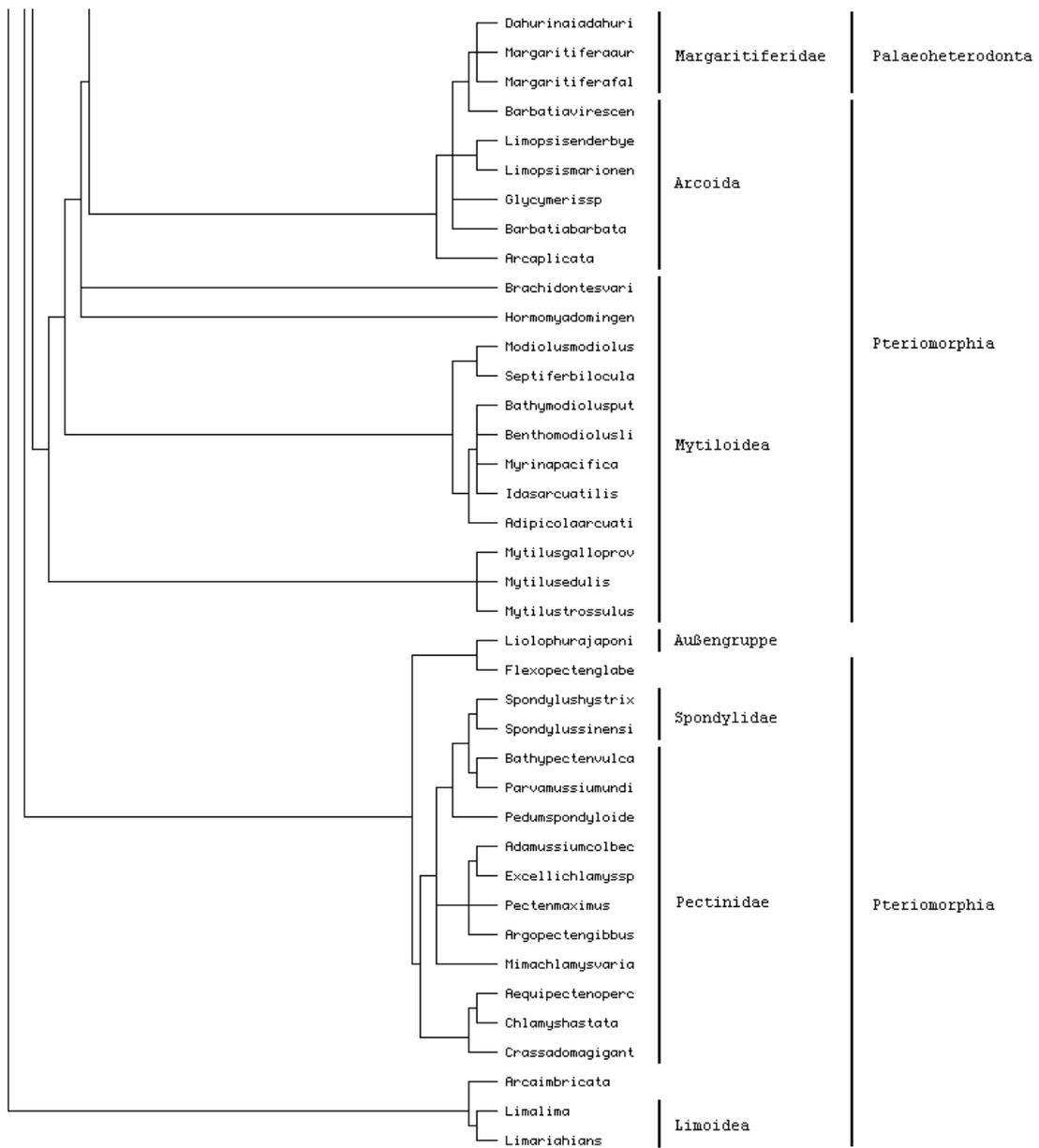
7.1.5.3 *Bivalvia NNC Multi State*

7.1.5.3.1 Differenziert gewichtet nach Größe

Parsimony Tree für Bivalvia nach NNC
differenziert gewichtet nach Größe, multi-state Merkmale

