

**Die molekulare Evolution  
der Hämoglobin-Genfamilie  
in *Chironomus tentans***

Dissertation  
zur Erlangung des Grades  
„Doktor der Naturwissenschaften“  
am Fachbereich Biologie der  
Johannes Gutenberg-Universität in Mainz

**Heike Friedl**  
geboren in Frankfurt am Main

Mainz 2001

---

Dekan:  
1. Berichterstatter:  
2. Berichterstatter:  
Tag der mündlichen Prüfung: 18.9.2001

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>7</b>
<b>Tabellenverzeichnis</b>	<b>9</b>
<b>Verzeichnis der Abkürzungen</b>	<b>11</b>
<b>1 Einleitung</b>	<b>13</b>
1.1 Genome und Genfamilien . . . . .	13
1.2 Mechanismen der Genduplikation . . . . .	15
1.3 Die Evolution duplizierter Gene . . . . .	17
1.4 Die Genfamilie der Globine . . . . .	18
1.5 Die Evolution der Hämoglobin-Genfamilie der Chironomiden . . . . .	20
1.6 Die Intron-/Exon-Struktur der Hämoglobin-Gene . . . . .	23
1.7 Problemstellung und Zielsetzung der Arbeit . . . . .	25
<b>2 Material und Methoden</b>	<b>27</b>
2.1 Versuchstiere . . . . .	27
2.2 Standardmethoden . . . . .	27
2.3 Konstruktion von genomischen DNA-Bibliotheken . . . . .	28
2.4 Präparation von DNA . . . . .	28
2.4.1 Präparation genomischer DNA von Zuckmückenlarven . . . . .	28
2.4.2 Präparation von Phagen-DNA . . . . .	29
2.4.3 Präparation von Plasmid-DNA . . . . .	29
2.5 Polymerase-Kettenreaktion (PCR) . . . . .	30
2.5.1 PCR aus genomischer und klonierter DNA . . . . .	30
2.5.2 Inverse-PCR . . . . .	30
2.5.3 RT-PCR (Reverse Transkriptase-Polymerasekettenreaktion) mit intronspezifischen Primern . . . . .	31
2.6 Oligonukleotidsynthese . . . . .	31
2.7 Subklonierung in Plasmidvektoren . . . . .	31
2.8 Hybridisierungsmethoden . . . . .	32
2.8.1 „Southern-Blot“ Hybridisierung . . . . .	32
2.8.2 Plauefilterhybridisierung . . . . .	33
2.8.3 Koloniefilterhybridisierung . . . . .	34
2.8.4 In situ-Hybridisierung an Polytäanchrosomen . . . . .	34

2.9	DNA-Markierung . . . . .	35
2.9.1	Random-primed oligo-labelling . . . . .	35
2.9.2	Nick Translation . . . . .	35
2.10	Automatische DNA-Sequenzierung . . . . .	35
2.11	Computerauswertung von DNA- und Proteinsequenzen . . . . .	36
2.12	Puffer und Lösungen . . . . .	37
2.13	<i>E. coli</i> -Stämme: . . . . .	39
2.14	Materialien . . . . .	40
2.15	Bezugsquellen . . . . .	40
<b>3</b>	<b>Ergebnisse</b>	<b>43</b>
3.1	Isolierung und Charakterisierung von $\lambda$ -Klonen mit Hämoglobin-Genen	43
3.2	Die chromosomale Organisation der $\lambda$ -Klone . . . . .	44
3.2.1	Die Reihenfolge der Contigs 1 und 2 auf dem Chromosom . . .	44
3.2.2	Die Orientierung des Contigs 1 . . . . .	49
3.2.3	Sequenzierungsergebnisse . . . . .	49
3.3	Die genomische Struktur der Hb-Gengruppe D . . . . .	59
3.4	Allele Hämoglobingen-Varianten . . . . .	60
3.5	Introns in Hämoglobin-Genen von <i>C. tentans</i> . . . . .	64
3.5.1	Bestimmung der Intronposition . . . . .	64
3.6	Charakterisierung eines Gens aus der Familie der Glukosetransporter	68
3.7	Repetitive Elemente im Hämoglobingen-Cluster . . . . .	73
<b>4</b>	<b>Diskussion</b>	<b>79</b>
4.1	Die Klassifizierung der Hämoglobin-Varianten von <i>C. tentans</i> . . . . .	80
4.1.1	Die Klassifizierung der Hämoglobin-Varianten auf Protein-Ebene	81
4.1.2	Die Klassifizierung der Hämoglobin-Varianten auf Gen-Ebene	89
4.2	Die Primärstruktur und Funktion der Hämoglobine von Chironomiden	92
4.2.1	Die „invarianten“ und „semi-invarianten“ Aminosäure-Positionen der Chironomiden Hämoglobine . . . . .	94
4.2.2	Die Quartärstruktur der Hämoglobine von <i>C. thummi</i> und <i>C. tentans</i> . . . . .	103
4.2.3	Die Architektur der Hämoglobin-Gengruppe in <i>C. tentans</i> . .	107
4.2.4	Die Evolution der Hämoglobin-Genstruktur . . . . .	107
4.2.5	Vergleich der paralogen Hämoglobin-Varianten von <i>C. tentans</i> : Hämoglobin-Gen-Subfamilien und „single copy“-Gene . .	117
4.3	Mechanismen der Entstehung der Hämoglobin-Genfamilie in <i>C. tentans</i>	121
4.3.1	Die Entstehung von Genen durch Duplikation am Beispiel der Hämoglobin-Gene Cte 12 und Cte 7A . . . . .	121
4.3.2	Duplikation eines 3' nicht-translatierten Bereichs durch Trans- position? . . . . .	128
4.3.3	Konzertierte Sequenz-Evolution von Hämoglobin-Varianten . .	130

4.4	Vergleichende Analyse orthologer Hämoglobin-Gen- gruppen der Chironomiden . . . . .	135
4.4.1	Die VIIB-Region der Arten <i>C. tentans</i> , <i>C. thummi</i> und <i>C. piger</i>	136
4.4.2	Die IIB-Region der Arten <i>C. tentans</i> , <i>C. thummi</i> und <i>C. me-</i> <i>lanotus</i> . . . . .	140
4.5	Das evolutionäre „Schicksal“ von Genduplikaten . . . . .	143
4.6	Aussichten und offene Fragen . . . . .	150
<b>5</b>	<b>Zusammenfassung</b>	<b>153</b>
	<b>Literaturverzeichnis</b>	<b>157</b>
	<b>Anhang</b>	<b>173</b>



# Abbildungsverzeichnis

1.1	Definition der Begriffe „paralog“ und „ortholog“ . . . . .	14
1.2	Mechanismen der Duplikation von DNA-Abschnitten . . . . .	17
1.3	Stammbaum der Vertebraten-Globine . . . . .	19
1.4	Die Lokalisation der Hämoglobin-Gene im Genom von Zuckmücken . . . . .	23
3.1	Übersicht über die Lambda-Klone aus dem Bereich des Hämoglobinengen- Clusters D von <i>C. tentans</i> . . . . .	44
3.2	Chromosomale Anordnung der Klon-Contigs 1 und 2 . . . . .	48
3.3	Orientierung des Klon-Contigs 1 auf dem Chromosom . . . . .	50
3.4	Die Chromosomale Anordnung und Orientierung der Contigs 1 (50 kb)und 2 (24 kb) . . . . .	51
3.5	Genomische Struktur der Hämoglobin-Gengruppe D von <i>C. tentans</i> . . . . .	62
3.6	Ermittlung der Intronposition im Gen Cte N . . . . .	66
3.7	Intronposition im Hämoglobin-Gen Cte N . . . . .	67
3.8	Vergleich der Nukleotid-Sequenzen des Glukosetransport-Gens von <i>C. tentans</i> und <i>C. melanotus</i> . . . . .	71
3.9	Die Struktur des GLUT-Proteins . . . . .	72
3.10	Vorhergesagte Topologie des Cte GLUT-Proteins . . . . .	73
3.11	Struktur der Cp1-Retroposons in der Hämoglobin-Gengruppe D . . . . .	77
4.1	Gelelektrophoretische Auftrennung der von Braunitzer und Mitarbei- tern ermittelten Hämoglobin-Varianten von <i>C. thummi</i> . . . . .	81
4.2	Vergleich der Hämoglobin-Sequenzen von <i>C. tentans</i> , <i>C. thummi</i> und <i>D. melanogaster</i> . . . . .	83
4.3	Phylogenetische Bäume der Hämoglobin-Aminosäure-Sequenzen von <i>C. tentans</i> und <i>C. thummi</i> . . . . .	86
4.4	Phylogenetische Klassifizierung dimerer Hämoglobin-Varianten aus <i>C. tentans</i> in zwei Subfamilien . . . . .	88
4.5	Verwandtschaftsbeziehungen der Hämoglobin-Aminosäure-Sequenzen von <i>C. tentans</i> , <i>C. thummi</i> und <i>C. piger</i> . . . . .	90
4.6	Die Hämoglobin-Varianten von <i>C. tentans</i> und anderen Chironomiden- Spezies im Vergleich . . . . .	92
4.7	Die dreidimensionale, räumliche Struktur des Myoglobins in Diagramm- form . . . . .	93

4.8	Alignment der Aminosäure-Sequenzen der Komponenten von <i>C. tentans</i> und <i>C. thummi</i> . . . . .	103
4.9	Der GC-Gehalt des Hämoglobin-Gen-Clusters von <i>C. tentans</i> . . . . .	108
4.10	MP-Baum der Hämoglobine von <i>C. tentans</i> und <i>Glyptotendipes barbipes</i> . . . . .	111
4.11	Die Intron-Sequenzen der Hämoglobin-Gene von <i>C. tentans</i> und <i>Glyptotendipes barbipes</i> im Vergleich . . . . .	112
4.12	Vergleich der Struktur der Hämoglobin-Gengruppen in <i>C. tentans</i> und <i>Glyptotendipes barbipes</i> . . . . .	113
4.13	Die Exon-/Intron-Struktur des Hämoglobin-Vorläufer-Gens von Vertebraten, Invertebraten und Pflanzen ist unbekannt . . . . .	114
4.14	Die Evolution der Genstruktur bei Zuckmücken durch Intron-Insertion . . . . .	115
4.15	Die Position von Introns in Hämoglobin-Genen verschiedener Taxa . . . . .	116
4.16	Phylogenetische Bäume der paralogen Hämoglobin-Gene und Hämoglobin-Proteine von <i>C. tentans</i> . . . . .	121
4.17	Die Entstehung der Hämoglobin-Gene Cte 12.1 und Cte 12.2 durch Duplikation . . . . .	123
4.18	Vergleich der Nukleotid-Sequenzen der Gene Cte 12.1 und Cte 12.2 sowie ihrer flankierenden Intergen-Regionen . . . . .	125
4.19	Die Expansion der 7A-Subfamilie durch Duplikation . . . . .	126
4.20	Die mögliche Evolution der 7A-Gen-Subfamilie in <i>C. tentans</i> . . . . .	127
4.21	Die Duplikation eines 3' nicht-translatierten Bereichs . . . . .	129
4.22	Sequenz-Homogenisierung innerhalb der 7A-Gen-Subfamilie . . . . .	133
4.23	Der Mechanismus der Genkonversion . . . . .	135
4.24	Organisation der 7B-Region im Vergleich . . . . .	138
4.25	Phylogenetischer Baum orthologer Hämoglobine aus der 7B-Region . . . . .	139
4.26	Die Organisation der II $\beta$ -Region im Vergleich . . . . .	141
4.27	Phylogenetischer Baum orthologer Hämoglobine aus der II $\beta$ -Region . . . . .	142
4.28	Synonyme- und nicht-synonyme Substitutionsraten im Vergleich . . . . .	148
4.29	Das mögliche Schicksal von Genkopien nach einem Duplikations-Ereignis . . . . .	149



## Tabellenverzeichnis

1.1	Beispiele für die Verbreitung von Hämoglobinen in den fünf Organismenreichen (nach Dewilde (1999), verändert). . . . .	20
3.1	Lokalisierung der Sonden für die <i>in situ</i> -Hybridisierung . . . . .	45
3.2	Sequenzierungsstatistik . . . . .	54
3.3	Liste der „offenen Leserahmen“ der Contigs 1 und 2 . . . . .	59
3.4	Sequenzunterschiede alleler Hämoglobin-Genvarianten . . . . .	63
3.5	Sequenz-Vergleich überlappender Klone . . . . .	64
3.6	Charakterisierung der Introns in den Hämoglobin-Genen Cte N und Cte P . . . . .	68
3.7	Struktur des Glukosetransporter-Gens in den Spezies <i>C. tentans</i> und <i>C. melanotus</i> . . . . .	69
4.1	Unterschiede der Hämoglobin-Sequenzen von <i>C. tentans</i> , <i>C. thummi</i> und <i>D. melanogaster</i> (p-Distanzen) . . . . .	82
4.2	Invariante Aminosäure-Positionen der beschriebenen Hämoglobine von <i>C. tentans</i> und <i>C. thummi</i> . . . . .	96
4.3	Semi-invariante Aminosäure-Positionen der beschriebenen Hämoglobine von <i>C. tentans</i> und <i>C. thummi</i> . . . . .	99
4.4	Konstitution der für eine Dimerisierung wichtigen Aminosäure-Positionen von <i>C. thummi</i> (Braunitzer et al., 1968) im Vergleich zu den orthologen Genen aus <i>C. tentans</i> . . . . .	105
4.5	Mögliche Quartärstrukturen der in dieser Arbeit gefundenen, neuartigen <i>C. tentans</i> -Hämoglobine . . . . .	106
4.6	Aminosäure-Sequenz-Übereinstimmungen innerhalb der Hämoglobin-Subfamilien von <i>C. tentans</i> . . . . .	118
4.7	Evolutionäre Raten der Hämoglobin-Gene des Clusters D . . . . .	144



## Verzeichnis der Abkürzungen

Abb.	Abbildung
abs.	absolut
A. bidest	aqua bidestillata
AMPPD	3-(2'-Spiroadamantan)-4-methoxy-4-(3'-phosphoryloxy)-phenyl-1,2-dioxetan
AP	alkalische Phosphatase
AS	Aminosäure
ATP	Adenosintriphosphat
BLAST	„Basic Local Alignment Search Tool“
bp	Basenpaar(e)
BSA	Rinderserumalbumin
bzw.	beziehungsweise
cDNA	komplementäre DNA
Ci	Curie
d. h.	das heißt
DIG	Digoxygenin
DNA	Desoxyribonukleinsäure
dNTP	2'-Desoxynukleosidtriphosphat
EDTA	Ethylendinitrilotetraessigsäure
EGTA	Ethylenglykotetraessigsäure
EMBL	„European Molecular Biology Laboratories“
E-Puffer	Elektrophoresepuffer
et al.	und andere (et altera)
EtOH	Ethanol
FISH	Fluoreszenz in situ Hybridisierung
FITC	Fluoresceinisothiocyanat
ges.	gesättigt
Hb	Hämoglobin
IPTG	Isopropyl- $\beta$ -D-thio-galaktopyranosid
kb	Kilobasenpaare
kDa	Kilodalton
konz.	konzentriert
MB	Megabasenpaare
mRNA	„messenger RNA,,
NTP	Nukleosidtriphosphat
OD	optische Dichte

## Verzeichnis der Abkürzungen

---

ORF	„open reading frame“
PBS	„Phosphate-buffered Saline“
PCR	Polymerasekettenreaktion
PEG	Polyethylenglycol
PM	Präinkubationsmedium
RNA	Ribonukleinsäure
RT-PCR	Reverse Transkriptase- Polymerasekettenreaktion
SDS	„Sodiumdodecylsulfate “
SINE	“short interspersed nuclear element,,
SM	Suspensionsmedium
SSC	Standard-Saline-Citrat
Tab.	Tabelle
TBE	Tris-Borat-EDTA
TE	Tris-EDTA
TELT	Tris-EDTA-Lithiumchlorid-Triton
TM	„Trademark “
Tris	Tris(hydroxymethyl)aminomethan
TRITC	Tetramethylrhodaminisothiocyanat
U	„unit“
upm	Umdrehungen pro Minute
UV	ultraviolett
VE	voll entsalzt
w/v	“weight per volume,,
Vol.	Volumen
www	“world wide web,,
X-Gal	5-Brom-4-chloro-3-Indolyl- $\beta$ -Galaktosid
z.B.	zum Beispiel

### 1- und 3- Buchstabencode der Aminosäuren

A	Ala	Alanin	M	Met	Methionin
C	Cys	Cystein	N	Asn	Asparagin
D	Asp	Asparaginsäure	P	Pro	Prolin
E	Glu	Glutaminsäure	Q	Gln	Glutamin
F	Phe	Phenylalanin	R	Arg	Arginin
G	Gly	Glycin	S	Ser	Serin
H	His	Histidin	T	Thr	Threonin
I	Ile	Isoleucin	V	Val	Valin
K	Lys	Lysin	W	Trp	Tryptophan
L	Leu	Leucin	Y	Tyr	Tyrosin

# 1 Einleitung

## 1.1 Genome und Genfamilien

Im Laufe der Evolution haben sich Lebewesen in unzähligen ökologischen Nischen angesiedelt und sich spezifisch an die dort herrschenden Bedingungen angepaßt. Da die physiologischen Eigenschaften eines Organismus von dessen genetischer Ausstattung bestimmt werden, bilden Veränderungen im Genrepertoire die Grundlage für die Entstehung neuer Merkmale und somit für die Evolution von Lebewesen. Einer der wichtigsten Prozesse, der auf molekularer Ebene zur Evolution des Genoms beiträgt, ist zweifellos die Bereitstellung neuer DNA-Abschnitte durch Duplikations-Ereignisse. Hierbei können Genteile, einzelne Gene, Chromosomenabschnitte, Chromosomen oder auch komplette Genome vervielfältigt werden. Schon Haldane (1932) und Muller (1935) vermuteten, daß Genduplikationen und die dadurch entstehende Redundanz von Genen die Grundlage für die Entstehung komplexer Genome bildet.

Gruppen von Genen, die sich durch eine überdurchschnittlich hohe Sequenz-Verwandtschaft auszeichnen, findet man im Genom aller bekannten Organismen. Man geht davon aus, daß solche Gengruppen meist durch aufeinanderfolgende Duplikationsereignisse aus einem ancestralen Vorläufer-Gen entstanden sind. Gene mit einem gemeinsamen stammesgeschichtlichen Ursprung werden auch in Genfamilien oder im Falle sehr großer Gengruppen, in Multigenfamilien zusammengefaßt. Bekannte Beispiele sind die RNA-Gene, die Histon-Gene, die Collagen-Gene (Yamada et al., 1980; Boedtke et al., 1985) oder die Hox-Gene (Zhang und Nei, 1996), sowie die Gene für Aktine und Immunglobuline (Übersicht in Li (1997)). Auch die Alpha-Globingene, die Beta-Globingene und das Myoglobingen der Vertebraten sind Mitglieder einer Genfamilie. Gene, die aus einem gemeinsamen Vorläufer-Gen hervorgegangen sind, werden allgemein als homolog („ursprungsgleich“) bezeichnet. Zur Unterscheidung, ob es sich dabei um verwandte Gene innerhalb einer Spezies oder um homologe Gene verwandter Spezies handelt, sind die Begriffe „paralog“ und „ortholog“ eingeführt worden. Aus einem gemeinsamen Vorläufer-Gen hervorgegangene Genkopien im Genom einer Spezies werden in ihrer Beziehung zueinander als „paralog“ bezeichnet. Ursprungsgleiche Gene in unterschiedlichen Spezies werden hingegen als „ortholog“ definiert. Den Unterschied verdeutlicht die Abb. 1.1.

Die Ermittlung der Nukleotid-Sequenzen ganzer Genome hat gezeigt, daß der Anteil repetitiver Gene am gesamten Genbestand einer Art überraschend hoch ist. So besitzt das Archaeobakterium *Methanococcus jannaschii* etwa 100 Genfamilien (Bult et al., 1996). 47% aller potentiellen Gene des gram-positiven Bakteriums *Bacillus*

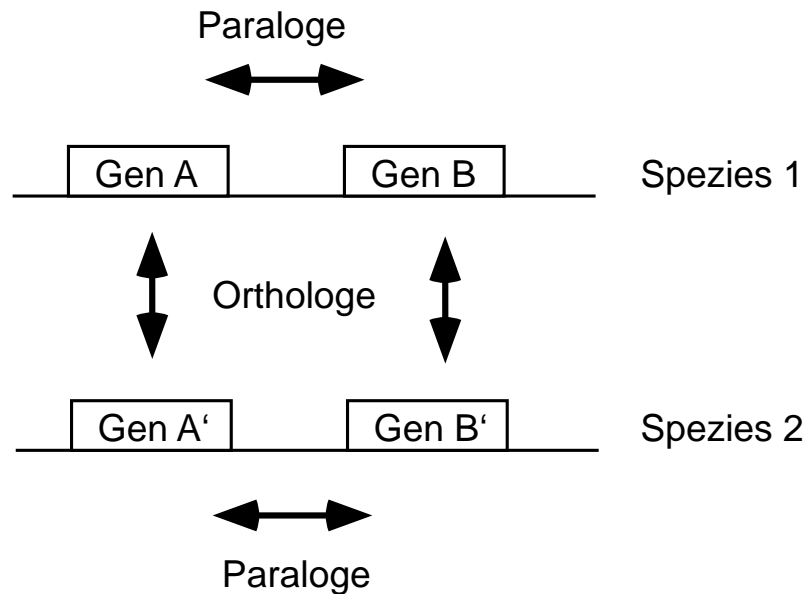


Abbildung 1.1: Definition der Begriffe „paralog“ und „ortholog“

*subtilis* besitzen ein oder mehrere paraloge Gegenstücke (Kunst et al., 1997). Auch die auf Plasmiden lokalisierten Gene des Borreliose-Erregers *Borrelia burgdorferi* sind zu 39% Mitglieder von insgesamt 47 Genfamilien (Fraser et al., 1997). Das Genom der Bäckerhefe *Saccharomyces cerevisiae* weist ebenfalls einen hohen Anteil paraloger Gene auf: mindestens 20% der Gene des Chromosoms IV haben paraloge Partner (Jacq et al., 1997). Eine Vielzahl langer, genhaltiger Chromosomenabschnitte liegen in der Hefe als Duplikationen vor. Solche Duplikationen ganzer Gengruppen erzeugen ein äußerst komplexes Muster von Verwandtschaften zwischen den einzelnen Hefe-Chromosomen (Mewes et al., 1997) und geben einen Einblick in das Ausmaß der Duplikationsprozesse, die eukaryotische Genome verändern.

Die Organisation der Mitglieder einer Genfamilie im Genom kann sehr vielfältig sein. So sind beispielsweise bei den Genen für ribosomale RNA oft mehr als 100 Kopien streng tandemartig hintereinander angeordnet und an einem Ort im Genom konzentriert (Gerbi, 1985). Bei anderen Genfamilien hingegen (wie bei den Alpha- und Beta-Globingenen der Säuger) liegen die Gene in Gruppen („Cluster“) in räumlicher Nähe zueinander auf den Chromosomen (Efstratiadis et al., 1980). Mitglieder der Familie der Geruchsrezeptorgene sind dagegen über das gesamte Genom verstreut (Sullivan et al., 1996). Trotz unterschiedlicher räumlicher Organisation weist jedoch alles darauf hin, daß Genvermehrung ein entscheidender Motor der Evolution der Genome darstellt.

## 1.2 Mechanismen der Genduplikation

Die Entstehung einer Genfamilie beginnt stets mit der Duplikation eines zunächst singulär vorliegenden Gens. Grundsätzlich gibt es alternative Möglichkeiten, durch die Gene vervielfacht werden können: (1) Die Duplikation des gesamten Genoms eines Organismus oder zumindest einzelner Chromosomen (Ohno, 1970), und (2) die Duplikation chromosomaler Teilabschnitte, z. B. einzelner Gene.

### Genom-Duplikationen

Werden die Schwesterchromatiden eines Chromosomensatzes nach erfolgter Replikation nicht auf die Tochterzellen verteilt, so führt dies zu einer Verdopplung des Gesamt-Genoms. In Pflanzen ist eine solche Polyploidisierung ein weit verbreitetes Phänomen (Müntzing, 1936; Lewis, 1980). So weisen beispielsweise die Genomgrößen verschiedener Monokotyledonen darauf hin, daß diese ihr Genom im Verlauf ihrer Evolution schrittweise dupliziert haben (Sparrow und Naumann, 1976).

Häufig resultiert die Duplikation des kompletten Genoms in Störungen der Genbalance und des Zellzyklus und ist somit von Nachteil für das Individuum (Stebbins, 1971; Dewey, 1980). Hingegen scheint bei Pflanzen die Bildung von Hybridformen durch die Kombination verschiedener Genom-Typen (z. B. Arten, Rassen) während der Evolution Vorteile bei der Anpassung an veränderte Lebensbedingungen oder neue Lebensräume gebracht zu haben. Bei sich bisexuell fortpflanzenden Tieren ist die Polyploidisierung stets von großem Nachteil und daher selten. Bereits 1925 wies Muller darauf hin, daß der Mechanismus der heterogametischen Geschlechtsbestimmung durch Polyploidisierung aus dem Gleichgewicht gebracht werden könnte.

Gegenwärtig wird davon ausgegangen, daß während der Evolution der frühen Vertebraten zumindest zwei Duplikationen des Gesamt-Genoms stattgefunden haben (Sidow, 1996). Phylogenetische Analysen von Genfamilien, wie beispielsweise der für die Entwicklung eines Organismus wichtigen Homöobox-Gene, unterstützen diese Annahme (Holland und Garcia-Fernandez, 1996; Lundin, 1993).

Auch für die Hefe *Saccharomyces cerevisiae* konnte gezeigt werden, daß 13% aller Gene paraloge Kopien besitzen (Wolfe und Shields, 1997). Da die identifizierten Blöcke, die die duplizierten Gene enthalten, nicht durch unabhängige Duplikations-Ereignisse erklärt werden können, wird von einer Verdopplung des Gesamt-Genoms im Verlauf der Hefe-Evolution ausgegangen. Während Genomduplikationen sich offensichtlich öfter während der Evolution ereignet haben, sind Verdopplungen einzelner Chromosomen weniger häufig. Aufgrund von Gendosiseffekten sind Aneuploidien in der Regel mit katastrophalen Auswirkungen für den betroffenen Organismus verbunden. Es ist daher zu vermuten, daß sie eine weniger bedeutende Rolle in der Evolution von Genomen gespielt haben.

### Duplikation von Teilabschnitten des Genoms

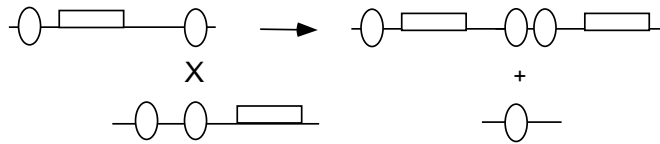
Duplikationen kurzer bis mittelgroßer Chromosomenabschnitte sind ein häufig zu beobachtendes Phänomen. Als Mechanismen kommen dafür das inäquale Crossing-over, sowie die Retroposition in Frage.

Der vermutlich wichtigste Prozeß, durch den funktionelle Genkopien initial entstehen, ist die Duplikation von Genomabschnitten durch inäquales Crossing-over. Hierbei führt die versetzte (inäquale) Paarung von Nichtschwester-Chromatiden oder Schwesterchromatiden bei nachfolgendem Crossing-over zu einem reziproken DNA-Austausch. Während die eine Chromatide die Duplikation eines DNA-Abschnittes aufweist, ist dieser auf der anderen Chromatide deletiert. In der Regel erfolgt inäquales Crossing-over zwischen DNA-Abschnitten, die eine deutliche Sequenzähnlichkeit besitzen, und wird dann auch als homologe Rekombination bezeichnet. So findet man z. B. innerhalb der menschlichen Beta-Globingen-Gruppe bevorzugt Rekombinations-Ereignisse an Orten, die eine etwa 50 bp lange identische Sequenz aufweisen (Metzenberg et al., 1991). Interspergiert organisierte, nichtkodierende repetitive Elemente schaffen nahezu überall in einem komplexen Genom Orte, an denen eine längere Sequenz-Übereinstimmung existiert. Diese Repetitionen dienen als Ausgangspunkte für Crossing-over-Ereignisse (siehe Abb. 1.2A). Eine Reihe menschlicher Erkrankungen sind auf inäquale Crossing-over zwischen interspergiert-repetitiven Alu-Elementen zurückzuführen, die in den Introns der betroffenen Gene vorkommen (z. B. Muskeldystrophie, Hu et al. (1991); Wachstumshormondefizienz, Vnencak-Jones et al. (1988)). In selteneren Fällen erfolgt Crossing-over auch an Positionen, an denen die beteiligten Stränge entweder keine erkennbare Sequenzähnlichkeit aufweisen, oder an denen lediglich kurze Homologien von 3-5 bp Länge zu finden sind (Ruley und Fried, 1983). Solche Rekombinations-Ereignisse werden als „illegitim“ bezeichnet und können ebenfalls zu Genvermehrung oder Genverlust führen (siehe Abb. 1.2B).

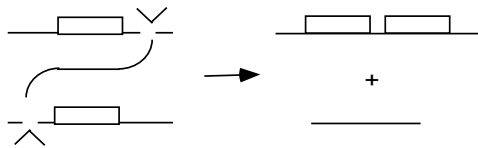
Die Verbreitung von Genen durch Retrotransposition erfordert das Überschreiben der mRNA des Gens in die cDNA durch das Enzym Reverse Transkriptase. Diese neugebildete cDNA kann anschließend in das Genom reintegrieren. Das Fehlen von Introns, der Promotoregion im 5'-Bereich des Gens, sowie der Besitz einer von der mRNA stammenden polyA-Sequenz am 3'-Ende, sind typische Erkennungsmerkmale solcher aus Retrotransposition hervorgegangener Genduplikate. Aufgrund dieser Eigenschaften sind diese Gene in der Regel nicht funktionstüchtig und werden daher auch als „prozessierte Retro-Pseudogene“ bezeichnet (Weiner et al., 1986). Je nach Integrationsort können Retro-Gene jedoch auch mit einem neuen Promotor ausgestattet werden. So finden sich z. B. in der Glycerinkinase- und in der GA733-Genfamilie des Menschen (Sargent et al., 1994; Linnenbach et al., 1993), sowie im Falle des Prä-Proinsulins in Maus und Ratte (Soares et al., 1985) solche transkriptionsaktiven Genkopien (siehe Abb. 1.2C).



A. In quale, homologe Rekombination zwischen repetitiven Elementen



B. In quale, illegitime Rekombination



C. Transposition ber RNA-Intermediate

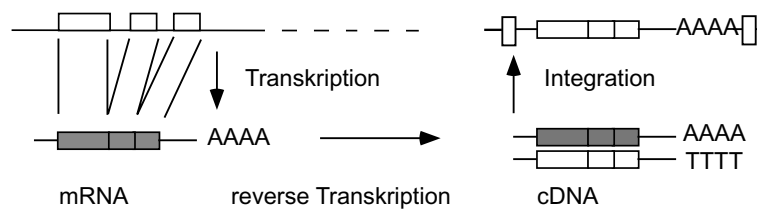


Abbildung 1.2: Mechanismen der Duplikation von DNA-Abschnitten

A) Inäquales, homologes Crossing-over zwischen repetitiven Elementen

B) Inäquales Crossing-over an Stellen ohne deutliche Sequenzhomologie

C) Genvermehrung durch RNA-vermittelte Transposition („Retropositon“) Nach Maeda und Smithies (1986).

### 1.3 Die Evolution duplizierter Gene

Das Schicksal duplizierter Gene kann vielfältig sein. Zunächst bietet eine zusätzliche Genkopie die Möglichkeit, eine größere Menge des entsprechenden Transkriptionsproduktes oder Proteins bereitzustellen. Wie sich am Beispiel der rRNA-, tRNA- und Histon-Genfamilie (Elgin und Weintraub, 1975) zeigt, werden die Gensequenzen der duplizierten Gene bei entsprechendem Bedarf am Genprodukt fixiert (Ohno, 1970). Die Genkopien unterliegen in einem solchen Fall einer stabilisierenden Selektion. Auch können duplizierte Gene durch die vermehrte Bereitstellung ihres Produktes

als Reaktion auf einen Umwelt-Stimulus die Überlebenschancen eines Organismus spezifisch verbessern. Beispielsweise bewirkt bei dem Californischen Mosquito *Culex pipiens* ein amplifiziertes Esterase-Gen eine erhöhte Resistenz gegenüber Insektiziden (Mouches et al., 1986).

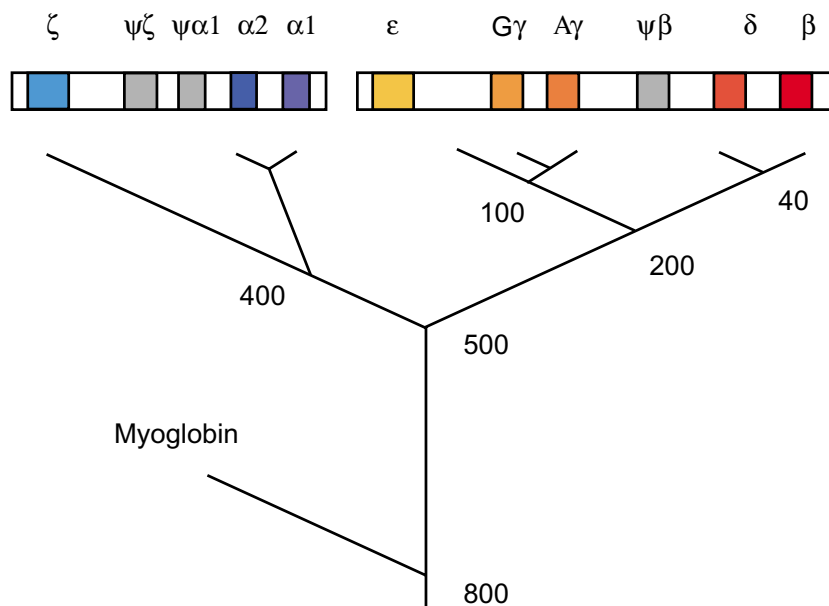
Prinzipiell können beide Genduplikate auch unabhängig voneinander evolvieren. So sind häufig zusätzliche Genkopien nicht von direktem Selektionsvorteil für die Spezies und degenerieren durch zufällig auftretende Mutationen wieder. Die resultierenden Pseudogene können in Abhängigkeit von der Populationsgröße, der Mutationsrate und der genetischen Drift in einer Population zufällig fixiert werden (Fisher, 1935; Nei und Roychoudhury, 1973; Bailey et al., 1978; Kimura und King, 1979; Li, 1997; Watterson, 1983). Aus Duplikationen hervorgegangene Pseudogene wurden bislang in allen genauer untersuchten Genfamilien wie beispielsweise den Immunglobulinen (Huang et al., 1981), den Globinen (Nishioka et al., 1980; Proudfoot und Maniatis, 1980; Little, 1982) oder den 5S rRNA-Genen (Jacq et al., 1977) nachgewiesen.

Mutationen führen jedoch nicht zwangsläufig zur Inaktivierung eines Gens, oft sind sie neutral und somit ohne Auswirkung auf die Funktion. In anderen Fällen können sie die Funktion eines Genproduktes verändern und somit den Organismus mit neuen Eigenschaften ausstatten. Ein Beispiel hierfür ist die Entstehung des trichromatischen Sehens bei Altweltaffen, Menschenaffen und Menschen. Dichromaten besitzen zwei Pigmentgruppen mit unterschiedlichen Absorptionsspektren, die von zwei paralogenen Genen kodiert werden. Durch die Duplikation des X-chromosomal gekoppelten Pigmentgens vor ca. 35-40 Millionen Jahren entstand im gemeinsamen Vorfahren von Altweltaffen, Menschenaffen und Menschen eine zusätzliche Genkopie. Diese führte im Verlauf der Evolution zu einem dritten Photopigment mit einer geänderten spektralen Empfindlichkeit (Nathans et al., 1986; Yokoyama und Yokoyama, 1989; Neitz et al., 1991).

### 1.4 Die Genfamilie der Globine

Auch die Globin-Genfamilie ist ein typisches Produkt von aufeinanderfolgenden Genduplikationen mit anschließender Diversifizierung der Gene. Durch Stammbaum-Rekonstruktion anhand von Sequenzdaten läßt sich ausgehend von einem „Urglobin“ eine phylogenetische Entwicklung verfolgen (siehe Abb. 1.3).

Während dieser Entwicklung hat sich das Myoglobin vor ca. 600-800 Millionen Jahren durch einen Duplikationsschritt von der Hämoglobin-Genfamilie abgetrennt (Dayhoff, 1972; Doolittle, 1987). Eine weitere Differenzierung des Hämoglobins in Säugern erfolgte durch die Trennung der  $\alpha$ - und  $\beta$ -Unterfamilie. Die Gene für  $\alpha$ - und  $\beta$ -Ketten sind als Gengruppen auf den Chromosomen 16 und 11 lokalisiert. Die verschiedenen Gene werden in unterschiedlichen Entwicklungsstadien exprimiert. Beim Menschen folgt auf das embryonale Hämoglobin ( $\zeta\epsilon 2$ ) das fötale Hämoglobin



**Abbildung 1.3: Stammbaum der Vertebraten-Globine**

Der Balken zeigt die Lage der Hämoglobin-Gene innerhalb der alpha- und beta-Gencluster. Pseudogene sind grau dargestellt. An den Verzweigungen des Stammbaumes sind die geschätzten Divergenzzeiten in Millionen Jahren gezeigt (nach Jeffreys et al. (1983)).

( $\alpha_2\gamma_2$ ), das schließlich durch das Hämoglobin des Erwachsenen ( $\alpha_2\beta_2$ ) abgelöst wird (Karlsson und Nienhuis, 1985). Das phylogenetisch jüngste Gen kodiert für das  $\delta$ -Globin, das nur beim Menschen und Menschenaffen vorkommt und den übrigen Primaten fehlt.

Daß es sich bei den Globin-Genen um eine sehr alte Genfamilie handelt, zeigt ihr Auftreten in allen fünf Organismenreichen. Sowohl in Bakterien, Einzellern, Pilzen und Pflanzen sind sporadisch Globin-Gene vorhanden (siehe Tabelle 1.1).

Im Gegensatz zu den Vertebraten-Hämoglobinen zeichnen sich die Globine der Invertebraten durch eine enorme strukturelle und funktionelle Vielfalt aus (zusammengefaßt in Dewilde (1999)). Der Vergleich ihrer Primärstrukturen zeigte, daß ausschließlich die beiden Aminosäuren, die für die Ausbildung der molekültypischen Faltung verantwortlich sind (Phe CD1 und His F8), konserviert vorliegen (Moens et al., 1996). Auch die Tertiär-Struktur der Invertebraten-Globine weist große Unterschiede auf. Sie können eine, zwei aber auch mehrere Sauerstoffbindende Domänen besitzen. Einige der Invertebraten-Globine werden in die Hämolymphe sekretiert, während andere intrazellulär vorliegen. Auch die Molekulargewichte schwanken zwischen ca. 17 kD und ca. 800 kD sehr stark (zusammengefaßt in Dewilde (1999)). Trotz all dieser Unterschiede ist anzunehmen, daß auch die Globine der Invertebraten auf

einen gemeinsamen phylogenetischen Ursprung zurückzuführen sind (Vinogradov, 1985; Vinogradov et al., 1993; Bolognesi et al., 1997)

Tiere	Pflanzen	Pilze	Einzeller	Prokaryoten
<b>Vertebraten</b> alle	<b>Leguminosen</b>	<b>Hefen</b>	<b>Ciliaten</b>	
<b>Invertebraten</b>	<i>Glycine</i>	<i>Saccharo-</i>	<i>Paramecium</i>	<i>E. Coli</i>
<b>Arthropoda:</b> <i>Chironomus</i>		<i>myces</i>	<i>Tetrahymena</i>	<i>Baccillus sub.</i>
<i>Drosophila</i>		<i>Pichia</i>		<i>Vitreoscilla</i>
<i>Gastrophilus</i>				
<b>Mollusca:</b> <i>Biomphalaria</i>	<b>Ulmaceen</b>		<b>Algen</b>	
<i>Aplysia</i>			<i>Chlamydom.</i>	<i>Synechocystis</i>
<i>Barbartia</i>			<i>Nostoc</i>	<i>Mycobacterium</i>
<i>Lucina</i>				<i>Erwinia chrys.</i>
<b>Annelida:</b> <i>Lumbricus</i>				
<i>Arenicola</i>				
<b>Nematoda</b> <i>Ascaris</i>				
<i>Mermis</i>				
<i>C. elegans</i>				
<b>Trematoda</b> <i>Fasciola</i>				
<i>Paramphist.</i>				

**Tabelle 1.1:** Beispiele für die Verbreitung von Hämoglobinen in den fünf Organismenreichen (nach Dewilde (1999), verändert).

## 1.5 Die Evolution der Hämoglobin-Genfamilie der Chironomiden

Zuckmücken (Diptera, Chironomidae) stellen mit einer geschätzten Anzahl von mehr als 10000 Arten eine der erfolgreichsten Insektenfamilien dar. Ihre Larven leben im Sediment meist sauerstoffarmer Gewässer. Es zeichnet sie aus, daß sie sich an sehr verschiedene, zum Teil extreme Lebensbedingungen angepaßt haben. In Thermalgewässern Nordamerikas, Islands und Europas wurden Chironomiden-Larven bei Temperaturen von 40-50 °C gefunden. Im Gegensatz zu diesen wärmeliebenden Formen gibt es auch Arten, die in arktischen Gletscherseen bei Temperaturen

knapp über dem Gefrierpunkt existieren und sogar monatelanges Einfrieren überleben können. Auch extreme chemische Bedingungen werden von vielen Arten toleriert. So treten selbst in schwefelsauren Gewässern bei einem pH-Wert von 1,4 noch Chironomiden auf. Auch in Schwefelwasserstoffquellen mit einem Gehalt von 25 mg Schwefel pro Liter Wasser sowie in Seen, die einen Salzgehalt von 28 % aufweisen, sind Chironomiden-Larven zu finden. Viele Chironomiden-Arten überleben unter extrem sauerstoffarmen Bedingungen im Sediment stehender Gewässer. Bei Hypoxie ( $pO_2 < 15$  Torr) können die Larven sogar zu einem anaeroben Stoffwechsel übergehen, bei dem sie als Produkt der alkoholischen Gärung Ethanol ausscheiden (Redecker und Zebe, 1988). Als eine Anpassung der Larven an sauerstoffarme Lebensbedingungen ist auch der Besitz des roten Blutfarbstoffes aufzufassen, der erstmals von Rollet im Jahre 1861 beschrieben (zitiert in Weber (1980)) und von Lankaster 1867 als „Hämoglobin“ bezeichnet wurde.

Das Hämoglobin der Chironomiden wird in den Zellen der Fettkörper aller vier Larvenstadien synthetisiert (Shilova, 1958; Bergtrom et al., 1976) und kann schließlich bis zu 90 % der gesamten Hämolympheproteine ausmachen (English, 1969; Tichy, 1980). Desweiteren konnte Hämoglobin in Ovarien erwachsener Mücken sowie in Eiern und Embryonen nachgewiesen werden (Trewitt und Bergtrom, 1981; Trewitt et al., 1986). Eine zusätzliche Funktion des Hämoglobins als mögliche Nahrungsquelle während der Embryonalentwicklung wurde in diesem Zusammenhang diskutiert. Der enzymatische Abbau des Hämoglobins beginnt mit der Verpuppung der Larve (Laufer und and, 1971; Schin et al., 1974), so daß in der Hämolymphe der erwachsenen Zuckmücke kein Hämoglobin mehr zu finden ist.

Die fundamentale Bedeutung des Chironomiden-Hämoglobins für die Aufnahme von Sauerstoff unter hypoxischen Bedingungen wird durch die Beobachtung unterstützt, daß der Hämoglobingehalt der Larven in Abhängigkeit vom Sauerstoffgehalt des Wassers schwankt. So weisen Larven der Sommergeneration von *C. thummi* sowie *C. plumosus* etwa die doppelte Menge an Hämoglobin auf als die Larven der entsprechenden Frühjahrsgeneration (Czeczuga, 1960; Leyko und Osmulski, 1985). Einen weiteren Hinweis auf die Funktion des Hämoglobins als Sauerstoff-Bindungs- und Transportmolekül liefert der Befund, daß die respiratorische Aktivität von Larven eingeschränkt ist, deren Hämoglobine bei geringem Sauerstoffpartialdruck ( $\leq 50$  Torr) durch Kohlenmonoxid inhibiert wurden (Zebe, 1991).

Bereits 1934 zeigten Svedberg und Eriksson-Quensel durch Ultrazentrifugation, daß es sich bei den Hämoglobinen der Chironomiden um dimere Moleküle mit einem vergleichsweise geringen Molekulargewicht von 34 000 Da handelt. Dies ist insofern bemerkenswert, als die Blutfarbstoffe anderer Invertebraten häufig hochmolekulare Komplexe bilden (Überblick in Vinogradov et al. (1993)). Auch werden die Hämoglobine der Chironomiden sekretiert und liegen gelöst in der Hämolymphe der Larven vor. Zu diesem Zweck besitzen die primären Translationsprodukte der Hämoglobin-mRNAs an ihrem N-Terminus ein meist 16 Aminosäuren langes, hydro-

phobes Signalpeptid, das die Sekretion des Moleküls aus den Fettzellen ermöglicht und dabei abgespalten wird (Baldwin et al., 1981; Antoine und Niessing, 1984).

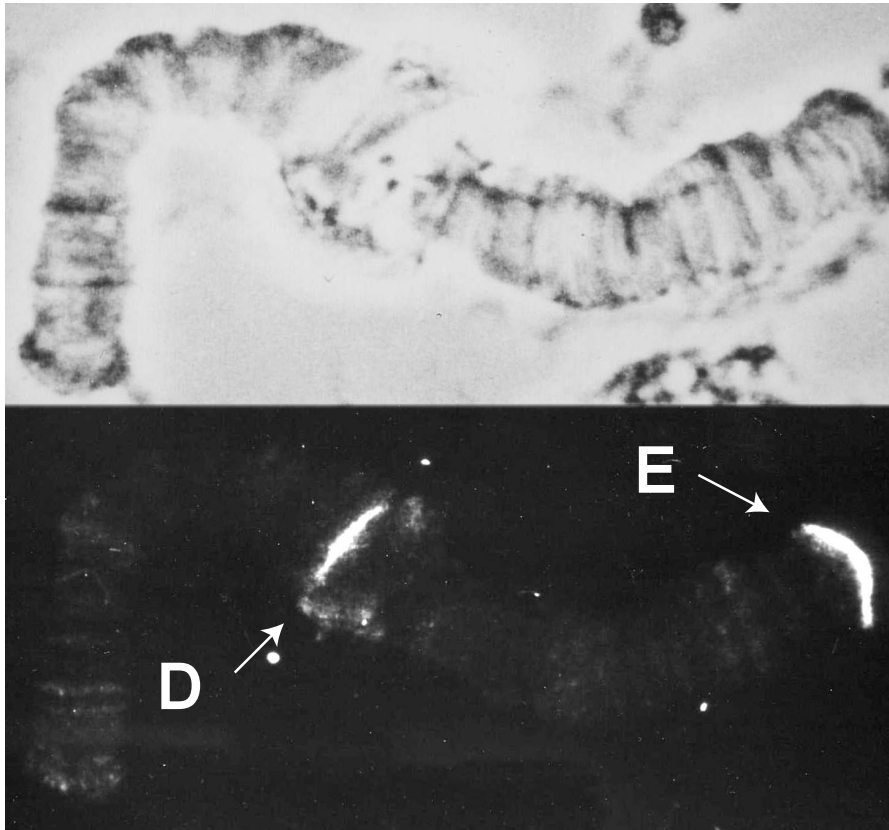
Thompson und English konnten 1966 zeigen, daß das „Chironomiden-Hämoglobin“ in Wirklichkeit eine Mischung unterschiedlicher Protein-Varianten darstellt, die nicht auf allele Polymorphismen, sondern auf multiple Gene zurückzuführen ist. Der Gruppe um Gerhard Braunitzer gelang es in den folgenden Jahren, 12 unterschiedliche Hämoglobin-Komponenten aus *C. thummi* zu isolieren und deren Aminosäuresequenz zu bestimmen (Braun et al. (1968), Überblick in Goodman et al. (1983, 1988)). Während fünf dieser Varianten monomere Hämoglobine ausbilden (I, IA, III, IIIA, IV), liegen die Varianten II $\beta$ , VI, VIIA, VIIB, VIII und IX als Homodimere in der Hämolymphe vor. Das Hämoglobin X kann in Abhängigkeit von den pH-Bedingungen sowohl in der monomeren, als auch in der dimeren Form auftreten. Von diesen 12 Proteinen haben sich nach Analyse der entsprechenden Gene nur die Varianten I und IA nachträglich als Allele herausgestellt (Kao et al., 1995).

Erste phylogenetische Vergleiche der Aminosäure-Sequenzen deuteten darauf hin, daß vermutlich die monomeren Hämoglobine die stammesgeschichtlich älteren Varianten darstellen (Goodman et al., 1983, 1988). Der Vergleich der rekonstruierten, hypothetischen Vorläufer monomerer und dimerer Hämoglobin-Varianten von Chironomus mit den Alpha- und Beta-Hämoglobin-Ketten des Menschen zeigt, daß über einen Bereich von 140 Aminosäuren noch 20 funktionell bedeutende Positionen konserviert sind (Goodman et al., 1983).

Zahlreiche Untersuchungen der chromosomalen und molekularen Organisation der Hämoglobin-Gene in den letzten 15 Jahren sollten die genetischen Grundlagen der Hämoglobin-Multiplizität in Chironomus aufklären. Gleichzeitig erhoffte man sich eine Antwort auf die Frage nach der funktionellen Bedeutung dieser außergewöhnlichen Proteinvielfalt.

Durch cytogenetische Kartierung und erste in situ-Hybridisierungsexperimente war die Existenz von mindestens zwei Genorten im Genom der Chironomiden vorhergesagt worden (Tichy, 1970, 1975; Laufer et al., 1982). Nachdem entsprechende Gensonden kloniert vorlagen, konnten tatsächlich zwei getrennte Genloci den Chromosomenarmen E und D nachgewiesen werden (Hankeln et al., 1988; Schmidt et al., 1988). Hier liegen die Hämoglobin-Gene dicht gepackt in Gengruppen, die entsprechend ihrer Zugehörigkeit zu dem einen oder anderen Chromosomenarm als Cluster E bzw. Cluster D bezeichnet werden (siehe Abb. 1.4).

Die Analyse auf genomischer Ebene zeigte, daß die Vielfalt der Hämoglobin-Gene die der bekannten Protein-Varianten noch bei weitem übersteigt. Der genaue Umfang dieser Hämoglobin-Multigenfamilie ist bislang unbekannt. Die von Braunitzers Gruppe beschriebenen Hämoglobine konnten nahezu alle auf Genebene identifiziert werden. Es zeigte sich sogar, daß einige dieser Varianten durch mehrere z.T. nur leicht unterschiedliche Gene kodiert werden (Kao und Bergtrom, 1995; Kao et al., 1995; Antoine et al., 1987; Rozynek et al., 1991; Kao et al., 1994; Gruhl et al., 1997; Hankeln et al., 1988; Hankeln und Schmidt, 1991; Hankeln et al., 1998; Trewitt et al.,



**Abbildung 1.4:** Die Lokalisation der Hämoglobin-Gene im Genom von Zuckmücken  
Die Hämoglobin-Gene sind im Chironomiden-Genom in zwei Gengruppen organisiert. Diese werden entsprechend den Chromosomen-Armen als Gen-Cluster E und Gen-Cluster D bezeichnet. Die Abbildung zeigt ein Polytäanchromosom von *C. tentans*. Als Sonde diente das Hämoglobin-Gen 4 aus *C. thummi*. Es hybridisiert unter relaxierten Bedingungen mit Hämoglobin-Genen der Gengruppe E sowie der Gengruppe D.

1995). Zusätzlich wurden zahlreiche neue Hämoglobin-Gene identifiziert, deren Proteine in der Hämolymphe bislang nicht nachgewiesen werden konnten (Antoine et al., 1987; Hankeln und Schmidt, 1991; Hankeln et al., 1998; Rozynek et al., 1991; Gruhl et al., 1997).

## 1.6 Die Intron-/Exon-Struktur der Hämoglobin-Gene

Bei den Hämoglobin-Genen handelt es sich um eine sehr alte Genfamilie, die im gesamten Organismenreich verbreitet ist. Ihre Intron/Exon-Struktur ist daher seit langem von besonderem Interesse, um das stammesgeschichtliche Alter von Introns einzugrenzen. Die meisten der bislang untersuchten Hämoglobin-Gene bei Pflanzen,

Vertebraten und auch Invertebraten besitzen zumeist mehrere Introns (Reviews: Hardison (1996); Moens et al. (1996); Vinogradov et al. (1992)). Die Intronverteilung in Vertebraten-Hämoglobin-Genen ist ausgesprochen einheitlich. Es existiert durchgängig eine drei Exon/zwei Intron-Struktur, wobei die Introns in den Helices B und G des korrespondierenden Proteins liegen (Hardison, 1991; Blanchetot et al., 1983, 1986; Weller et al., 1984). Die Hämoglobine der Pflanzen besitzen zusätzlich ein drittes Intron in der Helix E, das aufgrund dieser Lage auch als „zentrales“ Intron bezeichnet wird (Go, 1981; Wiborg et al., 1982; Jensen et al., 1981). Im Gegensatz zu den Hämoglobinen der Vertebraten weisen die Hämoglobine der Invertebraten eine erstaunliche Variabilität in ihrer Struktur auf. Während der Annelide *Lumbricus terrestris* die für Vertebraten typischen zwei Introns besitzt (Jhiang und Riggs, 1989), finden sich bei den Nematoden *Pseudoterranova decipiens*, *Ascaris suum* und *Caenorhabditis elegans* ein für den Pflanzentypus charakteristisches „zentrales“ Intron innerhalb der Proteinhelix E (Dixon et al., 1992; Sherman et al., 1992; Kloek et al., 1993).

Molekulare Analysen von Hämoglobin-Genen der Chironomiden zeigten, daß die meisten der untersuchten Gene keinerlei Introns besitzen (Antoine und Niessing, 1984; Antoine et al., 1987; Trewitt et al., 1987; Hankeln et al., 1988; Hankeln und Schmidt, 1991; Hankeln et al., 1997; Rozynek et al., 1989, 1991). Drei Hämoglobin-Gene (2 $\beta$ , 9 und 7A) weisen jedoch in einigen *Chironomus*-Arten ein Intron innerhalb der Protein-Helix E auf (Kao et al., 1994; Gruhl et al., 1997; Hankeln et al., 1997). Diese Intronverteilung ist insofern überraschend, als die orthologen Gene in vielen anderen Chironomiden-Spezies kein Intron enthalten (Rozynek et al., 1991; Hankeln et al., 1997). Es stellte sich daher die Frage, ob die Hämoglobin-Gen-Introns der Chironomiden den „zentralen“ Introns der anderen Organismen entsprechen und damit einen Beweis für das phylogenetisch hohe Alter dieses Hämoglobin-Introns liefern („intron-early“ Hypothese; Moens et al. (1992)). In diesem Fall müßte ein Verlust dieses ursprünglichen Introns in der Entwicklungslinie der Vertebraten postuliert werden. Alternativ könnte das Vorhandensein dieser Introns durch mehrere, unabhängige Integrationsereignisse erklärt werden („intron-late“-Hypothese). Um die stammesgeschichtliche Herkunft dieses Introns einzugrenzen, wurden daher orthologe Gene in einer Vielzahl europäischer und australoasiatischer Chironomiden Spezies auf den Besitz dieses Introns hin untersucht (Hankeln et al., 1997). Es zeigte sich, daß die Introns in den untersuchten Genen dieser Spezies nur sporadisch auftreten und auch die Intronpositionen nicht konserviert sind. Es ist deshalb fraglich, ob die vielfältigen „zentralen Introns“ in den Hämoglobin-Genen von Invertebraten und Pflanzen tatsächlich von einem Vorläufer-Intron im Ur-Hämoglobin-Gen abstammen. Vielmehr weisen die Ergebnisse darauf hin, daß Introns durch ein Integrationsereignis Teil von Hämoglobin-Genen werden. Die Analyse der Intron-/Exon-Struktur insbesondere aus Chironomiden ist für die Absicherung dieser Hypothese von großer Wichtigkeit.



