

**Phylogenetische Entwicklung asiatischer  
Wasserbüffel anhand Polymorphismen in der  
mitochondrialen D-loop Region**

Dissertation  
zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultäten  
der Georg-August-Universität zu Göttingen

vorgelegt von  
Gerold Kierstein  
aus Essen

Göttingen 2001

D7

Referent: Prof. Dr. R. Willmann

Korreferent: Prof. Dr. U. Ehlers

Tag der mündlichen Prüfung:

Die Wissenschaft von heute ist der Irrtum von morgen.

(Jacob von Uexküll, Zoologe)

## Danksagung

Die vorliegende Arbeit wurde unter der Anleitung von Prof. Dr. Dr. Bertram Brenig am Tierärztlichen Institut der Georg-August-Universität Göttingen angefertigt. Die Arbeit wurde von Prof. Dr. Rainer Willmann im Sinne der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultäten betreut und vor der Fakultät vertreten.

In erster Linie möchte ich mich ganz herzlich bei Prof. Brenig bedanken. Er hat mir nicht nur das Thema der Arbeit überlassen, sondern es war auch immer sein Anliegen jederzeit (selbst wenn uns 8.000 km trennten) ansprechbar zu sein, selbst für die scheinbar belanglosesten Fragen. Ohne die ausgezeichneten Arbeitsbedingungen in seinem Labor, besonders das Untere, wäre diese Arbeit auch nicht so zustande gekommen. Bei ihm und in seiner Arbeitsgruppe habe ich in den fast drei Jahren außerordentlich viel gelernt.

Bei Herrn Prof. Dr. R. Willmann bedanke ich mich für die Übernahme des Referats, ebenso bei Herrn Prof. Dr. U. Ehlers für die Übernahme des Korreferats im Fachbereich Biologie.

Und was wäre diese Arbeit ohne die geduldige – und manchmal auch ungeduldige – Sequenzierarbeit von Frau Alex? Danke für das „Drüberjagen“ der vielen Büffel über die Sequenale und für die Einführungen in das automatische Sequenzieren.

Nicht zu vergessen „minha preta“ Marcelo, der mir nicht nur beim „PAUPen“ geholfen hat, sondern mit ihm hatte ich auch viele fruchtbare Diskussionen (sem sacanagem!).

Frau Friedrichs, die gute Seele des Instituts, erleichterte die Arbeit nicht nur durch das Abnehmen vieler zeitraubender Arbeiten - nein, sie versüßte gar oft die Arbeit mit selbstgemachten Kuchen oder gab der Arbeit die rechte Würze mit selbstgemachter deftiger Wurst.

Ferner gilt mein Dank jenen (stellvertretend sei Christoph und Reza genannt), die es mit stoischer Ruhe ertrugen, wenn ich mal wieder gleich für mehrere Tage den Computer in Beschlag genommen habe, um aufwendige phylogenetische Kalkulationen durchzuführen.

Oft wird immer nur „der Arbeitsgruppe“ gedankt. Deshalb Danke an die, die noch nicht genannt wurden: Brigitte, Leonard, Stephan, Uli, Tosso, Andrea, Anne, "der Alex", Ina, Steffi, Melli, Paula, Artur, Leonardo (nicht DiCaprio) und alle anderen die ich jetzt doch noch vergessen habe.

Sonja, dir danke ich nicht nur für die vielen Hilfestellungen die ich im Laboralltag unbedingt brauchte - ich kam mit nur minimalen Kenntnissen in die Molekularbiologie und du nahmst mich nicht nur unter die fachlichen Fittiche. Daher ganz besonderer Dank für die Unterstützung, den Zuspruch und die Liebe auch nach Feierabend.

Meinen Eltern danke ich auch hier, die den Grundstein für meine Ausbildung gelegt haben und damit dieses hier letztendlich ermöglichten.

Teile diese Arbeit sind zur Veröffentlichung eingereicht:

Kierstein, G., Vallinoto, M., Silva, A., Schneider, M. P., Iannuzzi, L., & Brenig, B. (2001). Comparative analysis of the complete mitochondrial D-loop region casts new light on domestic water buffalo (*Bubalus bubalis*) phylogeny.

Die Sequenzen sind in der Gen Datenbank unter der Zugriffsnummer AF197196 – AF197223 abgelegt; EMBL AJ250508 – 250535.

## INHALTSVERZEICHNIS

<b>1</b>	<b><u>EINLEITUNG</u></b>	<b><u>1</u></b>
<b>2</b>	<b><u>LITERATURÜBERSICHT</u></b>	<b><u>3</u></b>
<b>2.1</b>	<b>MOLEKULARE EVOLUTION</b>	<b>3</b>
2.1.1	DIE GESCHICHTE DER MOLEKULAREN EVOLUTION	3
2.1.2	TERMINOLOGIE	4
2.1.2.1	Beschreibung eines phylogenetischen Baumes	4
2.1.2.2	Was ist ein Merkmal?	5
2.1.3	GENETISCHE DISTANZEN	5
2.1.3.1	Arten der Substitutionen	6
1.1.4	HOMOLOGIE UND HOMOPLASIE	7
1.1.5	EVOLUTION VON MOLEKÜLEN	8
1.1.5.1	Neutrale Evolution	9
1.1.5.2	Molekulare Uhr	9
1.1.6	MODELLE DER SEQUENZEVOLUTION	10
1.1.6.1	Jukes-Cantor (JC) -Modell	11
1.1.6.2	Kimuras 2-Parameter (K2P) -Modell	12
1.1.6.3	Felsenstein (F81) -Modell	12
1.1.6.4	Hasegawa, Kishino und Yano (HKY85) -Modell	12
1.1.6.5	Tamura und Nei (TrN) -Modell	13
1.1.6.6	Gammaverteilung	13
<b>1.2</b>	<b>ABLEITUNG MOLEKULARER PHYLOGENIE</b>	<b>14</b>
1.2.1	DISTANZ VERSUS DISKRETEM MERKMAL	14
1.2.2	GRUPPIERUNGSMETHODE VERSUS SUCHMETHODE	15
1.2.3	SUCHE NACH DEM OPTIMALEN BAUM	16
1.2.3.1	Exakte Suche	16
1.2.3.2	Heuristische Suche	18
1.2.4	DENDROGRAMME VERSUS SPLIT-ZERLEGUNG	19
1.2.5	DISTANZVERFAHREN	22
1.2.5.1	Unkorrigierte Distanzen	22
1.2.5.2	Minimum Evolution	22
1.2.5.3	Neighbor-joining	22
1.2.6	DISKRETE VERFAHREN	23
1.2.6.1	Maximum Parsimonie	23

---

1.2.6.2	Maximum Likelihood	23
1.2.7	ABSCHÄTZUNG DES STICHPROBENFEHLERS	24
1.2.7.1	Bootstrap-Test	24
1.2.7.2	Jackknifing	24
<b>1.3</b>	<b>MITOCHONDRIALE DNA</b>	<b>25</b>
1.3.1	DIE ENDOSYMBIONTENTHEORIE	26
1.3.2	MITOCHONRIALE DNA IN DER EVOLUTIONSFORSCHUNG	26
<b>1.4</b>	<b>PHYLOGENESE DER BÜFFEL</b>	<b>27</b>
<b>2</b>	<b>MATERIAL, METHODEN UND SOFTWARE</b>	<b>29</b>

---

<b>2.1</b>	<b>MATERIAL</b>	<b>29</b>
2.1.1	PROBENSAMMLUNG	29
2.1.2	ENZYMEN, NUKLEOTIDE, CHEMIKALIEN UND KITS	29
2.1.2.1	Enzyme	29
2.1.2.2	Nukleotide	29
2.1.2.3	Chemikalien	29
2.1.2.4	Kits	30
2.1.3	GLASWAREN UND VERBRAUCHSARTIKEL	30
2.1.4	LABORGERÄTE	31
<b>2.2</b>	<b>METHODEN</b>	<b>31</b>
2.2.1	DNA-EXTRAKTION AUS VOLLBLUT	31
2.2.1.1	DNA-Isolierung und phenolische Extraktion	31
2.2.1.2	DNA-Extraktion nach Miller <i>et al.</i> (1988)	32
2.2.2	AMPLIFIKATION DER MITOCHONDRIALEN D-LOOP REGION	32
2.2.3	DNA-SEQUENZIERUNG	34
2.2.3.1	Sequenzreaktion	34
2.2.3.2	Automatische Sequenzierung	34
2.2.4	DATENANALYSE	35
<b>2.3</b>	<b>SOFTWARE</b>	<b>35</b>
2.3.1	CLUSTALX	35
2.3.2	MACCLADE	37
2.3.3	MODELTEST	37
2.3.4	PAUP* (4.0B4A)	38
2.3.5	NETWORK 2.0C	38
2.3.6	SWAN V1.0	39

<b>3</b>	<b>ERGEBNISSE</b>	<b>40</b>
<b>3.1</b>	<b>DIE BUBALINEN HAPLOTYPEN</b>	<b>40</b>
3.1.1	DIVERSITÄT IN DEN EINZELNEN RASSEN	40
3.1.2	SEQUENZINFORMATIONEN DER BUBALINEN D-LOOP REGION	41
3.1.2.1	Sequenzlänge	41
3.1.2.2	Sequenzalignment	43
3.1.2.3	Basenfrequenz	43
3.1.2.4	Transitionen und Transversionen	43
3.1.2.5	Verteilung variabler Positionen im bubalinen Alignment	46
3.1.2.6	Längenpolymorphismen	47
3.1.2.7	Deletionen / Insertionen	48
3.1.2.8	Der Büffel Mur481	50
3.1.3	STRUKTURELEMENTE DER BUBALINEN D-LOOP REGION	50
3.1.3.1	Die 3'- und 5'-variablen Bereiche der D-loop Region	51
3.1.3.2	Sekundärstrukturen der D-loop Region	53
<b>3.2</b>	<b>PHYLOGENETISCHE ANALYSE</b>	<b>55</b>
3.2.1	MODELTEST	56
3.2.2	DENDROGRAMME FÜR 37_940	56
3.2.2.1	Maximum Likelihood	56
3.2.2.2	Maximum Parsimonie	57
3.2.2.3	Genetische Distanz	61
3.2.3	DENDROGRAMME FÜR 42_159	64
3.2.3.1	Maximum Parsimonie	65
3.2.3.2	Minimum Evolution	65
3.2.4	NETZWERKDARSTELLUNG	68
3.2.5	DIVERGENZ DER GENERA <i>BOS</i> UND <i>BUBALUS</i>	70
<b>4</b>	<b>DISKUSSION</b>	<b>72</b>
<b>4.1</b>	<b>ERSTMALIGE UNTERSUCHUNG ZUR DIVERSITÄT BRASILIANISCHER WASSERBÜFFEL</b>	<b>72</b>
<b>4.2</b>	<b>FLUß- UND SUMPFBÜFFEL LASSEN SICH DEUTLICH ANHAND DER POLYMORPHEN D-LOOP REGION UNTERSCHIEDEN</b>	<b>72</b>
4.2.1	DIE DIVERSITÄT DER RASSEN IST IN DEN POPULATIONEN VERSCHIEDEN	72
4.2.2	EIN HAPLOTYP IN DREI RASSEN	73
4.2.3	PHYLOGENESE DER WASSERBÜFFEL	74
4.2.4	WAS EINEN FLUßBÜFFEL ZUM SUMPFBÜFFEL MACHT	74



---

<b>4.3</b>	<b>FLUß- UND SUMPFBÜFFEL TRENNTEN SICH LANGE VOR DER DOMESTIKATION</b>	<b>75</b>
<b>4.4</b>	<b>PROBLEMATIK DER MOLEKULAREN EVOLUTION</b>	<b>77</b>
4.4.1	EINFLUß DES VERFÜGBAREN DATENUMFANGS AUF DIE AUSSAGEFÄHIGKEIT EINER PHYLOGENETISCHEN ANALYSE	78
4.4.2	PROBLEMATIK DES ART- UND RASSEBEGRIFFS	79
4.4.3	VERRAUSCHEN DER DATEN	79
<b>4.5</b>	<b>BUBALINE STRUKTURELEMENTE SIND HOMOLOG ZU DENEN ANDERER VERTEBRATEN</b>	<b>80</b>
4.5.1	MULTIPLE DEL/INS EREIGNISSE IN DER ETAS REGION	81
4.5.2	DIE FUNKTION DER ZENTRALEN DOMÄNE IST NOCH UNBEKANNT	81
4.5.3	DIE BUBALINE D-LOOP REGION ENTHÄLT DEN CSB3	81
4.5.4	DIE BUBALINEN D-LOOP REGION WEIST KEINE SEQUENZWIEDERHOLUNGEN AUF	82
<b>4.6</b>	<b>FORM DER SEKUNDÄRSTRUKTUREN IN DER D-LOOP REGION IST FRAGLICH</b>	<b>82</b>
4.6.1	MINIMALE FREIE ENERGIE FAVORISIERT NICHT DIE KLEEBLATTSTRUKTUR	83
<b>4.7</b>	<b>DIE WILDFORM DES <i>BUBALUS BUBALIS</i></b>	<b>84</b>
4.7.1	VERTEILUNG UND DERZEITIGER STATUS DER WILDFORM	84
<b>5</b>	<b>ZUSAMMENFASSUNG</b>	<b>86</b>
<hr/>		
<b>6</b>	<b>LITERATURVERZEICHNIS</b>	<b>88</b>
<hr/>		
<b>7</b>	<b>ANHANG</b>	<b>95</b>
<hr/>		

# 1 Einleitung

Bis heute ist relativ wenig über die Genetik der Wasserbüffel (*Bubalus bubalis* L.) bekannt. Das mag u. a. an ihrem in Europa relativ geringem ökonomischen Wert, verglichen mit dem der Rinder, liegen. Lediglich in Italien gibt es eine längere Tradition in der Büffelzucht. In den Entwicklungsländern besitzt der Wasserbüffel hingegen ein gewaltiges ökonomisches Potential. Der Bestand an domestizierten Wasserbüffeln wird weltweit auf mindestens 130 Mio. geschätzt - etwa  $\frac{1}{9}$  der weltweiten Rinderpopulation - von denen mehr Menschen abhängig sind als von jeder anderen domestizierten Spezies. Der Wasserbüffel liefert Zugkraft, Fleisch und/oder Milch. Um effiziente Zuchtprogramme zu entwickeln ist das Ausmaß und das Muster der genetischen Diversität innerhalb einer Rasse oder einer Population essentiell. Außerdem ist dieses Wissen Voraussetzung zum Erhalt genetischer Ressourcen.

Die Unterfamilie der Bovinae setzt sich aus den vier Gattungen *Bison*, *Bos*, *Bubalus* und *Syncerus* zusammen. Die beiden letzteren gehören zur Gruppe der Büffel, der asiatische Wasserbüffel bzw. afrikanische Kaffernbüffel. Diese Gattungen sind differente Linien, die sich von den anderen Bovinae schon während des Pliozäns getrennt haben. So finden sich Überreste von *Syncerus* aus dem späten Pliozän in Südafrika (Savage & Russell, 1983). Während des Pleistozäns erstreckte sich das Verbreitungsgebiet des *Bubalus* vom südlichen Asien bis nach Europa. Mit ansteigendem trockenerem Klima schrumpfte das Verbreitungsgebiet auf Indien, Indonesien und Südostasien (Nachtsheim & Stengel, 1977).

Der Wasserbüffel wird in die Subgenera *Bubalus* und *Anoa* mit vier lebenden Spezies unterteilt. Die Gruppe der Anoa umfasst den Flachland Anoa (*B. depressicornis*) und Hochland Anoa (*B. quarlesii*) auf Sulawesi, Indonesien und den Tamaraw (*B. mindorensis*) auf Mindoro, Philippinen. Der Wasserbüffel (*B. bubalis*) besiedelt einige Gebiete Indiens und wird als die Stammform des domestizierten Wasserbüffel angesehen. Dieser wird nicht nur aufgrund morphologischer, sondern auch aufgrund karyotypischer Merkmale in den Sumpf- ( $2n = 48$ ) und Flußtyp ( $2n = 50$ ) unterschieden (Fischer & Ulbrich, 1968; Ulbrich & Fischer, 1967). Morphologisch ist der Sumpftyp dem wilden Asiatischen Wasserbüffel (*B. arnee*) ähnlicher als es der Flußtyp ist.

Die Domestikation des Wasserbüffels fand wahrscheinlich in den Zivilisationen am Indus, am Yangtze und in der Euphrat und Tigris Ebene im dritten Jahrtausend v. Chr. statt (Cockrill, 1981; Nachtsheim & Stengel, 1977). Chen und Li schlagen eine Domestikation in China schon vor 7000 Jahren vor (Chen & Li, 1989). Aus dem asiatischen Raum gelangten die Büffel im 6. Jahrhundert nach Italien und erst in der jüngsten Vergangenheit wurden sie in Afrika, Australien, Süd- und Nordamerika eingeführt. Heute werden 19 Rassen unterschied-

den, von denen 18 zu den Flußbüffeln zählen; der Sumpfbüffel ist nur mit einer Rasse vertreten (Cockrill, 1981).

Sumpfbüffel sind hauptsächlich im östlichen Asien, von den Philippinen westwärts bis Indien verbreitet. Sie suhlen in jedwedem Gewässer oder in Schlamm. Primär dienen sie als Arbeitstiere, z. T. auch als Fleischproduzenten aber so gut wie nie als Milchlieferant. Flußbüffel sind in der westlichen Hälfte Asiens vertreten, von Indien bis Ägypten und Europa. Sie suhlen nur im klaren Wasser und dienen speziell als Milchlieferant.

Die Abschätzung der Phylogenie innerhalb der Bovinae ist noch immer Gegenstand zahlreicher und kontroverser Diskussionen. Groves schätzte einen gemeinsamen Vorfahren für alle Bovini 2-4 Mio. Jahre vor heute (Groves, 1981). Basierend auf fossilen Funden wird der Genus *Bubalus* erstmals im jungen Pliozän und frühen Pleistozän in Asien angetroffen, also ca. 1,8 Mio. bis 2,5 Mio. Jahre vor heute (Kowalski, 1986; Savage & Russell, 1983). In der jüngeren pleistozänen Fauna Südchinas sind unter den Huftieren viele rezente Arten bekannt, darunter auch *Bubalus bubalis*. Die molekularbiologischen Abschätzungen über die Trennung von Fluß- und Sumpfbüffel schwanken je nach Autor (und Methode) zwischen 10.000 Jahren und 1,7 Mio. Jahren vor heute (siehe Kapitel 2.4 Phylogenie der Büffel).

Die Bestimmung der Divergenz von Fluß- und Sumpfbüffel ist von Interesse, da die Abschätzungen einer jüngeren Trennung (also vor etwa 10.000 Jahren) beider Genera etwa mit dem Domestikationsbeginn der Büffel zusammenfällt. Dies würde bedeuten, daß die beiden Büffeltypen aus dem Domestikationsprozeß hervorgegangen sind. Eine frühere Trennung gäbe Hinweise auf mehrere Domestikationsprozesse.

Um eine möglichst genaue Abschätzung der bubalinen Phylogenie zu erhalten, wurde in dieser Arbeit die Sequenz der kompletten mitochondrialen D-loop Region von 80 Individuen verschiedener Rassen ermittelt. Die maternale Vererbung der mitochondrialen DNA (mtDNA) und damit das Ausbleiben von Rekombinationen, einhergehend mit einer 10-fach höheren Mutationsrate der D-loop Region verglichen mit der nukleären DNA, macht die mtDNA zu einem extrem wichtigen Untersuchungsobjekt in der Populationsgenetik und Systematik.

## 2 Literaturübersicht

### 2.1 Molekulare Evolution

#### 2.1.1 Die Geschichte der molekularen Evolution

Die molekulare Phylogenetik ist das "Studium evolutiver Verwandtschaft zwischen Organismen unter Zuhilfenahme molekularbiologischer Techniken" (Li & Graur, 1991). Im allgemeinen basieren phylogenetische Studien auf der Annahme, daß der Grad der Divergenz zwischen homologen Molekülen (DNA oder Proteine) verschiedener Taxa mit der Zeit korreliert, da diese Taxa sich getrennt evolutiv entwickelt haben.

Erste Untersuchungen in diese Richtung unternahm der Biologe George Nuttall um die Jahrhundertwende (Nuttall, 1904), als er Blutsera und Antisera verschiedener Spezies miteinander mischte, um eine "Blutsverwandtschaft" zwischen diesen zu zeigen. Die Idee war, daß näher verwandte Spezies eine stärkere Serum-Antiserum Kreuzreaktivität zeigen als entfernt verwandte Arten. Im Vergleich zu heute mutet diese Methode recht grob an, doch wurde damit der Grundstein zu Verwandtschaftsuntersuchungen auf (letztendlich) molekularbiologischer Ebene gelegt.

In den folgenden 50 Jahren wurden nur wenig weitere Fortschritte im Bereich der molekularen Evolution erzielt. Hauptsächlich weil sich das Datenmaterial nicht erweiterte. Dies änderte sich erst in den 50er Jahren. Hier waren zwei historische Entdeckungen von überragender Wichtigkeit: im Jahre 1953 die Entschlüsselung der molekularen Struktur der DNA durch James Watson und Francis Crick und zwei Jahre später die erste Publikation eines Vergleiches von Aminosäuresequenzen verschiedener Spezies von Frederick Sanger und Kollegen. In den folgenden Jahren wurden weitere Aminosäuresequenzen verschiedener Proteine bestimmt. Aus diesen Sequenzinformationen versuchte man nun evolutive Informationen zu extrahieren. Es galt also eine mathematische Beschreibung zu finden, die den Prozeß der Gensequenzveränderung über die Zeit beschreibt. Das einfachste Modell nimmt an, daß die molekulare Evolution ein hauptsächlich stochastischer Prozeß ist. Leider zeigt sich, daß die Sequenzunterschiede, die wir heute sehen, nicht unbedingt die tatsächlichen Sequenzunterschiede sind, die stattgefunden haben. Hier muß man zusätzlich zwischen Aminosäuresequenzen und DNA-Sequenzen unterscheiden.

Aminosäuren werden durch Basentriplets kodiert. Aufgrund des degenerierten genetischen Codes kann eine Aminosäure durch mehrere verschiedene Basentriplets codiert werden. Daraus resultiert die Problematik, daß auf der Ebene der Aminosäuresequenz Mutationen in der DNA-Sequenz verdeckt bleiben kann (stille Mutationen). Aber auch auf der Ebene der DNA können Mutationen durch multiple Substitutionen an einer Position unerkant bleiben. So bleibt es bis heute eines der wichtigsten und schwierigsten Ziele der Molekularevolution die wahre Anzahl akkumulierter Gensequenzunterschiede festzustellen. Ein weiteres und noch heute kontrovers diskutiertes Problem ist das Modell der "molekularen Uhr". Unter der Annahme, daß sich Mutationen in zeitlich regelmäßigen Intervallen manifestieren, hätte man ein Werkzeug zur Hand die evolutive Geschichte in einen Zeitrahmen zu stellen. Es zeigt sich jedoch, daß sich verschiedene Abschnitte der DNA verschieden schnell entwickeln; ein Punkt, der berücksichtigt werden muß.

## 2.1.2 Terminologie

In diesem Kapitel sollen einige Begriffe, die zur Beschreibung der molekularen Evolution dienen, erläutert werden.

### 2.1.2.1 Beschreibung eines phylogenetischen Baumes

Die Abbildung 2-1 illustriert die Terminologie die in dieser Arbeit zur Beschreibung phylogenetischer Bäume verwendet wird. Das ist nötig, da in der Literatur die Terminologie stark unter den Autoren variiert, ferner in den verschiedenen Disziplinen wie Mathematik und Biologie. In Klammern wird die synonyme englischsprachliche Terminologie benannt.

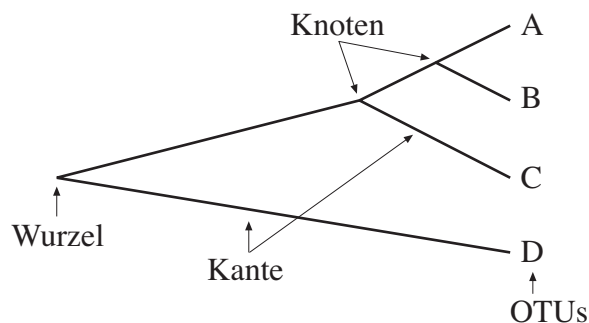


Abbildung 2-1 Baumterminologie.

Ein einfacher, wurzelloser phylogenetischer Baum mit der assoziierten Terminologie. Für Einzelheiten siehe Text.

Ein Stammbaumdiagramm (tree) ist ein mathematisches Konstrukt, um die evolutive Geschichte einer Gruppe von Sequenzen bzw. Organismen darzustellen. Synonyme Begriffe sind: Dendrogramm, Phylogramm, Kladogramm, Baumgraphik. Solch ein Baum besteht aus Knoten (nodes) die durch Kanten (branches oder edges) verbunden sind. Knoten symbolisieren Speziationsereignisse. Die genaue Bedeutung ergibt sich aus dem Kontext der Untersuchung, ob an diesem Punkt Genera, Arten oder bestimmte Grundmuster entstehen

bzw. entstanden sind. Dendrogramme werden fast immer als dichotome Diagramme dargestellt, d. h. jeder Speziation folgt die Separation in zwei divergierende Populationen. Multiple Speziationen, sie würden als polytome Verzweigung dargestellt, kommen in der Natur nur selten vor. Beispielsweise brachte eine sehr schnelle Radiation im Pliozän die heutigen Landkrabben der Gattung *Sesarma* auf Jamaika hervor. Die Speziationereignisse ließen sich nicht trennen, so daß fast gleichzeitig divergierende Linien als wahrscheinlich angenommen werden mußten (Schubart *et al.*, 1998).

Die terminalen Knoten (terminal nodes oder leaves) werden als terminales Taxon oder auch als OTUs (Operational Taxonomic Units) bezeichnet und repräsentieren die Sequenzen bzw. die Organismen für die entsprechende Daten vorhanden sind. Die internen Knoten (internal nodes) repräsentieren hypothetische Ancestoren. Der Ancestor aller Sequenzen die den Baum bilden ist die Wurzel (root) des Baumes. Oft ist der Ancestor, bzw. dessen DNA-Sequenz, nicht bekannt; dann sind die Bäume wurzellos ("unrooted"). Die Anordnung aller Taxa zueinander in einem Baum ist durch die Topologie des Baumes wiedergegeben.

#### 2.1.2.2 Was ist ein Merkmal?

Von der Grundbedeutung des Wortes her ist alles was wir an einem Objekt oder Organismus "bemerken" ein Merkmal. Das können physikalische Strukturen sein, Verhaltensmuster, biochemische Komponenten etc. In der Umgangssprache kann ein Merkmal verschiedene Bedeutungen haben. Das reicht vom real wahrnehmbaren Eigenschaften eines Organismus bis zur Homologie bzw. Übereinstimmung, die bei mehreren Organismen festgestellt werden. Man beachte, daß der Begriff Homologie schon eine weitreichende Hypothese beinhaltet (siehe auch Kapitel 1 in (Wägele, 2000)).

Für die molekularbiologische Untersuchung in dieser Arbeit werden die Nukleotide der untersuchten Sequenz als Merkmal bezeichnet. So besteht eine untersuchte Sequenz z. B. aus 940 Merkmalen, oder: das Individuum A hat bei Position 300 das Merkmal A (Adenin) während Individuum B das Merkmal G (Guanin) trägt.

Merkmalsänderungen beruhen entweder auf eine Zustandsänderung einer Homologie oder auf das Auftreten eines neuen Merkmals.

#### 2.1.3 Genetische Distanzen

Wie schon oben erwähnt, beruht die Idee der molekularen Phylogenetik auf der Annahme, daß die evolutive Distanz zweier Taxa mit deren genetischer Distanz korreliert. Das Problem ist allerdings, daß man die Geschichte einer Sequenz nicht kennt, sondern nur das Resultat aus der Geschichte. Damit sind DNA Sequenzen in gewisser Weise nicht sehr informativ: Vergleich zweier homologer Sequenzen offenbart lediglich, daß sie gleich sind oder nicht. Hinzu kommt die Schwierigkeit, daß eine Position nur vier verschiedene Merkmale aufweisen kann: Adenin (A), Cytosin (C), Guanin (G) und Thymin (T). Finden an einer Stelle mehr als eine Substitution statt, kann man keine Aussage über vorherige Substitutionen machen.

## 2.1.3.1 Arten der Substitutionen

Biochemisch betrachtet lassen sich Basensubstitutionen wie folgt klassifizieren. Basen sind entweder Purine (Adenin und Guanin) oder Pyrimidine (Cytosin und Thymin). Werden Purine bzw. Pyrimidine untereinander ausgetauscht spricht man von Transition. Wird ein Purin gegen ein Pyrimidin ausgetauscht (oder umgekehrt) spricht man von einer Transversion (Abbildung 2-2).

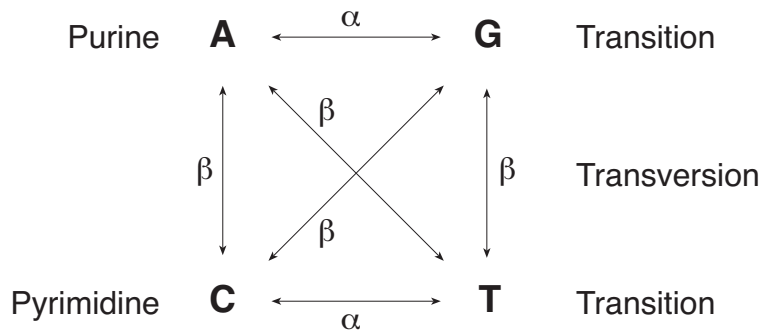


Abbildung 2-2 Mögliche Mutationen einer Nukleotidposition.

Transition ( $\alpha$ ): Austausch von Purinen oder Pyrimidinen. Transversion ( $\beta$ ): Austausch von Purinen gegen Pyrimidine und umgekehrt.

Betrachtet man zwei DNA-Sequenzen, können unterschiedliche oder identische Merkmale auf verschiedene Art erklärt werden. An einer Nukleotidposition findet man entweder einen Unterschied (verschiedene Basen) oder keinen Unterschied. An einem hypothetischem Beispiel soll dargestellt werden, wie ein Unterschied eines Merkmals zustande kommen kann. Es sei Adenin (A) das Merkmal an einer gegebenen Position eines gemeinsamen Vorfahren zwei Spezies. Eine einzelne Substitution in der Vergangenheit einer Art (z. B. eine A→C Transversion) führt zu einem Unterschied (Abbildung 2-3 a). Dies ist wahrscheinlich für z. B. nahe verwandte Arten. Bei entfernt verwandten Arten können in einer Linie auch multiple Substitutionen stattgefunden haben, z. B. eine A→C Transversion und später eine C→T Transition. Das Ergebnis ist auch hier ein Unterschied in der Nukleotidsequenz. Man würde hier die evolutive Veränderung unterschätzen (Abbildung 2-1 b). Bei einer koinzidentalen Substitution haben in der Vergangenheit beider Linien Nukleotidaustausche stattgefunden, in der einen Linie eine A→C und in der anderen Linie eine A→G Substitution (Abbildung 2-3 c).

Bei den drei Substitutionstypen in Abbildung 2-3 d)-f) ist die Interpretation der zu beobachtenden Merkmalszustände weitaus schwieriger. Abstammend von einer gemeinsamen ancestralen Spezies zeigen beide Abkömmlinge trotz multipler Nukleotidaustausche (zwei und mehr) identische Merkmale. Man spricht hier von Homoplasie, also identische Sequenzen, die nicht direkt von einem gemeinsamen Vorfahren abzuleiten sind.

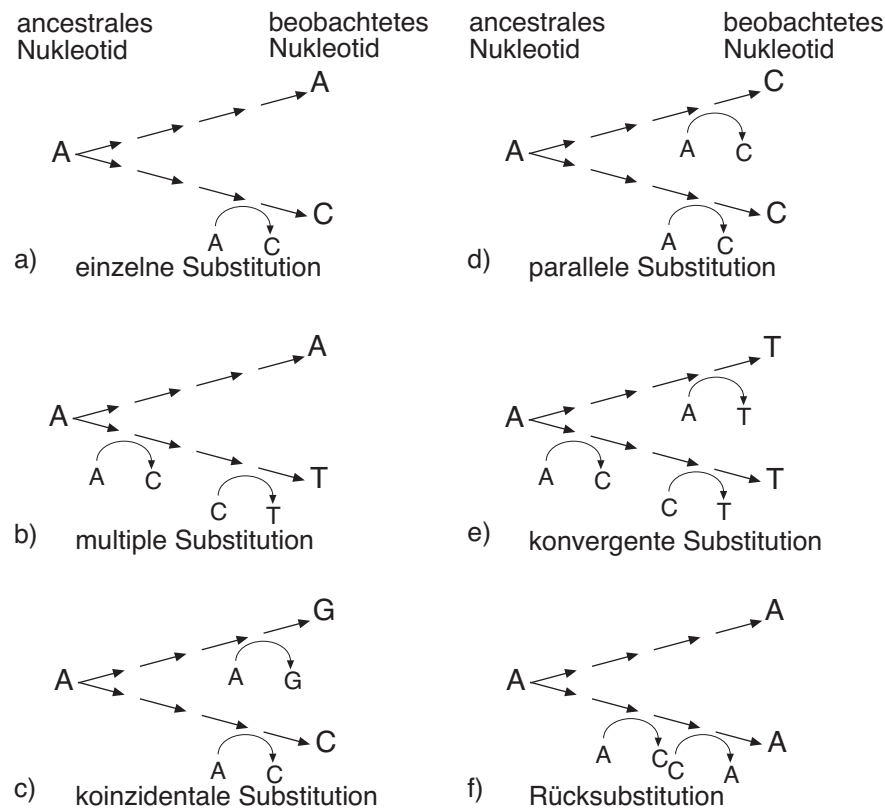


Abbildung 2-3 Sechs Möglichkeiten der Substitution. Dargestellt sind die möglichen Veränderungen an einer Position, ausgehend von einem A(denin). Bei den drei Formen auf der rechten Seite handelt es sich um multiple Substitutionen, bei denen die Nachfahren gleiche Nukleotide aufweisen. Sie machen die Interpretation von Sequenzvergleichen ungleich schwieriger (siehe Text). Verändert nach (Page & Holmes, 1998b)

#### 2.1.4 Homologie und Homoplasie

Selbst mit dem Beispiel aus Abbildung 2-3 bleiben die Begriffe der Homologie und Homoplasie nicht eindeutig. Deshalb sollen hier eingehend diese Begriffe erläutert werden. Homologie kann definiert werden als „eine genetisch fixierte Information [...], die von einem gemeinsamen Vorfahren der Arten, die das Merkmal aufweisen, geerbt wurde“ (Wägele, 2000). Diese Aussage impliziert gleichzeitig, daß der Begriff Homologie bereits eine Hypothese beinhaltet. So kann es vorkommen, daß ein Merkmal am Anfang einer Analyse als homolog bezeichnet wird, sich später aber als homoplasisch herausstellt. Solche Homoplasien sind mit einem Dendrogramm nicht kompatibel, lassen sich also nicht eindeutig darstellen. Es seien die informativen Merkmale der Spezies 1-4 wie folgt aufgeteilt: CT(1), CA(2), GA(3) und GA(4). Damit lassen sich zwei gleichwertige Dendrogramme konstruieren (Abbildung 2-4).

Zum einen eine  $\{\{1, 2\} \{3, 4\}\}$ , zum anderen eine  $\{1 \{2 \{3, 4\}\}\}$  Phylogenie. Hier sind in der ersten Topologie das Merkmal A, in der Zweiten das Merkmal C Homoplasien. Diese Konflikte in Verwandtschaftshypothesen lassen sich durch Konsensusdiagramme oder besser durch Netzwerkdiagramme (Kapitel 2.2.4 Dendrogramme versus Split-Zerlegung) darstellen.



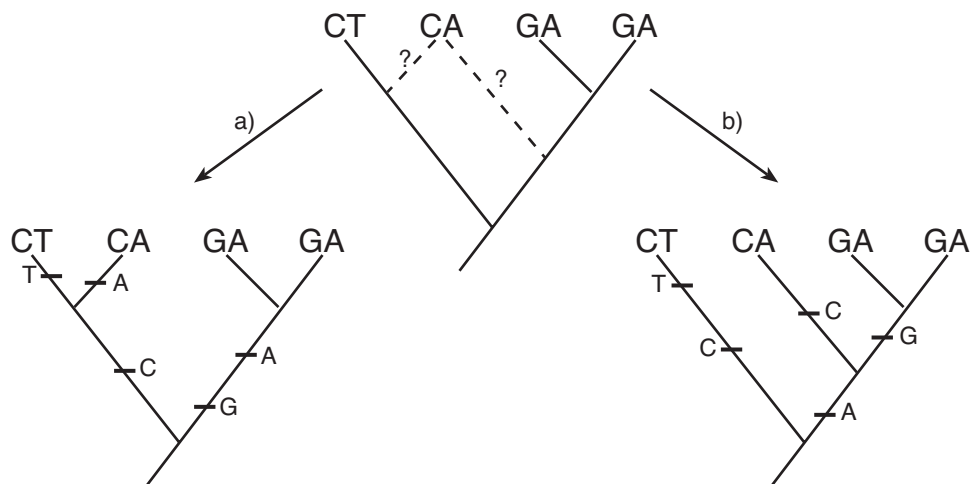


Abbildung 2-4 Beispiel für Homoplasie.

Für eine gegebene Merkmalsverteilung (oben) lassen sich zwei verschiedene Dendrogramme konstruieren, die die Merkmalsverteilung gleich gut beschreiben. Hier sind in der Topologie a) das Merkmal A, in der Topologie b) das Merkmal C Homoplasien. Die horizontalen Balken markieren das Auftreten eines Merkmals. (Verändert nach (Wägele, 2000)).

### 2.1.5 Evolution von Molekülen

Die Annahme, daß Evolution stattgefunden hat und noch immer stattfindet, begründet die phylogenetische Forschung. Dabei ist die Evolution wertungsfrei die Veränderung auf der Zeitachse. Zu unterscheiden ist die morphologische von der molekularen Veränderung.

Die Morphologie von Organismen unterliegt in hohem Maße der Selektion, damit also den Umwelteinflüssen, wie Klimaveränderung, Auftreten von neuen Feinden, Krankheitserregern oder Nahrungsquellen. Da diese Randbedingungen nicht vorhersagbar sind, ist folglich auch die morphologische Evolution, insbesondere ihre Geschwindigkeit, nicht vorhersagbar. Beispiele für diesen chaotischen Prozeß gibt es genug in der Natur. Fossile Funde belegen, daß bis in die Kreidezeit Dinosaurier und Säugetiere koexistierten. Mit dem Aussterben der Saurier wurden offenbar Nischen frei, die eine rapide Entwicklung der Säugetiere nach sich zog; es entstanden in einem kurzen Zeitrahmen die Vorfahren der meisten heute lebenden Mammalia. Ebenso finden sich in der rezenten Fauna Beispiele für artspezifische morphologische Variabilität. Hier sind besonders Organe betroffen, die nicht primär der Lebenserhaltung dienen. Während die lebenswichtigen Augen bei allen Wirbeltieren einen ähnlichen Aufbau haben - für das Sehen optimiert sind -, findet sich z. B. bei den Bovidae ein Vielzahl verschieden geformter Hörner.

Der Ursprung aller morphologischen Veränderungen liegt in den Molekülen, letztendlich in den Veränderungen der Moleküle der Erbsubstanz, der DNA. Hier ist es interessant, daß es offenbar Sequenzabschnitte oder Positionen gibt die nicht oder nur wenig dem Selektionsdruck unterliegen, also selektionsneutral sind und solche, die extrem selektiert werden. Das zeigt sich an so extremen Beispielen, wie die ribosomale DNA, die bei Einzellern, Pflanzen

und Tieren homologisiert werden kann, während auf der anderen Seite die hypervariablen Bereiche der mitochondrialen D-loop Region individuelle Variationen aufweisen. Dieses Erkenntnis ist wichtig bei der Wahl für phylogenetische bzw. populationsgenetische Untersuchungen.

#### 2.1.5.1 Neutrale Evolution

Bis in die Mitte des 20ten Jahrhunderts beschränkte sich die phylogenetische Forschung auf die Analyse der morphologischen Veränderungen. Aus diesem Zusammenhang war dann auch die Evolution ein Prozeß, der hauptsächlich durch Mutation und Selektion bestimmt war. Je mehr molekulare Daten zur Verfügung standen, um so mehr wurde klar, daß Moleküle ohne Einwirkung der Selektion evolvieren, erkennbar an Polymorphismen auf Ebene der Moleküle, die sich in der Morphologie aber nicht widerspiegelten. Daraus entstand die Theorie der neutralen molekularen Evolution, die erstmals von Motoo Kimura 1968 und 1969 von Jack King und Tom Jukes vorgeschlagen wurde (Kimura, 1968; King & Jukes, 1969). Wohlgermerkt, die Evolution ist nicht grundsätzlich neutral, aber es gibt selektionsneutrale Mutationen. Dazu gehören Mutationen, die in der Aminosäuresequenz keine Veränderungen bewirken (synonyme Substitutionen), die die Transkription oder die Funktionen von RNA-Molekülen nicht beeinflussen. Ebenso sind Mutationen von funktionslosen Sequenzpositionen neutral (z. B. intronische Bereiche, repetitive Elemente oder Abschnitte der mitochondrialen D-loop Region). Eine wichtige Folgerung aus der neutralen Mutation ist der fast ausschließliche Einfluß der genetischen Drift auf die Allelfrequenz, und damit die relative Gleichförmigkeit der Substitutionsraten.