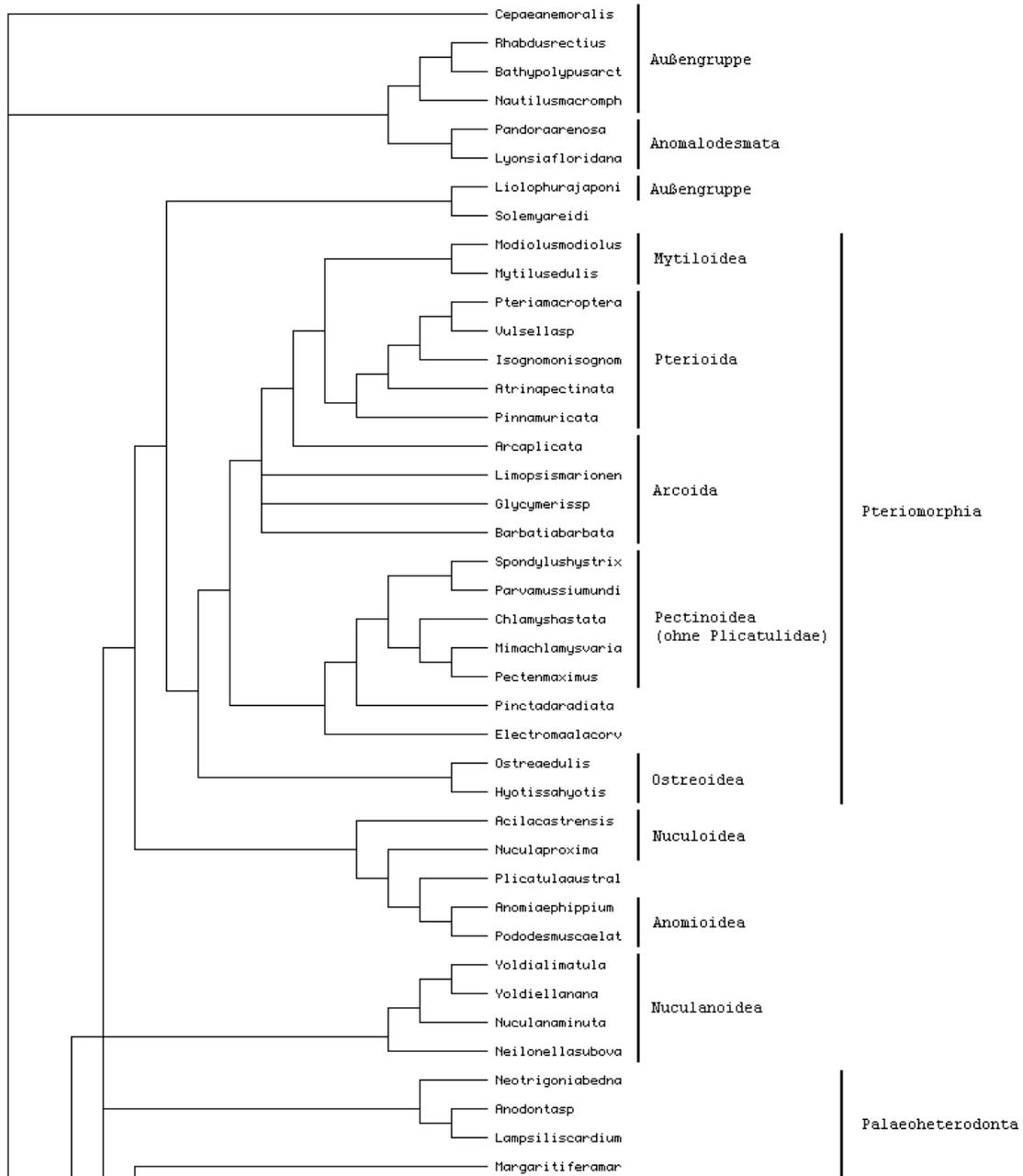


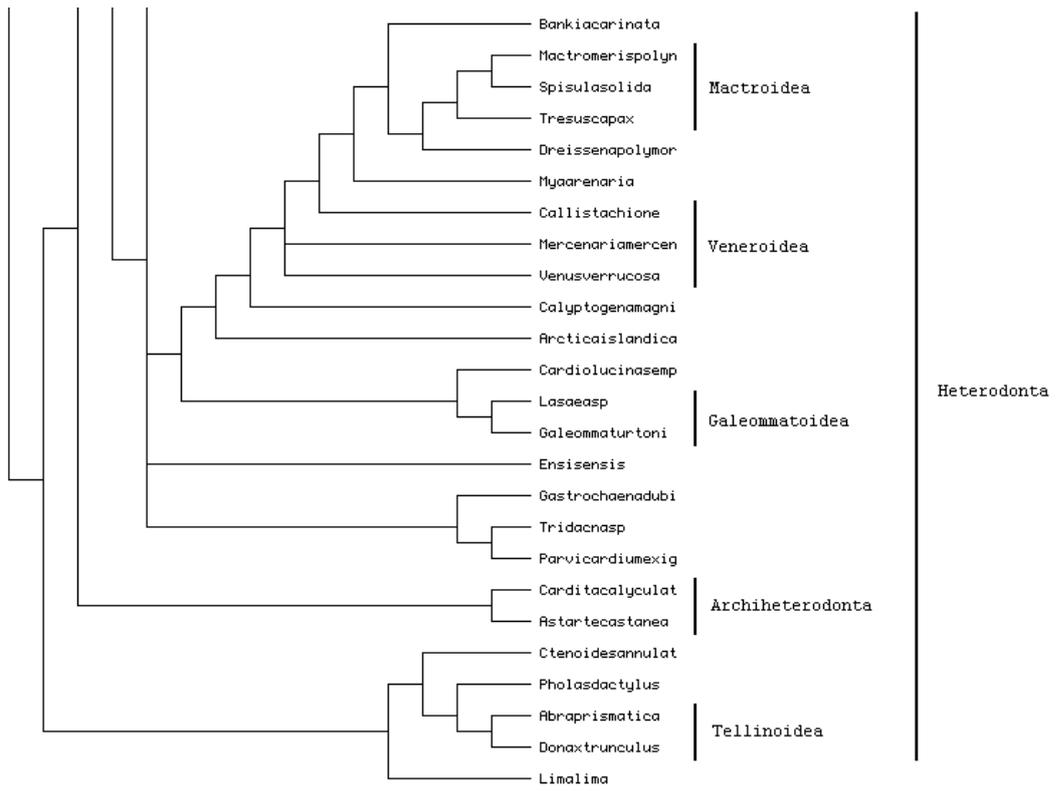




7.1.5.3.2 Gleichgewichtet

Parsimony Tree für Bivalvia nach MMC
gleichgewichtete, multi-state Merkmale





7.2 Glossar

18S rDNA

Die 18S rRNA ist als ein Baustein der Ribosomen Bestandteil aller eukaryotischen Zellen. Während die DNA die Erbinformation speichert, sind die verschiedenen RNA-Typen für die Übersetzung der symbolischen Zeichenfolge der DNA in eine Sequenz von Aminosäuren zuständig (Transkription), d.h. für die Steuerung der Proteinsynthese.

Es gibt mehrere RNA-Typen, die nach ihren Funktionen und ihrer Größe benannt sind: die Messenger-RNA (mRNA), die Transfer-RNA (tRNA), die ribosomale RNA (rRNA) und die mitochondriale RNA (mtRNA).

18S rRNA gehört zu den Ribonucleinsäuren, aus denen eukaryotische Ribosomen aufgebaut sind. Ribosomen sind für die Proteinsynthese innerhalb der Zelle zuständige Strukturen. Sie bestehen aus einer großen und einer kleinen Untereinheit. Die große Untereinheit besteht ihrerseits aus jeweils einem Strang 28S rRNA und 5,8S rRNA; die 18S rRNA bildet die kleine Untereinheit.

rRNA wird von der RNA-Polymerase I von der DNA transkribiert. Zunächst entsteht eine 45S rRNA, die anschließend durch RNA-Prozessierung in die einzelnen Untereinheiten aufgespalten wird.

Die 18S rRNA besteht im Durchschnitt aus 1800 beteiligten Basen, diese Zahl kann um ca. 150 variieren. Ihren Namen erhielt die 18S rRNA aufgrund ihres Sedimentationskoeffizienten von 18 Svedberg (S).

Die 18S rRNA ist aufgrund ihrer essentiellen Funktion ein von Art zu Art sehr konservatives Molekül, das eine sehr geringe Mutationsrate aufweist. Im Sinne der Molekularen Uhr lassen sich mit der 18S rRNA nur Veränderungen über Zeiträume von mehr als 100 Mio Jahren analysieren.

Alignment

Um funktionelle oder konservative Abschnitte in einer Gensequenz und Übereinstimmungen zwischen den Genen verschiedener Arten zu finden, werden die Primärsequenzen der gleichen Gene verschiedener Arten durch die Analysesoftware gegeneinander ausgerichtet („aligniert“). Im Alignment werden „Leerstellen“ in die Sequenz „eingefügt“, um Indels auszugleichen. Die vom Nutzer gesetzten Parameter legen die Höhe der Strafen für eingefügte Leerstellen, und die Zahl der Pluspunkte für übereinstimmende Basen fest. Entsprechend wird ein „bestes“ Alignment durch maximale Punktzahl gemäß den gesetzten Parametern ermittelt (Abbildung 19 und Abbildung 21).