## 1.7 Problemstellung und Zielsetzung der Arbeit

Chironomiden gehören zu den wenigen Insekten-Familien, die den roten Blutfarbstoff Hämoglobin besitzen. Bereits frühe Arbeiten an Hämolymphe der Zuckmücken haben gezeigt, daß eine Vielzahl unterschiedlicher Hämoglobin-Varianten darin enthalten sind (Thompson und English, 1966; Tichy, 1970; Goodman et al., 1983). Diese werden entsprechend von vielen verschiedenen Genen kodiert. Die Hämoglobin-Gene der Chironomiden sind chromosomal in zwei Gen-Gruppen organisiert (Tichy, 1970, 1975; Laufer et al., 1982; Hankeln et al., 1988; Schmidt et al., 1988). Durch in situ-Hybridisierung an Polytänchromosomen wurde eine Gen-Gruppe auf dem Chromosomen-Arm E lokalisiert, die daher auch als Gen-Cluster E bezeichnet wird (Hankeln et al. (1988); Schmidt et al. (1988); siehe Abb. 1.4). Sie enthält je nach Spezies fünf oder sechs Gene und wurde im Rahmen eines vergleichenden Ansatzes bereits von den Arten *C. thummi*, *C. piger*, *C. pallidivittatus* und auch *C. tentans* molekular und phylogenetisch charakterisiert (Hankeln et al. (1998); Hankeln et al., unpubliziert). Die zweite, umfangreichere Gen-Gruppe wird aufgrund ihrer Lage auf dem Chromosomen-Arm D als Cluster D bezeichnet. Die Ausdehnung dieser Gen-Gruppe, sowie die Anzahl ihrer Hämoglobin-Gene ist noch nicht vollständig bekannt. Zu Beginn dieser Arbeit war nur klar, daß das Hb-Cluster D eine große Anzahl von Hb-Genen enthalten mußte. Das primäre Ziel dieser Arbeit ist es daher das Hb-Cluster D von *C. tentans* molekular zu charakterisieren. Exemplarisch wurde hierfür die Art *C. tentans* ausgewählt, die neben *C. thummi* als Referenzspezies für molekularbiologische Analysen bei Chironomus gilt. Bereits vorliegende Teil-Sequenzdaten aus dem Bereich des Hämoglobingen-Clusters D von *C. tentans* (Rozynek et al., 1991) ermöglichen, einen raschen experimentellen Zugang zu der genomischen Region.

Ausgehend von zwei bereits vorliegenden Lambda-Phagenklonen kann ein Chromosomenmarsch durchgeführt werden, um möglichst große Bereiche des Hämoglobingen-Clusters D zusammenhängend zu klonieren. Durch vollständige Sequenzierung ist es anschließend möglich die Anzahl der vorhandenen Hämoglobin-Gene sowie ihre chromosomale Organisation zu ermitteln. Die Identifizierung der Hämoglobin-Gene soll hierbei aufgrund ihrer im wesentlichen gut konservierten Nukleotid-Sequenz durch einfache Sequenz-Vergleiche z. B. im Rahmen von Datenbanksuchen erfolgen. Die Ergebnisse der bisherigen Untersuchungen auf Genebene legen nahe, daß bislang nicht alle bei Chironomien vorhandenen Hämoglobin-Varianten beschrieben wurden. Durch den Vergleich aller verfügbaren Hämoglobin-Gene der Gen-Gruppe D von *C. tentans* mit bereits publizierten orthologen Gen-Varianten anderer Chironomiden-Spezies sollen neue Hämoglobin-Gene identifiziert werden. Auf Grundlage der DNA-Sequenzen können die gefundenen Gene auf ihre Funktionalität hin untersucht werden und eventuell vorhandene Pseudogene identifiziert werden. Durch die Sequenzierung längerer genomischer Abschnitte wird es ferner möglich,

auch großräumig die Architektur dieses Gen-Clusters mitsamt seinen Intergen-Regionen inklusive aller Gene sowie möglicher repetitiver DNA-Sequenzen zu erfassen.

Es sollte möglich sein, durch den Vergleich paraloger Hämoglobin-Gene von *C. tentans* die Sequenz-Verwandtschaften der unterschiedlichen Hämoglobin-Varianten zu ermitteln. So kann beispielsweise überprüft werden, ob sich die Genkopien der Gruppe D bestimmten Subfamilien zuordnen lassen und ob es Hinweise auf die rezente Entstehung von Hämoglobin-Genkopien z.B. durch Genduplikation gibt.

Eine weiterführende Aufklärung über die molekulare Evolution der Gen-Gruppe ist durch den interspezies Vergleich orthologer Hämoglobine möglich. Hierdurch können die Veränderungen des Genrepertoires und der Clusterstruktur nach der Trennung der Arten aufgezeigt und Mechanismen der molekularen Evolution rekonstruiert werden. Es soll gezeigt werden, welchen Anteil Prozesse wie Insertion, Duplikation, Deletion, Genkonversion und Transposition an der Evolution der Hämoglobin-Gengruppe haben.

Die Intron-/Exon-Struktur der Invertebraten-Hämoglobine wird von den Befürwortern der „intron-early“-Hypothese häufig als Hinweis auf die Existenz eines ursprünglichen, zentralen Introns im Hämoglobin-Vorläufer-Gen gewertet. Auch in Chironomiden Hämoglobin-Genen wurden Introns an zentralen Positionen gefunden (Hankeln et al., 1997). Die Art *C. tentans* ist insofern ungewöhnlich, als daß keines der bisher analysierten Hämoglobin-Gene dieser Spezies Introns besitzt. Alle neu identifizierten Hämoglobin-Gene werden daher auf ihren Gehalt an Introns hin überprüft. Um zu beurteilen, ob neu identifizierte Introns ursprünglicher Natur sind, können ihre Positionen mit denen anderer Invertebraten- und Pflanzen-Introns verglichen werden. Identische Positionen sprechen hierbei für die Orthologie von Introns. Eine zufällige Intronverteilung würde hingegen auf unabhängige Insertionsereignisse im Sinne der „intron-late“-Hypothese hinweisen.

Die funktionelle Bedeutung der Hämoglobin-Multiplizität bei Chironomiden ist nach wie vor unklar. Gewährleistet die Vielzahl unterschiedlicher Hämoglobin-Varianten ausschließlich die Bereitstellung einer ausreichenden Hämoglobindosis oder besitzen die verschiedenen Hämoglobine aufgrund unterschiedlicher Eigenschaften (z. B. unterschiedlicher O<sub>2</sub>-Affinitäten) funktionell differenzierte Aufgaben? Es sollte möglich sein, Gene, die in mehreren Arten konserviert sind und daher als essentiell betrachtet werden können, zu identifizieren. Die Anzahl der Nukleotid- und Aminosäure-Austausche zwischen diesen Hämoglobin-Varianten kann hierbei als Maß für die funktionelle Konservierung einzelner Genkopien dienen.

## 2 Material und Methoden

### 2.1 Versuchstiere

Für die vorliegende Untersuchung wurden Zuckmückenlarven der Art *C. tentans* (Chironomidae, Diptera) verwendet. Die Gelege wurden von Herrn Prof. M. Lezzi zur Verfügung gestellt und stammten aus der permanenten Laborzucht des Inst. für Zellbiologie der Eidgenössischen Technischen Hochschule in Zürich. Nach Vereinzelung der Gelege wurden die Larven bis zum vierten Larvenstadium in belüfteten Schalen bei Raumtemperatur aufgezogen. Als Futter diente insektizidfreies Brennesselpulver und Zellstoff. Die Spezies wurde anhand ihres Chromosomensatzes bestimmt (Keyl und Strenzke, 1956; Keyl, 1962).

### 2.2 Standardmethoden

Die elektrophoretische Auftrennung von DNA-Fragmenten erfolgte nach Zugabe von 0,2 Vol. DNA-Stoppuffer (siehe Kapitel 2.12) auf 0,8-1,6%igen vertikalen Agarosegelen (2-4 mm dick) bei einer Stromstärke von 25-30 mA. Als Laufpuffer diente 1x TBE- oder 1x E-Puffer. Anschließend wurden die Agarosegele in einer Ethidiumbromidlösung (5 µg/ml in 1x E-Puffer) gefärbt. Im UV-Licht (302 nm) konnte die DNA sichtbar gemacht und mit einer digitalen Kamera fotografiert werden. Als Molekulargewichtstandards dienten mit HindIII restringierte Lambda-DNA und mit AluI restringierte pBR322-DNA. Die Wiedergewinnung der DNA-Fragmente aus dem Agarosegel erfolgte entweder durch Elektroelution (McDonell et al., 1977) oder alternativ mit Hilfe des „QIAquick Gel Extraktion Kits“ der Firma QIAGEN (Hilden) nach Angaben des Herstellers. Die Fällung von DNA erfolgte nach Zugabe von 0.1 Vol. 10x Dialysepuffer und 2 Volumen Ethanol abs. für 10-30 min bei -20°C. Nach 15-30 minütigem Abzentrifugieren der DNA bei 1000 g und 4°C wurde das Pellet mit 70%igem Ethanol gewaschen und erneut 5 min zentrifugiert. Die in einer Vakuumzentrifuge getrocknete DNA wurde in Aqua bidest gelöst. Die aus Geweben, Bakterien oder Phagen isolierte DNA wurde durch mehrfaches Extrahieren mit 1 Vol. Phenol (Tris pH 7,6-gesättigt) gereinigt. Um gelöstes Phenol aus der wässrigen DNA-Phase zu entfernen, folgte hierauf eine Extraktion mit 1 Vol. eines Chloroform/Isoamylalkohol (24:1) Gemisches. Die Restriktion von DNA wurde unter den von dem Enzymhersteller empfohlenen Bedingungen durchgeführt.

RNA-Moleküle wurden mit Hilfe von vertikalen 0,8-1,2%igen Agarosegelen mit einer Endkonzentration von 5,5% Formaldehyd elektrophoretisch aufgetrennt. Als

Laufpuffer diente 1x MOPS. Zur Denaturierung von Sekundärstrukturen wurde die aufzutrennende RNA-Probe mit 1  $\mu$ l Ethidiumbromid-Stammlösung (5 mg/ml) sowie mit 3 Vol. RNA-Lade-Puffer vermischt und für 10 min bei 70°C inkubiert. Nach Abkühlen der Probe auf Eis und Zugabe von 1/10 Vol. RNA-Färbelösung wurde sie auf das Gel geladen. Nach erfolgter Elektrophorese konnte das Agarosegel unter UV-Licht mit einer digitalen Kamera aufgenommen werden.

Restriktionskarten von Lambda- bzw. Plasmid-Klonen wurden durch kombinierten Verdau mit verschiedenen Restriktionsendonukleasen und anschließender Southern-Hybridisierung (siehe Kapitel 2.8.1) ermittelt.

### 2.3 Konstruktion von genomischen DNA-Bibliotheken

Für die Konstruktion genomischer DNA-Bibliotheken stand hochmolekulare DNA von *C. tentans* Larven des 3.-4. Larvenstadiums zur Verfügung. Diese DNA wurde mit dem Restriktionsenzym Sau3A unvollständig verdaut und auf einem 0,8%igen Agarosegel mit 1x E-Puffer/0,1% SDS elektrophoretisch aufgetrennt. DNA-Fragmente einer Größe zwischen 14 kB und 24 kB wurden durch Elektroelution wiedergewonnen und für die Ligation in die Sau3A-kompatible BamHI-Schnittstelle des Lambda-vektors EMBL 3 eingesetzt (Frischauf et al., 1983). Hierbei wurden äquimolare Mengen der mit Sau3A restringierten Genom-DNA und ligationsbereite Lambda-Arme (Stratagene Cloning Systems, Heidelberg) in einem Volumen von 20  $\mu$ l mit Hilfe der T4-Ligase (Boehringer, Mannheim) nach den Empfehlungen des Herstellers verknüpft.

Der Phagentiter der Genbanken konnte durch Ausplattieren verschiedener Verdünnungsstufen und Auszählen der Plaques bestimmt werden. Als Wirtsbakterium zur Vermehrung rekombinanter EMBL3 Phagen diente der Bakterienstamm *Escherichia Coli* P2392 (Stratagene Cloning Systems, Heidelberg). Die Lambda-phagen wurden mit 250  $\mu$ l der Wirtsbakteriensuspension ( $OD_{600} = 0,7$ ; gelöst in 10 mM  $MgCl_2$ ) für 20 min bei 37°C inkubiert. Mit 3 ml Weichagar (max. 50°C) gemischt wurde dieser Präadsorptionsansatz auf Agar-Platten ausgegossen und über Nacht bei 37°C bebrütet. Für die Selektion von Phagenklonen wurden jeweils  $10^6$  Phagen einer Genbank ausgeplattet, ihre Identifizierung erfolgte durch Plaquefilterhybridisierung (siehe Kapitel 2.8.2).

### 2.4 Präparation von DNA

#### 2.4.1 Präparation genomischer DNA von Zuckmückenlarven

Die Isolierung von genomischer DNA aus Chironomiden des 4. Larvenstadiums erfolgte mit geringen Modifikationen nach der von Schmidt et al. (1980) entwickelten Methode. Abweichend hierzu erfolgte die Kernlyse in 1% SDS/1x Dialysepuffer. Nach

zweistündiger Inkubation bei 60°C wurden 0,1 Vol. gesättigter Tris-Lösung pH 8,5 und 0,25 Vol. 5M NaClO<sub>4</sub> zugegeben. Zur Reinigung der DNA wurden mehrere Extraktionen mit Tris-gesättigtem Phenol pH 7,6 und eine Extraktion mit Chloroform-Isoamylalkohol (24:1) durchgeführt. Die Fällung der DNA erfolgte bei Raumtemperatur.

#### 2.4.2 Präparation von Phagen-DNA

Phagen-DNA wurde nach einer von Marzluff entwickelten Methode (unveröffentlicht) gewonnen. Durch Plauefilterhybridisierung identifizierte Phagenklone wurden aus der Agar-Platte ausgestochen und zusammen mit einigen Tropfen Chloroform in 500 µl SM-Medium (siehe Kapitel 2.12) überführt. Nach Amplifikation der Klonen durch erneutes Ausplattieren einer solchen Phagensuspension und Waschen der Platten mit SM-Medium erfolgte die Präadsorption der Phagen an die Wirtsbakterien wie in Kapitel 2.3 beschrieben (Phagen: Bakterien-Verhältnis 10<sup>6</sup>:10<sup>8</sup>). Der gesamte Präadsorptionsansatz wurde in 50-250 ml NZY-Medium oder L-Medium mit 10 mM MgCl<sub>2</sub> überführt und bei 37°C bis zur Lyse der Zellen geschüttelt (ca. 12-20 h). Das Abtöten der Bakterien erfolgte durch Zugabe einiger Tropfen Chloroform. Nach der Inkubation des Lysats für mindestens 30 min bei 4°C konnten die Bakterienreste durch Zentrifugation (3800 x g, 15 min, 4°C) abgetrennt werden. Die Fällung der Phagenpartikel aus dem klaren Überstand erfolgte durch Zugabe von NaCl (0,06 g/ml) und PEG 6000 (0,07 g/ml) in Eiswasser für mindestens 1 Stunde. Das durch Zentrifugation (9000 Upm, 4°C, 30 min) gewonnene Phagenpellet wurde in 2 ml SM-Medium/10mM MgCl<sub>2</sub> resuspendiert. Bakterielle DNA und RNA wurde durch Zugabe von 25 µl DNase I (1 mg/ml) und 5 µl RNase A (10 mg/ml) 30 min bei 37°C verdaut. Durch 10 minütige Inkubation mit 0,1 Vol. SDS-Mix bei 37°C und anschließenden Extraktionen mit 1 Vol. Tris-gesättigtem Phenol wurden die Phagenköpfe aufgebrochen und die Phagen-DNA von den Phagenhüllproteinen getrennt. Nach Extraktion mit 1 Vol. eines Chloroform-/Isoamylalkohol-Gemisches (24:1) zur Beseitigung von Phenolresten aus der DNA-Lösung erfolgte die Fällung der DNA mit 0,1 Vol. 2M NaCl und 2 Vol. EtOH.

Alternativ zu dieser Methode wurde Phagen-DNA mit dem Lambda-Kit der Firma QIAGEN (Hilden) präpariert. Die so gewonnene Phagen-DNA konnte direkt zur Sequenzierung eingesetzt werden.

#### 2.4.3 Präparation von Plasmid-DNA

Kleinere Mengen Plasmid-DNA zur Verifizierung rekombinanter Bakterienklone wurden mit Hilfe des TELT-Puffers ( Triton x 100/ EDTA/ Lithiumchlorid/ Tris-HCl) nach der Methode von He et al. (1989) gewonnen. Nach der von Clewell und Helinski (1969) entwickelten und durch Schmidt et al. (1982) modifizierten 'cleared lysate'-Methode wurden größere Mengen Plasmid-DNA für die Restriktionskartierung und

Subklonierung gewonnen. Diese Methode beruht auf dem Prinzip der schonenden Lyse der Bakterienzellwand durch Lysozym bei Anwesenheit von EDTA.

Plasmid-DNA für die nicht-radioaktive Sequenzierung wurde mit Hilfe des ‘QIAwell-8 Plasmid Kits‘ bzw. des ‘QIAprep Spin Plasmid Kits‘ der Firma QIAGEN (Hilden) nach Angaben des Herstellers gewonnen. Diese DNA-Isolierung beruht auf dem Prinzip der alkalischen Lyse (Birnboim und Doly, 1979).

## 2.5 Polymerase-Kettenreaktion (PCR)

### 2.5.1 PCR aus genomischer und klonierter DNA

Mit Hilfe der PCR-Technik (Saiki et al., 1988) wurden durch Einsatz spezifischer Primer genomische sowie klonierte DNA-Abschnitte selektiv amplifiziert. 50-100 ng genomische DNA (bzw. 10 ng Plasmid-DNA) wurden in 50 µl Ansätzen mit 1x Taq-Puffer (40 mM KCl; 10 mM Tris-HCl pH 8,3; 0,1 mg/ml Gelatine; 1,5 mM MgCl<sub>2</sub>) je 0,2 mM der vier dNTPs, je 10 nmol der beiden Primer und 2,5 U Taq-Polymerase amplifiziert. Vor Zugabe der Primer, der dNTPs und der Polymerase wurde die DNA für 3 min bei 96°C denaturiert und der Ansatz in Eiswasser überführt. Die Amplifikation erfolgte im Standardfall in dem Thermocycler PTC-200 (MJ Research, USA) unter folgenden Bedingungen:

Denaturierung: 94°C, 1,0 min;  
„Annealing“: 52°C, 1,0 min;  
Polymerisation: 72°C, 2,5 min;  
33 Zyklen

Zur Überprüfung der Amplifikation wurden 1/10 Vol. des Reaktionsansatzes gelelektrophoretisch aufgetrennt. Die Subklonierung der Amplifikationsprodukte erfolgte nach Ausschneiden der Banden und Elution oder direkt nach Fällung des PCR-Ansatzes. Mit Hilfe des ‘QIAquick PCR Purification Kits‘ (QIAGEN, Hilden) gereinigte DNA-Fragmente konnten direkt für die Sequenzierung eingesetzt werden.

### 2.5.2 Inverse-PCR

Um Sequenzen zu erhalten, die an bereits klonierte DNA-Abschnitte angrenzen, wurde die Methode der Inversen-PCR (Innis et al., 1990) durchgeführt. Dabei wurde ca. 1 µg genomischer *C. tentans* DNA in getrennten Ansätzen mit Restriktionsendonukleasen verdaut, die eine 4 bp lange Erkennungssequenz aufweisen. Bevorzugt wurden hierbei Enzyme ausgewählt, die bereits eine Schnittstelle nahe der unbekannt Sequenz aufwiesen. Die entstandenen Restriktionsfragmente wurden mit

sich selbst religiert. Daraufhin konnte mit einem Primerpaar aus der bereits bekannten Sequenz die angrenzende genomische DNA in einer PCR-Reaktion amplifiziert werden. Die Reaktionsbedingungen entsprachen hierbei den unter 2.5.1 beschriebenen. Die Sequenzierung der PCR-Produkte erfolgte auch hier nach Aufreinigung mit dem „QIAquick PCR Purification Kit“ (QIAGEN, Hilden).

### 2.5.3 RT-PCR (Reverse Transkriptase-Polymerasekettenreaktion) mit intronspezifischen Primern

Zur exakten Lokalisation der Intronengrenzen in dem Hämoglobin-Gen Cte N wurden spezifisch cDNA-Fragmente dieses Gens in einer PCR-Reaktion amplifiziert. Dazu wurde zunächst eine einzelsträngige cDNA aus ca. 1 µg larvaler *C. tentans* gesamt RNA mit Hilfe des Thermoscript RT-PCR Systems (Gibco BRL Life Technologies GmbH, Karlsruhe) unter Verwendung eines genspezifischen Primers (pten61, siehe CD-Datenträger im Anhang; Ordner: Primer ) nach Angaben des Herstellers synthetisiert. Bei dem verwendeten Enzym „Superscript RT“ handelt es sich um eine gentechnisch modifizierte Reverse Transkriptase aus M-MLV (Moloney Murine Leukemia Virus) ohne RNase Aktivität. Die cDNA wurde anschließend mit dem ‘QIAquick PCR Purification Kits‘ (QIAGEN, Hilden) gereinigt. Ein Aliquot (0,1-0,2 Vol.) dieser cDNA diente dann als Matrize für die Amplifikation mit weiteren genspezifischen Primerpaaren (pten131/pten165 bzw. pten131/pten166, siehe CD im Anhang) unter den in Kapitel 2.5.1 beschriebenen Bedingungen.

## 2.6 Oligonukleotidsynthese

Die Synthese der für die Sequenzierung und PCR verwendeten Primer erfolgte auf den Oligonukleotidsynthesegeräten „Gene Assembler Plus“ (Pharmacia, Freiburg) oder „392A RNA/DNA Synthesizer“ (Applied Biosystems GmbH, Weiterstadt). Nach erfolgter Synthese wurden die Oligonukleotide durch eine 16 stündige Inkubation bei 55°C in konzentrierter Ammoniak-Lösung abgespalten und anschließend auf einer NAP-10-Säule (Pharmacia Biotech GmbH, Freiburg) entsalzt. Die Ausbeute der Synthese wurde photometrisch ermittelt und belief sich auf 50-300 µg/ml DNA. Zusätzlich wurden Oligonukleotide bei der Firma Gibco BRL Life Technologies GmbH (Karlsruhe) bezogen. Die Nukleotid-Sequenzen aller, in der vorliegenden Arbeit verwendeten Primer, sind auf der beiliegenden CD aufgeführt.

## 2.7 Subklonierung in Plasmidvektoren

Zur Subklonierung von Restriktionsfragmenten sowie von PCR-Produkten wurde der Plasmidvektor pUC18 verwendet (Yanisch-Perron et al., 1985). pUC 18 ist doppelsträngig und zirkulär mit einer Länge von 2686 bp. Er trägt Gene, die für das Enzym

$\beta$ -Lactamase (Ampicillinresistenz) und für einen Teil der  $\beta$ -Galaktosidase (*lacZ'*-Gen) codieren. Wird das *lacZ'*-Gen durch Integration von Fremd-DNA in eine der Klonierungsschnittstellen zerstört, sind transformierte Bakterien des Wirtsstammes *E. coli* RRI (M15 Rütter (1982)) nicht in der Lage, durch Alpha-Komplementation eine funktionstüchtige  $\beta$ -Galaktosidase zu bilden (Langley et al., 1975). Dies ermöglicht die Farbselektion rekombinanter Bakterien auf IPTG / X-Gal-haltigen Nährböden.

Restriktionsfragmente mit nicht-kompatiblen Schnittstellen sowie PCR-Produkte wurden durch das Klenow-Fragment der *Escherichia coli* DNA-Polymerase I mit glatten Enden versehen (Maniatis et al., 1982).

Die Ligation von 0,5-1  $\mu$ g geschnittener Vektor DNA mit 1-5  $\mu$ g Fremd-DNA erfolgte mittels T4-DNA-Ligase (Boehringer, Mannheim) in einem Volumen von 15-20  $\mu$ l über Nacht bei Raumtemperatur. Der Ligationsansatz wurde anschließend mit Hilfe des „QIAquick Purification Kits“ (QIAGEN, Hilden) gereinigt.

Die Transformation der Bakterien erfolgte entweder durch eine modifizierte  $\text{CaCl}_2$ -Methode (Mandel und Higa, 1970) oder durch Elektroporation (1500-2000 Volt, 5 ms Impulsdauer). Für die Herstellung von elektrokompenten Bakterien wurden 2 ml einer Übernacht-Kultur in 500 ml L-Medium überimpft und bis zu einer optischen Dichte  $\text{OD}_{600} = 0,6-0,8$  bei 37°C geschüttelt. Die Bakterien wurden abzentrifugiert (4000 x g, 15 min, 4°C) und mehrmals mit 100 ml eiskaltem A. bidest gewaschen. Der letzte Waschschritt erfolgte mit eiskaltem A. bidest/10% Glycerin. Das Bakterienpellet wurde schließlich in 1 Vol. eiskaltem A. bidest/10% Glycerin resuspendiert, in 50  $\mu$ l Portionen aliquotiert und in flüssigem Stickstoff schockgefroren. Die elektrokompenten Bakterien wurden bei -70°C gelagert. Die Selektion positiver Klone erfolgte mittels Restriktionsanalyse der Plasmid-DNA oder durch Koloniefilterhybridisierung (siehe Kapitel 2.8.3). Bakterienkulturen positiver Klone wurden in einem Glycerin/ 2x L-Medium (1:1) Gemisch bei -20°C als Dauerkultur gelagert.

## 2.8 Hybridisierungsmethoden

### 2.8.1 „Southern-Blot“ Hybridisierung

In Anlehnung an die von Southern (1975) entwickelte Technik wurden DNA- Restriktionsfragmente von Agarosegelen auf Nitrocellulose oder Nylonmembranen übertragen und anschließend mit radioaktiv markierter DNA hybridisiert. In der Regel wurden mit der in Sambrook et al. (1989) beschriebenen ‘Sandwich’-Technik parallel zwei Filter aus einem Gel hergestellt. Die DNA wurde bei 80°C für 2 Stunden auf dem Nitrocellulosefilter fixiert und dieser in 2x PM (Denhardt, 1966) / 1% SDS für mindestens 3 Stunden zur Absättigung freier Bindungsstellen bei 60°C präinkubiert. In Anschluß an die Hybridisierung mit einer in 1x PM/ 0,5% SDS gelösten, radioaktiv markierten DNA Sonde (65°C für 6-15 Stunden) wurden die Filter dreimal in 2x SSC und einmal in 0,1x SSC/ 0,1% (w/v) SDS bei 65°C für jeweils 30 min gewaschen, um freie radioaktive Sondenmoleküle zu entfernen. Für die Hybridisierung



von Chironomiden-DNA mit einer heterologen Sonde aus anderen Spezies wurden weniger stringente Temperaturbedingungen (55-60°C) gewählt. Gewaschen wurde in diesem Fall ausschließlich mit 2x SSC bei 60°C. Zur Detektion der Signale wurden die getrockneten Blots mit einem Röntgenfilm (Kodak Xomat AR oder Fuji-RX) und Verstärkerfolie (Dupont Cronex Hi-Plus, Dr. Groos-Suprema) bei -80°C exponiert. Für die nichtradioaktive Hybridisierung wurde die DNA auf eine Nylon-Membran (Boehringer Mannheim) transferiert und bei 120°C für 2 Stunden fixiert. Nach Präinkubation in Hybridisierungspuffer ohne Formamid (hergestellt nach Anweisung des AMPPD-Herstellers Roche, Mannheim) bei 65°C erfolgte die Hybridisierung mit 10-20 ng/ml DIG-markierter DNA (siehe Kapitel 2.9) bei 65°C für 14 Stunden. Als Hybridisierungspuffer diente 1x PM/ 0,5% SDS. Das Waschen der Blots erfolgte unter stringenten Bedingungen (2x 30 min in 2x SSC, 2x 30 min in 0,1x SSC/ 0,1% (w/v) SDS, bei 65°C). Der Nachweis DIG-markierter DNA Sonden erfolgte mit dem Chemilumineszenzsubstrat AMPPD (3-(2'Spiroadamantan)-4-methoxy-4-(3''-phosphoryloxy)-phenyl-1,2-dioxetan, Boehringer, Mannheim) . Dieses wird durch die alkalische Phosphataseaktivität des Anti-Digoxigenin-AP-Konjugats (Boehringer, Mannheim) in das instabile AMPD überführt, dessen Zerfallsprodukt Licht der Wellenlänge 477 nm emittiert. Die Durchführung erfolgte nach dem Protokoll des Reagenzienherstellers. Nach Inkubation mit dem Anti-DIG-AP-Konjugat und der AMPPD-Lösung wurde der in Frischhaltefolie eingeschlagene Blot mit einem Kodak X Omat AR-Film 30 min exponiert.

### 2.8.2 Plauefilterhybridisierung

Nach der von Benton und Davis (1977) entwickelten Methode konnten Lambda-Klone mit Hilfe radioaktiver Sonden aus Genbanken selektiert werden. Zum Transfer der Phagen-DNA aus den Plaques wurden ein Nitrozellulose Filter für 4 min auf die gekühlte Agarplatte gelegt und seine Orientierung markiert. Nach Denaturierung der Phagen-DNA auf mit 0,1 M NaOH/ 1,5 M NaCl getränktem Filterpapier für 2 min und anschließender Neutralisierung mit 0,2 M Tris/ HCl pH7,5/ 2x SSC (2 min) wurden die Filter zweimal 15 min in 2x SSC gewaschen, 2 Stunden bei 80°C fixiert und wie unter 2.8.1 beschrieben hybridisiert, gewaschen und autoradiographiert. Plaque-Bereiche, die nach der Autoradiographie mit einem Signal korrespondierten, wurden aus dem Weichagar ausgestochen und in 500 µl SM-Medium überführt, dem einige Tropfen Chloroform zugesetzt waren. Diese Phagensuspension wurde erneut ausplattiert und einer weiteren Plaquefilterhybridisierung unterzogen, bis eindeutig Einzelplaques isoliert werden konnten. Diese wurden wiederum aus dem Weichagar ausgestochen und in 500 µl SM-Medium mit CHCl<sub>3</sub> überführt und bei 4°C dauerhaft aufbewahrt.

Zur Verifizierung positiver Phagenklone wurden „Plaque dots“ angefertigt. Hierzu wurden 200 µl des Bakterienstammes p2392 (in 10 mM MgCl<sub>2</sub> ) mit 3 ml Weichagar ausgeplattiert. 20 µl der zu testenden Phagensuspension wurden dann auf die

erhärtete Oberfläche an definierte Stellen aufgetropft. Die Platten wurden über Nacht bei 37°C bebrütet. Anschließend wurde eine erneute Plaquefilterhybridisierung durchgeführt.

### 2.8.3 Koloniefilterhybridisierung

Rekombinante Plasmidklone konnten in größerer Anzahl nach der von Grunstein und Hogness (1975) entwickelten Methode selektiert werden. Es wurden zu untersuchende Bakterienkolonien an identische Positionen auf zwei Nitrozellulosefilter überimpft, die zuvor auf Ampicillin-haltige (100 µg/ml) Nährböden gelegt worden waren und über Nacht bei 37°C angezüchtet. Einer der beiden Replikafilter wurde zur Denaturierung der Bakterien-DNA für 7 min auf ein mit 0,5 M NaOH getränktes Filterpapier gelegt, getrocknet und für 5 min auf einem mit 1,5 M Tris-HCl pH 7,4 getränkten Filterpapier neutralisiert. Der Nitrozellulosefilter wurde 5 min in 0,5 M Tris/ 1,5 M NaCl pH7,4 untergetaucht und erneut getrocknet. Es folgte der Abbau der bakteriellen Proteine durch Proteinase K (1 mg/ml in 1xSSC). Durch Tauchen des Nitrozellulosefilters in EtOH abs. wurde die Reaktion nach 30 min gestoppt. Der erneut getrocknete Koloniefilter wurde für 5 min in 0,3 M NaCl inkubiert und anschließend zwischen zwei Filterpapierbögen kurz gepreßt, um Bakterienreste zu entfernen. Nach dem Fixieren der DNA für 2 Stunden bei 80°C erfolgte die Hybridisierung und Detektion wie unter 2.8.1 beschrieben.

### 2.8.4 In situ-Hybridisierung an Polytänchromosomen

Zur Lokalisation klonierter DNA sowie von PCR-Produkten auf den Polytänchromosomen von Chironomiden wurde die Technik der FISH (fluorescence in situ hybridisation) angewandt.

Die Durchführung erfolgte entsprechend dem von Schmidt (1992) veröffentlichten Protokoll. Essigsäure- (50%) fixierte Chromosomenpräparate aus Tieren des vierten Larvenstadiums wurden hitzestabilisiert (Bonner und Pardue, 1976) und mit NaOH denaturiert (Singh et al., 1977). Auf eine Behandlung der Chromosomen mit RNase wurde verzichtet. Die Hybridisierung der so vorbereiteten Chromosomen mit Biotin- oder Digoxigenin-markierten DNA-Sonden (siehe Kapitel 2.9) erfolgte bei 55°C für 6-12 Stunden entsprechend der von Langer-Safer et al. (1982) beschriebenen Methode, die von Schmidt et al. (1988) für Chironomus optimiert wurde.

Nach erfolgter Hybridisierung und Detektion wurden 10 µl eines Gemisches aus p-Phenyldiamin und Glycerin auf das Präparat gebracht, diese mit einem Deckgläschen verschlossen. Die Präparate wurden mit einem confokalen Mikroskop betrachtet (Leica). Aufnahmen der Präparate wurden an dem Leitz Aristoplan-Photomikroskop (Film Ilford HP5) bzw. mit der digitalen SenSyn™-Kamera (Photometrics, München) gemacht. Die Bildverarbeitung erfolgte mit den Programmen „IPLab™ Spectrum

version 3.0.1“ (Signal Analytics Corporation, USA) und „Adobe Photoshop 4.0“ (Adobe Systems Inc., USA) für Macintosh Power PCs.

## 2.9 DNA-Markierung

### 2.9.1 Random-primed oligo-labelling

Die Herstellung radioaktiv oder Digoxigenin-markierter DNA-Sonden erfolgte nach einer von Feinberg und Vogelstein (1983) entwickelten Technik mit Hilfe von Markierungssystemen der Firma Boehringer (Mannheim). Diese Methode beruht auf der enzymatischen Synthese eines DNA-Stranges in Anwesenheit markierter dNTPs. Als Matrize diente 1-10 µg hitzedenaturierte DNA. Der Synthesestart wird durch Primer mit zufälliger Basenzusammensetzung garantiert. Bei radioaktiven Markierungen wurden 35 µCi <sup>32</sup>P-dATP (3000 Ci/mMol) eingesetzt. Zur Entfernung nicht-eingebauter radioaktiver Nukleotide wurde die DNA im Anschluß an die Markierungsreaktion gefällt. Die in Aqua bidest. gelöste und mit Carrier-DNA aus Kalbsthymus (Endkonzentration 200 µg/ml ) versehene Probe konnte nach der Hitzedenaturierung zur Hybridisierung eingesetzt werden. Die nicht-radioaktive DNA Markierung wurde mit dem „DIG DNA Labelling Kit“ nach Angaben des Herstellers Boehringer (Mannheim) durchgeführt. Eingesetzt wurden 1-3 µg DNA und 0,15 M Digoxigenin-11-dUTP.

### 2.9.2 Nick Translation

Bei dieser von MacGregor und Mizuno (1976) beschriebenen Methode werden durch die 5'-3' Exonukleaseaktivität der DNA Polymerase I entstehende einzelsträngige DNA-Bereiche durch Einbau z.T. markierter Nucleotide zum Doppelstrang ergänzt. Etwa 1 µg DNA wurde mit 10 µCi <sup>32</sup>P-dATP (3000Ci/mmol), 1µl einer 0,2 M Stammlösung der übrigen nicht-radioaktiv markierten Triphosphate, 10 µl 10x Salze, 1 µl einer DNase I Lösung (1 µg/ml) und 12 Einheiten DNA Polymerase I in einem Volumen von 100 µl für 1 Stunde bei Raumtemperatur inkubiert. Im Anschluß an die Markierungsreaktion wurde der Ansatz wie unter 2.8.1 beschrieben behandelt. Für die nicht-radioaktive DNA Markierung wurde das BioNick Labelling System der Firma Bibco BRL nach Anweisung des Herstellers verwendet.

## 2.10 Automatische DNA-Sequenzierung

Die auf der von Sanger et al. (1977) entwickelten Kettenabbruch-Methode beruhende automatische DNA-Sequenzierung wurde mit den Sequenziergeräten 373 stretch und PRISM™ 377 der Firma ABI (Weiterstadt) durchgeführt. Als Matrize dienten dabei sowohl superzirkuläre doppelsträngige Plasmid-DNA (Chen und Seeburg, 1985),

als auch gereinigte PCR-Produkte. Die Sequenzierung erfolgte als lineare DNA-Amplifikation durch „Cycle Sequencing“. Verwendet wurden dabei der „PRISM™ Ready Reaction Dye Deoxy Terminator Cycle Sequencing Kit“ (PE Applied Biosystems, Weiterstadt) und der „Thermo Sequenase™ dye terminator cycle sequencing pre-mix Kit“ (Amersham Life Science Inc., USA). Die Sequenzierreaktion erfolgte in der Regel nach einem 2-Schritt-Protokoll (0:15 min 96°C, 4 min 55°C; 25 Zyklen). Bei Sequenzierung von DNA-Bereichen, die aufgrund von repetitiven Elementen in der Lage sind, Sekundärstrukturen zu bilden, sowie für die direkte Sequenzierung von  $\lambda$ -DNA wurden das antifold-Programm (0:20 min 98°C, 0:15 min 55°C, 4 min 60°C; 25 Zyklen) sowie die doppelte Menge an Reaktionsmix benutzt. Die Anzahl der Reaktionszyklen wurde bei geringen Signalstärken erhöht (bis zu 60 Zyklen). Als Primer (10 pmol/Ansatz) dienten entweder die Standardprimer M13 „universal“ und M13 „reverse“ oder aber sequenzspezifische Oligonukleotide (siehe CD-Datenträger im Anhang; Ordner: Primer). Die Aufreinigung der Sequenzierproben sowie die Vorbereitung der Sequenzgele und die Gelelektrophorese erfolgten nach den Empfehlungen der Hersteller.

### 2.11 Computerauswertung von DNA- und Proteinsequenzen

Die Auswertung der DNA-Sequenzdaten erfolgte mit Hilfe der 373A-Software für Apple Macintosh-Computer (ABI, Weiterstadt) sowie mit der „Lasergene Sequence Analysis Software“ (DNASTAR Inc., USA). Das Programm „Sequencher™ 4.0“ (Gene Codes Corporation, USA) diente zur Bildung von „Sequenz-Contigs“ durch das Zusammenfügen überlappender Einzelsequenzen. Paarweise Sequenz-Vergleiche wurden mit dem Programm „SIM“ (Huang und Webb (1991); <http://www.expasy.ch/tools/sim-nucl.html>) auf dem ExPASy-Server (<http://www.expasy.ch/tools>) durchgeführt. Mit dem Programm LALNVIEW (Duret et al. (1996); <http://www.expasy.ch/tools/lalnview.html>) konnten die Ergebnisse der Sequenz-Vergleiche graphisch dargestellt werden.

Das Programm „Seqed“ („Lasergene Sequence Analysis Software“, DNASTAR Inc., USA) diente zur Übersetzung von Nukleotid-Sequenzen in Aminosäure-Sequenzen.

Ferner wurden Datenbankrecherchen mit den Algorithmen BLASTN, BLASTX, TBLASTX und FASTA (Altschul et al., 1990) über das World Wide Web (www) am „National Center for Biotechnology Information“ (NCBI, Bethesda, USA) durchgeführt. Durchsucht wurden dabei die Datenbanken GenBank/EMBL/DDBJ/PDB. Die „Homepage“ des NCBI findet sich unter der Adresse <http://www.ncbi.nlm.nih.gov/>.

Die phylogenetischen Auswertungen der Sequenzdaten erfolgte mit dem Programmpaket „MEGA 1.02“ (Molecular Evolutionary Genetics Analysis, Kumar et al.

(1993)). Als Grundlage für die Ermittlung der Sequenz-Verwandtschaften dienten Sequenz-Vergleiche, die mit Hilfe des Programms „Clustal W“ (Higgins und Sharp, 1988, 1989) erstellt und anschließend per Auge optimiert wurden. Ausgehend von diesem „Sequenz-Alignment“ wurden die Distanzwerte der einzelnen Sequenzen mit dem Program „MEGA“ (Kumar et al., 1993) berechnet. Die Hämoglobin-Sequenzen wurden hierbei paarweise verglichen und der Anteil der Aminosäure-Austausche an der Gesamtsequenz ermittelt (p-Distanzen).

Anteil der AS-Unterschiede  $p = \text{Anzahl der AS-Austausche} / \text{Sequenzlänge } n$   
 Mit der Variation  $V(p) = p(1-p)/n$

Die Ermittlung der Nukleotid-Distanzen erfolgte nach Jukes und Cantor (1969). Dieses Modell geht von einheitlichen Substitutionsraten aller vier Nukleotide sowie konstanten Evolutionsraten aus. Die Distanzwerte für synonyme- und nicht-synonyme Substitutionen wurden getrennt berechnet (Kumar et al., 1993). Lücken im „Sequenz-Alignment“, die auf Insertionen bzw. Deletionen von Nukleotiden zurückzuführen sind, wurden bei keiner der Methoden berücksichtigt. Die Konstruktion von Verwandtschaftsdentrogrammen erfolgte mit dem Program „PAUP“ (Swofford, 1993) basierend auf der „maximum parsimony“ Methode (Eck und Dayhoff, 1966; Fitch, 1977), sowie mit dem „neighbour joining“ Algorithmus (Saitou und Nei, 1987) aus dem Program „MEGA“ (Kumar et al., 1993).

Die Analyse von Protein-Strukturen erfolgte auf dem „Internet-Server“ des „Atelier BioInformatique“ der Universität Aix-Marseille (ABIM; <http://www-biol.univ-mrs.fr/english/logligne.html#predi>). Transmembrandomänen wurden mit dem Program „SOSUI“ (Hirokawa et al. (1998); [http://www.tuat.ac.jp/~mitaku/adv\\_sosui](http://www.tuat.ac.jp/~mitaku/adv_sosui)) vorhergesagt. Das Programm „Blocks“ (Henikoff und Henikoff (1994); Henikoff et al. (1999); <http://www.blocks.fhcrc.org/>) diente zur Identifikation funktioneller Proteindomänen.

## 2.12 Puffer und Lösungen

Denaturierungspuffer (Blot)	1,5 M NaCl 0,5 M NaOH
10x Dialyse-Puffer	0,25 M Tris 3 M NaCl 0,1 M Na <sub>2</sub> EDTA
DNA-Stoppuffer	4 M Harnstoff 0,1 M Na <sub>2</sub> EDTA 0,1 % w/v Bromphenolblau

## 2 *Material und Methoden*

---

	50 % w/v Saccharose
10x E-Puffer	0,36 M Tris 0,3 M NaH <sub>2</sub> PO <sub>4</sub> 0,1 M Na <sub>2</sub> EDTA
Ethidiumbromid Färbelösung	5 µg Ethidiumbromid / ml in 1x E-Puffer
L-Medium	10 g Trypton 5 g Hefeextrakt 5 g NaCl ad 1000 ml a. dest.
MOPS	0,2 M 3-(N-morpholin) Propan- Sulfonsure 0,05 M Na-Acetat 0,01 M Na <sub>2</sub> EDTA pH 7,0
Neutralisierungslösung (Blot)	0,5 M Tris-HCl pH 7,5 3 M NaCl
NZY-Medium, pH 7,5	10 g NZ Amin 5 g NaCl 5 g „Yeast Extract“ 2 g MgSO <sub>4</sub> x 7 H <sub>2</sub> O ad 1 l. a. bidest.
10x PBS (FISH)	200 g NaCl 5 g KCl 5 g KH <sub>2</sub> PO <sub>4</sub> 27,8 g Na <sub>2</sub> HPO <sub>4</sub> ad 1 l. a. bidest.
1x PM nach Denhardt (1966)	0,02% w/v Ficoll 400 0,02%w/v Polyvinylpyrrolidon 0,02% w/v BSA in 3x SSC
RNA-Färbelösung	50% Glycerin 0,001 M Na <sub>2</sub> EDTA pH 8,0 0,25% (w/v) Bromphenolblau 0,25% (w/v) Xylencyanol FF

RNA-Lade-Puffer	500 $\mu$ l Formamid (deionisiert) 100 $\mu$ l 10x MOPS-Puffer 150 $\mu$ l Formaldehyd (filtriert)
SDS-Mix	2,5% w/v SDS 0,25 M Na <sub>2</sub> EDTA 0,5 M Tris-HCl pH 8,0
SM-Medium	0,1 M NaCl 50 mM Tris-HCl pH 7,5 8 mM MgSO <sub>4</sub> 0,01% w/v Gelatine
20x SSC	3 M NaCl 0,3 M Natriumcitrat
10x Taq-Puffer	400 mM KCl 100 mM Tris-HCl pH 8,3 15 mM MgCl <sub>2</sub>
Tris-Borat-Elektrophorese-Puffer 20x TBE	1,8 M Tris 1,8 mM Borsäure 25 mM Na <sub>2</sub> EDTA
TE-Puffer	10 mM Tris-HCl pH 8,0 1 mM Na <sub>2</sub> EDTA
Tris-EDTA-Lithiumchlorid-Triton X-Puffer TELT-Puffer	50 mM Tris-HCl 62 mM Na <sub>2</sub> EDTA 2,5 M LiCl 4% v/v Triton X-100
Triton X-Mix	50 mM Tris-HCl pH 8,0 62,5 mM Na <sub>2</sub> EDTA pH 8,0 0,2% v/v Triton X-100

### 2.13 *E. coli*-Stämme:

HB101: *supE44 ara14 galK2 lacY1*  $\Delta$ (*gpt-proA*)62 *rpsL20* (Str<sup>r</sup>) *xyl-5 mtl-1 recA13*  
 $\Delta$ (*mcrC-mrr*) HsdS<sup>-</sup>(r<sup>-</sup> m<sup>-</sup>)

Dh5 $\alpha$ : *supE44*  $\Delta$ *lacU169* ( $\phi$ 80 *lacZ* $\Delta$ *M 15*) *hsdR17 recA1 endA1 gyrA96 thi-1*

P2392: *LE392* [*e14*<sup>-</sup>( *McrA hsdR514 supE44 supF58 lacY1* or  $\lambda$ (*lacIZY*)6 *galK2 galT22 metB1 trpR55*] P2-lysogen *relA1*

RR1: HB101 *recA*<sup>+</sup>

### 2.14 Materialien

Agar-Platten:

15 g Agar-Agar ad 1000 ml L-Medium. Hinzugefügt wird je nach Bedarf: Ampicillin 100  $\mu$ g/ml oder Kanamycin 10  $\mu$ g/ml oder X-Gal 0,01%, IPTG 0,2 mM.

Weichagar:

10 g Trypton, 5 g Hefeextrakt, 5 g NaCl, 5 g Agarose ad 1000 ml A. dest.

Dialyse-Schläuche (Medicell, London) für Elektroelution:

Durchschnitts-Porenradius: 28 Å, Durchmesser: 15,9 mm. Es wurden ca. 20 cm lange Streifen geschnitten, diese mit Siedesteinen im Becherglas in Natriumcarbonat 50 g/l A. bidest. (2x 10 Minuten) aufgekocht, zwischendurch mit A. bidest. gespült und in 3,72 g EDTA pro l. A. bidest. aufbewahrt.

### 2.15 Bezugsquellen

Nicht aufgeführte Chemikalien der Qualitätsstufen „p. A.“ bzw. „reinst“ wurden von den folgenden Firmen bezogen: Roche (Mannheim), Difco (USA), Fluka (Schweiz), Life Technologies (Eggenstein), Merck (Darmstadt), Pharmacia (Freiburg), Riedel-de Häen (Seelze), Roth (Karlsruhe), Serva (Heidelberg), Sigma (USA).

Agarosen	MP: Roche (Mannheim)
Ampicillin	Höchst (Frankfurt)
AMPPD	Roche (Mannheim)
DIG DNA Labeling Kit	Roche (Mannheim)
DIG-dUTP	Roche (Mannheim)
DNase I	Sigma (USA)
Elektroporationsküvetten	Eurogentec
IPTG	Biomol (Hamburg)
Klenow-Enzym	Roche (Mannheim)
T4-DNA-Ligase	Pharmacia (Freiburg), Roche (Mannheim)
Lysozym	Serva, (Heidelberg)
Mineralöl	Sigma (USA)
3 MM-Papier	Schleicher & Schüll (Dassel)



Nitrocellulosefilter	Schleicher & Schüll (Dassel)
Nylonmembranen	Roche (Mannheim)
$\alpha^{32}\text{PdATP}$ , $\alpha^{32}\text{PdCTP}$	ICN (Meckenheim)
Proteinase K	Merck (Darmstadt)
Random Primed DNA Labeling Kit:	Roche (Mannheim)
Restriktionsendonukleasen	Roche (Mannheim)
	Pharmacia (Freiburg)
	AGS (Heidelberg)
	Stratagene (Heidelberg)
	USB (USA)
RNase A	Serva (Heidelberg)
Röntgenfilm	X-O-MAT AR: Kodak
Röntgenfilm-Entwicklerkonzentrat	Adefo (Nürnberg)
Röntgenfilmfixierkonzentrat	Adefo (Nürnberg)
Verstärkerfolie (Cronex Hi-Plus)	Dupont; Dr. Goos-Suprema (Heidelberg)
X-Gal	Biomol (Hamburg)



## 3 Ergebnisse

### 3.1 Isolierung und Charakterisierung von $\lambda$ -Klonen mit Hämoglobin-Genen

Aus einer genomischen Genbank von *C. tentans* im Vektor EMBL3 wurden  $\lambda$ -Klone des Hämoglobin-Genclusters D isoliert. Als Sonde wurde eine Mischung bereits bekannter Hämoglobin-Gene von *C. tentans* verwendet, die aus den bereits isolierten und teilweise charakterisierten Klonen hb5 und hb10 stammten (Rozynek et al., 1991). Die so gewonnenen Klone hb20.1, hb2, hb15, hb11, hb3.2, hb18, hb9, hb4a und hb21 konnten mittels *in situ*-Hybridisierung an Polytänchromosomen der Hämoglobingen-Gruppe D zugeordnet werden. Ihre Kartierung erfolgte durch Restriktionsanalysen und Endsequenzierung. Zusammengefasst repräsentieren die Klone zwei Ausschnitte des Hämoglobin-Genclusters D von 50 kb (Contig 1) und 24 kb (Contig 2). In Abb. 3.1 ist die Anordnung der Lambda-Klone in diesen beiden Contigs dargestellt.

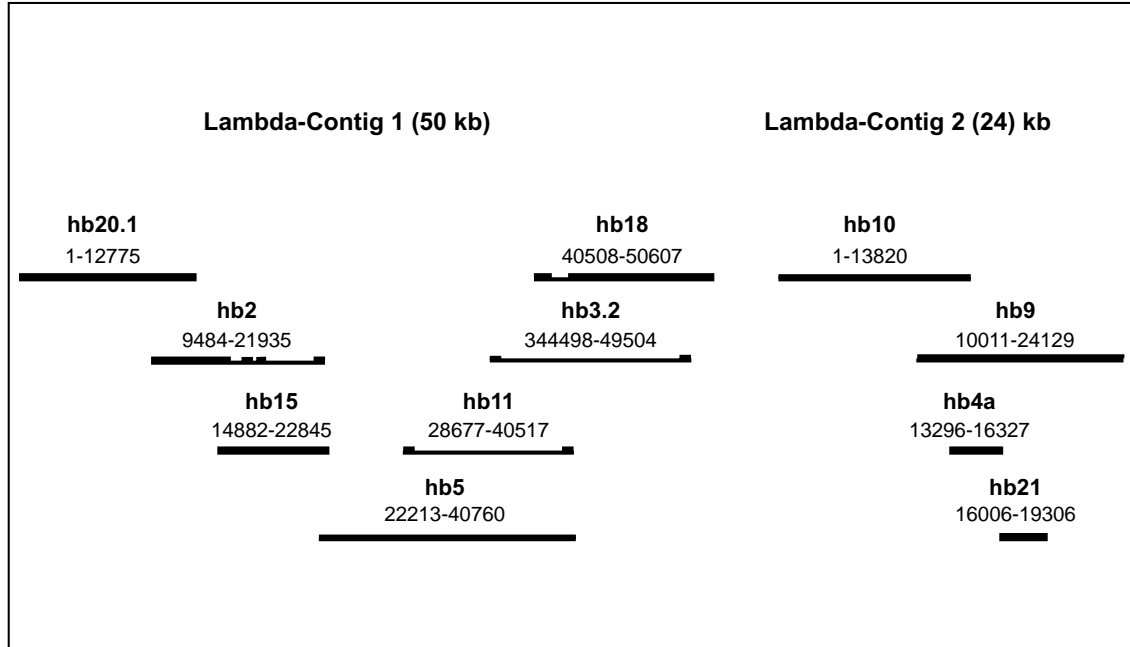


Abbildung 3.1 Legende siehe folgende Seite.

**Abbildung 3.1: Übersicht über die Lambda-Klone aus dem Bereich des Hämoglobingen-Clusters D von *C. tentans***

Ausgehend von den vorliegenden Klonen hb5 und hb10 (Rozynek et al., 1991) wurden weitere Klone aus einer genomischen Genbank von *C. tentans* isoliert. Diese bilden zwei zusammenhängende Sequenzbereiche (Contigs) von 50 kb bzw. 24 kb. Die Positionen der einzelnen Klone innerhalb der Contigs sind in Basenpaaren angegeben. Beide Sequenzbereiche wurden vollständig sequenziert. Die sequenzierten Abschnitte der einzelnen Klone sind durch breite Linien kenntlich gemacht. Die beiden Contigs überlappen nicht, der Abstand zwischen ihnen ist unbekannt.