#### 2.1.5.2 Molekulare Uhr

Erfolgen Substitutionen in konstanter Gleichförmigkeit, kann man dies gleichsetzen mit einem stochastischen, regelmäßigen Zufallsprozeß. So wie der radioaktive Zerfall von  $^{14}\text{C}$  zur Datierung fossiler Funde genutzt wird, hat der Phylogenetiker mit der molekularen Uhr ein einfaches Instrument in Händen, mit der die Divergenz verschiedener Spezies einfach durch Vergleich der DNA-Sequenzen abgeschätzt werden kann. Dabei ist die Zeit  $t$  seit der Divergenz zweier Linien proportional zur genetischen Distanz  $d$  dieser Linien (genauer: zur  $\frac{1}{2} d$ , um die Anzahl der Substitutionen beider Linien zu berücksichtigen) und umgekehrt proportional zur Substitutionsrate  $k$ . Es gilt:

$$t = \frac{d}{2k}$$

Das Konzept der molekularen Uhr sagt nicht, daß alle Gene mit der selben Rate evolvieren. Vergleiche verschiedener Proteine zwischen verschiedenen Spezies zeigen eine starke Variabilität der Substitutionsraten. Das Ubiquitin hat eine Substitutionsrate von  $0,01 \cdot 10^{-9}$  Aminosäuresubstitutionen pro Position und Jahr, während das Fibrinopeptid mit fast  $9 \cdot 10^{-9}$  Substitutionen pro Position und Jahr evolviert (Nei, 1987). In der Diskussion über die Existenz der molekularen Uhr konzentriert man sich auf die Frage, ob die Substitutionsrate in einem Gen konstant über die evolutive Zeit ist. Der langzeitliche Durchschnitt schließt dabei nicht die Möglichkeit kurzfristiger Fluktuationen in der Substitutionsrate aus. Der Analyst

macht sich die Beobachtung der genspezifischen Substitutionsraten zu eigen und wählt für seine spezifische Untersuchung die Sequenz (oder den Sequenzabschnitt), die ihm die meiste Information liefert. Die Phylogenie stark divergierender Taxa würden z. B. anhand kodierender Sequenzen untersucht werden, die unter dem Druck stabilisierender Selektion stehen. Neutral evolvierende Sequenzen würden bei entfernt verwandten Taxa relativ schnell durch multiple Substitutionen verrauschen. Sie sind aber hingegen für nahe verwandte Taxa geeignet. Ein Indiz für verrauschte Daten gibt das Verhältnis der Transitionen zu Transversionen gegen die unkorrigierte Distanz an. Transversionen akkumulieren wesentlich langsamer und haben einen linearen Verlauf. Transitionen treten hingegen sehr viel häufiger auf, haben einen schnellen Anstieg, bis es durch multiple Substitutionen bald zu einer Sättigung kommt und ein Plateau erreicht wird.

### 2.1.6 Modelle der Sequenzevolution

Vergleicht man zwei Sequenzen miteinander, muß man sich aufgrund der oben beschriebenen Problematik fragen, wie groß der tatsächliche Unterschied zwischen ihnen ist. Der einfachen Annahme, die Anzahl der Nukleotidunterschiede zwischen zwei Sequenzen gebe eine gute Annäherung an den evolutiven Unterschied beider Sequenzen wieder, kann man nach Kapitel 2.1.3.1 (Arten der Substitutionen) nur bedingt folgen. Vielmehr muß man sich vorstellen, daß im Lauf der Evolution Nukleotidsubstitutionen immer häufiger auch an schon substituierten Positionen stattfinden ("multiple hits"). Aus der Arbeit von Janecek und Kollegen (Janecek *et al.*, 1996), die die Evolution der Nukleotidsequenz des mitochondrialen Cytochrom C Oxidase II Gens innerhalb der Bovinae untersuchten, läßt sich eine nichtlineare Evolution ableiten. Spezieslinien die sich vor 20 Mio. Jahren trennten, zeigen nicht mehr Sequenzunterschiede als solche, die sich erst vor 15 Mio. Jahren trennten. Je mehr Substitutionen sich zwischen zwei Sequenzen akkumulieren, desto mehr kommt es zu einer Sättigung. Daraus folgt, daß die beobachteten Sequenzunterschiede die tatsächlichen evolutiven Sequenzveränderungen unterschätzen. Deshalb sind in der Vergangenheit viel Methoden entwickelt worden, um eine Distanzkorrektur zu berücksichtigen. Die verschiedenen Modelle sind stark miteinander verwandt und unterscheiden sich hauptsächlich in der Anzahl der berücksichtigten Parameter. Im Laufe der Jahre, mit zunehmenden Rechenkapazitäten der Computer, konnten immer mehr Parameter in den Modellen berücksichtigt werden.

Unter der Voraussetzung, daß die Nukleotidsubstitution über die Zeit konstant bleibt und die Basenzusammensetzung der Sequenz im Equilibrium ist, kann man in einer einfachen Matrix die Substitutionswahrscheinlichkeit als

$$p_t = \begin{bmatrix} p_{AA} & p_{AC} & p_{AG} & p_{AT} \\ p_{CA} & p_{CC} & p_{CG} & p_{CT} \\ p_{GA} & p_{GC} & p_{GG} & p_{GT} \\ p_{TA} & p_{TC} & p_{TG} & p_{TT} \end{bmatrix} \quad (1)$$

darstellen, wobei  $p_{AC}$  die Wahrscheinlichkeit wiedergibt, daß im Zeitintervall  $t$  an einer gegebenen Position ein Austausch von A nach C stattfindet, u.s.w. Besondere Beachtung finden die diagonalen Werte ( $p_{AA}, p_{CC}, p_{GG}, p_{TT}$ ). Sie beschreiben die Wahrscheinlichkeit, daß (anscheinend) keine Substitution stattgefunden hat; eine bestimmte Position trägt z. B. zum Zeitpunkt 0 das Merkmal A und zum Zeitpunkt  $t$  ebenfalls ein A. Es ist nicht möglich zu entscheiden, ob hier keine Substitution stattgefunden hat, oder das Beobachtete Ergebnis multipler Substitutionen ist. Die Wahrscheinlichkeit dafür ist 1 minus die Wahrscheinlichkeit für die Substitution des A durch ein C, G oder T. Formell läßt sich diese Wahrscheinlichkeit als

$$p_{ii} = 1 - \sum_{j \neq i} p_{ij} \quad (2)$$

beschreiben.

Ferner kann die Basenzusammensetzung als Vektor dargestellt werden als

$$\mathbf{f} = [f_A f_C f_G f_T] \quad (3)$$

wobei  $f_A$  die Equilibriumfrequenz der Base A beschreibt, u.s.w. Die Modelle, die die folgenden Kapitel ausmachen, sind die bekanntesten und sollen deshalb kurz erläutert werden.

#### 2.1.6.1 Jukes-Cantor (JC) -Modell

Das JC-Modell war eines der ersten und das vielleicht einfachste Modell, welches gleiche Basenfrequenzen annimmt und die Wahrscheinlichkeit von Transition und Transversion gleich setzt (Jukes, 1969). Anders ausgedrückt: der Einfluß von Selektion auf einzelne Genabschnitte wird nicht berücksichtigt und die Sequenzevolution wird als mechanischer Zufallsprozeß behandelt. Die Divergenz  $d$  zwischen zwei Sequenzen wird über den natürlichen Logarithmus  $\ln$  berechnet.

$$d = -\frac{3}{4} \ln\left(1 - \frac{4}{3} p\right) \quad (4)$$

wobei  $p$  den Anteil der Merkmale angibt, die in den Frequenzen unterschiedlich sind. Ausgedrückt in der Form wie in Gleichungen (1) und (3) beschrieben, läßt sich die Matrix der Substitutionswahrscheinlichkeit und der zugehörige Basenvektor wie folgt darstellen:

$$p_t = \begin{bmatrix} . & \alpha & \alpha & \alpha \\ \alpha & . & \alpha & \alpha \\ \alpha & \alpha & . & \alpha \\ \alpha & \alpha & \alpha & . \end{bmatrix}, \mathbf{f} = \left[ \frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{1}{4} \right] \quad (5)$$

Hier ist  $\alpha$  die Wahrscheinlichkeit einer Substitution (für alle gleich groß); die Werte für die Diagonale ( $p_{AA}, p_{CC}, p_{GG}, p_{TT}$ ) wird durch Gleichung (2) beschreiben. Jedes der vier Merkmale (A, C, G und T) ist mit einer Frequenz von 25 % vertreten.

### 2.1.6.2 Kimuras 2-Parameter (K2P) -Modell

Es zeigt sich allerdings bei Sequenzvergleichen, daß Transitionen häufiger vorkommen als Transversionen (Janecek *et al.*, 1996). Dies gilt speziell für die mtDNA, wie es auch in dieser Arbeit gezeigt wird. Transitionen akkumulieren wesentlich schneller als Transversionen und erreichen eine Sättigung, während Transversionen sehr viel langsamer, annähernd linear, über die Zeit akkumulieren. Im K2P-Modell (Kimura, 1980) wird dieser Beobachtung Rechnung getragen, indem die Transitionsrate ( $\alpha$ ) je Position und die Transversionsrate ( $\beta$ ) je Position zu einer totalen Substitutionsrate von  $\lambda = \alpha + 2\beta$  aufaddiert wird. Anders ausgedrückt: für ein Nukleotid gibt es drei Möglichkeiten der Substitution, von denen eine eine Transition ist, und zwei Transversionen sind (siehe Abbildung 2-2).

Die Substitutions-Wahrscheinlichkeitsmatrix und der Basenfrequenz Vektor hat dann folgende Form:

$$p_t = \begin{bmatrix} . & \beta & \alpha & \beta \\ \beta & . & \beta & \alpha \\ \alpha & \beta & . & \beta \\ \beta & \alpha & \beta & . \end{bmatrix}, \mathbf{f} = \left[ \frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{1}{4} \right] \quad (6)$$

### 2.1.6.3 Felsenstein (F81) -Modell

Ein weiterer Punkt der beachtet werden muß ist die Variation der Basenzusammensetzung. Zum Beispiel ist die mtDNA der Insekten AT reicher als die der Vertebraten. Man kann daher annehmen, daß Substitutionen der Base Guanin in einer Sequenz relativ selten sind, wenn diese Base in einem relativ geringem Anteil vorliegt. Dieser Umstand wird in Felsensteins Modell (F81) berücksichtigt (Felsenstein, 1981). Da die Basenzusammensetzung zwischen den zu vergleichenden Sequenzen unterschiedlich sein kann, wird eine Näherung in Form der durchschnittlichen Basenfrequenz gewählt. Mit  $\pi_i$  als durchschnittliche Frequenz der Base  $i$  über die Sequenzen hat das F81-Modell folgende Form

$$p_t = \begin{bmatrix} . & \pi_C \alpha & \pi_G \alpha & \pi_T \alpha \\ \pi_A \alpha & . & \pi_G \alpha & \pi_T \alpha \\ \pi_A \alpha & \pi_C \alpha & . & \pi_T \alpha \\ \pi_A \alpha & \pi_C \alpha & \pi_G \alpha & . \end{bmatrix}, \mathbf{f} = [\pi_A \quad \pi_C \quad \pi_G \quad \pi_T] \quad (7)$$

### 2.1.6.4 Hasegawa, Kishino und Yano (HKY85) -Modell

Hasegawa und Kollegen (Hasegawa *et al.*, 1985) haben durch Verbindung der Modelle K2P und F81 das HKY85-Modell entwickelt. Dieses gestattet unterschiedliche Transitions- und Transversionsraten, kombiniert mit variabler Basenfrequenz. Es ergibt sich:

$$p_t = \begin{bmatrix} . & \pi_C \beta & \pi_G \alpha & \pi_T \beta \\ \pi_A \beta & . & \pi_G \beta & \pi_T \alpha \\ \pi_A \alpha & \pi_C \beta & . & \pi_T \beta \\ \pi_A \beta & \pi_C \alpha & \pi_G \beta & . \end{bmatrix}, \mathbf{f} = [\pi_A \quad \pi_C \quad \pi_G \quad \pi_T] \quad (8)$$

#### 2.1.6.5 Tamura und Nei (TrN) -Modell

Tamura und Nei (Tamura & Nei, 1993) haben das HKY85-Modell um eine Variable zum TrN-Modell erweitert. Bei der Untersuchung der humanen mitochondrialen D-loop Region stellten sie fest, daß die Transitionsrate zwischen Purinen bzw. zwischen Pyrimidinen unterschiedlich ist. Die relative Substitutionsrate zwischen Purinen (A↔G) in den humanen hypervariablen Regionen der D-loop Region ist nur etwa halb so hoch, wie die Pyrimidin Substitutionen. Sie erweiterten daraufhin das Modell um einen Parameter für Transitionen zwischen Purinen ( $\alpha_1$ ) und Transitionen zwischen Pyrimidinen ( $\alpha_2$ ). Entsprechend ist:

$$p_t = \begin{bmatrix} . & \pi_C \beta & \pi_G \alpha_1 & \pi_T \beta \\ \pi_A \beta & . & \pi_G \beta & \pi_T \alpha_2 \\ \pi_A \alpha_1 & \pi_C \beta & . & \pi_T \beta \\ \pi_A \beta & \pi_C \alpha_2 & \pi_G \beta & . \end{bmatrix}, \mathbf{f} = [\pi_A \quad \pi_C \quad \pi_G \quad \pi_T] \quad (9)$$

Zusammenfassend läßt sich sagen, daß alle Modelle auf dem JC-Modell aufbauen, indem sie mehr oder weniger viele Parameter in das Modell mit einbeziehen. Somit lassen sich die Modelle fast unbegrenzt erweitern und modifizieren. Es stellt sich daher die Frage, welches das beste Modell für eine vorliegende Untersuchung ist. Die einleuchtendste Antwort wäre, das realistischste Modell zu wählen. Realität hat jedoch seinen Preis. Denn je mehr Parameter für eine phylogenetische Untersuchung berücksichtigt werden müssen, desto aufwendiger wird die Berechnung.

#### 2.1.6.6 Gammaverteilung

Um die evolutiven Distanzen zwischen Sequenzen zu beschreiben, werden in allen Modellen Annahmen über die Wahrscheinlichkeit einer Substitution gemacht. Dabei werden Art des Nukleotids, seine relative Häufigkeit etc. beachtet. Eine Annahme ist aber bei allen Modellen gleich, daß jede Nukleotidposition die gleiche Wahrscheinlichkeit einer Substitution besitzt. Diese Annahme vereinfacht die Mathematik auf Kosten der biologischen Wirklichkeit. Ein Vielzahl von Untersuchungen hat gezeigt, daß die relativen Substitutionsraten in verschiedenen Abschnitten von Genen und Pseudogenen unterschiedlich sind (Graur, 1984; Li *et al.*, 1981; Li *et al.*, 1984).

Die variablen Substitutionsraten haben einen starken Einfluß auf die Sequenzdivergenz. Es kann zu Unter- oder Überschätzung der genetischen Divergenz führen.

Die Entwicklung eines geeigneten Modells zur Berücksichtigung der Variation war gefordert. Die Gamma- ( $\Gamma$ -) verteilung (Yang, 1994) ist die meistgenutzte Herangehensweise. Die

Verteilung hat einen Kurvenparameter  $\alpha$  der die Variation über die Positionen spezifiziert. Kleine  $\alpha$  Werte führen zu einer L-förmigen Verteilung mit extremer Variationsrate; also viele Positionen sind invariabel aber einige wenige Positionen haben eine hohe Substitutionsrate. Es ist klar, der Wert für den Kurvenparameter  $\alpha$  läßt sich nicht aus den beobachteten Daten direkt ableiten. Man kann aber (mit Hilfe entsprechender Software) versuchen diesen Wert zu optimieren, damit die Modelle und die daraus konstruierten Dendrogramme die Daten möglichst gut widerspiegeln.

## 2.2 Ableitung molekularer Phylogenie

Die Aufgabe des molekularen Phylogenetikers ist es beobachtete molekulare Information (z.B. homologe Nukleotidsequenzen verschiedener Spezies oder Individuen) in einen evolutiven Zusammenhang zu stellen. Die gebräuchlichste Darstellungsweise dieser Ergebnisse erfolgt als ein evolutiver Stammbaum. Mittlerweile gibt es eine Vielzahl von Methoden, um diese Aufgabe lösen. Daraus ergibt sich allerdings die Frage, welche für einen gegebenen Datensatz die richtige ist. Zwei Wege sind gebräuchlich die Methoden zu unterscheiden. Einmal nach den Methoden wie sie die Daten verwerten oder nach der Herangehensweise wie sie Bäume konstruieren. Eine andere Form der Darstellung ist das Netzdiagramm, welches sich aus einer Splitzerlegung ergibt.

### 2.2.1 Distanz versus diskretem Merkmal

Diese Unterscheidung basiert auf der Art wie die Daten verwendet werden. Distanz Methoden verwandeln ein Sequenzalignment in eine Paarweise Distanzmatrix und nehmen diese Werte zur Berechnung in der Baumkonstruktions Methode. Die diskrete Methode nutzt die gesamte Sequenz (bzw. ihre relevanten Positionen) zur phylogenetischen Ableitung. Zur Veranschaulichung sei hier das Beispiel einer Sequenzmatrix und einer Distanzmatrix dargestellt (Abbildung 2-5).

Die Matrizes führen sowohl bei der Parsimonie Methode (Diskret) als auch bei der "Minimum Evolution" Methode (Distanz) zu einer gleichen Topologie und Verzweigungslänge des resultierenden Baumes. Trotzdem liefert der Parsimonie Baum zusätzliche Informationen darüber, welche Sequenzposition zu welcher Kantenlänge führt. Bei der Distanzmatrix ist diese Information verloren.

## Sequenzen

	Position							
	1	2	3	4	5	6	7	
Sequenz	A	T	T	A	T	T	A	A
B	A	A	T	T	T	A	A	
C	A	A	A	A	A	T	A	
D	A	A	A	A	A	A	T	

## Distanzen

	A	B	C
Sequenz	A		
B	3		
C	5	4	
D	5	4	2

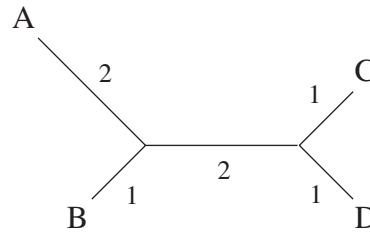
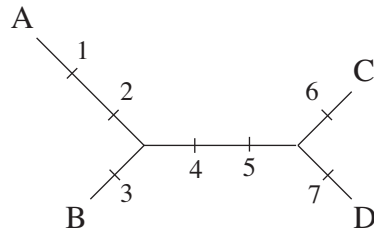


Abbildung 2-5 Vergleich von Sequenz- und Distanzmatrix.

Dargestellt sind vier Sequenzen A-D mit entsprechenden (unkorrigierten) Distanzen. Die daraus erhaltenen Bäume nach dem Parsimoniekriterium (Diskret, linke Seite) und der Minimum Evolution Methode (Distanz, rechte Seite) haben die selbe Topologie und Kantenlänge. Der linke Baum zeigt jedoch zusätzlich, welche Sequenzposition zur Länge einer jeden Kante beiträgt. Verändert nach (Page & Holmes, 1998a).

### 2.2.2 Gruppierungsmethode versus Suchmethode

Die zweite Form der Unterscheidung ist die Art wie die Methoden phylogenetische Bäume konstruieren. Die Gruppierungsmethode folgt einem bestimmten Algorithmus und erhält am Ende einen Baum. Man beginnt mit drei Sequenzen, da es nur eine Möglichkeit gibt drei OTUs in einem wurzellosen (unrooted), dichotomen Baum anzuordnen. In der ersten Runde wird berechnet wo die vierte Sequenz zu plazieren sei. Man erhält so einen Baum für vier Sequenzen, der als Ausgangsbaum für die nächste Runde dient u. s. w. Der Vorteil dieser Methode ist die einfache Ausführung, die zu schnellen Computerprogrammen führt. Ferner wird fast immer ein einziger phylogenetischer Baum kalkuliert. Die Kombination von Schnelligkeit und eindeutigem Ergebnis ist sehr verlockend. Man muß jedoch die analytischen Einschränkungen dieser Methode beachten. Die Form der endgültigen Phylogenese hängt oft von der Reihenfolge der addierten Sequenzen ab. Hier ist darauf zu achten, daß die Sequenzen in zufälliger Reihenfolge dazu gruppiert werden. Der größte Nachteil ist jedoch, daß die Evaluierung kompetitiver Hypothesen nicht durchgeführt werden kann, man erhält lediglich einen phylogenetischen Baum. Es könnte aber sein, daß zwei unterschiedliche Bäume die beobachteten Daten gleich gut erklären können. Wird aber keine Güte zwischen dem konstruierten Baum und den beobachteten Daten berechnet, wird man es nicht bestimmen können.

Die zweite baumkonstruierende Methode verwendet ein Optimalitätskriterium, um aus einer Vielzahl von Möglichkeiten die Beste zu wählen. Dieses Kriterium teilt jedem Baum einen



Wert zu, als Funktion der Verwandtschaft zwischen Baum und Daten, wie es bei Maximum Parsimonie und Maximum Likelihood der Fall ist. Ferner erlaubt diese Methode die Abschätzung der Qualität eines jeden Baumes. Konkurrierende Hypothesen über die evolutive Verwandtschaft können daraufhin überprüft werden, wie gut sie zu den Daten passen. Nachteil dieser Methode ist der enorme Rechenaufwand. Es konnte nachgewiesen werden, daß die Suche nach dem kürzesten Baum nicht durch einen effektiven Algorithmus zu lösen ist (Foulds & Graham, 1982). Es müssen daher folgende Aufgaben gelöst werden. Für einen gegebenen Datensatz müssen alle möglichen dichotomen, ungewurzelten Topologien konstruiert werden. Für jeden dieser Bäume wird der Wert unter dem bestehenden Optimalitätskriteriums bestimmt, also z. B. die Länge eines jeden Baums. Welcher von allen möglichen Bäumen hat dann den höchsten Wert des Kriteriums, also welches ist der kürzeste Baum? Im Falle, daß mehrere Bäume die selbe Länge aufweisen, wird ein Konsensusbaum konstruiert. Die Lösung dieser Probleme ist möglich für eine geringe Anzahl zu untersuchender Taxa, wird aber ein nicht zu lösendes Problem darstellen, wenn die Anzahl der OTUs zu hoch wird (siehe weiter unten).

### 2.2.3 Suche nach dem optimalen Baum

Die Methoden, die unter einem Optimalitätskriterium (ML, MP) arbeiten, bearbeiten das Problem einen bestimmten Baum unter dem gegebenen Kriterium zu evaluieren, indem sie versuchen den optimalen Baum, bzw. die optimalen Bäume, zu finden.

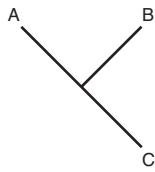
#### 2.2.3.1 Exakte Suche

Die konzeptionell einfachste Herangehensweise den optimalen Baum zu finden, ist die, unter allen möglichen (dichotomen) Bäumen den Optimalen zu suchen. Diese Methode ist allerdings nur für eine begrenzte Anzahl Taxa möglich. Es soll hier kurz das Prinzip erläutert werden. Für drei Taxa (A, B und C) gibt es nur eine Möglichkeit einen dichotomen Baum zu konstruieren

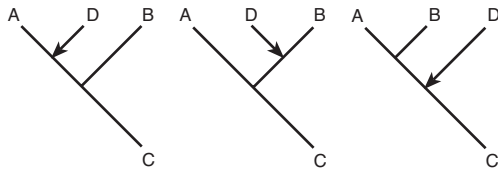
Addiert man ein viertes Taxon (D), hat man die Möglichkeit, dieses an drei Kanten anzubringen; in die Nähe von A, B oder C. Daraus resultieren drei verschiedene Bäume, mit jeweils fünf Kanten.

Die Addition eines fünften Taxons (E) resultiert in 15 mögliche Bäume (3 Bäume x 5 Kanten).

3 Taxa:



4 Taxa:



5 Taxa:

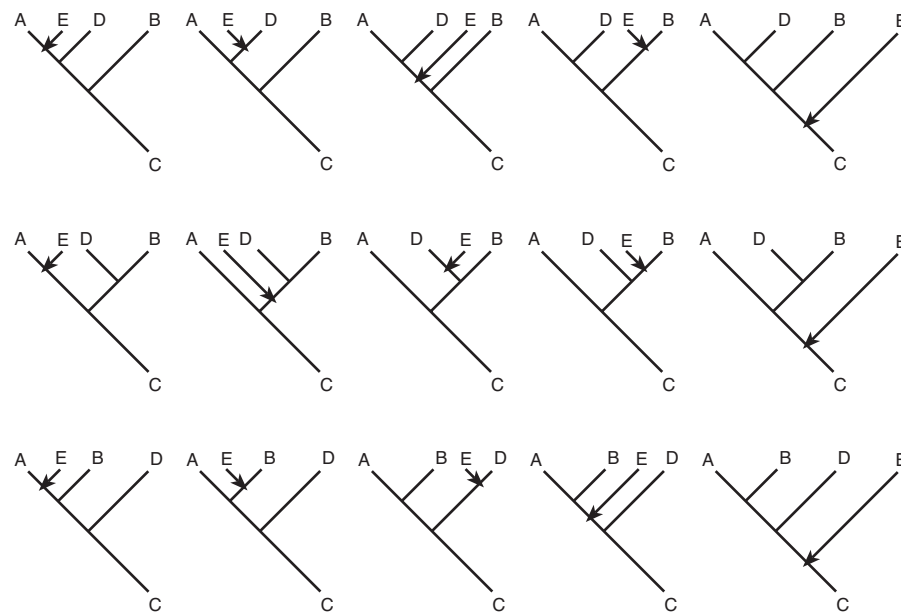


Abbildung 2-6 Graphische Darstellung wie alle ungewurzelten, dichotome Bäume konstruiert werden, aus denen der optimale gefunden werden muß. Nähere Erläuterungen im Text.

Diese 15 Dendrogramme sind jeweils wieder Ausgang für die Addition des nächsten Taxons ( $15 \times 7 = 105$ ). In dieser Weise wird jedes weitere Taxon addiert. So erreicht diese Methode schnell ihre Grenze, die von einem herkömmlichen Computer bewältigt werden kann. Für die Anzahl terminaler Taxa  $n$  ( $n \geq 3$ ) gibt es in einem dichotomen, ungewurzelten Baum  $2n-3$  Kanten. Die Zahl möglicher dichotomer, ungewurzelter Topologien  $N$  für  $n$  Taxa wird bestimmt durch  $N(n) = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n - 5)$ , oder um es mathematisch zu formulieren:

$$N(n) = \prod_{n=3}^n (2n - 5); (n \geq 3) \quad (10)$$

Der schnelle Anstieg möglicher Bäume sei nachfolgend verdeutlicht:

7 Taxa:	945 mögliche Bäume
10 Taxa:	ca. $2 \cdot 10^6$ mögliche Bäume
20 Taxa:	ca. $2 \cdot 10^{20}$ mögliche Bäume

Mehr als 11 Taxa lassen sich auf einem kommerziellen Computer nicht mehr in dieser Art bearbeiten.

Eine Möglichkeit, die Anzahl der Topologien zu beschränken, aber trotzdem eine exakte Suche durchzuführen ist die sogenannte "branch-and-bound" Methode. Man geht von einem zufällig generierten "Suchbaum" aus (oder einem schnellen NJ-Baum) und bestimmt dessen Kantenlänge  $L$ . Dann weiß man, es gibt einen Baum der Länge  $L$ , der als Obergrenze dient. D. h. der optimale Baum hat die Länge  $L$  oder ist gar kürzer (unter dem Optimalitätskriterium Parsimonie zum Beispiel). Man versucht nun die Länge zu optimieren bzw. zu minimieren indem man entlang des Suchbaumes "wandert" und alternative Verzweigungen wählt. Praktisch wird ein neuer Baum durch Addition der Taxa (wie oben beschrieben) konstruiert. Erreicht man auf den Weg zu den terminalen Taxa den Wert  $L$ , so wird die weitere Suche entlang dieses Pfades abgebrochen und man begibt sich wieder zurück zu einem früheren internen Knoten und versucht einen anderen Pfad. Erreicht man die terminalen Taxa und der Wert ist kleiner als  $L$ , setzt man eine neue obere Grenze fest und der Baum wird als ein optimaler gespeichert. Dadurch, daß viele Bäume ausgeschlossen werden können (die, die größer als  $L$  sind) ist der Rechenaufwand etwas geringer als bei der exakten Suche, wo alle möglichen Bäume konstruiert werden. Es ergibt sich aber auch daraus, daß der Rechenaufwand kaum minimiert wird, wenn die Entscheidung erst in der Nähe der terminalen Taxa fällt.

### 2.2.3.2 Heuristische Suche

Für entsprechend große Datensätze ( $\geq 12$  Taxa in PAUP\*) muß man auf heuristische Suchverfahren zurückgreifen. Dies sind Näherungsverfahren, die die Berechnungen beschleunigen, aber keine Garantie geben, den optimalen Baum zu finden. Das Prinzip läßt sich mit einer Metapher beschreiben: Ein kurzsichtiger Pilot stürzt in einem hügeligen Gebiet ab, kann sich mit einem Fallschirm retten, verliert aber seine Brille und ist nun blind wie ein Maulwurf. Er vermutet, daß auf dem höchsten Berg (= optimaler Baum) eine Funkstation ist, von der ein SOS gesendet werden kann. Ohne Weitsicht kann er nur losgehen und den Weg wählen der ihn mit jedem Schritt ein Stück aufwärts bringt (= Addition eins Taxons). In dem Moment wo ihn ein Schritt - egal in welche Richtung weder aufwärts noch abwärts bringt, hat er ein Maximum (= Optimum) erreicht. Der Pilot findet aber keine Funkstation. Zwei Erklärungen gibt es dafür: a) er ist auf dem höchsten Gipfel, aber seine Annahme, daß es dort eine Station gibt war falsch, oder b) er hat den höchsten Gipfel nicht erreicht sondern nur ein lokales Maximum. Blind wie ein Maulwurf kann er es nicht beurteilen.

Aufgrund dieser Metapher spricht man auch von der "hill-climbing" Methoden. Man kann prinzipiell zwischen zwei Herangehensweisen wählen, um den sparsamsten Baum zu finden. Zum einen die schrittweise Addition von Taxa, bei der ausgehend von einer Topologie mit drei Taxa schrittweise ein Taxon addiert wird. Dabei wird jeweils die Topologie beibe-

halten, die unter einem Optimalitätskriterium die sparsamste ist. Am Ende erhält man genau einen Baum, wobei man nicht sagen kann, ob nur ein lokales oder ein globales Optimum erreicht ist. Das zweite Verfahren (Star Decomposition Methode) geht von einem Sterndiagramm aus (alle Taxa sind in einem Knoten verbunden). Ein zufällig ausgewähltes Taxon wird nun mit allen andern zu einer Gruppe verbunden. Die Gruppe aus zwei terminalen Taxa, die unter einem Optimalitätskriterium den besten Wert erhält wird beibehalten.

#### *Verlagerung von Ästen*

Bei der „hill-climbing“ Methode hängt das Resultat des optimalen Baumes ganz entscheidend von der Reihenfolge der addierten Taxa ab. Um bei der Metapher zu bleiben: je nachdem wo im Gelände der Pilot landet und in welche Richtung er losgeht kann er zu einem ganz anderen Hügel (lokales Optimum) gelangen. Deshalb werden die Taxa in zufälliger Reihenfolge addiert. Die so erhaltene Topologie eines Baumes kann dann durch Modifikation (Verlagerung von Kanten; engl.: branch swapping) zu verbessern versucht werden. Handelt es sich bei dem Anfangsbaum um ein lokales Optimum, ist die Chance das globale Optimum zu finden mit dieser Methode deutlich verbessert. Drei Algorithmen finden Verwendung, von denen der TBR („tree bisection reconnection“) der aufwendigst ist und die meisten möglichen Bäume abdeckt. Dazu wird eine Topologie an einer inneren Kante in zwei Bäume geteilt (tree bisection), von denen dann zufällig gewählte innere Kanten wieder verknüpft werden (reconnection). Für jede neue Topologie muß die Länge berechnet werden und mit der Ausgangstopologie verglichen werden. Ist eine Topologie unter dem gegebenen Optimalitätskriterium besser als das Ausgangsmodell, wird dieses nun beibehalten und dient nun als „Referenzbaum“. Andere Algorithmen sind das SPR (subtree pruning and regrafting) und NNI (nearest neighbor interchange). Bei der letzten Variante werden jeweils zwei von vier mit einer inneren Kante verknüpften Äste ausgetauscht. Der SPR Algorithmus trennt einen Ast von der Ausgangstopologie ab und verbindet ihn an einer zufälligen inneren Kante.

Vorteil der Verlagerung von Ästen ist sicher, daß die Wahrscheinlichkeit, den optimalen Baum zu finden, recht hoch ist. Aber auch dieses Verfahren ist sehr langsam und kann zum Teil vom Rechner nicht bewältigt werden.

#### 2.2.4 Dendrogramme versus Split-Zerlegung

Eine andere Form der Darstellung phylogenetischer Verwandtschaften ist die Splitzerlegung. Dieses Verfahren eignet sich zur Berücksichtigung und Darstellung inkompatibler Merkmalsausprägungen, sogenannter Homoplasien. Tatsächlich veranschaulicht die aus der Splitzerlegung resultierende Netzwerkdarstellung Konfliktpunkte zwischen den OTUs. Dazu wird eine Datenmatrix Merkmal für Merkmal überprüft und jeweils eine Spaltung der OTUs vorgenommen. Zur Veranschaulichung soll ein Beispiel gewählt werden, bei dem jedes Merkmal nur zwei Zustände aufweist (Abbildung 2-7).

		Merkmale								
		1	2	3	4	5	6	7	8	9
Taxon	M	T	C	C	T	T	A	A	A	A
	C	T	T	C	T	A	T	A	A	A
	G	T	T	A	C	A	A	T	A	A
	O	C	C	A	C	A	A	A	T	A
	B	C	C	A	C	A	A	A	A	T

Abbildung 2-7 Beispiel einer Merkmalmatrix.

Diese ist Ausgang für die Konstruktion eines Netzwerkdiagramms (Abbildung 2-8). Zur Vereinfachung hat jedes Merkmal nur zwei Zustände.

Jedes Merkmal weist zwei Zustände auf, korrespondiert also zu einem Split. Das Merkmal 1 in der Beispielsequenz fordert den Split  $\{M, C, G\} \{O, B\}$  (siehe Abbildung 2-8a). Merkmal 2 fordert den Split  $\{C, G\} \{M, O, B\}$ . Die Kombination ergibt  $\{\{M\} \{C, G\}\} \{O, B\}$  (Abbildung 2-8b). Das folgende Merkmal stellt einen Konfliktpunkt dar. Die Trennung  $\{M, C\}$  und  $\{G, O, B\}$  läßt sich nicht mit dem vorherigen Split kombinieren. Bei der Konstruktion eines Baumes müßte man sich entscheiden, ob das Taxon G näher zu C oder zur Gruppe  $\{O, B\}$  steht. Hier im Netzdiagramm ist der Konflikt darstellbar: Die Splits 2 und 3 lassen sich durch Einfügen eines Parallelogramms darstellen (Abbildung 2-8d). G hat nun über eine Kante sowohl Verbindung zum C als auch (über einen Knoten; leerer Kreis in der Abbildung) zur Gruppe  $\{O, B\}$ . Mit den weiteren Merkmalen wird in gleicher Weise verfahren. Letztendlich kommt man zu einem Netzdiagramm wie in Abbildung 2-8h.

Um nochmals herauszustellen, was in dem Netzdiagramm dargestellt ist läßt man das Parallelogramm kollabieren. In der vertikalen Richtung entsteht der Split  $\{\{M, C\} \{G\}\} \{O, B\}$ , in der horizontalen Richtung der Split  $\{\{M\} \{C, G\}\} \{O, B\}$ . In dem Netzwerkdiagramm lassen sich also mehrere alternative Dendrogramme wiedergeben.

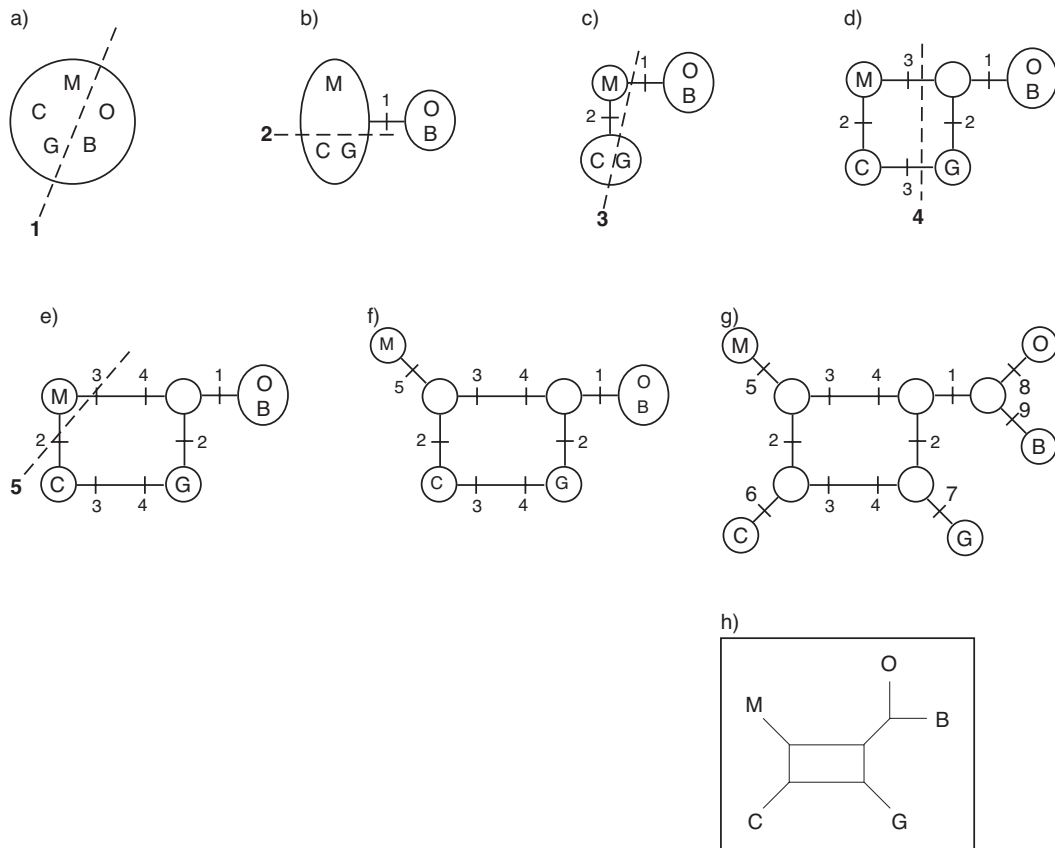


Abbildung 2-8 Beispiel für eine Splitzerlegung. Die unterbrochene Linie markiert die Splitzerlegung, die durch das Merkmal (fette Zahl) gefordert ist und eine im folgenden eine Kante zwischen den Taxa einfügt. Auf den Kanten sind die Merkmale (aus Abbildung 2-7) notiert, die den Split bewirkten. In h) ist das Netzdiagramm in seiner endgültigen Form dargestellt; die Kanten haben proportionale Längen. Einzelheiten sind im Text erläutert. Verändert nach (Page & Holmes, 1998a).

Ergänzung: Es sei hier nochmals auf das Beispiel aus Kapitel 2.1.4 Homologie und Homoplasie eingegangen. Die Merkmale {CT} {CA} {GA} {GA} lassen sich in eine hypothetische Sequenzmatrix transferieren und dann, wie oben beschrieben, in ein Netzdiagramm darstellen (Abbildung 2-9).

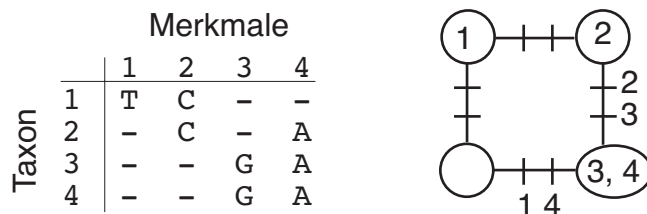


Abbildung 2-9 Homoplasie als Netzdiagramm dargestellt. Diese Abbildung greift das Beispiel aus Kapitel 2.1.4 Homologie und Homoplasie auf und stellt die beiden möglichen Verwandtschaftsbeziehungen als Netzdiagramm dar. An den Kanten sind die induzierenden Merkmale markiert.

## 2.2.5 Distanzverfahren

### 2.2.5.1 Unkorrigierte Distanzen

Distanz Methoden beruhen auf der Idee, daß wenn wir die evolutive Distanz zwischen allen terminalen Taxa eines Datensatzes kennen, dann läßt sich leicht die evolutive Geschichte dieser Sequenzen rekonstruieren.

Die Distanzmessung assoziiert einen numerischen Wert mit einem Paar Sequenzen, wobei niedrige Werte eine hohe Ähnlichkeit indizieren. Gewöhnlich folgt die Distanzmessung dem mathematischen, metrischen Axiom, d. h. die Werte werden nicht negativ.

Die einfachste Form der Distanzmessung ist die sogenannte Hamming-Distanz ( $d_H$ ). Sie gibt für zwei Sequenzen mit gleicher Länge die Anzahl der unterschiedlichen Merkmalsausprägungen der analysierten Sequenzen an. So haben die zwei Sequenzen

$$s_1 = AAT$$

$$s_2 = TAA$$

eine Hamming-Distanz von  $d_H = 2$  (unterschiedliche Positionen 1 und 3) und entsprechend haben die Sequenzen

$$t_1 = AGCACACA$$

$$t_2 = ACACACTA$$

einen Hamming-Distanzwert von  $d_H = 6$ .

Wird die Hamming-Distanz in Bezug zur Länge des Alignments  $N$  gesetzt, erhält man die  $p$ -Distanz, also:

$$p = \frac{d_H}{N} \quad (11)$$

Eine ausführliche Beschreibung der zugrunde liegenden Mathematik findet man bei (Dress, 1995).

### 2.2.5.2 Minimum Evolution

Bei einem gegebenen wurzellosen metrischen Baum für  $n$  Sequenzen gibt es  $(2n-3)$  Kanten, jeweils mit einer Länge  $e_i$ . Die Summe aller Kantenlängen  $L$  des Baumes ist

$$L = \sum_{i=1}^{2n-3} e_i \quad (12)$$

Der Minimum Evolution Baum (ME) ist der Baum mit dem niedrigsten  $L$ -Wert.

### 2.2.5.3 Neighbor-joining

Bei der Neighbor-joining (NJ) Methode (Saitou & Nei, 1987) handelt es sich um die oben beschriebene Gruppierungsmethode und wird gerne benutzt, da sie schnell ist und fast immer zu einem einzigen Baum führt. Diese Methode hat die oben besprochenen Nachteile,

eignet sich aber als gute heuristische Methode zur Evaluierung eines ME Baums, indem man erst einen NJ Baum konstruiert und dann versucht durch lokale Neuordnungen einen kürzeren Baum zu bekommen.