Ein Alignment von Sequenzen ist eine Vermutung, welche Abschnitte der Gensequenz analog zueinander sind, bzw. wo Mutationen stattgefunden haben.

Ein Alignment von mehreren Sequenzen verschiedener Arten hilft dabei, die funktionellen Abschnitte innerhalb einer Gensequenz zu identifizieren. Bereits bekannte funktionelle/analoge Abschnitte können ein Alignment verbessern.

Ein Alignment ist die Grundlage für einen Consensus Constraint.

Bootstrapping

In der Phylogenetischen Analyse ist die Bootstrap-Methode ein bewährter Weg, um die Robustheit und Fehler-Empfindlichkeit eines berechneten phylogenetischen Stammbaums zu bestimmen. Dabei werden aus den Originaldaten vergleichbare Surrogat-Daten durch zufällige Modifikationen künstlich hergestellt und im anschließenden Vergleich die Variation zwischen verschiedenen gleichgroßen Stammbäumen genutzt, um die Güte des ursprünglichen Stammbaums abzuschätzen. Bootstrap-Werte (die Häufigkeit des wiederholten Vorkommens einer ausgewählten Gruppe in allen Surrogat-Stammbäumen, in Prozent) die größer als 95 sind, werden dabei als signifikant angesehen.

Die statistische Methode des Bootstrapping wurde 1979 von Efron vorgestellt, 1985 wurde sie erstmals von Felsenstein für phylogenetische Stammbäume angewandt.

Constraint

Ein Constraint ist in der Software die Bezeichnung für eine Einschränkung der möglichen Sekundärstrukturen einer Gensequenz, die durch Faltungsalgorithmen vorhergesagt werden können. Der Constraint gibt vor, welche Basen in Doppelbindungen festgelegt sind. Das heißt, bestimmte Abschnitte der Gensequenz müssen auf eine vorher festgelegte Weise gefaltet werden. Der Constraint hat Vorrang vor MFE-Faltungen. Meist sind funktionelle Abschnitte durch einen Constraint festgelegt.

Consensus Constraint

Um Sekundärstrukturen nach dem Prinzip NNC von Sequenzen zu erstellen, von denen kein Constraint bekannt ist, kann man einen sogenannten Consensus Constraint nutzen. Voraussetzung für einen Consensus Constraint ist ein Constraint von mindestens einer nah verwandten Art. Die Sequenz der Constraint-Art wird in einem Alignment mit den anderen zu untersuchenden Sequenzen ausgerichtet. Als Resultat kann der bekannte Constraint für die unbekanntenen Sequenzen als Vorlage dienen – es ist ein allen gemeinsamer Constraint. Ein solcher Consensus Constraint ist eine Vermutung.

Nexus-Datei

Eine Nexus-Datei ist eine Textdatei im ASCII-Format, in der mittels einer nach dem Nexus-Format definierten Syntax eine Merkmalsmatrix sowie Bearbeitungs- oder Analysebefehle festgelegt werden. Diese Datei kann von entsprechender Software interpretiert werden, um phylogenetische Stammbäume oder Abstandsmatrizen, basierend auf der übergebenen Merkmalsmatrix, zu erstellen.

Das Format wurde von Maddison et al., 1997 vorgestellt.

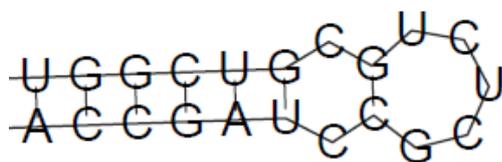
DCSE

DCSE ist eine Software und ein Format, um Alignments mit einer großen Anzahl Sequenzen zu erstellen. Es bietet außerdem die Möglichkeit, Sekundärstrukturen in das Alignment mit einzubeziehen. Alignments und Sekundärstrukturen die mit diesem Programm generiert wurden, können in RNAViz weiterprozessiert werden.

DCSE wurde 1993 von Rijk & Wachter vorgestellt.

Dot-Bracket-Notation

Die Sekundärstruktur von RNA wird häufig in der sogenannten Dot-Bracket-Notation dargestellt: Dabei steht parallel zur Primärsequenz eine gleich lange Reihe aus Punkten (./ Dot) und öffnenden und schließenden Klammern ((,)) / Bracket). Jeder Punkt steht für eine ungebundene Base, jede Klammer für eine Base in einer Doppelbindung. Eine Doppelbindung der Basen an Position i und j wird durch „(,“ an Position i und „,)“ an Position j dargestellt. Die Anzahl der öffnenden und schließenden Klammern muss dementsprechend gleich sein.



Die links abgebildete Struktur würde in der Dot-Bracket-Notation folgendermaßen aussehen:

```
UGGCUGGCGUCGCUAGCCA
(((((((.(.....).))))))
```

Pseudoknoten werden mit Hilfe von {} oder [] beschrieben. Sie stellen jedoch keine eigentliche Doppelbindung dar.

EST / Expressed Sequence Tag

Ein EST ist ein Marker für DNA-Sequenzen, mit dem der Anfang (5' EST) und das Ende (3' EST) eines gesuchten Genabschnitts in einer nicht näher identifizierten Sequenz lokalisiert, identifiziert und isoliert werden kann. ESTs können 80 bis 500 Nukleotide enthalten.

FASTA-Format

Das FASTA-Format ist eine Möglichkeit Primärstruktur und Nukleinsäuresequenz von Genen und Proteinen in Textformat darzustellen, zu speichern und zwischen Programmen zu transferieren.

```
>Pecten_maximus  
CAACCCUGGUUGAUCUGCCAGUAGUCA ...  
....((((((((.....))))))(((.( ...
```

Das Format erlaubt es, der Sequenz einen Namen oder Kommentar voranzustellen. In der zweiten Zeile erfolgt dann die Zeichenkette der Sequenz. Zusätzlich ist es möglich im Falle von DNA oder RNA Constraint-Informationen in einer dritten Zeile in der Dot-Bracket-Notation hinzuzufügen. Die Zeilen des Formats sind durch Steuerzeichen (Zeilenumbruch) voneinander separiert.

Fossilbericht / Fossil Record

Der so genannte Fossilbericht umfasst den gesamten Datenbestand von Fossilien. Er beginnt im Kambrium mit der Entwicklung von Skeletten und Gehäusen aus Kalziumkarbonat durch vielzellige Tiere. Im Idealfall kann eine ununterbrochene Entwicklungslinie bestimmter Organismen von ihrer Entstehung bis zu ihrer heutigen Form im Fossil Record nachgewiesen werden.