## 3.2 Die chromosomale Organisation der $\lambda$ -Klone

Die beiden DNA-Abschnitte Contig 1 und Contig 2 überlappen nicht. Es ist derzeit unbekannt, wie groß der Abstand zwischen Contig 1 und Contig 2 im Genom von *C. tentans* ist. Ihre Anordnung auf dem Chromosom sowie ihre relative Orientierung wurden mittels Mehrfarben-*in situ*-Hybridisierung ermittelt. Dazu wurden in den Randbereichen der Contigs liegende Intergenregionen durch PCR amplifiziert und als Sonden wahlweise mit Digoxigenin bzw. Biotin markiert. Je zwei unterschiedlich markierte PCR-Fragmente wurden dann in Kombination zur Hybridisierung verwendet. Die so erstellten Sonden sind in Tabelle 3.1 charakterisiert. Die Hybridisierung erfolgte an Polytänchromosomen aus Speicheldrüsen von *C. tentans*-Larven des dritten oder vierten Larvalstadiums. Digoxigenin-markierte Sonden wurden mit TRITC-gekoppelten Antikörpern (rote Fluoreszenz), Biotin-markierte Sonden mit FITC-gekoppelten Antikörpern (grüne Fluoreszenz) nachgewiesen. Als zusätzlicher „natürlicher“ Marker zur Orientierung der Signale auf dem Chromosom wurde der Nukleolus verwendet. Dieser ist als Verdickung des Chromosoms im Mikroskop sichtbar. Er befindet sich zwischen dem Hämoglobin-Gencluster D und dem Centromer und ermöglicht es somit, die Lage der Hybridisierungssignale relativ zum Centromer zu bestimmen, auch wenn dieses nicht sichtbar ist.

### 3.2.1 Die Reihenfolge der Contigs 1 und 2 auf dem Chromosom

Um die Anordnung der beiden Contigs auf dem Chromosom festzulegen, wurden die Sondenkombinationen 1, 2 und 3 zur Hybridisierung verwendet. Sie bestehen jeweils aus Abschnitten beider Contigs mit unterschiedlicher Fluoreszenz-Markierung (siehe Tab. 3.1). Keine der Sondenkombinationen lieferte zwei vollständig voneinander getrennte Signale. Die zu beobachteten Signale setzen sich in allen drei Fällen jeweils aus einem roten und einem grünen Anteil zusammen, die extrem dicht beieinander liegen. Der gelbe Anteil des Signals ergibt sich aus der Überlagerung der roten und grünen Farbmarkierungen. Trotzdem war es bei genauer Betrachtung möglich, die

### 3.2 Die chromosomale Organisation der $\lambda$ -Klone

Sondenkombination	Position der PCR Produkte			PCR-Primer	Fragment-Länge (bp)	Markierung
	Contig:	$\lambda$ -Klon:	bp:			
1	Contig 1:	hb20.1,	379- 1306	tenp146/tenp221 tenp 49/tenp49x	927	FITC (grün)
	Contig 2:	hb10,	37- 837		800	TRITC (rot)
2	Contig 1:	hb20.1,	379- 1306	tenp146/tenp221 tenp255/tenp254	927	FITC
	Contig 2:	hb21,	17319-17783		464	TRITC
3	Contig 1:	hb20.1,	379- 1306	tenp146/tenp221 tenp 73/tenp 74	927	FITC
	Contig 2:	hb9,	14249-16069		1820	TRITC
14	Contig 1:	hb20.1,	379- 1306	tenp146/tenp221 tenp 88/tenp 98	927	TRITC
	Contig 1:	hb18,	47717-49123		1406	FITC

Sondenkombination	Abstand der PCR-Produkte (bp)	Orientierung der Signale	Abbildung
1	$\geq 51065$ *	grün zum Centromer (Contig 1)	3.2
2	$\geq 56379$ *	grün zum Centromer (Contig 1)	3.2
3	$\geq 59913$ *	grün zum Centromer (Contig 1)	3.2
14	=48744	rot zum Centromer (hb20.1)	3.3

**Tabelle 3.1:** Lokalisierung der Sonden für die *in situ*-Hybridisierung

Die Anordnung der DNA-Abschnitte von Contig 1 und Contig 2 auf den Chromosomen von *C. tentans* erfolgte mittels *in situ*-Hybridisierung an Polytänchromosomen der Speicheldrüsen. Flankierende Bereiche der Contigs wurden durch PCR amplifiziert. Die Lage der amplifizierten Bereiche sowie die verwendeten PCR-Primer sind in den ersten beiden Spalten aufgeführt (siehe CD-Datenträger im Anhang für die Primer-Sequenzen; Ordner: Primer). Je zwei unterschiedlich markierte PCR-Produkte wurden vereinigt und in Kombination als Sonde eingesetzt. Der Nachweis erfolgte mittels der fluoreszenzmarkierten Antikörper Anti-Dig- TRITC (rot) und Anti-Bio-FITC (grün).

\* Der Abstand zwischen diesen PCR-Produkten ist nicht bekannt, da zwischen Contig 1 und Contig 2 eine Lücke unbekannter Größe besteht. Die angegebene Distanz ist daher ein Mindestwert.

Reihenfolge der Farbmarkierungen eindeutig festzulegen. Die Ergebnisse der Hybridisierung sind in Abb. 3.2 dargestellt. Es zeigt sich, daß der Contig 1 dem Centromer des Chromosoms zugewandt ist (Sonden mit grüner Fluoreszenz), wohingegen der Contig 2 telomerwärts orientiert ist (Sonden mit roter Fluoreszenz).

A.

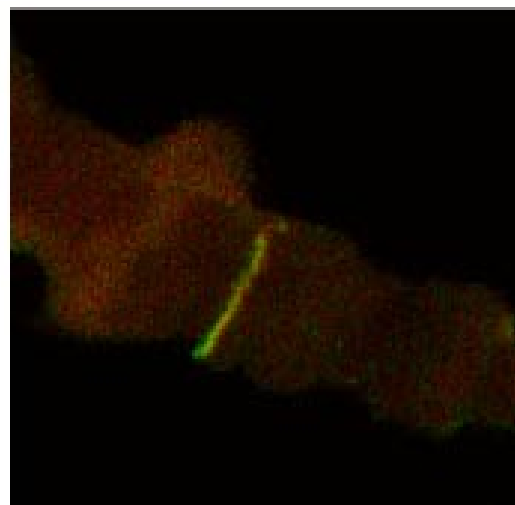
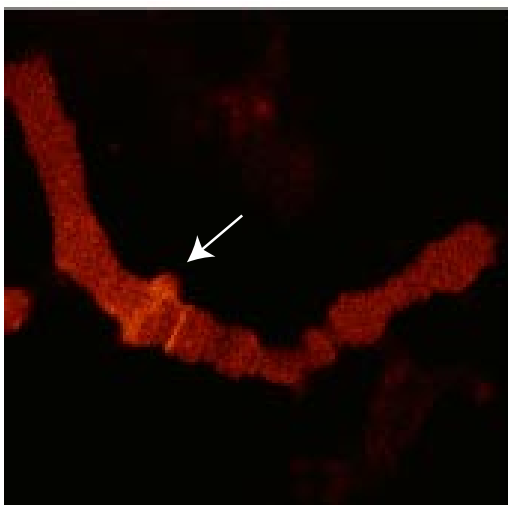
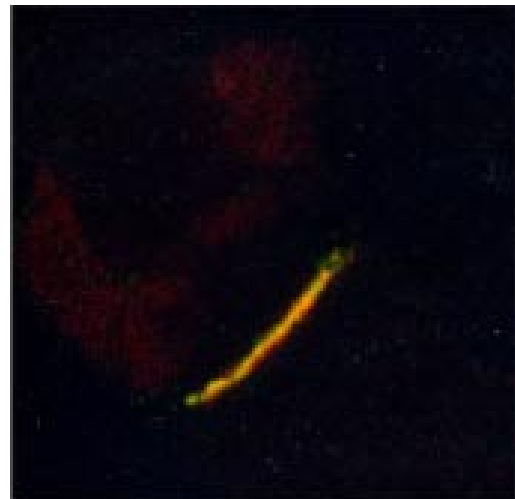
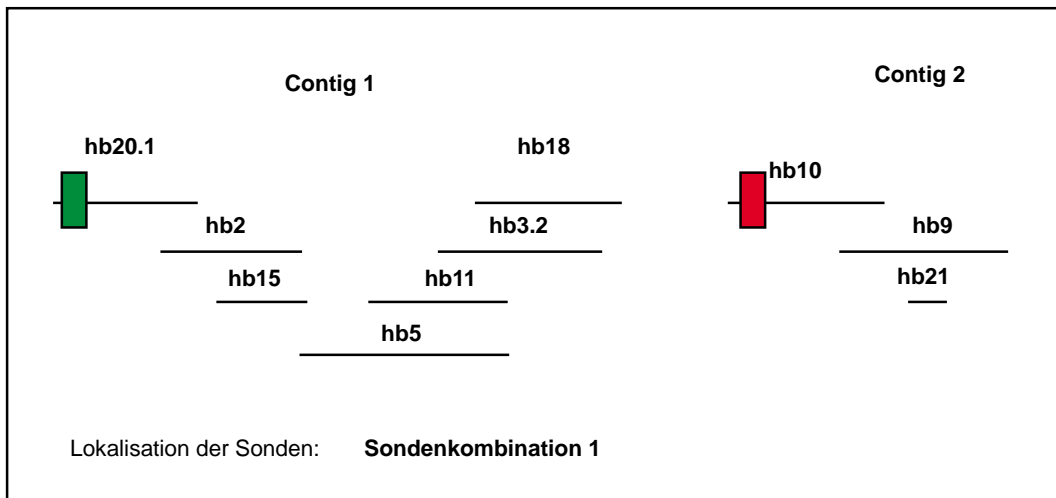


Abbildung 3.2: Legende siehe Seite 48

B.

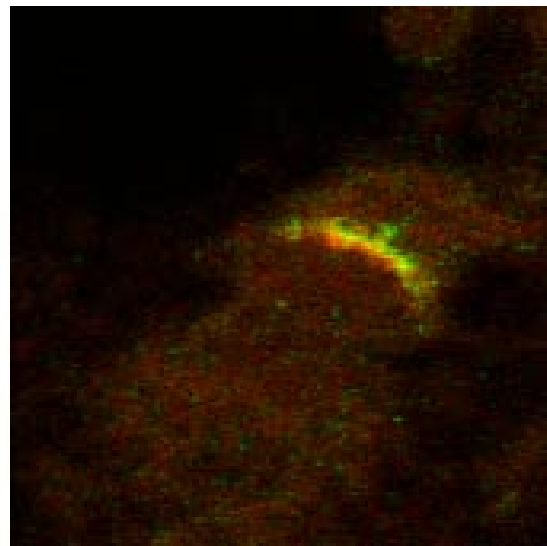
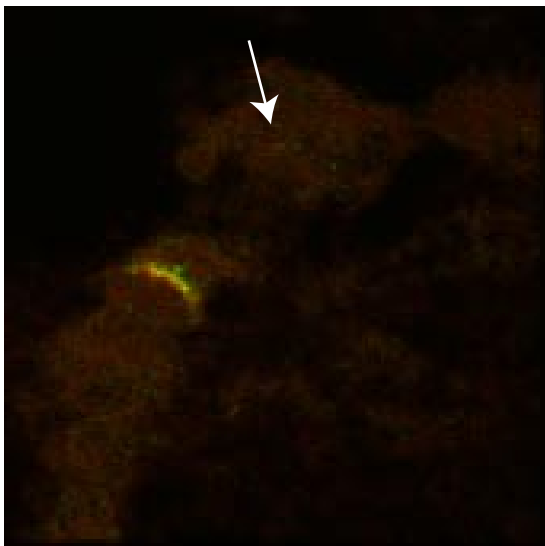
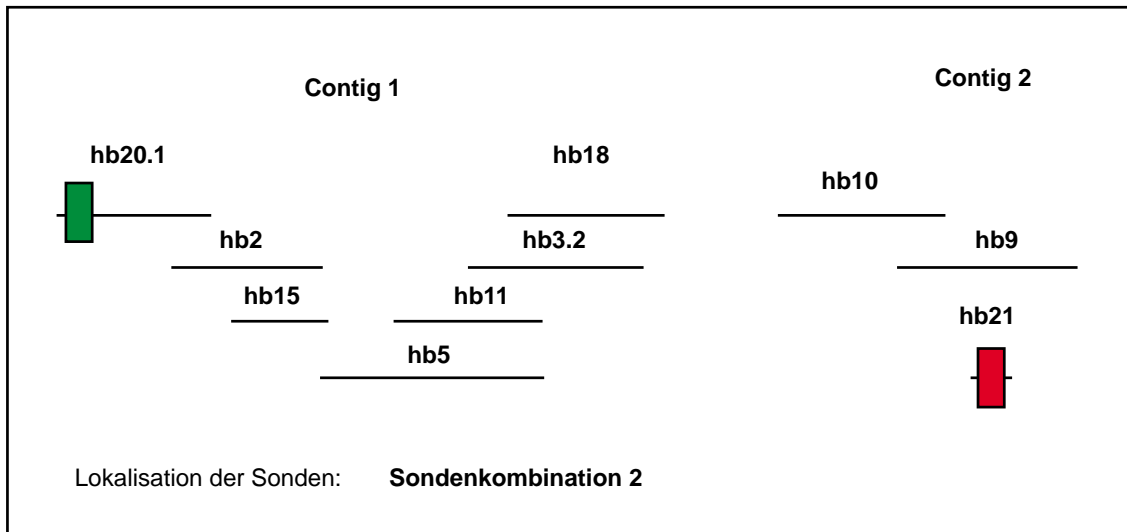


Abbildung 3.2: Legende siehe Seite 48

C.

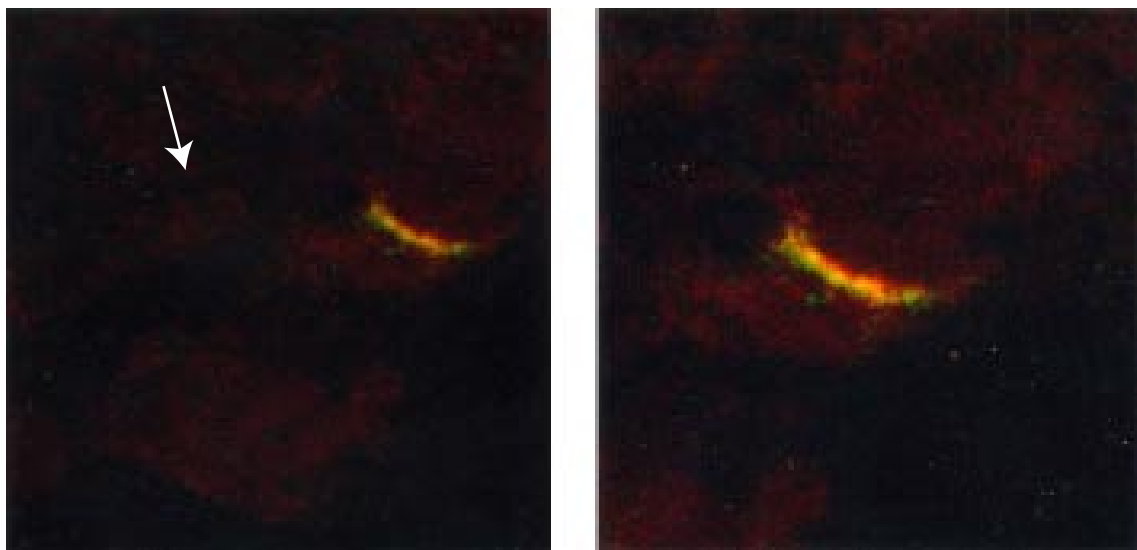
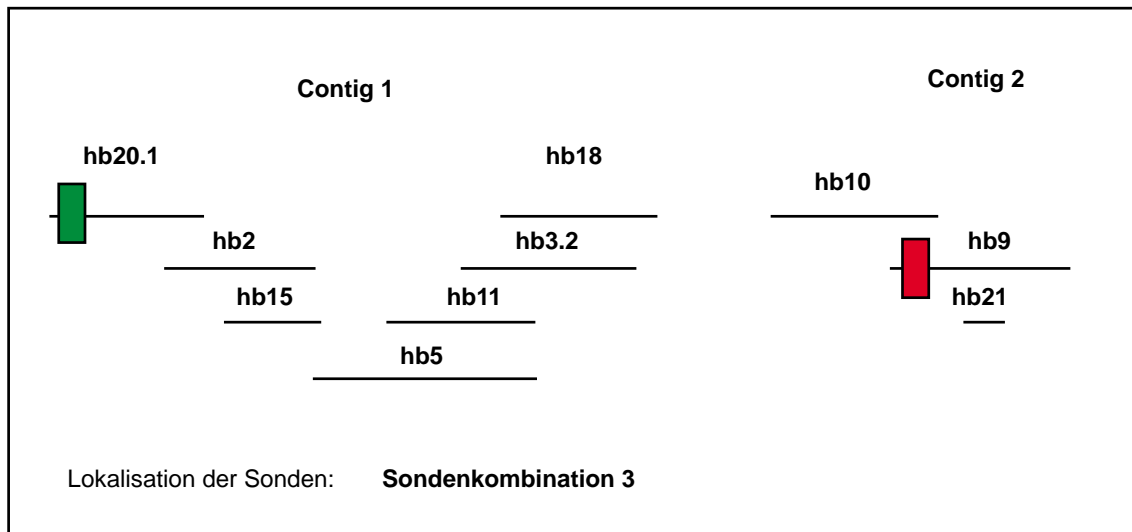


Abbildung 3.2: Chromosomale Anordnung der Klon-Contigs 1 und 2

Die Abbildungen A, B und C zeigen die Ergebnisse der *in situ*-Hybridisierung mit den Sondenkombinationen 1, 2 und 3 (zur Charakterisierung der Sonden vergleiche Tab. 3.1). Im oberen Bereich der Abbildungen ist jeweils die Lage der verwendeten Sonden innerhalb der beiden Contigs schematisch dargestellt (Kasten). Mit dem Fluoreszenz-Farbstoff FITC nachgewiesene Sonden erzeugen eine grüne, mit TRITC nachgewiesene Proben hingegen eine rote Fluoreszenz. Die Reihenfolge der Signale im Bild des confokalen Mikroskops zeigt, daß der Contig 1 näher zum Nukleolus (Pfeil) liegt und somit dem Centromer zugewandt ist, während der Contig 2 zum Telomer hin orientiert ist. Das rechte Bild stellt stets einen vergrößerten Ausschnitt des linken Übersichtsbildes dar.



### 3.2.2 Die Orientierung des Contigs 1

Um die Orientierung des 50 kb Contigs auf dem Chromosom von *C. tentans* festzulegen, wurden Intergenregionen aus den Klonen hb20.1 und hb18 amplifiziert, die die äußeren Enden des Contigs 1 darstellen. Diese wurden unterschiedlich markiert (mit Digoxigenin bzw. mit Biotin) und in Kombination zur Hybridisierung eingesetzt (siehe Tab. 3.1, Sondenkombination 14). Die Ergebnisse der Hybridisierung sind in Abb. 3.3 dargestellt.

Im confocalen Mikroskop lassen sich die zwei verschiedenen Signale unterscheiden. Das rote Signal kennzeichnet das Contig-Ende, das durch den Klon hb20.1 charakterisiert ist, das grüne Signal markiert das Ende, an dem der Klon hb18 den Contig begrenzt. Die Anordnung der beiden Signale auf dem Chromosom zeigt, daß der Klon hb20.1 dem Nukleolus und damit dem Centromer zugewandt ist, während der Klon hb18 telomerwärts ausgerichtet ist. In Abb. 3.4 sind die Ergebnisse schematisch zusammengefaßt.

### 3.2.3 Sequenzierungsergebnisse

Zu Beginn der Arbeiten lagen zwei  $\lambda$ -Klone aus *C. tentans* vor, die Hämoglobin-Gene enthalten. Aus dem Klon hb5 war ein Abschnitt von 7 kb mit 5 Hämoglobin-Genen bereits publiziert (Rozynek et al., 1991), aus Klon hb10 lagen weitere 7 kb unpublizierte Sequenzdaten vor (Rozynek, Hankeln und Schmidt, unveröffentlicht). Ausgehend von diesen zwei bereits analysierten Sequenzabschnitten wurden die Nukleotid-Sequenzen der beiden erstellten Contigs vollständig ermittelt. Daraus ergibt sich für Contig 1 eine Länge von 50607 bp und für Contig 2 eine Länge von 24129 bp. Insgesamt wurden somit 74736 bp genomischer DNA sequenziert (siehe Abb. 3.1 und Gesamtsequenz der Contigs auf dem beigelegten CD-Datentäger; Dateien: 24 kb Contig 1-24129.pdf, 50 kb Contig 1-21400.pdf, 50 kb Contig 21401-50607.pdf).

Die kompletten Nukleotid-Sequenzen beider Contigs sowie die Lage aller identifizierten Gene sind im Anhang abgebildet. Überlappende Bereiche der Klone innerhalb der Contigs wurden nur insofern sequenziert, als es zur Bestimmung ihrer genauen Lage erforderlich war. Sequenzunterschiede in verschiedenen Klonen, die aufgrund ihrer Herkunft aus unterschiedlichen Allelen auftreten, sind ebenfalls im Anhang aufgeführt.

### Sequenzierungsstrategie und Sequenzqualität

Zur Sequenzierung der einzelnen DNA-Abschnitte wurde eine aus Subklonierung, PCR, „Primer-walking“ und „Shotgun“-Sequenzierung kombinierte Strategie verwendet. Eine reine „Shotgun“-Sequenzierung, wie sie in der Regel für größere Sequenzierungsprojekte angewendet wird, konnte hier nicht durchgeführt werden. Bei

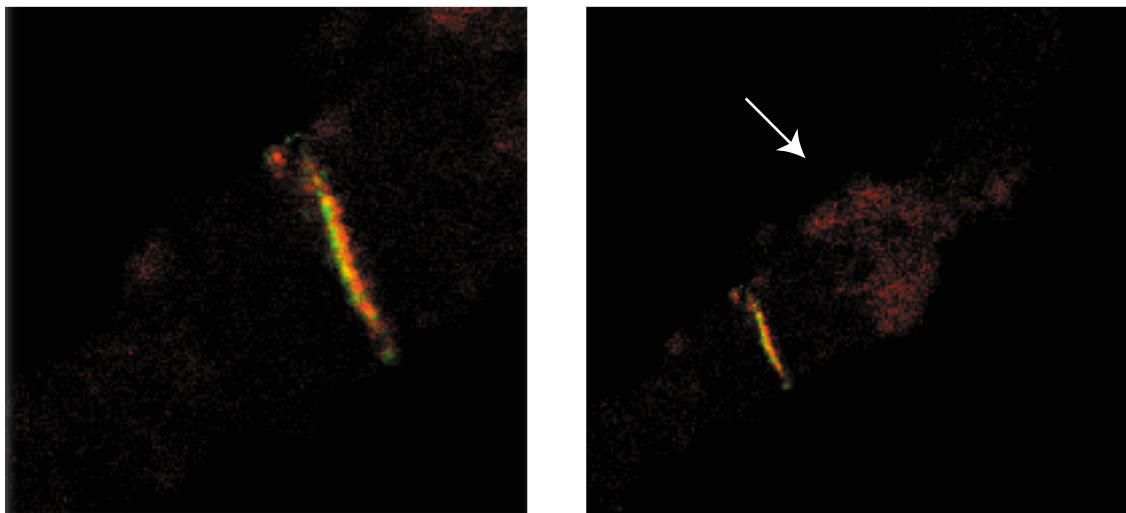
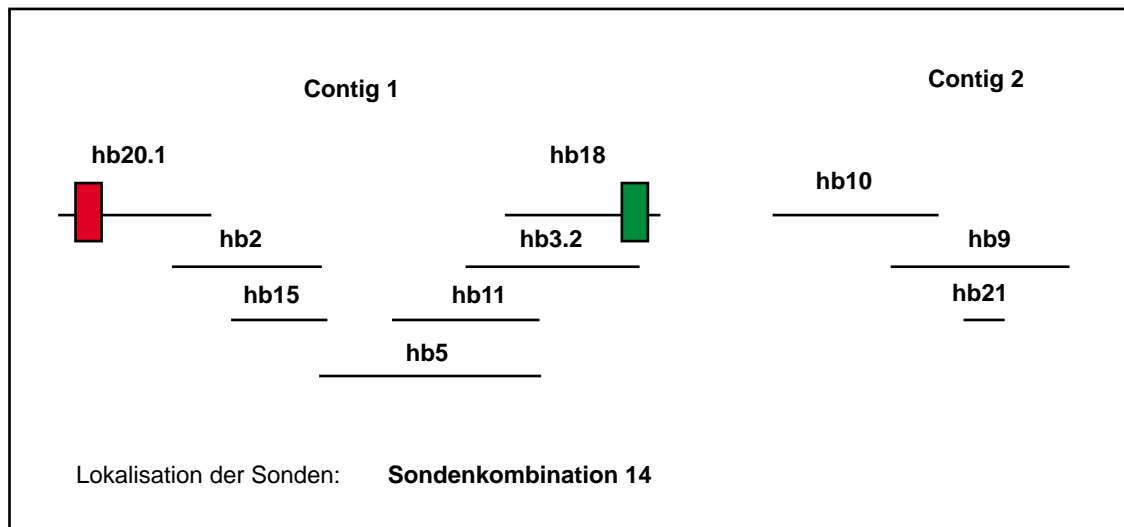


Abbildung 3.3: Orientierung des Klon-Contigs 1 auf dem Chromosom

Das Schema zeigt die Lage der Sondenkombination 14 innerhalb des Contigs 1. Im unteren Bereich der Abbildung ist das Ergebnis der Hybridisierungsreaktion dargestellt. Der Nukleolus (Pfeil) dient als Marker für die Lage des Centromers auf dem Chromosom. Anhand der lasermikroskopischen Aufnahme läßt sich erkennen, daß der Klon hb20.1, charakterisiert durch das rote Signal, dem Centromer zugewandt ist. Der Klon hb18, repräsentiert durch die grüne Fluoreszenz, liegt telomerwärts. Das linke Bild zeigt einen vergrößerten Ausschnitt des rechten Bildes.

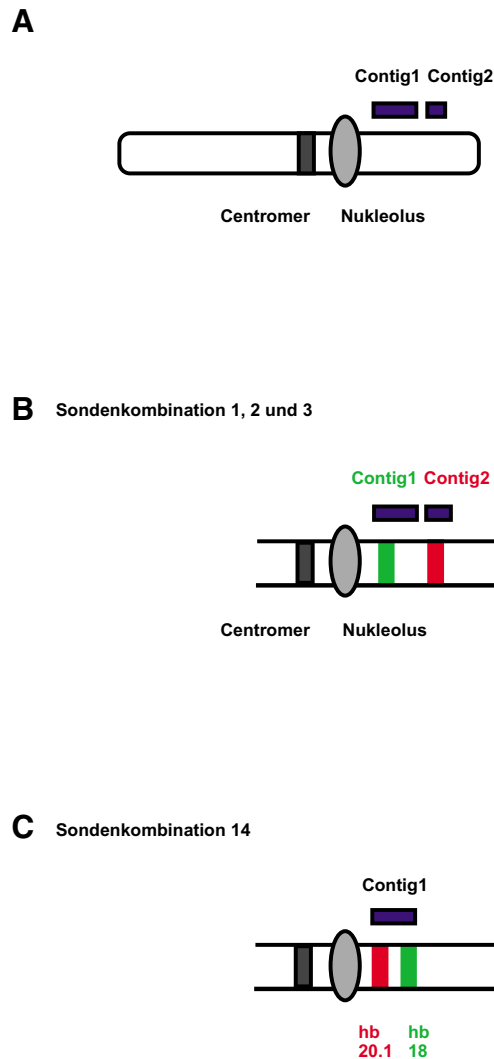


Abbildung 3.4: Die Chromosomale Anordnung und Orientierung der Contigs 1 (50 kb) und 2 (24 kb)

Die Abbildung faßt die Ergebnisse der Hybridisierungsexperimente schematisch zusammen. Der Contig 1 liegt dem Centromer zugewandt, während der kleinere Contig 2 telomerwärts orientiert ist (A). Die Anordnung der beiden Contigs erfolgte mittels der Sondenkombinationen 1, 2 und 3 (vergl. Tab. 3.1), wobei der Contig 1 jeweils durch die grüne, der Contig 2 durch die rote Fluoreszenz gekennzeichnet ist (B). Die Orientierung des Contig 1 auf dem Chromosom ist durch *in situ*-Hybridisierung mit den Klonen hb20.1 und hb18 festgelegt worden. Hierbei weist das Contig-Ende, das durch den Klon hb20.1 charakterisiert ist, die rote und das hb18-Contig-Ende die grüne Fluoreszenzmarkierung auf. Die Anordnung der Signale auf dem Chromosom zeigt, daß der Klon hb20.1 zum Centromer, der Klon hb18 hingegen zum Telomer hin orientiert ist (C).

dieser Methode wird das zu sequenzierende DNA-Fragment durch Scherung unspezifisch in kleinere Fragmente zerlegt, die dann subkloniert und sequenziert werden. Die zufällig erhaltenen Sequenz-Abschnitte werden per Computeralignment zusammengefügt. Dabei verbleibende Sequenzlücken werden in der Regel anschließend durch „Primer-walking“ geschlossen. Diese Strategie ermöglicht einen hohen Grad an Automatisierung, so daß große Datenmengen in kürzester Zeit relativ kostengünstig erhoben werden können.

Zwei Gründe verhindern eine konsequente Anwendung der „shotgun“-Methode bei der Sequenzierung von Hämoglobin-Genen in *Chironomus*: zum einen macht der hohe Vektoranteil (in der Regel 3/4 der Gesamtlänge) eine „Shotgun“-Sequenzierung von  $\lambda$ -Klonen ineffizient, zum anderen erschwert der hohe Anteil an Hämoglobin-Genen das korrekte Zusammenfügen der Einzelsequenzen per Computer. Aufgrund ihrer hohen Übereinstimmung wirken die Hämoglobin-Gene wie eine Ansammlung repetitiver Sequenzen. Versuche mit dem Programm „Sequencher“ (Gene Codes Corp., 1995), das routinemäßig für das Zusammensetzen der Teilsequenzen verwendet wurde, zeigten, daß ein computerunterstütztes Zusammenfügen aller Einzelsequenzen eines Hämoglobin-Gen-haltigen  $\lambda$ -Klonintegrats auch unter stringenten Bedingungen nie die korrekte Anordnung der Sequenzen lieferte. Eine gerichtete Sequenzierungsstrategie war aus diesen Gründen für alle  $\lambda$ -Klone die einzige Möglichkeit, die korrekte Gesamtsequenz zu ermitteln.

Durch gezielte Subklonierung und „Primer-walking“ ist es möglich, Sequenzen zu erstellen, deren Lage exakt definiert ist. Diese Strategien bildeten daher die Grundlage der Sequenzierungsarbeit. Bei Klonen mit einer hohen Gendichte war es zunächst erforderlich, Gene mit hoher Sequenzähnlichkeit voneinander zu isolieren, um eine mehrfache Bindung von Sequenzier-Primern beim „Primer-walking“ auszuschließen. Aus diesem Grund wurden die Integrate der Klone zunächst kartiert und die einzelnen Restriktionsfragmente subkloniert. Diese Subklone wurden anschließend durch „Primer-walking“ sequenziert und mit Hilfe überlappender Sequenzbereiche oder durch PCR zueinander orientiert. Die PCR-Reaktionen wurden an der  $\lambda$ -DNA als Matrize durchgeführt und die Amplifikate direkt sequenziert. Diese Sequenzierungsstrategie wurde bei allen  $\lambda$ -Klonen mit Ausnahme des Klons hb20.1 angewendet. Dieser Klon zeichnet sich durch seinen geringen Anteil an Hämoglobin-Genen aus. Er enthält nur zweieinhalb Hämoglobin-Gene mit einer relativ geringen Sequenz-Übereinstimmung. Diese Hämoglobin-Gene unterscheiden sich zusätzlich dadurch, daß eines der Gene ein Intron besitzt. Es war daher möglich, diesen Klon durch „Primer-walking“ direkt an der  $\lambda$ -DNA als Matrize zu sequenzieren. Um diese Strategie zu beschleunigen, wurde die Anzahl der Startpunkte für die „Primer-walking“-Reaktionen erhöht. Dies geschah durch einen Verdau des Inserts mit dem Restriktionsenzym Sau 3A und anschließender „shot-gun“ Klonierung der resultierenden Fragmente. Sau 3A besitzt eine vier Basenpaare lange Erkennungssequenz und liefert daher eine große Anzahl kurzer Restriktionsfragmente. Diese ergaben nach Sub-

klonierung und Sequenzierung mit Standard-Primern eine große Anzahl zusätzlicher Startpunkte, die parallel für das „Primer-walking“ genutzt werden konnten.

Über 93 % der Gesamtsequenz von 74000 bp wurden doppelsträngig sequenziert (hierzu zählen alle Gene). Einzelsträngige Sequenzbereiche wurden stets mehrfach ermittelt. Die Integrate der Klone hb5, hb10, hb9 und hb20.1 liegen vollständig sequenziert vor. Bei den Klonen hb2, hb15, hb18 und hb21 wurde die Basensequenz nur in Teilbereichen ermittelt, da überlappende Bereiche häufig nur auf einem der beiden Klone vollständig sequenziert wurden. Tab. 3.2 gibt Auskunft über die für jeden Klon benötigte Anzahl an Sequenzierungsreaktionen (Spalte a). Aus der Länge der Einzelsequenzen errechnet sich die Gesamtsumme der ermittelten Sequenz-Rohdaten in Spalte b. Die durchschnittliche Leseweite variiert in Abhängigkeit von der verwendeten Sequenzierungschemie und Sequenzierungsstrategie. Die Klone hb5 und hb10 wurden mit der ersten nichtradioaktiven Sequenzierungschemie auf der Basis von Dye Terminatoren der Firma Applied Biosystems (Weiterstadt) sequenziert. Leseweiten von 350-400 bp entsprechen hier den Erwartungen. Mit fortschreitender Weiterentwicklung der Sequenzierungschemikalien und der Programme für die Sequenzauswertung wurden bei den zuletzt sequenzierten Klonen hb15 und hb18 durchschnittliche Leseweiten von mehr als 600 bp erzielt. Die geringere durchschnittliche Leseweite von 500 bp bei der Sequenzierung des Klons hb20.1 ist auf die erschwerten Bedingungen der Sequenzierung direkt an der  $\lambda$ -DNA als Matrize zurückzuführen. Insgesamt wurden 557 Sequenzierungsreaktionen mit 254 internen Primern benötigt, um 263368 bp an Sequenz-Rohdaten zu ermitteln. Dies entspricht bei einer endgültigen Gesamtsequenz von 74736 bp einer durchschnittlichen Redundanz von 3,4.

#### Die Nomenklatur der Hämoglobin-Gene

Die Nomenklatur der Chironomiden-Hämoglobine geht zurück auf die von Braunitzer eingeführten Bezeichnungen für Hämoglobin-Protein-Varianten (Übersicht: Goodman et al. (1983)) und wurde von anderen Autoren erweitert (Gruhl et al., 1997; Hankeln und Schmidt, 1991; Hankeln et al., 1998). Der Gen-Name beginnt stets mit dem abgekürzten Artnamen (z.B. Cte: *C. tentans*, Ctt: *C. thummi*, Ctp: *C. piger* und Cme: *C. melanotus*. Entsprechend werden die korrespondierenden Proteine CTE, CTT, CTP, CME genannt). Hämoglobin-Proteine werden mit römischen, Hämoglobin-Gene mit arabischen Buchstaben bezeichnet (z. B. CTT IX, Ctt 9). Neuartige Hämoglobin-Gene, zu denen bislang kein Protein beschrieben wurde, erhalten Buchstaben (z.B. Ctp W, Hankeln und Schmidt (1991)) oder weitere Nummern (z.B. Ctt 11, Gruhl et al. (1997)). Mitglieder von Gen-Subfamilien werden durch fortlaufende Nummern im Anschluß an den Gen-Namen voneinander unterschieden (z.B. Cte 7A1, Cte 7A2). Buchstaben, die an den Gen-Namen angehängt werden, kennzeichnen Allele dieses Gens (z.B. Cte 7A1, Cte 7A1a).

$\lambda$ -Klon	a. Anzahl der Einzelsequenzen	b. Summe der ermittelten Basen	c. durchschnittliche Leseweite = $b/a$ in bp	d. Integratlänge	e. Redundanz = $b/d$	f. benötigte Primer
hb20.1	93	44215	508	12775	3,5	57
hb2	33	18576	563	5568	3,3	20
hb15	45	27665	614	7963	3,5	26
hb5	141	46738	331	11647	4,0	44
hb18	57	36131	634	11116	3,2	36
hb10	77	30629	398	6811	4,5	19
hb9	77	41519	539	13879	2,9	44
hb21	18	9285	516	3289	2,8	8
Summe	557	263368	508	74736	3,4	254

**Tabelle 3.2:** Sequenzierungsstatistik

In Spalte a sind die für jeden Klon benötigten Sequenzierungsreaktionen aufgeführt. Die Summe aller ermittelten Basenpaare ist in Spalte b dargestellt. Aus dem Quotient der Werte aus Spalte b und a ergibt sich die durchschnittliche Leseweite der Sequenzierungsreaktionen in Spalte c. Spalte d gibt die sequenzierte Länge der Klon-Integrate an. Aus den Werten b und d läßt sich somit die Redundanz der Endsequenzen errechnen (Spalte e). Die Anzahl der jeweils benötigten Primer ist in Spalte f gezeigt.

### Identifizierung und Charakterisierung von Hämoglobin-Genen

Die große Ähnlichkeit der Mitglieder dieser Multigenfamilie untereinander ermöglicht das Auffinden der Hämoglobin-Gene innerhalb einer Nukleotid-Sequenz durch Datenbanksuchen. Diese wurden mit allen ermittelten Sequenzen durchgeführt (siehe Kapitel Material und Methoden). Die Identifikation bereits bekannter Hämoglobin-Varianten erfolgte durch (i) den Vergleich der abgeleiteten AS-Sequenzen mit vorliegenden Daten von *C. thummi* (zusammengestellt in Goodman et al. (1983)), (ii) phylogenetische Untersuchungen auf Nukleotid- und Aminosäure-Ebene, (iii) das Kriterium der Lage innerhalb der Gen-Cluster, (iv) die Analyse der Intergenregionen im paralogen sowie im orthologen Vergleich (vergl. Kapitel Diskussion).

In den vorliegenden 74,7 kb genomischer Sequenz des Hämoglobingen-Clusters D von *C. tentans* sind 37 vollständige Hämoglobin-Gene sowie ein Pseudogen enthalten. Die Sequenzen der Hämoglobin-Gene Cte Y, Cte W, Cte V, Cte Z, Cte ORFA, Cte 2 $\beta$ , Cte 9, Cte 7A-01 sowie des Pseudogens Cte ORFB waren bereits bekannt (Rozynek et al. (1991); Rozynek, Hankeln und Schmidt, unveröffentlicht), so daß insgesamt 28 bei *C. tentans* bislang unbekannte Hämoglobin-Gene in der vorliegenden Arbeit sequenziert wurden. Sechs dieser Hämoglobin-Gene sind bisher

bei keiner weiteren Chironomiden Art beschrieben worden, es handelt sich somit um neue Hämoglobin-Genvarianten. Sie wurden als Cte N, Cte P, Cte Q, Cte R, Cte S sowie Cte T bezeichnet.

Ein weiteres Hämoglobin-Gen ist nur unvollständig am Ende des Contig 1 vorhanden. Aufgrund des fehlenden 5'-Bereichs konnte es nicht sicher den bereits bekannten Hämoglobin-Genen zugeordnet werden. Sequenz-Übereinstimmungen weisen jedoch darauf hin, daß es sich hier möglicherweise um das Hämoglobin-Gen Cte 6 handelt, das bereits für *C. thummi* beschrieben wurde (Kao und Bergtrom, 1995). Der Aminosäurevergleich zeigt eine Sequenz-Identität von 84 % zu dem Hämoglobin CTT VI, während die Übereinstimmung zu allen weiteren Hämoglobin-Varianten von *C. thummi* zwischen 26 % und 41 % (durchschnittlich 29 %) liegt. Auf Nukleotidebene beträgt die Sequenz-Übereinstimmung 81 %. Die Identitäten zu anderen Hämoglobin-Genen sowohl von *C. thummi* als auch von *C. tentans* bewegen sich hingegen zwischen 42 % und 55 % über die vorhandenen 306 bp. Der nicht kodierende 3'-Bereich weist mit 73 % Identität über einen 165 bp langen Bereich ebenfalls eine überdurchschnittlich gute Übereinstimmung zu Ctt 6 auf. Solch gute Sequenz-Übereinstimmungen innerhalb nicht kodierender Bereiche sind als deutlicher Hinweis auf echte Orthologie von Sequenzen zu werten. Aufgrund der fehlenden Sequenzinformation sowie der bislang nicht näher definierten Lage des Ctt 6 Hämoglobin-Gens innerhalb der Gengruppe D, wird das vermutlich orthologe *C. tentans* Hämoglobin-Gen vorläufig als Cte U bezeichnet .

Alle potentiellen Genkandidaten wurden in Hinblick auf ihre mögliche Funktionalität hin überprüft. Hierbei waren „offene Leserahmen“ (ORFs) und das Vorhandensein von Promotorsequenzen und Polyadenylierungs-Signalen Kriterien. Eine TATA-Box etwa 60 bp aufwärts des Startcodons sowie ein Konsensus-Polyadenylierungssignal ca. 60 bp abwärts des Stoppkodons kennzeichnen funktionelle Hämoglobin-Gene. Alle identifizierten Hämoglobin-Gene besitzen die für eine Transkription erforderlichen Signalsequenzen. Die Ergebnisse dieser Auswertungen sind in Tab. 3.3 zusammengefaßt.

Innerhalb der Cte 7B-Gengruppe ist ein weiterer ORF in der für Hämoglobin-Gene typischen Größe von 489 bp vorhanden. Allerdings besitzt weder seine Nukleotid- noch die abgeleitete AS-Sequenz eine Ähnlichkeit zu bekannten Genen bzw. Proteinen. Auch bisher beschriebene funktionelle Domänen (Henikoff und Henikoff, 1994) sind in der abgeleiteten AS-Sequenz nicht vorhanden. Das Fehlen der erforderlichen Promotorsequenzen deutet darauf hin, daß es sich hier nicht um ein transkribiertes Gen handelt. Dennoch ist das Vorhandensein eines derart langen ORFs ungewöhnlich, zumal reine Intergenregionen bei Chironomiden aufgrund ihres AT-Reichtums in der Regel zahlreiche Stoppkodons enthalten. Der AT-Gehalt dieses ORFs liegt mit 66,0 % nur geringfügig unter dem AT-Gehalt der flankierenden Sequenzen (68 % und 71 %).

Ein weiterer offener Leserahmen, dessen Sequenz keinem bisher bekannten Gen oder Protein zugeordnet werden konnte, ist am Anfang des Contigs 1 lokalisiert. Die

komplette Länge dieses ORF 1 sowie das Vorhandensein promotortypischer Signalsequenzen ist unbekannt, da nur der 3'-Bereich auf dem Klon liegt.



### 3.2 Die chromosomale Organisation der $\lambda$ -Klone

Contig 1 50 kb						
ORF	Gen	ORF Position	TATA-Box Position	CAP Position	PolyA AATAAA	Länge (bp)
ORF 1 $\Rightarrow$	?	1-385	?	?	nein	?
ORF 2 $\Rightarrow$	Hb-Gen Cte P	8510 -9071	GTTTAAA 8414-8420 GTATAT 8454	AGTT 8499	(T)AATAAA 9105-9109	562
ORF 3 $\Leftarrow$	Hb-Gen Cte Q	10548- 11057	ATATAAAA 11143-11149 11149-11155	AGTT 11084	10491-10496	510
ORF 4 $\Rightarrow$	Hb-Gen Cte 8	12331- 12831	GTATAAAA 12251-12257	AGTT 12292	12874-12879	501
ORF 5 $\Leftarrow$	Hb-Gen Cte R	13018- 13521	GTATAAAA 13582-13588	AGTT 13551	12981-12986	504
ORF 6 $\Leftarrow$	Hb-Gen Cte S	14150- 14653	CTATAAAA 14714-14720	AGTT 14683	14081-14086	504
ORF 7 $\Rightarrow$	Cte GLUT 1	15850- 17450	GTATAATA 15787-15794	AGTT 15836	17530-17535	
ORF 8 $\Leftarrow$	Hb-Gen Cte 11	17575- 18081	GTATAAAA 18146-18152	AGTT 18115	17494-17499	507
ORF 9 $\Leftarrow$	Hb-Gen Cte 12.1	18844- 19347	GTATAA 19413-19419	AAGTT 19382	?19696-18701	504
ORF 10 $\Leftarrow$	Hb-Gen Cte 12.2	20449- 20946	GTATAAAA 21015-21021	AGTT 10984	?20293-20297	498
ORF 11 $\Rightarrow$	Hb-Gen Cte ORFA	22264- 22755	GTATAAAA 22189-22195	AGTT 22222	22804-22808	492
ORF 12 $\Rightarrow$	Pseudogen ORFB	23272- 23646	ATATAAAA 23714-23720	nein	nein	375
ORF 13 $\Leftarrow$	Hb-Gen Cte 2 $\beta$	24181- 24663	GTATAAAA 24736-24742	AGTT 24705	24086-24091	483
ORF 14 $\Leftarrow$	Hb-Gen Cte 9	26212- 26697	GTATAAAA 26767-26773	AGTT 26736	26146-26151	486
ORF 15 $\Leftarrow$	Hb-Gen Cte 7A1	27366- 27851	GTATAAAA 27921-27927	AGTT 27890	?27262-27267	486
ORF 16 $\Leftarrow$	Hb-Gen Cte 7A2	29970- 30455	GTATAAAA 30525-30531	AGTT 30494	29880-29885	486
ORF 17 $\Rightarrow$	Hb-Gen Cte7A3	32126- 32611	GTATAAAA 32050-32056	AGTT 32084	32697-32702	486
ORF 18 $\Leftarrow$	Hb-Gen Cte 7A4	33513- 33998	GTATAAAA 34068-34074	AGTT 34041	33443-33448	486
ORF 19 $\Leftarrow$	Hb-Gen Cte 7A5	35834- 36319	ATATAAAA 36389-36395	AGTT 36358	35760-65765	486
ORF 20 $\Rightarrow$	Hb-Gen Cte 7A6	37144- 37629	GTATAAAA 37068-37074	AGTT 37102	37698-37703	486
ORF 21 $\Leftarrow$	Hb-Gen Cte 7A7	38761- 39246	GTATAAAA 39316-39322	AGTT 39285	38661-38666	486

### 3 Ergebnisse

Contig 1 50 kb						
ORF	Gen	ORF Position	TATA-Box Position	CAP Position	PolyA AATAAA	Länge (bp)
ORF 22 ⇒	Hb-Gen Cte 7A8	39961- 40446	GTATAAA 39885-39891	AGTT 39919	40540-40546	486
ORF 23 ⇐	Hb-Gen Cte 1	43073- 43564	GTATAAA 43636-43632	AGTA 43606	42992-42997	492
ORF 24 ⇐	Hb-Gen Cte T	46415- 46903	GTATAAA 46975-46981	AGTT 46945	46346-46351	489
ORF 25 ⇐	Hb-Gen Cte U	50302- 50607	?	?	50174-50179	306

Contig 2 24 kb						
ORF	Gen	ORF Position	TATA-Box Position	CAP Position	PolyA AATAAA	Länge (bp)
ORF 26 ⇒	Hb-Gen Cte Y	1037- 1531	GTATAAA 969-975		1573-1578	495
ORF 27 ⇒	Hb-Gen Cte W	2282- 2761	GTATATA 2220-2226	AGTT 2256	2799-2804	480
ORF 28 ⇐	Hb-Gen Cte V	3628- 4119	GTATAAA 4188-4194	AGTT 4158	3568-3573 3603-3608	492
ORF 29 ⇒	Hb-Gen Cte Z	4542- 6935	GTATAAA 4856-4862		5456-5461	492
ORF 30 ⇐	Hb-Gen Cte7B12	6827- 7318	CTATAAA 7378-7393	AGTT 7357	6770-6775	492
ORF 31 ⇒	Hb-Gen Cte 7B13	8297- 8788	CTATAAA 8222-8228	AGTT 8255	8849-8854	492
ORF 32 ⇐	Hb-Gen Cte 7B7	9262- 9750	GTATAAA 9819-9825	AGTT 9789	9208-9213	489
ORF 33 ⇐	Hb-Gen Cte 7B6	10553- 11038	GTATAAA 11107-11113	AGTT 11077	10495-10500	486
ORF 34 ⇐	Hb-Gen Cte 7B5	12930- 13415	GTATAAA 13484-13490	AGTT 13454	12872-12877	486
ORF 35 ⇐		14609- 15097	nein	nein	nein	489
ORF 36 ⇐	Hb-Gen Cte7B10	16667- 17152	GTATAAA 17221-17227	AGTT 17191	16609-16614	486
ORF 37 ⇒	Hb-Gen Cte 7B9	17921- 18406	ATATAAA 17847-17852	AGTT 1 17879	8459-18464	486
ORF 38 ⇐	Hb-Gen Cte 7B8	19143- 19628	GTATAAA 19697-19703	AGTT 19667	19081-19086	486
ORF 39 ⇒	Hb-Gen	20454-	GTATAAA	AGTT	21001-21006	486

Contig 1 50 kb						
ORF	Gen	ORF Position	TATA-Box Position	CAP Position	PolyA AATAAA	Länge (bp)
ORF 39	Cte 7B4	20939	20379-20385	20412		
ORF 40 ⇒	Hb-Gen Cte 7B11	22828- 23319	GTATAAA 22750-22756	AGTT 11795	23369-23374	492
ORF 41 ⇐	Hb-Gen Cte N	23560- 24129	?	?	23689-23694	211

Tabelle 3.3: Liste der „offenen Leserahmen“ der Contigs 1 und 2

Insgesamt wurden 41 „offene Leserahmen“, die für Peptide größer 100 AS kodieren, gefunden. Datenbank-Analysen identifizierten insgesamt 38 dieser ORFs als Hämoglobin-Gene, wobei das Hämoglobin-Gen Cte U nur unvollständig vorhanden ist. Bei dem ORF 7 des Contig 1 handelt es sich um das Gen für einen Glukosetransporter. Das Vorhandensein der für die Transkription erforderlichen Signalsequenzen bei allen Hämoglobin-Genen sowie dem Glukosetransporter-Gen, weist darauf hin, daß es sich wahrscheinlich um aktive Gene handelt. Nur bei dem ORF B handelt es sich vermutlich um ein Pseudogen. Innerhalb der Cte 7B-Gengruppe ist ein weiterer ORF in der für Hämoglobin-Gene typischen Größe vorhanden, bei dem es sich voraussichtlich nicht um ein transkribiertes Gen handelt. Seine Sequenz weist keine Ähnlichkeiten zu bekannten Genen oder funktionellen Protein-Domänen auf. Ebenso fehlen die für die Transkription erforderlichen Signalsequenzen. Auch bei dem ORF 1 handelt es sich nicht um ein Hämoglobin-Gen. Seine volle Länge sowie sein Transkriptionspotential sind unbekannt.

Die angegebenen Positionen der Promotor-Boxen beziehen sich auf die Gesamtsequenz der beiden Contigs (siehe CD-Datenträger im Anhang). Der Pfeil in Spalte 1 gibt die Transkriptionsrichtungen der putativen Gene an.

### 3.3 Die genomische Struktur der Hb-Gengruppe D

Die Anordnung der Gene innerhalb der Gengruppe D ist in Abb. 3.5 dargestellt. In der Regel liegen die Hämoglobin-Gene dicht gepackt innerhalb des Clusters, wobei durchschnittlich alle 1,9 kb ein Hämoglobin-Gen zu finden ist. Benachbarte Gene weisen häufig eine entgegengesetzte Orientierung auf. Dies ermöglicht es, klonierte Abschnitte bereits im Elektronenmikroskop auf Gehalt und Anordnung ihrer Hämoglobin-Gene hin zu untersuchen (Schmidt et al., 1988). Dabei wird die DNA eines Klonen zunächst denaturiert. Bei der Renaturierung bilden die invertierten Gene aufgrund des palindromischen Charakters der Sequenzen „Haarnadelschleifen“ aus, deren Ausdehnung und Abstand gemessen werden kann. Diese Arbeiten wurden von Prof. Keyl (Ruhr Universität, Bochum) mit den hier beschriebenen  $\lambda$ -Klonen durchgeführt und deren Gehalt an Hämoglobin-Genen vorhergesagt. Mit Hilfe die-

ser Methode war es ebenfalls möglich, die komplexe Verteilung und Orientierung der Hämoglobin-Gene des Klons hb9 exakt zu bestimmen.

Auffällig bei der Organisation der Hämoglobin-Gengruppe ist, daß manche Genvarianten als „Einzelkopien“ vorhanden sind, während andere Gene für nur leicht unterschiedliche bis identische Hämoglobin-Protein-Varianten kodieren. So gibt es im Contig 1 eine Cte 7A-Subfamilie mit acht Mitgliedern und eine Cte 12 Subfamilie, die lediglich zwei Gene umfaßt. Der Contig 2 beinhaltet die zehn Mitglieder große Cte 7B-Subfamilie. Diese Hämoglobingen-Subfamilien sind in Gruppen innerhalb des Clusters organisiert. Von den insgesamt 38 identifizierten Hämoglobin-Genen sind somit 20 Hämoglobin-Gene Angehörige einer Subfamilie, während 18 Gene „single copy“-Varianten darstellen. Die Entstehung und funktionelle Bedeutung dieser Unterteilung von Hämoglobin-Genen in „single copy“-Varianten und Subfamilien ist bislang unklar.

Funktionslose Pseudogene sind in der Hämoglobin-Genfamilie bei Chironomiden bislang selten aufgetreten. Das Gen Cte ORF B, ein Hämoglobin-Gen ohne Startkodon und mit vorzeitigem Stoppkodon (Rozynek et al., 1991), liegt im Contig 1. Trotz der Multiplizität der Hämoglobin-Gene innerhalb des Clusters D konnten keine weiteren Pseudogene gefunden werden.

Die Mehrzahl der Hämoglobin-Gene in der Gruppe D von *C. tentans* enthalten keine Introns. Nur die Hämoglobin-Gene Cte P und Cte N besitzen je ein kurzes Intron (vergl. Kapitel 3.5). Die dichte Abfolge der Hämoglobin-Gene innerhalb der Gengruppe D wird nur an einer Stelle durch ein nicht-Hämoglobin-Gen unterbrochen, das für ein Glukosetransporter-Protein kodiert (vergl. Kapitel 3.6). Über die Bedeutung von repetitiven Elementen innerhalb der Chironomiden Hämoglobin-Gene ist wenig bekannt (Gruhl et al., 1997). Innerhalb der Cte 7A Gen-Subfamilie finden sich zwei Kopien des Retrotransposons CP1, das von Hong et al. (1995) ursprünglich in der Centromerregion der Chromosomen entdeckt wurde (vergl. Kapitel 3.7).

### 3.4 Allele Hämoglobingen-Varianten

Durch die Sequenzierung überlappender Klone konnten die Sequenzen von fünf allen Hämoglobin-Genen bestimmt werden. Es handelt sich hierbei um vier Hämoglobin-Gene der Cte 7B-Subfamilie (7B6A, 7B5A, 7B10A und 7B9A) sowie um das Hämoglobin-Gen Cte Q (vergl. Abb. 3.5). Erwartungsgemäß zeichnen sich alle allelen Hämoglobin-Varianten durch einen hohen Grad an Sequenz-Übereinstimmung aus. In Tabelle 3.4 sind die Unterschiede der Nukleotid-Sequenzen sowie deren Auswirkungen auf die AS-Sequenzen gezeigt. Es wäre dennoch voreilig, allein aufgrund der extrem hohen Sequenz-Übereinstimmungen kodierender DNA-Bereiche eine allele Beziehung dieser Gene zu postulieren. Auch manche paralogen Hämoglobin-Gene weisen im Vergleich nur wenige Nukleotid-Austausche auf (Trewitt et al., 1988).