### 2.2.6 Diskrete Verfahren

Diese Methoden, im Gegensatz zu den Distanzverfahren, operieren direkt mit den Sequenzdaten. Folglich können hier Informationen genutzt werden, die bei der Konvertierung zu Distanzdaten verloren gehen. Man unterscheidet die zwei Hauptmethoden Maximum Parsimonie (MP) und Maximum Likelihood (ML). Maximum Parsimonie wählt den Baum (oder Bäume), der die wenigsten evolutiven Schritte benötigt. Maximum Likelihood wählt den Baum (oder Bäume), der am wahrscheinlichsten die beobachteten Daten erklärt.

#### 2.2.6.1 Maximum Parsimonie

Bei dieser Methode wird jedes informative, individuelle Merkmal verglichen. Für jede Position ist es das Ziel die Evolution dieser Position in einem Baum zu rekonstruieren unter Erzwingung möglichst weniger evolutiver Veränderungen. Die Gesamtzahl evolutiver Veränderungen eines Baumes ist die Summe der Veränderungen an jeder Position. Also hat die Sequenz  $k$  Nukleotidpositionen, jeweils mit der Länge  $l$ , dann ist die Baumlänge

$$L = \sum_{i=1}^k l \quad (13)$$

Nukleotidpositionen die in allen Sequenzen gleich sind oder nur bei einer Sequenz unterschiedlich sind, sind phylogenetisch nicht informativ, da sie nicht zwischen verschiedenen alternativen Bäumen diskriminieren. Man unterteilt die variablen Positionen daher in Parsimonie-informativ und in Parsimonie nicht-informativ.

Bei der Parsimonie Methode können noch verschiedene Parameter berücksichtigt werden. Man kann das Substitutionsmodell berücksichtigen, d. h. die "Kosten" für eine Substitution. So kann man die Kosten für Transversionen höher bewerten als für Transitionen. (Siehe 2.1.6)

#### 2.2.6.2 Maximum Likelihood

Likelihood ist die Wahrscheinlichkeit die beobachteten Daten mit einem bestimmten Modell zu erklären. Unterschiedliche Modelle machen die beobachteten Daten mehr oder weniger wahrscheinlich. Dabei ist es wichtig zu unterscheiden, zwischen der Wahrscheinlichkeit die beobachteten Daten zu bekommen und der Wahrscheinlichkeit, daß das zugrundeliegende Modell das Richtige ist. Mit einem Modell, welches die Wahrscheinlichkeit verschiedener Ereignisse zu beobachten beschreibt, kann man die Wahrscheinlichkeit  $L$  die beobachteten Daten zu erhalten berechnen. Es gilt:

$$L = \Pr(D | H) \quad (14)$$



Dabei ist  $L = \Pr(D | H)$  die Wahrscheinlichkeit die Daten  $D$  zu erhalten unter der Hypothese  $H$ . Da der Wert  $L$  meist sehr klein ist, wird er als natürlicher Logarithmus dargestellt. Diese Berechnung erlaubt es verschiedene Modelle auf ihre Wahrscheinlichkeit zu testen. Dies ist dem Programm Modeltest verwirklicht.

## 2.2.7 Abschätzung des Stichprobenfehlers

Es ist nicht auszuschließen, daß das Modell zwar richtig gewählt ist, trotzdem eine schwache, also unsichere phylogenetische Ableitung entsteht. Dies kann an den Daten selbst liegen. Ein Beispiel bietet die Arbeit von Årnason und Johnsson (Årnason & Johnsson, 1992), die 15 verschiedene mitochondriale Gene in sechs verschiedenen Säugetieren untersucht haben. Obwohl die mitochondrialen Gene eines Individuums jeweils die selbe phylogenetische Geschichte besitzen, da sie maternal ohne Rekombination vererbt wurden, ließen sich fünf verschiedene Phylogeneseen konstruieren.

### 2.2.7.1 Bootstrap-Test

Es gibt zwei Methoden den Stichprobenfehler vom Populationsmittel abzuschätzen. Die Eine ist die Untersuchung weiterer Stichproben aus dieser Population. Die Streuung aus diesen Proben läßt eine Abschätzung des Stichprobenfehlers zu. Diese Form ist sehr kostspielig und oft nicht durchführbar. In Computerprogrammen werden diese wiederholten Stichproben nachgeahmt in Form von Stichproben aus der originalen Stichprobe, sogenannte Pseudoreplikationen. Die Pseudoreplikate werden aus dem Sequenzdatensatz durch zufällige Entnahme und Zurücklegen einer Position generiert. Anders ausgedrückt, eine zufällige Position im Sequenzalignment wird kopiert und bildet die Position Eins des ersten Pseudoreplikats. Eine weitere zufällige Position wird kopiert (es kann theoretisch die Selbe sein) und bildet Position Zwei des Pseudoreplikats. Das wird so lange wiederholt bis ein Pseudoreplikat mit der gleichen Länge wie das original Alignment entstanden ist. Durch dieses Verfahren können manche Positionen mehrmals im Pseudoreplikat vorhanden sein, andere gar nicht. Folglich enthält das Pseudoreplikat nur Positionen wie im Original Datensatz aber mit einer veränderten Frequenz. Auf Grundlage dieses Pseudoreplikats wird ein phylogenetischer Baum konstruiert. Dieser Prozeß des Generierens eines Pseudoreplikats wird 100-1000 mal wiederholt und man erhält so 100-1000 sogenannte Bootstrap Bäume. Dieser Datensatz gibt Auskunft über den Stichprobenfehler unserer Daten und wird oft als Mehrheitsregel-Konsensus Baum (50 % majority rule consensus tree) ausgegeben. Dieser zeigt an mit welcher Frequenzen eine bestimmte Verzweigungen in allen Bootstrap Bäumen vorkommt. Es ist wichtig zu betonen, daß der Bootstrap Wert nur etwas über die Genauigkeit, nicht über die Korrektheit aussagt. Eine Verzweigung kann einen hohen Bootstrap Wert haben aber völlig falsch sein.

### 2.2.7.2 Jackknifing

Der Bootstrap-Test ist in Verbindung mit dem Parsimoniekriterium und Verlagerung von Ästen bei einem umfangreichen Datensatz kaum durchzuführen, da die kommerziellen

Rechnerleistung dafür nicht ausreichen. Eine gute Alternative ist der sogenannte Eliminierungs-Test („jackknifing“). Dabei wird aus dem original Datensatz ein Pseudoreplikat konstruiert, welches aber nur aus einem gewissen Prozentsatz der Merkmale besteht. Wie beim „bootstrapping“ werden viele Wiederholungen (z. B. 1000) durchgeführt. Die Wahrscheinlichkeit für die Eliminierung eines Merkmals darf nicht zu hoch sein, da es sonst keine sinnvolle Beziehung zwischen stützenden Merkmalen und der Frequenz von wiedergefundenen Artgruppen gibt. Farris und Kollegen schlagen einen Wert von  $e^{-1}$  (also etwa 37 %) vor (Farris *et al.*, 1996).

### 2.3 Mitochondriale DNA

Die mitochondriale DNA (mtDNA) der Vertebraten ist ein zirkuläres Molekül, das gewöhnlich kleiner als 20 kb (Kilobasenpaare) ist. Es besteht aus einer nicht-translatierten Kontrollregion, auch D-loop Region genannt, in der sich der Replikationsstart für dieses Molekül befinden. Es gibt 13 für Proteine codierende Gene, zwei ribosomale RNA Gene (12S- und 16S-rRNA), und 22 verschiedene Transfer RNA (tRNA) Gene. Die beiden Stränge der zirkulären mtDNA haben eine asymmetrische Verteilung von Gs (Guaninbasen) und Cs (Cytosinbasen), wodurch ein schwerer (H, heavy) und leichter (L, light) Strang unterschieden werden kann. Jeder Strang hat seinen eigenen Promotor (HSP, heavy strand promoter und LSP, light strand promoter), die in der D-loop Region (D, displacement) liegen. Obgleich der HSP vorzugsweise den H-Strang und der LSP vorzugsweise den L-Strang transkribiert, sind doch beide bi-direktional (Chang *et al.*, 1986).

Die D-loop Region ist ein dreisträngiger Bereich, generiert durch die Synthese eines kurzen Stücks H-Strang DNA, der 7S DNA. Die H-Strang DNA Replikation ist initiiert innerhalb der D-loop in vier Haupt- und drei Nebenpositionen. Drei von diesen korrespondieren zu den L-Strang Transkriptionsstop Sequenzen der "conserved sequence blocks" CSB1 bis 3. Der vorherrschendste 7S DNA Start ist im CSB2 Bereich. Die Primer für die 7S DNA Synthese werden vermutlich durch Spaltung des L-Strangtranskripts durch RNAse MRP (mitochondriale RNA Prozessierung) generiert. Das beinhaltet eine nukleär kodierte RNA, die scheinbar die Spaltung der RNA an der CSB2 Position lenkt (Chang & Clayton, 1987a; Chang & Clayton, 1987b; Chang & Clayton, 1989).

Alle 7S DNA Moleküle enden an Nukleotiden kurz hinter den "extended termination associated sequences" (ETAS). Die H-Strang Replikation beginnt mit der 7S DNA und geht entlang des L-Strangs, den parental einelsträngigen H-Strang verdrängend ("displacement"); dadurch entsteht der D-loop. Nach  $\frac{2}{3}$  des Genoms wird der Replikationsstart des L-Strangs frei. Die L-Strangreplikation wird dann durch eine spezifische Primase initiiert (Wong & Clayton, 1986) und verläuft zurück entlang der verdrängten H-Strangmatrize.

Die genomische Organisation der mtDNA ist unter den Vertebraten recht gut konserviert, obgleich einige Reorganisationen beschrieben wurden (Pääbo, 1991). Der genetische Code der mtDNA ist, verglichen mit dem nukleären universellen Code einfacher, da weniger

Anti-codons (nur 22) genutzt werden können. Manche Codonsequenzen haben eine andere Funktion als in der nukleären DNA.

### 2.3.1 Die Endosymbiontentheorie

Allgemein akzeptiert ist die Hypothese über die Abstammung der Mitochondrien von freilebenden Eubakterien, welche zum Zeitpunkt der Symbiose ihr eigenes Genom besaßen. Die Entstehung der eukaryotischen Zellen wurde 1971 in einem 2-Schritte Modell postuliert (Margulis, 1971). Danach kam zuerst der Nukleus von einem Archebakterium und dann entstand eine Symbiose mit einem eubakteriellem Vorläufer der heutigen Mitochondrien.

Neuer ist die sogenannte Hydrogen Hypothese (Martin & Muller, 1998; Muller & Martin, 1999), die eine simultane Entstehung von eukaryotischem Zellkern und Mitochondrium postuliert. Die Hypothese stellt das Szenario als eine Fusion von hydrogenbedürftigen methanogenen Archebakterien (Wirt) mit hydrogenproduzierenden  $\alpha$ -Proteobakterien (Symbiont) dar.

### 2.3.2 Mitochondriale DNA in der Evolutionsforschung

Was die mitochondriale DNA für evolutive Untersuchungen interessant macht ist die Tatsache, daß einige Sequenzabschnitte schneller evolvieren als nukleäre Sequenzen. So liegt die Evolutionsrate für die mtDNA etwa 10mal höher als für single-copy DNA (Brown *et al.*, 1979b). Die nicht codierende D-loop Region hat eine noch höhere Evolutionsrate, die ein 100faches der nukleären entsprechen kann (Greenberg *et al.*, 1983). Es gibt viele mögliche Erklärungen für die schnelle Evolution der mtDNA, die wichtigste ist sicher die der ineffizienten Reparaturmechanismen der mtDNA.

Typisch für Vertebraten mtDNA ist auch das hohe Transitionen/Transversionen Verhältnis, wie es für viele Spezies beschrieben wurde (Excoffier & Yang, 1999; Giuffra *et al.*, 1994; Lanave *et al.*, 1986; Lyrholm *et al.*, 1996; Tamura & Nei, 1993).

Ein weiterer wichtiger Aspekt für phylogenetische Untersuchungen ist die prinzipiell ausschließlich maternale Vererbung. Die mtDNA wird immer nur über die Eizelle an die Nachkommen weitergegeben. Heteroplasmie, also das Vorkommen von zwei oder mehreren mtDNA Typen in einem Individuum ist in einigen Fällen beobachtet worden (Arnason & Rand, 1992), wird aber im allgemeinen mit einer Mutation innerhalb der maternalen Linie erklärt, nicht mit einer Übertragung paternaler mtDNA (Awise, 1991).

Trotzdem kann eine paternale Vererbung nicht ausgeschlossen werden. Untersuchungen an Mauslinien zeigten die Übertragung von paternaler mtDNA auf die Nachkommen mit einer relativen Frequenz von  $10^{-4}$  mittels der PCR Technik (Gyllensten *et al.*, 1991).

Viele der erwähnten Charakteristika machen die mtDNA zu einem interessanten Untersuchungsobjekt in der molekularen Evolution. Wegen ihrer hohen Kopienzahl pro Zelle, kann man sie leicht auch aus sehr kleinen Proben gewinnen. Trotzdem einige Bereich der mtDNA einer schnellen Veränderung unterliegen, gibt es doch viele Gene, die in verschiedenen

Spezies stark konserviert sind, so daß man sie mit "universellen" Oligonukleotiden amplifizieren kann (Kocher *et al.*, 1989). Verglichen mit der nukleären DNA besitzt die mtDNA Vorteile zur Rekonstruktion phylogenetischer Stammbäume. Es ist schwierig einen individuellen Haplotypen von der nukleären DNA eines diploiden Individuums zu gewinnen, und die analysierten Loci könnten Rekombinationen eines DNA Segments mit unterschiedlicher Geschichte sein. Das mitochondriale Genom hingegen besteht gewöhnlich aus nur einer Linie je Individuum, die ohne Rekombination weitergegeben wird. Vorteilhaft ist ferner die unterschiedlich schnelle Evolutionsrate verschiedener Abschnitte der mtDNA.

Es ist also möglich (und notwendig) sich den geeignetsten Abschnitt für eine spezifische Untersuchung zu wählen. Einige mitochondriale Gene entwickeln sich zu langsam, um eine geeignete Sequenzdivergenz zu liefern, während andere so schnell Mutationen akkumulieren, daß selbst intraspezifische Unterschiede nachgewiesen werden können. Analysen phylogenetisch entfernter Taxa werden deshalb den Gebrauch sich langsam entwickelnder mitochondrialer Regionen fordern. Im Gegensatz dazu wird die Untersuchung naher verwandter Taxa eher auf sich schnell entwickelnde Abschnitte zielen. Für die Untersuchung molekularer Evolution sind vielleicht tatsächlich die sich schnell entwickelnden Abschnitte der mtDNA von größter Bedeutung. Da extensive Sequenzunterschiede sogar innerhalb einer Art Aufschluß geben über Verwandtschaftsverhältnisse, können phylogenetische Prinzipien in die Populationsgenetik einbezogen werden. Damit liefert die mtDNA eine molekulare "Brücke zwischen Populationsgenetik und Systematik", die zur "Erweiterung des phylogenetischen Denkens auf der Ebene der Mikroevolution" veranlaßt (Avice, 1986).

## 2.4 Phylognese der Büffel

Eine der ersten molekulargenetischen Untersuchungen zur Phylognese der Wasserbüffel wurde 1994 von Amano und Kollegen veröffentlicht. Sie versuchten mittels Restriktions Fragmentlängen Polymorphismus (RFLP) die genetische Verwandtschaft zwischen Fluß- und Sumpfbüffel zu klären (Amano *et al.*, 1994). Es wurden 73 Sumpfbüffel und drei Flußbüffel untersucht und die Diversität in der ribosomalen DNA (rDNA) und in der mitochondrialen DNA (mtDNA) bestimmt. Dabei wurden drei mitochondriale Haplotypen (zwei Sumpftypen, ein Flußtyp) und keine Unterschiede in der rDNA gefunden. Die Divergenz von Rind und Wasserbüffel wurde auf 4-6 Mio. Jahre geschätzt. Eine Abschätzung der Divergenz zwischen Fluß- und Sumpftyp war anhand dieser Daten nicht möglich. Ein Jahr später untersuchten Tanaka und Kollegen die Nukleotiddiversität der mitochondrialen DNA (mtDNA) zwischen Fluß- und Sumpftyp des domestizierten Wasserbüffels (Tanaka *et al.*, 1995). Zur Feststellung dieser Unterschiede nutzten sie ebenfalls die RFLP Technik. Die mtDNA von 13 Fluß- und 10 Sumpfbüffeln wurde mit 15 verschiedenen Restriktionsendonukleasen verdaut, von denen 9 einen Polymorphismus aufwiesen. Hier wurden fünf mtDNA Typen identifiziert (zwei im Fluß- und drei im Sumpfbüffel). Aufgrund der errechneten Nukleotiddiversität diskutiert die Gruppe eine Divergenz beider Typen vor mindestens 700.000 Jahren.

Ein Jahr später veröffentlicht die gleiche Gruppe eine Untersuchung zur phylogenetischen Verwandtschaft aller Arten des Genus *Bubalus* (Tanaka *et al.*, 1996). Für je einen Vertreter jeder Art wurde die komplette Sequenz des mitochondrialen Gens Cytochrom b und die resultierenden Nukleotidunterschiede bestimmt. Dabei fanden sie eine unerwartet hohe Divergenz zwischen Fluß- und Sumpftyp mit einer kalkulierten Trennung von etwa 1,7 Mio. Jahren. Die Divergenz *Bubalus* – *Bos* wurde aufgrund fossiler Daten (Savage & Russell, 1983) auf ca. 10 Mio. Jahre festgesetzt.

Ein Jahr später untersuchte die Gruppe um Barker die genetische Diversität innerhalb der Wasserbüffel. Zum einen wurden genetische Variationen in 53 Protein kodierenden Loci analysiert. Es wurden Allelfrequenzen und Heterogenitätsgrade bestimmt, die eine genetische Differenzierung von Fluß- und Sumpfbüffeln zeigen, wie man es von anderen Rassen anderer Spezies kannte (Barker *et al.*, 1997b). Auch eine Mikrosatelliten Untersuchung differenzierte beide Typen deutlich, mit einer Trennung vor mindestens 10.000 – 15.000 Jahren (Barker *et al.*, 1997a). Im Gegensatz dazu führte die Untersuchung der Sequenzunterschiede im mitochondrialen Cytochrom b Gen zu einer geschätzten Trennung von beiden Büffeltypen vor etwa 1 Mio. Jahren (Kikkawa *et al.*, 1997).

Wieder zu anderen Ergebnissen kamen Lau und Kollegen, die die genetische Diversität der Büffel abschätzten, anhand von Sequenzunterschieden in der D-loop Region und im Cytochrom b Gen (Lau *et al.*, 1998). Mit diesen Daten wurde die Trennung von Fluß- und Sumpftyp auf vor 28.000 – 87.000 Jahren datiert.

Eine neuere Studie aus dem letzten Jahr untersuchte, ob die generell akzeptierte Phylogenie der Bovini mit den Polymorphismen 20 boviner Mikrosatelliten korrespondiert (Ritz *et al.*, 2000). Mit den ermittelten Daten zur genetischen Distanz berechneten Ritz und Kollegen eine Trennung von Fluß- und Sumpfbüffel vor 0,74 – 1,96 Mio. Jahren. Die Trennung von *Bos* und *Bubalus* soll im Zeitraum 1,85 – 4,93 Mio. Jahre vor heute stattgefunden haben.

---

## 3 Material, Methoden und Software

### 3.1 Material

#### 3.1.1 Probensammlung

Insgesamt wurden 63 Blutproben von Murrah-, Jafarabadi- und Carabao-Rassen von verschiedenen Farmen in Brasilien genommen. Dazu wurden 5 ml Blut durch punktieren der Jugularvene in 12 ml Röhrchen aufgefangen, in die 300  $\mu$ l EDTA vorgelegt waren. Die Proben stammen von fünf verschiedenen Populationen in Brasilien

Von Prof. Iannuzzi (Italien) wurden uns freundlicherweise 17 Blutproben der Rasse Mediterraneo zur Verfügung gestellt, sowie fertig präparierte DNA. Diese stammen von drei verschiedenen Populationen, respektive Farmen.

Ein Zusammenstellung über Herkunft, Rasse und Art der verschiedenen Proben sind in Tabelle 1 aufgeführt.

#### 3.1.2 Enzyme, Nukleotide, Chemikalien und Kits

##### 3.1.2.1 Enzyme

New England Biolabs GmbH (Schwalbach): Restriktionsenzyme

Qiagen GmbH (Hilden): *Taq*-Polymerase

##### 3.1.2.2 Nukleotide

Interactiva Biotechnologie GmbH (Ulm): unmarkierte Oligonukleotide

MWG Biotech AG (Ebersberg): IRD-800 fluoreszenzmarkierte Oligonukleotide, unmarkierte Oligonukleotide

Roth GmbH (Karlsruhe): dNTPs

##### 3.1.2.3 Chemikalien

Die eingesetzten Chemikalien wurden von folgenden Firmen bezogen: Amersham-Pharmacia Biotech Europe GmbH (Freiburg), Biozym Diagnostic GmbH (Hess. Oldendorf), Calbiochem (Bad Soden), ICN Biomedicals GmbH (Eschwege), Life Technologies GmbH

(Heidelberg), Merk AG (Darmstadt), QIAGEN GmbH (Düsseldorf), Roth GmbH (Karlsruhe) und Sigma-Adrich Vertriebs GmbH (Deisenhofen).

### 3.1.2.4 Kits

Amersham-Pharmacia Biotech Europe GmbH (Freiburg): Thermo Sequenase fluorescent labelled primer cycle sequencing kit

QIAGEN GmbH (Hilden): QiaexII Kit

Tabelle 1 Herkunft der Blut- und DNA-Proben.

Die Herkunft der Blut- und DNA-Proben von verschiedenen Büffelrassen. Die Individuen sind nummeriert.

Herkunft	Rasse	Art der Probe	Tiernummer
Fazenda Itaquí - Castanhal, Pará (Brasilien)	Jafarabadi	Frischblut	01, 02, 03, 04, 19, 23, 25, 27, 30, 32, 40, 44, 47, 49, 50, 51, 53, 55, 60, 66 105, 110, 117
	Murrah	Frischblut	
EMBRAPA/CPATU - Belém, Pará (Brasilien)	Murrah	Frischblut	143, 181, 182, 183
Fazenda São Luís - Marajó, Pará (Brasilien)	Murrah	Frischblut	196, 267, 304, 310, 314
	Carabao	Frischblut	315, 320, 325, 327, 330, 335, 336, 341, 342, 344, 348, 349, 352, 359, 366, 370, 375, 377, 380
Fazenda Jari - Monte Dairado, Amapá (Brasilien)	Murrah	Frischblut	481, 497
Registro, São Paulo (Brasilien)	Murrah	Frischblut	IZ278, IZ192, IZ358, IZ244, IZ369, IZ143, IZ364, IZ287, IZ368, WB1649
IABBAM (Italien) <sup>a)</sup>	Mediterraneo	Frischblut	B05, C01, C02, C03, C04, C05, C07, C08
		DNA	D01, D02, D03, D04, D05, D06, D07, D09, D10

<sup>a)</sup> Gleiche Buchstaben indizieren die Herkunft von einer Farm.

### 3.1.3 Glaswaren und Verbrauchsartikel

Glaswaren wurden von den Firmen Schütt Labortechnik (Göttingen) und Kranich GmbH & Co. KG (Göttingen) bezogen.

Verbrauchsartikel wurden von folgenden Firmen bezogen: Greiner Labortechnik GmbH (Solling), Integra Biosciences GmbH (Fernwald), Kranich GmbH & Co. KG (Göttingen), Nerbe plus GmbH (Winsen/Luhe), Sarstedt (Nürnbrecht), Schütt Labortechnik (Göttingen) und Schleicher & Schuell (Dassel).

### 3.1.4 Laborgeräte

Bio-Rad Laboratories GmbH (München): Agarose-Gelkammern

Heraeus Instruments GmbH (Hannover): Zentrifugen

Millipore GmbH (Eschborn): Reinwasseranlage

MWG Biotech AG (Ebersberg): LI-COR Sequenzierer 4000L

Sauer Laborgeräte GmbH (Reutlingen): Heizblock, UV-Tische

Schütt Labortechnik (Göttingen): Bioclav, Kühlschränke und Gefrierschränke, Haake Wasserbad DC3, Eppendorf Pipetten, Eppendorf Tischzentrifuge, Eppendorf Thermostat 5320,

IKA Magnetrührer, Savant Speed-Vac SC110A, Vortex Genie

Stratagene GmbH (Heidelberg): Eagle Eye II, RoboCycler Gradient 2000

## 3.2 Methoden

### 3.2.1 DNA-Extraktion aus Vollblut

Zur Isolation von genomischer DNA aus Blut wurde zum einen die klassische phenolische Extraktion angewandt. Die Proben aus Italien wurden mit einer einfachen Aussalzungs- methode behandelt.

#### 3.2.1.1 DNA-Isolierung und phenolische Extraktion

Das Blut (5 ml in 300  $\mu$ l EDTA) wurde auf 12 ml mit 1x SSC-Puffer aufgefüllt, gemixt und zentrifugiert. Das Leukozytenpellet wurde mehrmals in 12 ml 1x SSC-Puffer resuspendiert und zentrifugiert, bis der Überstand klar war. Das Leukozytenpellet wurde mit Na-acetat (0,2 M, pH 7,0) resuspendiert, bevor 0,5 ml 10 % SDS zugegeben wurde. Die Lösung wurde vorsichtig geschwenkt bis sie klar wurde. Die gesamt genomische DNA wurde mit einem gleichen Volumen Phenol/CIA extrahiert. Die wäßrige Phase wurde mit einem gleichen Volumen CIA (24/1) re-extrahiert. Die DNA wurde durch Zugabe eines gleichen Volumens von Isopropanol präzipitiert. Die DNA wurde in ein Eppendorf-Reaktionsgefäß überführt, mit Ethanol gewaschen, luftgetrocknet und in 1 ml TE Puffer (10/1, pH 8,0) gelöst.

1x SSC	150 mM NaCl 15 mM Na-Citrat
Phenol/CIA	Phenol, Chloroform und Isoamylalkohol wurden im Verhältnis 25:24:1 gemischt
CIA (24/1)	Chloroform und Isoamylalkohol wurden im Verhältnis 24:1 gemischt



---

TE Puffer	10 mM	Tris-HCl, pH 7,5
	1 mM	EDTA

### 3.2.1.2 DNA-Extraktion nach Miller *et al.* (1988)

Die Blutproben der italienischen Büffel wurden in 5 ml Vakuumröhrchen geliefert. Die DNA wurde nach einer einfachen Aussalzungs-Methode extrahiert (Miller *et al.* 1988), die hier kurz beschrieben wird.

Zum Aufbrechen der Leukozyten wurden 3 ml Blut mit ddH<sub>2</sub>O auf 10 ml aufgefüllt und leicht geschwenkt. Die Probe wurde 10 min bei T<sub>amb</sub> und 9000 g zentrifugiert. Der Überstand wurde mit einer Wasserstrahlpumpe abgesaugt und das Sediment in 1 % NP40 (Detergenz) resuspendiert. Die Suspension wurde mit 1 % NP40 auf 10 ml aufgefüllt, geschwenkt und wie oben zentrifugiert. Nach Dekantieren des Überstands wurde das Pellet mit 1 ml Proteinase K Puffer resuspendiert, 100 µl Proteinase K und 70 µl 10 % SDS zugegeben und das Ganze bei 56 °C für zwei Stunden inkubiert. Zugabe von 333 µl 6M NaCl und kräftiges Schütteln führte zur Fällung der DNA. Zelltrümmer etc. wurden durch 15minütige Zentrifugation bei 6000 g und 4 °C abzentrifugiert. Der Überstand wurde in 5 ml 100 % Ethanol dekantiert und durch vorsichtiges Schwenken die DNA präzipitiert. Die gut sichtbare DNA wurde mit einer zugeschmolzenen Pasteurpipette herausgefischt, luftgetrocknet und in 500 µl TE Puffer gelöst.

### 3.2.2 Amplifikation der mitochondrialen D-loop Region

Zur Amplifikation der bubalinen D-loop Region wurden Oligonukleotide "designed", die von der bekannten bovinen mtDNA Sequenz (Anderson *et al.*, 1982) abgeleitet wurden. Zwei Oligonukleotide lagen in den die D-loop Region flankierenden konservierten Bereichen des tRNA<sup>Phe</sup> und des Cytb Gens. Bei der Sequenzanalyse der Carabao-Proben zeigte sich ein poly-G Abschnitt, der sich nicht korrekt sequenzieren ließ. Nach Vorlage der bekannten Teilsequenzen dieser Tiere wurde weitere vier Oligonukleotide synthetisiert. Folgendes Schema verdeutlicht die Strategie der D-loop Amplifikation bei den Carabaos.

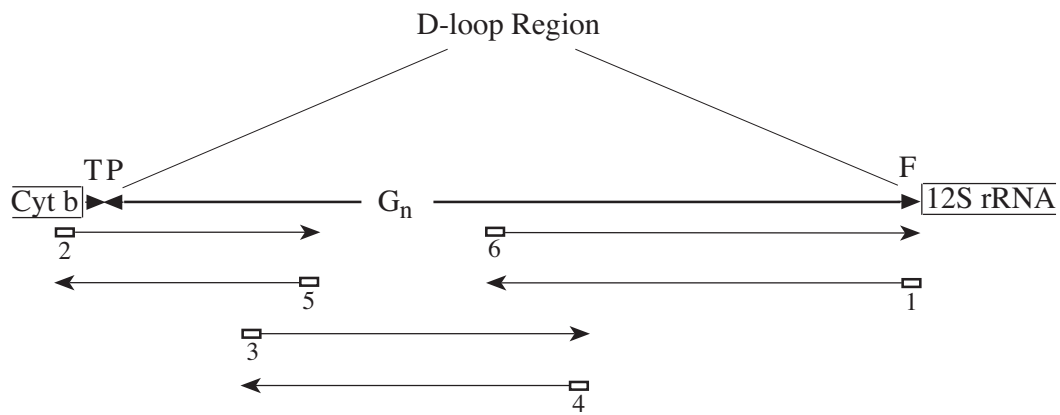


Abbildung 3-1 Schematische Darstellung der Strategie zur Amplifikation der D-loop Region.

Diese Strategie war speziell für die Carabao Proben notwendig, die einen poly-G Abschnitt ( $G_n$ ) aufwiesen. Dargestellt ist die D-loop Region mit ihren flankierenden Bereichen, die für Cytb, tRNAs (dargestellt sind die entsprechenden Aminosäuren im Ein-Buchstaben-Code; Spitze des Dreiecks gibt Transkriptionsrichtung an) und 12S rRNA kodieren. Die Oligonukleotide 1-6 sind der Tabelle 2 zu entnehmen.

Tabelle 2 Oligonukleotide zur Amplifikation der D-loop Region.

Die Nummern, die den einzelnen Oligonukleotiden vorangestellt sind, entsprechen der Beschriftung aus Abbildung 3-1.

Name	Sequenz
1: tRNA <sup>Phe</sup>	5'-AGG CAT TTT CAG TGC CTT GC-3'
2: Cytb	5'-TAG TGC TAA TAC CAA CGG CC-3'
3: G-stretch-uni	5'-CCA TCA ACA CAC CTG ACC-3'
4: G-stretch-rev	5'-GCG AGG ACG GAT TTG ACT-3'
5: D-loop-int-uni	5'-CCA TTC GGA GTA GTA GGG TC-3'
6: D-loop-int-rev	5'-CAT AAC ATT AAT GTA ATA AGG GC-3'

Allen Oligonukleotide wurde am 5' Ende eine Sequenz angehängt, die komplementär zu Standard-Sequenzieroligonukleotide waren (Tabelle 3). Diese Oligonukleotide waren IRD 800 markiert, und wurden in die Sequenzreaktion (Kapitel 3.2.3.1) eingesetzt.

Tabelle 3 Standard-Sequenzieroligonukleotide, IRD 800 markiert

Name	Sequenz
M13universal	5'-TGT AAA ACG ACG GCC AGT-3'
M13revers	5'- CAG GAA ACA GCT ATG ACC-3'

Die PCR wurde in einem Gesamtvolumen von 100  $\mu$ l angesetzt und enthielt 10 ng genomische DNA, 100 pmol eines jeden Oligonukleotids, 200  $\mu$ M dNTPs, 2,5 U (Einheiten) *Taq*-Polymerase, 10 mM Tris-HCl (pH 8,3), 50 mM KCl und 1,5 mM  $MgCl_2$ . Die besten Amplifikationen wurden unter folgenden Bedingungen in einem RoboCycler Gradient 2000 erreicht:

95 °C (1 min), 35 Zyklen bei 94 °C (1 min), 56 °C (1 min), 72 °C (1 min), gefolgt von einer abschließenden Extension bei 72 °C (4 min).

Die PCR-Produkte wurden elektrophoretisch auf einem 1%igen Agarosegel aufgetrennt und mit dem QIAEX II Kit aus dem Gel eluiert.

### 3.2.3 DNA-Sequenzierung

#### 3.2.3.1 Sequenzreaktion

Für Sequenzreaktionen wurde das "Thermo Sequenase fluorescent labelled primer cycle sequencing kit" benutzt. Die Sequenzreaktion beruht auf der Synthese neuer DNA Stränge, beginnend an einer spezifischen Bindungsstelle für Oligonukleotide und endend mit dem Einbau eines Didesoxynukleosid-triphosphats (Sanger *et al.*, 1977). Die relative Konzentration von dNTPs und ddNTPs ist so ausgewogen, daß im Durchschnitt an jeder Position innerhalb der Sequenz der Einbau eines ddNTPs gleich groß ist. Durch den Zusatz von 7-deza-dGTPs in dem Nukleotidmix dieses Kits werden Kompressionsartefakte während der Gelelektrophorese effektiv eliminiert. Aufgetragen auf einer geeigneten Gelmatrix erhält man die Sequenzinformation.

Für die Reaktionen wurden ca. 200 ng DNA/kb mit 2 pmol IRD 800 markiertem Oligonukleotid und 1 µl DMSO (Dimethylsulfat) versetzt und mit ddH<sub>2</sub>O auf 26 µl aufgefüllt. Je 6 µl dieses Ansatzes wurden zu 2 µl der Terminationsmische pipettiert, die in einer Mikrotiterplatte vorgelegt waren und anschließend mit Mineralöl überschichtet.

Die nachfolgende PCR erfolgte in einem RoboCycler Gradient 2000 unter folgenden Bedingungen: 94 °C (2 min), 30 Zyklen bei 94 °C (45 s), 58 °C (45 s) und 68 °C (45 s). Zur Vermeidung unspezifischer Kettenbrüche an den 7-deza-dGTPs und Degradierung des Fluoreszenzfarbstoffs wurden 5 µl Auftragspuffer erst kurz vor dem Auftrag auf das Sequenzgel dazugegeben.

Auftragspuffer	97,5 % (v/v)	Deionisiertes Formamid
	10 mM	EDTA
	0,1 % (w/v)	Bromphenolblau

#### 3.2.3.2 Automatische Sequenzierung

Zur Auftrennung der Sequenzreaktionen wurde ein LI-COR<sup>®</sup> 2000L und die zugehörige Steuersoftware BasemagelR v2.30 verwendet. Die Sequenzreaktionen wurden auf einem 41 cm langen, 6%igen Sequagel XR (unter Zusatz von 1 % (v/v) DMSO) aufgetragen. Die Auftrennung erfolgte bei 1500 V, 35 mA und 31 W für ca. 14 Stunden. Als Laufpuffer wurde 1x TBE long run verwendet. Nach dem Lauf wurden die Sequenzen mittels der Image Analysis Software v2.30 automatisch gelesen. Eine manuelle Korrektur erfolgte beim direkten Vergleich der gelesenen Sequenz mit dem aufgezeichneten Bild. Die Sequenzen wur-

den alsdann in das SCF-Dateiformat umgewandelt, um sie später mit dem Programm Sequencher™ 3.1.1 zu bearbeiten und auszuwerten.

Sequagel XR 6 %	24 ml	Sequagel XR
	6 ml	Sequagel complete Buffer Reagent
	300 $\mu$ l	DMSO
	350 $\mu$ l	10 % (w/v) APS
	10 $\mu$ l	TEMED
10x TBE long run	1,34 M	Tris Base
	450 mM	Borsäure
	25 mM	EDTA

### 3.2.4 Datenanalyse

Die SCF-Dateien wurden im Sequencher™ 3.1.1 als Sequenzdatei inklusive des zugehörigen Chromatogramms importiert. Hier konnten weitere Unsicherheiten in der Sequenz manuell ausgebessert werden. Als wichtigster Schritt wurden hier die Teilsequenzen der D-loop Region zu einer Gesamtsequenz zusammengefügt, zum sogenannten Contig.

Für die Murrah, Jafarabadi und Mediterraneo Proben wurden die universal und revers Sequenzen (siehe Kapitel 3.2.2) zu einem Contig zusammengefügt, die im mittleren Teil über einige hundert Basenpaare überlappten.

Für die Carabao Proben standen entsprechend mehr Teilsequenzen zur Verfügung, um einen individuellen Contig zusammzusetzen. So wurde für jedes Individuum ein Sequenz-Contig erstellt. Diese Contigs war Ausgangspunkt für die nachfolgenden phylogenetischen Analysen.

## 3.3 Software

Für die phylogenetische Analyse von Sequenzdaten stehen eine Vielzahl von Programmen zur Verfügung. Ein Großteil der Software ist im Internet frei verfügbar (z. B. unter <http://evolution.genetics.washington.edu/phylip/software.html>). Die für diese Arbeit relevante Software soll hier kurz in ihrer Anwendung beschrieben werden.

### 3.3.1 ClustalX

Mit dem Programm ClustalX (1.64b) wurden Multiple Sequenz Alignments durchgeführt, in die alle 80 Sequenzen einbezogen wurden. Die Contigsequenz aus dem Sequen-

cher™ 3.1.1 Programm wurde für die weitere phylogenetische Analyse in das, für das ClustalX Programm lesbare FASTA-Format gebracht. Dieses Format ist definiert durch ein ">" direkt vor der Bezeichnung der Sequenz (z.B. Taxon Name). Der folgende Absatz enthält die zugehörige Sequenz ohne Leerzeichen und/oder Absätze. Die Datei, die alle OTUs enthält wird als reine Text Datei abgespeichert.

```
>Taxon1¶
ATGCATGCGGTAGCGATTTCGCACAAATACGGTGGCCTTTTGGCTATATGC
ATGAGCGATTTCGCACAAATACGGTGG¶
>Taxon2¶
ATGCATGCGGTACCGATTTCGCACAAAACGGTGGGCTTTTGCTATATGCA
TGAGCGATTTCGCACAAATACGGTGG¶
Abbildung 3-2 Beispiel für das FASTA Format.
Dieses Format wurde zur Eingabe von Sequenzen in ClustalX gewählt. (¶ = Absatzmarke)
```

Diese Datei wird in ClustalX geöffnet, in der alle Sequenzen untereinander aufgelistet erscheinen. Für das multiple Alignment werden die Parameter festgelegt.

Tabelle 4 Parameter für das multiple Alignment im ClustalX.  
Beschrieben sind die Parameter für das multiple Alignment in ClustalX (1.64b) mit ihren Funktionen und die für die vorliegende Analyse gewählten Werte.

Parameter	Funktion	Wert (0-1)
Gap opening	Kosten für das Einfügen von Lücken. Hoher Wert reduziert die Frequenz von Lücken	0,7
Gap extension	Kosten für die Länge einer Lücke. Hohe Werte bevorzugen kurze Lücken	0,7
Verzögerung Sequenzen	divergenter Verzögert das Alignment divergenter Sequenzen, bis die näher verwandten Sequenzen ausgerichtet wurden.	0,8
Gewichtung der Transition	Gibt der Transition eine Gewichtung zwischen 0 und 1. 0 impliziert eine Transition als Fehlpaarung. Für nahe verwandte Sequenzen wird ein hoher Wert vorgeschlagen.	0,8

Es folgt ein komplettes multiples Alignment, daß in drei Stufen durchgeführt wird.

- Alle Sequenzen werden miteinander verglichen (paarweises Alignment)
- Ein Dendrogramm (ähnlich einem phylogenetischen Baum) wird konstruiert
- Das endgültige multiple Alignment wird durchgeführt, wobei das Dendrogramm als Leitstruktur dient.

Durch die farbliche Darstellung der verschiedenen Merkmale ist das Alignment am Bildschirm leicht zu analysieren. Eventuell wurden Teilbereiche des Alignments durch Variation der

Parameter (z.B. Erniedrigung der Gap opening und extension Parameter) erneut "re-aligned". Ebenso war es möglich das Alignment direkt im Textformat manuell zu korrigieren.

Das endgültige Alignment wurde im PAUP\* kompatiblen Format (NBRF/PIR) gespeichert.

### 3.3.2 MacClade

Die NBRF/PIR Datei wurde in PAUP\* importiert und ausgeführt, um sie dann in ein MacClade kompatibles Format zu exportieren. MacClade 3.08a ist ein Computerprogramm für phylogenetische Analysen. Seine Stärke liegt insbesondere in der Analyse der Nukleotid Evolution. Es bietet eine Vielzahl von Hilfsmitteln, um die Daten und Phylogenien zu editieren, phylogenetische Bäume zu konstruieren und auch zu manipulieren, um Hypothesen zu testen. Ein weiteres Hilfsmittel ist die Suche nach redundanten Sequenzen, die dann aus dem Datensatz entfernt werden. Ferner bietet es die Option die Sequenzmatrix vereinfacht darzustellen. Die Sequenz an erster Position dient als Referenz, in allen anderen Sequenzen des Alignments werden die Merkmale eliminiert, die identisch zur Referenz sind; es verbleiben die von ihr abweichenden Merkmale. In Kombination mit der Eliminierung nicht variablen Merkmale (siehe 2.2.6.1 Maximum Parsimonie) ergeben sich übersichtliche Darstellungen der relevanten Nukleotid Positionen. Für jeden verifizierten mitochondrialen Haplotypen wurde eine Sequenz als repräsentativ gewählt und mit diesen die phylogenetischen Studien fortgeführt. Das Alignment mit der endgültigen Anzahl der Haplotypen wird gespeichert und kann dann wieder in PAUP\* geöffnet und analysiert werden.

### 3.3.3 Modeltest

Das Programm Modeltest 3.0 (Posada & Crandall, 1998) bestimmt in Verbindung mit PAUP\* auf Ebene von Likelihood Werten ( $-\log L$ ), das Modell der DNA Evolution, welches am besten zu den Daten paßt.