Aus dem Präkambrium gibt es nur wenige Fossilien. Hartteile aus Kalziumkarbonat haben ein wesentlich besseres Fossilisationspotenzial als Weichteile allein. Dies erklärt, warum im Kambrium plötzlich so viele Tierstämme auftreten, über deren Vorfahren nichts bekannt ist.

Um zu belegen, dass Menschen und Affen von den gleichen Vorfahren abstammen, wurde ebenfalls der Fossil Record herangezogen.

GIBBS

Siehe MFE

InDels

Eine mögliche Mutationsvariante in Gensequenzen sind die sogenannten „InDels“. Mit InDels sind „Insertions“ und „Deletions“ einzelner oder mehrerer Basen gemeint, d.h. eine Anzahl beliebiger Basen wurde neu eingefügt oder aus der Sequenz herausgeschnitten. Dadurch verändert sich die Länge der Sequenz.

Maximum Expected Pair Accuracy

Im Gegensatz zu der Berechnung von Sekundärstrukturen nach dem MFE-Prinzip (bei dem nur die erwarteten Doppelstrangabschnitte ausgewertet werden), wird bei „Maximum Expected Pair Accuracy“ nicht nur die Wahrscheinlichkeit von Basenpaarungen bestimmt, sondern auch die Wahrscheinlichkeit, dass Nukleotide ungebunden in einem Einzelstrangabschnitt vorkommen.

Diese Methode wurde vorgestellt von *Lu et al., 2009*.

Maximum Parsimony

Das Prinzip „Maximum Parsimony“ oder „Maximale Sparsamkeit“ wird bei der Berechnung phylogenetischer Stammbäume genutzt. Dabei wird jeweils der Stammbaum bevorzugt, in welchem die geringste Anzahl an Merkmalsänderungen benötigt wird, um die beobachteten Merkmale zu erklären.

MFE / Minimum Free Energy

Begriffsnutzung: In dieser Arbeit wird im Einklang mit der englischen Literatur nur von „Minimum Free Energy“ oder abgekürzt MFE gesprochen. Die direkte deutsche Übersetzung „Minimale Freie Energie“ wäre im thermodynamischen Sinn nicht korrekt, sondern müsste „Minimale Freie Enthalpie“ heißen. Im Deutschen wird zwischen den Begriffen „Freie Enthalpie“ (*G*) und „Freie Energie“ (*F*) unterschieden. Freie

Enthalpie $G(T,P)$ bedeutet, dass Druck P und Temperatur T gegeben sind, das Volumen $V(T,P)$ dagegen eine Funktion dieser Größen ist. Freie Energie $F(T,V)$ dagegen setzt die Temperatur und das Volumen als bekannt voraus, und der Druck $P(T,V)$ ergibt sich dann daraus.

Für die Freie Enthalpie sind außerdem, häufiger im Englischen als im Deutschen, auch die Bezeichnungen „Gibbs-Energie“ oder „Freie Gibbs-Energie“ (nach J. W. Gibbs) gebräuchlich. Für die Freie Energie (F), sind auch die Begriffe „Helmholtz-Energie“ oder „Freie Helmholtz-Energie“ (nach H. v. Helmholtz) in der Literatur zu finden.

Die MFE-Struktur eines Moleküls, und damit auch einer RNA-Struktur, besitzt ein Minimum an Freier Enthalpie (bei gegebenen Werten von Temperatur und Druck). Ein Molekül mit geringer Freier Enthalpie ist stabiler als ein Molekül mit hoher Freier Enthalpie. Im Sinne dieser Arbeit und als Annahme, dass RNA-Strukturen relativ stabile Moleküle sind, werden durch die verwendeten Algorithmen immer die Prinzipien von MFE für die Berechnung der Sekundärstruktur zugrunde gelegt. Bei der Nutzung eines Constraints für die Strukturberechnung werden diese Prinzipien für die im Constraint festgelegten Bindungen ignoriert. Die im Anschluss noch freien Basen werden wiederum nach MFE gebunden.

Molekulare Uhr

Je länger zwei Taxa getrennt sind, desto größer werden die Unterschiede in ihren Nucleotidsequenzen. 1961 wurde auf dieser Grundlage von Linus Pauling und Emile Zuckerkandl das Konzept der Molekularen Uhr postuliert. Sie bezieht sich auf die Mutationsraten von Genen oder Proteinen, wobei angenommen wird, dass deren Taktfrequenz konstant ist und z.B. durch Fossilienfunde geeicht werden kann.

Molekulare Uhren lassen relative Zeitaussagen zu – je nachdem welche Proteine bzw. Gene verglichen werden, können verschiedene Zeiträume untersucht werden.

Gene, deren Funktion überlebenswichtig ist, haben eine sehr geringe Evolutionsgeschwindigkeit. Im Gegensatz dazu haben Gene mit nicht-lebenswichtigen Funktionen deutlich höhere Mutationsraten. Dieses Prinzip gilt sowohl für Nucleotidsubstitution als auch Aminosäuresubstitution: Histone sind hoch konserviert, Hämoglobine verändern sich relativ schnell; rDNA ist stabiler als nicht-kodierende DNA.

Near-Natural

Im Gegensatz zu MFE-Strukturen von RNA-Sequenzen werden die so genannten „Near Natural“-Strukturen mit einem Constraint gefaltet. Diese Einschränkung in den Faltungsmöglichkeiten gewährleistet, dass die resultierende 3D-Struktur in den funktionellen Bereichen möglichst „naturnah“ dargestellt wird.

Pseudo-Knoten

Ein RNA-Einzelstrang faltet in Abschnitten auf sich selbst zurück und bildet Doppelstrangregionen, Hairpin Loops, Multibranching Loops und Bulge Loops aus (Abbildung 26). Als Pseudoknoten bezeichnet man eine RNA-Sekundärstruktur, bei der die Schleife eines Loops einen Teil des Stamms eines anderen Loops bildet (Abbildung 27). Diese Struktur wurde 1982 von *Rietveld et al.* an einem Mosaikvirus nachgewiesen.

Derartige Formen kommen in der Natur vor, können aber durch automatische Faltungsalgorithmen kaum vorhergesagt werden. Aus diesem Grund und weil die generierten Sekundärstrukturen eine vereinfachte Darstellung und nicht unbedingt eine naturgetreue Abbildung sind, wurden Pseudoknoten aus der Betrachtung herausgelassen. Im unwahrscheinlichen Fall, dass ein Pseudoknoten doch als Ergebnis des Faltungsalgorithmus auftritt, sollte dieser bei der Auswertung ignoriert werden. Praktisch kam dieser Fall bei den untersuchten Strukturen aber nicht vor.