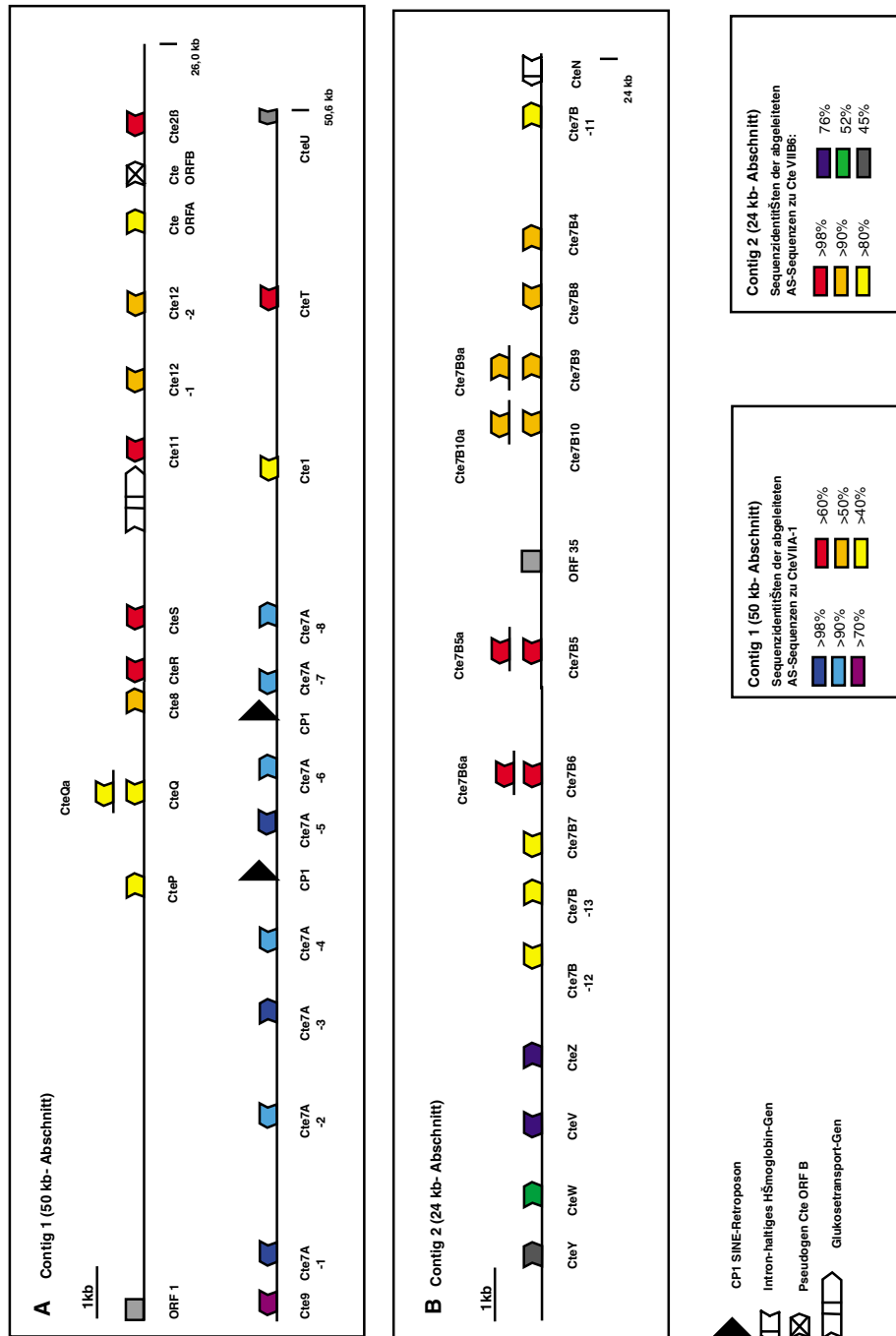


Abbildung 3.5. Legende siehe folgende Seite

**Abbildung 3.5: Genomische Struktur der Hämoglobin-Gengruppe D von *C. tentans***

Innerhalb des Hämoglobingen-Clusters D wurden zwei zusammenhängende Sequenzbereiche ermittelt: Zum einen der 50 kb umfassende Contig 1 (A) und zum anderen der 24 kb lange Contig 2 (B). Farbige Pfeile symbolisieren die Gene. Die Pfeilrichtung entspricht jeweils der Transkriptionsrichtung. Einige der Hämoglobin-Gene liegen als „Einzelkopien“ vor und kodieren jedes für eine andere Proteinvariante. Die Hämoglobin-Varianten VII A und VII B werden hingegen von einer Vielzahl nur leicht unterschiedlicher Gene kodiert. Sie bilden Subfamilien innerhalb der Hämoglobin-Gengruppe. Ihre Mitglieder sind stets in räumlicher Nähe zueinander im Gen-Cluster angeordnet. Um die Sequenz-Übereinstimmungen der Gene zu veranschaulichen, wurden die Identitäten der abgeleiteten Aminosäure-Sequenzen zu je einem Mitglied der Gen-Subfamilien farblich kodiert. Zusätzlich sequenzierte Hämoglobingen-Allele sind oberhalb der „standard-Gene“ gezeigt. Der Buchstabe a am Ende der Genbezeichnung kennzeichnet sie als Allele. Introns sind durch einen Balken im Gen dargestellt. Das Gen für ein Glukose-Transportprotein (GLUT) unterbricht die ansonsten dichte Abfolge von Hämoglobin-Genen. Zwei weitere offene Leserahmen, deren Größe 300 Nukleotide übersteigt, sind als schraffierte Kästchen abgebildet. Innerhalb der VIIA Gen-Subfamilie befinden sich zwei Retroposons aus der Familie der SINE-Elemente (Hong et al., 1995).

Durch den Vergleich der nicht-kodierenden Intergenregionen kann jedoch gezeigt werden, daß es sich bei den hier beschriebenen Genen tatsächlich um Allele und nicht um paraloge Gene mit ungewöhnlich hoher Sequenz-Übereinstimmung handelt. Außerhalb der Hämoglobin-Gensequenzen nimmt die Sequenz-Übereinstimmung zwischen paralogen Abschnitten in der Regel drastisch ab. Der Vergleich der nicht-kodierenden DNA-Bereiche ist in Tabelle 3.5 dargestellt. Es zeigt sich, daß der hohe Identitätsgrad der Nukleotid-Sequenzen auch in den Intergenregionen nahezu unvermindert ist (99,2-99,6 % Identität der Gene gegenüber 98,1-99 % Identität der Intergenregionen).

Bei der überwiegenden Mehrzahl der Nukleotid-Austausche handelt es sich um „stille“ Substitutionen, die ohne Auswirkung auf die resultierende Aminosäure-Sequenz sind. Die allelen Protein-Varianten unterscheiden sich jeweils durch nur eine Aminosäure. Der Austausch der funktionell gleichwertigen Aminosäuren Leucin und Isoleucin im Signalpeptid der Proteine CTE VIIB5 und CTE VIIB9 bleibt sehr wahrscheinlich ohne Auswirkung auf die Sekretion der Hämoglobine und ist daher als neutral zu werten. Gleiches gilt für die Hämoglobine CTE Q und CTE VIIB6. Die Aminosäuren Isoleucin und Valin finden sich alternativ bei einer Vielzahl der Hämoglobine an den hier betroffenen Positionen. Bei der Proteinvariante CTE VIIB10 hingegen ist eine ansonsten konservierte Aminosäure in einer funktionell wichtigen Region ausgetauscht. Die Position C6 wird von den beiden Hämkontaktstellen C4 und C7/C8 flankiert (Steigemann und Weber, 1979). Ob der Austausch von Alanin gegen Threonin an dieser Stelle die Struktur und damit möglicherweise auch die Funktion dieser Hämoglobin-Variante beeinflusst, ist jedoch ungewiß. Eine analoge

Situation wurde jedoch bereits für die Hämoglobin-Variante CTT I beschrieben (Kao et al., 1995). Auch die allelen Hämoglobine CTT I und CTT IA unterscheiden sich durch das Vorhandensein von Alanin bzw. Threonin an einer konservierten Position. Dieser anscheinend geringe Unterschied führt hier zu einem veränderten Laufverhalten der Moleküle in nativen Polyacrylamid-Gelen (Goodman et al., 1983). Es wird daher spekuliert, daß Threonin in der Lage ist, die Struktur und damit möglicherweise auch die Ladungsverteilung des Gesamtmoleküls zu beeinflussen.

Hb-Gen	Klon	Identität Nt (%)	Nukleotid-Substitutionen (Position) *				Identität AS (%)	AS-Substitution (Position) *
			C	C	A	C		
Cte 7B6 Cte 7B6A	hb10 hb9	99,2	C T (180)	C T (264)	A G (298)	C A (408)	99,4	I (100) V E-F 11
Cte 7B5 Cte 7B5A	hb10 hb9	99,6	A C (25)		C T (471)		99,4	I L („leader“)
Cte 7B10 Cte 7B10A	hb21 hb9	99,6	A G (160)		C T (216)		99,4	T A (54/C6)
Cte 7B9 Cte 7B9A	hb21 hb9	99,4	C A (25)	T C (264)	T C (396)		99,4	L I („leader“)
Cte Q Cte QA	hb2 hb20.1	99,2	C T (274)	C A (318)	G A (472)	T A (483)	99,4	V I (158/H17)

**Tabelle 3.4:** Sequenzunterschiede aller Hämoglobin-Genvarianten

Durch die Sequenzierung überlappender Klone wurden die Sequenzen fünf aller Hämoglobin-Gene ermittelt. Diese zeichnen sich erwartungsgemäß durch einen hohen Grad an Sequenz-Übereinstimmung aus. Maximal vier Nukleotide sind ausgetauscht. Die überwiegende Mehrheit dieser Austausch sind „synonym“ (blau); nur jeweils eine der Substitutionen hat einen Effekt auf die Aminosäure-Sequenz (rot). Alle allelen Hämoglobin-Protein-Varianten unterscheiden sich somit an einer einzigen Aminosäure-Position. \* Die Nummerierung beginnt mit dem ersten Nukleotid des Startkodons bzw. mit der ersten Aminosäure des Leader-Peptids.

<b>A</b>	Klone	Überlappender Bereich (bp)	Substitutionen	Insertionen/Deletionen	Übereinstimmung (%)
	hb20.1/ hb2	3291	22	7	99,1
	hb10 / hb9	3809	44	11	98,6
	hb9 / hb21	3310	20	9	99,1
<b>B</b>	Klone	Hb-Gene	Intergenregion -600bp/ Hb-Gen	Substitutionen/Insertionen/Del.	Übereinstimmung (%)
	hb20.1/ hb2	Cte Q	2691	18/7	99,0
	hb10 / hb9	Cte 7B6, Cte 7B5	2609	38/11	98,1
	hb9 / hb21	Cte 7B10, Cte 7B9	2110	15/9	98,8

**Tabelle 3.5** Legende siehe folgende Seite

**Tabelle 3.5: Sequenz-Vergleich überlappender Klone**

Zunächst wurden die Übereinstimmungen der Nukleotid-Sequenzen der gesamten überlappenden Bereiche inklusive der Hämoglobin-Gene ermittelt (A). Die Sequenz-Übereinstimmungen der Intergenregionen sind separat in Tab. B gezeigt. In der ersten Spalte sind jeweils die sich überlappenden Klone aufgeführt, die zweite Spalte gibt Auskunft über die in diesem Bereich vorhandenen Hämoglobin-Gene. Pro Hämoglobin-Gen wurden 600 bp von der Gesamtsequenz abgezogen. Dies entspricht der eigentlichen Gensequenz zuzüglich 100 bp für den in der Regel ebenfalls konservierten Promotor. Bei der Berechnung der Sequenz-Übereinstimmungen (%) wurden Substitutionen und Insertionen/Deletionen vereinfachend als gleichwertige Ereignisse angesehen. Der Vergleich der Sequenzen unabhängiger Klone zeigt, daß nicht nur die Hämoglobin-Gene, sondern auch alle Intergenregionen nahezu identisch sind. Dies entspricht den Erwartungen für alle Sequenzen (vgl. z. B. Kao et al. (1995)).

### 3.5 Introns in Hämoglobin-Genen von *C. tentans*

Die Hämoglobin-Gene Cte N und Cte P besitzen je ein kurzes Intron. Das Gen Cte P ist auf dem Klon hb20.1 des Contigs 1 lokalisiert (siehe Abb. 3.1). Das Gen Cte N ist hingegen auf dem Klon hb9 des Contigs 2 nur unvollständig enthalten. Die fehlenden 186 bp bis zum 5'-Ende des Gens und weitere 260 bp der Promotorregion wurden daher mit Hilfe „inverser PCR“ aus genomischer *C. tentans*-DNA amplifiziert und durch Sequenzierung des PCR-Produktes ermittelt.

#### 3.5.1 Bestimmung der Intronposition

Das Intron des Hämoglobin-Gens Cte P konnte auf Grundlage der Nukleotid-Sequenz eindeutig lokalisiert werden. Im Falle des Gens Cte N konnte nur durch RT-PCR zwischen zwei möglichen Introns entschieden werden. Die Erststrang-cDNA-Synthese wurde an Gesamt-RNA aus *C. tentans* Larven durchgeführt. Der verwendete Primer (ten p61 5'-GTTGAATTCAGTGAAGAG-3') bindet ca. 100 bp abwärts des Introns. Das cDNA-Produkt dieser Reaktion wurde in der anschließenden PCR-Reaktion als Matrize eingesetzt. Die Position des Introns wurde auf zwei Arten ermittelt: Die PCR-Primer wurden so gewählt, daß sie spezifisch für die eine oder andere der beiden möglichen mRNAs sind (ten p165 5'-CCAAGCGTGTAAGAC/ATG-3' und ten p166 5'-CTTTTCGTCCATT/TCAAC-3'). Beide Primer enthalten ausschließlich Exonsequenzen und überspannen die möglichen Exon-/Introngrenzen. Der Schrägstrich in der Primersequenz markiert die Position, an der die fraglichen Intron-Sequenzen in der genomischen DNA vorliegen (siehe Abb. 3.6 A). Der zweite Reaktions-Primer stammte aus dem 5'-Bereich des Gens (ten p131 5'-GTGTCAGTCTAAGCCAC-3'). Erwartungsgemäß konnten mit keinem der beiden Primerpaare Amplifikate mit genomischer DNA als Matrize erzeugt werden. Ausgehend von der cDNA-Matrize



konnte hingegen ein 400 bp großes Fragment mit der Primerkombination ten p166/ten p131 amplifiziert werden. Die Primerkombination ten p165/ten p131 lieferte kein Produkt (siehe Abb. 3.6 B). Dies bedeutet, daß die Sequenz des Primers ten p166 in der mRNA existiert und nur durch das Spleißen eines Introns in diesem Sequenzabschnitt generiert worden sein kann.

In einem zweiten Ansatz wurden die PCR-Primer so gewählt, daß sie das Intron weiträumig flankieren (ten p61 und ten p107 5'-ATTACATCCTCCTTTTGGG-3'). Die Amplifikationsprodukte wurden sequenziert und die erwartete Intronposition bestätigt (siehe Abb. 3.7). Die Introns der Gene Cte N und Cte P sind in Tabelle 3.6 charakterisiert.

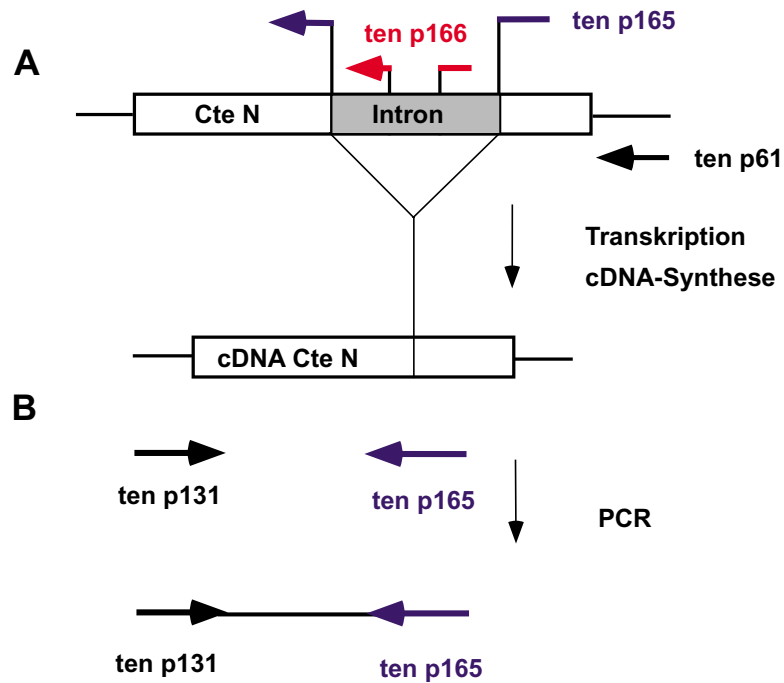


Abbildung 3.6 Legende siehe folgende Seite

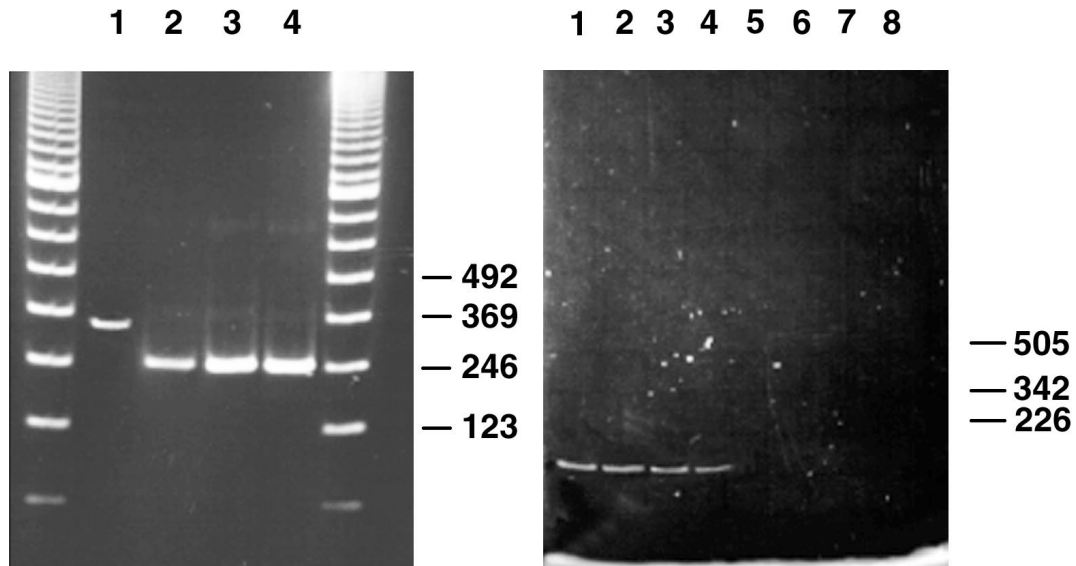


Abbildung 3.6: Ermittlung der Intronposition im Gen Cte N

Auf Grundlage der Nukleotid-Sequenz konnten die Introngrenzen des Hämoglobin-Gens Cte N nicht eindeutig bestimmt werden. Mittels reverser Transkription von larvaler mRNA wurde daher eine cDNA des Hämoglobin-Gens Cte N synthetisiert und die cDNA-Sequenz mit der ermittelten genomischen Sequenz verglichen. Es wurden zwei Arten von PCR-Reaktionen mit der cDNA als Matrize durchgeführt.

Die Abbildung A zeigt schematisch die Lage der Primer, die für die Bestimmung der Intronposition im Cte N Gen verwendet wurden. Die Primer ten p165 und ten p166 setzen sich jeweils aus den flankierenden Sequenzen der beiden möglichen Introns zusammen (blau, bzw. rot). Die cDNA-Synthese, ausgehend von dem Primer ten p61, liefert ein Produkt, das in der folgenden PCR-Reaktion nur einem der Primer als Matrize dienen kann.

Das rechte Gelbild zeigt die Ergebnisse der Amplifikationen mit der Primerkombination ten p131/ ten p166 (Spuren 1- 4) und ten p131/ ten p165 (Spuren 5- 8). Es sind jeweils zwei PCR-Reaktionen von zwei unabhängigen cDNA-Synthesen gezeigt. Bei allen Reaktionen mit dem Primer ten p165 konnte ein 400 bp langes PCR-Produkt detektiert werden. Die Ansätze mit dem Primer ten p166 zeigen dagegen kein PCR-Amplifikat. Dies bedeutet, daß die Introngrenzen, die durch den Primer ten p165 vorhergesagt wurden, der cDNA entsprechen.

Primer, die das Intron weiträumig flankieren, liefern mit genomischer Matrizen-DNA von *C. tentans*-Larven ein Amplifikat von ca. 330 bp (linkes Gelbild, Spur 1). Die entsprechenden PCR-Reaktionen mit cDNA als Matrize ergaben ein ca. 250 bp langes Produkt (linkes Gelbild, Spuren 2-3). Dieses wurde sequenziert und die Sequenz mit der genomischen Sequenz verglichen (siehe Abb. 3.7).

### 3.5 Introns in Hämoglobin-Genen von *C. tentans*

```

Cte N      ATGAAGTTTA TAGTAATAGC ATTGTGTATA GCTTGTGTCA CTGCTAACCC
                                     ten p131 ->
Cte N      ACATTGGGAT AATCTGGATG AAGTAGAAT  TGCCAGCATC  AAGAGGACAT
Cte N      GGAATCAAGT GAAATGAAT  GAAATTGAAA  TTCTTTACAA  TATTTTCAAA
Cte N      GATCATCCGG AAATTATGTC TAAATCCCA  CAATTCGCTG  GAAAGGATCT
Cte N      CGAAGAATTG AAAGGAACGA ACCAATTTGC AATTCATGCT ACAAGAATTG

Cte N      TCTCATTFFF CTCACATTAC ATCCTCCTTT TGGGAAAGGA ATCAAGTAAA
                                     ten p107 ->
Seq. 1073          TTAC ATCCTCCTTT TGGGAAAGGA ATCAAGTAAA
Seq. 1074          ATCCTCCTTT TGGGAAAGGA ATCAAGTAAA
Seq. 1075          TAC  ATCCTCCTTT TGGGAAAGGA ATCAAGTAAA
Seq. 1076          CATTAC ATCCTCCTTT TGGGAAAGGA ATCAAGTAAA
.....

Cte N      CCATGCCTTA GAACAATTAC TAATGTRATG GCGGAAGTGC ACAAGGCTCG
Seq. 1073  CCATGCCTKA GAACAATTAC TAATGTRATG GCGGAAGTGC ACAAGGCTCG
Seq. 1074  CCATGCCTKA GAACAATTAC TAATGTRATG GCGGAAGTGC ACAAGGCTCG
Seq. 1075  CCATGCCTKA GAACAATTAC TAATGTRATG GCGGAAGTGC ACAAGGCTCG
Seq. 1076  CCATGCCTKA GAACAATTAC TAATGTRATG GCGGAAGTGC ACAAGGCTCG
.....*.....

Cte N      AGGTGTACACA AAAGACGAAT TCATTAAATT TAAAGACTCA ATCTTCAAGT
Seq. 1073  AGGTGTMACA AAAGAYGART TCATTAAATT TAAAGACTCA ATCTTCAAGT
Seq. 1074  AGGTGTMACA AAAGAYGART TCATTAAATT TAAAGACTCA ATCTTCAAGT
Seq. 1075  AGGTGTMACA AAAGAYGART TCATTAAATT TAAAGACTCA ATCTTCAAGT
Seq. 1076  AGGTGTMACA AAAGAYGART TCATTAAATT TAAAGACTCA ATCTTCAAGT
.....*.....*.....

Cte N      ATTTAAAGGC TCATGTTGAG TAAGTGCTTG ATTTTAATAA AAATAGTCAT
                                     <-ten p165 CAT
                                     <-ten p166 GTGA
Seq. 1073  ATTTAAAGGC TCATGTTGA: .....: .....: .....:
Seq. 1074  ATTTAAAGGC TCATGTTGA: .....: .....: .....:
Seq. 1075  ATTTAAAGGC TCATGTTGA: .....: .....: .....:
Seq. 1076  ATTTAAAGGC TCATGTTGA: .....: .....: .....:
.....

Cte N      TAATCTAAAA TAATTTTAAAT TCCAGAATG GACGAAAAGG TCTTACACGC
                                     G TCTTACACGC
                                     AATG GACGAAAAC <-ten p166
Seq. 1073  .....: .....: .....: AATG GACGAAAAGG TCTTACACGC
Seq. 1074  .....: .....: .....: AATG GACGAAAAGG TCTTACACGC
Seq. 1075  .....: .....: .....: AATG GACGAAAAGG TCTTACACGC
Seq. 1076  .....: .....: .....: AATG GACGAAAAGG TCTTACACGC
.....

Cte N      TTGGAATGAC TGCTTTGATA GPATGTATTA CATCATTTTC GCCAATTTTG
                                     TTGG <-ten p165
Seq. 1073  TTGGAATGAC TGCTTTGATA GPATGTATTA CATCATTTTC GCCAATTTTG
Seq. 1074  TTGGAATGAC TGCTTTGATA GPATGTATTA CATCATTTTC GCCAATTTTG
Seq. 1075  TTGGAATGAC TGCTTTGATA GPATGTATTA CATCATTTTC GCCAATTTTG
Seq. 1076  TTGGAATGAC TGCTTTGATA GPATGTATTA CATCATTTTC GCCAATTTT
.....

Cte N      ATGGAAAACC AGTTAGCTAA AAATGAAAATCAACTTCAAATCCATAAACTC
Seq. 1073  ATGGAAAACC AGTTAGCTAA AAATGAAAAT
Seq. 1074  ATGGAAAACC AGTTAGCTAA AAATGAAA
Seq. 1075  ATGGAAAACC AGTTAGCTAA AAATGAAA
.....

```

Abbildung 3.7: Intronposition im Hämoglobin-Gen Cte N

Ausgehend von dem Primer ten p61 wurde eine reverse Transkription an Gesamt-RNA von *C. tentans*-Larven durchgeführt. Die resultierende cDNA wurde anschließend in einer PCR-Reaktion mit den Primern ten p61 und ten p107 als Template eingesetzt. Das PCR-Produkt wurde sequenziert (Seq. 1073-1076). Der Vergleich der genomischen Sequenz des Gens Cte N mit der cDNA-Sequenz legt die Intronposition eindeutig fest. Die Intron-Sequenz ist in blau, die Primersequenzen in rot, bzw. grün dargestellt. Unterschiedliche cDNAs, die von allen Cte N-Genen stammen, führen zu heterogenen Basen in der Sequenzierungs-Reaktion, so daß an vier Stellen der cDNA-Sequenz zwei verschiedene Basen an identischer Position auftreten (K = T und G, M = A und C, Y = C und T, R = A und G). Diese Positionen sind mit einem Stern gekennzeichnet.

	Cte P (hb20.1)	Cte N (hb9)
Intronposition NT-Sequenz	240-297	420-479
Intronposition AS-Sequenz	80 (2)	140 (2)
Intronposition Protein	Helix E 12-2	Interhelix G-H 2-2
Intronlänge (bp)	58	60
AT-Gehalt	84%	75%
Spleiß-Signale 5'/3'	GTAAG/CAG	GTAAG/CAG

**Tabelle 3.6:** Charakterisierung der Introns in den Hämoglobin-Genen Cte N und Cte P

### 3.6 Identifizierung und Charakterisierung eines Gens aus der Familie der Glukosetransporter

Bisherige molekulare Untersuchungen von Hämoglobinen-Gruppen bei Chironomiden zeigten, daß die Hämoglobin-Gene mit meist geringen Abständen aufeinander folgen (Hankeln und Schmidt, 1991; Hankeln et al., 1998; Rozynek et al., 1989; Gruhl et al., 1997; Trewitt et al., 1995, 1988). Bislang konnte lediglich ein Nicht-Hämoglobin interspergiert innerhalb eines Hämoglobinen-Clusters identifiziert werden. Hierbei handelt es sich um das Gen eines Zuckertransport-Proteins, das sowohl bei *C. melanotus* (Hankeln, Güttler und Schmidt, unveröffentlicht), als auch bei *C. tentans* die Hämoglobinen-Gruppe unterbricht (vergl. Abb. 3.5). Glukosetransport-Proteine (GLUT) sind Transmembranproteine, die den Hexosetransfer vermitteln und von einer Multigenfamilie codiert werden (Übersicht siehe Mueckler (1993)).

Die Ausdehnung des Leserahmens sowie die Exon-/Introngrenzen des GLUT-Gens von *C. tentans* wurden durch Vergleich mit der Nukleotid- und Aminosäure-Sequenz des orthologen Gens aus *C. melanotus* festgelegt. Die resultierende Genstruktur ist in Abb. 3.8 dargestellt. Das GLUT-Gen von *C. tentans* ist 1601 bp lang und kodiert für ein Protein von 481 AS. Es besitzt 3 Exons und 2 Introns. Die exakten Intronpositionen sind in Tab. 3.7 angegeben. Das GLUT-Gen ist an Position 15850-17450 in der Nukleotid-Sequenz des Contigs 1 zwischen den Hämoglobin-Genen Cte S und Cte 11 lokalisiert (siehe CD im Anhang; Datei: 50 kb Contig 1-21400.pdf; Seite 7-8 ).

Der Vergleich der Nukleotid-Sequenzen sowie der abgeleiteten Aminosäure-Sequenzen des GLUT-Gens der Arten *C. tentans* und *C. melanotus* zeigt den hohen Grad der Konservierung (siehe Abb. 3.8). Die Identität der kodierenden Bereiche beträgt 87 %, die der Aminosäure-Sequenzen 89 %. Werden Substitutionen durch eindeutig funktionell ähnliche Aminosäuren als Übereinstimmung gezählt, so steigt dieser Wert sogar auf 97,5 %.

Der Hydrophobizitäts-Plot zeigt den typischen Aufbau eines GLUT-Proteins mit 12 Transmembrandomänen. Die mit dem Program SOSUI (Hirokawa et al., 1998) festgelegten transmembranen Bereiche sowie vier für Glukosetransporter typische funktionelle Blöcke (identifiziert mit dem Programm Blocks; Henikoff und Henikoff (1994)) sind in Abb. 3.9 dargestellt. Abb. 3.10 zeigt schematisch den vermutlichen Aufbau des Glukosetransporters: Amino- und Carboxy-Terminus liegen im Cytoplasma, die vier funktionellen Domänen sind in der Zellmembran verankert und ragen in die plasmatischen Bereiche, wo sie mit dem Substrat interagieren können.

	Gen-Länge (bp)	Intron 1 (Länge in bp)	Intron 2 (Länge in bp)	TATA-Box*	Cap-Signal*	Poly-A-Signal*
Cte GLUT	1601	571-667 (97)	819-876 (58)	-63	-14	+70
Cme GLUT	1600	571-663 (93)	815-872 (58)	-63	?	+40

**Tabelle 3.7:** Struktur des Glukosetransporter-Gens in den Spezies *C. tentans* und *C. melanotus*

\* Die Angabe zur Position der TATA-Box und des Cap-Signals bezeichnet dessen Abstand zum Startkodon, entsprechend bezieht sich die Angabe zur Position des Poly-A-Signals auf das Stoppkodon.

### 3 Ergebnisse

Cte GLUT	TCTCCCAGAT	AAAGTAAAAA	CACAAAAAGT	TCTCACAGTA	TAATATCCAG
Cme GLUT	TGTCTCAGAT	AAAATACAAA	TGCGAAAAGT	TCTCTCAGTA	TAATAGCCAG
				TATA- Box	
Cte GLUT	TTAATATTTA	AACTTCATTG	AGTACATTTT	GATTAAAGTT	CATTGTCAAA
Cme GLUT	TTAAGACTTG	AACTTGATTG	TGATCATTTC	AATTAAAACT	CAGCAGCACA
				CAP	
Cte GLUT	ATGAAAATTT	CATTAAAAAA	TTTTTTTAGT	AAAAATCCT	TCAGTTCGAA
Cme GLUT	ATGAAAATTT	CATTTAAAAA	TTTCTTCAGT	AAAAACACT	TCAACTTAGA
		*	* *	* * *	* * * * *
Cte GLUT	TCCAAAAATG	GAGAATCTAA	CATCGAAGGC	AAAATACATT	CAAATTGGAT
Cme GLUT	CCCAAATAAA	GACAATTTAA	CAAGCAAAGC	AAAATACATT	CAAATTGGAT
	* * *	* *	* * *	*	*
Cte GLUT	TTGCTATGCT	TGCAAATTTT	TCAATATTAT	CAACTGGAAT	GGGTCTTGG
Cme GLUT	TTGCAATGCT	CGCAAATTTT	TCAATTTTAT	CAACAGGAAT	GGGTCTTGG
	*	*	*	*	*
Cte GLUT	TATCCAGCAA	TTACGACTCA	ATTGCTAGCA	AAAGATAAGA	CACTGATATT
Cme GLUT	TATCCAGCAA	TTACGACTCA	ATTACTAGCA	AAAGATGAGA	CAGTATATT
			*	*	*
Cte GLUT	AACAGAATCT	CAAGTTTCGT	GGTTTGCATC	GATCACAGCA	ATTGCATGTC
Cme GLUT	AACAGAATCT	CAAGTTTCTT	GGTTTGCATC	AATCACAGCT	ATCGCATGTC
		*		* *	*
Cte GLUT	CATTTCGGAGG	GCCAATAACC	GCTTATCTCA	CAAATAAGTT	TGGAAGAAAA
Cme GLUT	CTTTTGGAGG	ACCTATAACA	GCTTATTTCA	CAAATAAGTT	TGGAAGAAAA
	* *	* * *	*		*
Cte GLUT	GGAACCTTAA	TAATTATTGA	CGTAATTTCT	ATTATTCAGT	GGATTATTAT
Cme GLUT	GGAACCTTAA	TAATAATAGA	CGTAATTTCC	ATTATTCAGT	GGATAATTAT
		* *	*		*
Cte GLUT	TGGATTTTCA	AGTCGCAGTG	ATGCTCAGAT	ATTTTTTATT	CAATTAATGA
Cme GLUT	TGGATTTTCA	AGTCGTACTG	ATGCTCAAAT	ACTTTTTTATT	CAACTAATGT
		* *	*	*	* *
Cte GLUT	TTGCCAGAGT	GTTAACTGGA	CTTACGATCG	GAATGATAAC	AACTCCGGCT
Cme GLUT	TTGCAAGAGT	ATTAACCTGGA	CTTACGATCG	GAATGATAAC	AACTCCGGCT
	*	*	*		
Cte GLUT	GTGTTGTATT	CAGCAGAAAT	ATGCCATCCA	TCGTTGAGAG	GCAGAATGAT
Cme GLUT	GTTTTATATT	CAGCAGAGAT	ATGTCATCCG	TCGTTGAGAG	GCAGAATGAT
	* *	*	*	*	
Cte GLUT	GGTTTTATCG	ACTCCATTTT	TTGTGGCAAT	TGGAACTTTA	CTGATTTATT
Cme GLUT	GGTTTTATCG	ACTCCATTTT	TTGTATCAGT	TGGAACTTTG	TTGATTTATT
			* * *		* *
Cte GLUT	TGCTAGGGTA	TCTGATTGAG	GTAATGAAAT	TGAGGCCTTT	TACTTATTTT
Cme GLUT	TGCTGGGTTA	TTTAATACAG	GTAATAATAT	TGGTGATTTG	GATGTCATTC
	* *	* * *	* *	* * * * *	* * * * * *
Cte GLUT	GAATTCAAAG	ATTAGTTTCA	TGTGTATAAC	CTTGAATTTT	TATTTTATAG
Cme GLUT	ACTAATGACG	TTTGTCAAAA	CTGGGACATC	AATTAGTTTT	ATCTTTTTTT
	* * * * * *	* * * * * *	* * * * *	* * * * *	* * * * *
Cte GLUT	TGCATATATT	TTGACAGGAA	GACTATAGAT	TAGTAGCAAT	TTACTCCCTT
Cme GLUT	ATCCACATTT	::: CAGGAA	GACTATAGAT	TGGTAGCAAT	CTACTCTCTT
	* * * * * *	* * * * *		*	*
Cte GLUT	TGGATCACAA	TATTAACACT	ATTGATCTGT	TTTTTCATTC	CTGAATCTCC
Cme GLUT	GGGATCACAA	TCTTAACACT	CTTGATCTGC	TTTTTCATTC	CTGAATCTCC
	*	*	*	*	
Cte GLUT	AGTTCATCTT	ATTCTTAAGA	ATAAGTTGAA	AACAGCTCGT	CAAGTACTTT
Cme GLUT	AGTTCATCTT	ATTCTTAAAG	ATAAGCTGAA	AAGTGCTCGT	CATGTGCTCT
		* *	*	* *	* * *
Cte GLUT	CAAAGTTAG	AAATTTAGGT	AAAATTTATT	TAAAATTTCT	ATGCTTTTTA
Cme GLUT	CAAAGTTAG	AAATTTATGT	AAGTTGATTG	GAAATTCCTT	CAAATATTTG
		*	* * * * *	* * * * *	* * * * * *

Abbildung 3.8 Legende siehe folgende Seite

### 3.6 Charakterisierung eines Gens aus der Familie der Glukosetransporter

Cte GLUT	AAAATAATTT	AATCTCATT	ATTTAGTCAA	AGACGATCCA	AAAATTGATG
Cme GLUT	AAATTATTTA	AATTTCTGTT	ATCCAGCCAA	AGATGATGCA	AAAGTTGATG
	* * *	* ** *	** *	* *	*
Cte GLUT	AAGAAATTCT	CCAAATTCAA	CAAAACGCAC	AAAGA:::AC	TTCCGAAAAG
Cme GLUT	AAGAAATTCT	CCAAATTCAA	CAAAATGCTC	AAAGTTCAGC	ATCTAAAACA
			* *	*****	* ** **
Cte GLUT	TCAAATCTTT	TAATTTTCAG	AGAATTTGGG	AAGCCACAGC	TCTACAAACC
Cme GLUT	TCAAAACTGT	CAATATTCAG	AGAGTTTGCC	AAACCTCAAC	TCTATAAACC
	* * *	* * *	* **	* * *	*
Cte GLUT	ATFCATGATA	ATGCTTGCAT	TCTTCACAAT	TCAGCAACTT	TCCGGAATAT
Cme GLUT	ATTCTTGATT	ATGCTTGCAT	TCTTCACAAT	TCAACAACCTG	TCTGGAATAT
	* *			* *	*
Cte GLUT	TTGTGATTTT	CGTATATACT	GCACAGTTTT	CAATAGAAGC	TGGGGTATCA
Cme GLUT	TTGTGATTTT	TGTATATACT	GCACAATTTT	CCATTGAAGC	TGGCGTGCAG
		*	*	* *	* ** *
Cte GLUT	ATTGATGCAT	TTTTAAGTAC	AGTAATTATT	GGAATAATTC	GTTGTGTGAC
Cme GLUT	ATTGATGCCT	TTTTAAGCAC	AGTAATTATT	GGCACAATTC	GTTGTGTGAC
	*	*		* *	
Cte GLUT	AACTTTTGGG	ACTGCATTTT	TATCAGATAA	AGTTGGAAGA	AAGCCATTGG
Cme GLUT	GACATTTACA	ACTGCATTTT	TAAGTGATAA	AGTAGGAAGA	AAGCCATTAG
	* * **		* *	*	*
Cte GLUT	CCATAGTCTC	AAGTATTGGA	ATGTTCTTTA	GCATGTCAGG	ATTGGCATTG
Cme GLUT	CTATTGTCTC	AAGTATTGGA	ATGTTTTTCA	GCATGTCAGG	ATTGGCATTG
	* *		* **	*	*
Cte GLUT	AGCTCTGAAT	TTGCTTTAAA	AGACACAAAA	TTCTTTTGGG	TACCAGCAGC
Cme GLUT	AGCTCTGAAT	TTTCTGTGAA	AGACACAAAA	TTCTTTTGGG	TACCAGCAGC
		* * *			
Cte GLUT	ACTTTTGTAC	TTTTTCATTT	TTATTGGTAC	TTCTGGAGTT	TTAGTCCTTC
Cme GLUT	ACTTTTGTAC	TTTTTTGTCG	TCATTGGCAC	TTCTGGAATT	TTGGTACTTC
		** ** *	* *	*	* *
Cte GLUT	CATTTTCTAT	GGTCGCTGAA	ATGTATCCAC	AAAAGTCAAG	AAGTCTTGCT
Cme GLUT	CGTTTTTCGAT	GGTCGCTGAG	ATGTATCCAC	AAAAGTCAAG	AAGTCTTGCT
	* *	*			
Cte GLUT	GTTGGACTTT	CCCTTTCTTA	CTGCTTTATC	GTGAGCTTTT	TAACATTTAA
Cme GLUT	GTTGGACTTT	CACTTTCTTA	TTGCTTTTAT	ATTAGCTTTT	TAACAATTTAA
		*	*	* * *	*
Cte GLUT	ATTTTTCTCA	ACAGCTTTTC	TATTTTTTGG	CAGTGCATATA	GTGTTTACAT
Cme GLUT	GTTTTTCTCA	ACAGCATTTT	TGTTTTTGGG	AAGTGCCATA	TTATTTACAT
	*	* * *	*	* *	* *
Cte GLUT	TTTTTGCTAT	AATTTCAATTG	ATTAGCATAT	TTTCTCAAT	TTATGTTTTG
Cme GLUT	TTTTTGCTGT	AATTTCATTG	ATTGCAATAT	TTTTTTCGAT	TTATATCTTG
	*	*	***	* *	* *
Cte GLUT	CCTGAGACGA	AAGGGAAGTC	ACTTCATGAG	ATTGAGAAAA	ATTTTCAGAA
Cme GLUT	CCTGAGACGA	AAGGGAAGAC	ACTTAATGAG	ATTGAGAAAT	ATTTTCAGAA
		*	*	*	*
Cte GLUT	ATGATCTGAG	ATTTAATATC	TGATCGATAT	TTTTTATATT	TGATATATTT
Cme GLUT	ATAATGAGAG	ATTTAA	TAAAT	AATTTTTTAC	TGTTGAAATA
	*				TAAAAATAAA
Cte GLUT	ATTTGCTCTT	GGGAATCATT	AAATTTATAA	TGAATAAA	TAAAACATG
Cme GLUT	TCATACTTAA	GGGTGAATGA	CCCCTCAAAT	TAAAAATAAG	GTTTCAGTTA

Abbildung 3.8: Vergleich der Nukleotid-Sequenzen des Glukosetransport-Gens von *C. tentans* und *C. melanotus*

Durch den Vergleich der Nukleotid-Sequenzen konnte das Start- und Stoppkodon sowie die Positionen der Introns festgelegt werden. Das GLUT-Gen von *C. tentans* ist 1601 bp lang. Start und Stopp sind in grün dargestellt. Es besitzt zwei kurze Introns von 97 und 58 Basenpaaren (blau), außerdem weist es die für die Expression nötigen Signalsequenzen auf (rot). Die Sequenzunterschiede zwischen den beiden Spezies sind durch Sterne gekennzeichnet. Die Sequenz-Übereinstimmung der kodierenden Genbereiche beträgt 87%.

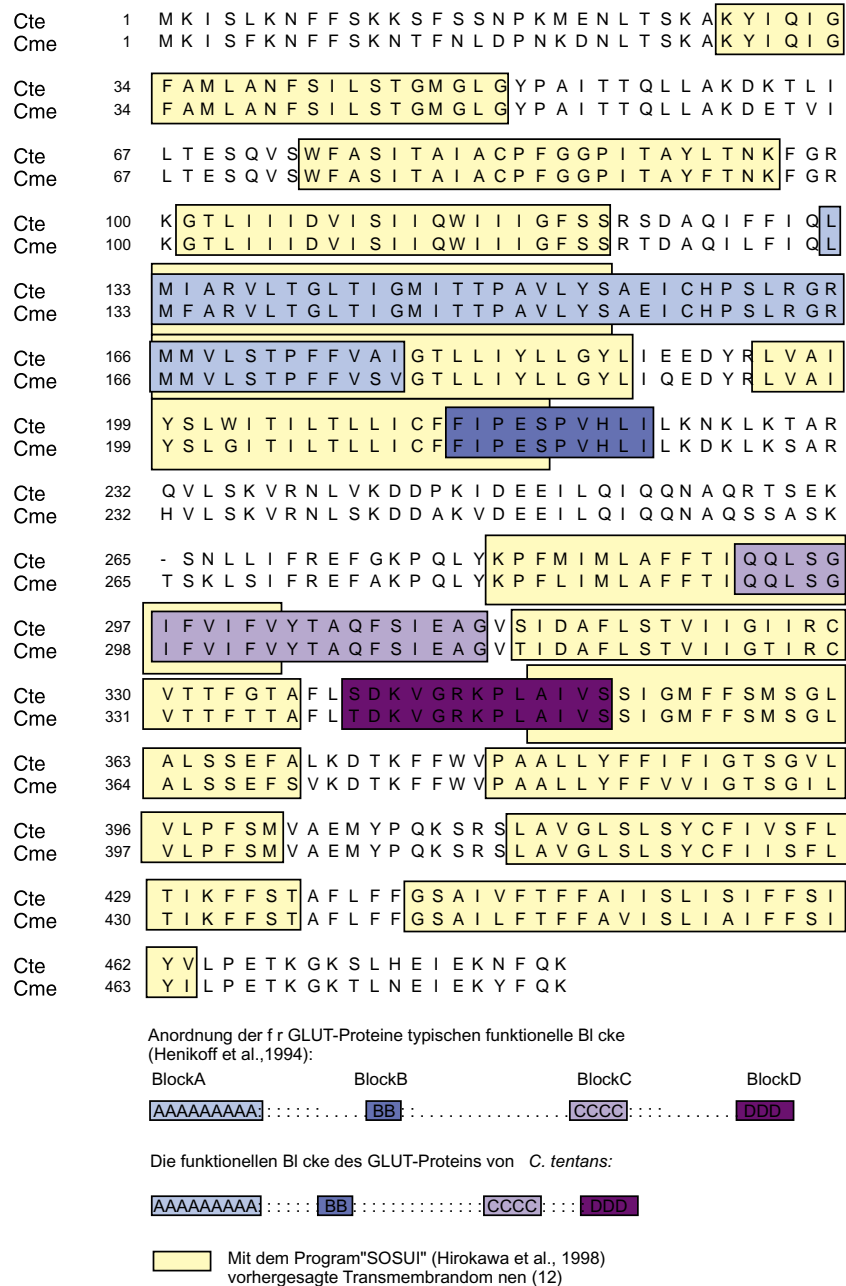


Abbildung 3.9: Die Struktur des GLUT-Proteins

In den Aminosäure-Sequenzen der GLUT-Varianten von *C. tentans* und *C. melanotus* (Hankeln, Güttler und Schmidt, in Vorbereitung) können 12 transmembrane Bereiche festgelegt werden (gelb). Sie umfassen zwischen 21 und 23 Aminosäuren und zeichnen sich durch den überdurchschnittlich hohen Anteil hydrophober Seitenketten aus. Die für Glukosetransporter charakteristischen funktionellen Domänen sind blau dargestellt. Sie stimmen in Sequenz, Größe und Anordnung mit denen bekannter GLUTs überein und belegen, daß es sich bei dem putativen Protein tatsächlich um einen membranständigen Glukosetransporter handelt.



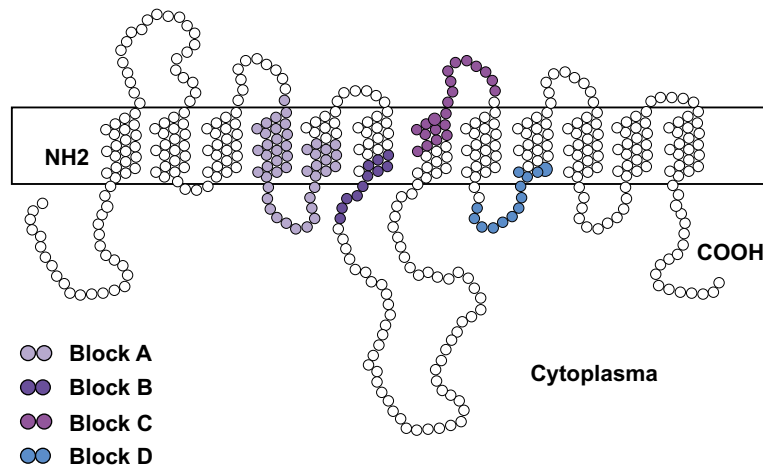


Abbildung 3.10: Vorhergesagte Topologie des Cte GLUT-Proteins

Der Glukosetransporter von *C. tentans* besitzt die für GLUT-Proteine typischen 12 Transmembrandomänen (vorhergesagt mit SOSUI, Hirokawa et al. (1998)). Die für den Hexosetransport wichtigen funktionellen Blöcke A-D sind mit den zentralen Transmembrandomänen assoziiert (farbige Bereiche). Sowohl der Carboxy- als auch Amino-Terminus sind im Cytoplasma lokalisiert. In Zusammenhang mit der höheren Struktur der Glukosetransporter werden unterschiedliche Hypothesen diskutiert. Möglicherweise bilden die membranständigen Aminosäuren amphiphatische  $\alpha$ -Helices, deren hydrophile Bereiche sich zu einem Membrankanal zusammenlagern (Mueckler et al., 1985).

### 3.7 Repetitive Elemente im Hämoglobingen-Cluster

In den Intergenbereichen des Contigs 1 sind zwei Retroposons, sogenannte Cp1-Elemente integriert. Es handelt sich hierbei um SINES („short interspersed nuclear elements“) die zuerst von Hong et al. (1995) aus der Centromerregion von *C. pallidivittatus* isoliert wurden. Charakteristisch für CP1-Elemente ist ihr variabler Aufbau aus zwei Sequenzmodulen, dem SCA- und dem B-Element. Diese kommen in verschiedenen CP1-Kopien in unterschiedlicher Anzahl, Kombination und Orientierung vor (Hong et al., 1995). Die von He und Mitarbeitern durch den Vergleich mehrerer Cp1 Sequenzen unterschiedlicher chromosomaler Herkunft von *C. pallidivittatus* und *C. tentans* erstellte Konsensussequenz diente hier zur Identifizierung der SCA- und B-Elemente.

Cp1-Retroposons liegen in der Nukleotid-Sequenz des Contigs 1 an den Positionen 34539-34988 und 38173-38620 innerhalb der Subfamilie der 7A-Hämoglobin-Gene. Es handelt sich hierbei um Abschnitte der Intergenregion 18, flankiert von den Hämoglobin-Genen Cte 7A4 und Cte 7A5 und der Intergenregion 20 zwischen den Hämoglobin-Genen Cte 7A6 und Cte 7A7 (siehe auch Abb. 3.5, die Positions-

angaben beziehen sich auf die Gesamtsequenz auf der beiliegenden CD; Datei:50 kb Contig 21401-50607.pdf; Seiten 4-5 und 7-8).

Die Abb. 3.11 A zeigt den Aufbau des Cp1-Elements aus der Intergenregion 18 (Cp1/ 18). Es beginnt mit einem 36 bp großen Fragment aus dem 3'-Bereich eines B-Elements in umgekehrter Orientierung. Nach 44 bp unbekannter Herkunft folgt ein weiteres B-Element, dem 29 bp des 5'-Bereiches fehlen. Das 3'-Ende bildet ein vollständiges SCA-Element. Eine mögliche, 13 bp lange „target-site“-Duplikation flankiert diesen Bereich. Der Aufbau des Cp1-Elements der Intergenregion 20 (Cp1/ 20) ist in Abb. 3.11 B dargestellt. Es beginnt mit einem unvollständigen SCA-Element, dem 57 bp des 5'-Bereichs fehlen. Hierauf folgt ein B-Element, das eine zentrale Deletion von 81 bp aufweist. Die 16 bp, die diese Deletion am 5'-Bereich flankieren, liegen dupliziert vor. Das 3'-Ende des Cp1/ 20 wird von einem vollständigen SCA-Element gebildet. Flankiert wird es von einer „target-site“-Duplikation, deren Länge zwischen neun und zwölf Basenpaaren liegt. Die exakte Größe kann nicht eindeutig festgelegt werden, da die drei Adenin-Basen am 3'-Ende des Cp1-Elements sowohl aus dem SCA-Element als auch aus der „Target“-Sequenz stammen können. Die von Hong et al. (1995) beschriebenen CP1-„target-site“-Duplikationen besitzen Längen zwischen zehn und vierzehn Basenpaaren. Auch hier konnte in einigen Fällen die exakte Grenze der Cp1-Sequenz nicht eindeutig bestimmt werden. Dies hängt möglicherweise damit zusammen, daß Retrotransposons bevorzugt in A/T-reiche DNA-Abschnitte integrieren (Rogers, 1985) und auch die Enden der Cp1-Elemente stets A/T-reich sind (Hong et al., 1995). Alle B-Elemente besitzen die erforderlichen Promotor- und Terminationssequenzen für die Polymerase III Transkription (Hong et al., 1995). Dies läßt darauf schließen, daß es sich hierbei um ehemalige tRNA-Sequenzen handelt. Den SCA-Elementen fehlen jegliche bekannten Promotorsequenzen. Aufgrund der Sequenzähnlichkeiten, die zwischen B- und SCA-Elementen vorliegen, wurde spekuliert, daß beide auf einen gemeinsamen Ursprung zurückgehen (Hong et al., 1995).