Jede phylogenetische Methode basiert auf einem Modell der DNA Substitution (Felsenstein, 1988). Folglich hängt jede phylogenetische Ableitung von dem zugrunde liegenden Modell ab. Um also Vertrauen in der Ableitung zu haben, muß man Vertrauen in die Modelle haben (Goldman, 1993). Es gilt zu prüfen, welches Modell die beobachteten Daten am besten beschreibt bzw. erklärt. Die weitgehend akzeptierte statistische Methode, um die Güte eines Modells zu messen ist der Wahrscheinlichkeitsverhältnis Test  $\delta = -2 \log \Lambda$ , mit

$$\Lambda = \frac{\max[L_0(\text{Null Model} / \text{Daten})]}{\max[L_1(\text{alternatives Model} / \text{Daten})]}$$

wobei  $L_0$  die Wahrscheinlichkeit unter der Null Hypothese (also das einfache Modell) ist, und  $L_1$  die Wahrscheinlichkeit unter der alternativen Hypothese (also das komplexere Modell mit mehr Parametern). In Modeltest sind 56 verschiedene Modelle implementiert, die getestet werden. Es ist aber zu beachten, daß das Testsystem bestimmte Modelle bevorzugen kann. Der Nutzer des Programms ist daher gehalten, das Ergebnis kritisch zu betrachten.

Zur Verfahrensweise: der zu untersuchende Datensatz wird in PAUP\* geöffnet und ausgeführt. Durch Doppelklick der Programmuntereinheit „modelblock3“ wird diese in PAUP\* geöffnet und man führt sie dann dort aus. Modeltest beginnt nun mit der Analyse das optimale Modell zu finden. Während der Analyse werden auch der  $\Gamma$ -Parameter, die Basenzusammensetzung sowie das Transitions-/Transversionsverhältnis abgeschätzt. Das Ergebnis wird in der Datei "model.score" gespeichert und läßt sich im Textverarbeitungsprogramm öffnen und lesen.

### 3.3.4 PAUP\* (4.0b4a)

Das Programm PAUP\* Version 4.0b4a bietet unter anderem die Analyse der Daten nach den drei Optimalitätskriterien Distanz, Maximum Likelihood (ML) und Maximum Parsimonie (MP) (Swofford, 2000). Die Datei wird in dem Programm ausgeführt, es werden die phylogenetische Außengruppe (hier: *Bos taurus*), das gewünschte Analyse Kriterium und die Parameter definiert.

Distanz: Zur einfachen Darstellung der Distanzunterschiede zwischen den Sequenzen wurde die unkorrigierte p-Distanz (siehe 2.2.5.1 Unkorrigierte Distanzen) gewählt. Zur Abschätzung der evolutiven Distanz wurde das TrN-Modell (siehe 2.1.6.5 Tamura und Nei (TrN) -Modell) gewählt. Die hieraus errechneten Distanzwerte wurden zur Abschätzung der Divergenz herangezogen und die neighbor-joining Bäume (Saitou & Nei, 1987) konstruiert.

ML: Die Minimum Evolution Methode wurde auf die Daten angewandt unter Verwendung des Optimalitätskriterium Maximum Likelihood. Die Parameter wie Substitutionsmodell, Basenfrequenz und Gammaverteilung basierten auf das von Modeltest vorgeschlagene HGY+ $\Gamma$ -Modell und die dort ermittelten Parameter (Hasegawa *et al.*, 1985; Yang, 1994).

MP: Eine heuristische Suche mit bis zu 1000 Wiederholungen und zufälliger schrittweise Addition der Sequenzen unter Maximum Parsimonie wurde durchgeführt.

### 3.3.5 Network 2.0c

Das Programmpaket Network 2.0 (Bandelt *et al.*, 1999) wurde komplett in Borland Pascal 7.0 geschrieben und ist für die Analyse intraspezifischer, rekombinationsfreier Daten, wie mtDNAs (in Form von Punktmutationen, Aminosäuresequenzen und RFLP-Schnittstellen) konzipiert. Es kann auf jedem DOS-Rechner ab 80386er Prozessor installiert werden. Allerdings sollten mindestens 1 MB Hauptspeicher und ca. 3 MB auf der Festplatte verfügbar sein.

Durch den Aufruf 'network' innerhalb des entsprechenden Verzeichnisses auf der Festplatte wird eine Benutzeroberfläche gestartet, von der aus alle implementierten Werkzeuge und Algorithmen aufgerufen werden können. Das komplette Programmpaket ist mit Ausnahme einiger Optionen des Zeichenalgorithmus ausschließlich mit der Tastatur zu bedienen. Die wesentlichen Funktionstasten sind die Cursortasten und die Escapetaste.

In Network 2.0 sind Werkzeuge zur Generierung der Datensätze enthalten, die die benötigten Daten abfragen und in einem entsprechendem Format speichern. Da das importieren von Daten kompliziert ist (nur wenige Dateiformate werden erkannt) macht es weniger Mühe und ist vermutlich fehlerfreier, die Daten manuell einzugeben. Es ist auch nicht so arbeitsaufwendig, wie es erst erscheint, da nur eine Sequenz eingegeben wird. Diese wird für alle Taxa übernommen und man ändert später die unterschiedlichen Merkmalszustände.

Der „median-joining (MJ) Algorithmus (Bandelt *et al.*, 1999) ist in dem Programm implementiert. Es können bis zu 14.000 Sequenzen mit 500 Merkmalen in die Datenmatrix eingegeben und berechnet werden.

Die Ausgabe des Netzwerks kann wahlweise über einen Drucker, den Bildschirm oder in eine Datei erfolgen. Bei der Ausgabe in eine Datei werden zusätzlich die Frequenzen und die Gewichtung ausgegeben. Weiterhin enthält der Dateikopf die Einstellungen der Parameter.

Grundsätzlich ist eine Ausgabe in eine Datei zu empfehlen, da diese später immer noch ausgedruckt werden kann, und die Ausgabedatei des MJ-Algorithmus als Eingabedatei für das Graphikprogramm dient. Die Datei wird in dem Graphikprogramm geöffnet und ein Netzdiagramm generiert. Es stehen Werkzeuge zur Verfügung, um Knoten zu rotieren und die Ansicht zu vergrößern oder zu verkleinern. Das Diagramm läßt sich als Bitmap-Datei speichern und in entsprechenden Graphikprogrammen öffnen und bearbeiten.

### 3.3.6 Swan v1.0

Das Programm Swan v1.0 (Proutski & Holmes, 1998) nutzt eine "gleitende Fenster" Methode, um die Variabilitätsverteilung in einem Alignment zu analysieren. Die Variabilität  $Var$  in jeder Spalte des Alignments wird als Entropiefunktion der Merkmalsvariation der jeweiligen Spalte nach folgender Gleichung berechnet:

$$Var = - \sum_{i=A,C,G,T} \frac{n_i}{N} \ln \frac{n_i}{N} \quad (15)$$

Dabei ist  $n_i$  die Anzahl eines jeden der Merkmale (A, C, G oder T) in einer Spalte des Alignments und  $N$  die Anzahl der analysierten Sequenzen.

Weiter berechnet das Programm die Variabilitätsdichte, wobei die Fenstergröße und die Bewegung des Fensters variiert werden kann. Die Analyse wird in einem Fenster graphisch dargestellt und kann als pict-Datei gespeichert werden.



## 4 Ergebnisse

Die DNA-Sequenz der D-loop Region wurde für 80 Individuen von *Bubalus bubalis* über eine Länge von etwa 916 bp bestimmt. Im einzelnen handelte es sich um 19 Sumpfbüffel (Carabao) und 61 Flußbüffel (24 Murrah, 20 Jafarabadi, 17 Mediterraneo). Die Individuen wurden mit den ersten drei Buchstaben ihrer Rasse in Verbindung mit ihrer individuellen Nummer aus Tabelle 1 bezeichnet (z. B. Jaf30).

Aufgrund Polymorphismen im sequenzierten Bereich wurden Haplotypen definiert. Diese lagen der phylogenetischen Analysen und der Abschätzung der Divergenz zugrunde.

### 4.1 Die bubalinen Haplotypen

#### 4.1.1 Diversität in den einzelnen Rassen

Der Vergleich der Sequenzen innerhalb der einzelnen Rassen zeigt eine geringe Diversität innerhalb der Jafarabadi und der Carabaos. Von den 20 sequenzierten Jafarabadi Proben ließen sich 5 Haplotypen bestimmen. Das Alignment der polymorphen Positionen (Abbildung 4-2) zeigt, daß auch diese Sequenzen z.T. nur durch eine Substitution zu unterscheiden sind. Ähnliches gilt für die 10 Carabao Haplotypen, die hauptsächlich Unterschiede in den Längenpolymorphismen aufweisen (siehe unten). Im Gegensatz zu diesen beiden Rassen ist die untersuchte Gruppe der Murrahs und Mediterraneos sehr viel diverser, mit 11 und 14 Haplotypen.

Im Programm MacClade konnten alle redundanten Sequenzen ermittelt werden. Es ergaben sich daraus 36 Haplotypen. Die folgende Auflistung (Tabelle 5) gibt alle redundanten Haplotypen wieder. Bei den redundanten Haplotypen fanden sich einige, die in mehreren Rassen vorkamen, gekennzeichnet durch unterstreichen.

Ein spezieller Haplotyp ist Mur481: phänotypisch ein Flußbüffel, zeigt er auf Ebene der mtDNA deutliche Anzeichen zur Zugehörigkeit zu den Sumpfbüffeln. Darauf wird im weiteren noch näher eingegangen werden (speziell aber in Kapitel 4.1.2.8 Der Büffel Mur481).

Tabelle 5 Auflistung der redundanten Haplotypen. Aufgeführt sind alle Haplotypen, die in mindestens zwei Individuen gefunden wurden. Die redundanten Haplotypen aus einer anderen Rasse sind unterstrichen.

Haplotyp	Gleich wie:
Jaf30	Jaf25
Jaf32	Jaf40, 19, 01, 04, 27, 23
Mur314	Mur143, 196, IZ143, <u>Jaf02</u> , 03, 44, 47, 50, 51, 53, 55, 60, <u>MedC8</u> , D6
Mur110	Mur105, 183, 117, 267
Mur310	MurIZ244, <u>MedC7</u> , D2
MurIZ369	MurIZ287, IZ364, IZ368
MurIZ192	MurIZ358, IZ278, <u>MedC3</u>
MedC1	MedD9
Car320	Car348, 335, 341
Car315	Car366, 359, 325, 380, 349
Car344	Car375

#### 4.1.2 Sequenzinformationen der bubalinen D-loop Region

##### 4.1.2.1 Sequenzlänge

Die Länge der untersuchten Sequenzen betrug zwischen 910 bp und 916 bp. Die Längenunterschiede sind hauptsächlich auf Längenpolymorphismen in der 5'- und 3'-hypervariablen Region zurückzuführen. Die gesamte D-loop Region ist jedoch noch etwa 14 bp länger, war aber nur für einige Individuen bekannt. Die Sequenzen wurden deshalb für die phylogenetische Analyse auf eine einheitliche Länge gekürzt (Abbildung 4-1).

Direkt an die D-loop Region schließt sich die Sequenz für die tRNA<sup>Phe</sup> an. Die Sequenzunterschiede zwischen Rind und Büffel sind am Beispiel von Mur481 und *B. taurus* in Abbildung 4-1 dargestellt. In der bovinen Sequenz folgt gleich nach dem C<sub>12</sub> Motiv die Sequenz für die tRNA<sup>Phe</sup>. Die bubaline Sequenz hat keine homologe Sequenz zu dem poly-C Motiv. Satt dessen folgt eine AC-reiche Region bis zur homologen Sequenz für die tRNA<sup>Phe</sup>.

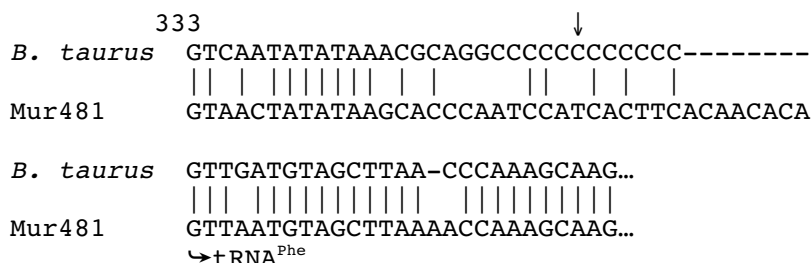


Abbildung 4-1 Übergang von der D-loop Region zur tRNA<sup>Phe</sup>. Der Vergleich der Sequenzen zwischen Büffel (hier Mur481) und *B. taurus* am Übergang von der D-loop Region zur tRNA<sup>Phe</sup>. Senkrechte Striche (|) zwischen den Sequenzen kennzeichnen übereinstimmende Basen. Querstriche (-) markieren Deletionen. Position nach (Anderson *et al.*, 1982). Der Pfeil (↓) markiert die Position bis zu der die Sequenz für alle Haplotypen bekannt ist.



#### 4.1.2.2 Sequenzalignment

Das komplette Sequenzalignment findet sich in Anhang 2. Zur übersichtlichen Darstellung sind in Abbildung 4-2 nur die variablen Positionen der 36 Haplotypen in einem Alignment dargestellt. Es wurden 128 variable Positionen gefunden (Abbildung 4-2).

In der phylogenetischen Software werden gewöhnlich Lücken wie fehlende Basen betrachtet. Mit anderen Worten, ist ein Merkmal im Alignment konstant für einen, oder alle bis auf einen Haplotypen und die restlichen Haplotypen haben eine Lücke, so wird diese Position als konstant betrachtet. Beispiel hierfür ist die Position 169 in Abbildung 4-2: während die Carabaos an dieser Position ein T haben, haben die anderen Tieren eine fehlende Base (-). Diese Position wird daher als konstant betrachtet. Unter dieser Bedingung waren dann 825 Nukleotide konstant und 99 variable, von denen 72 unter dem Parsimoniekriterium informativ waren (siehe auch Kapitel 2.2.6.1 Maximum Parsimonie).

#### 4.1.2.3 Basenfrequenz

Anhand der 36 Haplotypen wurde eine Analyse der Basenfrequenz durchgeführt. Das Ergebnis ist in Tabelle 6 zusammengefaßt. Wie es für andere Säugetiere schon gezeigt wurde, ist auch die D-loop Region der Büffel sehr AT-reich, wobei die Flußbüffel einen etwas höheren AT Anteil haben als die Carabaos. Verglichen mit der bovinen Sequenz ist der AT Gehalt bei den Büffeln etwa um 2,75 % niedriger.

Tabelle 6 Basenfrequenzen für die D-loop Region.  
Ermittelt wurde der prozentuale Anteil der Nukleotidbasen für Fluß- und Sumpfbüffel. Zum Vergleich ist die selbe Analyse für die bovine Sequenz angeführt.

	Adenin (%)	Cytosin (%)	Guanin (%)	Thymin (%)
Flußbüffel	31,3 ± 0,2	25,4 ± 0,3	14,7 ± 0,2	28,6 ± 0,2
Sumpfbüffel	30,7 ± 0,2	26,2 ± 0,1	15,4 ± 0,2	27,7 ± 0,1
<i>Bos taurus</i>	32,9	24,4	13,7	29,0

#### 4.1.2.4 Transitionen und Transversionen

Der Sequenzvergleich innerhalb der 36 Haplotypen zeigte 48 C↔T und 37 A↔G Transitionen (mehr Substitutionen zwischen Pyrimidinen als zwischen Purinen), aber nur 9 Transversionen; das entspricht etwa einem Verhältnis von 9,4:1. Dieses Ergebnis demonstriert das starke Ungleichgewicht zugunsten der Transitionen. Fünf variable Positionen zeigten drei verschiedene Merkmale (Tabelle 7).

Tabelle 7 Auflistung der Transversionen.

Die Anzahl und Position von Transversionen im bubalinen Sequenzalignment ist hier aufgelistet. Die rechte Seite zeigt Merkmalspositionen auf, die drei Status aufweisen.

Position	Anzahl	Status	Position	Anzahl	Status
16	1	CG	162, 188	2	CGT
54, 732, 751, 756	4	AC	271	1	AGT
190	1	GT	739	1	ACG
297, 448, 791	3	AT	920	1	ACT

Das Sequenzalignment mit *Bos taurus* umfaßt 940 Merkmale (Anhang 3), von denen 747 konstant sind. 118 variable Positionen sind unter dem Parsimoniekriterium nicht informativ, 75 Merkmale sind Parsimonie-informativ. Die hohe Zahl Parsimonie nicht informativer variabler Positionen beruht auf den Unterschieden zwischen der bovinen Sequenz und den konstanten Positionen im bubalinen Sequenzalignment. Beispiel hierfür sind die Positionen 3, 17, 22 etc. des Alignments aus Anhang 3.

Die paarweise Sequenzdivergenz wurde in PAUP\* unter dem TrN-Modell (Tamura-Nei) berechnet und in Microsoft-Excel graphisch dargestellt (Abbildung 4-3). Aufgetragen ist die Anzahl variabler Merkmale (unterschieden nach Transitionen und Transversionen) gegen die paarweise Sequenzdivergenz. Der linke Teil der Graphik stellt die variablen Merkmale innerhalb der untersuchten Büffel gegen die paarweise Sequenzdivergenz dar. Der Unterschied innerhalb der Fluß- bzw. der Sumpfbüffel ist gering und liegt bei unter 5 %. Der Unterschied zwischen Sumpf- und Flußbüffel liegt bei etwa 10 %. Der paarweise Vergleich von Mur481 gegen die Carabaos ergibt einen Unterschied von etwa 5 %.

Eine evolutive Divergenz von 30 % - 40 % ergibt sich für den Vergleich der Flußbüffel gegen die Außengruppe *Bos taurus*. Die Divergenz zwischen der Außengruppe und Sumpfbüffel liegt nochmals etwa 10 % höher. Interessanterweise zeigt die paarweise Sequenzdivergenz zwischen Büffel und Rind eine Sättigung für Transversionen, aber einen annähernd linearen Zuwachs für Transitionen (rechter Teil der Abbildung 4-3).

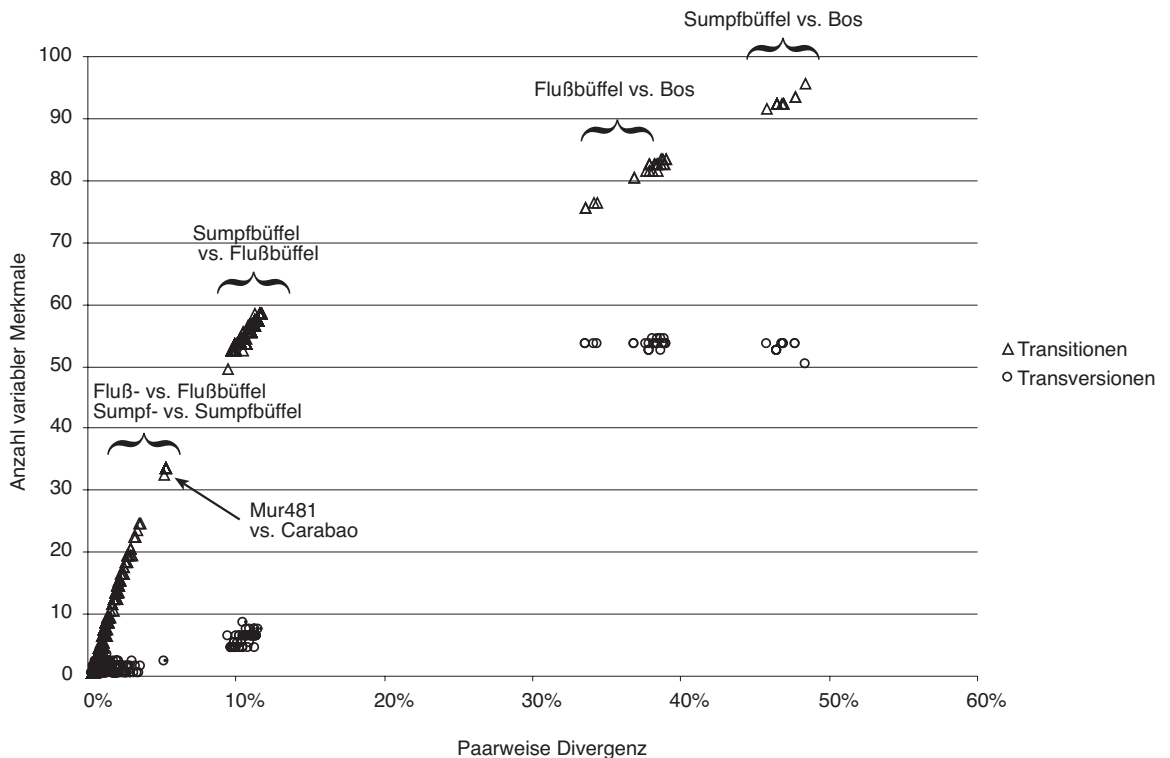


Abbildung 4-3 Variable Merkmale gegen paarweise Divergenz, mit Außengruppe. Graphische Darstellung der Anzahl von Transitionen und Transversionen gegen die prozentuale paarweise Divergenz. Mit geschweiften Klammern sind deutlich getrennte Gruppierungen markiert. Der Berechnung der paarweisen Divergenz lag das TrN-Modell zugrunde.

Betrachtet man den linken Teil der Abbildung 4-3 isoliert, so kommt man zu Abbildung 4-4. Hier sind Transitionen und Transversionen gegen die paarweise Divergenz ohne Außengruppe (nach dem TrN-Modell) aufgetragen. Für die Werte der Transition und Transversion wurde jeweils eine Trendlinie und der  $R^2$ -Wert berechnet. Für die Transition wurde eine polynome, für die Transversion eine lineare Trendlinie berechnet.

Der  $R^2$ -Wert gibt, einfach formuliert, die Güte der berechneten Trendlinie an. Je näher der Wert an 1 liegt, desto besser entsprechen die Werte der berechneten Linie. Für die Transitionen ergab sich ein Wert von 0,9992. Die Werte der Transversion streuen etwas, so daß der  $R^2$ -Wert bei etwas über 0,9 liegt.

Wie in der Abbildung 4-4 (und auch in Abbildung 4-3) zu sehen ist, ist für die Transitionen innerhalb der Büffel keine Sättigung zu erkennen. Ein wichtiger Aspekt, um Aussagen über die Güte phylogenetischer Berechnungen zu machen (siehe Kapitel 2.1.5.2 Molekulare Uhr).

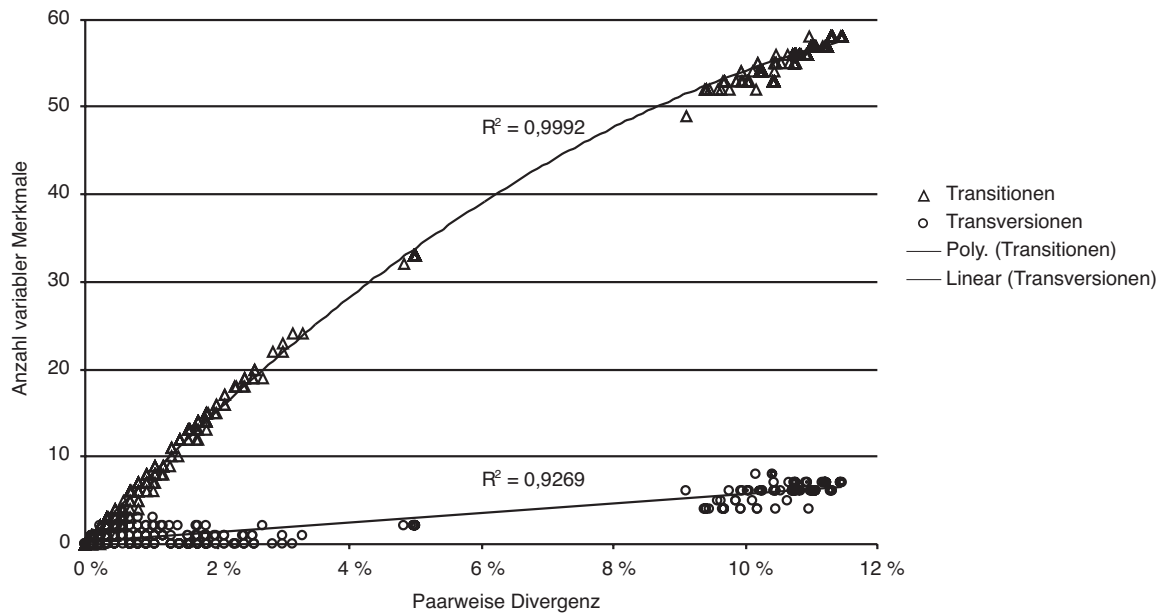


Abbildung 4-4 Variable Merkmale gegen paarweise Divergenz, ohne Außengruppe. Graphische Darstellung der Anzahl von Transitionen und Transversionen gegen die prozentuale paarweise Divergenz innerhalb der untersuchten Büffel. Für die Transitionen wurde eine polynome, für die Transversionen eine lineare Trendlinie mit  $R^2$ -Wert berechnet.

#### 4.1.2.5 Verteilung variabler Positionen im bubalinen Alignment

Die Verteilung der variablen Positionen und ihre relative Dichte wurde mit Hilfe des Programmes SWAN (v1.0) bestimmt. Die Verteilung variabler Positionen auf die Länge der D-loop Region ist in Abbildung 4-5a dargestellt. Zur Berechnung der Variabilität siehe Kapitel 3.3.6 Swan v1.0. Für die Dichteanalyse (Abbildung 4-5b) wurde eine Fenstergröße von 10 bp gewählt. Bei der Analyse der Verteilung variabler Positionen innerhalb der D-loop Region, lassen sich zwei hypervariable Bereiche ausmachen.

Der erste Bereich, in der 5'-Region, umfaßt etwa die Positionen 120 bis 540. Dieser Bereich enthält 66 variable Positionen, von denen 52 Parsimonie informativ sind. Das heißt,  $\frac{2}{3}$  aller variablen Positionen der bubalinen D-loop Region befinden sich in einem Bereich von 420 bp.

Die zweite hypervariable Region findet sich im 3'-Bereich (Pos. 730-920). Hier finden sich 24 variable Positionen, von denen 14 Parsimonie informativ sind. Beide Bereiche zusammen (etwa 610 bp) decken also etwa 90 % aller variablen Positionen ab.

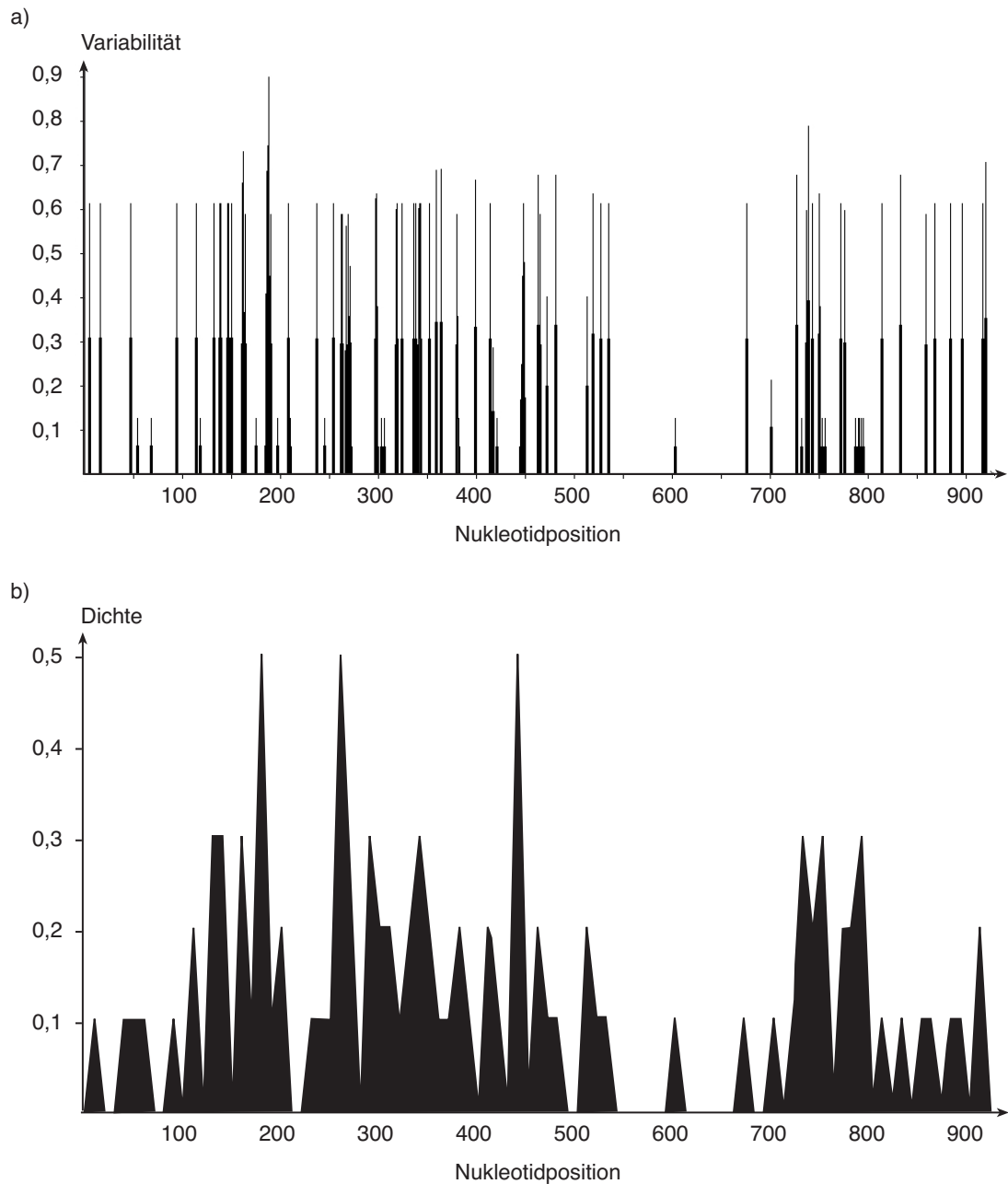


Abbildung 4-5 Variabilität und Dichte der Substitutionen in der bubalinen D-loop Region. Graphische Darstellung der Variabilitäts- (a) und der Dichteverteilung (b) innerhalb der bubalinen D-loop Region. Die Variabilität wurde als Entropiefunktion der Nukleotidvariation in einer Spalte des Alignments berechnet (siehe Kapitel 3.3.6 Swan v1.0), die Dichteverteilung als Anzahl der variablen Positionen in einem Fenster von 10 bp.

#### 4.1.2.6 Längenpolymorphismen

Wie schon erwähnt gibt es in den hypervariablen Regionen Längenpolymorphismen. Im 5'-Bereich (Position 179-190) besitzen alle Büffel ein poly-Guanin Motiv. Jedoch gibt es einen deutlichen Unterschied zwischen Fluß- und Sumpfbüffel, der durch die Konsensussequenz dieses Bereiches deutlich wird (siehe Abbildung 4-6)



```

Flußbüffel  --GGGGAATGTA
              |||||  | |
Sumpfbüffel  GGGGGGGGGGA

```

Abbildung 4-6 Vergleich des poly-G Motivs.

Vergleich der Sequenzunterschiede zwischen Fluß- und Sumpfbüffel im 5'-Bereich der D-loop Region (Position 178-190). Sequenzen entsprechen den Konsensussequenzen aus Abbildung 4-9.

Die Länge des poly-G Motivs variiert bei den Flußbüffeln zwischen 4 und 6 Guaninbasen; z.T. ist das angrenzende Adenin durch ein Guanin substituiert (Position 186 in Abbildung 4-2 bzw. Anhang 2). Bei allen Carabaos ist das poly-G Motiv einheitlich, mit Längenunterschieden zwischen 10 und 12 Guaninbasen. Wie schon weiter oben erwähnt soll auch hier Mur481 gesondert betrachtet werden. Während in vielen Eigenschaften Mur481 eher den Carabaos ähnelt, ist das poly-G Motiv nicht wie bei den Carabaos, sondern ähnlich dem der Flußbüffel.

Zwei weitere Längenpolymorphismen finden sich im 3'-Bereich der D-loop Region zwischen den Positionen 779 und 806. Hier gibt es zwei poly-Cytosin Motive. Das erste zwischen 6-8 bp, der zweite zwischen 7-9 bp lang. Beide Bereiche sind durch ein TCC TAT CAA A Motiv getrennt; manche Merkmale können dabei individuell substituiert sein. In diesem Bereich gibt es keine spezifischen Unterschiede zwischen Fluß- und Sumpfbüffel.

Die gleichen Längenpolymorphismen finden sich ebenfalls im Alignment mit *B. taurus* als Außengruppe wieder (Anhang 3). Das bovine poly-G Motiv gleicht dabei dem der Flußbüffel. Die beiden poly-C Abschnitte sind homolog zum bubalinen Bereich, allerdings nur mit 6 und 5 Cytosinbasen. Die bovine Sequenz zwischen den poly-C Motiven unterscheidet sich von der bubalinen: C<sub>6</sub>TTC ATA AAA AAT TTC<sub>5</sub>.

#### 4.1.2.7 Deletionen / Insertionen

Da in den seltensten Fällen eine ancestrale Sequenz bekannt ist, die als Referenz dient zu entscheiden, ob eine Deletion oder Insertion stattgefunden hat, spricht man i. a. von Deletions/Insertions (Del/Ins) Ereignissen.

Abgesehen von den Del/Ins Ereignissen, die durch Längenpolymorphismen entstehen, gibt es im Mur182 an mehreren Stellen Deletionen (Positionen 359, 364, 378, 398, 410, 442) und fast gleich viele Insertionen (Positionen 482, 485, 693, 741) einzelner Merkmale.

Ansonsten gibt es noch weitere vereinzelte Del/Ins Ereignisse (siehe Anhang 2):

MedB5:	Deletion von drei Basen (Position 735-737)
Jaf66:	Deletion einer Base (771)
Mur481:	Deletion von vier Basen (775-778)
MedD10:	Deletion einer Base (775)

Der Vergleich mit der bovinen Sequenz offenbart größere Del/Ins Ereignisse (Anhang 3). So gibt es zwischen Position 129 und 151 einige Deletionen von je 1 bis 4 Nukleotide. Hier zeigt sich deutlich die Schwierigkeit, daß korrekte Alignment zu finden. Der entsprechende Abschnitt ist in Form einer Dotplotmatrix in Abbildung 4-7 zu sehen. In solch einer Matrix sind die Merkmale zweier zu alignende Sequenzen als Ordinate und Abszisse aufgetragen.

Punkte (dots) werden an Schnittpunkten korrespondierender Merkmale eingezeichnet (plot). Die kürzeste Verbindung vom Ursprung (links unten) nach rechts oben über möglichst viele "dots" gibt ein optimales Alignment wieder.

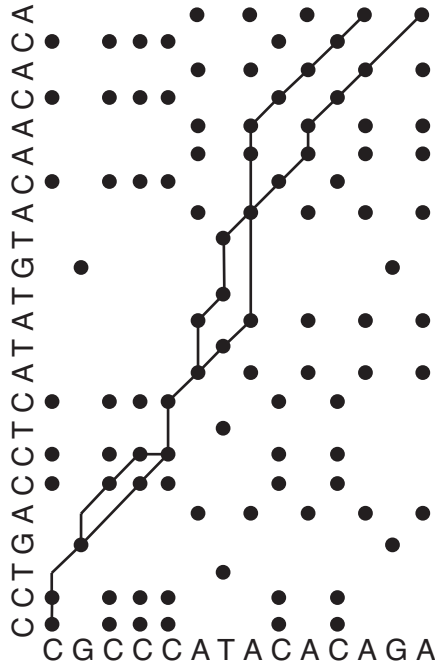


Abbildung 4-7 Dotplotmatrix für zwei Sequenzen. Die Matrix verbindet den Sequenzbereich von Position 129 bis 151 des Alignments aus Anhang 3. Eingezeichnet sind alternative Wege für das Alignment beider Sequenzen.

Das Problem des korrekten Alignments wird hier deutlich: Es gibt viele Kombinationen die beiden Sequenzen zu alignen (Linien, die die Punkte verbinden). Zwei Möglichkeiten sind in Abbildung 4-8 dargestellt. Dabei sind nicht alle Möglichkeit einer Substitution berücksichtigt. In Abbildung 4-8a) zum Beispiel ist eine C→G Substitution angenommen, um die Anzahl von Del/ins Ereignissen zu verringern.

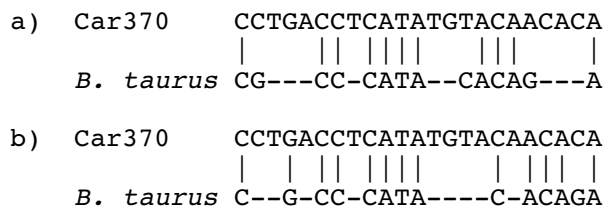


Abbildung 4-8 Vergleich zweier möglicher Alignments. Alignment der bubalinen mit der bovinen Sequenz zwischen Position 129 und 151. Die beiden Möglichkeiten in a) und b) demonstrieren das Problem, daß richtige Alignment zu finden.

Es ist klar, man kann nicht feststellen, welche Merkmale homolog sind. Man muß sich aber entscheiden, mit welchen Kosten man die Anzahl und Ausdehnung von Deletionen und die

Möglichkeit für Transversionen und Transitionen belegen will (vergleiche Tabelle 4). In dieser Arbeit wurde das alternative Alignment wie in Abbildung 4-8a) gewählt, um den Kosten für Transversionen und zusätzlichen Del/Ins Ereignissen gerecht zu werden.

Zwei weitere Del/Ins Ereignisse an den Position 168 und 169 zeigen spezifische Deletionen für Fluß- und Sumpfbüffel (Anhang 3). Die Sequenz CCTA wird bei allen Flußbüffeln zu CC-A und bei den Carabaos zu C-TA.

Zwischen den Positionen 357 und 364 haben alle Büffel eine Deletion des bovinen Sequenzabschnittes TAT ATA TT.

Weitere einzelne Del/Ins Ereignisse lassen sich dem Alignment aus dem Anhang 3 entnehmen. Durch diese Deletionen bzw. Insertionen umfaßt das Alignment mit *B. taurus* 940 bp.

#### 4.1.2.8 Der Büffel Mur481

Der Vergleich der Sequenzen, und, wie später noch weiter ausgeführt wird, die phylogenetische Analyse, läßt die Sequenz des Flußbüffels Mur481 besonders erscheinen. Die Abbildung 4-2 zeigt eine deutlich größere Ähnlichkeit der Mur481-Sequenz zu denen der Carabaos als zu denen der anderen Flußbüffel. Eine Analyse der variablen Positionen zwischen Mur481 und der Konsensussequenzen von Fluß- bzw. Sumpfbüffel (Abbildung 4-9) bestätigt diesen Eindruck. Mur481 hat im Vergleich der Konsensussequenzen 3 Merkmal gemeinsam mit *Bos* aber nicht mit Fluß- oder Sumpfbüffel. In 14 Positionen zeigt Mur481 Merkmale, welches nicht in den anderen Individuen beobachtet wurde. 18 variable Positionen teilt sich Mur481 mit den Flußbüffeln, hingegen 43 Positionen mit den Sumpfbüffeln.

Aus diesen Beobachtungen, inklusive der phylogenetischen Analysen, folgte die Annahme, daß es sich bei dem Tier Mur481 auf mitochondrialer Ebene um einen Sumpfbüffel handelt, d. h. in der maternalen Linie besteht Verwandtschaft zu Sumpfbüffel. Es gilt zu beachten, daß das Probenmaterial der Sumpfbüffel nicht sehr umfangreich, und zudem von einer einzigen Farm stammte. Über deren Verwandtschaftsverhältnisse ist nichts bekannt. Da aber die Sequenzunterschiede innerhalb der Carabaos minimal sind, muß eine nahe Verwandtschaft angenommen werden. Deshalb, um weitere Klarheit über den Mur481-Haplotypen zu erlangen, wurde eine weitere Untersuchungen angeschlossen: Die hier vorliegenden Daten wurden mit publizierten Daten aus der Datenbank verglichen und insgesamt analysiert (Lau *et al.*, 1998). Siehe Kapitel 4.2.3.

#### 4.1.3 Strukturelemente der bubalinen D-loop Region

Für einen einfacheren Vergleich der bubalinen mit der bovinen D-loop Region, wurde eine Konsensussequenz für Fluß- und Sumpfbüffel ermittelt (Abbildung 4-9). Wegen des speziellen Charakters von Mur481, ist diese Sequenz ebenfalls in das Alignment der Konsensussequenzen einbezogen worden. An einer gegebenen Position wurde eine Base als Konsensusbase bezeichnet, wenn sie von mehr als 50 % der Individuen geteilt wurde.

Anhand der Konsensussequenzen wurden strukturelle Elemente identifiziert, die innerhalb der Mammalia hochkonserviert sind. Solche Elemente wie CSB (conserved sequence block) oder ETAS (extended termination associated sequence) sollen spezifische Erkennungstellen für regulatorische Proteine sein (Saccone *et al.*, 1987; Sbisà *et al.*, 1997). Diese Sequenzbereiche sind in Abbildung 4-9 durch Unterstreichung markiert. Ein weiterer Bereich ist bei den Mammalia hochkonserviert. Er unterliegt so gut wie keinen Mutationen und wird daher in der Literatur als zentrale Domäne (central domaine; CD) bezeichnet. Dieser Bereich ist in der Abbildung grau unterlegt.

#### 4.1.3.1 Die 3'- und 5'-variablen Bereiche der D-loop Region

Die bubaline ETAS1 Region ist mit 61 bp um eine Base länger als die taurine. Innerhalb der untersuchten Büffel finden sich acht variable Positionen, zwei invariable Positionen innerhalb der Büffel sind variabel gegenüber *B. taurus*. Die ETAS2 erstreckt sich über 59 bp (ein mehr als *B. taurus*) und enthält in den Büffeln acht variable Positionen; zehn konstante Positionen sind gegenüber der taurinen Sequenz variabel. ETAS1 und ETAS2 sind durch 2 bp von einander getrennt.

Die zentrale Domäne (in der Abbildung grau unterlegt) erstreckt sich über eine Länge von 316 bp, die selbe Länge wie beim Rind. Verglichen mit *Bos taurus* enthält der Bereich 34 variable Positionen, von denen 19 konstant innerhalb der Büffel sind. Insgesamt zeichnet sich die zentrale Domäne durch längere konservierte Abschnitte aus.