Sekundärstruktur

Von einer Sekundärstruktur spricht man insbesondere bei biologischen Makromolekülen, die kettenförmig aus einander ähnlichen Bausteinen zusammengesetzt sind. Zu solchen Makromolekülen gehören DNA und die verschiedenen Ausprägungen von RNA und Proteinen. Durch physikalische Wechselwirkungen zwischen den Bausteinen bildet sich aus der eindimensionalen Primärstruktur

Anhang

(Reihenfolge der Nukleinsäuren, bzw. Aminosäuren) eine übergeordnete, oft zweidimensional dargestellte Sekundärstruktur (*Linderstrøm-Lang, 1952*).

Von der DNA ist bekannt, dass sie aus zwei (aus Nukleinsäuren aufgebauten) komplementären Einzelsträngen besteht. Durch die o.g. Wechselwirkungen bildet sich als resultierende Sekundärstruktur eine chemisch stabile Doppelhelix aus (Abbildung 24).

Die RNA liegt zumeist als Einzelstrang vor. Die Sekundärstruktur der RNA entsteht dadurch, dass der Einzelstrang unter bestimmten physiko-chemischen Bedingungen und unter Bildung intramolekularer Basenpaare auf sich selbst zurückfaltet. Unter der Sekundärstruktur einer RNA-Sequenz versteht man im Wesentlichen eine Liste der gepaarten Basen (*Fontana, 1996*).

Eine dreidimensionale Tertiärstruktur kann sich wiederum aufgrund von Wechselwirkungen zwischen den Clustern/Regionen/Abschnitten der Sekundärstruktur herausbilden (Abbildung 44).

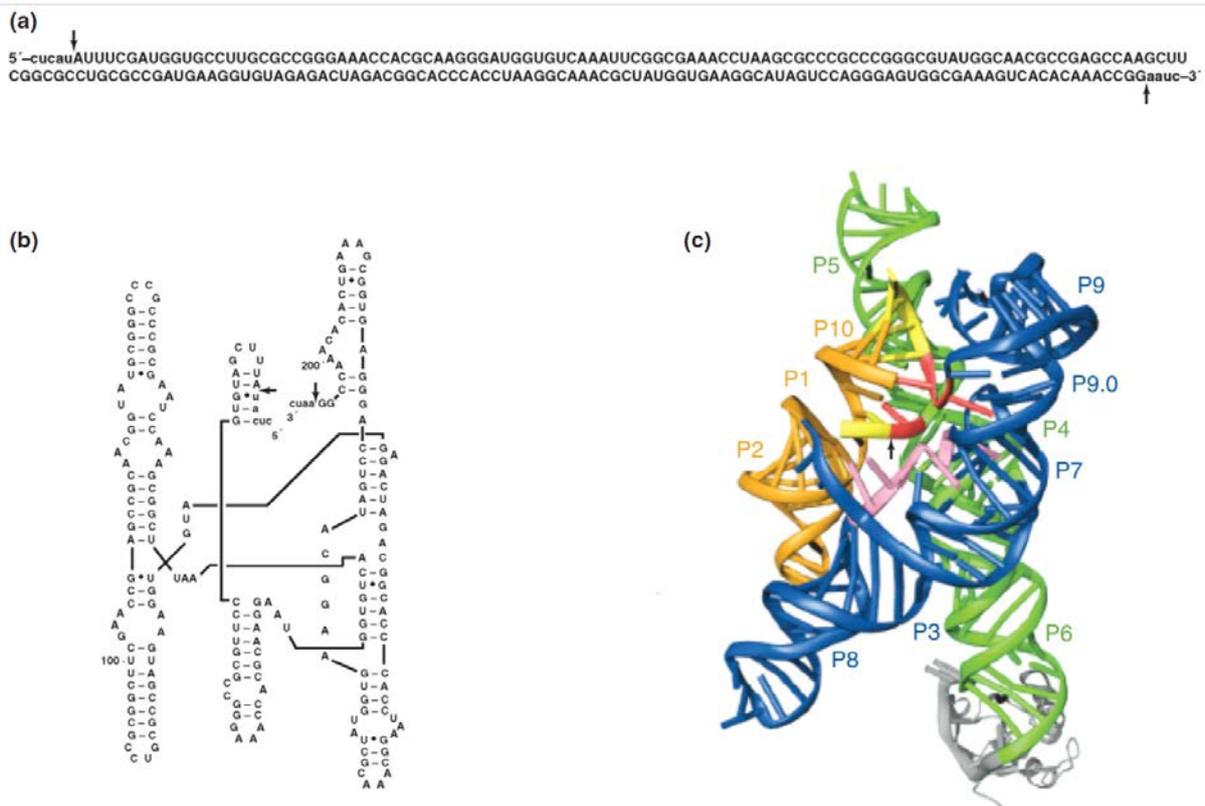


Abbildung 44: RNA Strukturdarstellungen. (a) Die Primärstruktur stellt die Reihenfolge der Nukleotide bzw. die Basensequenz der RNA dar. (b) Die Sekundärstruktur zeigt die gefaltete Sequenz mit den Paarbindungen zwischen den Basen, innerhalb eines Einzelstrangs. (c) Die Tertiärstruktur ist die Darstellung der dreidimensionalen, übergeordneten Anordnung der (in der Sekundärstruktur geformten) Doppelstrangregionen. Die RNA in dieser Abbildung ist ein Abschnitt aus der Isoleucin-tRNA des Purpurbakteriums *Azoarcus* (aus *Lehman, 2010*).

Skeleton-Datei

Eine Skeleton-Datei ist eine Vorlage, mit der das Programm RNAViz die Sekundärstruktur mehrerer ähnlicher Sequenzen darstellen kann.

SVG-Datei

Eine SVG-Datei ist ein vektorbasiertes Grafikformat, welches einfache, interaktive Zeichnungen ermöglicht. Diese können dynamisch durch Algorithmen (zum Beispiel auf einem Server) erzeugt und von einer großen Anzahl Internetbrowser ohne weitere Hilfsmittel interpretiert und angezeigt werden. Dieses Format eignet sich sehr gut zur Erstellung und zum Vergleich von Sekundärstrukturen.

Tertiärstruktur

Siehe Sekundärstruktur.