### 3.7 Repetitive Elemente im Hämoglobinen-Cluster

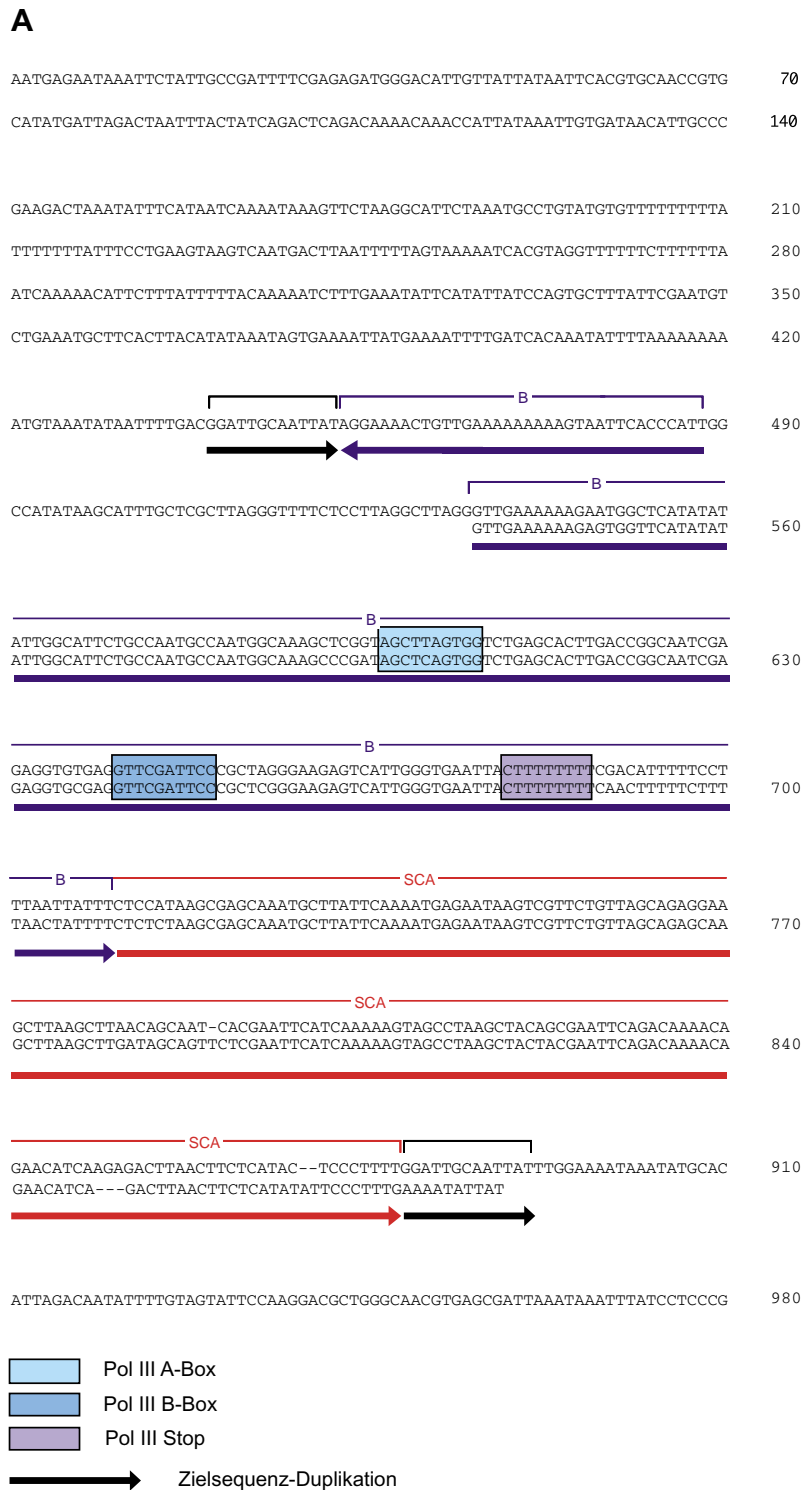
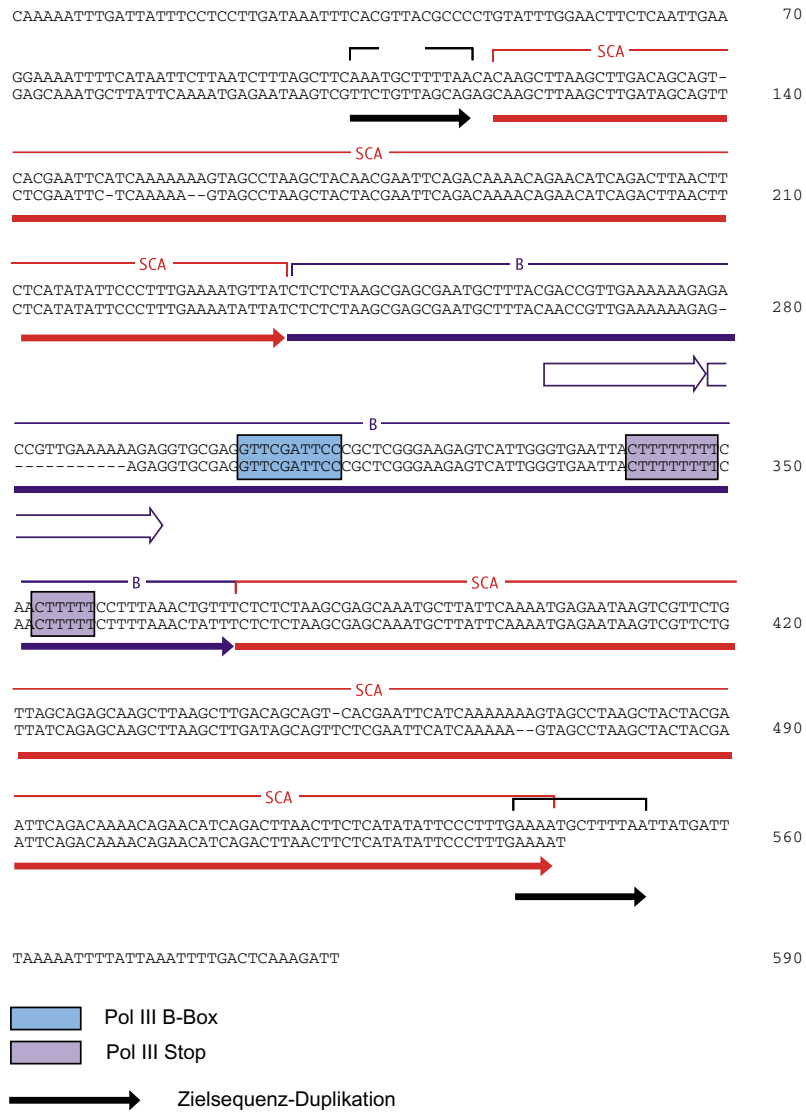


Abbildung 3.11 Legende siehe Seite 77

**B**



**Abbildung 3.11: Struktur der Cp1-Retroposons in der Hämoglobin-Gengruppe D**

Zwei dieser als Cp1 bezeichneten SINE-Elemente sind in der 7A-Hämoglobin-Gengruppe lokalisiert. Nach ihrer Integrationsstelle in verschiedenen Intergenregionen des Contigs 1 werden sie im folgenden als Cp1/18 (A), bzw. als Cp1/20 (B) bezeichnet. Sie bestehen aus je zwei Sequenzmodulen, dem SCA-Element (rot) und dem B-Element (blau). Die Konsensus-Sequenz dieser Module wurde von He und Mitarbeitern (1995) ermittelt und ist unterhalb der Cp1-Sequenzen aus dem Hämoglobin-Cluster dargestellt. Die Orientierung der einzelnen Sequenz-Elemente ist durch die Pfeilrichtung angegeben.

Cp1-Elemente sind generell durch ihren sehr variablen Aufbau gekennzeichnet. SCA- und B-Elemente sind in unterschiedlicher Zusammensetzung, Orientierung, sowohl komplett als auch in Teilen, vorhanden. Ungewöhnlich bei den hier identifizierten Cp1-Elementen ist die Insertion einer Sequenz unbekannter Herkunft zwischen zwei B-Abschnitten in Cp1/18 (siehe A). Innerhalb des Cp1/20 (in B gezeigt) ist ein Teil der B-Element-Sequenz dupliziert. Das Vorhandensein der PolIII-Promotorsequenzen weist auf die Herkunft der Cp1-Elemente aus tRNA-Sequenzen hin (Kästen). Beide Cp1-Elemente besitzen „target site“-Duplikationen (schwarze Pfeile).



## 4 Diskussion

Im Rahmen dieser Arbeit konnten innerhalb der Hämoglobin-Gengruppe D von *C. tentans* 28 Hämoglobin-Gene identifiziert werden. Der gesamte, zur Zeit bekannte Bereich dieser Gen-Gruppe umfaßt somit insgesamt 38 Hämoglobin-Gene (Rozynek et al. (1991), Rozynek, Hankeln und Schmidt, unveröffentlicht, diese Arbeit) und erstreckt sich über eine Länge von ca. 75000 Basen. Trotz der großen Redundanz innerhalb der Hämoglobin-Genfamilie der Chironomiden sind funktionslose Pseudogene offenbar sehr selten. Bis auf das Gen ORF B (Rozynek et al., 1991) weisen alle Hämoglobin-Gene die für ihre Expression erforderlichen Signal-Sequenzen auf. Die Gengruppe D von *C. tentans* ist somit die zur Zeit am besten charakterisierte Hämoglobin-Multigenfamilie in Chironomiden.

Voraussetzung für eine mögliche Rekonstruktion evolutionärer Beziehungen der Hämoglobine von *C. tentans* ist die Identifizierung orthologer Varianten in anderen Chironomus-Arten. Nur diese lassen sich auf einen gemeinsamen Gen-Vorläufer zurückführen und ermöglichen es somit, Veränderungen der Gen-Sequenz, -Struktur und -Organisation seit der Trennung der Spezies *C. tentans* von Schwesterarten zu dokumentieren. Die Klassifizierung der Hämoglobin-Gene von *C. tentans* erfolgt zunächst durch den Vergleich ihrer abgeleiteten Aminosäure-Sequenzen mit den bei Goodman et al. (1983) beschriebenen 11 Protein-Hauptkomponenten. Hämoglobin-Gen-Sequenzen weiterer Varianten stehen aus den Spezies *C. thummi* und *C. piger* für weitere Vergleiche zur Verfügung.

Die Hämoglobine zählen zu den strukturell und funktionell am besten untersuchten Proteinen. Aus der Ermittlung ihrer dreidimensionalen Struktur durch Röntgen-Kristallographie leiteten sich erste Erkenntnisse über die komplexen Beziehungen zwischen Proteinstruktur und Funktion ab (Dickerson und Geis, 1969, 1983).

Eine Bewertung der Unterschiede in den Aminosäure-Sequenzen von Hämoglobin-Varianten bei Chironomiden ist möglich, da die Röntgenstruktur der Hämoglobin-Komponente III aus der Spezies *C. thummi* bekannt ist (Steigemann und Weber, 1979), und eine Reihe funktionell wichtiger Aminosäure-Positionen bereits festgelegt werden konnte. Obwohl zur Zeit die 3D-Strukturen der in dieser Arbeit identifizierten Hämoglobine von *C. tentans* nicht vorliegen, ist es auf der Grundlage dieser Daten in einigen Fällen möglich, die Auswirkungen einzelner Aminosäure-Austausche auf die Proteinstruktur abzuschätzen.

Um die Hämoglobin-Familie der Chironomiden auf der Ebene der Gene zu charakterisieren, soll die Architektur des Hämoglobin-Gen-Clusters D in *C. tentans* umfassend analysiert werden. Hieraus lassen sich allgemeingültige Erkenntnisse über

die Evolution der Hämoglobin-Genfamilie gewinnen. So wird beispielsweise der Ursprung der Introns in den Hämoglobin-Genen seit langem kontrovers diskutiert. Besaß das ancestrale Hämoglobin-Gen zwei Introns, wie es bei den Vertebraten zu finden ist, oder wies es zusätzlich ein drittes Intron auf, wie es dem Typus der heutigen Pflanzen-Hämoglobin-Gene entspräche? Es soll daher untersucht werden, welchen Beitrag die Verteilung der Introns in den Hämoglobin-Genen der Chironomiden zur Klärung dieser Frage leisten kann.

Durch die molekulare Analyse der paralogen Hämoglobin-Gene von *C. tentans* können Gen-Subfamilien und „single copy“-Gene identifiziert und Sequenz-Verwandtschaften aufgezeigt werden.

Anhand von einzelnen Gen-Beispielen sollen mögliche Mechanismen der molekularen Evolution von Genfamilien aufgezeigt werden. Die vergleichende Analyse von Hämoglobin-Gengruppen in unterschiedlichen Chironomiden-Spezies ermöglicht es schließlich, Veränderungen des Genrepertoires und der Clusterstruktur nach Aufspaltung der Taxa aufzuzeigen. Das evolutionäre Schicksal einzelner Hämoglobin-Gene während der Speziation kann im Interspezies-Vergleich rekonstruiert werden. Indirekt sind hierdurch auch Rückschlüsse auf die funktionelle Bedeutung einzelner Hämoglobin-Varianten möglich.

### 4.1 Die Klassifizierung der Hämoglobin-Varianten von *C. tentans*

Bereits frühe Arbeiten zeigten, daß Chironomiden einen ausgeprägten Polymorphismus in der Zusammensetzung ihrer Hämolymphe besitzen (Thompson und English, 1966; Tichy, 1970). Seitdem konnten eine Vielzahl unterschiedlicher Hämoglobin-Varianten aus verschiedenen Spezies isoliert und molekular charakterisiert werden (Antoine et al., 1987; Rozynek et al., 1991; Kao et al., 1994; Kao und Bergtrom, 1995; Kao et al., 1995; Gruhl et al., 1997; Hankeln et al., 1988; Hankeln und Schmidt, 1991; Hankeln et al., 1998; Trewitt et al., 1995). Durch den Vergleich der Hämoglobin-Sequenzen von *C. tentans* mit den bereits vorliegenden Daten sollte es daher möglich sein, die Hämoglobin-Gene von *C. tentans* zu gruppieren und ihren möglichen orthologen Partnern in anderen Arten zuzuordnen. Dies soll primär durch einen phylogenetischen Vergleich der abgeleiteten Aminosäure-Sequenzen der Gene erfolgen.

In einem ersten Schritt werden alle ermittelten Hämoglobin-Sequenzen von *C. tentans* mit den von Braunitzer für *C. thummi* beschriebenen Protein-Sequenzen (zusammengefaßt in Goodman et al. (1983)) verglichen. Durch die Zuordnung der vermutlich orthologen Hämoglobin-Varianten wird ersichtlich, ob alle in *C. thummi* bekannten Hämoglobin-Proteine auch in *C. tentans* identifiziert wurden.

Die Klassifizierung der verbleibenden Hämoglobin-Varianten von *C. tentans* kann anschließend durch phylogenetische Vergleiche mit weiteren ausschließlich auf Nukleotid-Ebene beschriebenen Hämoglobin-Varianten erfolgen. Eine besonders umfassende

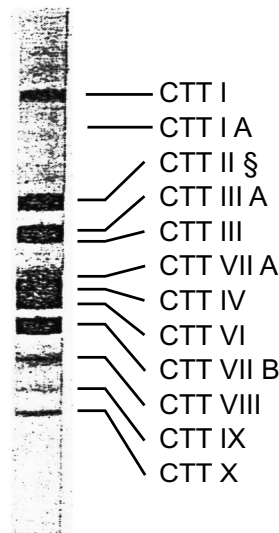


de Kenntnis über das Hämoglobin-Gen-Repertoire der Chironomiden liegt hierbei von den Arten *C. thummi* und *C. piger* vor (Hankeln und Schmidt, 1991; Trewitt et al., 1995; Gruhl et al., 1997). Weitere Kriterien, die zur Identifizierung von Genen beitragen können, sind ihre Lokalisation innerhalb des Genclusters, sowie die Sequenzen der sie umgebenden Intergenbereiche.

Am Ende diese Kapitels soll eine Übersicht Aufschluß über alle bislang bekannten Hämoglobin-Varianten bei Chironomus geben.

#### 4.1.1 Die Klassifizierung der Hämoglobin-Varianten auf Protein-Ebene

Auf Protein-Ebene liegen zur Zeit die Sequenzdaten von 11 Hämoglobin-Varianten aus der Spezies *C. thummi* vor (zusammengefaßt in Goodman et al. (1983), siehe Abb. 4.1).



**Abbildung 4.1:** Gelelektrophoretische Auftrennung der von Braunitzer und Mitarbeitern ermittelten Hämoglobin-Varianten von *C. thummi* (aus Goodman et al. (1983)).

Die Gene für die dimeren Hämoglobin-Varianten IX, IIB, VIIA und VIIB, sowie für die monomeren Varianten III und IV, konnten bereits in *C. tentans* nachgewiesen werden (Rozynek et al. (1991), Hankeln und Schmidt, unveröffentlicht). Der Vergleich aller 38 in der vorliegenden Arbeit beschriebenen Hämoglobin-Sequenzen von *C. tentans* mit den Protein-Varianten von *C. thummi* zeigt, daß zusätzlich auch die Gene Cte 1, Cte 8 und mit einiger Wahrscheinlichkeit das Gen Cte 6 in *C. tentans* vorhanden sind und für die entsprechenden Protein-Varianten CTE I,

CTE VIII und CTE VI kodieren. Die Hämoglobin-Sequenzen aus *C. thummi* sowie ihre vermutlich orthologen Partner aus *C. tentans* sind in Abbildung 4.2 im Vergleich gezeigt. Die Zuordnung der Aminosäuren zueinander (Alignment) erfolgte mit Hilfe des Programms Clustal W (Higgins und Sharp, 1988, 1989). Das Computer-Alignment konnte anschließend unter Einbeziehung der Nukleotid-Sequenzen, die mittlerweile von 9 der 11 *C. thummi* Hämoglobine vorliegen, optimiert werden (I/IA, Kao et al. (1995); III/IV, Antoine et al. (1987); IX, IIB, VIIA, Rozynek et al. (1991); Kao et al. (1994); Gruhl et al. (1997); VIIB, Hankeln et al. (1988); Hankeln und Schmidt (1991); Trewitt et al. (1995); VI, Kao und Bergtrom (1995); Kao et al. (1995); VIII, Hankeln et al. in Vorbereitung). Ausgehend von diesem Sequenz-Alignment wurden die Unterschiede der Sequenzen berechnet (Kumar et al. (1993); siehe Tab.4.1). Hierbei wurden die Sequenzen paarweise verglichen und der Anteil der Aminosäure-Austausche an der Gesamtsequenz ermittelt (p-Distanzen). Lücken im Sequenz-Alignment, die auf Insertionen bzw. Deletionen von Nukleotiden zurückzuführen sind, wurden nicht berücksichtigt.

	CTT VI	CTT VIII	CTT IX	CTT IIB	CTT VIIA	CTT VIIB	CTT I	CTT Ia	CTT III	CTT IV	CTT X
CTE U	0.158	0.574	0.525	0.566	0.515	0.505	0.622	0.622	0.544	0.587	0.614
CTE VIII	0.483	0.119	0.407	0.407	0.414	0.455	0.532	0.532	0.566	0.574	0.533
CTE IX	0.469	0.384	0.200	0.345	0.331	0.448	0.546	0.546	0.537	0.552	0.537
CTE IIB	0.483	0.397	0.352	0.186	0.352	0.441	0.559	0.559	0.478	0.537	0.537
CTE VIIA1	0.448	0.425	0.345	0.338	0.152	0.317	0.552	0.552	0.515	0.515	0.565
CTE VIIB1	0.408	0.480	0.448	0.428	0.386	0.221	0.566	0.566	0.544	0.537	0.550
CTE I	0.545	0.507	0.552	0.511	0.497	0.552	0.252	0.259	0.545	0.575	0.571
CTE E	0.526	0.551	0.591	0.562	0.504	0.547	0.585	0.585	0.348	0.341	0.583
CTE III1	0.522	0.540	0.544	0.529	0.485	0.581	0.537	0.537	0.082	0.163	0.580
CTE IV1	0.507	0.533	0.544	0.552	0.500	0.559	0.522	0.530	0.185	0.133	0.601

**Tabelle 4.1:** Unterschiede der Hämoglobin-Sequenzen von *C. tentans*, *C. thummi* und *D. melanogaster* (p-Distanzen)

Basierend auf dem Alignment der Aminosäuren (siehe Abb. 4.2) wurden die Sequenzen paarweise verglichen. Die angegebenen Distanzwerte stellen ein Maß für die Unterschiede zwischen den Sequenzen dar. Sie basieren auf dem Anteil der Aminosäure-Austausche bezogen auf die Länge der Gesamtsequenz. Die höchsten Sequenz-Übereinstimmungen sind rot hervorgehoben.

## 4.1 Die Klassifizierung der Hämoglobin-Varianten von *C. tentans*

	1	60
CTEU	-----	-----
CTEVIII	MKFLVLALCTAAASA--TPMTADQLALFKSSWAEVK---HNEVDILYAIFKDHDPDIQA	
CTEIX	MKFLAVLALCIVGAVAD--PISADQADLIRSSWAQVK---HNEVDILAAVFKDHPDIQA	
CTEII $\beta$	MKFL-VLALCIVGAI--PLSADQADLIRSSWAQVK---HNEVDILYAIFKDHDPDIQA	
CTEVIIA-01	MKFFAVLALCIVGAIAT--PLTADQASLVQSTWAQVR---NNEVDILAAVFTANPDIIQA	
CTEVIIB-01	MKFFAVLAVCIVGAIAT--PLTADQANLVQSSWAQVK---NNEVDILAAVFAAYPDIIQA	
CTEI	MKFL-ILALCVAYTMAA--P-SADQIAAAKASWATVK---GKQVDILAAIFKANPDIDK	
CTEE	MKFI-ILALCVAAAASA---LSADQIGLVQTTYGKVK---GDSTGILYAVFKADSSIIQA	
CTEIII-1	MKFL-ILALCFAAAASA---LSADQISTVQASFDKVK---GDSVGLYAVFKADPSIQA	
CTEIV-1	MKFI-VLALCFAAAASA---LTADQISTVQASFDKVK---GDSVGLYAVFKADPSIQA	
CTVI	-----A-VLTTEQADLVKKTWSTVK---FNEVDILYAVFKAYPDIIQA	
CTII $\beta$	-----A-PLSADQASLVQSSWAQVK---HNEVDILYAIFKDHDPDIQA	
CTTIX	-----D-PVSSDEANAIASWAGVK---HNEVDILAAVFSDDHPDIQA	
CTTVIII	-----AV-TPMSADQLALFKSSWNTVK---HNEVDILYAVFKANPDIIQA	
CTTVIIA	-----A-PLSADQASLVKSTWAQVR---NNEVDILAAVFTAYPDIIQA	
CTTVIIB	-----S-PLTADQASLVQSSWKAIVS---HNEVDILAAVFAAYPDIIQA	
CTTII	-----LSADQISTVQASFDKVK---GDSVGLYAVFKADPSIQA	
CTTIV	-----LTADQISTVQSSVQAVK---GDAVGLYAVFKADPSIQA	
CTTIIa	-----G-P-SGDQIAAAKASWNTVK---NNQVDILYAVFKANPDIIQT	
CTTI	-----G-P-SGDQIAAAKASWNTVK---NNQVDILYAVFKANPDIIQT	
CTTX	-----DPEWHTLDAHEVEQVQATWKAIVS---HDEVEILYVFKAHDPDIQA	
Dme glob1	-----NSDEVQLIKKTWEIPVATPTDGSAAILTQFFNRFPSNLE	
	120	
CTEU	-----DLESVKDSPAFATHATRIISFLSQVIVSLAGSDDTIPAIQNLAKDLATSHKPR	
CTEVIII	RFPPQFAGKDLDSLKDTAAAFATHAGRIVGFPSDVIIGLIGNSENRPALKTLIDGLASNHKAR	
CTEIX	RFPPQFAGKDVDSLKDTAAAFATHAGRIVGFISEIIVALVGNENRPMNTLTNELATNHHR	
CTEII $\beta$	KFPQFVQKDLDSLKDTAAAFATHAGRIVGFMSQI IALMGNEANRPAMVTLINEMAANKAR	
CTEVIIA-01	RFPPQFAGKDVAAKDTAAAFATHAGRIVGFVSEI IALVGNESNAPAVQTLVGLQAASHKAR	
CTEVIIB-01	KFPQFAGKDLAAVKTAAAFATHATRIIVSFLSELISLGSNANAAAINSLVTKLGSDDHQR	
CTEI	KFSQFAGKDVDSIKGTPEFATHAGRIVGFPGDVMDDLGNANTATIVSKAKDLGKTHKR	
CTEE	AFPQFVQKDLDAIKGAEFATHAGRIVGLFSGV---IDDLPIGKH---VDALVATHKPR	
CTEIII-1	KFTQFAGKDLDSIKGTAPFESHANRIVGFPSKI I---IGDLPSIEGD---VDTFVASHKPR	
CTEIV-1	KFSQFAGKDLDSIKGSAEFASHAGKI VGFPSKI I---VGDLPNIGD---VDTFVASHKPR	
CTVI	KFPQFAGKDLDSIKDSAAAFATHATRIIVSFLSEVIISLAGSDANI PAIQNLAKELATSHKPR	
CTII $\beta$	KFPQFAGKDLDTLKGTFATHAGRIVGFVSEI IALVGNESNANPAMETLIKDMAANKAR	
CTTIX	RFPPQFAGKDLASIKDTGAFATHAGRIVGFISEIIVALVGNESNAPAMATLINELSTSHHR	
CTTVIII	KFPQFAGKDLDSIKDSADFAVHSGRIVGFPSVIGLIGNPENRPAKTLIDGLASSHKAR	
CTTVIIA	RFPPQFAGKDVASIKDTGAFATHAGRIVGFVSEI IALVGNESNAPAVQTLVGLQAASHKAR	
CTTVIIB	KFPQFAGKDLASIKDTGAFATHATRIIVSFLSEVI IALVGNESNAAVQGLDLKGDHDKAR	
CTTII	KFTQFAGKDLDSIKGTAPFETHANRIVGFPSKI I---GELPNI EAD---VNTFVASHKPR	
CTTIV	KFTQFAGKDLDSIKGSADFAHANKI VGFPSKI I---GDLPI DGD---VTFVASHKPR	
CTTIIa	AFSQFAGKDLDSIKGTPDFSKHAGRVVGLFSEVMDLLGNANTPTIILAKAKDFGKSHKSR	
CTTI	AFSQFAGKDLDSIKGTPDFSKHAGRVVGLFSEVMDLLGNANTPTIILAKAKDFGKSHKSR	
CTTX	KFPKQFAGKDLDAIKDTADFAVHASRIIGFPGYVTLGSSGNQAARIITLLHDLGVFHKTR	
Dme glob1	KFP-PRDVPLEELSGNARFRAHAGRIIRVFDSEIQVLGQDGLKLEDEYWKIAVSHIPR	
	174	
CTEU	GVSKDQFTEFRALVSYLKSHITFDEPTETAWTLALETYYKMLFAAMDS----	
CTEVIII	GIEKAQFEEPRASLVSYLSKHLWNTMKSTWDLALNNMFFYILHALESAAH----	
CTEIX	GISKAQFNEFRASMTSYLSHHTWNTDATAAAWTHGLDNIFDAIPAHL----	
CTEII $\beta$	KIPQAQFNEFRASLVSYLSQSHVTWNTKLGAAWTQGLDNATMIFAKL----	
CTEVIIA-01	GISQAQFNEFRASLVSYLQANVAVGNDVNAAAWQGLDNIVYGLLPAAL----	
CTEVIIB-01	GVSAQFNEFRALVAYLQANVSWGNVAAAWQALDNITYAIAPTTLTA----	
CTEI	-ASPAQFDSFRKSLVVLKGTATWDAVTVESSWNAVLDVFGVLEQLAAAK----	
CTEE	GVTHAQFNNPRAAFIGYLGKGVHYTAAVEAAWATFDVDFGAVFAKM----	
CTEIII-1	GVTHDQLNNFRAGFVSYMKAHTDY--AASEAWGATLDTFFGMIFAKM----	
CTEIV-1	GATHDQLNNFRSAFVSYMKAHTDF--AGESAAGATLDAFFGAVFAKM----	
CTVI	GVSKDQFTEFRALFTYLKAHINFDGPTETAWTLALDITYYAMLFSAMDS----	
CTII $\beta$	GIPKAQFNEFRASLVSYLSQSVSWNTSLGAAWTQGLDNVFNMMPSYL----	
CTTIX	GITKQFNEFRSLSVSYLSHSAWNTADATWTHGLDNIFGMIFAHL----	
CTTVIII	GIEKAQFEEPRASLVSYLSHHLWNTMKSTWDLALNNMFFYILHALEVAQ----	
CTTVIIA	GISQAQFNEFRAGLVSYVSNVAAWNAASAWTAGLDNIFGLLPAAL----	
CTTVIIB	GVSAQFGEFRLALVAYLQAHVSWGNVAAAWSKALDNITYAIVVPRL----	
CTTII	GVTHDQLNNFRAGFVSYMKAHTDF--AGAEAAGATLDTFFGMIFSKM----	
CTTIV	GVTHDQLNNFRAGFVSYMKAHTDF--AGAEAAGATLDAFFGMVFAKM----	
CTTIIa	-TSPAQLDNFRKSLVYLVKGTAKWDSAVESSWAPVLDVDFVSTLKNEL----	
CTTI	-ASPAQLDNFRKSLVYLVKGTAKWDSAVESSWAPVLDVDFVSTLKNEL----	
CTTX	GITKAQFGEPRETMTAYLKGHNKWNADISHWDDAFDKAFSVIFEVLES----	
Dme glob1	TVSKESYNQLKGVILDLVLTAAACSLDESQAATWAKLVDHVYAIIPKAIDDDGNAK	

Abbildung 4.2: Vergleich der Hämoglobin-Sequenzen von *C. tentans*, *C. thummi* und *D. melanogaster*

Um orthologe Hämoglobine bei den Arten *C. tentans* und *C. thummi* zu identifizieren, wurden die Aminosäure-Sequenzen von *C. thummi* (zusammengefaßt in Goodman et al. (1983)) mit den Hämoglobin-Varianten von *C. tentans* verglichen. Die Sequenz des Hämoglobins CTE U liegt nur teilweise vor. Voraussichtlich fehlen 66 Aminosäuren des N-Terminus zur vollständigen Sequenz. Als entfernt verwandte Referenzgruppe („outgroup“) wurde ein Hämoglobin von *Drosophila melanogaster* (Dme glob1) in den Vergleich einbezogen.

Der Vergleich der Aminosäure-Sequenzen zeigt, daß die Distanzwerte der bekanntermaßen orthologen Gene von *C. tentans* und *C. thummi* zwischen 0,08 und 0,25 liegen, während die Werte für nicht-orthologe Gene von 0,31 bis 0,62 variieren. Die geringen Sequenz-Distanzen der Hämoglobine III und IV (0,16 und 0,19) spiegeln die nahe Verwandtschaft dieser nicht-orthologen Hämoglobin-Komponenten von *C. tentans* und *C. thummi* wider. Dennoch ist es möglich, sie anhand der Distanzwerte von 0,08 und 0,13 eindeutig ihren orthologen Partnern zuzuordnen. Die ermittelten Distanzen beweisen, daß im Rahmen der vorliegenden Arbeit die Hämoglobin-Gene für die Varianten CTE VIII und CTE I identifiziert wurden. Die im Interspezies-Vergleich ermittelten Werte von 0,12 und 0,25 sind signifikant für orthologe Gene. Auch das Hämoglobin CTE U weist mit einem Distanzwert von 0,16 eine sehr hohe Sequenz-Übereinstimmung zu der Variante CTT VI auf. Es ist daher wahrscheinlich, daß auch das Hämoglobin VI aus *C. tentans* isoliert wurde. Um mit Sicherheit von der Orthologie dieser beiden Varianten ausgehen zu können, ist es jedoch zunächst erforderlich, die fehlende Sequenz im 5'-Bereich des Gens von *C. tentans* zu ermitteln.

### ***Die phylogenetische Rekonstruktionen der Sequenz-Evolution***

*Basierend auf den Distanzwerten können die Verwandtschaftsverhältnisse ermittelt und in Form phylogenetischer Bäume dargestellt werden. Dies erfolgte parallel durch zwei Methoden, denen unterschiedliche Prinzipien zugrunde liegen:*

- *Der zur Rekonstruktion der Verwandtschaftsverhältnisse verwendete „neighbour-joining“ Algorithmus (Saitou und Nei, 1987) wählt unter allen möglichen Gruppierungen den Baum mit den kürzesten Astlängen (geringsten Distanzen), als den wahrscheinlichsten aus.*
- *Ein anderes Prinzip zur Konstruktion phylogenetischer Bäume wird durch die „maximum parsimony“-Methode genutzt (Eck und Dayhoff, 1966; Fitch, 1977). Auch hier wird zunächst ein Alignment aller Sequenzen angefertigt. Der Algorithmus wählt nun unter allen möglichen Bäumen denjenigen aus, der die vorhandenen Sequenz-Unterschiede durch die geringste Anzahl evolutionärer Ereignisse erklärt. Vorausgesetzt wird demnach, daß die sparsamste Erklärung die tatsächlichen Verhältnisse am wahrscheinlichsten widerspiegelt. Mit der MP-Methode ist es daher durchaus möglich, mehrere ähnlich sparsame Bäume aus einem Datensatz zu erhalten.*
- *Vorteile und Nachteile beider Methoden: Im Gegensatz zu den Methoden, die auf einer Distanzmatrix beruhen, werden bei der Berechnung von MP-Bäumen die Art des Austausches mit berücksichtigt und nur sogenannte „informative Stellen“, die einen der*

möglichen Bäume favorisieren, in die Kalkulation einbezogen. Für eine sinnvolle Anwendung dieser Methode ist daher eine ausreichende Anzahl solcher „informativen Stellen“ erforderlich. Ein Nachteil dieser Methode ist jedoch, daß Basensubstitutionen, die zu einem identischen Ergebnis in zwei Sequenzen führen, sogenannte Homoplasien, nicht identifiziert und somit aus der Berechnung ausgeschlossen werden. Hierzu zählen Rückmutationen in den Ausgangszustand, aber auch parallele- sowie konvergente Austausch. Diese verfälschen das Ergebnis in zunehmendem Maße, je länger die Sequenzen getrennt evolvieren und somit Homoplasien anhäufen können. In solchen Fällen sollten auf Distanzen basierende Methoden bevorzugt werden, da diese in der Lage sind, für Homoplasie-Effekte, aber auch für unterschiedliche Evolutionsraten und bevorzugte Substitutionsmuster zu korrigieren.

- Um die Glaubwürdigkeit der erstellten Topologie zu prüfen, wird in der Regel ein statistischer Test, ein sog. Bootstrapping, angewendet (Felsenstein, 1985). Hierbei werden phylogenetische Bäume für zufällig aus dem gesamten Datensatz ausgewählte Sequenz-Positionen erstellt. Unterstützen diese mit statistischer Signifikanz die zu überprüfende Topologie, so erhält die Verzweigung einen entsprechend hohen Wert. In Anlehnung an Hillis und Bull (1993) werden Knotenpunkte mit Bootstrap-Werten über 70 als glaubwürdig eingestuft, Werte über 50 werden stets an den Verzweigungen gezeigt. Zur Überprüfung der hier erstellten NJ- und MP-Bäume wurden je 500-1000 Bootstrap-Replikationen durchgeführt.
- Mit Hilfe einer Referenzgruppe („outgroup“) ist es möglich, die Wurzel (Root) eines phylogenetischen Baums festzulegen. Die „outgroup“ ist gekennzeichnet durch ihre nachgewiesenermaßen entfernte Verwandtschaft zu den zu untersuchenden Sequenzen. Das Hämoglobin von *Drosophila melanogaster* (Burmester und Hankeln, 1999) erfüllt hier diese Aufgabe.

### Die Phylogenetische Rekonstruktion der Hämoglobin Evolution

Für die Hämoglobin-Sequenzen von *C. tentans* und *C. thummi* wurden sowohl NJ- („neighbour-joining“) als auch MP- („maximum parsimony“) Bäume ermittelt. Die resultierenden Topologien sind in Abb. 4.3 dargestellt.

Es fällt auf, daß sich die mit unterschiedlichen Methoden errechneten Bäume in ihrer Topologie weitgehend entsprechen. So werden die bereits auf Grundlage der Sequenz-Distanzen als orthologe Varianten identifizierten Hämoglobine stets zusammen gruppiert. Ihre Zuordnungen werden mit deutlich signifikanten Bootstrap-

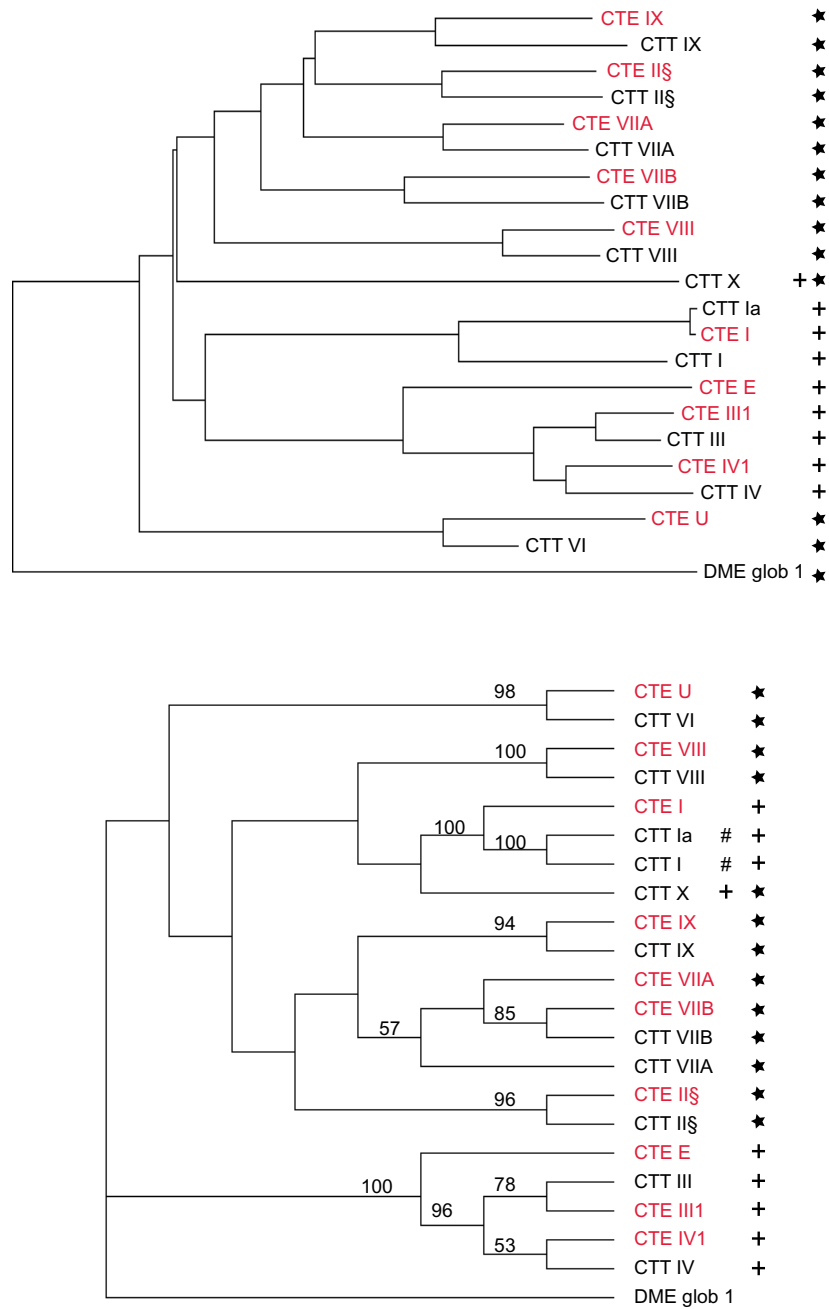


Abbildung 4.3: Phylogenetische Bäume der Hämoglobin-Aminosäure-Sequenzen von *C. tentans* und *C. thummi*

Die von Braunitzers Gruppe ermittelten Hämoglobin-Sequenzen von *C. thummi* wurden mit den abgeleiteten Aminosäure-Sequenzen der Gene von *C. tentans* verglichen (diese Arbeit, in rot dargestellt). Aus den Sequenzen mit den größten Übereinstimmungen wurde ein NJ-Baum (A) sowie ein MP-Baum (B) berechnet. Alle Bootstrap-Werte über 50 sind an den Ästen angegeben.

\* dimere Hämoglobine; + monomere Hämoglobine; # allele Hämoglobin-Varianten.

Werten (78-100) unterstützt. Unter anderem konnten somit in *C. tentans* die bisher dort unbekanntenen Hämoglobin-Gene Cte 1, Cte 8 und Cte 6 identifiziert werden.

Die Abfolge der Genduplikations-Ereignisse, die zeitlich offenbar deutlich vor der Speziation von *C. thummi* und *C. tentans* erfolgt sind, läßt sich nur schwer mit Sicherheit bestimmen. Die entsprechenden Verzweigungen der Bäume werden statistisch nicht sehr gut unterstützt (siehe Abb. 4.3). Sie entsprechen jedoch weitgehend der von Goodman et al. (1983) erstellten Stammbäume für die Hämoglobine von *C. thummi*. Die monomeren Hämoglobine (III, IV, E) sind dabei von den dimeren Varianten getrennt. In dem hier gezeigten MP-Baum mit den wenigsten Nukleotid-Austauschen (NA = 700) gruppiert einzig das monomere Hämoglobin I innerhalb der dimeren Varianten. Dies wurde auch von Braunitzer und Mitarbeitern beobachtet. Schon von diesen Autoren wird dieses Ergebnis mit der Begründung abgelehnt, daß, ausgehend von einem monomeren Hämoglobin-Vorläufer, dimere Varianten während der Evolution vermutlich nur einmal entstanden sind. Mit nur einem Austausch mehr erhält man denn auch einen MP-Baum, der das Hämoglobin I in die Nähe der sonstigen Monomeren plaziert, wie es im übrigen auch bei dem NJ-Baum der Fall ist. Desweiteren schien auch die chromosomale Organisation von „Monomer“- und „Dimer-Genen“ für eine phylogenetische Spaltung vor langer Zeit zu sprechen. Es wurde zunächst angenommen, daß sie getrennt auf unterschiedlichen Chromosomenarmen lokalisiert sind (Schmidt et al., 1988). Tatsächlich befindet sich jedoch das monomere Hämoglobin I im Cluster D zusammen mit den anderen dimeren Hämoglobinen (Kao et al., 1995). Unter dem Gesichtspunkt, daß sich benachbarte Gen-Sequenzen z.B. durch Genkonversion beeinflussen können (s.u.), ist eine Sequenz-Ähnlichkeit der Hämoglobin-Variante I zu dimeren Komponenten nicht unbedingt erstaunlich.

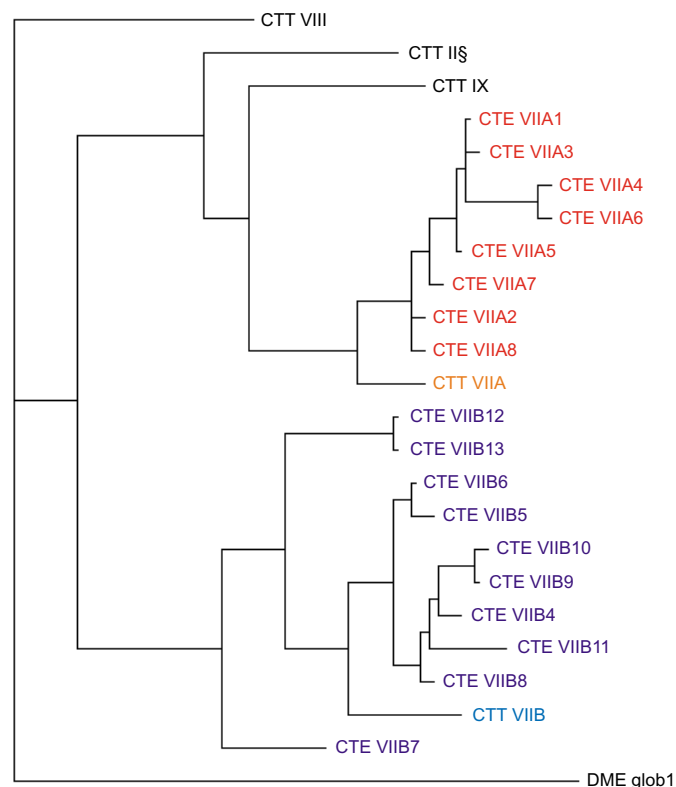
Wie zuvor schon bei Goodman et al. (1983) beschrieben, nimmt das Hämoglobin X in der Topologie beider Bäume eine Sonderstellung zwischen Dimeren und Monomeren ein. Dies steht im Einklang mit seiner Eigenschaft, pH-Wert-abhängig sowohl als Monomer, als auch als Dimer vorzuliegen. Ein orthologes Gen zu der Variante CTT X konnte in *C. tentans* nicht identifiziert werden. Das Hämoglobin X wurde somit bislang weder in *C. tentans* noch in *C. thummi* oder einer anderen untersuchten Spezies als Gen nachgewiesen.

#### **Die Hämoglobin 7A- und 7B-Gen-Subfamilien in *C. tentans***

Der Vergleich der ermittelten Hämoglobin-Sequenzen von *C. tentans* mit den dimeren Protein-Hauptkomponenten von *C. thummi* (Braunitzer und Mitarbeiter; zusammengefaßt in Goodman et al. (1983)) zeigt weiterhin, daß die Hämoglobine VIIA und VIIB in einer Vielzahl nur leicht unterschiedlicher Varianten in *C. tentans* vorliegen. Aufgrund ihrer nahen Sequenz-Verwandtschaft werden diese Varianten als VIIA- bzw. VIIB-Subfamilie klassifiziert. Während die 7B-Gen-Subfamilie bereits in den Spezies *C. thummi* und *C. piger* beschrieben wurde (Hankeln und Schmidt,

1991; Trewitt et al., 1995), war bislang unbekannt, daß auch das Hämoglobin VIIA von mehreren sehr ähnlichen Genen kodiert wird.

Die Rekonstruktion der Verwandtschaftsverhältnisse in einem MP-Baum zeigt die beiden Subfamilien als monophyletische Gruppen (siehe Abb. 4.4). Innerhalb dieser Hämoglobin-Subfamilien weisen die paralogen Varianten häufig eine größere Sequenz-Verwandtschaft zueinander als zu ihren jeweils orthologen Partnern aus *C. thummi* auf. Eine detaillierte Analyse der verwandtschaftlichen Beziehungen innerhalb der Gen-Subfamilien erfolgt im Rahmen des Vergleichs aller paralogen Hämoglobin-Gene von *C. tentans* (siehe Kapitel 4.2.5).



**Abbildung 4.4:** Phylogenetische Klassifizierung dimerer Hämoglobin-Varianten aus *C. tentans* in zwei Subfamilien

Auf Grundlage des Vergleichs dimerer Protein-Komponenten von *C. thummi* (Braunitzer und Mitarbeiter; zusammengefaßt in Goodman et al. (1983)) mit abgeleiteten Aminosäure-Sequenzen der Hämoglobin-Gene von *C. tentans* lassen sich zwei Hämoglobin-Subfamilien in *C. tentans* identifizieren. Aufgrund ihrer nahen Sequenz-Verwandtschaft zu den Varianten VIIA und VIIB von *C. thummi* werden sie als VIIA- und VIIB-Subfamilien bezeichnet (hier rot und blau dargestellt). Die abgeleitete Aminosäure-Sequenz des Globins 1 von *Drosophila melanogaster* (Burmester und Hankeln, 1999) diente bei der Erstellung dieses Mp-Baums als „outgroup“.



#### 4.1.2 Die Klassifizierung der Hämoglobin-Varianten auf Gen-Ebene

Um auch diejenigen Hämoglobin-Gene aus *C. tentans* zu klassifizieren, die nicht zu den von Braunitzer und Mitarbeitern (zusammengefaßt in Goodman et al. (1983)) bei *C. thummi* beschriebenen 11 Hauptkomponenten gehören, wurden die ermittelten Nukleotid-Sequenzen, sowie deren abgeleitete Aminosäure-Sequenzen mit bereits bekannten Hämoglobin-Genen weiterer Chironomiden-Spezies verglichen. Das Alignment aller Nukleotid- und Aminosäure-Sequenzen ist im Anhang gezeigt. Die auf dieser Grundlage erstellten NJ- und MP-Bäume der Aminosäure-Sequenzen sind in Abb. 4.5 dargestellt. Das Hämoglobin von *D. melanogaster* diente wiederum als „outgroup“.

Der NJ-Baum und der MP-Baum entsprechen sich vollständig in der Zuordnung der orthologen Hämoglobine. Die Topologie des Baumes wird größtenteils mit signifikanten Bootstrap-Werten unterstützt. Die bereits bekannten Hämoglobine CTE V, CTE Z, CTE Y und CTE W (Rozynek, Hankeln und Schmidt, unveröffentlicht) werden ihren orthologen Partnern aus *C. thummi* und *C. piger* eindeutig zugeordnet. Die Werte der p-Distanzen sind hierbei typisch für orthologe Hämoglobin-Gene dieser Arten (0,10-0,20). Das bei *C. tentans* beschriebene Hämoglobin-Gen mit dem Arbeitstitel „ORF A“ (Rozynek et al., 1991) weist die sehr geringe Distanz von 0,18 zu dem Hämoglobin XIII von *C. thummi* (Gruhl et al., 1997) auf. Es ist daher wahrscheinlich, daß hier ebenfalls orthologe Gene vorliegen. Zwei Hämoglobine von *C. tentans* werden zu der Variante CTT XII (Gruhl et al., 1997) gruppiert (p-Distanz = 0,23). Sie werden daher als CTE XIII1 und CTE XIII2 bezeichnet. Auch die Variante CTT XI konnte in *C. tentans* identifiziert werden (p-Distanz = 0,20). Somit wurden alle bislang bei *C. thummi* oder der Schwesterart *C. piger* beschriebenen Hämoglobin-Gen-Varianten auch in *C. tentans* nachgewiesen. Zusätzlich wurden in der vorliegenden Arbeit sechs neue Varianten gefunden, die bislang keine orthologen Gene in anderen Arten aufweisen. Sie wurden gemäß der in Hankeln et al. (1998) vorgeschlagenen Nomenklatur-Richtlinien als CTE N, CTE P, CTE Q, CTE R, CTE S und CTE T bezeichnet. Die Abbildung 4.6 zeigt eine Übersicht aller aus *C. tentans* isolierten Hämoglobin-Gene und ihrer orthologen Partnern aus *C. thummi*, *C. piger*, *C. pallidivittatus* und *C. melanotus*.

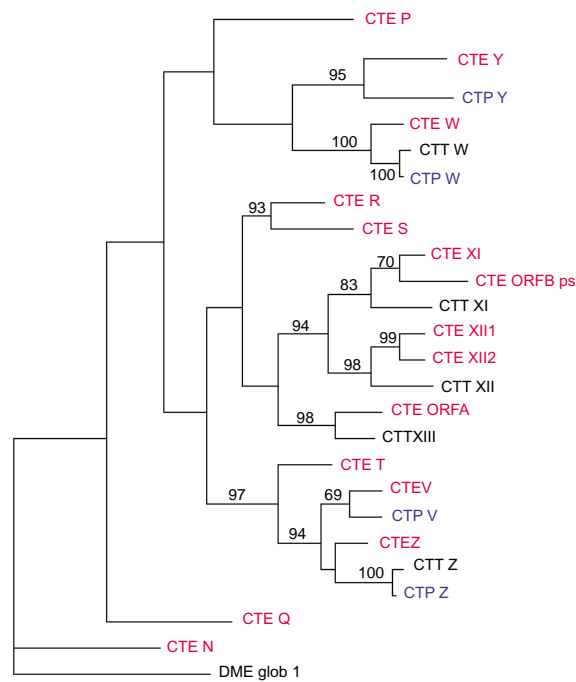
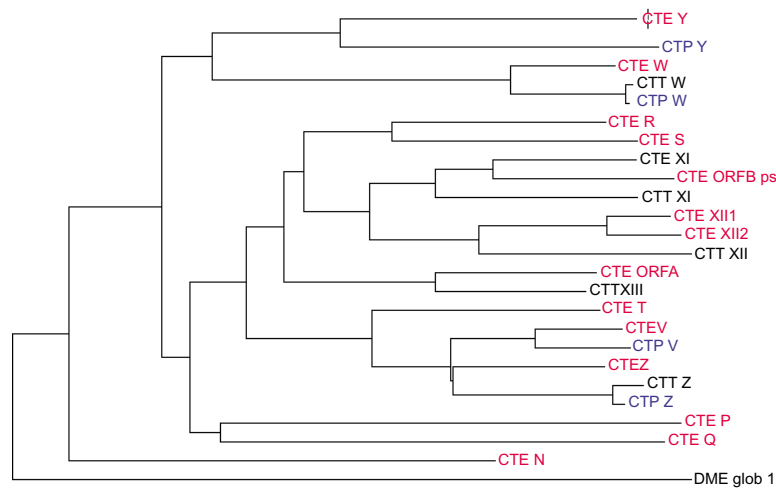


Abbildung 4.5: Verwandtschaftsbeziehungen der Hämoglobin-Aminosäure-Sequenzen von *C. tentans*, *C. thummi* und *C. piger*

Es wurden die Hämoglobin-Varianten in den Vergleich einbezogen, die nicht zu den von Braunitzer ermittelten Hauptkomponenten gehören. Basierend auf den p-Distanzen wurde der NJ-Baum (A) berechnet. Der MP-Baum mit der geringsten Anzahl an Austauschen (872) ist in B dargestellt. Die Ergebnisse des Bootstrap-Tests sind an den Ästen vermerkt, er wurde mit 1000 Replikationen durchgeführt. Die Hämoglobin-Varianten von *C. tentans* sind in rot, die von *C. thummi* in schwarz und die von *C. piger* in blau gezeigt. Das Pseudogen Ctt V sowie das nur unvollständig bekannte Hämoglobin-Gen Ctt Y wurden nicht berücksichtigt.

4.1 Die Klassifizierung der Hämoglobin-Varianten von *C. tentans*

	<i>C. t. thummi</i>	<i>C. tentans</i>	<i>C. t. thummi</i>	<i>C. t. piger</i>	
<b>Braunitzer Varianten: Proteine bekannt</b> s. Goodman et al., 1983	CTT VIII	Cte 8			<b>Gen-Gruppe</b>  <b>D</b>
	CTT II	Cte 2B	Ctt 2B		
	CTT IX	Cte 9	Ctt 9		
	CTT I/CTT IA	Cte 1	Ctt 1/1a		
	CTT VI	Cte U/Cte 6	Ctt 6		
	CTT VIIA	Cte 7A1	Ctt 7A		
		Cte 7A2			
		Cte 7A3	Ctt 13ps		
		Cte 7A4			
		Cte 7A5			
		Cte 7A6			
		Cte 7A7			
		Cte 7A8			
	CTT VIIB	Cte 7B12			
		Cte 7B13			
		Cte 7B7	Ctt 7B7	Ctp 7B7	
		Cte 7B6	Ctt 7B6	Ctp 7B6	
		Cte 7B5	Ctt 7B5	Ctp 7B5	
		Cte 7B10	Ctt 7B10	Ctp 7B8/10	
		Cte 7B9	Ctt 7B9	Ctp 7B4	
	Cte 7B8	Ctt 7B8			
	Cte 7B4	Ctt 7B4			
	Cte 7B11				
<b>Gene bekannt</b>		Cte ORFA			
		Cte ORFBps			
		Cte Y	Ctt Y	Ctp Y	
		Cte W	Ctt W	Ctp W	
		Cte V	Ctt V	Ctp V	
		Cte Z	Ctt Z	Ctp Z	
		Cte 11	Ctt 11		
		Cte 12	Ctt 12		
		Cte 12-1			
<b>Neu</b>		Cte N			
		Cte P			
		Cte Q			
		Cte R			
		Cte S			
		Cte T			
	CTT X ?	?	?	?	
	CTT III	Cte E	Ctt E		
	CTT IV	Cte 3-1	Ctt 3-1		
		Cte 4-1	Ctt 4-1		
		Cte 3-2	Ctt 3-2		
		Cte 4-2			
				<b>E</b>	

Abbildung 4.6 Legende siehe folgende Seite

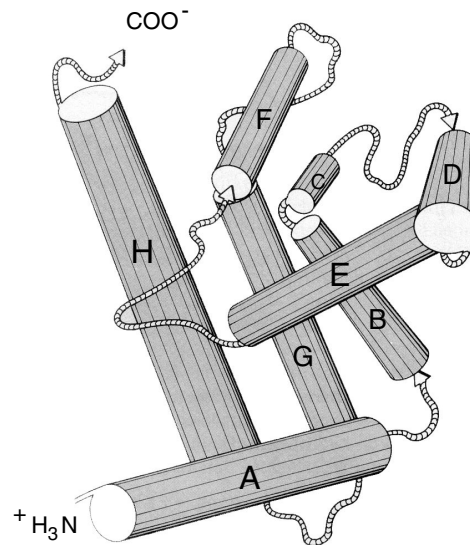
**Abbildung 4.6: Die Hämoglobin-Varianten von *C. tentans* und anderen Chironomiden-Spezies im Vergleich**

Dargestellt sind die 11 auf Protein-Ebene sequenzierten Hämoglobine (Goodman et al., 1983), sowie alle bislang isolierten Hämoglobin-Gene der Chironomus-Arten *C. tentans* (diese Arbeit; Rozynek et al. (1991)), *C. thummi* (Antoine et al., 1987; Kao et al., 1994; Kao und Bergtrom, 1995; Kao et al., 1995; Gruhl et al., 1997; Trewitt et al., 1995) und *C. piger* (Hankeln et al., 1988; Rozynek et al., 1989). Es zeigt sich, daß alle bekannten Hämoglobin-Gene der übrigen Chironomiden auch in *C. tentans* vorhanden sind. Nur die Gene für die Protein-Varianten CTT X und CTT IIIA konnte bislang in keiner Chironomus-Art nachgewiesen werden. Die überwiegende Mehrheit der für *C. thummi* auf Protein-Ebene beschriebenen Hämoglobine wird von Genen kodiert, die in nur einer Kopie vorliegen. Für die zwei Varianten VIIA und VIIB konnte hingegen eine Vielzahl sehr ähnlicher Gene beschrieben werden. Sie werden als Gen-Subfamilien bezeichnet. In *C. tentans* konnten sechs gänzlich neue Varianten identifiziert werden, die bislang weder auf Protein- noch auf Gen-Ebene in anderen Spezies nachgewiesen wurden.