Die "conserved sequence blocks" (CSB1 und 2) sind 25 bzw. 20 bp lang. CSB1 enthält vier variable Positionen, sowohl innerhalb der Büffel als auch gegenüber der bovinen Sequenz. CSB2 enthält eine A→T Transition in allen Büffel und eine C→T Transversion. Fluß- und Sumpfbüffel haben eine Insertion von fünf Basen (Mur481 eine Base) direkt vor einem poly-C Motiv.

Ferner sind noch der Ursprung der H-Strang Replikation ( $O_H$ ) eingezeichnet und die Promotorregionen für die H- und L-Strangsynthese (HSP und LSP).

---

#### Nächste Seite

Abbildung 4-9 Variable Positionen in der D-loop Region von Rind und Büffel. Gezeigt werden die Konsensussequenzen von Sumpf- und Flußbüffel sowie der Sequenz von Mur481 verglichen mit der bovinen Sequenz (Anderson *et al.*, 1982). Punkte stehen für übereinstimmende Basen, Striche für Deletionen. Strukturelle Elemente der D-loop Region sind markiert: ETAS1 und ETAS2 sind unterstrichen, die zentrale Domäne ist grau unterlegt, die konservierten Sequenz Blöcke (CSB) 1 und 2 sind ebenfalls unterstrichen. Der Replikationsursprung des H-Strangs ( $O_H$ ) sowie der L-Strang und H-Strang Promotor (LSP und HSP) sind durch abgeknickte Pfeile symbolisiert, die die Transkriptionsrichtung angeben. Zur Orientierung sind einige Position oberhalb der bovinen Sequenz notiert (Anderson *et al.*, 1982).



Der Vergleich der taurinen mit den bubalinen Konsensussequenzen zeigt ebenfalls die variablen Bereiche im ETAS- und CSB-Bereich. Bezogen auf die taurine Sequenz finden sich die meisten Polymorphismen (rein optisch betrachtet) etwa zwischen den Positionen 15900-16143 und 163-363 (die Nukleotidposition beziehen sich auf die bovine Nomenklatur von (Anderson *et al.*, 1982)). Die 3' hypervariable Region (ca. 343 bp), die die Strukturen ETAS1 und 2 und den Anfang der Zentralen Domäne einschließt, enthält 65 % aller Substitutionen. Das entspricht dem Ergebnis aus Kapitel 4.1.2.5 Verteilung variabler Positionen im bubalinen Alignment.

#### 4.1.3.2 Sekundärstrukturen der D-loop Region

Brown und Kollegen haben in ihrer Arbeit konservierte Elemente der D-loop Region in verschiedenen Spezies verglichen (Brown *et al.*, 1986). So zeigen sie u.a., daß im 3'- und im 5' Bereich der D-loop Region Sekundärstrukturen zu finden sind, die Haarnadel- bzw. Kleeblattstrukturen ausbilden (Abbildung 4-10). Homolog zur bovinen Sequenz lassen sich diese Strukturen auch in der bubalinen Sequenz nachweisen.

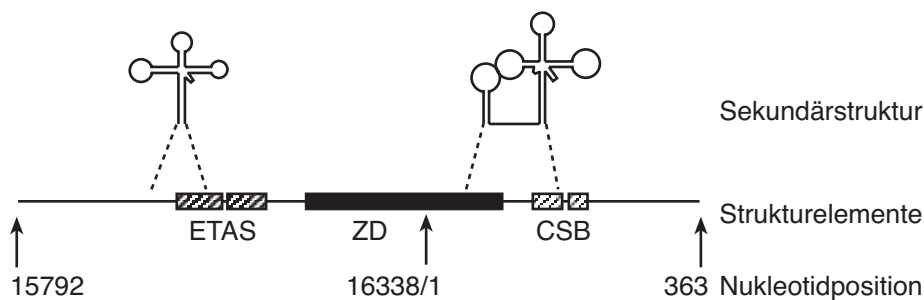


Abbildung 4-10 Schematische Darstellung der Sekundärstrukturen, die im 3'- und 5'- Bereich der D-loop Region ausgebildet werden. Es bilden sich sogenannte Kleeblatt- und Haarnadelstrukturen. Sie formieren sich in den Bereichen der ETAS1 bzw. der Zentralen Domäne (ZD) und im CSB1. Die Nukleotidpositionen beziehen sich auf die Nomenklatur von (Anderson *et al.*, 1982).

Im 3'-Bereich der D-loop Region bilden Teile der hypervariablen Region und der ETAS1 eine Kleeblattstruktur. Im 5'-Bereich führt ein längerer Sequenzabschnitt, bestehend aus dem Ende der zentralen Domäne und einem Großteil des CSB1 zu einer Haarnadel-, gefolgt von einer Kleeblattstruktur (Abbildung 4-10).

In den folgenden zwei Abbildungen (Abbildung 4-11 und Abbildung 4-12) sind diese Sekundärstrukturen in linearer Form dargestellt. Jeweils korrespondierende öffnende und schließende Klammern unterhalb der Sequenzen geben dabei die Basen an, die komplementär einen Stamm ausbilden. Die Basen dazwischen bilden einen sogenannten Loop. Im 3'-Bereich nimmt die Struktur die Form eines Kleeblatts an, hat also einen Hauptstamm (runde Klammern), an dem sich mehr oder weniger direkt drei Seitenstämme mit je einem Loop anfügen (vergleiche Abbildung 4-10). Diese Struktur wird teilweise durch die als ETAS1 bezeichnete Region gebildet. Die Sequenzunterschiede zwischen Rind und Büffel sind gering, es gibt eine hohe Homologie. Deutliche Unterschiede gibt es allerdings in der Komple-

mentarität des Hauptstamms. Auch beim Rind sind hier nicht alle Basen, die den Hauptstamm bilden komplementär. Die fünfte (T) und siebte (A) Base "paart" mit einem G bzw. A, illustriert durch unterstrichene Klammern in Abbildung 4-11. Der Hauptstamm hat beim Rind eine Länge von 8 Basen.

Bei der Konsensussequenz der Flußbüffel verkürzt sich dieser Stamm auf 7 Basen durch Deletion eines Thyminrestes. Dadurch kommt in diesem Stamm nur eine nicht komplementäre Paarung vor. Bei den Sumpfbüffeln fällt das weiter oben besprochene poly-G Motiv in den Hauptstamm. Die zehn Guaninreste finden nur vier komplementäre Cytosinreste und verkürzen so den Stamm. Der erste Seitenstamm (eckige Klammern) ist durch Substitution um eine Base verkürzt.

*Bos taurus*

GGGGTAATGTACATAACATTAATGTAATAAAGACATAATATGTATATAGTACATTAAATTATATGCCCC  
 (((((((( [ [ [ [ [ ] ] ] ] ] { { { } } } } \_ } } } } <<<<< >>>>> )))))))

Flußbüffel

GGGGAATGTACATAACATTAATGTAACAAGGACATAATATGTATATAGTACATTACATTATATGCCCC  
 (((((((( [ [ [ [ [ ] ] ] ] ] { { { } } } } \_ } } } } <<<<< >>>>> )))))))

Sumpfbüffel

GGGGGGGGGACATAACATTAATGTAATAAGGACATAATATGTATATAGTACATTATATTATATGCCCC  
 ((( [ [ [ [ [ ] ] ] ] ] { { { } } } } \_ } } } } <<<<< >>>>> )))))))

Abbildung 4-11 Vergleich der Sekundärstrukturen im 3' Bereich der D-loop Region in Rind, Fluß- und Sumpfbüffel.

Die Klammern markieren korrespondierende Basen, die einen "Stamm" bilden. Dazwischen befinden sich die sogenannten Loops. Die runden Klammern bezeichnen den Hauptstamm. Insgesamt ergibt sich eine Kleeblattstruktur. Die bovine Sequenz umfaßt die Nukleotide 15.958 bis 16.026. Unterstrichen sind nicht komplementäre Basen. In der bovinen Sequenz ist der Teil der ETAS1 eingekastet, der an der Ausbildung der Sekundärstruktur beteiligt ist.

Im 5'-Bereich ist die Homologie zwischen Rind, Fluß- und Sumpfbüffel gut erhalten (Abbildung 4-12). Nach einer Haarnadel folgt immer im Abstand von 9 Basen eine Kleeblattstruktur. In der Klammernotation werden die Hauptstämme durch runde Klammern dargestellt. Nicht korrespondierende Basen eines Stammes sind durch unterstreichen der Klammern gekennzeichnet. Diese Strukturen werden durch Bereiche der konservierten zentralen Domäne und dem CSB1 gebildet. Die Variabilitäten der Sequenzen sind hauptsächlich auf die "Loop"-Bereiche begrenzt. Nur wenige beeinflussen die Positionen, die die Stämme bilden.





#### 4.2.1 Modeltest

Die Sequenzalignments der Datensätze "37\_940" und "42\_159" wurden mittels Modeltest (Posada & Crandall, 1998) auf das optimale Modell getestet. Die Ergebnisse sind in der nachfolgenden Tabelle 8 aufgeführt. In beiden Fällen wurde das HKY85-Modell (Hasegawa *et al.*, 1985) mit  $\Gamma$  Korrektur (Yang, 1996) als ausgewähltes Modell berechnet.

Tabelle 8 Ergebnisse aus dem Programm Modeltest.

Die Tabelle gibt Aufschluß über das ausgewählte Modell, seine Wahrscheinlichkeit, Transitions/Transversionsrate, die Basenfrequenzen und den  $\Gamma$  Kurvenparameter  $\alpha$ .

Daten	Modell	Wahrscheinlichkeit $[-\ln L]$	Ti/Tv	Basenfrequenz	$\Gamma$ Kurvenparameter ( $\alpha$ )
37_940	HKY+ $\Gamma$	2462,75	7,51	A=0,3171 C=0,2560 G=0,1422 T=0,2846	0,1298
42_159	HKY+ $\Gamma$	692,99	6,80	A=0,2538 C=0,2847 G=0,1551 T=0,3064	0,1952

#### 4.2.2 Dendrogramme für 37\_940

##### 4.2.2.1 Maximum Likelihood

Bei dem Kriterium Maximum Likelihood (ML) handelt es sich um eine sogenannte diskrete Methode. Hier werden die Merkmale einer Sequenz miteinander verglichen und ein Baum konstruiert, der die vorliegenden Daten am wahrscheinlichsten erklärt. Aufgrund des umfangreichen Datensatzes (37 OTUs) ist eine exakte Suche nicht möglich. Aus diesem Grund muß man auf heuristische Methoden ausweichen. Da unter dem ML Kriterium alle Merkmale in die Kalkulation einbezogen werden, ist selbst diese Verfahrensweise sehr zeit- und rechenaufwendig.

Die Analyse erfolgte mit dem Programm PAUP\*. Aus der Datei 37\_940 wurde ein Taxon (*Bos taurus*) als Außengruppe definiert. Das Optimalitätskriterium wurde auf Maximum Likelihood gesetzt. Merkmalsfrequenz, Transitions/Transversionsverhältnis und Gammakurvenparameter, wurden definiert wie es im Programm Modeltest abgeschätzt wurde. Diese Einstellungen korrespondieren zum HKY85-Modell. Es wurde die Star Decomposition Methode mit zufälliger Lösung ursprünglicher Verknüpfungen gewählt.

Abbildung 4-13 zeigt für die oben beschriebene Einstellung einen möglichen ML-Baum mit einem ML-Wert von 2442,412. Dieser Wert liegt nahe an dem, der im Programm Modeltest kalkuliert wurde ( $-\ln L = 2462,75$ ). Das Dendrogramm zeigt zwei deutliche Verzweigungen. Ein Zweig besteht aus allen Carabaos und Mur481, wobei Mur481 deutlich separiert von

den Carabaos ist. Der andere Zweig beinhaltet alle Flußbüffel. Diese bilden z. T. kleinere Gruppierungen, die aber nicht unbedingt aus nur einer Rasse bestehen. Die Sequenzpolymorphismen sind also nicht rassenspezifisch. Dies zeigte sich schon bei der Untersuchung redundanter Haplotypen (siehe Tabelle 5).

#### 4.2.2.2 Maximum Parsimonie

Aus der Datei 37\_940 wurde ein Taxon (*Bos taurus*) als Außengruppe definiert. Das Optimalitätskriterium war als Maximum Parsimonie definiert. Dabei waren 747 Merkmale konstant und 118 variabel, von denen 75 Parsimonie-informativ waren.

Es wurden 100 Bootstrap Wiederholungen mit randomisierter Taxonaddition (je 10 Wiederholungen) durchgeführt. Die Verlagerung von Ästen (branch-swapping) folgte dem „tree-bisection-reconnection“ (TBR)–Algorithmus. Es wurde ein 50% majority-rule Konsensusbaum konstruiert. Die Gesamtlänge des Baumes betrug 249. (Abbildung 4-14)

Die Bootstrap-Werte demonstrieren die deutliche Aufspaltung in Sumpfbüffel (inklusive Mur481) und Flußbüffel; jeweils 100 %. Weitere Aufspaltungen sind der Abbildung zu entnehmen. Deutlich zeigt sich, daß innerhalb einer Gruppe verschiedene Rassen vorkommen. Eine Differenzierung nach Rassen anhand der D-loop Sequenz war also auch hier nicht möglich.

Die Analyse mittels der Star Decomposition Methode ist sehr schnell unter den Kriterium Maximum Parsimonie. In wenigen Sekunden wird ein möglicher Baum berechnet, so daß man diese Analyse mehrmals hintereinander durchführen kann. Wie schon eingangs erwähnt ist die Topologie eines Baumes im wesentlichen von der Reihenfolge der addierten Taxa abhängig; hier von der Reihenfolge des Lösen einzelner Taxa aus dem ursprünglichen Sterndiagramm. Für die Datei 37\_940 wurden so Kantenlängen zwischen 242 und 248 berechnet. Nur der Baum mit der sparsamsten Topologie (242) wurde gespeichert (Abbildung 4-15).

Ein Vergleich beider Dendrogramme (Abbildung 4-14 und Abbildung 4-15) ist auch ein Vergleich der verschiedenen Methoden. Hier lassen sich keine gravierenden Unterschiede feststellen. Die Star Decomposition Methode liefert jedoch keine Bootstrap-Werte.

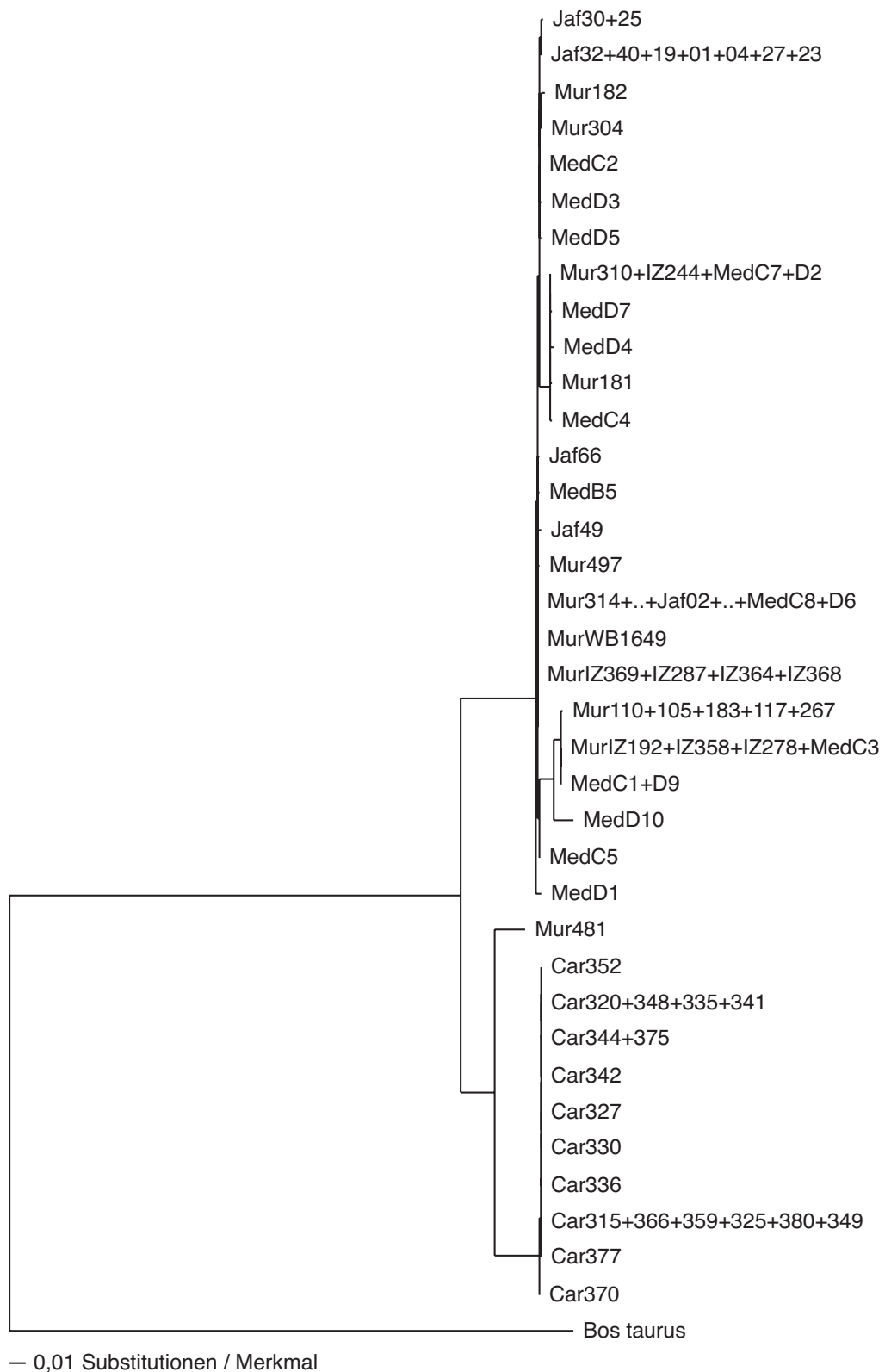


Abbildung 4-13 Möglicher ML-Baum für 37\_940. Unter Verwendung der heuristischen Star Decomposition Methode wurde dieser ML-Baum konstruiert. Die ursprünglichen Verknüpfungen wurden zufällig gelöst. Der ML Wert lag bei  $-\ln L = 2442,412$ . *B. taurus* war als Außengruppe definiert.

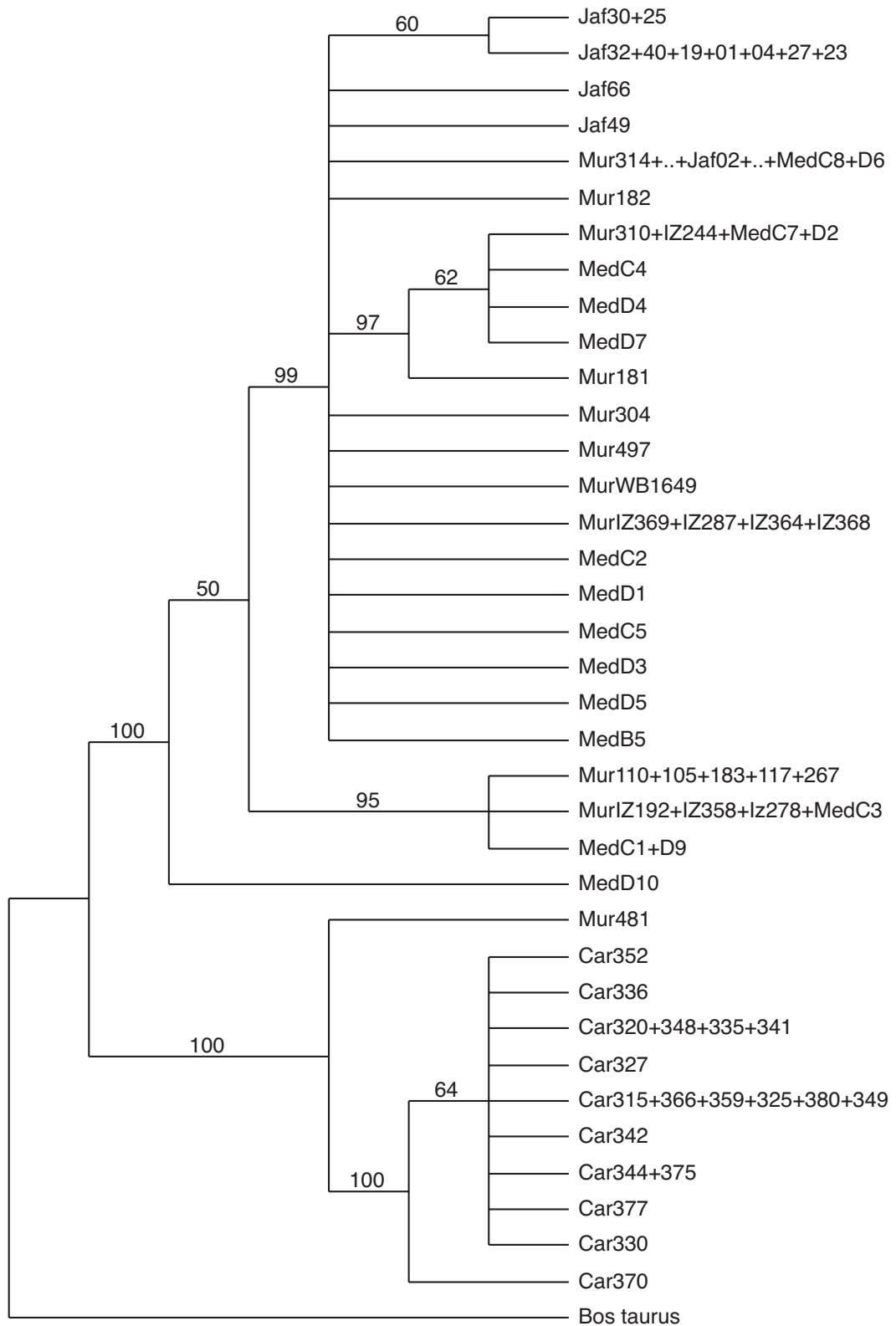


Abbildung 4-14 MP-Konsensusbaum für 37\_940. Unter Verwendung des TrN+Γ-Modells wurde dieser MP-Baum konstruiert. Zahlen oberhalb der Kanten sind Bootstrap Werte in Prozent für 100 Wiederholungen. Nur Werte größer oder gleich 50 % sind angezeigt. *Bos taurus* war als Außengruppe definiert. Die Gesamtbaumlänge betrug 249.

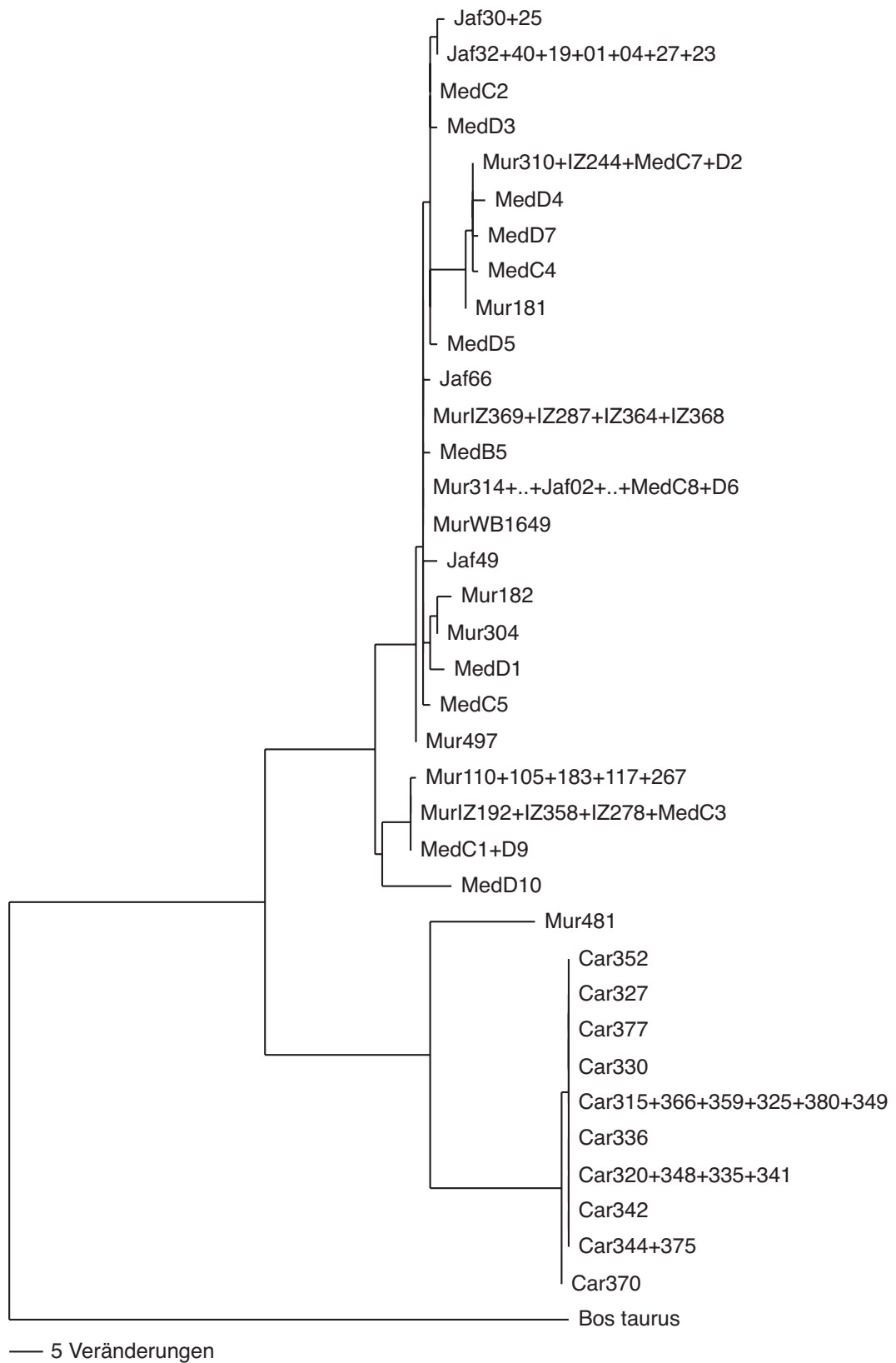


Abbildung 4-15 Möglicher MP-Baum für 37\_940.  
 Unter Verwendung der heuristischen Star Decomposition Methode wurde dieser Baum konstruiert. *B. taurus* war als Außengruppe definiert. Die Kantenlänge des Baumes wurde zu 242 berechnet.

#### 4.2.2.3 Genetische Distanz

Zur Berechnung eines NJ-Baumes wurden die Distanzparameter auf das TN+ $\Gamma$ -Modell gesetzt. Die Bootstrap Analyse wurde mit 1000 Wiederholungen durchgeführt. Die Berechnung erfolgte nach der NJ-Methode.

Auch hier zeigt sich das gleiche Bild wie bei den beiden anderen Methoden (ML und MP). Eine 100%ige Verzweigung separiert die Carabaos und Mur481 von den Flußbüffeln wie aus Abbildung 4-16 (Kladogramm) und Abbildung 4-17 (Phylogramm). Auch die Gruppierungen innerhalb der Flußbüffel

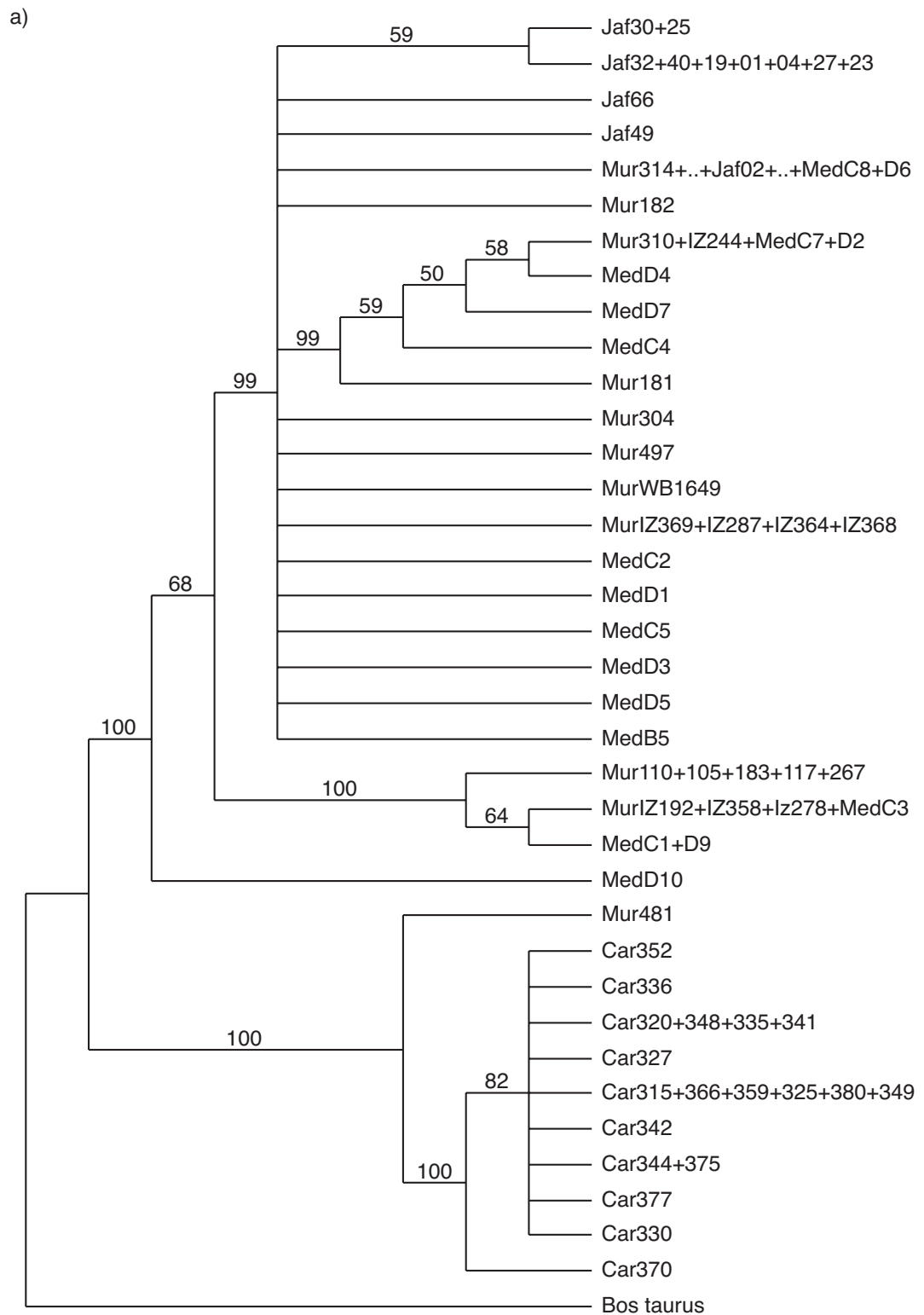


Abbildung 4-16 NJ Baum unter Verwendung des TN+Γ-Modells.  
 Zahlen oberhalb der Kanten sind Bootstrap Werte in Prozent für 1000 Wiederholungen.  
 Nur Werte über 50 sind angezeigt. *Bos taurus* war als Außengruppe definiert.

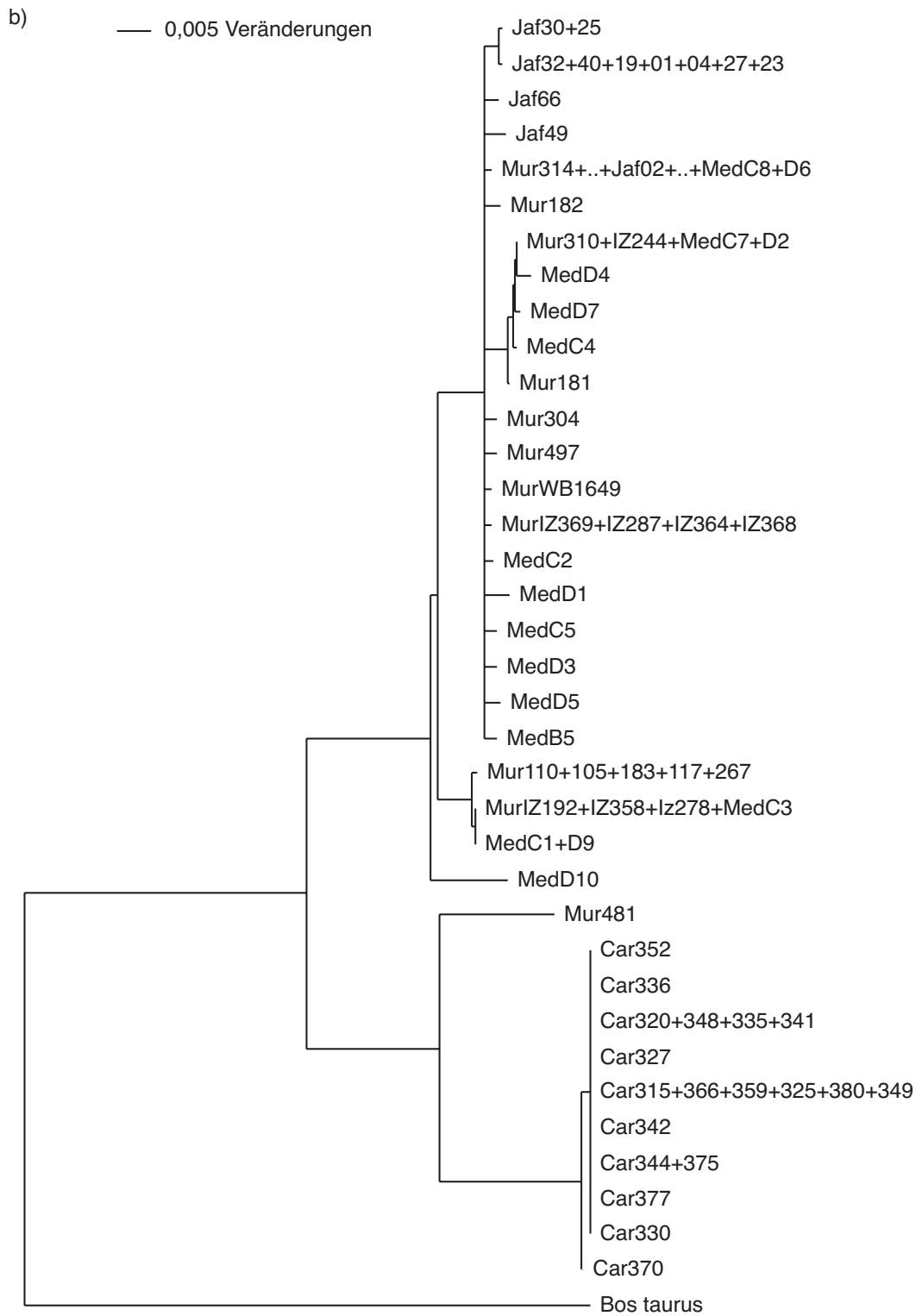


Abbildung 4-17 Darstellung des gleichen NJ-Baums wie in Abbildung 4-16, hier als Phylogramm.



### 4.2.3 Dendrogramme für 42\_159

Dies bisherigen Ergebnisse zeigten eine deutliche Separierung der Flußbüffel von den Carabaos und Mur481. Bei genauer Betrachtung der Sequenzunterschiede in der Gruppe der Carabaos zeigt sich jedoch eine geringe Diversität, die fast ausschließlich auf Längenpolymorphismen im poly-G und poly-C Motiven beruht. Ferner stammten alle diese Tiere von einer Farm und ihre Verwandtschaftsbeziehung war nicht bekannt. Ausgehend von der Hypothese, daß alle Carabaos einer maternalen Linie entstammen, war die Evidenz für Mur481 zu den Sumpfbüffeln zu gehören nicht sehr stark.

Zur Prüfung des Haplotyps Mur481, wurden alle Haplotypen dieser Arbeit mit veröffentlichten Daten verglichen, die in den Genbanken zur bubalinen D-loop Region hinterlegt sind (NCBI-Genbank: AF016397; (Lau *et al.*, 1998)). Aus dieser Sequenz wurde mit Hilfe der zugehörigen Veröffentlichung alle Haplotypen rekonstruiert und mit den hier vorliegenden Daten verglichen und phylogenetisch analysiert.

In der Arbeit von Lau und Kollegen wurden 80 Tiere (Sumpfbüffel und Murrah) aus 11 Populationen (Asien und Australien) in eine phylogenetische Untersuchung einbezogen. Neben der Analyse einer partiellen Sequenz des Cytochrom C Gens wurden 158 bp aus dem hypervariablen Bereich der D-loop Region sequenziert und analysiert. Die Sequenz korrespondierte zu den bovinen Nukleotidpositionen 16089 - 16245. Das entspricht dem Ende der ETAS2 bis hinein in die zentrale Domäne.

Durch die Reduktion der Sequenzinformation auf 158 bp der eigenen Daten ergaben sich weitere redundante Haplotypen im eigenen Datensatz. Von den 36 Haplotypen sind nunmehr sieben übrig geblieben, zwei (MedD10 und Mur481) waren einzig, die anderen fünf sind in Tabelle 9 aufgeführt. Zum Teil sind einige Haplotypen aus der Arbeit von Lau und Kollegen identisch mit Haplotypen dieser Arbeit, bezogen auf die untersuchte Region von 158 bp. Das gesamte Sequenzalignment, inklusive der Außengruppe *B. taurus*, umfaßte nun 42 OTUs und 159 bp aufgrund einer Insertion.

Tabelle 9 Auflistung redundanter Haplotypen.

Haplotypen aus der eigenen Arbeit und ihre redundanten Haplotypen bei Betrachtung eines 159 bp langen Abschnitts, z. T. aus der Arbeit von (Lau *et al.*, 1998). Diese sind unterstrichen. MedD10 und Mur481 sind einzig.

Haplotyp	Redundante Sequenzen
Jaf30	Jaf32, Jaf66, Jaf49, Mur314, Mur304, Mur497, MurlZ369, MedC2, MedC5, MedD3, MedD5, MedB5, <u>Aust2</u>
Mur110	MurlZ192, MedC1, <u>Aust1</u>
Mur310	MedC4, MedD4, MedD7,
Mur181	<u>Thai1</u>
Car352	Car370, Car330, Car377, Car344, Car342, Car315, Car327, Car320, Car336, <u>Phil2</u>

#### 4.2.3.1 Maximum Parsimonie

Für den Datensatz 42\_159 ergaben sich unter dem Parsimoniekriterium 47 variable Positionen, von denen 28 Parsimonie informativ sind.

Es wurde die schnelle heuristische Star Decomposition Methode gewählt. Auch hier wurden wieder mehrere Analysen durchgeführt (es ergaben sich Kantenlängen zwischen 100 und 107) und die sparsamste Topologie gespeichert (Abbildung 4-18).

Der MP-Baum zeigt für diese Untersuchung, wie für die Analysen der Datei 37\_940, eine deutliche Spaltung in zwei Gruppen. Die eine enthält die eigenen Haplotypen Mur481 und Car315 (Sumpfbüffel), die andere die Haplotypen Jaf01, MedD1, Mur181, Mur310, MurIZ192 und MedD10. Auch in dieser Datei, wie in 37\_940, findet sich keine populations- bzw. rassenspezifische Gruppierung, erkennbar an der Nomenklatur der Lau-Daten.

Die Dendrogramm-Topologie dieser Abbildung gleicht der aus der Lau Publikation, natürlich mit zusätzlicher Addition der sieben brasilianischen OTUs. Lau und Kollegen bemerken in ihrer Arbeit, daß die phänotypische Unterscheidung zwischen Fluß- und Sumpfbüffel nicht mit der genotypischen konform geht. In den 33 Haplotypen dieser Arbeit, gibt es zwei Haplotypen (Aust1 und Aust2) die sowohl in Sumpf- als auch in Flußbüffel gefunden wurden. Proben von Flußbüffel stammten lediglich aus den Population von Sri Lanka und Malaysia (SriL und Mal). Alle anderen Haplotypen sind Sumpfbüffel (Phil, Thai, Aust, Sula, Sabah, Bogor).

Nach Aussage von Lau und Kollegen läßt sich somit keine Differenzierung zwischen Fluß- und Sumpfbüffel treffen, allein anhand der D-loop Sequenz. Es fällt aber auf, daß im unteren Zweig tatsächlich nur Sumpfbüffel vertreten sind; ebenfalls die brasilianischen Haplotypen Mur481 und Car315. Im oberen Zweig finden sich nur brasilianische Haplotypen (Jaf, Med, Mur) in "phylogenetischer Nähe" zu Fluß-, aber auch Sumpfbüffel aus Asien und Australien (Abbildung 4-18).

#### 4.2.3.2 Minimum Evolution

Für eine weitere Untersuchung der Datei 42\_159 wurde das Optimalitätskriterium Distanz (Minimum Evolution) gewählt. Eine heuristische Suche mit 100 Bootstrap Wiederholungen unter dem TrN-Modell wurde gestartet. Der  $\Gamma$ -Kurvenparameter wurde mit  $\alpha = 0,1952$  definiert. Alle anderen Einstellungen entsprachen den Standardvoreinstellungen des Programms.

Die Suche wurde allerdings nach 14 Bootstrapwiederholungen (und fast 45 Stunden Rechenzeit) abgebrochen und ein 50% Majority-rule Konsensusbaum konstruiert (Abbildung 4-19).

Im Gegensatz zur MP Analyse ist in diesem Fall keine deutliche Separierung von Haplotypen erkennbar. Fast alle Haplotypen inserieren an einer internen Kante. Es gibt allerdings zwei Gruppen, die sich durch deutlich verlängerte Kanten auszeichnen. Eine Gruppe, zu der auch Mur481 zählt und die andere Gruppe mit Car315. Diese beiden Gruppen finden sich auch als Teil des unteren Zweiges in Abbildung 4-18. Man beachte, daß dieser Baum einen Konsensus aus nur 14 Bootstrapreplikaten darstellt.