7.3 Artenliste Bivalvia

<i>Acila castrensis</i>	<i>Limopsis marionensis</i>	<i>Atrina pectinata</i>
<i>Nucula proxima</i>	<i>Glycymeris</i> sp	<i>Pinna muricata</i>
<i>Nucula sulcata</i>	<i>Arca imbricata</i>	<i>Malvifundus regulatus</i>
<i>Yoldia limatula</i>	<i>Arca noae</i>	<i>Vulsella</i> sp
<i>Yoldia myalis</i>	<i>Arca plicata</i>	<i>Tropidomya abbreviata</i>
<i>Yoldiella nana</i>	<i>Barbatia barbata</i>	<i>Cuspidaria cuspidata</i>
<i>Nuculana minuta</i>	<i>Barbatia virescens</i>	<i>Myonera</i> sp
<i>Nuculana pella</i>	<i>Ctenoides annulata</i>	<i>Pandora arenosa</i>
<i>Nuculana pernula</i>	<i>Lima lima</i>	<i>Lyonsia floridana</i>
<i>Neilonella subovata</i>	<i>Limaria hians</i>	<i>Bankia carinata</i>
<i>Acharax</i> sp	<i>Limatula simillima</i>	<i>Pholas dactylus</i>
<i>Solemya velum</i>	<i>Crassostrea gigas</i>	<i>Mya arenaria</i>
<i>Solemya reidi</i>	<i>Crassostrea virginica</i>	<i>Varicorbula dissimilis</i>
<i>Solemya togata</i>	<i>Lopha cristagalli</i>	<i>Hiatella arctica</i>
<i>Neotrignonia bednalli</i>	<i>Ostrea edulis</i>	<i>Callista chione</i>
<i>Neotrignonia margaritacea</i>	<i>Saccostrea cucullata</i>	<i>Mercenaria mercenaria</i>
<i>Anodonta</i> sp	<i>Hyotissa hyotis</i>	<i>Venus verrucosa</i>
<i>Lampsilis cardium</i>	<i>Hyotissa numisma</i>	<i>Abra prismatica</i>
<i>Elliptio complanata</i>	<i>Plicatula australis</i>	<i>Donax trunculus</i>
<i>Psilunio littoralis</i>	<i>Plicatula plicata</i>	<i>Ensiculus cultellus</i>
<i>Dahurinaia dahurica</i>	<i>Spondylus crassisquamatus</i>	<i>Ensis ensis</i>
<i>Margaritifera auricularia</i>	<i>Spondylus hystrix</i>	<i>Mactromeris polynyma</i>
<i>Margaritifera falcata</i>	<i>Spondylus sinensis</i>	<i>Mulinia lateralis</i>
<i>Margaritifera laevis</i>	<i>Bathypecten vulcani</i>	<i>Spisula solida</i>
<i>Margaritifera margaritifera</i>	<i>Parvamussium undisonum</i>	<i>Spisula solidissima</i>
<i>Margaritifera marrianae</i>	<i>Adamussium colbecki</i>	<i>Spisula subtruncata</i>
<i>Musculus lateralis</i>	<i>Aequipecten opercularis</i>	<i>Tresus capax</i>
<i>Musculus senhousei</i>	<i>Chlamys hastata</i>	<i>Tresus nuttali</i>
<i>Musculus discors</i>	<i>Chlamys islandica</i>	<i>Diplodonta subrotundata</i>
<i>Lithophaga lithophaga</i>	<i>Crassadoma gigantea</i>	<i>Cardiolucina semperiana</i>
<i>Lithophaga nigra</i>	<i>Excellichlamys spectabilis</i>	<i>Codakia orbiculata</i>
<i>Modiolus modiolus</i>	<i>Flexopecten glaber</i>	<i>Ctena divergens</i>
<i>Modiolus americanus</i>	<i>Mimachlamys varia</i>	<i>Calyptogena magnifica</i>
<i>Modiolus auriculatus</i>	<i>Pecten maximus</i>	<i>Gastrochaena dubia</i>
<i>Bathymodiolus puteoserpentis</i>	<i>Pedum spondyloideum</i>	<i>Gastrochaena stimpsonii</i>
<i>Bathymodiolus thermophilus</i>	<i>Argopecten gibbus</i>	<i>Lasaea</i> sp
<i>Bathymodiolus childressi</i>	<i>Argopecten irradians</i>	<i>Galeomma takii</i>
<i>Bathymodiolus heckerae</i>	<i>Placopecten magellanicus</i>	<i>Galeomma turtoni</i>
<i>Benthomodiolus lignocola</i>	<i>Anomia ephippium</i>	<i>Dreissena polymorpha</i>
<i>Brachidontes variabilis</i>	<i>Pododesmus caelata</i>	<i>Basterotia elliptica</i>
<i>Hormomya exustus</i>	<i>Electroma alacorvi</i>	<i>Sphaerium striatinum</i>
<i>Hormomya domingensis</i>	<i>Pinctada albina</i>	<i>Corbicula fluminea</i>
<i>Myrina pacifica</i>	<i>Pinctada fucata</i>	<i>Chama gryphoides</i>
<i>Mytilus galloprovincialis</i>	<i>Pinctada imbricata</i>	<i>Cardita calyculata</i>
<i>Mytilus edulis</i>	<i>Pinctada maculata</i>	<i>Carditamera floridana</i>
<i>Mytilus trossulus</i>	<i>Pinctada margaritifera</i>	<i>Cardites antiquata</i>
<i>Adipicola arcuatilis</i>	<i>Pinctada martensi</i>	<i>Hippopus porcellanus</i>
<i>Idas arcuatilis</i>	<i>Pinctada maxima</i>	<i>Tridacna</i> sp
<i>Idas macdonaldi</i>	<i>Pinctada nigra</i>	<i>Fulvia mutica</i>
<i>Idas washingtonia</i>	<i>Pinctada radiata</i>	<i>Parvicardium exiguum</i>
<i>Geukensia demissa</i>	<i>Pteria hirundo</i>	<i>Arctica islandica</i>
<i>Tamu fisheri</i>	<i>Pteria macroptera</i>	<i>Astarte castanea</i>
<i>Septifer bilocularis</i>	<i>Isognomon isognomum</i>	
<i>Limopsis enderbyensis</i>	<i>Isognomon legumen</i>	

7.4 Inhalt des digitalen Anhangs (CD)

Dieser Dissertation ist eine CD beigelegt, die den folgenden Inhalt hat:

```

|__Dissertation
|__Sequenzen
|__Sekundaerstrukturen
|   |__MFE
|   |__NNC
|
|__Stammbaeume
|   |__MFE
|   |   |__Bivalvia
|   |   |__Heterodonta
|   |   |__Palaeoheterodonta
|   |   |__Protobranchia
|   |   |__Pteriomorphia
|   |__NNC
|   |   |__Bivalvia
|   |   |__Heterodonta
|   |   |__Palaeoheterodonta
|   |   |__Protobranchia
|   |   |__Pteriomorphia