## 4.2 Die Primärstruktur und Funktion der Hämoglobine von Chironomiden

Die dreidimensionale Gestalt von Myoglobin und Hämoglobinen ist entscheidend für die biologische Funktion dieser Moleküle. Myoglobin war das erste Protein, dessen Struktur man bis in den atomaren Bereich aufklärte. Im Jahre 1957 konnte Kendrew mittels Röntgenstrukturanalyse erstmals ein dreidimensionales Bild des Proteinmoleküls erstellen. Ungefähr 75 % der Hauptkette besitzt die Konformation einer  $\alpha$ -Helix. Es gibt im Molekül acht helikale Hauptsegmente (Helices A-H), die allesamt rechtsgängig sind (siehe Abb. 4.7). Dieses Strukturprinzip, die Globinfaltung, ist den Myoglobinen und Hämoglobinen aller Vertebraten gemeinsam und läßt sich sogar auf die Hämoglobin-Varianten von Invertebraten und Pflanzen ausweiten (Review: Hardison (1996)).

Die Globin-typische Faltung ist das fundamentale Bauprinzip der Natur für sauerstofftransportierende Moleküle: Es plaziert die Hämgruppe in eine Umgebung, die sie befähigt, Sauerstoff reversibel zu binden. Gewisse Hinweise auf die physiologische Funktion der verschiedenen Hämoglobin-Varianten können durch Vergleiche der von den Genen abgeleiteten Aminosäure-Sequenzen erhalten werden. Beispielsweise können die Auswirkungen von Aminosäure-Austauschen auf die Quartärstruktur der Proteine abgeschätzt werden. Hierbei erweist es sich als vorteilhaft, daß in der Art *C. thummi* verschiedene Globin-Komponenten bereits detailliert auf Protein-Ebene bis hin zur 3D-Struktur untersucht worden sind. Von der Gruppe um Braunitzer konnten 11 verschiedene Komponenten identifiziert und auf Aminosäure-Ebene



**Abbildung 4.7:** Die dreidimensionale, räumliche Struktur des Myoglobins in Diagrammform

Die Helix-Bereiche sind durch die Buchstaben A-H bezeichnet, vom Amino-Ende links unten beginnend (aus Karlson (1994), verändert).

sequenziert werden (Braun et al., 1968). Vier dieser Hämoglobine (CTT I, CTT III/IIIA, CTT IV) sind Monomere. Eine Variante, das CTT X, liegt ebenfalls als Monomer vor, bildet aber abhängig vom pH-Wert auch Homodimere. Die Hämoglobine CTT II $\beta$ , CTT VI, CTT VIIA, CTT VIIB, CTT VIII und das CTT IX bilden unter physiologischen Bedingungen obligat Homodimere.

Ungeachtet der im weiteren diskutierten Unterschiede in der Primärstruktur ist die Faltung der Chironomiden-Hämoglobine in Globin-typischer Form erhalten (Aronson et al., 1994). Die acht  $\alpha$ -Helices A bis H sind in ihrer charakteristischen Abfolge vorhanden. Dies kann auch für die Hämoglobin-Varianten von *C. tentans* vorausgesetzt werden, da in der Regel keine der  $\alpha$ -Helices einen Prolinrest (Helixbrecher) enthält. Die Aminosäure Prolin findet sich bevorzugt direkt am Ende der helikalen Faltungen. So endet beispielsweise bei 36 von 40 untersuchten Hämoglobinen von *C. tentans* die Helix C an einem Prolin. Bei den monomeren Hämoglobin-Varianten E, III und IV sowie bei dem dimeren Hämoglobin VI (CTE U) begrenzt Prolin die Helix F. Eine Ausnahme bildet die Hämoglobin-Variante Cte I, die interessanterweise innerhalb der Helix H einen Prolinrest aufweist. Welche Auswirkungen dies auf die Faltung der Polypeptidkette hat und ob diese die Funktion dieser Hämoglobin-Variante beeinflussen, ist auf Grundlage der vorliegenden Daten nicht zu beurteilen. Die Analyse der Röntgenstruktur sowie physiologische Untersuchun-

gen, z.B. der Sauerstoff-Bindungseigenschaften, könnten hier dazu dienen, mögliche funktionelle Unterschiede zu anderen Hämoglobin-Komponenten aufzudecken.

#### 4.2.1 Die „invarianten“ und „semi-invarianten“ Aminosäure-Positionen der Chironomiden Hämoglobine

Um Unterschiede in der Aminosäure-Sequenz von Hämoglobin-Varianten bei Chironomiden bewerten zu können, ist es außerordentlich hilfreich, daß die Röntgenstruktur der monomeren Hämoglobin-Komponente III aus der Spezies *C. thummi* (Steigemann und Weber, 1979) bekannt ist. Eine Reihe funktionell wichtiger Aminosäure-Positionen konnten so festgelegt werden:

1. Positionen, die an der Ausbildung von Hämkontakten bzw. der Sauerstoffbindungsstelle beteiligt sind.
2. Aminosäurereste, die durch interhelikale Kontakte oder durch Ausrichtung ins Innere des Moleküls zur Stabilisierung der Tertiärstruktur beitragen.

Auf der Grundlage dieser Daten konnten von Goodman und Mitarbeitern (1983) durch Sequenz-Vergleiche besonders konservierte Aminosäure-Positionen innerhalb der Hämoglobine von *C. thummi* ermittelt werden. Je nach Konservierungsgrad wurden diese in invariante und semi-invariante Positionen klassifiziert. Die Tabelle 4.2 zeigt die als invariant eingestuft Aminosäure-Positionen mit ihren möglichen Funktionen und ermöglicht dabei den Vergleich von 51 verschiedenen Hämoglobin-Komponenten und sechs allelen Gen-Varianten aus den beiden Spezies *C. thummi* und *C. tentans* (siehe auch das Aminosäure-Alignment in Abb. 4.8). Die in der vorliegenden Arbeit ermittelten Hämoglobin-Varianten sind hervorgehoben.

Der detaillierte Vergleich der Aminosäure-Sequenzen der Hämoglobin-Komponenten von *C. thummi* (nach Goodman et al. (1983)) mit den abgeleiteten Aminosäure-Sequenzen der Hämoglobin-Gene von *C. tentans* macht dabei deutlich, daß auch bei *C. tentans* diese funktionell bedeutenden Bereiche im wesentlichen konserviert vorliegen. So sind die Aminosäure-Positionen E7 (His) und F8 (His) für alle bisher untersuchten und physiologisch aktiven Hämoglobine absolut invariant. Während das proximale Histidin (Position F8) kovalent an das Eisenatom der Hämgruppe gebunden ist, steht das distale Histidin (Position E7) über Wasserstoffbrücken-Bindung mit dem Eisenporphyrin in Kontakt. Das distale Histidin ist für die Funktion des Hämoglobins als Sauerstofftransporter von entscheidender Bedeutung, da es die Affinität des Proteins für CO stark vermindert. Ein isoliertes Häm in Lösung bindet CO 25000 mal stärker als O<sub>2</sub>. Unter physiologischen Bedingungen ist die Bindungsaffinität für CO beim Hämoglobin dagegen nur 200 mal höher als für O<sub>2</sub>. Durch die sterische Behinderung einer CO-Bindung wird eine Vergiftung des Hämoglobins verhindert, die schon alleine durch das beim Hämabbau endogen entstehende CO verursacht werden würde (zusammengefaßt in Stryer (1997)). Eine

Variation der Aminosäuren an diesen Positionen kann daher nicht toleriert werden, sie hätte den Funktionsverlust der betroffenen Hämoglobin-Variante als Sauerstoff-Transportmolekül zur Folge.

Jedoch nicht alle als invariant eingestuften Aminosäure-Positionen weisen auch bei *C. tentans* funktionell gleichwertige Aminosäure-Reste auf. Ausnahmen von dem von Goodman und Mitarbeitern (1983) aufgestellten Schema bilden hierbei die Hämoglobin-Varianten CTE P, CTE Q, CTE S, CTE XII und das monomere Hämoglobin CTE I. Sie weisen je eine, zwei (CTE S) oder auch drei (CTE Q) Abweichungen von der konservierten Aminosäure-Sequenz auf. Nur bei einem einzigen der insgesamt acht Austausch ist eine Aminosäure gegen eine funktionell gleichwertige ersetzt (CTE I, Position G15, Tyr (Trp)). Dies ist umso erstaunlicher, als es sich bei den hier untersuchten Positionen um die vermeintlich invarianten und somit für die Struktur und Funktion der Hämoglobine relevanten Aminosäuren handeln sollte. Um zuverlässige Aussagen über Funktionsveränderungen durch nicht konservative Aminosäure-Austausche treffen zu können, sind neben detaillierten physiologischen besonders vergleichende strukturelle Analysen der einzelnen Komponenten notwendig. Besonders die Variante CTE Q, die bei weitem die auffälligsten Abweichungen an den konservierten Aminosäure-Positionen aufweist, ist ein vielversprechender Kandidat für weiterführende Untersuchungen (z.B. Aufklärung der Röntgenstruktur, aber auch der Sauerstoff-Affinität und der Kooperativität der Sauerstoffbindung).

Die Tabelle 4.3 zeigt die „semi-invarianten“ und die „variablen“ Aminosäure-Positionen, für die nach Goodman und Mitarbeitern (1983) zwei bzw. drei oder mehrere Alternativen bestehen. Dabei fällt auf, daß die meisten Substitutionen in den Hämoglobin-Varianten von *C. tentans* konservativ sind. Man kann vermuten, daß derartige Substitutionen, wenn überhaupt, nur geringe Unterschiede in der Funktion bewirken.

## 4 Diskussion

Helix/ AS-Position	invariante Aminosäure	Hb-Proteine mit alternativen Aminosäuren		Kontakte und Funktion
B5	Ile (I); [hydrophob]			B/EG- Hämkontakt
B10	Phe (F); [hydrophob]			B/DE
C4	Ile (I); [hydrophob]			C/G- Hämkontakt
CD1	Phe (F); [hydrophob]			C/E- Hämkontakt
CD4	Phe (F); [hydrophob]			Hämtasche?
CD6	Gly (G)	<b>CTE S</b>	Arg (R) ; [polar]	β-Krümmung?
CD7	Lys (K); [+]			Hämtasche?
D1	Asp (D); [-]	<b>CTE P</b>	Asn (N); [polar]	stabilisiert Helix D?
		<b>CTE Q</b>		
		<b>CTE S</b>		
D6	Lys (K); [+]	<b>CTE Q</b>	Arg (R) ; [polar]	
E4	Phe (F); [hydrophob]			
E7	His (H); [+]			distales Histidin
F8	His (H); [+]			proximales hämgeb. Histidin
FG3	Arg (R); [+]			Hämkontakt
G8	Phe (F); [hydrophob]			G/B- Hämkontakt
G15	Tyr (Y); [polar]	<b>CTE Q</b>	Phe (F); [hydrophob]	G/B
		<b>CTE I</b>	Trp (W); [polar]	
H7	Trp (W); [polar]	<b>CTE XII</b>	Phe (F); [hydrophob]	H/AB

**Tabelle 4.2:** Invariante Aminosäure-Positionen der beschriebenen Hämoglobine von *C. tentans* und *C. thummi*

Die überwiegende Mehrzahl der untersuchten *C. tentans*-Hämoglobin-Varianten besitzen die für *C. thummi* beschriebenen konservierten Aminosäuren (Goodman et al., 1983). Ausnahmen bilden die Hämoglobine CTE P, CTE Q, CTE S, CTE XII und CTE I (Spalte 3). Vorhandene Austausche sind erstaunlicherweise in der Regel nicht konservativ (rot).



#### 4.2 Die Primärstruktur und Funktion der Hämoglobine von Chironomiden

Helix	Aminosäure	Hb-Proteine mit alternativen Aminosäuren		Kontakte und Funktion
A5	Ala (A); [hydrophob]	<b>CTE R</b>	<b>Gly (G)</b>	A/E-Hämkontakt
		<b>CTE XII</b>	<b>Thr (T); [polar]</b>	
		CTT X	Val (V); [hydrophob]	
		<b>CTE I</b>	Ile (I); [hydrophob]	
		CTT I/IA		
		CTT III		
A12	Trp (W); [polar]	CTE E	Tyr (Y); [polar]	A/E
		CTT III	<b>Phe (F); [hydrophob]</b>	
		CTT IV		
A15	Val (V); [hydrophob]	CTE II $\beta$	Ala (A); [hydrophob]	?
		CTE Y	Ile (I); [hydrophob]	
		CTE Z		
B3	Val (V); [hydrophob]	<b>CTE P</b>	Asp (D); [-]	B/DE
		<b>CTE Q</b>	Ala (A); [hydrophob]	
		<b>CTE N</b>	Ile (I); [hydrophob]	
		CTE E	<b>Thr (T); [polar]</b>	
B7	Ala (A); [hydrophob]	<b>CTE P</b>	<b>Tyr (Y); [polar]</b>	B/DE
		<b>CTE Q</b>		
		<b>CTE R</b>		
		<b>CTE S</b>		
		<b>CTE T</b>		
		CTE Y		
		CTE W		
		CTE Z		
		<b>CTE VIIB11</b>		
		<b>CTE N</b>		
		CTT II $\beta$		
		CTE W		
		CTE Z		
		CTT X		
		CTT I/IA		
		CTE E		
CTT III				
CTTIV				
CD5	Ala (A); [hydrophob]	<b>CTE R</b>	<b>Ser (S); [polar]</b>	Hämtasche?
		<b>CTE P</b>	Val (V); [hydrophob]	
		<b>CTE Q</b>		
		CTE II $\beta$		
		CTE E		

#### 4 Diskussion

Helix	Aminosäure	Hb-Proteine mit alternativen Aminosäuren		Kontakte und Funktion
D2	Leu (L); [hydrophob]	<b>CTE XI</b>	<b>Val (V); [hydrophob]</b>	D/B
		CTE ORF A		
		CTE ORF B		
		CTEVIIA		
		CTE Y		
		CTT VIIA		
E1	Thr (T); [polar]	<b>CTE P</b>	Asn (N); [polar]	E/D
		CTE E	<b>Gly (G)</b>	
		<b>CTE Q</b>	Ser (S); [polar]	
		<b>CTE U</b>		
		<b>CTE XII</b>		
		CTT IV		
E10	Arg (R); [+]	CTT IV	Lys (K); [+]	E-Hämkontakt
E11	Ile (I); [hydrophob]	CTT I/IA	Val (V); [hydrophob]	E/BC-Hämkontakt
E12	Val (V); [hydrophob]	CTE W	Ala (A); [hydrophob]	E/B
		CTT X	Ile (I); [hydrophob]	
E14	Phe (F); [hydrophob]	<b>CTE P</b>	Ala (A); [hydrophob]	E-Hämkontakt
		<b>CTE Q</b>	Ile (I); [hydrophob]	
		CTT I/IA	Leu (L); [hydrophob]	
EF4	Asn (N); [polar]	<b>CTE U</b>	Thr (T); [polar]	?
		<b>CTE XI</b>	Ser (S); [polar]	
		<b>CTE N</b>		
		<b>CTE XII2</b>	<b>Gly (G)</b>	
		CTE ORF B		
F4	Leu (L); [hydrophob]	<b>CTE IIB</b>	Met (M); [hydrophob]	F/H-Hämkontakt
		<b>CTE VIIB11</b>		
		CTT IIB		
		CTT I/IA	Phe (F); [hydrophob]	
		CTT III		
		CTT IV		
		CTT X		
FG5	Ile (I); [hydrophob]	CTT IV	Val (V); [hydrophob]	Hämkontakt
		CTT VIIB		

## 4.2 Die Primärstruktur und Funktion der Hämoglobine von Chironomiden

Helix	Aminosäure	Hb-Proteine mit alternativen Aminosäuren		Kontakte und Funktion
FG5	Ile (I); [hydrophob]	<b>CTE N</b>	Val (V); [hydrophob]	Hämkontakt
		<b>CTE U</b>		
		<b>CTE T</b>		
		CTE V		
		CTE W		
		CTE Z		
		<b>CTE VIIB</b>		
		CTE E		
		CTT III		
		<b>CTE P</b>	Ala (A); [hydrophob]	
		<b>CTE I</b>		
		CTT I		
	CTT IA	Thr (T); [polar]		
G4	Gln (Q); [polar]	<b>CTE Q</b>	<b>Phe (F); [hydrophob]</b>	G/B-Hämkontakt
		<b>CTE N</b>	<b>Glu (E) ); [-]</b>	
		CTE Y		
		CTE W	<b>Leu (L); [hydrophob]</b>	
G5	Phe (F); [hydrophob]	CTTI/IA	Leu (L); [hydrophob]	G/H-Hämkontakt
		CTT III		
		CTT IV		

Tabelle 4.3: „Semi-invariante“ Aminosäure-Positionen der beschriebenen Hämoglobine von *C. tentans* und *C. thummi*

„Semi-invariante“ Aminosäure-Positionen (zwei Alternativen), bzw. variable Aminosäure-Positionen mit drei oder mehreren Alternativen bei den Hämoglobinen von *C. tentans* und *C. thummi* (nach Goodman et al. (1983)). Nicht-konservative Aminosäure-Substitutionen sind rot dargestellt. Die in dieser Arbeit ermittelten Hämoglobin-Varianten sind hervorgehoben.

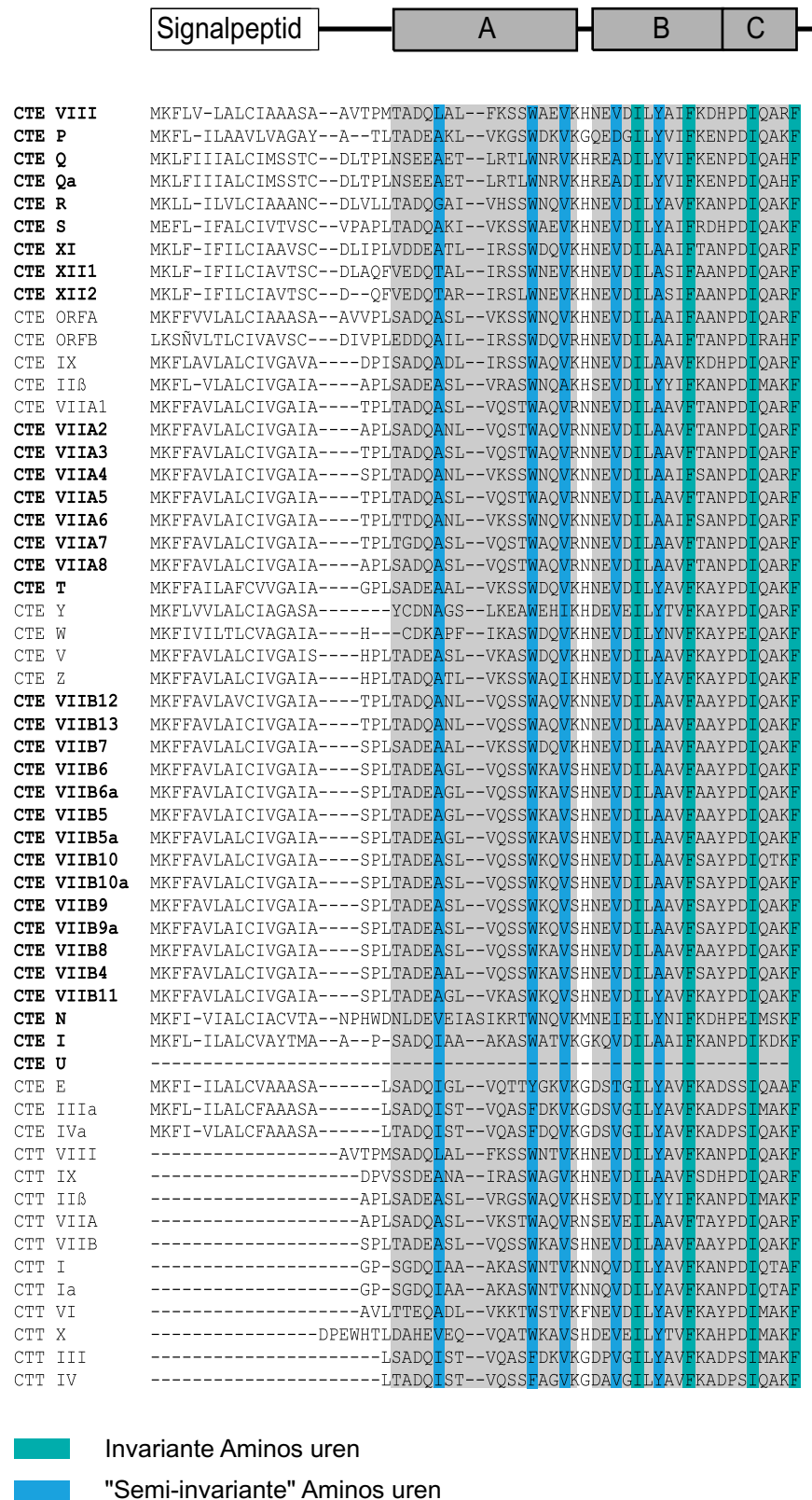


Abbildung 4.8 Legende siehe Seite 103

## 4.2 Die Primärstruktur und Funktion der Hämoglobine von Chironomiden

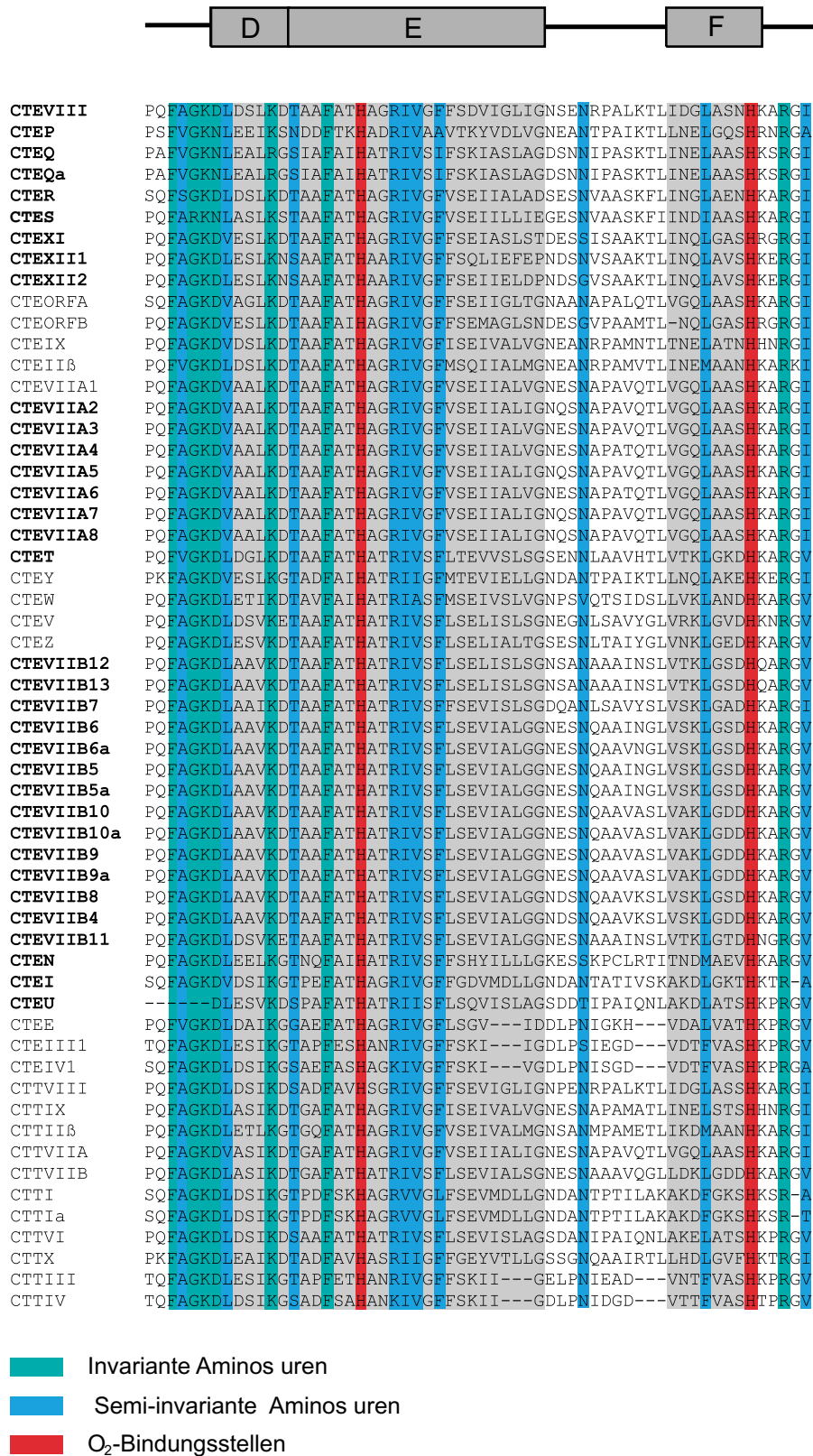


Abbildung 4.8 Legende siehe Seite 103

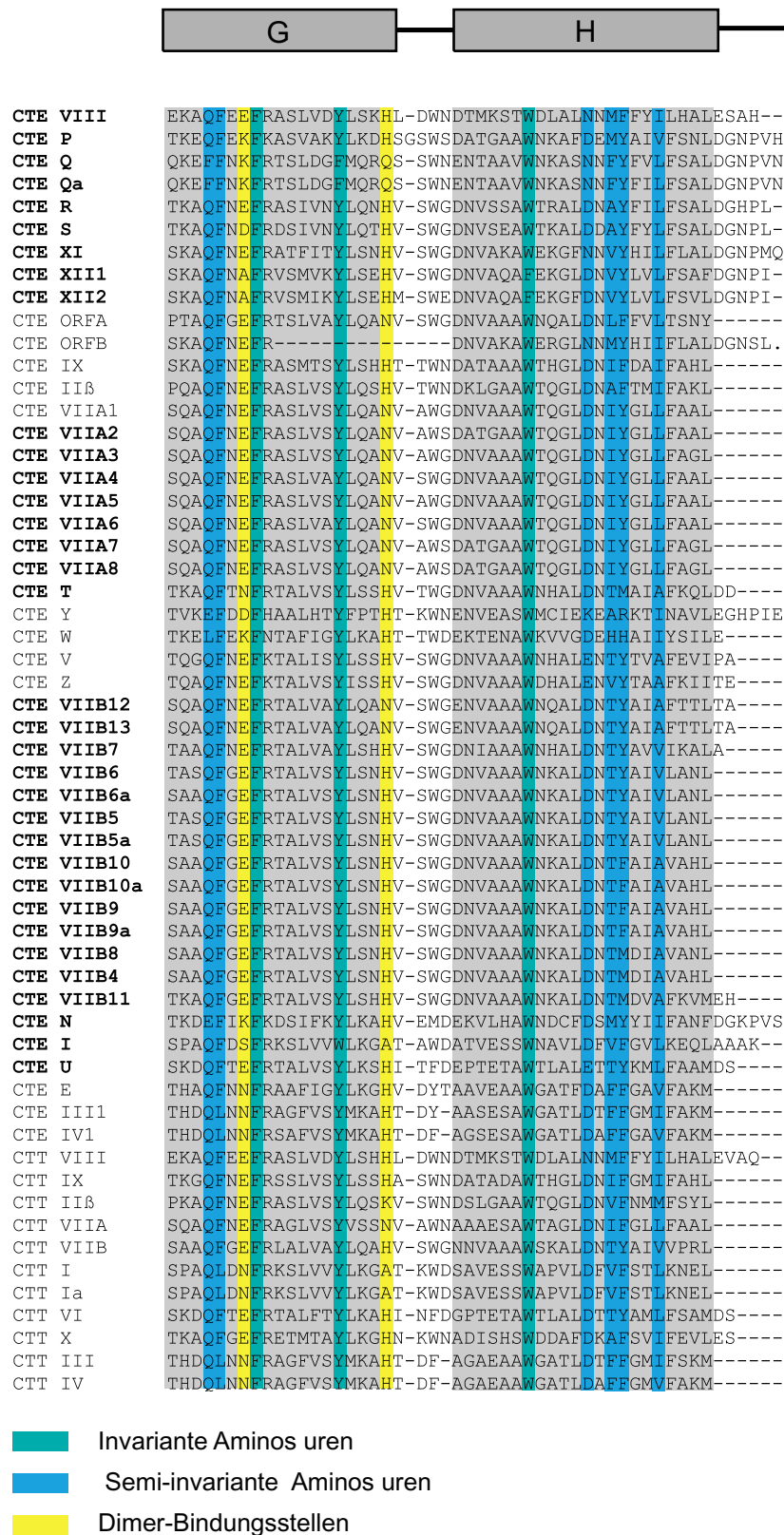


Abbildung 4.8 Legende siehe Seite 103

**Abbildung 4.8: Alignment der Aminosäure-Sequenzen der Komponenten von *C. tentans* und *C. thummi***

Den Aminosäure-Sequenzen von *C. thummi* fehlt das Signalpeptid, da sie auf Proteinebene ermittelt wurden (nach Goodman et al. (1983)). Farbig unterlegt sind die invarianten Aminosäuren, die „semi-invarianten“ Aminosäuren, für deren Position es zwei bis vier Alternativen gibt und die O<sub>2</sub>- bzw. Dimer-Bindungsstellen (nach Steigemann und Weber (1979)). Die  $\alpha$ -Helices A-H sind grau unterlegt.

### 4.2.2 Die Quartärstruktur der Hämoglobine von *C. thummi* und *C. tentans*

Für die Ausbildung der Quartärstruktur bei Chironomiden-Hämoglobinen sind vermutlich ionische Wechselwirkungen definierter Aminosäuren verantwortlich (Behlke et al., 1972). Durch Vergleiche der Primärstruktur verschiedener monomerer und dimerer Komponenten und durch selektive Modifikation einzelner Aminosäuren konnte gezeigt werden, daß die Aminosäure-Positionen G7 und G19 von entscheidender Bedeutung für die Quartärstruktur der Proteine sind (Kleinschmidt et al., 1981; Aschauer et al., 1982; Behlke et al., 1972). Bei der Homodimerbildung der Komponenten CTT IIB, CTT VI, CTT VIIA, CTT VIIB, CTT IX und CTT X in *C. thummi* interagieren die alternativ vorkommenden Aminosäuren Histidin (H), Lysin (K) oder Asparagin (N) an Position G19 mit der Glutaminsäure (E) an G7 (vgl. Tab.4.4). Den monomeren Komponenten der Art *C. thummi* (CTT I/IA, CTT III und CTT IV) fehlt die Glutaminsäure an G7. Sie weisen die polare Aminosäure Asparagin (N) an Position G7 auf und an G19 das konservierte Histidin (vgl. Tab. 4.4). Phylogenetische Untersuchungen der Protein-Varianten zeigten, daß dimere Komponenten, überdies auch aufgrund ihrer Aminosäure-Sequenzen, insgesamt von monomeren Komponenten unterschieden werden können (Goodman et al., 1983). Bei den in dieser Arbeit in *C. tentans* identifizierten Hämoglobin-Genen läßt sich ausschließlich die Variante Cte I aufgrund ihrer Aminosäure-Sequenz den monomeren Hämoglobinen zuordnen. Möglicherweise verhindern hier die Aminosäuren Serin (G7) und Alanin (G19) die Ionenbindung und somit die Dimerbildung. Die Hämoglobine CTE IIB, CTE U, CTE VIIA, CTE VIIB und CTE IX bilden mit großer Wahrscheinlichkeit Homodimere, da sie an den Positionen G7 und G19 ihren orthologen Partnern aus der Spezies *C. thummi* entsprechen (siehe Tab. 4.4).

Trotz ihrer phylogenetischen Verwandtschaft zu den dimeren Hämoglobinen ist die tatsächliche Quartärstruktur der Varianten CTE ORF A, CTE N, CTE P, CTE Q, CTE R, CTE S, CTE T, CTE V, CTE W, CTE Y, CTE Z, CTE XII1 und CTE XII2 bislang unbekannt. Auch die Aminosäuren an den Positionen G7 und G19 geben nur bedingt Aufschluß über eine mögliche Wechselwirkung (siehe Tab. 4.5).

Die Hämoglobine CTE R, CTE XI, CTE V, CTE Z und CTE ORF A entsprechen exakt dem Schema für dimere Hämoglobine. Bei den Hämoglobinen CTE S und CTE Y ist die Glutaminsäure an Position G7 gegen Asparaginsäure ausgetauscht, so daß vermutlich auch hier Homodimere gebildet werden können. Die übrigen Hämoglobin-Varianten von *C. tentans* weisen ungewöhnliche Aminosäure-Kombinationen an den entsprechenden Positionen auf. So finden sich hier beispielsweise bei den Varianten CTE P, CTE W und CTE N zwei positive Ladungen (Lysin an Position G7 und Histidin an Position G19), die eine Interaktion unterbinden dürften.

Die Hämoglobin-Variante CTE T entspricht an den Positionen G7 und G19 exakt den monomeren Hämoglobinen CTT III und CTT IV (siehe Tab. 4.4 und 4.5). Auch das CTE E, dessen Quartärstruktur noch nicht geklärt ist, hat an G7 ein Asparagin und an G19 ein Monomer-typisches Histidin (vgl. Tab. 4.5). Vorausgesetzt, die Positionen G7 und G19 sind tatsächlich ausschließlich für die Zusammenlagerung der Polypeptidketten obligat, wäre die Bildung von Homodimeren der Komponenten CTE T und CTE E nicht möglich.

Es fällt auf, daß an Position G19 Aminosäuren mit positiver Ladung konserviert sind (die einzige Ausnahme bilden die orthologen Hämoglobine CTT VIIA und CTE VIIA mit dem polaren Asparagin an G19). Theoretisch könnten Komponenten mit einer positiven Ladung an Position G19 mit Hämoglobinen, die an Position G7 eine Aminosäure mit negativer Ladung aufweisen Heterodimere bilden. Derartige Heterodimere wurden bislang jedoch nicht nachgewiesen. Prinzipiell stellt sich daher die Frage, ob neben den Positionen G7 und G19 Aminosäuren an anderer Stelle eine Funktion in Hinblick auf eine Multimer-Bildung ausüben können. Desweiteren muß untersucht werden, ob es allein ionische Wechselwirkungen sind, die für eine solche Formation der Proteine ausschlaggebend und ausreichend sind. Man kann vermuten, daß andere konservierte Positionen und verschiedene nicht-kovalente Molekül-Interaktionen für die Ausbildung der Quartärstruktur der Chironomiden-Hämoglobine von Bedeutung sind (Osmulski und Leyko, 1986).



## 4.2 Die Primärstruktur und Funktion der Hämoglobine von Chironomiden

Hb Protein mit bekannter Quartärstruktur	AS-Position G7	AS-Position G19	Quartärstruktur
CTT IIB	Glu (E)	Lys (K)	Dimer
CTE IIB	(-)	(+)	
CTT VI	Glu (E)	His (H)	Dimer
CTE U	(-)	(+)	
CTT VIIA	Glu (E)	Asn (N)	Dimer
CTE VIIA	(-)	(polar)	
CTE VIIB	Glu (E)	His (H)	Dimer
CTE VIIB	(-)	(+)	
CTE VIIB12		Asn (N)	
CTE VIIB13		(polar)	
CTT IX	Glu (E)	His (H)	Dimer
CTE IX	(-)	(+)	
CTT X	Glu (E)	His (H)	Dimer, bzw. Monomer (pH-abhängig)
	(-)	(+)	
CTT I/IA	Asn (N)(polar)	Ala (A) (hydrophob)	Monomer
CTE I	Ser (S) (polar)		
CTT III	Asn (N)	His (H)	Monomer
CTE III	(polar)	(+)	
CTT IV	Asn (N)	His (H)	Monomer
CTE IV	(polar)	(+)	

**Tabelle 4.4:** Konstitution der für eine Dimerisierung wichtigen Aminosäure-Positionen von *C. thummi* (Braunitzer et al., 1968) im Vergleich zu den orthologen Genen aus *C. tentans*

Die Aminosäure-Positionen G7 und G19 sind für eine Dimerbildung essentiell (Kleinschmidt und Braunitzer, 1976).

Hb Protein mit unbekannter Quartärstruktur	AS-Position G7	AS-Position G19	mögliche Quartärstruktur
CTE R	Glu (E) (-)	His (H) (+)	Dimer
CTE XI	Glu (E) (-)	His (H) (+)	Dimer
CTE V	Glu (E) (-)	His (H) (+)	Dimer
CTE Z	Glu (E) (-)	His (H) (+)	Dimer
CTE ORF A	Glu (E) (-)	Asn (N) (polar)	Dimer
CTE S	Asp (D) (-)	His (H) (+)	Dimer
CTE Y	Asp (D) (-)	His (H) (+)	Dimer
CTE W	Lys (K) (+)	His (H) (+)	?
CTE T	Asn (N) (polar)	His (H) (+)	Monomer?
CTE E	Asn (N) (polar)	His (H) (+)	Monomer?
CTE P	Lys (K) (+)	His (H) (+)	?
CTE Q	Lys (K) (+)	Gln (Q) (polar)	?
CTE XIII1	Ala (A) (hydrophob)	His (H) (+)	?
CTE XII2	Ala (A) (hydrophob)	His (H) (+)	?
CTE N	Lys (K) (+)	His (H) (+)	?

Tabelle 4.5: Mögliche Quartärstrukturen der in dieser Arbeit gefundenen, neuartigen *C. tentans*-Hämoglobine

Die Fähigkeit zur Dimerbildung von Chironomiden-Hämoglobinen ist vermutlich abhängig von definierten Aminosäuren. Die unterschiedliche Ladung zweier Aminosäuren der G-Helix (Positionen G7 und G19) spielt eine entscheidende Rolle bei der Bildung von homodimeren Hämoglobinen bei der Art *C. thummi*. Die in *C. tentans* identifizierten Hämoglobine unbekannter Quartärstruktur wurden daher auf das Vorhandensein der für Dimere charakteristischen Aminosäuren hin untersucht. Ungewöhnliche Aminosäure-Kombinationen sind rot hervorgehoben.

### 4.2.3 Die Architektur der Hämoglobin-Gengruppe in *C. tentans*

Die Gengruppe D von *C. tentans* umfaßt mindestens 38 Hämoglobin-Gene, von denen 28 im Rahmen dieser Arbeit identifiziert wurden (Rozynek et al. (1991), Rozynek, Hankeln und Schmidt, unveröffentlicht). Diese liegen sehr dicht gepackt innerhalb des Clusters, wobei durchschnittlich alle 1,9 kb ein Hämoglobin-Gen zu finden ist. Es ist hierbei bemerkenswert, daß benachbarte Gene häufig eine entgegengesetzte Orientierung aufweisen. Die auffällig homogene Abfolge der Hämoglobin-Gene wird nur an einer Stelle durch ein interspergiertes Gen aus der Familie der Glukosetransporter (GLUT) unterbrochen. Auch finden sich zwei Kopien des Retrotransposons CP1 (Hong et al., 1995) innerhalb der Gengruppe. Über das Vorkommen und die Bedeutung von repetitiven Elementen innerhalb der Chironomiden Hämoglobin-Gene sind allerdings nur sehr wenige Details bekannt (siehe z.B. Gruhl et al. (1997)).

Während die kodierenden Abschnitte für Hämoglobin-Gene häufig GC-reiche Nukleotid-Sequenzen aufweisen, zeichnen sich die Intergen-Regionen durch ihren hohen AT-Gehalt aus. Trägt man den errechneten GC-Gehalt über die Gesamtlänge des ermittelten Genomabschnitts auf, so spiegelt dieser sogenannte „GC-Blot“ eindeutig die Lage der einzelnen Hämoglobin-Gene wider (siehe Abb. 4.9).

### 4.2.4 Die Evolution der Hämoglobin-Genstruktur

Seit der Entdeckung der Mosaikstruktur eukaryotischer Gene (Jeffreys und Flavell, 1977) existieren kontroverse Vorstellungen über den Ursprung von Introns und ihrer Bedeutung für die Evolution der Gene (Hurst, 1994; Mattick, 1994; Hurst und McVean, 1996; Logsdon et al., 1995; Long et al., 1995). Die Hämoglobin-Gene spielen bei der Streitfrage um den Ursprung von Introns eine wichtige Rolle. Bei den Hämoglobin-Genen handelt es sich um eine sehr alte Familie, die im gesamten Organismenreich verbreitet ist (Reviews: Vinogradov et al. (1992); Hardison (1996); Moens et al. (1996)). Die vergleichende Analyse der Hämoglobin-Genstruktur bei Vertebraten, Invertebraten, Pflanzen und Protozoen kann daher zur Rekonstruktion der Intron-Evolution genutzt werden. Die Konservierung von Intronpositionen in den homologen Genen verschiedener Taxa gilt dabei als Meßwert für das relative stammesgeschichtliche Alter der Introns.

Erste molekulare Analysen von Hämoglobin-Genen der Art *C. thummi* zeigten, daß die untersuchten Gene keinerlei Introns besitzen (Antoine und Niessing, 1984; Antoine et al., 1987; Trewitt et al., 1987; Hankeln et al., 1988; Hankeln und Schmidt, 1991; Rozynek et al., 1989, 1991). Die meisten der bislang untersuchten Hämoglobin-Gene bei Pflanzen, Vertebraten und auch Invertebraten besitzen hingegen zumeist mindestens zwei Introns. Bei den Hämoglobin-Genen der Chironomiden ging man daher von einem Verlust aller Introns nach der Trennung der Entwicklungslinien von Vertebraten und Insekten aus (Lewin, 1984; Antoine und Niessing, 1984). Es war daher sehr überraschend, als in zwei Genen der Art *C. thummi* (Ctt 2ß und

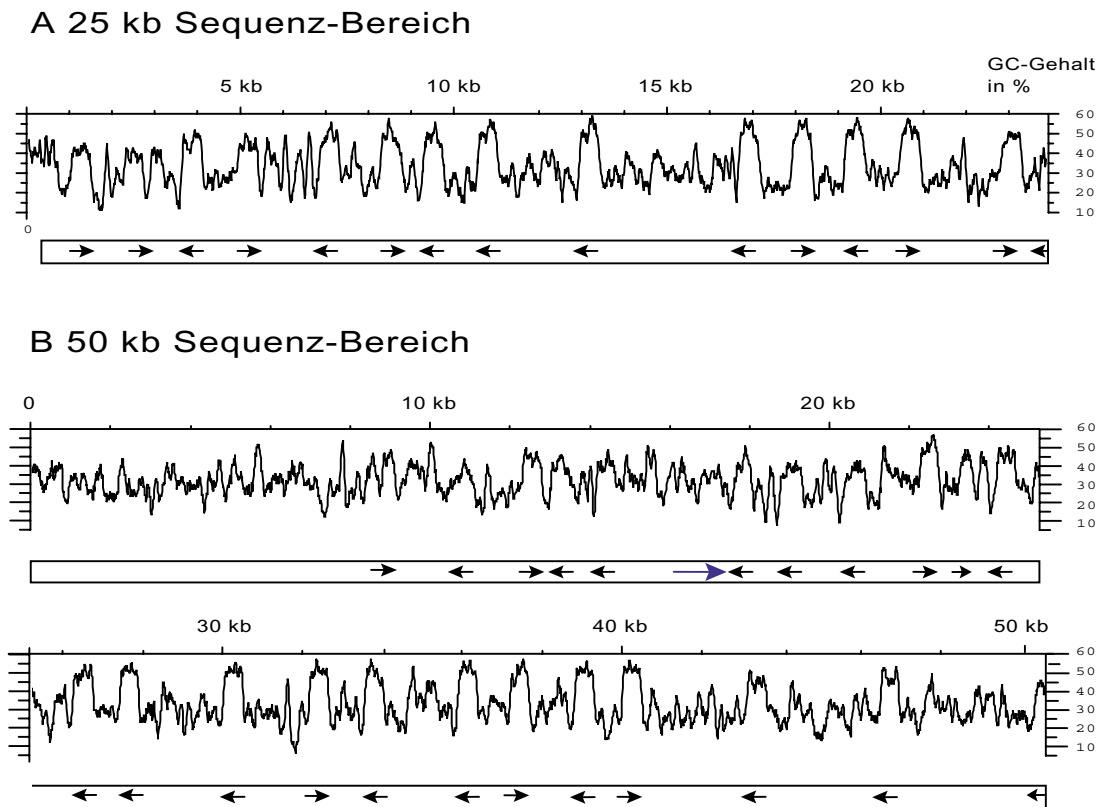


Abbildung 4.9: Der GC-Gehalt des Hämoglobin-Gen-Clusters von *C. tentans*

Der GC-Anteil, hier aufgetragen über die gesamte Länge der Nukleotid-Sequenz, schwankt innerhalb des Hämoglobin-Gen-Clusters sehr stark. Betrachtet man die Positionen der Hämoglobin-Gene innerhalb der Sequenz (hier durch Pfeile gekennzeichnet), so zeigt sich, daß die Gen-Sequenzen stets mit einem erhöhtem GC-Gehalt korreliert sind und sich zumeist deutlich von den AT-reichen Intergen-Regionen abgrenzen. Die Abb. A zeigt den 25 kb umfassenden Sequenz-Contig 2, der 50 kb lange Contig 1 ist in B dargestellt. Die Lage des Glukosetransporter-Gens (GLUT) ist durch einen blauen Pfeil markiert.

Ctt 9) jeweils ein Intron gefunden wurde, das innerhalb der Protein-Helix E liegt (Kao et al., 1994; Gruhl et al., 1997). Insbesondere, da die orthologen Genen in *C. tentans* bereits als intronlos bekannt waren (Rozynek et al., 1991; Hankeln et al., 1997). Um die stammesgeschichtliche Herkunft dieses Introns einzugrenzen, wurden daher orthologe Gene in einer Vielzahl europäischer und australoasiatischer Spezies auf den Besitz dieses Introns hin untersucht (Hankeln et al., 1997). Dabei zeigte sich, daß Introns in den Genen von Chironomiden eher sporadisch zu finden sind und zudem an unterschiedlichen Positionen lokalisiert sind.

Auch in dem Hämoglobin-Gencluster D von *C. tentans* konnten unter 38 Hämoglobin-Genen zwei identifiziert werden, die je ein kurzes Intron aufweisen: das Gen Cte P besitzt ein 58 bp langes Intron in dem für die Helix-E kodierenden Bereich (E12-2), während das 60 bp lange Intron des Gens Cte N den interhelikalen Bereich G-H 2-2 unterbricht. Dies ist umso erstaunlicher, als bislang keine Introns in Hämoglobin-Genen der Art *C. tentans* gefunden wurden. Selbst die Hämoglobin-Gene VIIA, IIB und IX, die in einer Vielzahl von Chironomiden ein „zentrales Intron“ aufweisen, sind in *C. tentans* ohne Intron (Rozynek et al., 1991; Hankeln et al., 1997). Dies wurde häufig als eine Bestätigung für den Verlust aller Introns in der Entwicklungslinie von *C. tentans* bewertet. Ob es sich bei den Introns in den Hämoglobin-Genen Cte N und Cte P um stammesgeschichtlich alte Introns handelt, ist nicht ohne weiteres zu beantworten. Um diese Frage zu klären, wäre es erforderlich, orthologe Gene in anderen Chironomus-Arten zu identifizieren und auf ihre Exon/Intron-Struktur hin zu untersuchen. Zwei Hämoglobin-Gene aus *Glyptotendipes barbipes* (Gba) besitzen in der Tat ebenfalls Introns an den Positionen E12-2 und G-H 2-2 (Hankeln, Ebersberger und Schmidt, unveröffentlicht). Eine Stammbaum-Rekonstruktion basierend auf Aminosäure-Sequenzdaten sollte zeigen, ob und welche der bekannten Hämoglobine von *C. tentans* zu den *Glyptotendipes barbipes* Hämoglobin-Varianten ortholog sind. (Abb. 4.10). Tatsächlich gruppieren die Hämoglobine CTE P und GBA 5.1.1-4, die beide ein Intron an Position E12-2 besitzen, zueinander. Die Nukleotid-Distanz zwischen ihnen beträgt nur 0,102 und entspricht somit der typischen Distanz orthologer Hämoglobine innerhalb der Gattung Chironomus. Die Sequenz-Übereinstimmung von 89,8 % zwischen den Gattungen Chironomus und *Glyptotendipes* ist hingegen unerwartet hoch. Die Sequenzen der Introns selbst geben erwartungsgemäß keinen Hinweis auf eine mögliche Verwandtschaft: sie weisen, abgesehen von den Konsensus-Spleiß-Signalen, keinerlei aussagekräftige Übereinstimmungen auf (siehe Abb. 4.11a). Das zu dem Hämoglobin-Gen Cte P im Hämoglobin-Gencluster von *C. tentans* benachbart liegende Cte Q gruppiert im Aminosäure-Vergleich mit dem Gen GBA 5.1.1-3. Diese beiden Gene sind intronlos, die Übereinstimmung ihrer Aminosäure-Sequenzen beträgt 60 %. Zusätzlich zu diesen guten Sequenz-Identitäten spricht auch die Anordnung, sowie die übereinstimmende Orientierung der Gene innerhalb der jeweiligen Gengruppe der beiden Arten für ihre Orthologie (siehe Abb. 4.12). Die im Genom von *G. barbipes* benachbart liegenden Hämoglobin-Gene Gba 5.1.1-2 und Gba 5.1.1-1 lassen sich hingegen nicht ganz problemlos den Hämoglobin-Varianten von *C. tentans* zuordnen. Der Protein-Stammbaum (Abb. 4.10) zeigt, daß sie entsprechend ihrer Aminosäure-Sequenz mit einiger Sicherheit den VIIA-Varianten der IIB-Unterfamilie zuzuordnen sind (71% Übereinstimmung). Dieses Ergebnis ist mit der genomischen Anordnung der Hämoglobin-Gene Gba 5.1.1-2 und Gba 5.1.1-1 jedoch nur in Einklang zu bringen, wenn von gravierenden Rearrangements in der Clusterstruktur nach der Trennung der Entwicklungslinien von Chironomus und *Glyptotendipes* ausgegangen wird (siehe Abb. 4.12).

Das Hämoglobin-Gen Cte N besitzt ebenfalls ein kurzes Intron, das in homologer Position in dem Hämoglobin-Gen Gba 16.4-2 aus *Glyptotendipes barbipes* ebenfalls vorhanden ist. Diese Hämoglobin-Varianten weisen eine Übereinstimmung von 66 % auf Ebene der Aminosäure-Sequenzen auf. Dieser Wert alleine reicht für eine sichere Bestimmung der Orthologie dieser Gene nicht aus. Es wird hier notwendig sein, die umliegenden Genom-Bereiche und die darin enthaltenen Gene zu analysieren. Zur Zeit liegen aus diesen Bereichen keine vergleichbaren Sequenzdaten vor. Auch die Übereinstimmung der Intron-Sequenzen von 60 % liefert keine eindeutige Aussage über die eventuelle Verwandtschaft, sie könnte eine zufällige Ähnlichkeit der stark AT-reichen Sequenz darstellen (siehe Abb.4.11 B).

### Der Ursprung des „zentralen Introns“ in Hämoglobin-Genen

Laut der „Exon-Theorie der Gene“ entstanden komplexe Proteine durch die Fusion funktioneller Einheiten. Diese funktionellen Domänen werden auf Gen-Ebene von separaten Exons kodiert, die im Urgen eigenständige, sogenannte Mini-Gene darstellten. Die untranslatierten Introns wären in diesem Sinne die ehemaligen Regionen zwischen solchen Minigenen (Gilbert, 1987; Gilbert und Glynias, 1993; Gilbert et al., 1997). Aufgrund dieser Herkunft kodieren moderne Exons häufig für definierte, abgegrenzte Proteinstrukturen, die als Domänen oder Module bezeichnet werden (Blake, 1978; Go, 1981). Solch eine funktionelle Einheit ein „mini-globin“ könnte die Häm-Domäne gewesen sein. Sie wird gebildet von dem Fragment 32-139 und korrespondiert daher mit dem Produkt des zweiten Exons und einigen Aminosäure-Resten des dritten Exons (Craik et al., 1980, 1981; DeSanctis et al., 1986, 1988). Es wurde daher postuliert, daß die Introns B12-2 und G7-0, die das zweite Exon flankieren, aus der Entwicklungsgeschichte des Hämoglobins resultieren und somit ursprünglicher Natur sind (Trotman, 1998; Hankeln et al., 1997; Tyshenko und Walker, 1997; Dewilde, 1999). Auch eine perfekte Übereinstimmung der Intronpositionen über große evolutionäre Distanzen ist ein eindeutiges Argument für die Ursprünglichkeit dieser Introns.