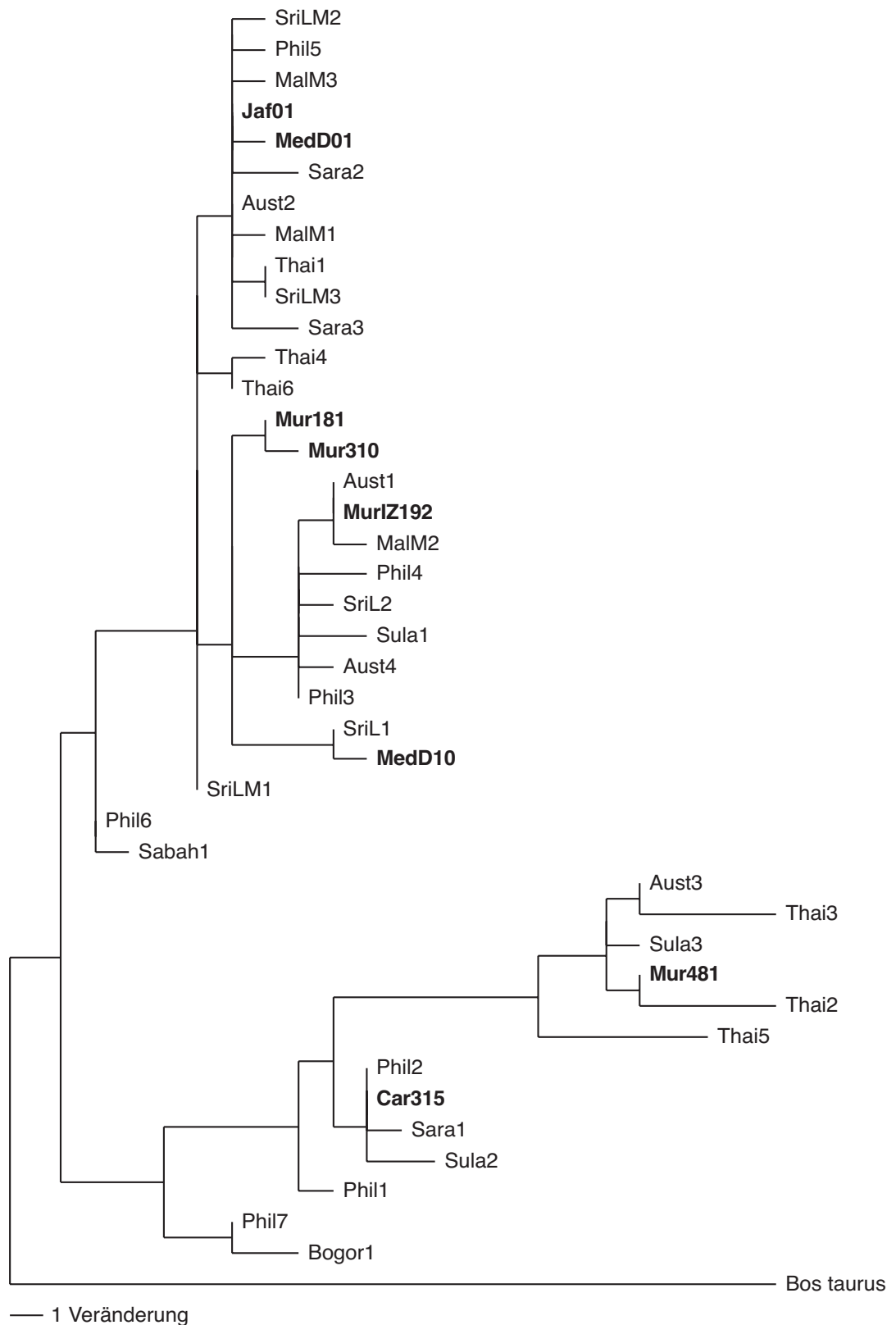


Abbildung 4-18 Möglicher MP-Baum für 42\_159. Unter Verwendung der heuristischen Star Decomposition Methode. *B. taurus* war als Außengruppe definiert. Die Haplotypen aus der eigenen Arbeit sind fett gedruckt. Die übrigen Haplotypen stammen aus (Lau *et al.*, 1998). Der Baum hat eine Kantenlänge von 100.

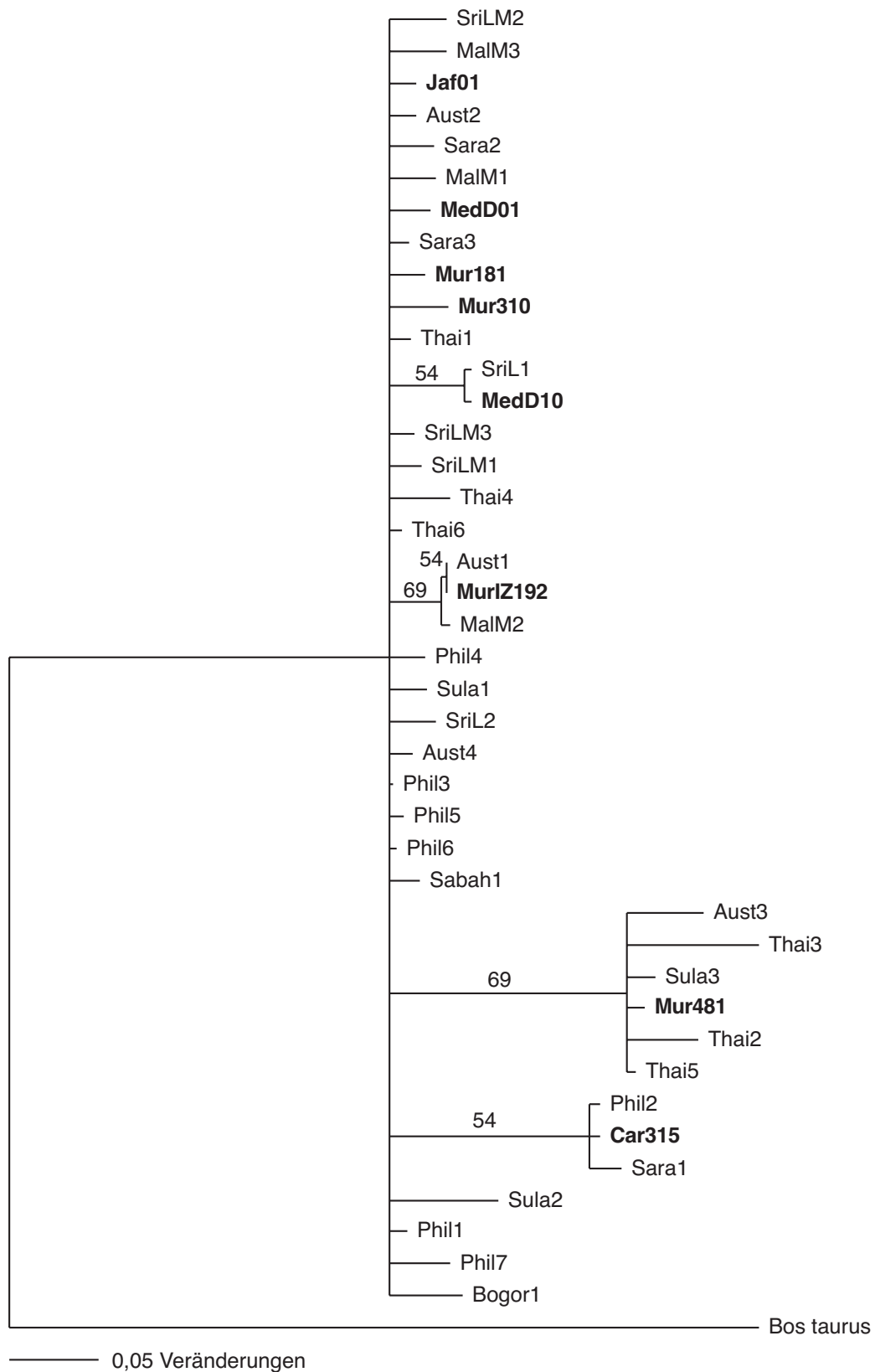


Abbildung 4-19 ME-Konsensusbaum für die Datei 42\_159. Nur Bootstrapwerte über 50 % sind vermerkt. Der ME-Wert für diesen Baum war 1,72479. Die Gesamtkantenlänge unter dem Parsimoniekriterium betrug 161. Die 8 Haplotypen aus der eigenen Arbeit sind fett gedruckt, alle anderen Haplotypen stammen aus (Lau *et al.*, 1998). Die Suche wurde nach 14 Bootstrapwiederholungen abgebrochen.

#### 4.2.4 Netzwerkdarstellung

Die graphische Darstellung von Netzwerken ermöglicht es inkompatible Merkmale darzustellen (siehe hierzu Kapitel 2.2.4 Dendrogramme versus Split-Zerlegung). Da die Software (Network 2.0c), die die Splitzerlegung kalkuliert, maximale Sequenzlängen von 500 Merkmalen bearbeiten kann und konstante Positionen keine Auswirkung auf die Splitzerlegung haben, wurden nur die variablen Positionen eingegeben.

Die graphische Darstellung der Splitzerlegung ist in Abbildung 4-20 wiedergegeben. Der obere Teil (a) gibt eine Übersicht über das gesamte Netzwerk. Der Außengruppe *Bos taurus* am nächsten finden sich alle Flußbüffel (Kasten). In großer Entfernung dazu sind die Carabaos und der Haplotyp Mur481 angesiedelt. Die Vernetzung der Flußbüffel untereinander ist in Abbildung 4-20b vergrößert dargestellt. Der Haplotyp MedD10 und die beiden Haplotypen Mur110 und MurlZ129 sind relativ weit von den anderen Haplotypen getrennt. Ihre separate Lage ist auch in den Baumdiagrammen (vergleiche z. B. Abbildung 4-15, ein MP-Baum) zum Ausdruck gekommen. Auch andere Gruppierungen sind parallel zum MP-Baum erkennbar, wie {Jaf30, Jaf32} oder {Mur181{Mur310, MedD7, MedC4, MedD4}}. Viel besser visualisiert im Netzdiagramm sind die anderen Beziehungen untereinander. Insbesondere sind inkompatible Verwandtschaftsverhältnisse darstellbar. So hat der Haplotyp MedD1 sowohl Ähnlichkeit zu Mur314 als auch zu Mur304. Diese Beziehung ist in keinem Dendrogramm zu finden. Die beiden Haplotypen MedC2 und Mur314 nehmen eine "zentrale" Rolle ein. Sie haben mehrere direkte und indirekte Verbindungen zu anderen Haplotypen. Dabei fällt auf, daß Mur314 (selbst ein Haplotyp, der in drei Rassen vorkommt) in direkter Nähe aller drei Flußbüffelryassen liegt.

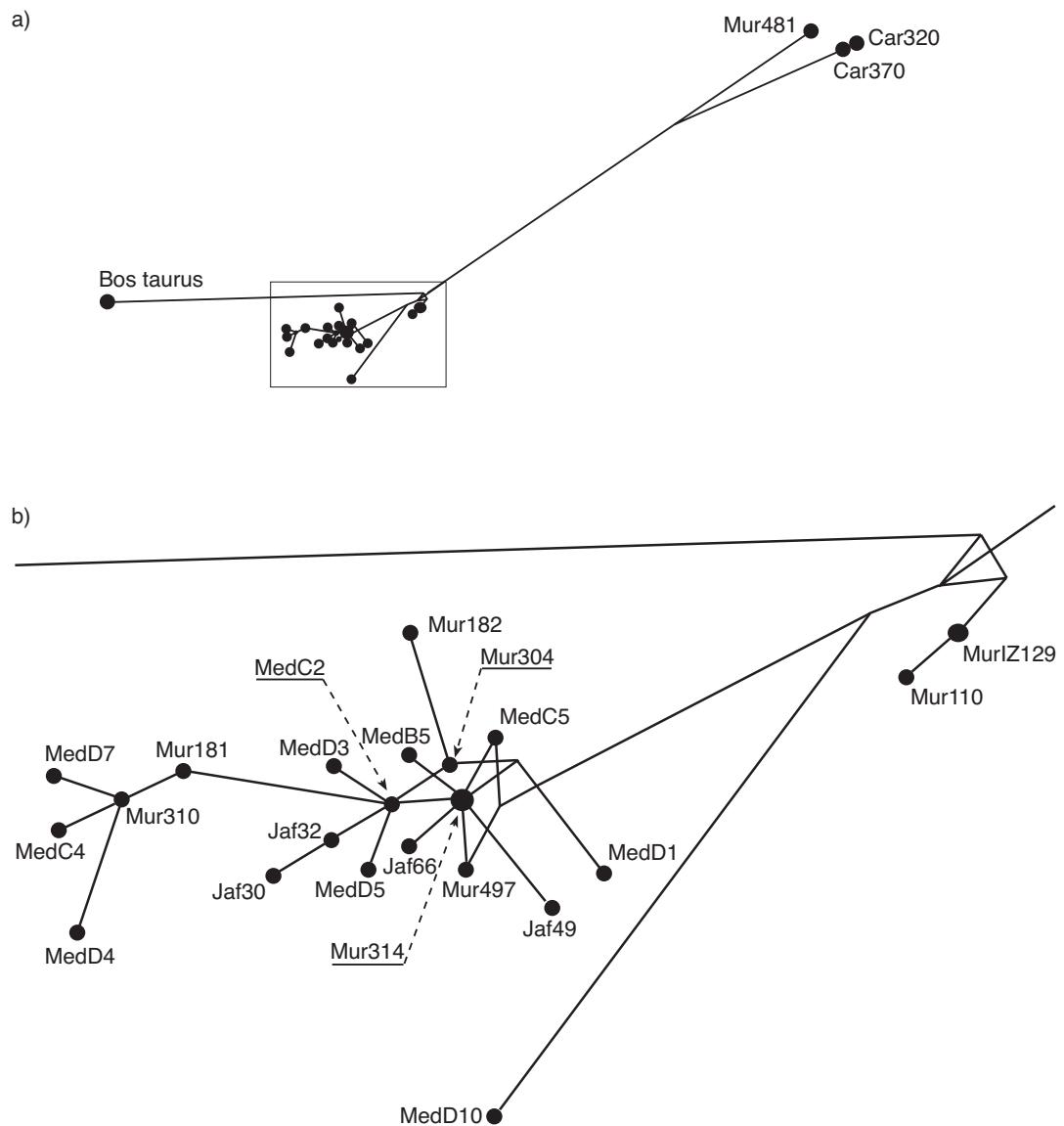


Abbildung 4-20 Netzwerkdiagramm.  
 Das Netzwerkdiagramm zeigt die Verwandtschaft zwischen den untersuchten Büffeln aufgrund der variablen Positionen innerhalb der D-loop Region. In a) gesamtes Netzwerk, dessen zentraler Bereich (Kasten) in b) vergrößert dargestellt ist.

Es fällt auf, daß einige Haplotypen im Netzdiagramm nicht berücksichtigt sind (MurWB1649, MurIZ369, MedC1 und verschiedene Carabao). Diese fanden in der Splitzerlegung keine Berücksichtigung, da sie sich von anderen Haplotypen nur durch Del/Ins Ereignisse unterscheiden. Diese sind im Programm Network nicht vorgesehen. Die Sequenz von MedC1 ist daher gleich wie die von MurIZ192, die anderen beiden Murrachs sind gleich wie Mur314. Die nicht aufgeführten Carabaos sind dem Haplotypen Car320 zugerechnet.

#### 4.2.5 Divergenz der Genera *Bos* und *Bubalus*

Die genetischen Distanzen wurden in PAUP\* berechnet (Anhang 4). Die nachfolgende Tabelle 10 gibt die unkorrigierten Werte als durchschnittliche Distanz zwischen den einzelnen Rassen wieder.

Tabelle 10 Berechnung der durchschnittlichen unkorrigierten Distanzen. Berechnet wurden die durchschnittlichen unkorrigierten Distanzen ( $\pm$  Standardabweichung) zwischen Flußbüffel, Sumpfbüffel (inkl. Mur481) und der Außengruppe *Bos taurus*.

Unkorrigiert („p“)	Flußbüffel	Sumpfbüffel
Flußbüffel	0,0085 $\pm$ 0,0063	
Sumpfbüffel	0,0691 $\pm$ 0,0024	0,0073 $\pm$ 0,0153
<i>Bos taurus</i>	0,1535 $\pm$ 0,0028	0,1711 $\pm$ 0,0009

Die Werte demonstrieren, daß die Diversität innerhalb der Flußbüffel recht gering sind (0,8 %), jedoch mit einer hohen Standardabweichung, da viele Tiere im untersuchten Bereich keine Unterschiede aufweisen. Erwartungsgemäß ist die Diversität innerhalb der Sumpfbüffel ebenfalls gering. Wie schon weiter oben bemerkt, handelt es bei den Carabao Haplotypen möglicherweise um nur eine oder zwei maternale Linien. Die Diversität zwischen Sumpf- und Flußbüffel liegt bei etwa 7 % (unkorrigiert), zwischen Büffel und Rind bei 15 % bis 17 %.

Die Zeit seit der Divergenz  $t(D)$  zweier Linien wurde nach folgender Formel berechnet:

$$t(D) = \frac{d}{2k} \quad (16)$$

Dabei ist  $k$  die Evolutionsrate in Substitutionen je Nukleotidposition und Jahr und  $d$  die durchschnittliche paarweise Distanz zwischen zwei Sequenzgruppen.

Für die vorliegende Arbeit wurde eine Evolutionsrate  $k = 7,5 \cdot 10^{-8}$  für die mitochondriale D-loop Region angenommen (Tamura & Nei, 1993). Entsprechend wurden die paarweisen Distanzen zwischen den Sequenzen nach dem Modell von Tamura und Nei (Tamura & Nei, 1993) unter Berücksichtigung des  $\Gamma$ -Parameters berechnet (Anhang 5). Zusammenfassend sind die durchschnittlichen korrigierten Distanzen in Tabelle 11 aufgeführt.

Tabelle 11 Berechnung der durchschnittlichen korrigierten Distanzen. Berechnet wurden die durchschnittlichen Distanzen ( $\pm$  Standardabweichung) zwischen den Büffelryassen Jafarabadi, Murrah, Mediterraneo, Carabao und der Außengruppe Rind. Der Berechnung lag das TrN-Modell mit  $\Gamma$  Verteilung zugrunde.

TrN+ $\Gamma$	Flußbüffel	Sumpfbüffel
Flußbüffel	0,0096 $\pm$ 0,0077	
Sumpfbüffel	0,1370 $\pm$ 0,0077	0,0107 $\pm$ 0,0225
<i>Bos taurus</i>	0,6090 $\pm$ 0,0361	0,9290 $\pm$ 0,0174

Der Vergleich mit Tabelle 10 zeigt deutlich den Effekt der durch die Korrektur unter einem bestimmten Modell erzielt wird. So ist die durchschnittliche Distanz innerhalb der Fluß-, bzw. Sumpfbüffel kaum erhöht. Mit der Korrektur steigt aber schon unter dem TrN-Modell die Divergenz zwischen Fluß- und Sumpfbüffel auf über 13 % (unkorrigiert: 8 %) an. Drastisch ist die Divergenz die sich dann zu *Bos taurus* ergibt; 60 % bzw. 90 % gegenüber Fluß- und Sumpfbüffel.

Mit den Werten aus Tabelle 11 und Gleichung (16) ergibt sich für die Divergenzabschätzung eine Trennung von Fluß und Sumpfbüffeln vor gut 913.226 Jahren mit einer Standardabweichung von etwa 51.000 Jahren. Die Trennung von *Bubalus* und *Bos* errechnet sich zu über 5,1 Mio. Jahren vor heute ( $\pm 178.276$  Jahren).



## **5 Diskussion**

### **5.1 Erstmalige Untersuchung zur Diversität brasilianischer Wasserbüffel**

Die vorliegende Arbeit stellt die Ergebnisse aus einer erstmaligen Analyse zur Phylogenese brasilianischer Wasserbüffel vor. Die Studie sollte Aufschluß über die Diversität der dort genutzten Sumpfbüffel (Carabao) und Flußbüffel (Jafarabadi, Murrah und Mediterraneo) geben. Auf verschiedenen Farmen wurden zur Analyse Blutproben von Büffeln genommen. Die Verwandtschaft unter den Tieren war dabei nicht bekannt, da keine Herdbücher geführt wurden. Das Führen von Herdbüchern und der Nachweis von Abstammungskontrollen ist in Brasilien noch nicht etabliert und wird bisher nur von einigen großen Farmen, speziell im Süden Brasiliens durchgeführt. Die meisten Farmen zeichnen sich durch eine extensive Büffelhaltung aus, so daß Anpaarungen schlecht kontrollierbar sind.

Da in Brasilien kein Zugang zu Mediterraneo Proben bestand, wurden Proben dieser Rasse aus Italien in die Analyse einbezogen.

Zum ersten mal wurde für die phylogenetische Analyse dieser Spezies die komplette D-loop Region sequenziert und analysiert.

### **5.2 Fluß- und Sumpfbüffel lassen sich deutlich anhand der polymorphen D-loop Region unterscheiden**

Die Sequenzanalyse der mitochondrialen D-loop Region zeigte zum einen, daß die Diversität innerhalb der Rassen stark schwanken kann. Zum anderen ist die genetische Distanz zwischen Fluß- und Sumpfbüffeln sehr hoch und separiert beide Typen deutlich.

#### **5.2.1 Die Diversität der Rassen ist in den Populationen verschieden**

Die Diversität innerhalb der einzelnen Rassen war unterschiedlich groß. So zeigten die Jafarabadis bezüglich der D-loop Region eine nur geringe genetische Variabilität. In 20 untersuchten Proben fanden sich nur 4 Haplotypen. Im Gegensatz dazu zeigten die Populationen von Murrah und Mediterraneo eine deutlich höhere Diversität (14 Mediterraneo Haplotypen, 11 Murrah Haplotypen). Auch die Sequenzen der Carabaos wiesen eine ge-

ringe Variabilität auf. Die Sequenzunterschiede beruhten hier hauptsächlich auf Längenpolymorphismen im 5'-poly-G Motiv und in den 3'-poly-C Motiven.

Der Unterschied in der Variabilität kann mit dem z. T. begrenzten Zugang zu den genetischen Ressourcen erklärt werden. Der Umkehrschluß ist die höhere Diversität von Murrah und Mediterraneo, die von verschiedenen Populationen (Farmen) kamen, während für die Jafarabadis und Carabaos nur Proben von jeweils einer Farm zu Verfügung standen. Keine der Farmen führte Herdbücher bzw. Abstammungskontrollen durch. Deshalb waren die Verwandtschaftsverhältnisse der Tiere nicht bekannt. Erst die hier vorliegende Untersuchung zeigte, daß einige Rassen weniger divers als andere waren. Die geringe Diversität der Jafarabadis läßt darauf schließen, daß die vier gefundenen Haplotypen vier maternalen Linien entstammen. Dies würde bedeuten, daß die Farm möglicherweise mit vier Büffelkühen begonnen hatte. Um also zu prüfen, ob es sich tatsächlich um einen sogenannten Gründereffekt handelt, oder ob die Jafarabadis allgemein eine geringe Diversität aufweisen, müßten ergänzende Untersuchungen an weiteren Jafarabadipopulationen durchgeführt werden.

Eine ähnliche Beobachtung findet man bei den untersuchten Carabaos. Auch diese Proben stammen von einer einzigen Farm, so daß man hier ebenfalls die Auswirkungen eines möglichen Gründereffekts vorfindet. Speziell fiel bei allen Carabaos das poly-G Motiv auf (ab Position 179 im bubalinen Alignment), welches charakteristisch für alle Carabaos war, aber bei sonst keinem Haplotypen gefunden wurde; auch nicht bei Mur481, der als potentieller Sumpfbüffel klassifiziert wurde. In den Nukleotid-Datenbanken findet man ebenfalls keine ähnlichen Sequenzen zu dem poly-G Motiv in anderen Spezies. Interessant wäre die Klärung, ob es sich bei diesem poly-G Motiv um ein Charakteristikum der Carabaos handelt, oder ob dies einen Hinweis auf *eine* maternale Linie ist. Die Möglichkeit einer allgemein niedrigen Diversität in diesen Rassen kann nur durch weitere Evaluierungen der Rassen geschehen. Proben aus verschiedenen Populationen (möglichst aus vielen geographischen Regionen) müssen genommen und analysiert werden.

Dennoch, diese erste Analyse brasilianischer Büffel zeigt deutlich, daß die Jafarabadi Population der Fazenda Itaquí (siehe Tabelle 1) eine geringe Diversität aufweist. Das gleiche gilt für die Fazenda São Luís, deren Carabaos eine geringe Diversität zeigten. Die Murrahs dieser Farm hingegen schienen deutlich diverser: innerhalb der sieben untersuchten Tiere konnten vier Haplotypen ausgemacht werden. Mit der Untersuchung der D-loop Region können Farmen auf Diversität der maternalen Linien durchmustert werden. Informationen über die Diversität einer Population sind Voraussetzung für die Entwicklung effizienter Zuchtprogramme.

### 5.2.2 Ein Haplotyp in drei Rassen

Ein interessanter Aspekt, der bei der Suche nach redundanten Sequenzen zum Vorschein kam, war die Tatsache, daß *ein* mitochondrialer Haplotyp in den *drei* untersuchten Flußbüffelfrassen auftrat. Der Mur314 Haplotyp fand sich nicht nur in weiteren Murrahtieren, sondern auch in einer Vielzahl von Jafarabadis und einigen italienischen Mediterraneo. Das

Netzdiagramm (Abbildung 4-20) stellt den Haplotypen Mur314 zusätzlich auch optisch in den Mittelpunkt; sechs weitere Haplotypen (auch aus allen untersuchten Flußbüffelrassen) sind direkt und zwei (Mur304, MedD1) über einen Knoten mit Mur314 verbunden. Damit könnte es sich bei diesem Haplotypen um eine ancestrale Sequenz handeln, die entstand noch bevor die Rassen sich trennten. Diese Hypothese schließt nicht aus, daß dieser Haplotyp erst in historischer Zeit durch Züchtungskreuzung in die verschiedenen Rassen eingebracht wurde.

Ansatzpunkte zukünftiger Arbeiten sollen prüfen, ob dieser Haplotyp auch in anderen Flußbüffelrassen (es gibt noch 15 weitere) zu finden ist. Hierzu müßten Proben insbesondere aus asiatischen Population gesammelt und untersucht werden. Notwendig erscheint besonders eine molekulargenetische Untersuchung am wilden Wasserbüffel. Dieser gilt als die Stammart der heutigen domestizierten Wasserbüffel. Ein Vergleich der D-loop Region von domestizierten und wilden Wasserbüffeln würde ein interessantes Ergebnis über die Diversität der Wildtiere und ihre Verwandtschaft zu den domestizierten Tieren liefern.

### 5.2.3 Phylogenie der Wasserbüffel

Alle Analysen zeigen das einheitliche Bild einer fast "sternförmigen" Phylogenie. Gerade bei der MP-Analyse zeigt der Konsensusbaum (Abbildung 4-14) bei dem viele Individuen an einem internen Knoten inserieren (hier dargestellt durch polytome Verzweigung). Solche "flachen" Phylogenien sind Anzeichen einer "Flaschenhals"-Population bzw. Anzeichen einer kleinen effektiven Populationsgröße.

Im Gegensatz dazu weist das Dendrogramm in Abbildung 4-18 (kombiniert mit Daten von (Lau *et al.*, 1998)) tiefere Verzweigungen auf. Ein Hinweis auf eine größere effektive Populationsgröße oder längere genetische Distanz. Da es sich allerdings um eine Spezies handelt, kann man letzteres ausschließen. Viel wahrscheinlicher ist die größere effektive Populationsgröße in den Daten von Lau und Kollegen (80 Tiere aus 11 Populationen). Die Proben der vorliegenden Arbeit stammen hauptsächlich von fünf verschiedenen Farmen in Brasilien. Es ist nicht bekannt wieviel Tiere nach Brasilien exportiert wurden. Der Anteil weiblicher Büffel definiert damit auch den Anteil maternaler Linien, die ohne Genfluß auch nicht verändert werden.

### 5.2.4 Was einen Flußbüffel zum Sumpfbüffel macht

Für die phylogenetische Analyse auf Ebene der mtDNA ergibt sich eine Problematik die auch in der hier vorliegenden Arbeit aufgetaucht ist: ein Phänotyp (hier Mur481) kann einen unerwarteten Genotypen aufweisen, da die mtDNA maternal vererbt wird. In der Züchtung ist die Kreuzung zwischen verschiedenen Rassen üblich. Sumpf- und Wasserbüffel lassen sich untereinander kreuzen und bringen auch fertile Nachkommen hervor, obwohl sie einen unterschiedlichen Chromosomensatz haben (Di Berardino & Iannuzzi, 1981; Gupta & Srivastava, 1984; Harisah *et al.*, 1989).

Der Haplotyp Mur481 ist phänotypisch ein Murrah, also ein Flußbüffel. Nun hat die genotypische Analyse der mtDNA deutlich gezeigt, daß diese Tier in der maternalen Linie auf einen Sumpfbüffel zurückgeht. Da Kreuzungen zwischen beiden Büffeltypen durchaus üblich sind, ist es sogar erstaunlich, daß in dem großen Probensatz kein weiterer Fall einer solchen "maternalen Vergangenheit" aufgetreten ist.

Wie für die anderen Tiere existieren keine Abstammungsunterlagen für diesen Mur481. Damit ist eine genaue Verifizierung der Hypothese nur schwer möglich. Selbst die Option das Individuum karyotypisch als Fluß- oder Sumpfbüffel zu verifizieren wird unmöglich, wenn es sich nicht um einen F1 Hybriden handelt. Eine Kreuzung eines Flußbüffels ( $2n=50$ ) mit einem Sumpfbüffel ( $2n=48$ ) bringt Hybriden mit  $2n=49$  Chromosomen hervor. Aber schon die F2 Generation zeigt alle drei möglichen Karyotypen ( $2n=48$ ,  $2n=49$  und  $2n=50$ ), so daß eine sichere Identifizierung nicht mehr möglich ist, wenn man nur ein einzelnes Tier untersucht (Harisah *et al.*, 1989).

In diesem Zusammenhang sind die Ergebnisse von (Lau *et al.*, 1998) interessant. Sie beschreiben in ihrer Arbeit das Vorhandensein eines Haplotypen (auf Grundlage eines 159 bp Fragments der D-loop Region) sowohl im Fluß- als auch im Sumpfbüffel. Dieses Ergebnis kommt auch in Abbildung 4-18 zum Ausdruck. Deutlich weist der MP-Baum zwei Gruppen aus, die eine, die die Flußbüffel aus der vorliegenden Arbeit enthält und die andere, die die Sumpfbüffel (Carabao und Mur481) enthält. Der untere Zweig definiert die Sumpfbüffel, nicht nur aufgrund der mitochondrialen Sequenz, sondern auch phänotypisch (Ausnahme Mur481). Der obere Zweig hingegen weist die Tiere genotypisch zu einer Gruppe (Flußbüffel), phänotypisch stammen die Tiere jedoch sowohl aus Sumpf- als auch aus Flußbüffelpopulationen (Lau *et al.*, 1998).

Diese Ergebnis kann so interpretiert werden, daß eine Unterscheidung zwischen Fluß- und Sumpfbüffel auf mitochondrialer Ebene nicht möglich ist. Was in gewisser Weise auch richtig ist. Hier muß man ganz klar unterscheiden zwischen den maternalen Erbgang der Mitochondrien und dem rekombinanten Erbgang, der den Phänotypen hauptsächlich bestimmt.

Die hier vorliegende Arbeit zeigt (und andere publizierten Arbeiten auch), daß die Divergenz zwischen beiden Büffeltypen sehr groß ist. Daher ist es unwahrscheinlich, daß ein mitochondrialer Haplotyp in beiden Typen vorkommt oder am phylogenetischen Dendrogramm direkt benachbart sind.

### **5.3 Fluß- und Sumpfbüffel trennten sich lange vor der Domestikation**

Zur Abschätzung der Divergenz beider Büffeltypen wurde die paarweise Sequenzdivergenz nach Tamura-Nei korrigiert. Die korrigierten Werte lieferten eine durchschnittliche Divergenz von Fluß- und Sumpfbüffel von etwa 13 %. Daraus berechnete sich unter der Annahme einer Evolutionsrate von  $k = 7,5 \cdot 10^{-8}$  Substitutionen je Nukleotidposition und Jahr (Tamura & Nei, 1993) eine Divergenz von Fluß- und Sumpfbüffel vor über 900.000 Jahren;

eine Zeitspanne, die weit über den Beginn der Domestikation (vor 10.000 Jahren) hinausgeht.

Diese Divergenz kann durch zwei Domestikationsprozesse erklärt werden. Beide Typen bevorzugten unterschiedliche Habitate. Während der Sumpfbüffel jede Form des Wasserlochs bis zum Schlammloch annimmt, suhlt der Flußbüffel nur in klarem Wasser. Die Annahme, daß ihre Vorfahren ebenfalls unterschiedliche Habitate bewohnten liegt nahe. So kann man sich vorstellen, daß die Domestikation beider Tiere in unterschiedlichen Regionen stattgefunden haben kann. So vermuten Cockrill und Nachtsheim eine Domestikation vor 5.000 Jahren in den Zivilisationen am Indus, Yangtze und Euphrat / Tigris Ebene (Cockrill, 1981; Nachtsheim & Stengel, 1977). Chen und Li geben eine Domestikation des Sumpfbüffels in China vor 9.000 Jahren an (Chen & Li, 1989). Es ist nicht auszuschließen, daß nachfolgend einem Domestikationsprozeß Einkreuzungen von verwandten Wildtieren stattfanden. Im Zuge von Migrationen, bei denen Tierzüchter mit ihren Tieren in andere Gebiete wanderten, könnten sie die dort lebenden Wildtiere zur Einkreuzung in ihren Bestand genutzt haben.

Die Divergenz von Fluß- und Sumpfbüffel vor gut 900.000 Jahren widerspricht deutlich dem Ergebnis von Lau und Kollegen. Diese hatten eine Trennung beider Typen vor 28.000 bis 87.000 Jahren abgeschätzt (Lau *et al.*, 1998). Der Hauptgrund für diese "kürzliche" Trennung beider Typen liegt primär in der Zusammenführung der mitochondrialen Haplotypen mit den Phänotypen der Tiere. Wie in dieser Arbeit gezeigt, ist es wichtig den maternalen Erbgang der Mitochondrien zu berücksichtigen. In der Arbeit von Lau und Kollegen wird ein phylogenetischer Baum gezeigt, der eine Separierung in zwei Gruppen zeigt, auch wenn diese mit einem Bootstrapwert unter 50 % unsicher ist. In der einen Gruppe finden sich jedoch phänotypische Fluß- und Sumpfbüffel. Folglich sind geringe Sequenzdivergenzen (der mtDNA) auf den Unterschied zwischen beiden Büffeltypen übertragen worden.

Die unsichere Trennung beider Gruppen in den oben erwähnten Baum, könnten auf die Kürze der analysierten Sequenz zurückzuführen sein. Wie schon weiter oben erläutert führt eine reduzierte Sequenzinformation (158 bp) auch zu einer reduzierten phylogenetischen Information.

Schon frühere Untersuchungen konnte eine deutliche Trennung zwischen Fluß- und Sumpfbüffel darlegen. Der Restriktions Fragmentlängen Polymorphismus (RFLP) konnte in einer Gruppe von 73 Tieren allerdings nur drei Haplotypen definieren (Amano *et al.*, 1994), zeigte aber, daß die Untersuchung der mtDNA für phylogenetische Untersuchungen geeignet wäre. Die gleiche Arbeitsgruppe untersuchte daraufhin die Sequenz des mitochondrialen Cytochrom C Gens und zeigten auch hier eine deutliche Trennung der beiden Büffel (Kikkawa *et al.*, 1997). Aufgrund der paarweisen Divergenz zwischen beiden Büffeltypen (insgesamt 15 Wasserbüffel Haplotypen) schätzten sie eine Divergenz vor etwa 1 Mio. Jahren, ein Wert, der durch die vorliegende Arbeit bestätigt werden konnte.

## 5.4 Problematik der molekularen Evolution

Viele phylogenetische Analysen beruhen auf die Auswertung von Sequenzdivergenzen innerhalb einer mehr oder weniger großen Region von kodierenden oder nicht kodierenden DNA Regionen. In der phylogenetischen Analyse hat sich die Untersuchung der mitochondrialen DNA (mtDNA) als wirkungsvolles Untersuchungsobjekt erwiesen. Vorteile sind die hohe Kopienzahl, einfache Zugänglichkeit, das Fehlen von Rekombinationen (Olivo *et al.*, 1983), die hohe Substitutionsrate (Brown *et al.*, 1979a) und die maternale Vererbung (Giles *et al.*, 1980). Gerade für die Untersuchung nahe verwandter Spezies hat sich die Untersuchung der D-loop Region wegen der äußerst hohen Substitutionsrate als aussagekräftig erwiesen.

Für unterschiedliche phylogenetische Analysen wurden unterschiedliche Regionen des D-loops sequenziert. Für Populationsstudien an freilebenden Gorillas reichte die Sequenzierung eines 250 bp bzw. 277 bp langes Bereichs der D-loop Region, um charakteristische Merkmale zu definieren, die eine Unterscheidung zwischen dem westlichen Tieflandgorilla und dem östlichen Tiefland- und Berggorilla erlaubten (Garner & Ryder, 1996; Saltonstall *et al.*, 1998). Ein ähnliches Ergebnis erhielten Vigilant und Kollegen in einer molekularanthropologischen Studie im südlichen Afrika (Vigilant *et al.*, 1989). Sie sequenzierten die zwei, die zentrale Domäne flankierenden variablen Bereiche der D-loop Region von Proben verschiedener Ethnien. Ihnen standen Sequenzinformation von zwei, etwa 400 bp großen Fragmenten zur Verfügung. Diese Daten offenbarten zum Teil identische Sequenzen zwischen nichtverwandten Individuen einer Ethnie. Allerdings wurden nie identische Sequenzen zwischen verschiedenen Ethnien (IKung, West- und Ostpygmäen) gefunden. Beiden Beispielen ist gemein, daß es sich um Populationen handelt, die entweder geographisch getrennt sind (Gorillas), oder die aufgrund ihrer Lebensweise (Jäger-Sammler) nur einen geringen Genfluß haben. Aufgrund der Diversität der verschiedenen Populationen und der Annahme eines gemeinsamen Ursprungs lassen sich Migrationen rekonstruieren (Ammerman & Cavalli-Sforza, 1985; Stoneking & Wilson, 1989).

Die Bestimmung von Populationen anhand der D-loop Sequenz wird dann schwierig, wenn die Populationen erst seit einem kurzen Zeitraum getrennt sind. So haben Di Rienzo und Wilson 400 bp der human D-loop Region von Individuen aus Sardinien und dem Mittleren Osten (Saudi Arabien, Israel, Jemen) sequenziert und phylogenetisch analysiert (Di Rienzo & Wilson, 1991). Der Sequenzvergleich offenbarte z. T. identische Haplotypen in beiden Gruppen. Schaut man sich das Sequenzalignment und das Dendrogramm an, findet man keine charakteristische Differenzierung zwischen kaukasischen und arabischen Populationen; beide kommen direkt nebeneinander im Stammbaum vor. Ein Ergebnis wie es in der vorliegenden Arbeit für die Büffel gezeigt wurde: unterschiedliche Rassen können sich ähnlicher sein als Individuen einer Rasse (bezogen auf die maternale Linie).

#### 5.4.1 Einfluß des verfügbaren Datenumfanges auf die Aussagefähigkeit einer phylogenetischen Analyse

Die Beispiele oben zeigen zwar, daß auch kurze Sequenzabschnitte die Information liefern können, um eine klare Aussage zu bekommen. In vielen Fällen kann aber eine gewisse (kurze) Information zu einem klaren Ergebnis kommen - die Frage ist, ob dieses auch das richtige ist.

Die Wichtigkeit der Länge einer untersuchten Sequenz läßt sich an der Arbeit von Brown und Kollegen leicht beispielhaft demonstrieren (Brown *et al.*, 1982). Sie haben die hominoid Phylogenese anhand der mitochondrialen DNA in fünf Primaten untersucht. Von den 896 Nukleotidpositionen waren 90 phylogenetisch informativ, wenn die Parsimonie Methode gewählt wurde. Der beste Baum unter diesem Kriterium ergab eine {{Mensch, {Schimpanse, Gorilla}}, Orang-Utan, Gibbon} Phylogenese. Betrachtet man hypothetisch nur die ersten 31 Basen der Sequenz kommt man nur auf 5 informative Positionen. Diese unterstützen dann eine {{{Mensch, Gorilla}, Schimpanse}, Orang-Utan, Gibbon} Phylogenese. Auf der anderen Seite kann man vermuten, daß z. B. die Analyse der gesamten mtDNA wiederum eine andere Phylogenese favorisiert.

Ein ähnlicher, wenn auch nicht so drastischer und deutlicher Effekt trat auch bei der hier vorliegenden Arbeit auf. Durch die Sequenzierung der gesamten D-loop Region ergaben sich 128 variable Positionen. Der Vergleich mit einem 158 bp langem Abschnitt (Lau *et al.*, 1998) führte zu einer "Beschneidung" der Sequenzinformation. Die Anzahl variabler Positionen wurde auf 47 reduziert. Die Anzahl identifizierter Haplotypen reduzierte sich von 36 auf 7. Allein dieser Effekt läßt die Deutlichkeit erkennen, wie viel Information verloren gehen kann. Die Schwierigkeit besteht hier, abzuwägen zwischen dem Arbeitsaufwand und dem zu erwarteten Informationsgehalt.

Trotzdem, es bleibt zu bedenken, daß auch die gesamte D-loop Region nur etwa 6 % der mtDNA ausmacht. Folglich würde die Analyse noch genauer, wenn die Sequenzinformation der kompletten mtDNA genutzt würde. Hier haben Ingman und Kollegen eine umfassende Analyse der humanen Evolution vorgelegt, als sie die komplette mtDNA 53 verschiedener Ethnien sequenzierten und diese phylogenetisch untersuchten (Ingman *et al.*, 2000). Der Arbeitsaufwand ist entsprechend hoch und sie haben auch nur die mtDNA eines Individuums je Ethnie sequenziert. Damit ist allerdings die Variabilität innerhalb einer Population bzw. Ethnie nicht berücksichtigt.

Ob sich mit zusätzlicher Information der mitochondrialen Gene eventuell auch Tierrassen identifizieren ließen, bliebe noch zu erforschen. Die D-loop Region liefert jedenfalls nicht die benötigte Information. Dies hat die hier vorliegende Arbeit gezeigt, aber auch frühere Untersuchungen in anderen Spezies ließen keinen Schluß von der D-loop Sequenz auf die Rasse zu. Hunde, die als das älteste Haustier des Menschen gelten, waren anhand der D-loop Sequenz nicht einer spezifischen Rasse zuzuordnen. Domestizierte Hunde zeigen einen hohen Grad an mtDNA Polymorphismen aus den ancestralen Wolfpopulationen (Okumura *et al.*, 1996; Tsuda *et al.*, 1997). Auch Pferderassen ließen sich aufgrund der D-loop Polymorphismen nicht spezifisch zuordnen (Kim *et al.*, 1999). In beiden Fällen sind

dies die Folgen der intensiven Kreuzungen zwischen maternalen Linien. Wobei die Kreuzungen der maternalen Linien wohl den größten Einfluß nehmen, da sie in einer Linie persistieren.