```

Dissertation	Enthält diese Arbeit in PDF-Format
Sequenzen	Enthält alle in dieser Arbeit einbezogenen Sequenzen. Quellen: GenBank, European Ribosomal Database
Sekundaerstrukturen	Enthält die in der vorliegenden Arbeit ausgewerteten Sekundärstrukturen, sortiert nach MFE (mit RNAfold erstellt) und NNC (mit RNAsalsa erstellt)
Stammbaeume	Enthält die in der vorliegenden Arbeit vorgestellten Stammbäume, basierend auf den in „Sekundaerstrukturen“ enthaltenen Faltungen. Die Stammbäume sind nach MFE und NNC sortiert, danach nach Großgruppe (Bivalvia, Heterodonta, Palaeoheterodonta, Protobranchia, Pteriomorphia) und Auswertungsart (single = Single State Merkmale, multi = Multi State Merkmale, weighted = differenziert gewichtet, bootstrap = mit Bootstrap-Analyse, size = differenziert gewichtet nach Größe, cmplx = differenziert gewichtet nach Komplexität)

7.5 Abbildungsverzeichnis

Abbildung 1: Klassische Phylogenie der Arthropoda nach morphologischen Merkmalen (Ax, 1999)	13
Abbildung 2: Neue Phylogenie der Arthropoda unter Einbeziehung von molekularen Daten, insbesondere der Analyse von Hox-Genen und rRNA.....	14
Abbildung 3: Stammbaum der eukaryotischen Großgruppen, ermittelt aus morphologischen und genetischen Informationen von 73 060 Arten. Stammbaum links: nur molekulare Daten; Stammbaum rechts: kombinierter Datensatz. Aus Goloboff (2009).	16
Abbildung 4: Verschiedene Muschelschalen: 1-3 Pteriomorphia (1 - Ostreoida, 2 - Pectinoidea, 3 - Mytiloidea), 4-6 Heterodonta (4 - Veneroida: Cardioidea, 5 - Myoidea: Myoidea, 6 - Veneroida: Solenoidea). Markiert sind charakteristische Merkmale: m – Muskelansatzstellen, ml – Mantellinie, s – Schloss mit Schlosszähnen.	22
Abbildung 5: Verschiedene Kiementypen bei den Bivalvia: A - Protobranchien, B - Septibranchien, C - Filibranchien, D - Eulamellibranchien (aus Westheide & Rieger, 1996).....	22
Abbildung 6: Ausgewählte Beispiele von hypothetischen Phylogenien der Bivalvia. Fünf Morphologie-basierte Stammbäume (A - E) und ein Stammbaum basierend auf der Analyse von 18S rDNA (F - Adamkewicz et al., 1997) (aus Adamkewicz et al., 1997).	25
Abbildung 7: Hypothetische Phylogenie der Bivalvia, angepasst nach verschiedenen Autoren.	26
Abbildung 8: Vergleich von fünf hypothetischen Stammbäumen der Pteriomorphia (aus Matsumoto, 2003). Oben: basierend auf morphologischen Merkmalen – Unten: basierend auf molekularen Untersuchungen.	28
Abbildung 9: Hypothetische Stammbäume der Palaeoheterodonta (aus Hoeh et al., 2009).....	30
Abbildung 10: Verschiedene Klassifizierungen der Heterodonta nach molekularen Untersuchungen (aus Taylor et al., 2007)	32
Abbildung 11: Hypothetische Phylogenien der Anomalodesmata. Links nach morphologischen Merkmalen, rechts nach 18S rDNA (aus Dreyer et al., 2003).	32
Abbildung 12: Die Eingabemaske des mfold Web Servers (http://mfold.rna.albany.edu/?q=mfold).....	45
Abbildung 13: Sekundärstruktur von <i>Pecten maximus</i> , nach RNAstructure. Links nach Prinzip MFE, rechts nach Prinzip MEA.....	47
Abbildung 14: Die Eingabemaske des RNAfold WebServers (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi)....	49
Abbildung 15: Faltung der 18S rRNA von <i>Pecten maximus</i> mit RNAsalsa unter Nutzung verschiedener Sequenzalignments.	53
Abbildung 16: Darstellung der Sekundärstruktur von <i>Pecten maximus</i> mit RNAViz.	53
Abbildung 17: Unterschiedliche Ergebnisse für die Sekundärstruktur der 18S rRNA von <i>Pecten maximus</i> nach Anwendung der verschiedenen Faltungsalgorithmen RNAfold, RNAsalsa, mfold, RNAstructure.	58
Abbildung 18: MFE-Sekundärstruktur der 18S rDNA von <i>Pecten maximus</i> nach dem RNAfold-Algorithmus	63
Abbildung 19: Ansicht der unalignierten Sequenzen in ClustalX.	69
Abbildung 20: Eingabemasken für Alignmentparameter bei ClustalX.	69
Abbildung 21: Ergebnis der Bearbeitung der Sequenzen aus Abbildung 19.....	72
Abbildung 22: NNC-Sekundärstruktur von <i>Pecten maximus</i> mit RNAsalsa.	73
Abbildung 23: Vergleich der NNC-Sekundärstruktur von <i>Abra prismatica</i> , gefaltet mit RNAsalsa (links) und RNAfold (rechts).	73
Abbildung 24: Doppelhelixstruktur der DNA aus Watson & Crick, 1953. Es wird schematisch dargestellt, wie die zwei Phosphat-Zucker-Ketten durch die horizontalen Bindungen der Basenpaare zusammengehalten werden.75	