Ausgehend von dieser Exon-Theorie der Gene wurde bereits 1981 aufgrund der Domänenstruktur des Hämoglobins das Vorhandensein eines weiteren ursprünglichen Introns im zentralen Bereich des Gens vorhergesagt (Go, 1981). Als kurz darauf solche „zentralen Introns“ in den Hämoglobin-Genen von Leguminosen und Ulmaceen nachgewiesen werden konnten (Jensen et al., 1981; Wiborg et al., 1982), galt dies als schlüssiger Beweis für ein hohes phylogenetisches Alter dieses Introns und somit als Beleg für die Exon-Theorie der Gene. Die Entdeckung weiterer „zentraler Introns“ in den Hämoglobin-Genen der Nematoden *Ascaris suum* (Sherman et al., 1992) und *Caenorhabditis elegans* (Moens et al., 1992; Kloek et al., 1993), dem Ciliaten *Paramecium caudatum* (Yamauchi et al., 1992) und der einzelligen Alge *Chlamydomonas reinhardtii* (Couture et al., 1994), schien diese Interpretation der Intron-Evolution in Hämoglobin-Genen weiter zu untermauern (Moens et al.,

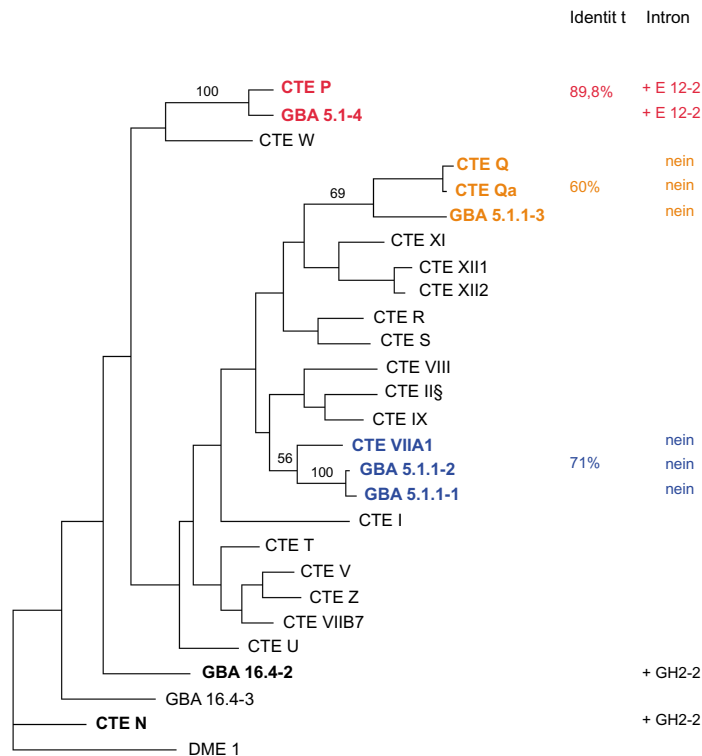


Abbildung 4.10: MP-Baum der Hämoglobine von *C. tentans* und *Glyptotendipes barbipes*

Beruhend auf dem Vergleich der Aminosäuren (siehe Anhang) wurden die Verwandtschaftsbeziehungen der Hämoglobin-Varianten von *C. tentans* (CTE) und *Glyptotendipes Barbipes* (GBA) nach der „maximum parsimony“-Methode ermittelt. Obwohl diese beiden Arten unterschiedlichen Gattungen angehören, bilden sowohl die Hämoglobine CTE P und GBA 5.1.1-4, als auch CTE Q und GBA 5.1.1-3 je eine separate Gruppe. Diese mit großer Wahrscheinlichkeit orthologen Hämoglobin-Paare sind rot hervorgehoben. Die verbleibenden Hämoglobin-Komponenten GBA5.1.1-1 und GBA5.1.1-2 sind möglicherweise Mitglieder der VIIA-Subfamilie (blau). Die Verwandtschaft zwischen den Hämoglobinen CTE N und GBA 16.4-2, deren Gene Introns an identischen Positionen aufweisen, wird durch diese Topologie nicht signifikant unterstützt (fett). Die Zahlen an den Verzweigungen geben die durch „Bootstrapping“ ermittelten Signifikanzen an.

1992). Selbst innerhalb der Vertebraten konnte im Neuroglobin, einem Mitglied der Globinfamilie, daß vorwiegend im Gehirn exprimiert wird, kürzlich ein „zentrales Intron“ nachgewiesen werden (Burmester et al., 2000). Insgesamt konnten elf Introns im zentralen Bereich, hiervon acht innerhalb der Helix E identifiziert werden (Kao et al., 1994; Hankeln et al., 1997; Dixon und Pohajdak, 1992; Blaxter et al., 1994; Hyldig-Nielsen et al., 1982; Wiborg et al., 1982; Brisson und Verma, 1982). Die Vertebraten-Hämoglobin-Gene hätten ausgehend von einem 4 Exon/3 Intron-

```

A
CteP      GTAAGTAAAATTTAATAAAAAATTATTCTACAAAAAATATTCCCTACT-----AAAAAACAG
Gba5.1.1-4  GTAAGTACAAAATGATATATCTTCATTAATCTAGTTCGTGACAATTTTAATTTAAAATGCAG
          *****-**-*-***-**-*-**-----*-----*-**-*-----*****
          *****-**-*-***-**-*-**-----*-----*-**-*-----*****

B
CteN      GTAAGTGCCTTGATTTTAATAAAAAAA--TAGTCA-TTAATm-----T
Gba16.4-2  GTAAGA--TT--TTTAATCCATAAAATTTTATTCAATTAATCGTTGGTT
          *****-**-*****-**-*****-**-*****-*****-----*
          *****-**-*****-**-*****-**-*****-*****-----*

CteN      AAAATAATTTTAATCCCAG
Gba16.4-2  TTATTCATTATAT----AG
          ---*-*-----*-----**

```

**Abbildung 4.11:** Die Intron-Sequenzen der Hämoglobin-Gene von *C. tentans* und *Glyptotendipes barbipes* im Vergleich

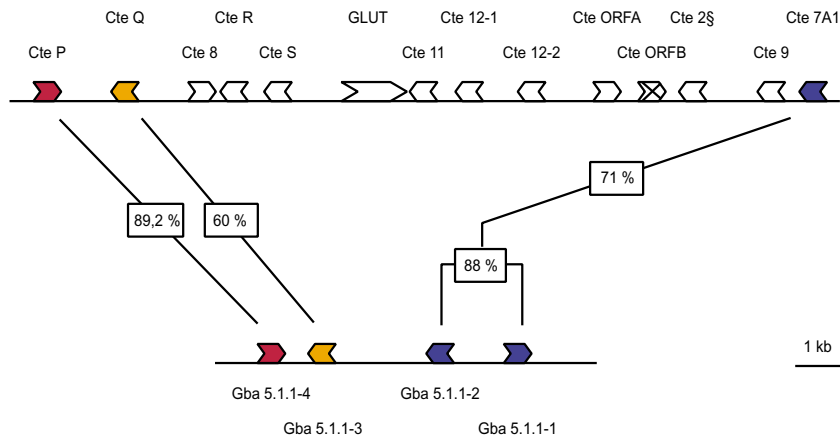
Der Vergleich der Intron-Sequenzen der Gene Cte P und Gba 5.1.1-4 ergibt ausschließlich rein zufällige Übereinstimmungen (A). Die Introns der Gene Cte N und Gba 16.4-2 weisen bei entsprechendem Alignment eine Sequenz-Identität von 60% auf (B). Die Aussagekraft dieses Wertes ist jedoch aufgrund des hohen AT-Gehaltes unklar. Es handelt sich wahrscheinlich um eine rein zufällige Ähnlichkeit, wie Sie bei stark AT-reichen Sequenzen zu erwarten ist. Sequenz-Übereinstimmungen sind mit einem Stern markiert.

Genvorläufer dieses ursprüngliche „zentrale Intron“ im Laufe ihrer Evolution bereits verloren (siehe Abb. 4.13).

Gegner dieser Hypothese weisen jedoch darauf hin, daß die Positionen der „zentralen“ Helix E-Introns in den verschiedenen Taxa keinesfalls streng konserviert sind. Damit besteht die Möglichkeit, daß diese Introns nicht homolog und damit ancestral sind, sondern aus unabhängigen Insertionsereignissen hervorgegangen sind (Stoltzfus et al., 1997). Es stellt sich daher die Frage, ob bereits das Ur-Chironomiden-Hämoglobingen die „zentralen Introns“ besaß. Nur wenn es sich bei den Introns der Chironomiden-Hämoglobin-Gene um phylogenetisch alte Introns handelt, können sie als Hinweis auf ein „zentrales Intron“ im gemeinsamen Vorläufer von Vertebraten- und Pflanzen-Hämoglobin-Genen gewertet werden. Eine mögliche Rekonstruktion der Intron-Evolution in Chironomus ist in Abb. 4.14 dargestellt. Der Vergleich der Intronverteilung in den Chironomiden-Hämoglobinen zeigt, daß die bislang beschriebenen Introns mit großer Wahrscheinlichkeit nicht ortholog sind (Hankeln et al. (1997); diese Arbeit). Alle vorhandenen Introns, auch die der Gene Cte P und Cte N, lassen sich am sparsamsten (im Sinne des „maximum parsimony“-Prinzips) durch unabhängige Insertionen in ein intronloses Ur-Chironomiden-Hämoglobingen erklären. Eine solche Intron-Insertion kann beispielsweise erfolgen, indem ein Intron am Spleißosom verbleibt und von diesem irrtümlich revers in eine neue, z. B. eine Hämoglobin-mRNA eingespleißt wird (Stoltzfus und Doolittle, 1993). Über reverse



## 4.2 Die Primärstruktur und Funktion der Hämoglobine von Chironomiden



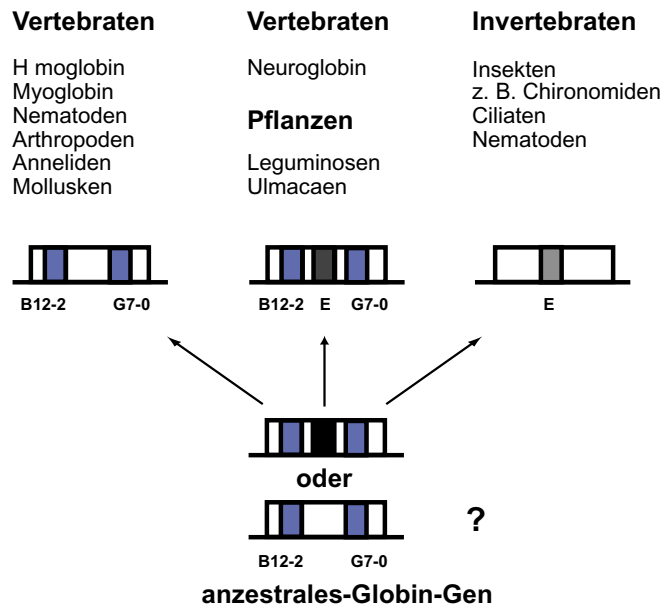
**Abbildung 4.12:** Vergleich der Struktur der Hämoglobin-Gengruppen in *C. tentans* und *Glyptotendipes barbipes*

Die Hämoglobin-Gene Cte P und Cte Q von *C. tentans* stimmen in ihrer genomischen Anordnung, sowie in der relativen Orientierung ihrer Transkriptionsrichtung (Pfeile) mit den Hämoglobin-Varianten Gba 5.1.1-4 und Gba 5.1.1-3 von *Glyptotendipes barbipes* überein. Neben der genomischen Organisation spricht auch ihre hohe Sequenz-Übereinstimmung auf Protein-Ebene, sowie eine konservierte Intronposition für die Orthologie der Varianten Cte P und Gba 5.1.1-4, sowie Cte Q und Gba 5.1.1-3. Die Gene Gba 5.1.1-1 und Gba 5.1.1-2 besitzen zu den direkten Nachbarn des Gens Cte Q keinerlei signifikante Sequenz-Übereinstimmung. Sie scheinen Mitglieder der VIIA Sub-Familie zu sein. Die %-Werte geben die jeweiligen Identitäten der korrespondierenden Aminosäure-Sequenzen an.

Transkription mit nachfolgender homologer Rekombination könnte die intronhaltige Sequenz in das Genom gelangen (Tani und Ohshima, 1991). Das Vorkommen eindeutig sequenzverwandter Introns in den Genen 2 $\beta$ , 9 und 7A der „thummi-Gruppe“ ist aller Wahrscheinlichkeit nach nicht das Ergebnis einer dreimaligen unabhängigen Insertion des gleichen Introns an identischer Position in alle drei Gene (Kao et al., 1994; Hankeln et al., 1997). Ein horizontaler Transfer der Intron-Sequenz, wie er beispielsweise durch Genkonversion stattfinden kann, bietet eine einfache Erklärung für diese Intronverteilung (Hankeln et al., 1997).

Weitaus schwieriger ist es, die herrschende Intronverteilung in Zuckmücken ausgehend von einem intronhaltigen Hämoglobin-Vorläufer zu erklären. Bei diesem Szenario sind mindestens acht unabhängige Intronverluste (in der „tentans“-Gruppe, in der „melanotus“-Gruppe und schließlich in den australo-asiatischen Arten) notwendig, um die intronlosen Gene zu erklären. Eine solche Anzahl unabhängiger Ereignisse muß als sehr unwahrscheinlich betrachtet werden.

Ausgehend von einem intronhaltigen Hämoglobin-Vorläufer-Gen stellt sich außerdem die Frage, wie die unterschiedlichen Intronpositionen innerhalb der Hämoglobin-Gen-Sequenzen zu erklären sind. Generell sollte davon ausgegangen werden,



**Abbildung 4.13:** Die Exon-/Intron-Struktur des Hämoglobin-Vorläufer-Gens von Vertebraten, Invertebraten und Pflanzen ist unbekannt

Die rezente Verteilung von Introns in den Hämoglobin-Genen unterschiedlicher phylogenetischer Taxa wirft die Frage auf, ob ein ancestrales Hämoglobin-Gen zwei oder mindestens drei Introns besaß.

daß homologe Introns in verwandten Genen an exakt identischen Positionen sitzen. Dennoch werden allein die Chironomiden-Hämoglobine an vier unterschiedlichen Stellen von „zentralen Introns“ unterbrochen (siehe Abb. 4.14 und Abb. 4.15). Es wurde postuliert, daß ein als „intron-sliding“ bezeichneter Prozeß die „Wanderung“ von Intron-Sequenzen im Verlauf der Gen-Evolution bewirken kann (Rogers, 1986; Cerff, 1995). Dieses „intron-sliding“ setzt jedoch eine Vielzahl von Ereignissen voraus und konnte bislang keine befriedigende Erklärung für die Vielzahl der Intronpositionen im zentralen Bereich der Hämoglobin-Gene liefern. So kann beispielsweise eine Intron-Verschiebung über 17 bp hinweg, wie es bei den 2 $\beta$ -Genen von *C. thummi* und *C. melanotus* zu beobachten ist, nicht ohne die zeitweilige Zerstörung des Leserahmens und damit der Genfunktion, erklärt werden. Auch liegen bislang keinerlei Evidenzen dafür vor, daß ein Verschieben der Intron-Sequenzen durch „reverses Spleißen“ (Morl und Schmelzer, 1990), eine Rolle bei der Evolution der Chironomiden-Hämoglobin-Gene gespielt haben könnte.

Zusammenfassend läßt sich somit festhalten, daß der Hämoglobin-Gen-Vorläufer in Chironomiden vermutlich kein „zentrales Intron“ besaß. Die erhobenen Daten stellen somit die von Go (1981) postulierte Hypothese der Ursprünglichkeit des „zentralen Introns“ in Hämoglobin-Genen in Frage. Die vorliegenden Daten deuten sogar darauf hin, daß selbst Introns an identischen Positionen, wie sie in einigen

## 4.2 Die Primärstruktur und Funktion der Hämoglobine von Chironomiden

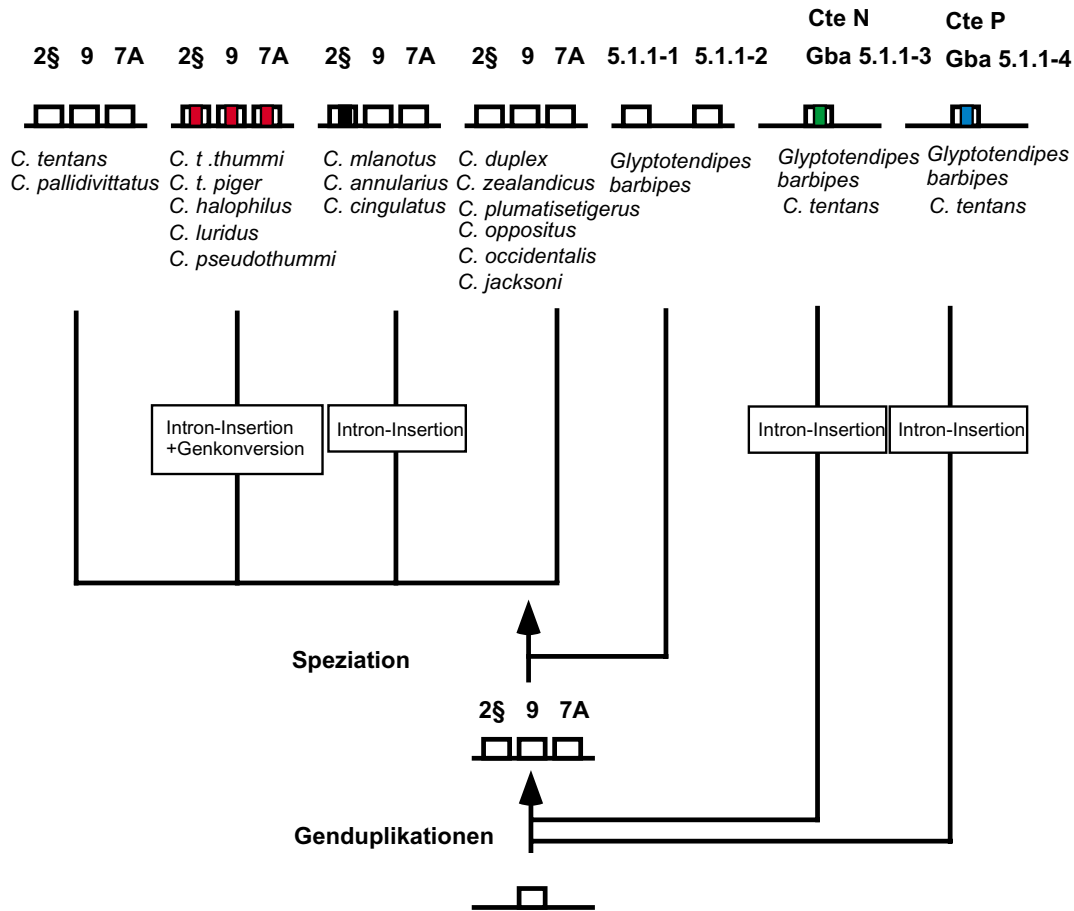


Abbildung 4.14: Die Evolution der Genstruktur bei Zuckmücken durch Intron-Insertion

Das hier abgebildete Modell erklärt die Verteilung der Introns in den Hämoglobin-Genen von Chironomiden durch vier unabhängige Intron-Insertionen. Das gemeinsame Vorläufergen aller Zuckmücken-Hämoglobin-Genes wäre somit intronfrei. Die Introns in den unterschiedlichen Positionen sind in verschiedenen Farben dargestellt. Die Genpaare Cte P/Gba 5.1.1-4 und Cte N/Gba 5.1.1-3 werden aufgrund ihrer Sequenz-Übereinstimmung (89,2% und 66% auf Aminosäure-Ebene), sowie der identischen Intron-Positionen hier als Orthologe Genpaare behandelt.

Hämoglobin-Genen von Chironomiden (Hankeln et al., 1997) und Pflanzen (Jensen et al., 1981; Wiborg et al., 1982) gefunden wurden, unabhängig voneinander entstanden sein müssen. Die in Invertebraten und Pflanzen beschriebenen „zentralen Introns“ sind demnach nicht homolog und dementsprechend auch nicht auf ein Intron im ancestralen Globin zurückzuführen. Die Variationen in der Lokalisation von „zentralen Introns“ bei Pflanzen und Invertebraten impliziert hingegen die unabhängige Insertion dieser Introns in die jeweiligen Globin-Genes nach der Auf-

spaltung der Taxa (siehe Abb. 4.15). Tatsächlich ist anzunehmen, daß zukünftige Analysen von Hämoglobin-Genen in verschiedenen Taxa eine noch weitaus größere Vielfalt der Intronpositionen zeigen werden.

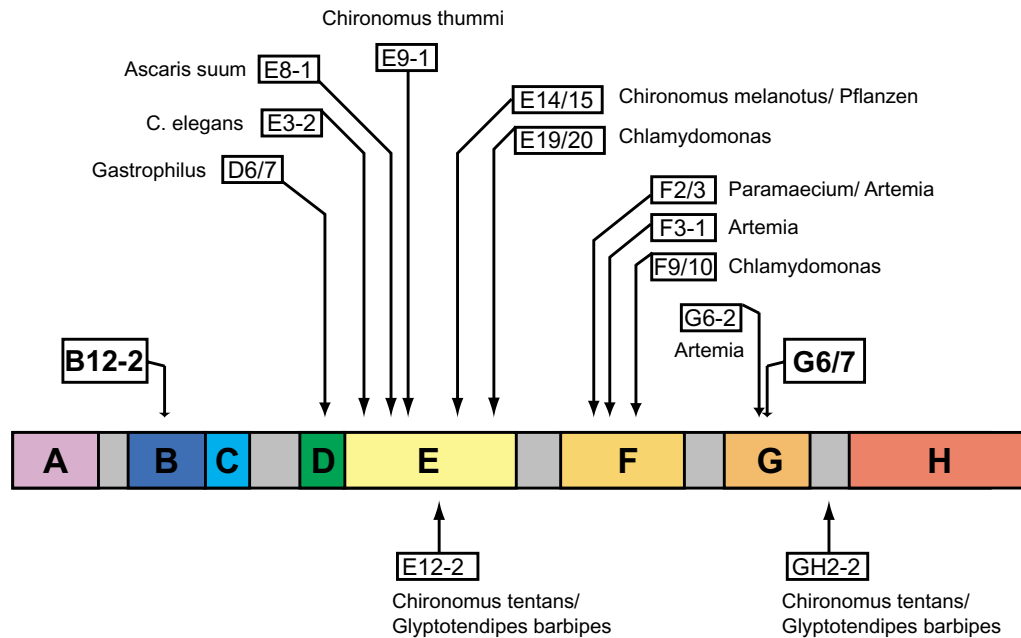


Abbildung 4.15: Die Position von Introns in Hämoglobin-Genen verschiedener Taxa

Die Intron-Positionen sind relativ zur Helixstruktur des Hämoglobins dargestellt. Die acht Helices sind mit den Buchstaben A bis H benannt. Die B- und G-Helix-Introns sind in Pflanzen und Tieren ihrer Lage nach konserviert und mögen daher phylogenetisch recht alt sein. Die zahlreichen anderen gefundenen Intronpositionen (insbesondere die umstrittenen Helix E-Introns) zeigen jedoch, daß Introns in den Hämoglobin-Genen prinzipiell sehr viele verschiedene Positionen einnehmen können. Dieser Befund läßt sich besser durch die Integration von Introns, als durch ihren differentiellen Verlust in zahlreichen stammesgeschichtlichen Linien während der Evolution erklären.

#### 4.2.5 Vergleich der paralogen Hämoglobin-Varianten von *C. tentans*: Hämoglobin-Gen-Subfamilien und „single copy“-Gene

Durch die molekulare Analyse der Hämoglobin-Gengruppe D von *C. tentans* konnten insgesamt 38 paraloge Hämoglobin-Gene identifiziert werden. Mit Ausnahme eines Pseudogens (Cte ORF B; Rozynek et al. (1991)) besitzen alle Gene, die für eine Transkription erforderlichen Merkmale. Es handelt sich daher aller Wahrscheinlichkeit nach um funktionelle Gene. Die große Anzahl bekannter Hämoglobin-Komponenten in *C. tentans* ermöglicht einen detaillierten Vergleich der paralogen Gen- und Aminosäure-Sequenzen. Hierdurch können die Verwandtschaftsverhältnisse der Hämoglobine dargestellt und mögliche Gen-Subfamilien identifiziert werden.

Innerhalb der Hämoglobin-Familie von *C. tentans* findet man zwei Gen-Gruppen, deren Mitglieder sich durch ihre sehr hohe Sequenz-Verwandtschaft auszeichnen. Sie kodieren für die Hämoglobin-Varianten VIIA und VIIB und werden daher als 7A- und 7B-Gen-Subfamilien bezeichnet. Während die 7B-Gen-Subfamilie bereits in den Spezies *C. thummi* und *C. piger* beschrieben wurde (Trewitt et al., 1995; Hankeln und Schmidt, 1991), war bislang unbekannt, daß auch das Hämoglobin VIIA von mehreren sehr ähnlichen Genen kodiert wird. Innerhalb der 7A-Subfamilie liegen die Sequenz-Übereinstimmungen der Proteine zwischen 87 % und 99,4 %, innerhalb der 7B-Subfamilie zwischen 76 % und 99,4 % (siehe Tabelle 4.6).

Die aus dem paarweisen Sequenz-Vergleich resultierenden Distanzwerte (p-Distanzen) dienen als Grundlage für die Darstellung der Verwandtschaftsverhältnisse in Form eines phylogenetischen Baums (NJ-Baum und MP-Baum). Zur Verifizierung der Verzweigungsmuster der Gen-Bäume wurden Verwandtschaftsdendrogramme basierend auf den Aminosäure-Sequenzen erstellt und mit Hilfe des Bootstrap-Tests überprüft. Der Gen-Baum (NJ-Baum) ist in Abb. 4.16a, der Aminosäure-Baum (MP-Baum) in Abb. 4.16b gezeigt. Das Auftreten von zwei Subfamilien innerhalb der Chironomiden-Hämoglobine spiegelt sich in der Topologie der phylogenetischen Bäume wider: alle Mitglieder einer Subfamilie lassen sich von einem gemeinsamen Vorfahren ableiten. Während die 7A-Gene einen monophyletischen Ursprung aufweisen, kann dies für die 7B-Hämoglobin-Gene nicht zweifelsfrei gezeigt werden. Die Verzweigung, die die 7B-Gengruppe von den Genen Cte T, V und Z trennt, wird durch den Bootstrap-Test nicht mit ausreichender Signifikanz unterstützt. Auf Aminosäure-Ebene gruppieren die Varianten CTE T, V und Z sogar innerhalb der VIIB-Gruppe. Sie sind demnach nahe Verwandte der VIIB-Gene. Innerhalb der 7B-Subfamilie bilden die Varianten 7B5/7B6, 7B4/7B8, 7B9/7B10 und 7B12/7B13 Paare. Dieses Verzweigungsmuster stimmt bei Gen- und Aminosäure-Bäumen überein, unabhängig davon, mit welcher Methode sie ermittelt wurden. Bis auf das Genpaar 7B8/7B4 werden alle diese Gruppierungen mit sehr hohen Bootstrap-Werten (98-100) unterstützt, so daß von einer engen Sequenz-Verwandtschaft innerhalb dieser Genpaare ausgegangen werden kann. Die Gene 7B7 und 7B11 zeigen hingegen keine vergleichbar nahe Verwandtschaft zu anderen Mitgliedern der 7B-Subfamilie.

A	CTE-VIIA2	CTE-VIIA3	CTE-VIIA4	CTE-VIIA5	CTE-VIIA6	CTE-VIIA7	CTE-VIIA8
CTEVIIA1	94,4	99,4	92,6	98,8	92,6	95,03	94,4
CTEVIIA2		93,8	88,8	95,7	88,2	96,9	98,8
CTEVIIA3			92	98,1	92	95,7	95
CTEVIIA4				91,3	98,8	87,6	87,6
CTEVIIA5					91,3	96,3	95,6
CTEVIIA6						88,2	87
CTEVIIA7							98,1

B	CTE-VIIB13	CTE-VIIB7	CTE-VIIB6	CTE-VIIB5	CTE-VIIB10	CTE-VIIB9	CTE-VIIB8	CTE-VIIB4	CTE-VIIB11
VIIB12	99,4	78,4	80,8	80,8	80,1	80,8	80,8	79,5	76
VIIB13		78,4	81,4	81,4	80,1	80,8	80,8	79,5	76
VIIB7			80,8	80,8	80,1	80,8	80,8	80,8	76,5
VIIB6				99,4	90	90,7	92,6	90,7	83,9
VIIB5					90	90,7	92,6	90,7	83,9
VIIB10						99,4	93,8	95	83,9
VIIB9							94,4	95,7	84,5
VIIB8								95,5	84,5
VIIB4									84,5

**Tabelle 4.6: Aminosäure-Sequenz-Übereinstimmungen innerhalb der Hämoglobin-Subfamilien von *C. tentans***

In Tabelle A sind die Sequenz-Übereinstimmungen der Mitglieder der VIIA-Subfamilie, in Tabelle B die der VIIB-Subfamilie in % angegeben. Die besten Übereinstimmungen sind rot hervorgehoben. Eine Übereinstimmung von 99,4 % entspricht dem Austausch einer Aminosäure.

Auch innerhalb der 7A-Subfamilie lassen sich in allen Bäumen übereinstimmende Gen-Verwandtschaften feststellen. So weisen die Varianten 7A2/7A8 und 7A7, sowie 7A4 und 7A6 eine signifikant hohe Identität, sowohl der Nukleotid-, als auch der Aminosäure-Sequenzen auf (Bootstrap-Werte zwischen 92 und 99). Das Auftreten von Gen-Subfamilien mit ihren extrem hohen Sequenz-Übereinstimmungen ist ein Hinweis darauf, daß rezente Ereignisse an der Evolution dieser Sequenzen maßgeblich beteiligt sind.

Neben diesen Subfamilien gibt es Hämoglobin-Varianten, die nur von einem einzigen Gen kodiert werden. Diese Gene werden dementsprechend als „single copy“-Gene bezeichnet. Hierzu zählen die Gene Cte 2 $\beta$ , 1, 8, 9, und 6 (U), deren Produkte in *C. thummi* bereits nachgewiesen wurden (zusammengefaßt in Goodman et al. (1983)), sowie die Gene Cte 11, 12, ORFA, ORFB (Pseudogen), N, P, Q, R, S, T, V, W, Y und Z. Die Tatsache, daß ein Teil dieser Gene auch in anderen Arten bereits identifiziert wurde, deutet darauf hin, daß ihr Ursprung vor der Aufspaltung

der Chironomiden in verschiedene Spezies liegt. Wie der Vergleich der Gen- und Aminosäure-Stammbäume zeigt, ist die Ermittlung der Verwandtschaftsverhältnisse dieser, über sehr lange Zeiten getrennt evolvierender Gene, häufig nicht mehr mit Sicherheit möglich. Nur die Hämoglobin-Paare R/S, 11/ORF B und V/Z werden im Aminosäure-Stammbaum durch signifikante Bootstrap-Werte unterstützt (82-92). Dies ist um so erstaunlicher, als es sich bei dem ORF B um ein Pseudogen handelt, das ohne Selektionsdruck eine hohe Evolutionsrate aufweisen sollte. Seine nahe Verwandtschaft mit dem Gen Cte 11 (81,5% Übereinstimmung) deutet demnach darauf hin, daß es seine Funktion erst rezent verloren haben muß. In beiden AS-Bäumen (NJ- und MP-Baum) bilden auch die Hämoglobine II $\beta$ /IX, P/Q und Y/W übereinstimmend Paare. Diese werden jedoch nicht durch signifikante Bootstrap-Werte gesichert.

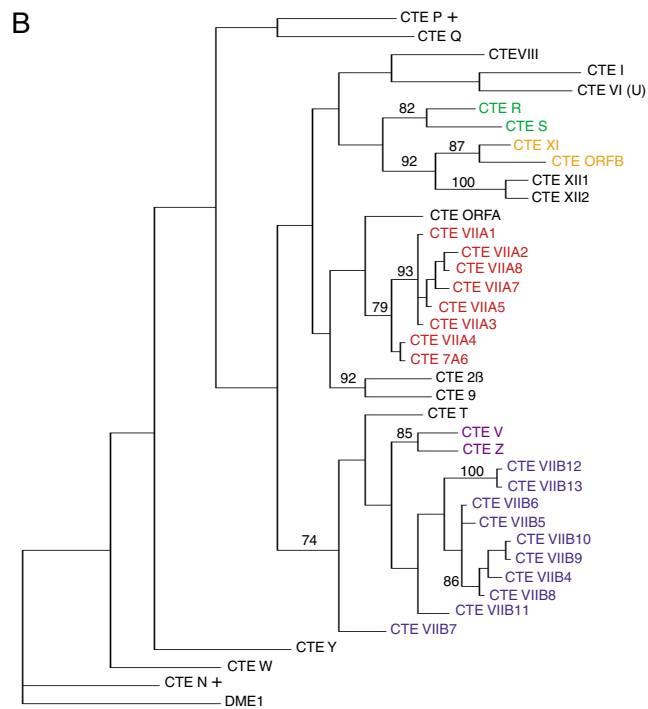
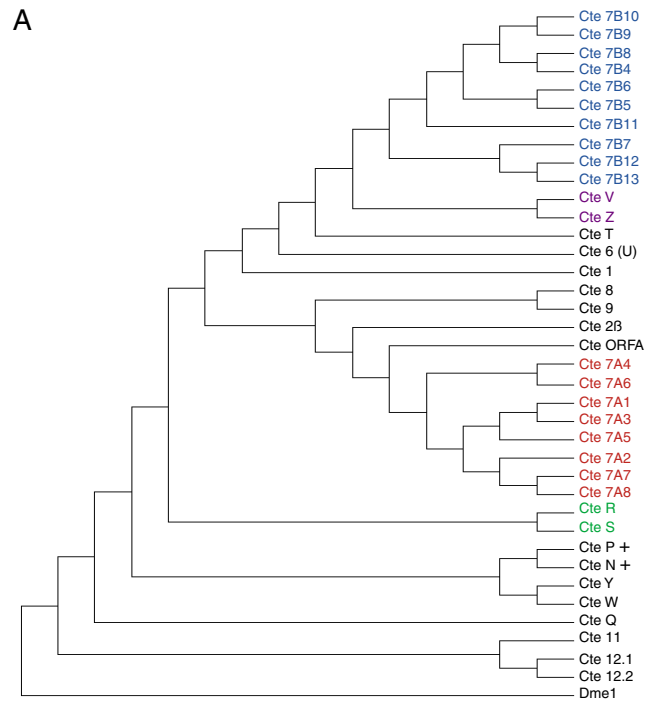


Abbildung 4.16 Legende siehe folgende Seite



**Abbildung 4.16:** Phylogenetische Bäume der paralogen Hämoglobin-Gene und Hämoglobin-Proteine von *C. tentans*

Basierend auf den p-Distanzen (siehe CD; Datei: Cte\_pDistanzen.pdf) wurde der NJ-Baum für die Nukleotid-Sequenzen berechnet (A). Der Aminosäure-Baum wurde nach dem „maximum parsimony“ Algorithmus ermittelt (MP-Baum, in B). Die Ergebnisse des Bootstrap-Tests sind an den Ästen vermerkt, er wurde mit 500 Replikationen durchgeführt. Die Verzweigungen innerhalb der Gen-Subfamilien werden, wie im Text erläutert, mit signifikanten Bootstrap-Werten unterstützt (nicht gezeigt). Die 7A-Subfamilie ist in rot, die 7B-Subfamilie in blau dargestellt. „Single copy“-Gene, deren Anordnung durch „boots trapping“ unterstützt wird, sind ebenfalls farbig hervorgehoben. Die intronhaltigen Varianten sind durch ein + gekennzeichnet.

### 4.3 Mechanismen der Entstehung der Hämoglobin-Genfamilie in *C. tentans*

Die ermittelten Verwandtschaftsverhältnisse der Hämoglobin-Varianten von *C. tentans* sollen dazu dienen, die dynamische Sequenz-Evolution dieser Gene zu beschreiben. Besonders die Mitglieder von Gen-Subfamilien, die sich durch ihre überdurchschnittlich hohen Sequenz-Übereinstimmungen auszeichnen, können Hinweise auf rezente Veränderungen des Genrepertoires und der Clusterstruktur geben. Weitere Vergleiche paraloger Sequenzen innerhalb der Hämoglobin-Gengruppe D, insbesondere der Intergen-Regionen, können dazu dienen, die Entstehungsgeschichte einzelner Hämoglobin-Gene zu rekonstruieren. Anhand dieser Daten sollen Mechanismen beschrieben werden, die bei der Sequenz-Evolution der Hämoglobin-Genfamilie in Chironomiden eine wichtige Rolle spielen.

#### 4.3.1 Die Entstehung von Genen durch Duplikation am Beispiel der Hämoglobin-Gene Cte 12 und Cte 7A

Der Vergleich paraloger Sequenzen ermöglicht das Auffinden rezent duplizierter Regionen, da diese sich durch eine überdurchschnittlich hohe Sequenz-Übereinstimmung auszeichnen. Insbesondere die Intergen-Regionen, die in der Regel sehr divergente Sequenzen aufweisen, liefern Hinweise auf Duplikations-Ereignisse. Um Sequenz-Wiederholungen innerhalb der Hämoglobin-Gengruppe aufzufinden, wurde eine Datenbank aus den vorhandenen Sequenzen des Clusters D von *C. tentans* erstellt und mit den Sequenzen der Intergen-Regionen auf Übereinstimmungen hin durchsucht. Interessante Bereiche wurden anschließend detailliert verglichen. Innerhalb der Gen-Gruppe D von *C. tentans* konnten hierdurch zwei Bereiche identifiziert werden, die aus einer Duplikation hervorgegangen sind.

Die paralogen Hämoglobin-Gene Cte 12.1 und Cte 12.2 besitzen mit 96,5 % eine überdurchschnittlich hohe Sequenz-Übereinstimmung. Innerhalb der Gengruppe liegen diese Gene benachbart und sind in gleicher Orientierung angeordnet. Die Abbildung 4.17 zeigt, daß nicht nur die Gene selbst, sondern auch ihre flankierenden Regionen über einen Bereich von ca. 500 Basen eine erstaunlich hohe Sequenz-Übereinstimmung aufweisen. Mit einem Wert von über 95% liegt diese sogar über dem Durchschnitt für paraloge, kodierende Bereiche. Sequenz-Abschnitte, die eine auffällig hohe Übereinstimmung aufweisen, können durch den Mechanismus der Genkonversion entstehen. Betrachtet man jedoch den Sequenz-Vergleich in Abb. 4.18, so fällt auf, daß in den Intergenregionen stromabwärts der Gene kurze Sequenz-Bereiche mit einer sehr hohen Übereinstimmung von einer Vielzahl von Deletionen oder Insertionen unterbrochen werden. Dies ließe sich nur durch eine sehr große Anzahl unabhängiger Genkonversions-Ereignisse erklären. Die Entstehung der hier vorliegenden Sequenz-Übereinstimmung durch Genkonversion erscheint in diesem Zusammenhang daher sehr unwahrscheinlich.

Intergen-Regionen können in der Regel ohne nennenswerten Selektionsdruck evolvieren und unterliegen somit schnellen Veränderungen. Eine derart gute Sequenz-Übereinstimmung läßt daher auf eine nahe Verwandtschaft der Sequenzen, hervorgerufen durch ein rezentes Duplikations-Ereignis, schließen. Der gesamte von der Duplikation betroffene Bereich erstreckt sich über eine Region von ca. 3300 Basen. Der Bruchpunkt des Duplikations-Ereignisses läßt sich auf eine 100 bp große Region zwischen den duplizierten Bereichen eingrenzen (siehe Abb. 4.18). Auch durch den Mechanismus der Genkonversion entstehen Sequenzen-Bereiche, die eine auffällig hohe Übereinstimmung aufweisen. Betrachtet man jedoch den Sequenz-Vergleich in Abb. 4.18, so fällt auf, daß in den Intergen-Regionen Sequenz-Bereiche mit einer sehr hohen Übereinstimmung zu finden sind, die von einer Vielzahl Deletionen bzw. Insertionen unterbrochen werden. Ein solcher Befund ließe sich durch eine sehr große Anzahl unabhängiger Genkonversions-Ereignisse erklären. Die Entstehung der Sequenz-Übereinstimmungen allein durch Genkonversion kann daher als unwahrscheinlich betrachtet werden.

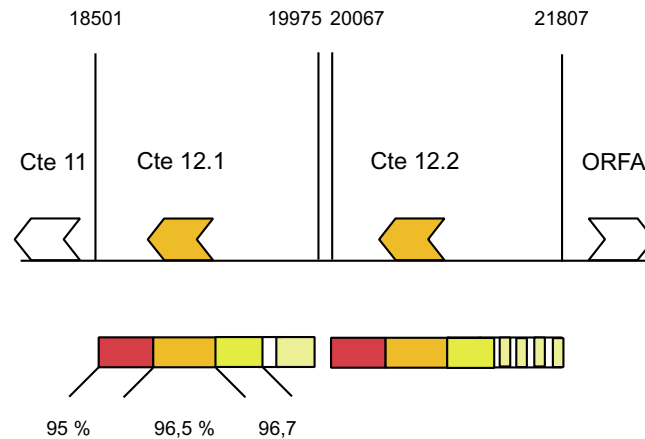


Abbildung 4.17: Die Entstehung der Hämoglobin-Gene Cte 12.1 und Cte 12.2 durch Duplikation

Die paralogen Gene Cte 12.1 und Cte 12.2 sind innerhalb des Clusters D von *C. tentans* benachbart und in gleicher Orientierung angeordnet. Sie besitzen eine überraschend hohe Sequenz-Übereinstimmung, die sich auch auf große Bereiche ihrer flankierenden Intergen-Regionen ausdehnt. Der farbige Balken gibt den Grad der Identität (in %) zwischen den beiden Sequenzbereichen an. Abschnitte mit verwandter Nukleotid-Sequenz sind durch identische Farben gekennzeichnet, weiße Bereiche bedeuten Insertionen oder Deletionen. Die extrem hohen Übereinstimmungen über weite Bereiche nichtkodierender Sequenz deuten auf eine rezente Duplikation dieser Region hin. Die Zahlen geben die Nukleotidpositionen in der Gesamtsequenz des Contigs 2 an (siehe CD; Datei: 50 kb Contig 1-21400.pdf; Seiten 8-9). Der Bruchpunkt ist voraussichtlich zwischen den Basen 19975 und 20067 lokalisiert.

Daß Sequenz-Duplikationen einen großen Anteil an der Entstehung und Evolution von Genfamilien haben, läßt sich auch am Beispiel der 7A-Gen-Subfamilie aufzeigen. Alle 7A-Gene liegen benachbart innerhalb der Hämoglobin-Genfamilie. Der Sequenz-Vergleich zeigt, daß mindestens drei der acht Gene aus einem einzigen Duplikations-Ereignis hervorgegangen sind. Im Gegensatz zu den Cte 12-Genen liegt der duplizierte Bereich hier invertiert vor (siehe Abb. 4.19). Die Duplikation umfaßt insgesamt eine 4 kb große Region. Die in Kapitel 4.2.5 festgestellte nahe Verwandtschaft der Hämoglobine 7A2 und 7A8, sowie 7A4 und 7A6 erklärt sich daher durch ihre Abstammung von gemeinsamen Vorläufer-Genen.

Die Expansion und Kontraktion von Genfamilien erfolgt häufig durch inäquale Rekombination. Dieser reziproke Sequenzaustausch resultiert stets in einer Duplikation in einem der Rekombinationspartner, sowie in einer Deletion innerhalb des anderen. Die hierdurch hervorgerufene Fluktuation der Repetitionseinheiten hat häufig phänotypische Auswirkungen. So ist beispielsweise der Verlust von X-chromosomalen Photopigment-Genen durch inäquale Rekombinationsprozesse eine der Mutationen, die zu Rot-Grün-Farbenblindheit führen (Nathans et al., 1986).

## 4 Diskussion

---

```

Cte 12.1      1  GAAAAAGTTTGAGAAAATCTGGACATTTTAAAGCTTAGCTAGGGCACAGCAGCTCGAGA
Cte 12.2      1  GAACATTTTGGGAAAATCTGGGACGTTTAAAGCTTAGCTAGGGCACAGCAGCTTGAGA
                *** *   ****   *****   ****   *****   *****   *****   ****

Cte 12.1      61  AAAAAA--TTCAGCAGTGTCCGGAAAAATGCAATAAATGATAACCTTAATAATAAAG
Cte 12.2      61  AAAAAAAAATTCAGCAGTGTCTCGGCCAAAATACAATAGATGATAACCTTAGTAATAAAG
                *****   *****   ****   *****   *****   *****   *****

Cte 12.1      118  AAACAACCCCTTATAGAAATTCAAAAATTTAAATATGAAAAATATCAAATCATTAATF
Cte 12.2      121  AAACAACCCCTTATAGAAATTCAAAAATTTAAATGTTGAAAAATATCAAATCATTAATF
                *****   *****   *****   *****   *****   *****

Cte 12.1      178  ATCAATTTTAAAAAATTTTATTTAAAAATAAATTAATTCAAAAATTTATTTATACAAAATA
Cte 12.2      181  ATCAATTTTATAATAATTTT-ATTTAAAAATAAATTAATTCAAAAATTTATTTATACAAAATA
                *****   **   *****   *****   *****   *****

Cte 12.1      238  AGTTCATAAACATGCTTTACTCAGAGGCATTTTAAGCTAA-----
Cte 12.2      240  AGTTCATAAACATGCTTTAATCAGAGGCATTTTAAGCAAATTTTCAACTAGCTTTTAA
                *****   *****   *****   **

Cte 12.1      279  -----TTTTAAAGTCGACTAAAAAATTTGCTTCGTAAGCGGAATF
Cte 12.2      300  ATTGGTTAATCTAAGCAAATTTTAAAGTCGACTAAAAA-TTTTCTTCGTAAGCGGAATF
                *****   *****   *****

Cte 12.1      320  TTAGCCGCTGAGTTAAGTGAAGGTTTATATCGGATTTCCATCAAAAAGCTGAGAAAAGAAC
Cte 12.2      359  TTGGCCGCTGAGTTAAGTGAAGGTTTATATCGGATTTCCATCAAGAAGCTGAGAAAAGAAC
                **   *****   *****   *****   *****   *   *****

Cte 12.1      380  TAAATAAACATTTATCAAGACCTTTTTCAAAAGCCTGAGCTACATTTATCTCCCATGACAC
Cte 12.2      419  AAGATAAACATTTATCAAAAACCTTTTTCAAAAGCCTGAGCTACATTTATCTCCCATGACAT
                *   *****   *****   *****   *****

Cte 12.1      440  ATGTTCTGATAGGTATTTAACCATCGAACTCGGAATGCATTTGAATGGGCCTTTGAAAT
Cte 12.2      479  ATGTTCTGAGAGGTATTTAATCATCGAACTCGGAATGCATTTGAATGGGCCTTTGAAAT
                *****   *****   *****   *****

Cte 12.1      500  ACCACGTTCTTTGTGACTAACAGCTAACTGATTGATGAGAGTTTTCGGCGGGAACATT
Cte 12.2      539  ACCACGTTCTTTGTGACTAACAGCTAACTGATTGATGAGAGTTTTCGGCGGGAACACC
                *****

Cte 12.1      560  TGAATCGTTTGGCTCAAAATTCATAAGTTGAGAAAAGAATCCAACAATTTGGCTGCATG
Cte 12.2      599  TGAATCGTTTGGATCAAGTTCAATAATTTTCAGAAAAGAATCCAACAATTTGGCTGCATG
                *****   ****   *****   **   *****

Cte 12.1      620  TGTAGCAAAAAGCCGCGCAATTTTAAATGATTCAAGATCCTTTCCAGCAAACGTGGAAA
Cte 12.2      659  TGTAGCAAAAAGCCGCGCAATTTTAAATGATTCAAGATCCTTTCCAGCAAACGTGGAAA
                *****   *****

Cte 12.1      680  ACGGCCTTGAATGCTCGGATTAGCAGCAAAAATGAAGCAAGAAATGTCAACTTCATTGTG
Cte 12.2      719  ACGAGCTTGAATGCTCGGATTAGCAGCAAAAATGAAGCAAGAAATGTCAACTTCATTGTG
                ***   *****

Cte 12.1      740  TTTAACTTCATTTCCATGATGATCTGATGAGAGCAGTTTGGTCCCTCCACAAATGAGCAAG
Cte 12.2      779  TTTAACTTCATTTCCATAATGATCTGATGCGAGCAGTTTGGTCCCTCCACAAATGGA----
                *****   *****   *****

Cte 12.1      800  ATCACAACCTGGTTACAGCAATGCACAAAATAAAAATAAATAAATTTTCATTTT---GCTT
Cte 12.2      834  -TCACAACCTGGTTACAGCAATGCACAAAATAAAAATAAATAAATTTTCATTTT---GCTT
                *****   *****

Cte 12.1      857  CTTAATAAACAAAAATTTTGTGGAAACTTAAATGAATTAATAGAAAAGCTGATCTTTTA
Cte 12.2      893  CTTAATAAACAAAAATTTTGTGGAAACTTAAATGAATTAATAGAAAAGCTGATCTTTTA
                *****

Cte 12.1      917  TACTCACAATTTATGTAACAAAAATTTAAATCGCTTCTGTACACTCGAATTTAGAAATA
Cte 12.2      953  TACTCACAAT--TGTAAACAAAAATTTAAATCGCTTCTGTACACTCGAATTTAGAAATA
                *****   *****

Cte 12.1      977  TATTATGCTPATATACAGCTAAATATTTAGTGAAAAATAAATAAATTTAATGAGATAAT
Cte 12.2      1010  AATTATGCTTTATACAGCTAAATATTTAGTGAAAAATAAATAAATTTAATGAGATAAT
                *****   *****

Cte 12.1      1037  GAGTGATTTTCTAAGTCAAGGACAT---TATTTTGTGAAAATATTCAAAA-----
Cte 12.2      1070  GAATGATTTTCTAAGTCAAGGACAGTCAATTTTGTGAAAATATTCAAAAATCTCA

```

Abbildung 4.18 Legende siehe folgende Seite



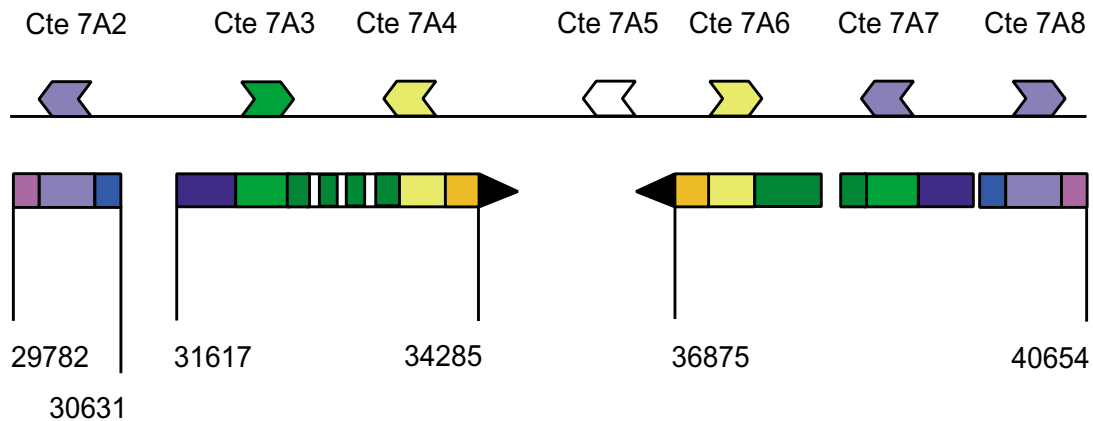
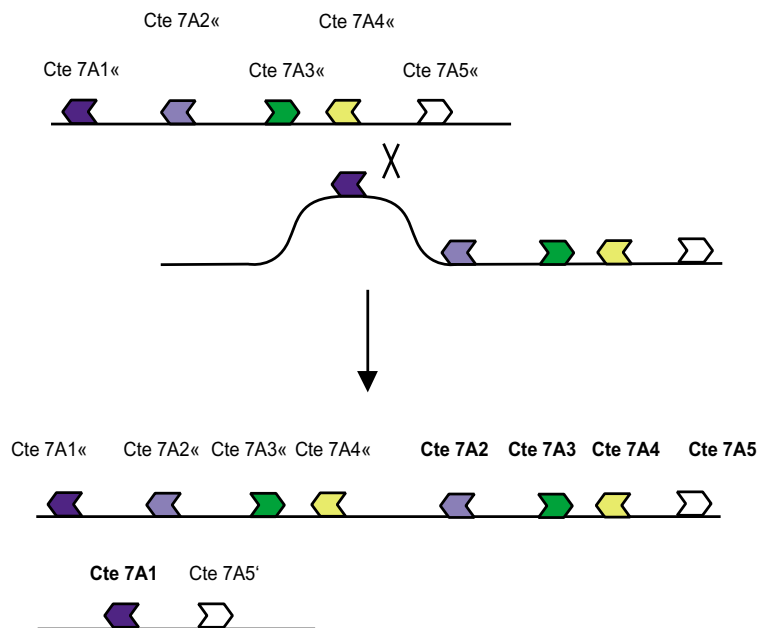


Abbildung 4.19: Die Expansion der 7A-Subfamilie durch Duplikation

Der Sequenz-Vergleich innerhalb der 7A-Subfamilie zeigt, daß ein drei Hämoglobin-Gene umfassender Bereich dupliziert und invertiert vorliegt. Identische Farben innerhalb der Balken kennzeichnen Regionen mit einer Sequenz-Übereinstimmung von über 85 %. Die weißen Bereiche bedeuten Insertionen oder Deletionen. Die extrem hohen Übereinstimmungen über weite Bereiche nichtkodierender Sequenz deuten auf eine rezente Duplikation dieser Region hin. Die Zahlen (29786-40654) geben die Nukleotidpositionen in der Gesamtsequenz des Contigs 2 an (siehe CD; Datei: 50 kb Contig 21401-50607.pdf).

Auch im Falle der Gene Cte 12, sowie in der Cte 7A-Subfamilie könnte inäquales Crossing-over zu Genduplikationen geführt haben. Die streng tandemartige Anordnung der Gene mit ihren hohen Sequenz-Übereinstimmungen sollte eine inäquale Rekombination innerhalb des Clusters begünstigen. Die Abbildung 4.20 zeigt hypothetisch die Entstehung der 7A-Subfamilie in *C. tentans*. Zunächst erfolgt die Duplikation der 7A-Vorläufergene durch inäquales Crossing-over zwischen Schwesterchromatiden (A). Sind repetitive Sequenz-Kopien auf einem DNA-Molekül gegensinnig angeordnet, so kann es bei einer Rekombination innerhalb einer Chromatide zu einer Inversion desjenigen DNA-Abschnittes kommen, der zwischen den Stellen des Crossing-over liegt (Abb. 4.20B). Die Sequenz-Übereinstimmungen innerhalb der 7A-Gen-Subfamilie in *C. tentans*, sowie die Reihenfolge und Orientierung der einzelnen Gene könnte somit durch zwei aufeinanderfolgende inäquale Crossing-over Ereignisse einfach erklärt werden.

A



B

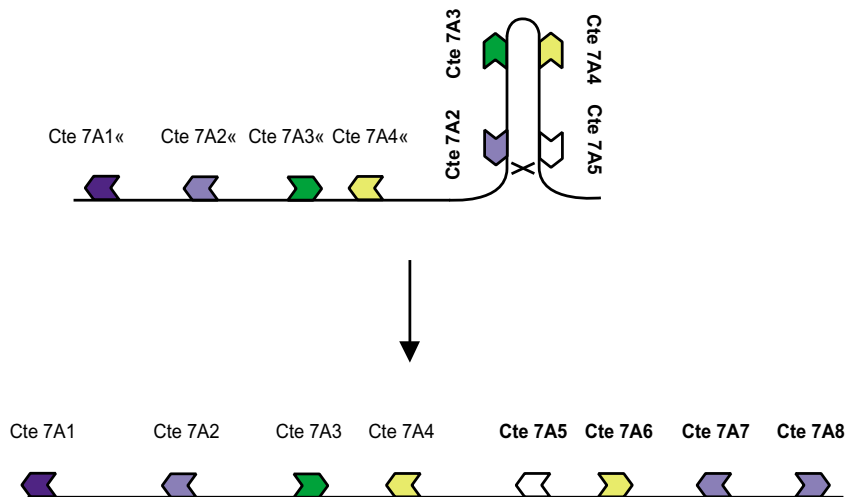


Abbildung 4.20: Die mögliche Evolution der 7A-Gen-Subfamilie in *C. tentans*

Die Entstehung der 7A-Gen-Subfamilie in *C. tentans* könnte hypothetisch durch zwei aufeinanderfolgende Rekombinations-Ereignisse einfach erklärt werden. Die inäquale Rekombination zwischen den Schwesterchromatiden resultiert zunächst in einer Verdopplung der Vorläufergene 7A2'-7A4' (A). Ein anschließendes intrachromatid Crossing-over könnte für die Inversion der Gene 7A2- 7A5 verantwortlich sein und den heute zu beobachtenden Zustand erzeugt haben (B).