#### 5.4.2 Problematik des Art- und Rassebegriffs

Die Untersuchung zur Phylogenese domestizierter Tiere birgt ein Problem, da durch den Eingriff des Menschen und durch Translokationen bzw. Migrationen Kreuzungen von Tieren erfolgten, die oft nicht mehr nachvollziehbar sind. Es tritt das Problem auf zu entscheiden, wo eine Art aufhört und eine Rasse beginnt. Die Problematik existiert auch bei den Büffeln. Taxonomisch gehören Fluß- und Sumpfbüffel zu einer Art (*Bubalus bubalis*) und werden deshalb oft als Fluß- bzw. Sumpftyp bezeichnet. Beide sind auch untereinander kreuzbar und liefern fertile Nachkommen. Auf der anderen Seite unterscheiden sie sich im Karyotyp und auch die D-loop Ergebnisse weisen auf eine Divergenz hin, die eine Trennung auf Speziesebene vermuten läßt.

Eine Definition zum Artbegriff sagt: eine Spezies umfaßt alle Populationen von Tieren, deren Angehörige faktisch oder potentiell kreuzbar und von den Angehörigen anderer Populationen reproduktiv isoliert sind. Das ist die Einheit, in der die Evolution (und Tierzucht) abläuft. Peter Ax hat den Begriff der evolutionären Art gewählt (Ax, 1988): "Evolutionäre Arten sind Linien von Populationen mit Individuen im Verhältnis Vorfahren-Nachkommen, welche zu jedem Zeitpunkt ihrer Lebensspanne geschlossene Fortpflanzungsgemeinschaften bilden. Arten entstehen durch Spaltung existierender Arten; Arten erlöschen unter Auflösung in Folgearten, in einem neuen Artbildungsprozeß oder aber durch Aussterben ohne Nachkommen."

Von Wildtieren stammen die domestizierten Tiere ab. Die züchterischen Einheiten innerhalb der Haustiere sind Rassen. Das sind vom Menschen künstlich generierte geschlossene Fortpflanzungsgemeinschaften. Die Individuen einer Rasse haben Gemeinsamkeiten bezüglich ihrer gemeinsamen Abstammung, phänotypischer Merkmale und Leistungsmerkmale bzw. ihres Nutzungszwecks. Die Definition ist aber nicht einheitlich. Auch lokale Populationen ohne züchterische Betreuung werden als Rasse bezeichnet.

Zusammenfassend läßt sich sagen, daß der Rassebegriff ein schwieriges Thema bleibt, weil mit verschiedenen Markersystemen unterschiedliche Ergebnisse erzielt werden. Genetische Marker der mtDNA ergeben oft ein anderes Bild von genetischen Strukturen einer Population als das, welches man von nukleären genomischen Markern herleiten kann (DeSalle & Giddings, 1986; Loftus *et al.*, 1994; Yonekawa *et al.*, 1988).

#### 5.4.3 Verrauschen der Daten

Die Untersuchung der Anzahl variabler Merkmale gegen die genetische Divergenz sollte abschätzen, ob die neutralen Substitutionen der D-loop Region noch informativ sind, da sie mit zunehmender Divergenzzeit verrauschen. Weil im Laufe der Zeit Substitutionen zwi-



schen zwei Sequenzen akkumulieren erreichen die Sequenzunterschiede eine Sättigung. Immer mehr Basen die substituiert werden, wurden schon mal substituiert.

Die Analyse hier zeigt jedoch keine Anzeichen für eine Sättigung innerhalb der Büffelpopulationen. Damit kann ein Verrauschen der Daten ausgeschlossen werden. Der Anstieg der Transitionen im Verhältnis zur Divergenz ist noch fast linear, wie aus Abbildung 4-4 zu sehen ist.

Interessant sind in dem Zusammenhang die Werte, die sich für die selbe Untersuchung zwischen Büffel und Rind ergeben. Die Abbildung 4-3 läßt einen linearen Anstieg der Transitionen bei offenbaren Erreichen eines Plateaus der Transversionen erkennen. Hier würde man ein umgekehrtes Ergebnis erwarten, da die Transitionen wesentlich schneller akkumulieren als die Transversionen. Nach einer gewissen Divergenz sollten die Transitionen, die ja wesentlich schneller akkumulieren, ein Sättigungsniveau erreichen, während die Transversionen langsam, aber stetig (linear) ansteigen sollten (Janecek *et al.*, 1996).

Hier muß kritisch in Betracht gezogen werden, daß *Bos* nicht die geeignetste Außengruppe für *Bubalus* ist und phylogenetisch schon so weit entfernt ist, daß eine sichere Analyse der D-loop Daten nicht gewährleistet ist. Die Daten von Janecek und Kollegen zeigen für Transitionen im mitochondrialen Cytochrom C Oxidase II Gen den Ansatz einer Sättigung schon zwischen *Bos taurus* und *Bubalus*. Die D-loop Region evolviert noch schneller und multiple Substitutionen zwischen Rind und Büffel können erwartet werden. Die Divergenz beider Genera machte sich auch bei den Schwierigkeit des Alignments bemerkbar. So zeigt der Dotplot (Abbildung 4-7) mehrere Möglichkeiten eines Alignment. Das heißt, mit Sicherheit können homologe Merkmale nicht bestimmt werden.

Sinnvollere Außengruppen wären die Wildformen der asiatischen Büffel (*Anoa*) oder der afrikanische Büffel (*Syncerus caffer caffer* oder *S. c. nanus*). Leider gibt es von keiner dieser Spezies komplette D-loop Sequenzen in den Datenbanken. Zukünftige Untersuchungen in diese Richtung sollen Proben von diesen Arten für die Außengruppe mit einschließen.

Aufgrund der Sättigung der Transitionen kann keine zuverlässige Aussage über die Divergenz zwischen *Bos* und *Bubalus* gemacht werden. Der errechnete Wert von über 5 Mio. Jahren muß daher als Untergrenze angesehen werden.

## 5.5 Bubaline Strukturelemente sind homolog zu denen anderer Vertebraten

Im Zuge des Vergleichs der D-loop Sequenzen von *Bubalus bubalis* mit *Bos taurus* wurde in den Ergebnissen auch auf konservierte Elemente und Strukturen eingegangen. (Abbildung 4-11 und Abbildung 4-12). Die D-loop Region kann danach in drei Regionen unterteilt werden: die zentrale Domäne und die flankierenden ETAS und CSB Regionen

### 5.5.1 Multiple Del/Ins Ereignisse in der ETAS Region

Die ETAS Region befindet sich am 3' Ende der D-loop Region, wo der neu synthetisierte dritte Strang (H) endet. Die Länge dieser Region kann in verschiedenen Vertebraten sehr unterschiedlich sein (zwischen 209 bp und 412 bp) und hat beim Rind eine Länge von 367 bp (Sbisà *et al.*, 1997). Dieser Bereich ist in den Flußbüffeln vier, in den Sumpfbüffeln sechs Basen länger. Dies ist insbesondere auf multiple Insertionen vor der ETAS1 Region zurückzuführen.

Zu Beginn der 1980er Jahre sind kurze Sequenzmotive identifiziert worden, die vermutlich mit der Termination der H-Strang Synthese in Verbindung stehen könnten, als TAS (termination associated sequences) bezeichnet (Doda *et al.*, 1981). Weitere Untersuchungen ließen diese Hypothese fraglich erscheinen und 1997 veröffentlichten Sbisà und Kollegen eine Strukturanalyse 26 kompletter D-loop Sequenzen, die in den Datenbanken verfügbar waren. Das multiple Alignment steht im Internet unter <http://www.ba.cnr.it/dloop.html> zur Verfügung (Sbisà *et al.*, 1997). In diesem Alignment identifizierten sie zwei neue konservierte Sequenzbereiche, von je etwa 60 bp Länge. Sie wurden als ETAS1 und ETAS2 ("extended TAS") benannt.

Danach werden den Regionen ETAS1 und 2 funktionelle Funktionen in der Replikations- und Transkriptionsregulation zugesprochen. Ihre Verteilung im 3'-Ende der D-loop Region ist in den verschiedenen Spezies unterschiedlich, ihr Abstand voneinander wird mit 1 bp bis 160 bp angegeben. In der bovinen und bubalinen D-loop sind beide Bereiche durch zwei Basen voneinander getrennt. Madsen und Kollegen führten *in vitro* und *in organello* Studien durch, um Protein - mtDNA Interaktionen zu studieren (Madsen *et al.*, 1993). Dabei lokalisierten sie eine spezifische Bindung eines 48 kDa großes Protein in dem Bereich, den Sbisà und Kollegen später als ETAS2 bezeichneten.

### 5.5.2 Die Funktion der zentralen Domäne ist noch unbekannt

Die zentrale Domäne ist, im Gegensatz zu den variablen Randbereichen hochkonserviert in verschiedenen Spezies und hat eine Länge zwischen 304-328 bp. Sowohl die bovine als auch die bubaline zentrale Domäne umfaßt 316 bp. Trotzdem die zentrale Domäne über die Speziesgrenze hoch konserviert ist, ist ihre Funktion noch immer unbekannt. Untersuchung der Region zeigten zwar offene Leserahmen, die abgeleiteten Aminosäuresequenzen waren in den untersuchten Spezies aber zu unterschiedlich, als das eine gemeinsame Funktion erkennbar wäre (Saccone *et al.*, 1987).

### 5.5.3 Die bubaline D-loop Region enthält den CSB3

Die CSB Region im 5'-Bereich der D-loop Region ist weniger konserviert als die ETAS Region (Sbisà *et al.*, 1990). Die Variabilität ist nicht unbedingt höher, wohl gibt es aber Spezies die drei CSB Bereiche haben, während anderen Spezies der CSB3 völlig fehlt (z. B. Delphin, Rind, Ziege).

Wie gezeigt sind die Sequenzunterschiede zwischen Büffel und Rind nicht sehr groß. Ein interessanter Aspekt in der Rinder D-loop Region ist das Fehlen der CSB3 Region. Es konnte gezeigt werden, daß die sogenannte RNase mitochondriale RNA Prozessierung (MRP) die beiden Sequenzdomänen CSB2 und CSB3 erkennt und ganz spezifisch das L-Strang Transkript zwischen diesen Domänen schneidet (Chang & Clayton, 1987b). Das Fehlen der CSB3 beim Rind scheint aber keinen Einfluß auf die Funktion der RNA MRP zu haben und schneidet statt dessen eine homologes Duplikat mit einem poly-C Motiv, das dem der CSB2 ähnelt (Dairaghi & Clayton, 1993). Diese homologe Duplikatsequenz ist in (Sbisà *et al.*, 1997) beschrieben und entspricht dem 5'-Ende der D-loop Region mit der Sequenz CAGG(C)<sub>12</sub>. Wie in Abbildung 4-1 gezeigt wurde, ist der Übergang von der D-loop Region zur tRNA<sup>Phe</sup> bei den Büffeln anders als beim Rind; dort findet sich statt des poly-C Motivs die Sequenz ACCCAATCCATCACTTCACAACACA. Da den Büffeln offensichtlich diese Homologie fehlt, muß man annehmen, das Büffel eine CSB3 Region besitzen. Diese ist im multiplen Alignment verschiedener Spezies zwischen 5 und 57 bp von der CSB2 entfernt und hat die Konsensussequenz TGC CAA ACC CCA AAA ACA (Sbisà *et al.*, 1997). In den Büffeln findet man in 5 bp Entfernung zum CSB2 die Sequenz TCA AAC CCC CCT TAA ATA (im Rind ist die homologe Sequenz: AAA ATT TCC CCC TTA AAT A), die eine gewisse Ähnlichkeit mit der Konsensussequenz hat. Anhand des Homologievergleichs läßt sich allerdings nicht bestimmen, ob es sich tatsächlich um die CSB3 Region handelt. Hier kann nur eine Funktionsstudie Aufschluß geben.

#### 5.5.4 Die bubalinen D-loop Region weist keine Sequenzwiederholungen auf

In vielen Spezies, Invertebraten und Vertebraten, sind Variationen in der Größe der mtDNA beschrieben worden (Fumagalli *et al.*, 1996; Hoelzel, 1993), die hauptsächlich auf multiple Sequenzwiederholungen (tandem repeats) zurückzuführen sind. Diese Insertionen von Sequenzwiederholungen finden in den beiden Randbereichen der D-loop Region statt und bestätigen damit, daß die D-loop Region Del/Ins Ereignisse toleriert und Sequenzwiederholungen akkumulieren kann, ohne in der Funktion beeinträchtigt zu werden. Solche Sequenzwiederholungen sind im Rind, und wie die vorliegende Arbeit zeigt, auch im Büffel nicht zu finden.

Die Promotorregion für die H- und L-Strang Synthese sind für die bovine mtDNA mittels Protein-"footprinting" nachgewiesen (Ghivizzani *et al.*, 1994). Die bubalinen Sequenzen, korrespondierend zu diesen funktionellen Regionen, sind homolog und dürften daher auch funktionell sein.

#### 5.6 Form der Sekundärstrukturen in der D-loop Region ist fraglich

Sowohl in der ETAS als auch in der CSB Region sind kompakt gefaltete Kleeblattstrukturen in der einzelsträngigen Polynukleotidkette von Mensch, Maus und Rind beschrieben worden (Brown *et al.*, 1986). In der vorliegenden Arbeit konnte gezeigt werden, das die betei-

ligten bubalinen Sequenzen homolog zu den bovinen sind. Die an der Ausbildung der Sekundärstruktur beteiligten Bereiche im Wasserbüffel (Abbildung 4-12) zeigen leichte Unterschiede, die aber die Ausbildung der Stämme, und damit die thermodynamische Stabilität, aber kaum beeinflussen dürften. Die meisten Substitutionen liegen in den Bereichen der Loops, die nicht an internen Basenbindungen teilhaben. Lediglich die Sequenz der Sumpfbüffel zeigte eine vermutlich starke Verkürzung des Hauptstamms. Inwieweit dadurch die thermodynamische Stabilität beeinflusst ist, kann aus der Sequenzinformation nicht abgelesen werden.

Die Vorhersagen über die Ausbildung von Sekundärstrukturen (Haarnadel, Kleeblatt) in der D-loop Region erfolgte in der Arbeit von Brown und Kollegen mittels Computerprogrammen, die nach einem Energiemodell Vorhersagen über die Faltung von tRNA und 5S RNA Molekülen treffen (Papanicolaou *et al.*, 1984). Über eine schrittweise Verbesserung ihres ursprünglichen Programms erreichten sie eine 80 %ige Sicherheit in der Vorhersage von Kleeblattstrukturen in tRNA und eine 60 %ige Sicherheit in der Vorhersage der Konsensusfaltung von 5S RNA Molekülen. Das heißt aber zum einen, daß eine Restunsicherheit in der Vorhersage bleibt und was viel entscheidender ist: dieses Programm wurde speziell für die Analyse von tRNA und 5S RNA entwickelt.

### 5.6.1 Minimale freie Energie favorisiert nicht die Kleeblattstruktur

Leider ist die Software von Papanicolaou und Kollegen nicht frei verfügbar (z. B. im Internet), um die Konstruktion der RNA Faltung zu wiederholen. Es gibt aber im Internet eine Vielzahl von Programmen bzw. Serverplattformen, die den Service einer Vorhersage zur RNA Faltung anbieten. Eine Plattform nutzt das vielverwendete Programmpaket von Ivo Hofacker "Vienna RNA Package" (Hofacker & Stadler, 1999; McCaskill, 1990; Wuchty *et al.*, 1999). Auf der Internetseite <http://www.tbi.univie.ac.at/cgi-bin/RNAfold.cgi> findet sich eine interaktive Version dieses Programms. Mit Hilfe dieser Software ließ sich die von Brown und Kollegen vorhergesagte Struktur in der D-loop Region nicht wiederholen.

Die tRNA Moleküle haben bekanntlich eine Kleeblattstruktur. Deshalb wurde willkürlich die Sequenz der tRNA<sup>Phe</sup> durch das "Vienna RNA Package" gefaltet. Die optimale sekundäre Struktur ist in der "Klammernotation" ausgegeben.

```
GTTGATGTAGCTTAACCCAAAGCAAGGCACTGAAAATGCCTAGATGAGTCTCCCAACTCCATAAACACCA
(((.((((..{{{.....}}})·[[[[[.[.....]]]]].....<<<<.....>>>>)))..)).....
```

Diese Darstellung zeigt eine Kleeblattstruktur mit einem Hauptstamm (runde Klammern) und drei Loops. Die minimale freie Energie betrug -12,50 kcal/mol, die freie Energie des thermodynamischen Ensemble war -15,11 kcal/mol.

Zur Wiederholung der Kleeblatffaltung wurde der Teil der bovinen Sequenz eingegeben, der ebenfalls ein Kleeblatt formieren soll, und unter den selben Bedingungen wie oben gefaltet. Die optimale Sekundärstruktur hatte folgende Form:

```

CTATGGCCGTCAAAGGCCCTGACCCGGAGCATCTATTGTAGCTGGACTTAACTG
....((((.....))) - ((((((.....{}}}}-}}}}-}}}}-}}}}{...
CATCTTGAGCACCAGCATAATGATAAGCATGGACATTACAGTCAATGGTCACAG
....}}}}}.}}}}}}}}}}}}}.}}}.<<.....>>...)))))...

```

Diese Sekundärstruktur hatte nichts mit der vorgeschlagenen Kleeblattstruktur gemein. Es gibt einen Bereich von etwa 24 Nukleotiden, die strickleiterartig miteinander verbunden sind und am Umkehrpunkt einen kleinen Loop bilden (geschweifte Klammern). Diese Struktur ist, wie die Werte zeigen, thermodynamisch stabiler als die tRNA Struktur: die minimale freie Energie betrug -31,10 kcal/mol, die freie Energie des thermodynamischen Ensemble war -32,44 kcal/mol.

Die bubaline Sequenz weist in diesem Bereich nur geringfügige Substitutionen auf. Ihre Faltung ergibt ein ähnliches Ergebnis wie für die bovine Struktur. Entsprechende Ergebnisse lieferte die Vorhersage zur Sekundärstruktur im 3'-Bereich der D-loop Region.

Mit Hinblick auf dieses Ergebnis ist es fraglich, welche Struktur die 3'- und 5'-Bereiche der D-loop Region tatsächlich annehmen. Dieses Problem zu lösen war aber nicht Aufgabe der vorliegenden Arbeit.

## 5.7 Die Wildform des *Bubalus bubalis*

Die Wildform des asiatischen Wasserbüffels wird als die Stammart der domestizierten Büffel angesehen (Cockrill, 1981; FAO, 2000). Aber nicht nur die domestizierte Art ist von großem wirtschaftlichem Interesse, da von keinem domestiziertem Tier so viele Menschen abhängig sind, sondern auch der Wildform kommt immense Bedeutung zu.

Daher müssen, mit Hinblick auf die momentane Situation, unbedingt weitere Untersuchungen zur Diversität der Büffel durchgeführt werden.

### 5.7.1 Verteilung und derzeitiger Status der Wildform

Wildbüffel werden heute nur noch in kleinen Bereichen ihrer früheren Ausbreitungsgebiete gefunden. Es gibt heute sicher weniger als 4000 Tiere und es ist wahrscheinlich, daß es gar weniger als 200 sind. Dabei bleibt fraglich, ob diese Tiere überhaupt noch reinrassig sind. Einige isolierte Populationen sollen noch in den Bastar und Raipor Bezirken des Madhya Pradesh, dem Royal Manas Wildlife Sanctuary / Projekt Tiger Reserve (Indien), Kosi Tappu Wildreservat (Nepal), Royal Manas Nationalpark (Bhutan) und dem Huai Kha Khoeng Wildlife Sanctuary (Thailand) vorkommen. Das sind zumindest die Populationen, von denen man annimmt, daß sie von Kreuzungen mit domestizierten Tieren bzw. verwilderten domestizierten Tieren am wenigsten betroffen sind. Tiere in zoologischen Gärten gelten alle als Hybriden und daher als ungeeignet für Untersuchungen zur Wildform. Exportierte oder verwilderte Büffel mit domestiziertem Ursprung kommen ebenfalls in Australien, Brasilien,

Indien, Indonesien, Laos, Thailand, Vietnam, auf den Philippinen, Timor, Italien und Sri Lanka vor (FAO, 2000).

## 6 Zusammenfassung

Die vorliegende Arbeit hatte zum Ziel die Phylogenese der Wasserbüffel (*Bubalus bubalis*) unter dem Aspekt der molekularen Evolution zu betrachten. Die Stammesgeschichte der Büffel wurde in der Vergangenheit schon mehrfach untersucht, führte aber zu keinem einheitlichen Ergebnis; Divergenzzeiten von 10.000 Jahren bis 1,7 Mio. Jahren für Fluß- und Sumpfbüffel sind beschrieben, abhängig vom Umfang des Probenmaterials und der Sensitivität der Tests.

Die vorliegende Arbeit stellt erstmalig die Analyse der kompletten bubalinen D-loop Region vor. 80 Individuen aus vier verschiedenen Rassen wurden in die Untersuchung einbezogen. Für jedes Tier wurde die Sequenz der mitochondrialen D-loop Region bestimmt und mittels entsprechender Software phylogenetisch analysiert.

Die molekulargenetische Analyse erlaubte eine klare Unterscheidung zwischen Sumpf- und Flußbüffeln. Innerhalb der Flußbüffel war keine rassenspezifische Differenzierung möglich. Im Gegenteil, es fanden sich ein Haplotyp, der in allen drei Flußbüfflerassen zu finden war. Die Möglichkeit, daß es sich dabei um einen ancestralen Haplotypen handelt wurde diskutiert.

Es wurde versucht eine Abschätzung der Divergenz von Fluß- und Sumpfbüffel vorzunehmen. Durch Einbeziehen der Sequenzinformation der gesamten D-loop Region ist eine hohe Genauigkeit erreicht, die eine gute Abschätzung erlaubt. Frühere Untersuchungen beschränkten sich nur auf einen kleinen Bereich der D-loop Region, auf die gesamte D-loop Region aber nur für wenige Individuen, auf nicht so schnell evolvierende Gene, oder auf Methoden, die nur einzelne Punktmutationen oder einige polymorphe Mikrosatelliten aufspürten. Unter der Annahme einer ähnlichen Substitutionsrate wie sie für Homioide bestimmt wurde, fand die Trennung von Fluß- und Sumpfbüffel vor ca. 900.000 Jahren statt; also deutlich früher als der Domestikationsbeginn. Die Bestimmung ihres gemeinsamen Vorfahren mit den Rindern wurde auf über 5 Mio. Jahre bestimmt. Dabei gibt dieser Wert die unterste Grenze an, da die Analyse des Transitions / Divergenzverhältnisses eine Sättigung für *Bos* vs. *Bubalus* ergab.

Die Untersuchung der mitochondrialen DNA zum Zwecke der phylogenetischen Abschätzung zwischen verschiedenen Spezies / Genera, aber auch auf Ebene der Populationsbiologie ist und bleibt ein wichtiges Werkzeug in der molekularen Evolutionsforschung. Radiative Speziation, geographische Verteilung, Diversität einer Art, Differenzierung von Linien und der Ursprung des Menschen sind Bereiche, in der dieses Werkzeug seine Anwendung findet. So kann, und soll, die vorliegende Arbeit Anstoß zu weiterer Forschung in

diese Richtung geben. Wie im vorigen Kapitel erwähnt, ist der Wasserbüffel eines der wichtigsten Nutztiere des Menschen, weshalb die Untersuchung der Diversität auch in Zukunft ein wichtiger Aspekt sein wird.

Die Wichtigkeit die genetischen Ressourcen der Wildformen unserer domestizierten Tiere zu konservieren wird heute von vielen (an)erkannt, da weitere Diversität den Nutztierpopulationen hinzugefügt werden kann. Dann lassen sich mit genaueren Ergebnissen zur Diversität effizientere Zuchtprogramme entwickeln. Ein "gesteuerter Genfluß" bietet auch den kleinen Farmern die Möglichkeit das Potential ihrer Farm zu verbessern.



## 7 Literaturverzeichnis

- Amano, T., Miyakoshi, Y., Takada, T., Kikkawa, Y. & Suzuki, H. (1994). Genetic variants of ribosomal DNA and mitochondrial DNA between swamp and river buffaloes. *Anim Genet* **25** Suppl 1, 29-36.
- Ammerman, A. J. & Cavalli-Sforza, L. L. (1985). *The neolithic transition and the genetics of populations in Europe*, Princeton University Press, Princeton, NJ.
- Anderson, S., de Bruijn, M. H., Coulson, A. R., Eperon, I. C., Sanger, F. & Young, I. G. (1982). Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *J Mol Biol* **156**(4), 683-717.
- Arnason, E. & Rand, D. M. (1992). Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic cod, *Gadus morhua*. *Genetics* **132**(1), 211-20.
- Årnason, U. & Johnsson, E. (1992). The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina*. *J Mol Evol* **34**(6), 493-505.
- Avise, J. C. (1986). Mitochondrial DNA and the evolutionary genetics of higher animals. *Philos Trans R Soc Lond B Biol Sci* **312**(1154), 325-42.
- Avise, J. C. (1991). Evolution. Matriarchal liberation [news; comment]. *Nature* **352**(6332), 192.
- Ax, P. (1988). *Systematik in der Biologie*, Gustav Fischer, Stuttgart.
- Bandelt, H. J., Forster, P. & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**(1), 37-48.
- Barker, J. S., Moore, S. S., Hetzel, D. J., Evans, D., Tan, S. G. & Byrne, K. (1997a). Genetic diversity of Asian water buffalo (*Bubalus bubalis*): microsatellite variation and a comparison with protein-coding loci. *Anim Genet* **28**(2), 103-15.
- Barker, J. S., Tan, S. G., Selvaraj, O. S. & Mukherjee, T. K. (1997b). Genetic variation within and relationships among populations of Asian water buffalo (*Bubalus bubalis*). *Anim Genet* **28**(1), 1-13.
- Brown, G. G., Gadaleta, G., Pepe, G., Saccone, C. & Sbisà, E. (1986). Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *J Mol Biol* **192**(3), 503-11.
- Brown, W. M., George, M. & Wilson, A. C. (1979a). Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A* **76**(4), 1967-71.

- Brown, W. M., George, M., Jr. & Wilson, A. C. (1979b). Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A* **76**(4), 1967-71.
- Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. (1982). Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol* **18**(4), 225-39.
- Chang, D. D. & Clayton, D. A. (1987a). A mammalian mitochondrial RNA processing activity contains nucleus-encoded RNA. *Science* **235**(4793), 1178-84.
- Chang, D. D. & Clayton, D. A. (1987b). A novel endoribonuclease cleaves at a priming site of mouse mitochondrial DNA replication. *Embo J* **6**(2), 409-17.
- Chang, D. D. & Clayton, D. A. (1989). Mouse RNAase MRP RNA is encoded by a nuclear gene and contains a decamer sequence complementary to a conserved region of mitochondrial RNA substrate. *Cell* **56**(1), 131-9.
- Chang, D. D., Hixson, J. E. & Clayton, D. A. (1986). Minor transcription initiation events indicate that both human mitochondrial promoters function bidirectionally. *Mol Cell Biol* **6**(1), 294-301.
- Chen, Y. C. & Li, X. H. (1989). New evidence of the origin and domestication of the Chinese swamp buffalo (*Bubalus bubalis*). *Buffalo Journal* **1**, 51-55.
- Cockrill, W. (1981). The water buffalo: A review. *Brit. Vet. J.* **137**(1), 8-16.
- Dairaghi, D. J. & Clayton, D. A. (1993). Bovine RNase MRP cleaves the divergent bovine mitochondrial RNA sequence at the displacement-loop region. *J Mol Evol* **37**(4), 338-46.
- DeSalle, R. & Giddings, L. V. (1986). Discordance of nuclear and mitochondrial DNA phylogenies in Hawaiian *Drosophila*. *Proc Natl Acad Sci U S A* **83**(18), 6902-6.
- Di Berardino, D. & Iannuzzi, L. (1981). Chromosome banding homologies in Swamp and Murrah buffalo. *J Hered* **72**(3), 183-8.
- Di Rienzo, A. & Wilson, A. C. (1991). Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci U S A* **88**(5), 1597-601.
- Doda, J. N., Wright, C. T. & Clayton, D. A. (1981). Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc Natl Acad Sci U S A* **78**(10), 6116-20.
- Dress, A. (1995). The mathematical basis of molecular phylogenetics.
- Excoffier, L. & Yang, Z. (1999). Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. *Mol Biol Evol* **16**(10), 1357-68.
- FAO. (2000). *World Watch List for domestic animal diversity*. 3rd edit (Scherf, B. D., Ed.), FAO, Rome.
- Farris, J. S., Albert, V. A., Källersjö, M., Lipscomb, D. & Kluge, A. G. (1996). Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12**, 99-124.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**(6), 368-76.

- Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet* **22**, 521-65.
- Fischer, H. & Ulbrich, F. (1968). Chromosomes of the Murrah buffalo and its crossbreeds with the Asiatic swamp buffalo (*Bubalus bubalis*). *Z Tierzuchtg* **84**, 110-114.
- Foulds, L. R. & Graham, R. L. (1982). The steiner problem in phylogeny is np-complete. *Proc. Natl. Academy Science* **3**, 43-49
- Fumagalli, L., Taberlet, P., Favre, L. & Hausser, J. (1996). Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Mol Biol Evol* **13**(1), 31-46.
- Garner, K. J. & Ryder, O. A. (1996). Mitochondrial DNA diversity in gorillas. *Mol Phylogenet Evol* **6**(1), 39-48.
- Ghivizzani, S. C., Madsen, C. S., Nelen, M. R., Ammini, C. V. & Hauswirth, W. W. (1994). In organello footprint analysis of human mitochondrial DNA: human mitochondrial transcription factor A interactions at the origin of replication. *Mol Cell Biol* **14**(12), 7717-30.
- Giles, R. E., Blanc, H., Cann, H. M. & Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* **77**(11), 6715-9.
- Giuffra, E., Bernatchez, L. & Guyomard, R. (1994). Mitochondrial control region and protein coding genes sequence variation among phenotypic forms of brown trout *Salmo trutta* from northern Italy. *Mol Ecol* **3**(2), 161-71.
- Goldman, N. (1993). Statistical tests of models of DNA substitution. *J Mol Evol* **36**(2), 182-98.
- Graur, D. (1984). Pattern of nucleotide substitution and the extent of purifying selection in retroviruses. *J Mol Evol* **21**(3), 221-31.
- Greenberg, B. D., Newbold, J. E. & Sugino, A. (1983). Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* **21**(1-2), 33-49.
- Groves, C. (1981). Systematic relationship in the bovini (*Artiodactyla*, *Bovidae*). *Zeitschrift für Zoologische Systematik und Evolutionsforschung*.
- Gupta, R. S. & Srivastava, R. K. (1984). Estimation of succinic dehydrogenase and cytochrome oxidase activities in the semen of buffalo, exotic and crossbreed bulls. *Cell Mol Biol* **30**(3), 253-5.
- Gyllensten, U., Wharton, D., Josefsson, A. & Wilson, A. C. (1991). Paternal inheritance of mitochondrial DNA in mice. *Nature* **352**(6332), 255-7.
- Harisah, M., Azmi, T. I., Hilmi, M., Vidyadaran, M. K., Bongso, T. A., Nava, Z. M., Momongan, V. & Basrur, P. K. (1989). Identification of crossbred buffalo genotypes and their chromosome segregation patterns. *Genome* **32**(6), 999-1002.
- Hasegawa, M., Kishino, H. & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* **22**(2), 160-74.

- Hoelzel, A. R. (1993). Evolution by DNA turnover in the control region of vertebrate mitochondrial DNA. *Curr Opin Genet Dev* **3**(6), 891-5.
- Hofacker, I. L. & Stadler, P. F. (1999). Automatic detection of conserved base pairing patterns in RNA virus genomes. *Comput Chem* **23**(3-4), 401-14.
- Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* **408**(6813), 708-13.
- Janecek, L. L., Honeycutt, R. L., Adkins, R. M. & Davis, S. K. (1996). Mitochondrial gene sequences and the molecular systematics of the artiodactyl subfamily bovinæ. *Mol Phylogenet Evol* **6**(1), 107-19.
- Jukes, T. H. (1969). Evolutionary pattern of specificity regions in light chains of immunoglobulins. *Biochem Genet.* **3**(2), 109-17.
- Kikkawa, Y., Yonekawa, H., Suzuki, H. & Amano, T. (1997). Analysis of genetic diversity of domestic water buffaloes and anoas based on variations in the mitochondrial gene for cytochrome b. *Anim Genet* **28**, 195-201.
- Kim, K. I., Yang, Y. H., Lee, S. S., Park, C., Ma, R., Bouzat, J. L. & Lewin, H. A. (1999). Phylogenetic relationships of Cheju horses to other horse breeds as determined by mtDNA D-loop sequence polymorphism. *Anim Genet* **30**(2), 102-8.
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* **217**(129), 624-6.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* **16**(2), 111-20.
- King, J. L. & Jukes, T. H. (1969). Non-Darwinian evolution. *Science* **164**(881), 788-98.
- Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, S., Villablanca, F. X. & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci U S A* **86**(16), 6196-200.
- Kowalski, K. (1986). *Die Tierwelt des Eiszeitalters*. Trans. Kubiak, H. Erträge der Forschung, 239, Wissenschaftliche Buchgesellschaft, Darmstadt.
- Lanave, C., Tommasi, S., Preparata, G. & Saccone, C. (1986). Transition and transversion rate in the evolution of animal mitochondrial DNA. *Biosystems* **19**(4), 273-83.
- Lau, C. H., Drinkwater, R. D., Yusoff, K., Tan, S. G., Hetzel, D. J. & Barker, J. S. (1998). Genetic diversity of Asian water buffalo (*Bubalus bubalis*): mitochondrial DNA D-loop and cytochrome b sequence variation. *Anim Genet* **29**(4), 253-64.
- Li, W.-H. & Graur, D. (1991). *Fundamentals of Molecular Evolution*, Sinauer Association, Sunderland MA.
- Li, W. H., Gojobori, T. & Nei, M. (1981). Pseudogenes as a paradigm of neutral evolution. *Nature* **292**(5820), 237-9.

- Li, W. H., Wu, C. I. & Luo, C. C. (1984). Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J Mol Evol* **21**(1), 58-71.
- Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M. & Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proc Natl Acad Sci U S A* **91**(7), 2757-61.
- Lyrholm, T., Leimar, O. & Gyllensten, U. (1996). Low diversity and biased substitution patterns in the mitochondrial DNA control region of sperm whales: implications for estimates of time since common ancestry. *Mol Biol Evol* **13**(10), 1318-26.
- Madsen, C. S., Ghivizzani, S. C. & Hauswirth, W. W. (1993). Protein binding to a single termination-associated sequence in the mitochondrial DNA D-loop region. *Mol Cell Biol* **13**(4), 2162-71.
- Margulis, L. (1971). The origin of plant and animal cells. *Am Sci* **59**(2), 230-5.
- Martin, W. & Muller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature* **392**(6671), 37-41.
- McCaskill, J. S. (1990). The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**(6-7), 1105-19.
- Miller, S. A., Dykes, D. D. & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16** (3), 1215
- Muller, M. & Martin, W. (1999). The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. *Bioessays* **21**(5), 377-81.
- Nachtsheim, H. & Stengel, H. (1977). *Vom Wildtier zum Haustier*, Paul Parey Verlag, Berlin (West), Hamburg.
- Nei, M. (1987). *Molecular Evolutionary Genetics*, Columbia University Press, New York.
- Nuttall, G. (1904). *Blood Immunity and Blood Relationship*, Cambridge University Press, Cambridge.
- Okumura, N., Ishiguro, N., Nakano, M., Matsui, A. & Sahara, M. (1996). Intra- and inter-breed genetic variations of mitochondrial DNA major non-coding regions in Japanese native dog breeds (*Canis familiaris*). *Anim Genet* **27**(6), 397-405.
- Olivo, P. D., Van de Walle, M. J., Laipis, P. J. & Hauswirth, W. W. (1983). Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature* **306**(5941), 400-2.
- Pääbo, S. (1991). Amplifying DNA from archeological remains: a meeting report. *PCR Methods Appl* **1**(2), 107-10.
- Page, R. D. M. & Holmes, E. C. (1998a). Inferring molecular phylogeny. In *Molecular evolution: a phylogenetic approach*, pp. 172-227. Blackwell Science Ltd., Oxford.

- Page, R. D. M. & Holmes, E. C. (1998b). Measuring genetic change. In *Molecular evolution: a phylogenetic approach*, pp. 135-171. Blackwell Science Ltd., Oxford.
- Papanicolaou, C., Gouy, M. & Ninio, J. (1984). An energy model that predicts the correct folding of both the tRNA and the 5S RNA molecules. *Nucleic Acids Res* **12**(1 Pt 1), 31-44.
- Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**(9), 817-8.
- Proutski, V. & Holmes, E. (1998). SWAN: sliding window analysis of nucleotide sequence variability. *Bioinformatics* **14**(5), 467-8.
- Rassmann, K. (1997). Evolutionary age of the Galapagos iguanas predates the age of the present Galapagos islands. *Mol Phylogenet Evol* **7**(2), 158-72.
- Ritz, L. R., Glowatzki-Mullis, M. L., MacHugh, D. E. & Gaillard, C. (2000). Phylogenetic analysis of the tribe Bovini using microsatellites [In Process Citation]. *Anim Genet* **31**(3), 178-85.
- Saccone, C., Attimonelli, M. & Sbisà, E. (1987). Structural elements highly preserved during the evolution of the D-loop- containing region in vertebrate mitochondrial DNA. *J Mol Evol* **26**(3), 205-11.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**(4), 406-25.
- Saltonstall, K., Amato, G. & Powell, J. (1998). Mitochondrial DNA variability in Grauer's gorillas of Kahuzi-Biega National Park. *J Hered* **89**(2), 129-35.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**(12), 5463-7.
- Savage, D. & Russell, D. (1983). *Mammalian paleofaunas of the world*, Addison-Wesley Publishing Company, Reading, MA.
- Sbisà, E., Nardelli, M., Tanzariello, F., Tullo, A. & Saccone, C. (1990). The complete and symmetric transcription of the main non coding region of rat mitochondrial genome: in vivo mapping of heavy and light transcripts. *Curr Genet* **17**(3), 247-53.
- Sbisà, E., Tanzariello, F., Reyes, A., Pesole, G. & Saccone, C. (1997). Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* **205**(1-2), 125-40.
- Schubart, C. D., Diesel, R. & Hedges, S. B. (1998). Rapid evolution to terrestrial life in Jamaican crabs. *Nature* **393**, 363-365.
- Stoneking, M. & Wilson, A. C. (1989). A genetic trail. In *The colonisation of the pacific* (Hill, A. V. S. & Serjeantson, S., eds.), pp. 215-245. Oxford University Press, Oxford.
- Swofford, D. L. (2000). PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods) Version 4.0b4a edit. Sinauer Associates, Sunderland, Massachusetts.

- Tamura, K. & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**(3), 512-26.
- Tanaka, K., Solis, C. D., Masangkay, J. S., Maeda, K., Kawamoto, Y. & Namikawa, T. (1996). Phylogenetic relationship among all living species of the genus *Bubalus* based on DNA sequences of the cytochrome b gene. *Biochem Genet* **34**(11-12), 443-52.
- Tanaka, K., Yamagata, T., Masangkay, J. S., Faruque, M. O., Vu-Binh, D., Salundik, Mansjoer, S. S., Kawamoto, Y. & Namikawa, T. (1995). Nucleotide diversity of mitochondrial DNAs between the swamp and the river types of domestic water buffaloes, *Bubalus bubalis*, based on restriction endonuclease cleavage patterns. *Biochem Genet* **33**(5-6), 137-48.
- Tsuda, K., Kikkawa, Y., Yonekawa, H. & Tanabe, Y. (1997). Extensive interbreeding occurred among multiple matriarchal ancestors during the domestication of dogs: evidence from inter- and intraspecies polymorphisms in the D-loop region of mitochondrial DNA between dogs and wolves. *Genes Genet Syst* **72**(4), 229-38.
- Ulbrich, F. & Fischer, H. (1967). The chromosomes of the Asiatic buffalo (*Bubalus bubalis*) and the African buffalo (*Syncerus caffer*). *Z Tierzuchtg* **83**, 219-223.
- Vigilant, L., Pennington, R., Harpending, H., Kocher, T. D. & Wilson, A. C. (1989). Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci U S A* **86**(23), 9350-4.
- Wägele, J.-W. (2000). *Grundlagen der phylogenetischen Systematik*, Verlag Dr. Friedrich Pfeil, München, Germany.
- Wong, T. W. & Clayton, D. A. (1986). DNA primase of human mitochondria is associated with structural RNA that is essential for enzymatic activity. *Cell* **45**(6), 817-25.
- Wuchty, S., Fontana, W., Hofacker, I. L. & Schuster, P. (1999). Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49**(2), 145-65.
- Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* **39**(3), 306-14.
- Yang, Z. (1996). Phylogenetic analysis using parsimony and likelihood methods. *J Mol Evol* **42**(2), 294-307.
- Yonekawa, H., Moriwaki, K., Gotoh, O., Miyashita, N., Matsushima, Y., Shi, L. M., Cho, W. S., Zhen, X. L. & Tagashira, Y. (1988). Hybrid origin of Japanese mice "*Mus musculus molossinus*": evidence from restriction analysis of mitochondrial DNA. *Mol Biol Evol* **5**(1), 63-78.