Abbildung 25: Ausschnitt einer Sekundärstruktur nach RNAfold. Grau markiert ist die "Wobble"-Bindung G-U.	75
Abbildung 26: Ausschnitt einer Sekundärstruktur nach RNAfold. Grau markiert sind die vier grundlegenden Strukturelemente der Sekundärstruktur.	78
Abbildung 27: Verschiedene Typen von Pseudoknoten aus <i>Staple & Butcher (2005)</i> . Gestrichelte Linien stellen Basenpaarungen dar.	78
Abbildung 28: Berechnete Sekundärstrukturen von <i>Chlamys islandica</i> und <i>Ch. hastata</i> nach RNAfold, MFE. Farblich markiert sind drei übereinstimmende Merkmalskomplexe.	81
Abbildung 29: Herausgestellt ist jeweils ein Merkmalskomplex mit unterschiedlich ausgeprägten Merkmalen (Sekundärstruktur von <i>Mytilus edulis</i> und <i>Yoldia limatula</i> nach RNAfold), rot markiert ist die ursächliche Punktmutation.	81
Abbildung 30: Herausgestellt ist jeweils ein Merkmal (B6) mit unterschiedlichen Grundelementen (Sekundärstruktur von <i>Acila castrensis</i> und <i>Pecten maximus</i> nach RNAsalsa), rot markiert sind Punktmutationen, bzw. eine Insertion von einer Base.	83
Abbildung 31: Ausschnitt aus der Rohfassung der Merkmalsmatrix von der MFE-Faltung mit RNAfold. Untersuchte Arten sind in den Zeilen, Merkmale in den Spalten. Rot markiert sind "vorhandene" Merkmale.	89
Abbildung 32: Größenverteilung der Merkmale in einem Satz MFE-Faltungen. Auf der x-Achse ist die Merkmalsgröße x aufgetragen, als Verteilungsfunktion auf der y-Achse der Prozentsatz derjenigen Merkmale, die kleiner oder gleich der Größe x sind. Kleine und mittelgroße Merkmale kommen am häufigsten vor. Große oder sehr große Merkmale, die aus mehr als 33 Basen bestehen, stellen nur etwa 20%.	92
Abbildung 33: Die fünf häufigsten gefundenen Merkmalskomplexe in den Sekundärstrukturen nach NNC, mit RNAsalsa gefaltet. Merkmalskomplexe im Uhrzeigersinn: A, E, D, C, B.	94
Abbildung 34: Near Natural (NNC) nach RNAsalsa. Verschiedene Ausprägungen von Merkmalskomplexen am Beispielkomplex A mit hier insgesamt 17 definierten Merkmalen in: v.l.n.r. <i>Acila castrensis</i> , <i>Yoldia limatula</i> , <i>Nuculana pernula</i> , <i>Ostrea edulis</i> , <i>Dahurinaia dahurica</i> .	95
Abbildung 35: Near Natural (NNC) RNAsalsa. Verschiedene Ausprägungen von Merkmalskomplexen am Beispielkomplex C mit fünf Merkmalen, z.T. mit verschiedenen Ausprägungen in: v.l.n.r. <i>Acila castrensis</i> , <i>Anodonta sp.</i> und <i>Bathypecten vulcani</i> .	95
Abbildung 36: Near Natural (NNC) RNAsalsa. Verschiedene Ausprägungen von Merkmalskomplexen am Beispielkomplex B mit neun Merkmalen, jeweils mit verschiedenen Ausprägungen in: v.l.n.r. <i>Acila castrensis</i> , <i>Ensis ensis</i> und <i>Flexopecten glaber</i> .	96
Abbildung 37: Near Natural (NNC) RNAsalsa. Verschiedene Ausprägungen von Merkmalskomplexen am Beispielkomplex D mit 11 Merkmalen, z.T. mit verschiedenen Ausprägungen in: v.l.n.r. <i>Acila castrensis</i> , <i>Nuculana pella</i> , <i>Anomia ephippium</i> und <i>Pecten maximus</i> .	97
Abbildung 38: Near Natural (NNC) RNAsalsa. Verschiedene Ausprägungen von Merkmalskomplexen am Beispielkomplex E mit 14 Merkmalen, z.T. mit verschiedenen Ausprägungen in: v.l.n.r. <i>Acila castrensis</i> , <i>Mytilus edulis</i> und <i>Pecten maximus</i> .	97
Abbildung 39: Merkmalsmatrix für 159 Sekundärstrukturen gefaltet nach NNC mit RNAsalsa. V.l.n.r.: Protobranchia, Palaeoheterodonta, Pteriomorphia, Heterodonta. V.o.n.u.: Merkmalskomplexe A - E mit 74 Merkmalen in 194 Merkmalsversionen.	99
Abbildung 40: Variabilität der Ausprägung homologer Merkmale. 50% der definierten Merkmale haben zwei oder weniger homologe Ausprägungen. 80% der Merkmale haben weniger als fünf verschiedene Ausprägungen.	103
Abbildung 41: Maximaler Größenunterschied zwischen verschiedenen Ausprägungen eines Merkmals. 50% der Merkmalsausprägungen unterscheiden sich durch weniger als zwei Basen. Maximale Varianz sind 40 Basen.	103
Abbildung 42: Korrelation zwischen Größenunterschied und Anzahl homologer Ausprägungen eines Merkmals. Der Großteil der Merkmale liegt im linken unteren Viertel und hat nur eine geringe Variabilität.	104

Anhang

Abbildung 43: Ausschnitte aus den Merkmalsmatrizen von MFE (oben) und NNC (unten). Horizontal sind die Arten aufgetragen (v.l.n.r.: Protobranchia, Palaeoheterodonta, Pteriomorphia, Heterodonta), vertikal stehen die definierten Merkmale. Rot markiert sind vorhandene Merkmale. Bei MFE ist ein deutliches Rauschen zu erkennen, die Merkmale bei NNC zeigen eine größere Stetigkeit. 105

Abbildung 44: RNA Strukturdarstellungen. (a) Die Primärstruktur stellt die Reihenfolge der Nukleotide bzw. die Basensequenz der RNA dar. (b) Die Sekundärstruktur zeigt die gefaltete Sequenz mit den Paarbindungen zwischen den Basen, innerhalb eines Einzelstrangs. (c) Die Tertiärstruktur ist die Darstellung der dreidimensionalen, übergeordneten Anordnung der (in der Sekundärstruktur geformten) Doppelstrangregionen. Die RNA in dieser Abbildung ist ein Abschnitt aus der Isoleucin-tRNA des Purpurbakteriums *Azoarcus* (aus *Lehman, 2010*). 186

8. Danksagung

Diese Arbeit wurde teilweise unterstützt durch die DFG im Rahmen des Projekts SPP 1174 an der Zoologischen Staatssammlung München. Mein Dank gilt in erster Linie meinem Betreuer, Prof. Gerhard Haszprunar, für zahlreiche Hilfestellungen und wissenschaftliche Diskussionen bis zur endgültigen Fertigstellung dieser Arbeit. Der LMU München bin ich zu Dank verpflichtet für die Bereitstellung eines Arbeitsplatzes im BioZentrum in München-Martinsried. Weiterhin danke ich vielen Kollegen und Kolleginnen an der LMU, allen voran PD Dr. Martin Heß und Heidemarie Gensler, die wesentlich zu einer freundschaftlichen, produktiven und kreativen Arbeitsatmosphäre beitrugen. Dankbar bin ich Jana Hertel und Roman Stocsits (Arbeitsgruppe Stadler) von der Universität Leipzig, sowie Harald Letsch (Arbeitsgruppe Misof) vom Museum Koenig, Bonn, für ihre Hilfe und die Bereitschaft, mir ihre Software zur Berechnung von Sekundärstrukturen zum Vergleich zu überlassen. Letztlich möchte ich an dieser Stelle auch meinen Eltern einen herzlichen Dank aussprechen für ihre ausdauernde Unterstützung über den gesamten Zeitraum meiner Arbeit.