## 8 Anhang

### Anhang 1: Abkürzungen

A	Adenin
APS	Amoniumperoxodisulfat
bp	Basenpaare
C	Cytosin
Car	Carabao (Sumpfbüffel)
CIA	Chloroform/Isoamylalkohol
CSB	conserved sequence block
Cytb	Cytochrom b
ddH <sub>2</sub> O	Doppeltdestiliertes Wasser
ddNTP	Didesoxynukleosid-triphosphat
Del/Ins	Deletion/Insertion
dGTP	Desoxyguanin-triphosphat
D-loop	displacement loop
DMSO	Dimethylsulfat
DNA	Desoxyribonukleinsäure
dNTP	2'-Desoxyribonukleosid-5'-triphosphat
EDTA	Ethylendiamintetraessigsäure
ETAS	extended termination associated sequence
fmol	femtomol
<i>g</i>	Erdbeschleunigung
G	Guanin
HSP	H-Strang Promotor
Jaf	Jafarabadi (Flußbüffel)
kb	Kilo-Basenpaare
kDa	Kilo-Dalton
LSP	L-Strang Promotor
M	Molar
ME	minimum evolution
Med	Mediterraneo (Flußbüffel)
MJ	median-joining
ML	Maximum Likelihood
MP	Maximum Parsimonie
MRP	mitochondriale RNA Prozessierung



---

mtDNA	mitochondriale DNA
Mur	Murrah (Flußbüffel)
NJ	neighbor joining
NNI	nearest neighbor interchange
O <sub>H</sub>	Replikationsursprung (origin) des H-Strangs
OTU	operational taxonomic unit
PCR	Polymerase Kettenreaktion
RFLP	Restriktions Fragment Längenpolymorphismus
RNA	Ribonukleinsäure
rRNA	ribosomale RNA
SDS	Natriumdodecylsulfat
SPR	subtree pruning and regrafting
T	Thymin
T <sub>amb</sub>	Raumtemperatur
TBE	Tris-Borat-EDTA
TBR	tree bisection reconnection
TE	Tris-EDTA
TEMED	N, N, N', N'-Tetramethylethyldiamin
Tris	Tris-(hydroxymethyl)-Aminomethan
tRNA	Transport RNA
v/v	Volumen per Volumen
w/v	Gewicht per Volumen









```

*****
Jaf30+25      CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCACCTTGAGCACCAGCATAAATG 713
Jaf32+40+19+01+04+27+23 CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCACCTTGAGCACCAGCATAAATG 713
Jaf66         CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
Jaf49         CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
Mur314+...+Jaf02+...+MedC8+D6 CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
Mur110+105+183+117+267 CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 714
Mur182        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 710
Mur481        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 715
Mur310+IZ244+MedC7+D2 CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
Mur181        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
Mur304        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
Mur497        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
MurWB1649    CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 711
MurIZ369+IZ287+IZ364+IZ368 CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
MurIZ192+IZ358+Iz278+MedC3 CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 714
Med4          CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
MedD4        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
MedC2        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
MedC1+D9    CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 714
MedD1        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
MedC5        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
MedD7        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
MedD10       CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 714
MedD3        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
MedD5        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
MedB5        CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 713
Car352       CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 716
Car336       CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 715
Car320+348+335+341 CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 716
Car327       CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 714
Car315+366+359+325+380+349 CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 715
Car342       CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 716
Car344+375  CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 715
Car377       CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 714
Car330       CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 715
Car370       CAGCTATGGCCGTCAAAGGCCCGACCCGGAGCATAAATTGTAGCTGGACTT-AACTGCATCTTGAGCACCAGCATAAATG 714
ruler        .....650.....660.....670.....680.....690.....700.....710.....720

```

```

*****
Jaf30+25      GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
Jaf32+40+19+01+04+27+23 GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
Jaf66         GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-GAACCC 789
Jaf49         GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGCCATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
Mur314+...+Jaf02+...+MedC8+D6 GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
Mur110+105+183+117+267 GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 791
Mur182        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 788
Mur481        GTAGGCCATTGGGACATTACAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AA-CCC 787
Mur310+IZ244+MedC7+D2 GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AA-CCC 789
Mur181        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AA-CCC 789
Mur304        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
Mur497        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
MurWB1649    GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 788
MurIZ369+IZ287+IZ364+IZ368 GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 791
MurIZ192+IZ358+Iz278+MedC3 GTAGGCCATTGGGACATTACAG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 791
Med4          GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AA-CCC 789
MedD4        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AA-CCC 789
MedC2        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
MedC1+D9    GTAGGCCATTGGGACATTACAG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 791
MedD1        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
MedC5        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
MedD7        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AA-CCC 789
MedD10       GTAGGCCATTGGGACATTACAG-TTAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-CCCCATC-AAACCC 790
MedD3        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
MedD5        GTAGGCCGTGGGACATTACGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 790
MedB5        GTAGGCCGTGGGACAA--CGG-TTAAATGGTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 787
Car352       GTAGGCCATTGGGACATTGCAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 793
Car336       GTAGGCCATTGGGACATTGCAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AA-CCC 791
Car320+348+335+341 GTAGGCCATTGGGACATTGCAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 794
Car327       GTAGGCCATTGGGACATTGCAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 791
Car315+366+359+325+380+349 GTAGGCCATTGGGACATTGCAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AA-CCC 791
Car342       GTAGGCCATTGGGACATTGCAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AA-CCC 792
Car344+375  GTAGGCCATTGGGACATTGCAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 793
Car377       GTAGGCCATTGGGACATTGCAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AA-CCC 790
Car330       GTAGGCCATTGGGACATTGCAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 792
Car370       GTAGGCCATTGGGACATTGCAG-TCAAATGGTTACAGGACATAAATAATATTATTCATTCTTCCCCCCC-TCCTATC-AAACCC 791
ruler        .....730.....740.....750.....760.....770.....780.....790.....800

```

```

**** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** *****
Jaf30+25          CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
Jaf32+40+19+01+04+27+23 CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
Jaf66            CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 867
Jaf49            CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
Mur314+...+Jaf02+...+MedC8+D6 CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
Mur110+105+183+117+267 CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 869
Mur182           CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 866
Mur481           CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 866
Mur310+IZ244+MedC7+D2 CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
Mur181           CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
Mur304           CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 869
Mur497           CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
MurWB1649       CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 866
MurIZ369+IZ287+IZ364+IZ368 CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 869
MurIZ192+IZ358+Iz278+MedC3 CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 870
Med4             CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 867
MedD4           CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
MedC2           CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
MedC1+D9        CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 869
MedD1           CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
MedD5           CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
MedD7           CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
MedD10          CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 869
MedD3           CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
MedD5           CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 868
MedB5           CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 865
Car352          CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 871
Car336          CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 871
Car320+348+335+341 CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 872
Car327          CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 871
Car315+366+359+325+380+349 CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 870
Car342          CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 871
Car344+375      CCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 871
Car377          CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 869
Car330          CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 872
Car370          CCCCC--TTAAATATTTACCACCATTTTAAACAGCCTTCTCCCTAGATGCTTATTCGAATTTATCACACTTCCAATACTCA 870
ruler           .....810.....820.....830.....840.....850.....860.....870.....880

```

```

*** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** *****
Jaf30+25          AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
Jaf32+40+19+01+04+27+23 AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
Jaf66            AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 911
Jaf49            AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
Mur314+...+Jaf02+...+MedC8+D6 AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
Mur110+105+183+117+267 AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 913
Mur182           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 910
Mur481           AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 910
Mur310+IZ244+MedC7+D2 AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
Mur181           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
Mur304           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCA- 912
Mur497           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
MurWB1649       AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 910
MurIZ369+IZ287+IZ364+IZ368 AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 913
MurIZ192+IZ358+Iz278+MedC3 AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 914
MedC4           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 911
MedD4           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
MedC2           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
MedC1+D9        AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 913
MedD1           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
MedC5           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
MedD7           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
MedD10          AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 913
MedD3           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
MedD5           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 912
MedB5           AATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAAACCAT 909
Car352          AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 915
Car336          AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 915
Car320+348+335+341 AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 916
Car327          AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 915
Car315+366+359+325+380+349 AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 914
Car342          AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 915
Car344+375      AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 915
Car377          AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 913
Car330          AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 916
Car370          AATCAGCACTCCAAATAAAGTAACTATATAAGCACCCAAACCCAT 914
ruler           .....890.....900.....910.....920....

```





```

*      ***  **      **      ***      *  *****  *****  *  *****  *****  *****  **
Jaf30+25      GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
Jaf32+40+19+01+04+27+23  GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
Jaf66      GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
Jaf49      GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
Mur314+...+Jaf02+...+MedC8+D6  GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
Mur110+105+183+117+267  GGTCTTACC-ACTCCGAAT---GGGGGATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 236
Mur182      GGTCTTACC-ACTCCGAAT---GGGGAACGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
Mur481      GGCCTTACC-ACTCCGAAT---GGGGGAACGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 237
Mur310+IZ244+MedC7+D2  GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
Mur181      GGTCTTACC-ACTCCGAAT---GGGGGATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
Mur304      GGTCTTACC-ACTCCGAAT---GGGGAACGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
Mur497      GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MurWB1649  GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MurIZ369+IZ287+IZ364+IZ368  GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MurIZ192+IZ358+Iz278+MedC3  GGTCTTACC-ACTCCGAAT---GGGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 236
MedC4      GGTCTTACC-ACTCCGAAT---GGGGGATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MedD4      GGTCTTACC-ACTCCGAAT---GGGGGATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MedC2      GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MedC1+D9  GGTCTTACC-ACTCCGAAT---GGGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 236
MedD1      GGTCTTACC-ACTCCGAAT---GGGGAACGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MedC5      GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MedD7      GGTCTTACC-ACTCCGAAT---GGGGGATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MedD10     GGTCTTACC-ACTCCGAAT---GGGGGATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MedD3      GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MedD5      GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
MedB5      GGTCTTACC-ACTCCGAAT---GGGGAATGTACATAAACATTAATGTAAACAAGGACATAAATATGTATAAGTACATTACAT 235
Car352     GACCCCTAC-TACTCCGAATGGGGGGGGGGGACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 238
Car336     GACCCCTAC-TACTCCGAATGGGGGGGGGGGACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 237
Car320+348+335+341  GACCCCTAC-TACTCCGAATGGGGGGGGGGGACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 238
Car327     GACCCCTAC-TACTCCGAAT---GGGGGGGGGGGACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 236
Car315+366+359+325+380+349  GACCCCTAC-TACTCCGAAT---GGGGGGGGGGGACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 237
Car342     GACCCCTAC-TACTCCGAATGGGGGGGGGGGACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 238
Car344+375  GACCCCTAC-TACTCCGAAT---GGGGGGGGGGGACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 237
Car377     GACCCCTAC-TACTCCGAAT---GGGGGGGGGGGACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 236
Car330     GACCCCTAC-TACTCCGAAT---GGGGGGGGGGGACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 237
Car370     GACCCCTAC-TACTCCGAAT---GGGGGGGGGGGACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 236
Bos_taurus  GAA-NTACCTAC--GCAA--GGGGTAATGTACATAAACATTAATGTAAATAAGGACATAAATATGTATAAGTACATTACAT 225
ruler      .....170.....180.....190.....200.....210.....220.....230.....240

```

```

*****  *****  *****  ***  *  *  *  *****  *****  ***  *  *  *****
Jaf30+25      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
Jaf32+40+19+01+04+27+23  TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
Jaf66      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
Jaf49      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
Mur314+...+Jaf02+...+MedC8+D6  TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
Mur110+105+183+117+267  TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 316
Mur182      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
Mur481      TATATGCCCCATGCGTATAAGCAAGTACATAAAGCATGCATGATAGTACATAGTACATTCGATTGTTAATCGTACATAGCG 317
Mur310+IZ244+MedC7+D2  TATATGCCCCATGCGTATAAGCAAGTACATAAAGCATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
Mur181      TATATGCCCCATGCGTATAAGCAAGTACATAAGACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
Mur304      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
Mur497      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MurWB1649  TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MurIZ369+IZ287+IZ364+IZ368  TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MurIZ192+IZ358+Iz278+MedC3  TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 316
MedC4      TATATGCCCCATGCGTATAAGCAAGTACATAAAGCATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MedD4      TATATGCCCCATGCGTATAAGCAAGTACATAAGACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MedC2      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MedC1+D9  TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 316
MedD1      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MedC5      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MedD7      TATATGCCCCATGCGTATAAGCAAGTACATAGACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MedD10     TATATGCCCCATGCGTATAAGCAAGTACATAAAGCATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 316
MedD3      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MedD5      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
MedB5      TATATGCCCCATGCGTATAAGCAAGTACATAAAACATGCATGATAGTACATAGTACATATAAATTTGATCGTACATAGCA 315
Car352     TATATGCCCCATGCATATAAGCGGGTATACAAACATGCATGATAGTACATAGTACATTCGAATTTGATCGTACATAGTG 318
Car336     TATATGCCCCATGCATATAAGCGGGTATACAAACATGCATGATAGTACATAGTACATTCGAATTTGATCGTACATAGTG 317
Car320+348+335+341  TATATGCCCCATGCATATAAGCGGGTATACAAACATGCATGATAGTACATAGTACATTCGAATTTGATCGTACATAGTG 318
Car327     TATATGCCCCATGCATATAAGCGGGTATACAAACATGCATGATAGTACATAGTACATTCGAATTTGATCGTACATAGTG 316
Car315+366+359+325+380+349  TATATGCCCCATGCATATAAGCGGGTATACAAACATGCATGATAGTACATAGTACATTCGAATTTGATCGTACATAGTG 317
Car342     TATATGCCCCATGCATATAAGCGGGTATACAAACATGCATGATAGTACATAGTACATTCGAATTTGATCGTACATAGTG 318
Car344+375  TATATGCCCCATGCATATAAGCGGGTATACAAACATGCATGATAGTACATAGTACATTCGAATTTGATCGTACATAGTG 317
Car377     TATATGCCCCATGCATATAAGCGGGTATACAAACATGCATGATAGTACATAGTACATTCGAATTTGATCGTACATAGTG 316
Car330     TATATGCCCCATGCATATAAGCGGGTATACAAACATGCATGATAGTACATAGTACATTCGAATTTGATCGTACATAGTG 317
Car370     TATATGCCCCATGCATATAAGCGGGTACACAAACATGCATGATAGTACATAGTACATTCGAATTTGATCGTACATAGTG 316
Bos_taurus  TATATGCCCCATGCATATAAGCAAGTACAT--GACCTCTATAGCAGTACATAAATATGTATAAATTTGACTGTACATAGTA 304
ruler      .....250.....260.....270.....280.....290.....300.....310.....320

```







```

*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
Jaf30+25      ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
Jaf32+40+19+01+04+27+23 ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
Jaf66         ATC - GAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 857
Jaf49         ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
Mur314+...+Jaf02+...+MedC8+D6 ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
Mur110+105+183+117+267 ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 859
Mur182        ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 856
Mur481        ATC - AA - CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 856
Mur310+IZ244+MedC7+D2 ATC - AA - CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
Mur181        ATC - AA - CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
Mur304        ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 859
Mur497        ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
MurWB1649    ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 856
MurIZ369+IZ287+IZ364+IZ368 ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 859
MurIZ192+IZ358+Iz278+MedC3 ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 860
MedC4         ATC - AA - CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 857
MedD4         ATC - AA - CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
MedC2         ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
MedC1+D9     ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 859
MedD1         ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
MedC5         ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
MedD7         ATC - AA - CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
MedD10        ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 859
MedD3         ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
MedD5         ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 858
MedB5         ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 855
Car352        ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 861
Car336        ATC - AA - CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 861
Car320+348+335+341 ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 862
Car327        ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 861
Car315+366+359+325+380+349 ATC - AA - CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 860
Car342        ATC - AA - CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 861
Car344+375    ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 861
Car377        ATC - AA - CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 859
Car330        ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 862
Car370        ATC - AAA CCCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 860
Bos_taurus   AAA - AAT TCCCTC -- TTAAATATTTACCACCA TTTTAAACACGCTTCTCCCTAGATGCTTATTGGAATTTATCACACTT 850
ruler        .....810.....820.....830.....840.....850.....860.....870.....880

```

```

***** *      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
Jaf30+25      CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
Jaf32+40+19+01+04+27+23 CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
Jaf66         CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 911
Jaf49         CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
Mur314+...+Jaf02+...+MedC8+D6 CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
Mur110+105+183+117+267 CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 913
Mur182        CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 910
Mur481        CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAATCCAT----- 910
Mur310+IZ244+MedC7+D2 CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
Mur181        CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
Mur304        CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
Mur497        CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
MurWB1649    CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 910
MurIZ369+IZ287+IZ364+IZ368 CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 913
MurIZ192+IZ358+Iz278+MedC3 CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 914
MedC4         CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 911
MedD4         CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
MedC2         CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
MedC1+D9     CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 913
MedD1         CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
MedD7         CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
MedD10        CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 913
MedD3         CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
MedD5         CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 912
MedB5         CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACCTAAACCAT----- 909
Car352        CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAACCAT----- 915
Car336        CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAACCAT----- 915
Car320+348+335+341 CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAACCAT----- 916
Car327        CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAACCAT----- 915
Car315+366+359+325+380+349 CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAACCAT----- 914
Car342        CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAACCAT----- 915
Car344+375    CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAACCAT----- 915
Car377        CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAACCAT----- 913
Car330        CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAACCAT----- 916
Car370        CCAATACTCAAATCAGCACTCCAAACAAAGTAACTATATAAGCACCAACCAT----- 914
Bos_taurus   CCAATACTCAAATTAGCACTCCAAACAAAGTAACTATATAAGCACGAGGCCCTCCCTC----- 910
ruler        .....890.....900.....910.....920.....930.....940

```

## Anhang 4: Tabelle der paarweisen genetischen Distanzen (unkorrigierte Werte)

	Jaf30	Jaf32	Jaf66	Jaf49	Mur314	Mur110
Jaf32+40+19+01+04+27+23	0,0010965					
Jaf66	0,004386	0,0032895				
Jaf49	0,0054825	0,004386	0,0032937			
Mur314+..+Jaf02+..+MedC8+D6	0,0032895	0,002193	0,0010965	0,002193		
Mur110+105+183+117+267	0,0153876	0,01649	0,0154121	0,016486	0,0142931	
Mur182	0,0055213	0,0044269	0,0055238	0,0066404	0,0044293	0,0166357
Mur481	0,0650926	0,0639996	0,0629667	0,0640154	0,0618082	0,0629522
Mur310+lZ244+MedC7+D2	0,0087841	0,0076888	0,0087922	0,0098879	0,0076926	0,0154227
Mur181	0,0076842	0,0065889	0,0076922	0,008788	0,0065927	0,0143165
Mur304	0,0032948	0,0021995	0,0032948	0,004399	0,0021995	0,0165342
Mur497	0,004386	0,0032895	0,0021972	0,0032895	0,0010965	0,0131902
MurWB1649	0,0033225	0,0022096	0,0011131	0,002193	0	0,014327
MurIz369+lZ287+lZ364+lZ368	0,0032859	0,0021906	0,0010953	0,0021953	0	0,014306
MurIz192+lZ358+lZ278+MedC3	0,0142328	0,0153353	0,0142661	0,0153401	0,0131471	0,0010929
MedC4	0,0087797	0,0076831	0,0087919	0,00988	0,0076869	0,0154066
MedD4	0,0109871	0,0098917	0,0109951	0,0120908	0,0098955	0,0176252
MedC2	0,002193	0,0010965	0,002193	0,0032895	0,0010965	0,0153872
MedC1+D9	0,0142388	0,0153424	0,014272	0,0153424	0,0131518	0,0010941
MedD1	0,006579	0,0054825	0,004386	0,0054825	0,0032895	0,0175973
MedC5	0,004386	0,0032895	0,0021972	0,0032895	0,0010965	0,0131902
MedD7	0,0098794	0,0087841	0,0098875	0,0109832	0,0087879	0,0165239
MedD10	0,0208404	0,0219427	0,0208819	0,0219516	0,0197545	0,0164832
MedD3	0,0032895	0,002193	0,0032895	0,004386	0,002193	0,01649
MedD5	0,0032895	0,002193	0,0032937	0,004386	0,002193	0,0164813
MedB5	0,0043898	0,0032933	0,0021968	0,0033018	0,0011003	0,0154572
Car352	0,0702992	0,0714157	0,0703878	0,0714157	0,0692299	0,0659492
Car336	0,0703633	0,0714743	0,0704476	0,0714834	0,0692907	0,0660374
Car320+348+335+341	0,0703229	0,0714382	0,0704115	0,0714428	0,0692548	0,0659872
Car327	0,0702645	0,0713653	0,0703374	0,0713744	0,0691795	0,0659354
Car315+366+359+325+380+349	0,0703396	0,0714518	0,070424	0,0714564	0,0692658	0,0659994
Car342	0,0703806	0,0714995	0,0704728	0,071504	0,0693159	0,0660442
Car344+375	0,0702818	0,0713905	0,0703626	0,071395	0,0692047	0,0659423
Car377	0,0702985	0,0714041	0,0703751	0,0714086	0,0692157	0,0659544
Car330	0,0703056	0,071413	0,0703863	0,0714221	0,0692296	0,0659804
Car370	0,0691462	0,0702481	0,069215	0,0702527	0,0680599	0,0648035
<i>Bos taurus</i>	0,1546523	0,1557563	0,1536657	0,1546426	0,153528	0,1479329

	Mur182	Mur481	Mur310	Mur181	Mur304	Mur497
Mur481	0,0610672					
Mur310+lZ244+MedC7+D2	0,0077053	0,0650175				
Mur181	0,006592	0,0639181	0,0010965			
Mur304	0,0022063	0,0617777	0,0076832	0,0065879		
Mur497	0,0055413	0,0629178	0,0087926	0,0076927	0,0033037	
MurWB1649	0,0044626	0,061996	0,0077417	0,0066417	0,0022161	0,0010965
MurIZ369+lZ287+lZ364+lZ368	0,0044245	0,0618025	0,0076889	0,0065902	0,0021971	0,0011
MurIZ192+lZ358+lZ278+MedC3	0,0154436	0,0616572	0,0164425	0,0153484	0,0153547	0,012053
MedC4	0,0077278	0,0651088	0,0010965	0,002193	0,0076969	0,0087834
MedD4	0,0099158	0,0672101	0,002193	0,0032895	0,0098815	0,0109955
MedC2	0,003323	0,0629013	0,0065935	0,0054936	0,0011042	0,002193
MedC1+D9	0,0154812	0,0617504	0,0164754	0,0153767	0,0153877	0,0120566
MedD1	0,0055201	0,0628942	0,0109785	0,0098786	0,0032859	0,004386
MedC5	0,0055413	0,0629178	0,0087926	0,0076927	0,0033037	0,002193
MedD7	0,0088117	0,0661281	0,0010965	0,002193	0,0087785	0,0098879
MedD10	0,0232132	0,0682324	0,0252287	0,0241346	0,0219609	0,0186603
MedD3	0,0044269	0,0617916	0,0076888	0,0065889	0,0021995	0,0032895
MedD5	0,0044342	0,0640062	0,0076889	0,0065889	0,0022037	0,0032895
MedB5	0,0055326	0,0630874	0,0088004	0,0077004	0,0032987	0,0022053
Car352	0,068646	0,0395631	0,0703395	0,0692434	0,0703816	0,0703228
Car336	0,0685985	0,0394997	0,0702769	0,069184	0,0703467	0,0703915
Car320+348+335+341	0,0686635	0,0395702	0,0703586	0,0692637	0,0704052	0,0703511
Car327	0,0684858	0,0395292	0,0702058	0,0691129	0,0702364	0,0702815
Car315+366+359+325+380+349	0,0685813	0,0394928	0,070258	0,0691639	0,0703234	0,0703633
Car342	0,0686319	0,0395504	0,0702992	0,0692063	0,0703723	0,0704121
Car344+375	0,068613	0,0395127	0,0703176	0,0692215	0,0703564	0,0703021
Car377	0,0685306	0,0395421	0,0702168	0,0691215	0,0702744	0,0703143
Car330	0,0685365	0,03948	0,0702469	0,0691552	0,0702854	0,0703304
Car370	0,0673698	0,0384122	0,0690928	0,0679987	0,0691148	0,0691586
<i>Bos taurus</i>	0,1558824	0,1695959	0,1537172	0,154855	0,1548767	0,1523953

	MurWB164	MurIZ369	MurIZ192	MedC4	MedD4	MedC2
	9					
MurIZ369+lZ287+lZ364+lZ368	0					
MurIZ192+lZ358+lZ278+MedC3	0,0131642	0,0131611				
MedC4	0,0077199	0,0076833	0,0164473			
MedD4	0,0099604	0,0098894	0,0186383	0,0032942		
MedC2	0,0011131	0,0010953	0,0142412	0,0065866	0,0087964	
MedC1+D9	0,013169	0,0131659	0	0,0164605	0,0186757	0,0142471
MedD1	0,0032895	0,0032859	0,0164294	0,0109765	0,0131814	0,004386
MedC5	0,0010965	0,0011	0,012053	0,0087834	0,0109955	0,002193
MedD7	0,0088373	0,008783	0,017545	0,002193	0,0032895	0,0076888
MedD10	0,0198049	0,0197707	0,015338	0,0252522	0,0274245	0,0208485
MedD3	0,0022096	0,0021906	0,0153353	0,0076831	0,0098917	0,0010965
MedD5	0,0022096	0,0021906	0,0153353	0,0076832	0,0098918	0,0010965
MedB5	0,0011003	0,0010991	0,0143063	0,0088044	0,0110033	0,0021968
Car352	0,0694907	0,0692205	0,0669011	0,070346	0,0725351	0,0703228
Car336	0,0695493	0,0692812	0,0668885	0,0703733	0,0724627	0,0703826
Car320+348+335+341	0,0695153	0,0691543	0,0669377	0,0703651	0,0725518	0,0703465
Car327	0,0694366	0,0691701	0,066776	0,0703022	0,0723994	0,0702724
Car315+366+359+325+380+349	0,0695248	0,0692564	0,0668523	0,0703542	0,0724462	0,0703589
Car342	0,0695764	0,0693064	0,0669041	0,0703952	0,072485	0,0704078
Car344+375	0,0694636	0,0691041	0,066886	0,0703241	0,0725132	0,0702976
Car377	0,069473	0,0692062	0,0668004	0,070313	0,0724074	0,07031
Car330	0,0694883	0,0692202	0,0668278	0,0703433	0,0724381	0,0703213
Car370	0,0683173	0,0680471	0,0656455	0,0691885	0,0712887	0,069154
<i>Bos taurus</i>	0,1538515	0,1544849	0,1466926	0,1547675	0,1559255	0,1546238

	MedC1	MedD1	MedC5	MedD7	MedD10	MedD3
MedD1	0,0164377					
MedC5	0,0120566	0,004386				
MedD7	0,0175778	0,0120738	0,0098879			
MedD10	0,0153621	0,0208396	0,0186603	0,026331		
MedD3	0,0153424	0,0054825	0,0032895	0,0087841	0,0219427	
MedD5	0,0153424	0,0054825	0,0032895	0,0087841	0,0219468	0,002193
MedB5	0,0143123	0,0043898	0,0022053	0,0098996	0,0209308	0,0032933
Car352	0,0669165	0,0714157	0,0703228	0,0692439	0,0713334	0,0692299
Car336	0,0669986	0,0714743	0,0703915	0,069184	0,0713209	0,0692818
Car320+348+335+341	0,0669531	0,0714382	0,0703511	0,0692642	0,0713697	0,0692502
Car327	0,0668862	0,0713653	0,0702815	0,069109	0,0712128	0,0691703
Car315+366+359+325+380+349	0,066962	0,0714518	0,0703633	0,0691639	0,071285	0,0692615
Car342	0,0670137	0,0714995	0,0704121	0,0692063	0,071332	0,0693115
Car344+375	0,0669013	0,0713905	0,0703021	0,069222	0,0713228	0,0692001
Car377	0,0669103	0,0714041	0,0703143	0,0691215	0,071238	0,0692113
Car330	0,0669379	0,071413	0,0703304	0,0691513	0,0712598	0,0692205
Car370	0,0657549	0,0702481	0,0691586	0,0679948	0,0700785	0,0680554
<i>Bos taurus</i>	0,1466459	0,1557688	0,1523953	0,1537283	0,1490394	0,1557563

	MedD5	MedB5	Car352	Car336	Car320	Car327
MedB5	0,0033018					
Car352	0,0714157	0,0694583				
Car336	0,0714745	0,0695224	0			
Car320+348+335+341	0,0714382	0,0694829	0	0		
Car327	0,0713653	0,0694219	0	0	0	
Car315+366+359+325+380+349	0,071452	0,0694978	0	0	0	0
Car342	0,0714997	0,0695406	0	0	0	0
Car344+375	0,0713905	0,0694401	0	0	0	0
Car377	0,0714043	0,0694549	0	0	0	0
Car330	0,071413	0,0694648	0	0	0	0
Car370	0,0702481	0,0682944	0,0010963	0,0010917	0,0010951	0,0010929
<i>Bos taurus</i>	0,1557384	0,155147	0,171527	0,170699	0,1724376	0,1715515

	Car315	Car342	Car344	Car377	Car330	Car370
Car342	0					
Car344+375	0	0				
Car377	0	0	0			
Car330	0	0	0	0		
Car370	0,0010929	0,0010917	0,0010963	0,0010941	0,0010917	
<i>Bos taurus</i>	0,1706288	0,1706868	0,1723814	0,1705707	0,1716085	0,1703434



## Anhang 5: Tabelle der paarweisen genetischen Distanzen (Tamura-Nei-Modell mit $\Gamma$ -Parameter)

	Jaf30	Jaf32	Jaf66	Jaf49	Mur314	Mur110
Jaf32+40+19+01+04+27+23	0,0011231					
Jaf66	0,0046536	0,0034176				
Jaf49	0,0057234	0,0045327	0,0033701			
Mur314+...+Jaf02+...+MedC8+D6	0,0034176	0,0022394	0,0011232	0,0022307		
Mur110+105+183+117+267	0,0179048	0,0193457	0,0178775	0,0191847	0,0164169	
Mur182	0,0057278	0,0045833	0,005756	0,0069235	0,0045544	0,0191539
Mur481	0,12547	0,1217276	0,1182728	0,1205423	0,1145169	0,1184287
Mur310+IZ244+MedC7+D2	0,0095532	0,0083187	0,0095724	0,0107587	0,0082716	0,0178855
Mur181	0,0082609	0,0070264	0,0083254	0,0094546	0,0070249	0,016433
Mur304	0,0034017	0,0022799	0,0034231	0,0045468	0,0022461	0,0194062
Mur497	0,0045744	0,0033961	0,0022437	0,0033978	0,0011162	0,0149696
MurWB1649	0,0034539	0,0022568	0,0011406	0,0022307	0	0,0164602
MurIZ369+IZ287+IZ364+IZ368	0,0034138	0,0022369	0,0011219	0,0022332	0	0,0164334
MurIZ192+IZ358+IZ278+MedC3	0,0164625	0,0178342	0,0163804	0,0177387	0,0149876	0,0011196
MedC4	0,0094944	0,0083038	0,0094585	0,0106988	0,0081993	0,0176967
MedD4	0,01207	0,010761	0,0121574	0,0133082	0,010778	0,0206904
MedC2	0,0022393	0,0011162	0,0023014	0,0033657	0,0011232	0,0178441
MedC1+D9	0,0164719	0,0178451	0,0163891	0,0177428	0,0149947	0,0011208
MedD1	0,0069328	0,0057421	0,0045326	0,0057593	0,0033977	0,0207673
MedC5	0,0045744	0,0033961	0,0022437	0,0033978	0,0011162	0,0149696
MedD7	0,0108498	0,0095542	0,0109347	0,0120703	0,0095693	0,0193985
MedD10	0,0258414	0,0273553	0,0256813	0,0272186	0,0241314	0,0194601
MedD3	0,003396	0,0022729	0,0034176	0,0045327	0,0022394	0,0193457
MedD5	0,0033656	0,0022307	0,0034346	0,0045086	0,0022395	0,0191186
MedB5	0,0045604	0,0033696	0,0022435	0,0033747	0,0011085	0,0177674
Car352	0,1413552	0,1453434	0,1412489	0,1441375	0,1372353	0,1270854
Car336	0,1416487	0,1456316	0,1415375	0,1444533	0,1375155	0,1274311
Car320+348+335+341	0,1414344	0,1454184	0,1413262	0,1442321	0,1373166	0,1272126
Car327	0,141421	0,1453625	0,1412651	0,144186	0,1372463	0,1272111
Car315+366+359+325+380+349	0,1415692	0,1455563	0,14146	0,1443584	0,1374339	0,1273036
Car342	0,1416158	0,1456221	0,1415295	0,1444291	0,1375101	0,1273683
Car344+375	0,1413875	0,1453522	0,1412563	0,1441611	0,1372401	0,1271478
Car377	0,1415232	0,1454912	0,1413912	0,1442884	0,1373583	0,1272393
Car330	0,1414671	0,1454278	0,1413341	0,1442562	0,137322	0,1272754
Car370	0,1369779	0,1409235	0,136975	0,1396855	0,1329767	0,1228695
<i>Bos taurus</i>	0,6301826	0,6413488	0,6151845	0,6202641	0,6118801	0,5352505

	Mur182	Mur481	Mur310	Mur181	Mur304	Mur497
Mur481	0,1097709					
Mur310+lZ244+MedC7+D2	0,0081425	0,125468				
Mur181	0,0069029	0,1219681	0,0011162			
Mur304	0,0022393	0,1144541	0,008312	0,0070252		
Mur497	0,0057496	0,1182296	0,0095645	0,0082719	0,0034114	
MurWB1649	0,0045898	0,115117	0,0083284	0,0070816	0,0022636	0,0011162
MurIZ369+lZ287+lZ364+lZ368	0,0045493	0,1144834	0,0082673	0,007022	0,0022436	0,0011199
MurIZ192+lZ358+lZ278+MedC3	0,0176327	0,1141967	0,0192893	0,0178563	0,017861	0,0135539
MedC4	0,0081461	0,1241872	0,0011046	0,0022307	0,0083195	0,0094993
MedD4	0,0106317	0,1320959	0,0022393	0,0033653	0,0107491	0,0120824
MedC2	0,0033969	0,1180736	0,0070317	0,005785	0,0011242	0,0022394
MedC1+D9	0,017683	0,1145024	0,0193332	0,0178927	0,0179068	0,013559
MedD1	0,0057087	0,1167421	0,0120988	0,0107471	0,0033619	0,0046073
MedC5	0,0057496	0,1182296	0,0095645	0,0082719	0,0034114	0,002273
MedD7	0,0094171	0,1298852	0,001123	0,0022391	0,0095475	0,0108622
MedD10	0,0288456	0,1370512	0,0323719	0,0305819	0,0273845	0,0224685
MedD3	0,0045833	0,1145119	0,0083187	0,0070264	0,0022799	0,0033961
MedD5	0,0045606	0,1204239	0,0082013	0,0069435	0,0022418	0,0033657
MedB5	0,0057221	0,1173576	0,0094686	0,0082104	0,0033753	0,0022435
Car352	0,1317079	0,0582974	0,1408762	0,1367076	0,1411899	0,1415959
Car336	0,1316198	0,0581825	0,1406717	0,1365245	0,1411392	0,1419107
Car320+348+335+341	0,131757	0,0583085	0,1409373	0,1367713	0,1412681	0,1416924
Car327	0,1313367	0,0582735	0,140584	0,1364376	0,1408665	0,141645
Car315+366+359+325+380+349	0,1315718	0,0581717	0,1406118	0,1364619	0,1410624	0,141814
Car342	0,131655	0,0582644	0,1406555	0,1365094	0,1411322	0,1418862
Car344+375	0,1316739	0,0582162	0,1408943	0,1367246	0,1411983	0,1416198
Car377	0,1314893	0,0582899	0,1405689	0,1364153	0,1409933	0,1417423
Car330	0,1314194	0,0581555	0,1406263	0,1364836	0,1409358	0,1417168
Car370	0,1274574	0,0561879	0,1363564	0,1323744	0,1365958	0,1371805
<i>Bos taurus</i>	0,6373525	0,9497339	0,6145926	0,6339202	0,6249998	0,5937451

	MurWB164	MurIZ369	MurIZ192	MedC4	MedD4	MedC2
	9					
MurIZ369+lZ287+lZ364+lZ368	0					
MurIZ192+lZ358+lZ278+MedC3	0,0150083	0,0150055				
MedC4	0,008236	0,008195	0,0190704			
MedD4	0,0108577	0,0107709	0,0221796	0,0033772		
MedC2	0,0011406	0,0011219	0,0163472	0,007004	0,0094625	
MedC1+D9	0,0150155	0,0150126	0	0,0190886	0,0222293	0,0163558
MedD1	0,0033977	0,0033938	0,019287	0,0120581	0,0146603	0,0045326
MedC5	0,0011162	0,0011199	0,0135539	0,0094993	0,0120824	0,0022394
MedD7	0,0096294	0,0095635	0,0208832	0,0022394	0,0034295	0,0082672
MedD10	0,0241991	0,0241546	0,0180273	0,032153	0,0355487	0,0256296
MedD3	0,0022568	0,0022369	0,0178342	0,0083038	0,010761	0,0011162
MedD5	0,0022572	0,0022371	0,0176054	0,0081794	0,0106627	0,0011046
MedB5	0,0011085	0,0011073	0,0163163	0,0094108	0,0120068	0,0022435
Car352	0,138094	0,137182	0,1298262	0,139665	0,1474611	0,1410088
Car336	0,13837	0,1374629	0,1298207	0,1397914	0,1472308	0,1412974
Car320+348+335+341	0,1381745	0,1369316	0,1299493	0,139727	0,1475159	0,1410862
Car327	0,1380942	0,1371939	0,1295625	0,1396928	0,1471722	0,1410248
Car315+366+359+325+380+349	0,1382892	0,1373809	0,1296992	0,1397289	0,1471771	0,1412198
Car342	0,1383699	0,137457	0,1297834	0,1397781	0,1472098	0,1412895
Car344+375	0,1380935	0,1368548	0,1298654	0,1396783	0,1474841	0,1410161
Car377	0,1382091	0,1373053	0,1296154	0,1396805	0,1471455	0,1411507
Car330	0,1381744	0,1372696	0,1296462	0,1397407	0,1472031	0,141094
Car370	0,1338239	0,1329146	0,1252646	0,1354085	0,1429145	0,1367592
<i>Bos taurus</i>	0,614775	0,6282302	0,5244159	0,6227455	0,6338126	0,6222886

	MedC1	MedD1	MedC5	MedD7	MedD10	MedD3
MedD1	0,0193003					
MedC5	0,013559	0,0046073				
MedD7	0,0209271	0,0134104	0,0108622			
MedD10	0,0180619	0,0254447	0,0224685	0,0341333		
MedD3	0,0178451	0,0057421	0,0033961	0,0095542	0,0273553	
MedD5	0,0176155	0,005686	0,0033657	0,00945	0,0269816	0,0022307
MedB5	0,016325	0,0045431	0,0022435	0,010785	0,0255856	0,0033696
Car352	0,1299072	0,1440997	0,1415959	0,1371059	0,14591	0,1368274
Car336	0,1302436	0,1443814	0,1419107	0,1369086	0,1459099	0,1370834
Car320+348+335+341	0,1300299	0,1441772	0,1416924	0,137171	0,1460416	0,1368906
Car327	0,1299857	0,1441142	0,141645	0,1368024	0,145651	0,1368074
Car315+366+359+325+380+349	0,1301206	0,1443036	0,141814	0,1368447	0,1457802	0,13702
Car342	0,1302044	0,1443742	0,1418862	0,136896	0,1458464	0,1370935
Car344+375	0,1299459	0,1441063	0,1416198	0,1371204	0,1459756	0,1368167
Car377	0,1300372	0,1442335	0,1417423	0,1367941	0,1457147	0,1369472
Car330	0,1300691	0,1441843	0,1417168	0,1368525	0,1457165	0,1368804
Car370	0,1256675	0,1396363	0,1371805	0,1325708	0,1409562	0,1327275
<i>Bos taurus</i>	0,5240573	0,639044	0,5937451	0,6150507	0,5499665	0,6413488

	MedD5	MedB5	Car352	Car336	Car320	Car327
MedB5	0,0033853					
Car352	0,1435729	0,1367734				
Car336	0,1438616	0,1370584	0			
Car320+348+335+341	0,1436478	0,1368562	0	0		
Car327	0,1435874	0,1368258	0	0	0	
Car315+366+359+325+380+349	0,1437865	0,1369753	0	0	0	0
Car342	0,1438545	0,1370322	0	0	0	0
Car344+375	0,1435795	0,136799	0	0	0	0
Car377	0,1437192	0,136919	0	0	0	0
Car330	0,1436549	0,1368823	0	0	0	0
Car370	0,1392864	0,1324677	0,001116	0,0011112	0,0011147	0,0011124
<i>Bos taurus</i>	0,6308363	0,6265622	0,9275061	0,9204876	0,9537838	0,9293792

	Car315	Car342	Car344	Car377	Car330	Car370
Car342	0					
Car344+375	0	0				
Car377	0	0	0			
Car330	0	0	0	0		
Car370	0,0011125	0,0011113	0,0011159	0,0011137	0,0011112	
<i>Bos taurus</i>	0,919314	0,9195576	0,9535601	0,9190745	0,9296117	0,8971578

## LEBENS LAUF

### **PERSÖNLICHE DATEN**

Name, Vorname	Kierstein, Gerold
Anschrift	Springstraße 46a 37077 Göttingen
Geburtsdatum	07. Okt. 1966
Geburtsort	Essen

### **BERUFLICHER WERDEGANG**

Aug. 1973 – Jun. 1986	Grundschule und Gymnasium in Essen, Abschluß: Allgemeine Hochschulreife
Jun. 1986 – Sep. 1987	Wehrdienst, Goslar und Delmenhorst
Okt. 1989 – Feb. 1995	Hochschulstudium an der Ruhr-Universität Bochum, Fakultät für Biologie. Abschluß: Dipl.-Biol.
Feb. 1995 – Aug. 1995	Wissenschaftliche Hilfskraft am Lehrstuhl für Allgemeine Zoologie und Neurobiologie, Bochum
Nov. 1995 – Mär. 1996	Mitarbeiter im Artenschutzprojekt: „ <i>Species conservation as an integral part of forest maintenance in the Philippines</i> “ auf Negros / Panay, Philippinen. Leitung Prof. Dr. E. Curio, Bochum
Okt. 1996 -	Aufbaustudium „Agrarwissenschaften der Tropen und Subtropen“ an der Georg-August-Universität, Göttingen. Abschluß: M. Sc. Agr.
Jan. 1998 – Jul. 1998	Forschungsarbeit im Rahmen des Aufbaustudiums an der Universidade Federal do Pará, Belém, Brasilien. Leitung: Prof. Dr. P. Schneider
Aug. 1998 -	Fortführung der Forschungsarbeit für das Aufbaustudium am Tierärztlichen Institut der Georg-August-Universität, Göttingen. Leitung: Prof. Dr. Dr. B. Brenig
Jan. 1999 – Jul. 2000	Vorarbeiten für die vorliegende Dissertation
Aug. 2000 – Mär. 2001	Dissertation am Tierärztlichen Institut der Georg-August-Universität, Göttingen. Leitung: Prof. Dr. Dr. B. Brenig