### 4.3.2 Duplikation eines 3' nicht-translatierten Bereichs durch Transposition?

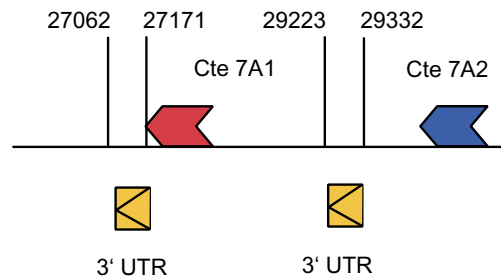
In der Intergenregion zwischen den Genen Cte 7A1 und Cte 7A2 befindet sich die Sequenz eines 7A-Gen-spezifischen 3'nicht-translatierten Bereichs (3'UTR). Der komplette 3'UTR des Cte 7A1-Gens liegt hier in einer nahezu perfekten Kopie vor: die Sequenz-Übereinstimmung beträgt 90 % über eine Länge von 110 bp. Der dazugehörige kodierende Bereich fehlt jedoch vollständig. Am Ende dieser Sequenz-Kopie befindet sich ein zehn Basenpaar langer poly-A Abschnitt. Der gesamte Bereich ist von direkten Sequenzwiederholungen mit einer Länge von sieben Basenpaaren flankiert. Die Sequenz dieser sogenannten „direct repeats“ ist bis auf einen Basenaustausch konserviert (siehe Abb. 4.21).

Diese Befunde legen die Vermutung nahe, daß es sich bei dem hier vorhandenen 3'UTR um eine durch Retrotransposition entstandene Sequenz-Kopie handeln könnte. Die Verbreitung von Sequenzen durch Retrotransposition erfordert das Umschreiben der mRNA in die cDNA durch das Enzym Reverse Transkriptase. Diese neugebildete cDNA kann anschließend in das Genom reintegrieren (siehe Abb. 1.2). Der hier vorhandene 3'UTR weist die typischen Erkennungsmerkmale solcher aus einer Retrotransposition hervorgegangenen Genduplikate auf: er besitzt die von der Polyadenylierung der mRNA herrührende polyA-Sequenz am 3'-Ende, sowie die flankierenden sogenannten „Zielsequenz-Duplikationen“, die gewöhnlich bei der Integration mobiler DNA entstehen.

Bereits 1984 entstand die Vorstellung, die Hämoglobin-Gene der Chironomiden seien eigentlich prozessierte Retrogene. Die Integration gespleißter RNA-Intermediate in das Genom schien eine geeignete Erklärung für die weitgehende Intronlosigkeit dieser Gene zu bieten (Antoine und Niessing, 1984; Lewin, 1984; Sargent et al., 1994). Der hier beschriebene 3' nicht-translatierte Bereich eines Hämoglobin-Gens ist bislang jedoch der einzige Hinweis auf eine, zumindest theoretisch mögliche, Entstehung von Hämoglobin-Genen durch Retrotransposition. Trotz dieses Befundes erscheint es weiterhin mehr als fraglich, ob die Multiplizität der Hämoglobin-Gene in Chironomiden primär durch Retrotransposition entstanden ist. Wenngleich es in der Literatur wenige Beispiele für funktionelle Retrogene gibt (Soares et al., 1985; Boer et al., 1987; Sargent et al., 1994), scheint dieses Ergebnis zu zeigen, daß die Entstehung funktioneller Gene auf diesem Wege höchst unwahrscheinlich ist. Selbst die Integration einer vollständigen cDNA resultiert in aller Regel in einem Pseudogen, weil die für eine Transkription essentiellen Promotorregionen fehlen. Es ist daher eher wahrscheinlich, daß Genduplikationen, wie sie innerhalb der 7A-Subfamilie und bei den Cte 12 Genen nachgewiesen werden konnten, primär für die Multiplizität der Chironomiden Hämoglobin-Gene verantwortlich sind.

Sind Gene repetitiv und gleichsinnig auf einem DNA-Molekül angeordnet, so kann ein Crossing-over innerhalb der Chromatide zu einer Exzision von DNA und somit zum Verlust von Genen führen. Ein solcher Prozeß könnte auch hier für den





obere Sequenz: Intergenregion Cte 7A1- Cte 7A2

untere Sequenz: Cte 7A1 3' UTR

```

      →
AAGCTTTTCA TATTTT TTAGATCTTT TAATTTTAT
TAATATAATA AACAATAAAT TTTGATTTCGT AGATTTTAT
• • • • • • • • • • • • • • • • • •
TAA AATTCAT CCAAATCATG GACTATTACT ATGCACATGC
TAA AATTCAT CCAAATCATG GACTTTTACT ATGCACATGC
•
TTTGAGCTGT TAATGAATAC ATGCTCTTAT TGATTCTATG
TTTGAGCTGT TAATAAATAC ATGTTCTTAT TGATTCTATA
• • • • • • • • • • • • • • • • • •
      →
AAATCACATA ATTAATTTAT AAATATTTT
AAATCACAAT TTACAAAATT TACAAAGCGG
• • • • • • • • • • • • • • • • • •
    
```

Abbildung 4.21: Die Duplikation eines 3' nicht-translatierten Bereichs

In der Intergenregion zwischen den Hämoglobin-Genen Cte 7A1 und Cte 7A2 befindet sich die Sequenz eines Hämoglobin-Gen typischen 3' nicht-translatierten Bereichs (3'UTR). Die Zahlen geben seine exakte Position innerhalb der Gesamtsequenz des Contig 2 an (siehe CD; Datei: 50 kb Contig 21401-50607.pdf; Seite 2). Der Sequenz-Vergleich zeigt, daß der 3'UTR des Gens 7A1 hier in einer nahezu perfekten Kopie vorliegt. Die Duplikation ist von einer sieben Basenpaar langen direkten Sequenzwiederholung flankiert (rote Pfeile) und besitzt an ihrem 5'-Ende zusätzlich eine sieben Basenpaar lange Poly-A Region (grün). Sequenzunterschiede sind durch Punkte gekennzeichnet.

Verlust eines Gens verantwortlich sein. Ebenso wäre ein Genverlust durch inäquales Crossing-over zwischen Schwesterchromatiden theoretisch denkbar. Beide Prozesse erklären jedoch nicht das Vorhandensein der direkten Sequenz-Wiederholungen sowie der Poly-A Region. Demnach kann hier von dem seltenen Fall einer Duplikation durch Retrotransposition ausgegangen werden.

### 4.3.3 Konzertierte Sequenz-Evolution von Hämoglobin-Varianten

Die beiden Gene 7A3 und 7A7, die aus einer Duplikation innerhalb der 7A-Subfamilie hervorgegangen sind, zeigen nur eine durchschnittliche Übereinstimmung der Sequenzen (Nt: 96,7 %, AS: 95,6 %). Das Hämoglobin-Gen 7A7 besitzt hingegen eine sehr hohe Sequenz-Übereinstimmung zu seinem benachbarten Gen 7A8 (Nt: 98,8 %, AS: 98,1 %). Dies läßt die Schlußfolgerung zu, daß hier eine Angleichung der Sequenzen zwischen diesen beiden Genen nach dem Duplikations-Ereignis stattgefunden hat, wobei eine Sequenzübertragung von dem Gen 7A8 auf das benachbarte Gen 7A7 erfolgt sein muß. Um diese Hypothese zu verifizieren und den homogenisierten Sequenzbereich einzugrenzen, wurden die Sequenzen der Hämoglobin-Gene Cte 7A8, 7A7, 7A3 und 7A2 nukleotidweise verglichen. Auch die 5' und 3' flankierenden Regionen wurden in den Vergleich einbezogen, sofern sie ausreichende Sequenz-Übereinstimmungen aufweisen. Hierdurch können die Nukleotidpositionen bestimmt werden, an denen die Gene 7A7 und 7A8 identisch sind, sich jedoch von ihren jeweiligen Duplikationspartnern 7A3 und 7A2 unterscheiden. Diese Positionen sind als das Produkt einer Sequenz-Homogenisierung zwischen den paralogen Genen 7A7 und 7A8 zu werten (Fitch et al., 1990). Das Ende des homogenisierten Bereichs ist dementsprechend durch eine größere Übereinstimmung der aus der Duplikation hervorgegangenen Sequenzen gekennzeichnet. Positionen, an denen drei der vier verglichenen Sequenzen identisch sind, lassen keine Aussage über die Verwandtschaftsverhältnisse zu und wurden deshalb aus dem Vergleich ausgeschlossen.

Der nukleotidweise Vergleich der Gene 7A2, 7A3, 7A7 und 7A8 zeigt folgendes Ergebnis: Die vier Sequenzen weisen innerhalb des kodierenden Bereichs einen hohen Anteil informativer Nukleotidpositionen auf. Von insgesamt 19 Unterschieden sind 15 Positionen informativ und geben über die Sequenz-Verwandtschaft Auskunft. Im kodierenden 5'-Bereich besitzen die Duplikationspartner 7A3 und 7A7 sowie 7A2 und 7A8 eine größere Übereinstimmung zueinander als zu ihren benachbart liegenden Genen. Diese Region ist somit aller Wahrscheinlichkeit nach nicht von einer Sequenz-Homogenisierung zwischen den benachbarten Genen 7A3 und 7A2, sowie 7A7 und 7A8 betroffen (siehe Abb. 4.22). In der zweiten Genhälfte (ab Nukleotidposition 248) finden sich hingegen 12 Positionen, die auf eine Sequenz-Homogenisierung zwischen den Hämoglobin-Genen 7A7 und 7A8 im Anschluß an die Genduplikation hindeuten (Sterne in Abb. 4.22C). Hierbei muß das Gen 7A8 seine Sequenz auf das Gen 7A7 übertragen haben, da letzteres sich von seinem Duplikationspartner 7A3 unterscheidet. Im nicht-kodierenden 3'-Bereich sind die drei Sequenzen über 115 bp nahezu identisch (97,4%), danach bricht die Übereinstimmung der Intergen-Sequenzen der Gene 7A7 und 7A8 plötzlich ab. Die Sequenzen 7A7 und 7A3 zeigen hingegen darüber hinaus nahezu identische Abschnitte. Auch im Anschluß an das inserierte CP1-Element setzt sich die eine für Intergen-Regionen ungewöhnlich hohe Übereinstimmung (88,7 % über 221 bp; 5 Insertionen/Deletionen) fort. Es kann daher gefolgert werden, daß eine Sequenzübertragung, ausgehend von dem Gen 7A7

auf das Gen 7A8, die zweite Genhälfte und möglicherweise einen kurzen Bereich (maximal 115 bp) der 3'-flankierenden Region der beiden benachbarten Gene angeglichen hat. Der 5'-Bereich der Gene ist von diesem Ereignis nicht betroffen (Abb. 4.22 B und C).

Die Analyse der 5'-flankierenden Intergen-Regionen dieser vier 7A-Gene zeigt, daß sich das Gen 7A8 an 20 Positionen von seinem Duplikationspartner 7A2 unterscheidet. An diesen Positionen stimmt es hingegen mit seinem benachbarten Gen 7A7 und dessen Duplikationspartner 7A3 überein (Quadrate in der Abb. 4.22 C). Dieses Ergebnis deutet auf eine zweite Sequenz-Homogenisierung hin, bei der, unabhängig von dem ersten Ereignis, ein Sequenztransfer ausgehend von dem flankierenden 5'-Bereich des 7A7 Gens auf den flankierenden 5'-Bereich des 7A8 Gens stattgefunden haben muß. Wiederum ist der 5'-Bereich der kodierenden Sequenzen von diesem Ereignis aller Wahrscheinlichkeit nach nicht betroffen.

Daß die paralogen Mitglieder repetitiver Sequenz-Familien größere Übereinstimmungen zueinander aufweisen können als zu ihren orthologen Partnern in verwandten Arten, beobachteten Edelman und Gally (1970). Sie folgerten aus diesem Befund, daß die paraloge Mitglieder einer Sequenz-Familie sich in ihrer Evolution gegenseitig beeinflussen können, also einer Koevolution unterliegen. Zimmer bezeichnete dieses Phänomen der „horizontalen Evolution“ mit dem heute gebräuchlichen Begriff der konzertierten Evolution (Zimmer et al., 1980). Zwei Mechanismen sind für die Angleichung repetitiver Sequenzen von besonderer Bedeutung. Beide beruhen auf Rekombinations-Ereignissen während der Meiose oder Mitose und können zum Austausch von Sequenzen innerhalb einer Chromatide, zwischen den Schwesterchromatiden, zwischen homologen Chromosomen oder auch zwischen verschiedenen Chromosomen führen. Der Mechanismus des inäqualen Crossing-over resultiert hierbei in einem reziproken Sequenztransfer, da er eine Duplikation in einem der beiden Rekombinationspartner und eine Deletion in dem anderen erzeugt. Die Fixierung aufeinander folgender Crossing-over-Ereignisse führt schließlich zu Multigen-Familien mit fluktuierender Größe, deren paraloge Mitglieder eine hohe Sequenz-Übereinstimmung aufweisen können (Zimmer et al., 1980; Smith, 1976). Da dieser als „cross over fixation“ bezeichnete Mechanismus nur bei gleichsinnig orientierten Genen stattfinden kann und stets zu einer Deletion und Insertion von Genen in den Rekombinationspartnern führt, ist er als Erklärung für die konzertierte Evolution der Hämoglobin-Gene Cte7A7 und Cte7A8 ungeeignet. Der zweite wichtige Mechanismus zur Homogenisierung von Sequenzen ist die sogenannte Genkonversion (Slightom et al., 1980; Smithies und Powers, 1986). Auch die Genkonversion ist mechanistisch eng an die Rekombination zwischen DNA-Molekülen gekoppelt. Sie resultiert aus einem Crossing-over zwischen ähnlichen, jedoch nicht identischen Sequenzen. Die anschließende Reparatur der entstandenen Heteroduplex-DNA führt zur Überschreibung des einen Stranges mit der Sequenz des anderen (Bollag et al. (1992)). Die Richtung der Korrektur ist hierbei vermutlich zufällig. Der Mechanismus der Genkonversion ist für die Hämoglobin-Gene 7A7 und 7A8 schematisch

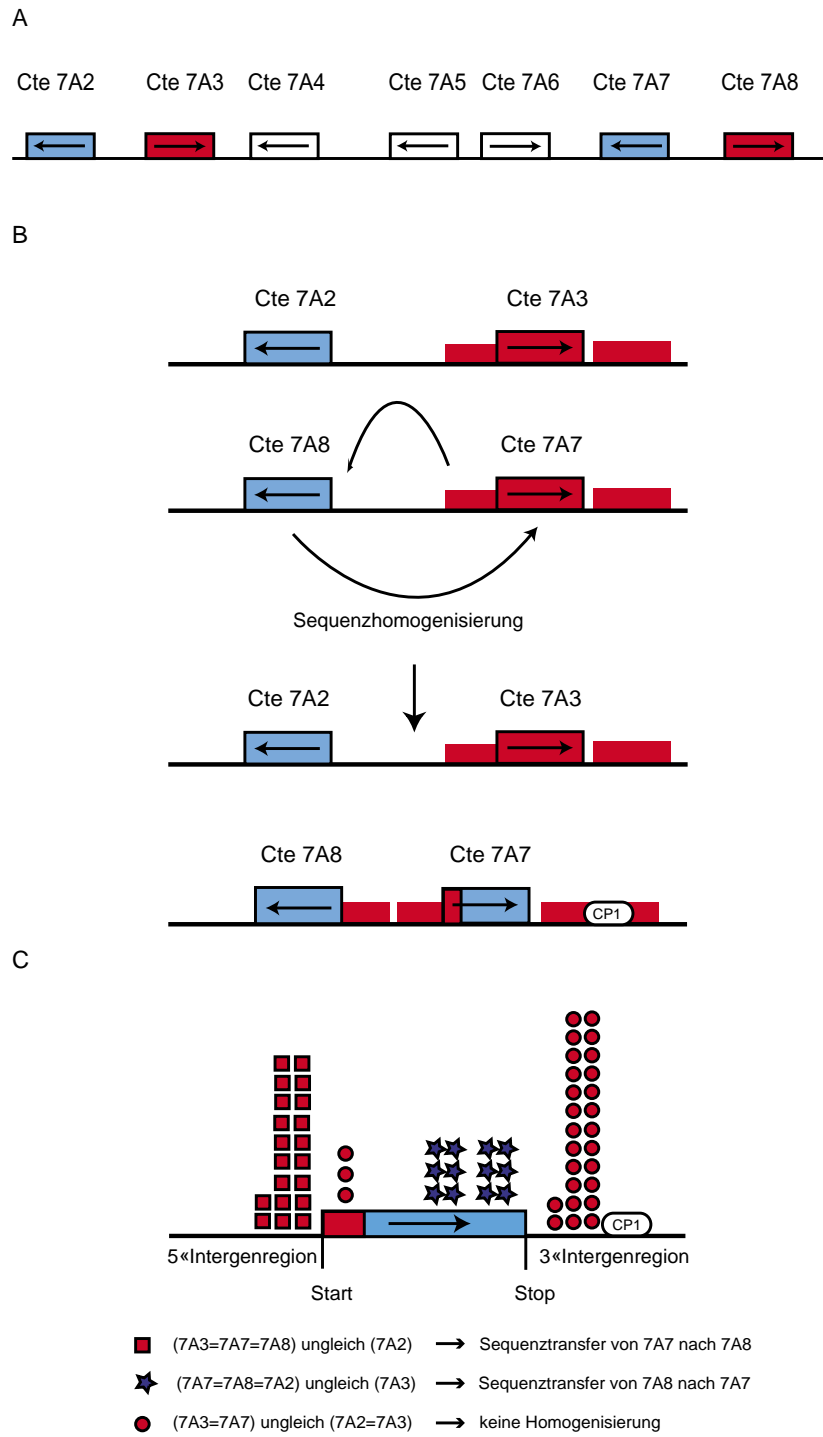


Abbildung 4.22 Legende siehe folgende Seite.

**Abbildung 4.22: Sequenz-Homogenisierung innerhalb der 7A-Gen-Subfamilie**

Die Hämoglobin-Gene Cte 7A2 und Cte 7A8 sind durch Duplikation aus einem gemeinsamen Vorläufer-Gen entstanden. Das Gleiche gilt auch für die Gene Cte 7A3 und Cte 7A7. Die Abbildung A zeigt die Verhältnisse nach diesem Duplikationsereignis. Sequenz-Verwandtschaften sind durch identische Farben gekennzeichnet. Durch zwei Sequenzübertragungen zwischen den benachbarten Genen Cte 7A8 und Cte 7A7 wurden ihre Sequenzen aneinander angeglichen (B). Die Richtung der Sequenzübertragung ist jeweils durch einen gebogenen Pfeil angegeben.

Die Abbildung C symbolisiert das Alignment aller vier Gen-Sequenzen und ihrer flankierenden Regionen. Durch den nukleotidweisen Vergleich konnten homogenisierte Sequenzbereiche ermittelt werden. Unterschiede zwischen den Nukleotid-Sequenzen wurden in drei Klassen eingeordnet, die Auskunft über die Sequenz-Verwandtschaften zwischen den Genen geben (Quadrat, Kreis und Stern). So kennzeichnet beispielsweise ein Stern eine Position innerhalb des Alignments, an der die Gene Cte 7A7, Cte 7A8 und Cte 7A2 eine identische Base besitzen, das Gen Cte 7A3 jedoch einen Unterschied aufweist. Eine Übereinstimmung des Gens 7A7 mit seinem benachbarten Gen 7A8, kombiniert mit einem Unterschied zu seinem Duplikationspartner 7A3, ist als Zeichen für eine Sequenz-Homogenisierung zu werten. Die Basen-Unterschiede wurden jeweils über 50 bp summiert und in einer Säule zusammengefaßt.

in Abbildung 4.23 gezeigt. Da die hier beschriebenen Konversions-Ereignisse nicht auf die kodierenden Regionen der Gene beschränkt sind, sondern auch die nicht-transkribierten 5'- und 3'-Bereiche umfassen, scheint die Möglichkeit, daß ein RNA- bzw. cDNA-Intermediat an dem Prozeß beteiligt war, ausgeschlossen zu sein. Die Rekombination zwischen einer chromosomalen Genkopie mit einer extrachromosomalen cDNA ist jedoch theoretisch denkbar und konnte bei Hefe nachgewiesen werden (Melamed et al., 1992; Derr und Strathern, 1992).

Genkonversionen sind in der Regel auf Bereiche mit guter Sequenz-Übereinstimmung begrenzt. Mitglieder von Genfamilien weisen häufig nur leicht unterschiedliche Sequenzen auf und sind somit ein bevorzugter Angriffspunkt für Konversions-Ereignisse. So sind beispielsweise bei den  $\alpha$ -Amylase Genen von *Drosophila* ausschließlich die kodierende Sequenz sowie unmittelbar flankierende nichtkodierende Bereiche betroffen (Hickey et al., 1991). Bei den Globin-Genen von *Caenorhabditis* sind Konversionen sogar auf die Exons beschränkt (Kloek et al., 1996). Sequenz-Unterschiede unterdrücken hingegen eine effiziente Rekombination. Die hohe Sequenz-Übereinstimmung der Gene Cte 7A7 und Cte 7A8, kombiniert mit ihrer inversen Orientierung und ihrem geringen Abstand, könnte eine Paarung ihrer Sequenzen zusätzlich begünstigt haben. Ein Crossing-over und die Reparatur der Basenfehlpaarungen führte anschließend zur Genkonversion. Die hohe Sequenz-Identität der resultierenden Gene könnte nun wiederum ein oder mehrere weitere Konversions-Ereignisse unterstützt haben. Daß kurze Abstände und invertierte Genanordnung in einer effi-

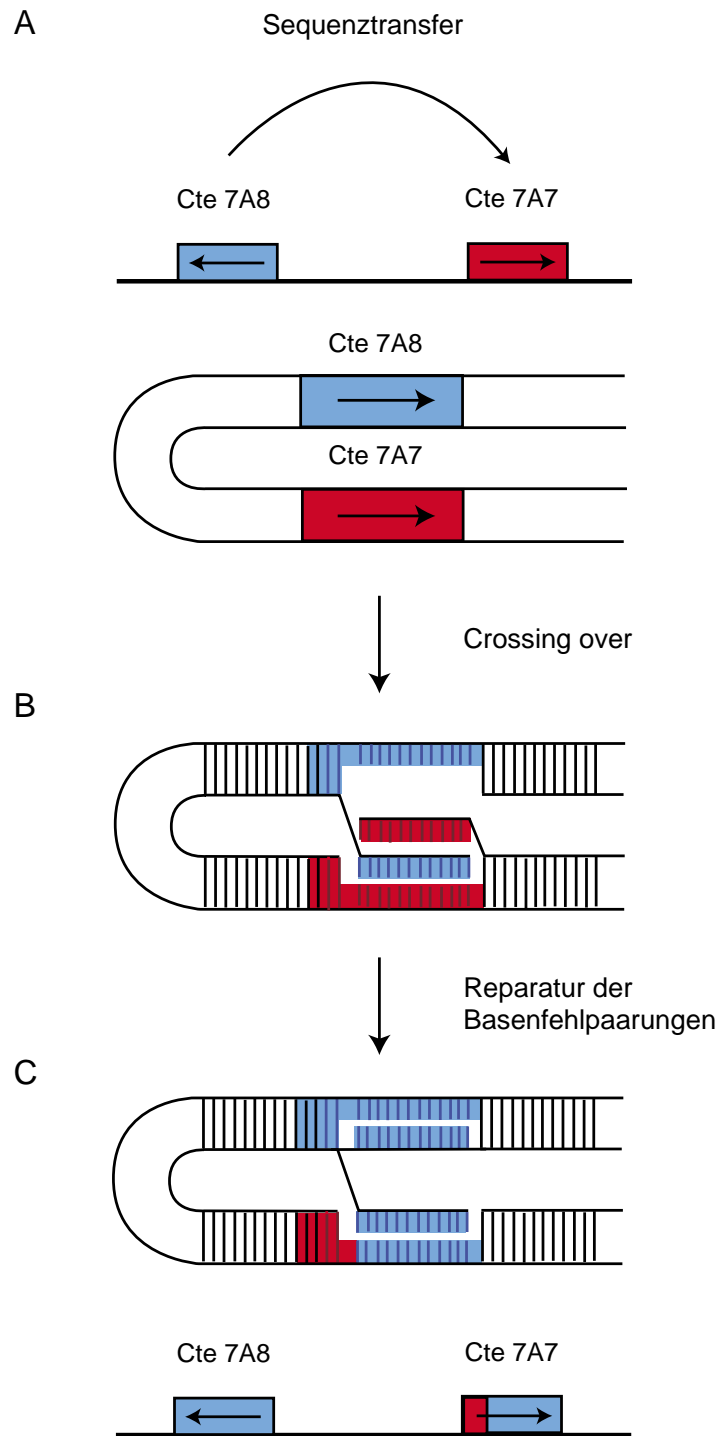


Abbildung 4.23 Legende siehe folgende Seite

##### Abbildung 4.23: Der Mechanismus der Genkonversion

Die Darstellung zeigt schematisch den Sequenztransfer ausgehend von dem Hämoglobin-Gen Cte 7A8 auf das benachbarte Gen Cte 7A7. Aufgrund der hohen Sequenz-Übereinstimmung und gegensinnigen Orientierung der beiden Gene erfolgt ihre Paarung innerhalb der Chromatide (A). Nach einem Einzelstrangbruch kann ein Strangaustausch zwischen den beiden Rekombinationspartnern stattfinden. Dieser resultiert in der Ausbildung einer Heteroduplex-DNA (B). Durch Entfernen eines der beiden Stränge und Schließen der entstandenen Lücke mittels DNA-Neusynthese werden Basenfehlpaarungen innerhalb des Heteroduplex-Bereichs eliminiert. Die Richtung der Sequenzübertragung ist hierbei zufällig (C). Durch diesen Reparaturmechanismus werden innerhalb des rekombinierten Bereichs die Sequenzen der beiden Gene aneinander angeglichen. Das Modell basiert in seinen Grundzügen auf dem von Robin Holliday (1964) postulierten Mechanismus der Rekombination zwischen homologen DNA-Molekülen.

zienten Homogenisation resultieren, zeigte sich bereits am Beispiel der Hämoglobin-Gene von *C. pallidivittatus* (Hankeln et al., 1998). Auch für die Homogenisierung der  $\alpha$ -Amylase Gene von *Drosophila*, die, anders als die Hämoglobin-Gene, durch einige tausend Basenpaare getrennt sind, soll die invertierte Orientierung verantwortlich sein (Hickey et al., 1991). Gen-Subfamilien wie die der 7A-Hämoglobin-Gene von *C. tentans* bieten unter diesen Aspekten das ideale Substrat für Genkonversionen und somit für eine konzertierte Evolution. Nicht immer resultieren Genkonversionen in einer Angleichung von Genen. Insbesondere beim Transfer kurzer Sequenzabschnitte kann die Diversität von Mitgliedern einer Sequenzfamilie erhöht werden (Baltimore, 1981). So entstehen beispielsweise in der Tomate durch Genkonversion neue Varianten von Genen, die eine Resistenz gegenüber Pilzen vermitteln (Parniske et al., 1997). In jedem Falle überlagern und verdecken derartige Prozesse diejenigen Unterschiede, die sich zwischen Genkopien durch zufallsgemäße Mutation anhäufen. Eine Interpretation der Verwandtschaftsverhältnisse zwischen Genen wird somit erschwert und ist nicht selten nur durch Einbeziehung weiterer orthologer Gene aus verwandten Spezies möglich.

#### 4.4 Vergleichende Analyse von Hämoglobin-Gengruppen in unterschiedlichen Chironomiden-Spezies

Neben dem Vergleich paraloger Sequenzen ermöglicht es häufig erst die vergleichende Analyse von Hämoglobin-Gengruppen aus verschiedenen Chironomiden-Arten, die Evolution dieser Multigenfamilie besser zu verstehen. Ausgehend von den bislang erhobenen Daten erscheint es sicher, daß die schrittweise erfolgten Genduplikationen mit einer anschließenden Diversifizierung der Gene der wichtigste Prozeß bei der

Entstehung der Hämoglobin-Genfamilie der Chironomiden ist. Ungleiches Crossing-over sollte dementsprechend zu Unterschieden im Genbestand und der Anordnung der Gene in unterschiedlichen Arten führen. Durch die klare Zuordnung der orthologen Gene in mehreren Spezies (siehe Kapitel 4.1) läßt sich das evolutionäre Schicksal einzelner Gen-Varianten während der Speziation verfolgen. Für die Gen-Gruppe D liegen Daten aus den Arten *C. thummi*, *C. piger* und *C. melanotus* vor, die größere zusammenhängende Cluster-Bereiche abdecken und zu Sequenz-Vergleichen mit *C. tentans* herangezogen werden sollen (Hankeln et al. (1998); Rozynek et al. (1991); Trewitt et al. (1995); Gruhl et al. (1997), Hankeln in Vorbereitung).

### 4.4.1 Die VIIB-Region der Arten *C. tentans*, *C. thummi* und *C. piger*

In den drei Chironomus-Arten *C. tentans*, *C. thummi* und *C. piger* liegen die orthologen Bereiche zwischen den Genen Y und 7B10 vollständig sequenziert vor (Trewitt et al., 1995; Hankeln und Schmidt, 1991). In Abb. 4.24 ist ihre genomische Organisation und der Genbestand im Vergleich gezeigt. Um die verwandtschaftlichen Beziehungen der Gene zueinander darzustellen, wurde ein MP-Baum berechnet (siehe Abb. 4.25). Die Anordnung der Hämoglobin-Gene innerhalb der Region zeigt, daß in allen drei untersuchten Arten die „single copy“-Gene räumlich getrennt von der 7B-Gen-Subfamilie liegen. Auch weisen „single copy“-Gene und die Mitglieder der 7B-Subfamilie eine Reihe sehr unterschiedlicher Charakteristika auf.

Die „single copy“-Gene sind zwischen den Arten sehr gut konserviert: orthologe Gen-Varianten sind einander ähnlich und weisen deutliche Unterschiede zu paralogen Hämoglobin-Genen auf. Im MP-Baum bilden die orthologen Varianten daher definierte Gruppen. Auch die Clusterstruktur und räumliche Ausdehnung ist im Bereich der singulären Varianten gut erhalten.

Ganz andere Verhältnisse sind innerhalb der Subfamilie zu finden. Die Sequenzen der paralogen Genkopien sind einander so ähnlich, daß es in der Regel nicht möglich ist, ihre orthologen Partner aus den anderen Spezies zu identifizieren. Im MP-Baum bilden die paralogen Varianten daher Gruppen. Nur das Gen 7B7 ist zwischen allen drei Arten konserviert und kann aufgrund seiner Sequenz den orthologen Partnern zugeordnet werden. Erst der Vergleich der Intergen-Regionen ermöglichte es, im Bereich der 7B-Subfamilie orthologe Sequenz-Abschnitte zwischen den drei Arten zu identifizieren. Ein solcher Befund war insofern unerwartet, als die schnell evolvierenden Intergen-Regionen keine signifikanten Übereinstimmungen zwischen den nur entfernt verwandten Arten *C. tentans* und *C. thummi* aufweisen sollten. Innerhalb der Region der „single copy“-Gene beschränken sich die Übereinstimmungen der Intergen-Regionen auch erwartungsgemäß auf die Promotorregionen und die 3' nicht-translatierten Bereiche der Gene. Die exakten Grenzen der orthologen Intergen-Regionen, sowie der Grad der Übereinstimmung ist im Anhang gezeigt. Die Abbildung 4.24 gibt das Ergebnis schematisch wieder. Aufgrund der Intergen-Regionen ist es demnach möglich, die 7B-Varianten aus *C. tentans* ihren orthologen



Genen aus *C. thummi* und *C. piger* zuzuordnen. Es zeigt sich, daß alle in *C. thummi* beschriebenen 7B-Gene auch in *C. tentans* vorhanden sind und in identischer Reihenfolge vorliegen. Diese Clusterstruktur spiegelt daher mit großer Wahrscheinlichkeit den ursprünglichen Zustand wider. Unterschiede in der Clusterstruktur der Art *C. piger* sind demnach abgeleitet. Bereits 1991 konnte gezeigt werden, daß *C. piger* zwei 7B-Gene durch eine Deletion verloren haben muß (Hankeln und Schmidt, 1991; Trewitt et al., 1995). Weiterhin bedeutet dieser Befund, daß die intraspezifische Ähnlichkeit zwischen Genen als starkes Indiz für eine innerartliche Sequenzangleichung gewertet werden muß. Wie bereits aus dem Vergleich der paralogenen Gene hervorging, sind es die Gen-Subfamilien, die in einem besonderen Maße der Sequenz-Homogenisierung unterliegen. Die Intergen-Regionen evolvieren offenbar nicht konzentriert. Ein möglicher Grund hierfür mag sein, daß die wahrscheinlich funktionslosen Intergen-Regionen nach einer Genduplikation vergleichsweise rasch ein initiales Maß an Sequenz-Unterschieden anhäufen, das sie effizient vor Angleichungsprozessen bewahrt. Ähnliches könnte auf die singulären Gen-Varianten zutreffen, die offenbar ebenfalls von Homogenisierung weitgehend unbetroffen bleiben. Man könnte spekulieren, daß diese Gene nach ihrer Duplikation möglicherweise einer Selektion auf adaptive Veränderungen hin unterlagen, die einer Homogenisierung entgegen wirkte.

Zusätzlich zu den 7B-Varianten von *C. thummi* finden sich in *C. tentans* die Gene 7B12 und 7B13. Sie besitzen eine selbst für Mitglieder einer Gen-Subfamilie überdurchschnittlich hohe Sequenz-Identität von 99,4 %. Die flankierenden Intergen-Regionen dieser Gene zeigen eine Sequenz-Übereinstimmung mit der Region zwischen den Genen Ctt Z und Ctt 7B7. Dies könnte als Hinweis auf eine nachträgliche Entstehung dieser beiden Gene in *C. tentans* nach Aufspaltung der Arten gewertet werden. Auch eine Deletion im gemeinsamen Vorfahren von *C. tentans* und *C. thummi* könnte diese Beobachtung erklären.

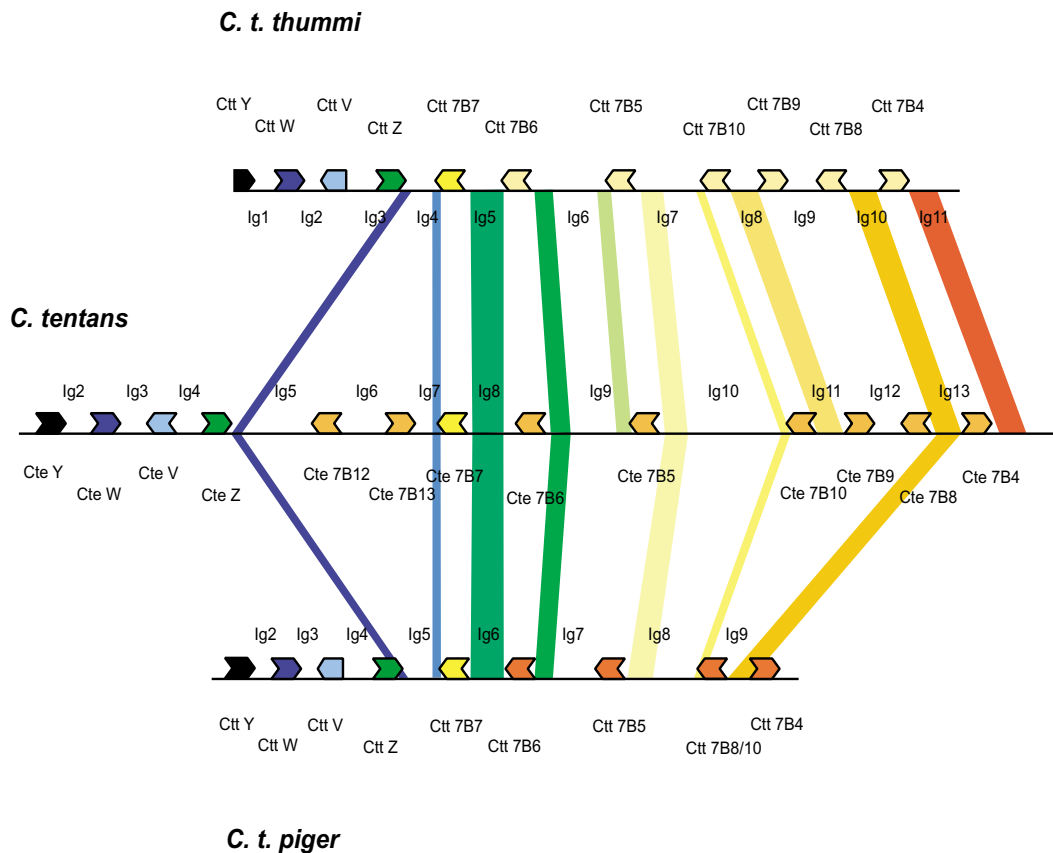
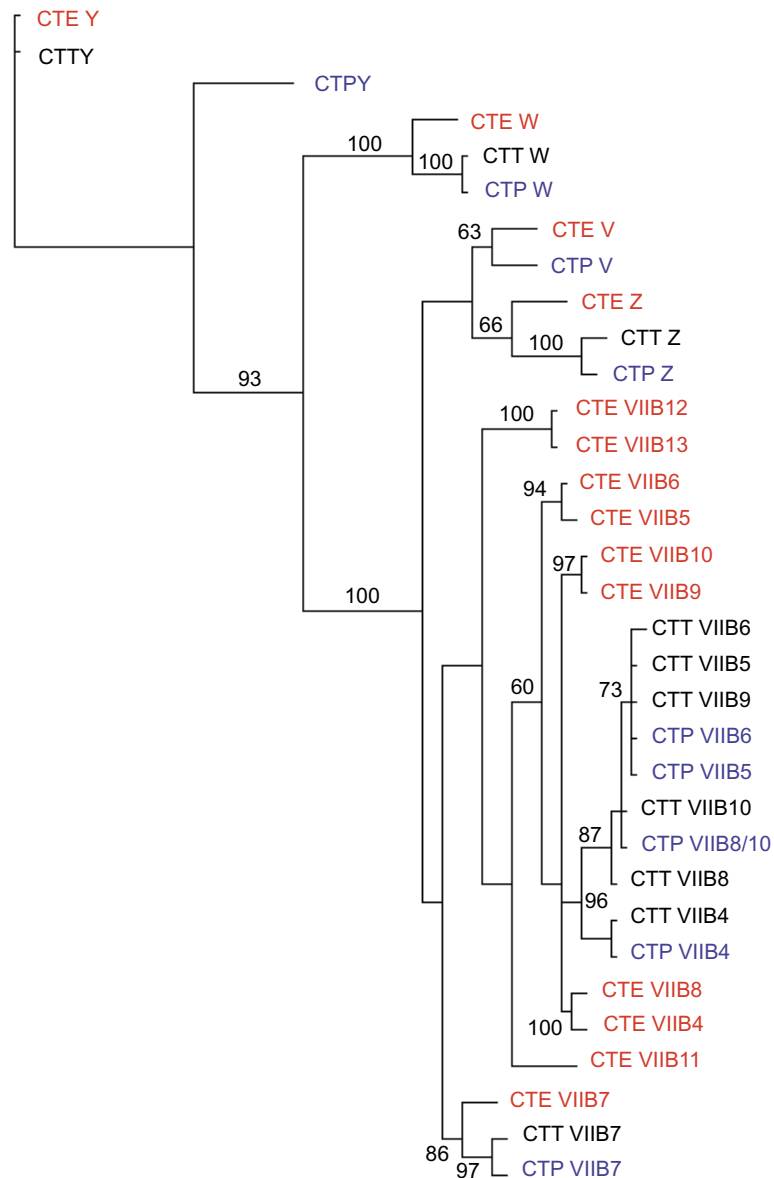


Abbildung 4.24: Organisation der 7B-Region im Vergleich

Aus den Spezies *C. tentans*, *C. thummi* (Trewitt et al., 1995) und *C. piger* (Hankeln und Schmidt, 1991) kann ein orthologer Bereich verglichen werden, der die Gene Y-7B10 umfaßt. In allen drei Spezies zeigen die 7B-Gene eine große innerartliche Sequenz-Übereinstimmung. Sie konnten nur auf Grundlage der Intergen-Regionen ihren orthologen Verwandten zugeordnet werden. Sequenz-Übereinstimmungen sind durch ähnliche Farben gekennzeichnet. Der Vergleich der Intergen-Regionen zeigt auch, daß die Gene der 7B-Subfamilie starken Rearrangements seit der Trennung der Arten unterlagen. Selbst die Schwester-Arten *C. thummi* und *C. piger* weisen tiefgreifende Unterschiede in der Anzahl und Anordnung der einzelnen 7B-Varianten auf. Im Gegensatz dazu sind die in nur einer Kopie vorliegenden Hämoglobin-Gene Y, W, V, und Z in ihrer Position stabil. Die orthologen „single copy“-Gene sind einander ähnlicher als die paralogen Varianten.



**Abbildung 4.25: Phylogenetischer Baum orthologer Hämoglobine aus der 7B-Region**  
 Basierend auf dem Vergleich der Aminosäure-Sequenzen wurde ein MP-Baum für die Spezies *C. tentans*, *C. thummi* und *C. piger* berechnet. Alle „single copy“-Varianten werden ihren orthologen Partnern zugeordnet. Diese Verzweigungen werden mit signifikanten Bootstrap-Werten unterstützt (500 Replikationen). Innerhalb der 7B-Subfamilie hingegen bilden die paralogenen Hämoglobine aus *C. tentans* eine separate Gruppe. Einzig die Variante 7B7 kann ihren orthologen Partnern aus *C. thummi* und *C. piger* zugeordnet werden. Die Hämoglobine aus den drei Arten sind durch unterschiedliche Farben gekennzeichnet.

#### 4.4.2 Die IIB-Region der Arten *C. tentans*, *C. thummi* und *C. melanotus*

In den drei Chironomus-Arten liegen orthologe Sequenzen aus dem Cluster-Bereich vor, der die Gene 8 bis 7A1 umfaßt (Gruhl et al. (1997), Hankeln, Güttler und Schmidt in Vorbereitung). Die Abb. 4.26 zeigt die genomische Organisation, sowie den Genbestand dieser Region im Interspezies-Vergleich. Um die verwandtschaftlichen Beziehungen der Gene zueinander darzustellen, wurde ein MP-Baum auf Grundlage der Aminosäure-Sequenzen berechnet (siehe Abb. 4.27). Es ist anhand dieses Stammbaumes möglich, die orthologen Gene zu identifizieren.

In *C. thummi* wurden die Gene Ctt 11, Ctt 12, Ctt 13, Ctt 2 $\beta$ , Ctt 9.1 und Ctt 7A beschrieben. Bei dem Hämoglobin-Gen Ctt 13 handelt es sich um ein Pseudogen, welches vermutlich durch die Insertion eines Transposons inaktiviert wurde (Gruhl et al., 1997). In *C. melanotus* sind die Gene Cme 8, Cme ORF7, Cme ORF5, Cme ORF2, Cme ORFB und Cme 2 $\beta$  bekannt. Die Sequenz des Gens Cme 9 liegt nur teilweise vor (Hankeln, Güttler und Schmidt, in Vorbereitung). Alle diese Gene besitzen mindestens einen orthologen Partner in *C. tentans*. Die Gene Ctt 11, Cme ORF2 und Cte ORF B bilden eine Verwandtschaftsgruppe. Die nahe Sequenz-Verwandtschaft der Gene Ctt 11 und Ctt 12 zu dem Pseudogen ORF B von *C. tentans* wurde bereits 1997 von Gruhl und Mitarbeitern beobachtet. Sie folgerten hieraus eine Orthologie dieser Gene. Die Anordnung der Gene innerhalb des Genclusters in umgekehrter Reihenfolge sollte durch ein Rekombinations-Ereignis erklärt werden. Zusätzlich zu dem ORF B besitzt *C. tentans* jedoch drei weitere Gene, welche eine signifikante Sequenz-Übereinstimmung zu den Hämoglobinen 11 und 12 von *C. thummi* aufweisen. Im Gegensatz zu dem ORF B stimmt ihre Anordnung und Orientierung innerhalb der Gengruppe mit den Genen Ctt 11 und Ctt12 überein. Sowohl die Clusterstruktur, als auch die Sequenz-Übereinstimmung sprechen somit dafür, daß diese Gene in *C. tentans* und *C. thummi* ortholog sind. Sie wurden dementsprechend in *C. tentans* als Cte 11, Cte 12-1 und Cte 12-2 bezeichnet. Auch bei dem als ORF B beschriebenen Gen aus *C. melanotus* scheint es sich um eine orthologe Variante der Gene Cte 12 und Ctt 12 zu handeln. In Übereinstimmung mit den Ergebnissen von Gruhl (1997) ist das Gen Cte ORF A und das Pseudogen Ctt 13 auf einen gemeinsamen Vorfahren zurückzuführen.

Der unterschiedliche Umfang der orthologen Regionen in den vier Spezies läßt bereits vermuten, daß Rearrangements der Genanordnung während der Speziation nicht selten sein dürften. In der Tat konnten in bislang allen ausgewerteten Interspezies-Vergleichen Unterschiede in der Zahl und Anordnung von Hämoglobin-Genen beobachtet werden (Hankeln und Schmidt, 1991; Hankeln et al., 1998; Trewitt et al., 1995). Auch die hier gezeigten Vergleiche belegen, daß Genvermehrung und -verlust ein charakteristisches Merkmal für die Evolution der Genfamilie ist. Dies hat dazu geführt, daß jede Zuckmücken-Spezies einen individuellen, charakteristischen Bestand an Hämoglobin-Genen aufweist. Ein Beispiel für eine auf die Art *C. tentans*

**C. tentans**

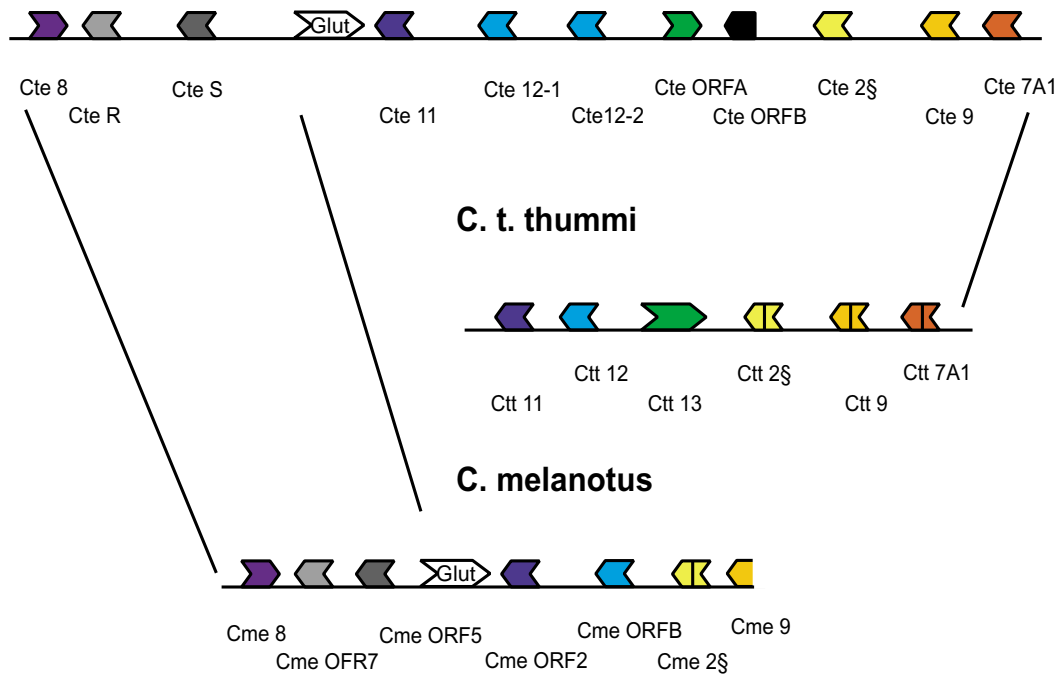
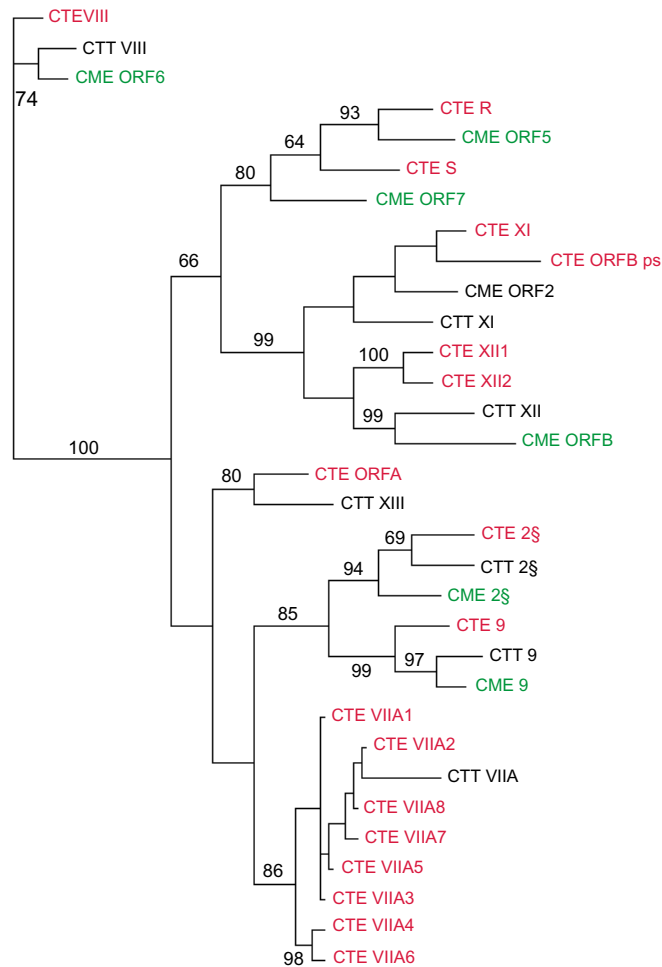


Abbildung 4.26: Die Organisation der IIB-Region im Vergleich

Aus den Spezies *C. tentans*, *C. thummi* (Gruhl et al., 1997) und *C. melanotus* (Hankeln, Güttler und Schmidt in Vorbereitung) kann ein orthologer Bereich verglichen werden. Dieser umfaßt die Gene 8-7A1. Orthologe Gene sind durch identische Farben kenntlich gemacht. Die drei Spezies weisen ein leicht unterschiedliches Repertoire an Hämoglobin-Genen auf: Während das Gen 12 in *C. tentans* dupliziert vorliegt, fehlt den Spezies *C. thummi* und *C. melanotus* aller Voraussicht nach der ORF B. Ein orthologes Gen der Variante Ctt13/CteORFA ist in *C. melanotus* nicht vorhanden. Bei den Genen Cte ORF B und Ctt 13 handelt es sich um Pseudogene. Die Anordnung der orthologen Gene ist während der Speziation erhalten geblieben. Die Verteilung von Introns ist unterschiedlich in den drei Spezies. Sie sind durch einen Balken im Gensymbol dargestellt. Das Gen für ein Glukose-Transportprotein (Glut) ist an identischen Positionen im Genom von *C. tentans* und *C. melanotus* lokalisiert.

beschränkte rezente Duplikation sind die Gene Cte 12-1 und Cte 12-2, die sowohl in *C. thummi* als auch in *C. melanotus* in nur einer Kopie vorhanden sind. Das orthologe Gen zu dem Genpaar Cte ORFA und Ctt 13 ist in *C. melanotus* offensichtlich verloren gegangen. Möglicherweise sind Crossing over-induzierte Schwankungen der Kopienzahl besonders häufig in Gen-Subfamilien, die mit ihrer Anzahl benachbarter, größtenteils sehr ähnlicher Genkopien ein bevorzugtes „Substrat“ für derartige Ereignisse darstellen sollten (Hankeln et al., 1998). Ein Beispiel für die Entstehung



**Abbildung 4.27: Phylogenetischer Baum orthologer Hämoglobine aus der II $\beta$ -Region**  
 Auf Grundlage der Aminosäure-Sequenzen wurde ein MP-Baum für Arten *C. tentans*, *C. thummi* und *C. melanotus* berechnet. Orthologe Gene aus der II $\beta$ -Region können hierdurch identifiziert werden. Ein Bootstrap-Test wurde mit 500 Replikationen durchgeführt. Die kalkulierten Werte sind an den Ästen vermerkt. Die Hämoglobine aus den drei Arten sind durch unterschiedliche Farben gekennzeichnet.

von drei neuen Genen durch Rekombination innerhalb einer Subfamilie sind die 7A-Varianten in *C. tentans*. Die sogenannten „single copy“-Hämoglobin-Gen-Varianten sind im Gegensatz dazu im interspezifischen Vergleich deutlich stärker in ihrer Anordnung konserviert.

## 4.5 Das evolutionäre „Schicksal“ von Genduplikaten

Innerhalb der Säugetiere weisen unterschiedliche Hämoglobine einen unterschiedlichen Konservierungsgrad zwischen verschiedenen Arten auf. Es gibt hoch konservierte Hämoglobin-Varianten, wie beispielsweise das embryonale  $\epsilon 2$  und variable Formen, wie die  $\beta$ -Varianten (Koop und Goodman, 1988; Cooper und Hope, 1993). Auch innerhalb der Chironomiden konnte dieses Phänomen beobachtet werden. So weist die Variante E einen weit überdurchschnittlich hohen Grad der Konservierung auf Aminosäure-Ebene zwischen den Arten *C. pallidivittatus* und *C. thummi* auf, während das Hämoglobin IV größeren Sequenz-Veränderungen unterlag (Hankeln et al., 1998). Die Anzahl der Basensubstitutionen zwischen zwei orthologen Genen kann als Maß für die Evolutionsgeschwindigkeit angesehen werden. Um den evolutionären Abstand zwischen Genen zu beschreiben, ist es erforderlich, zwischen synonymen Basenaustauschen, die keine Auswirkungen auf die Aminosäure-Sequenz haben und nicht-synonymen Substitutionen, die zu dem Austausch einer Aminosäure führen, zu unterscheiden. In vielen Genen häufen sich synonyme Substitutionen proportional zur Zeit an und dienen daher als Referenzwert für den evolutionären Abstand zweier Gene. Im Gegensatz hierzu unterliegen die nicht-synonymen Substitutionen dem Einfluß der Selektion. Für die Funktion des Proteins nachteilige Mutationen werden eliminiert, während adaptiv wertvolle Mutationen unter dem Einfluß der „positiven“ Selektion in der Population fixiert werden können. Das Verhältnis der nicht-synonymen Substitutionsrate  $K_A$  zur synonymen Substitutionsrate  $K_S$  reflektiert somit den Einfluß der Selektion beim Vergleich zweier Gene. Um die evolutionären Kräfte zu untersuchen, die auf die Mitglieder der Hämoglobin-Genfamilie einwirken, wurden die Raten von nicht-synonymen Substitutionen und synonymen Substitutionen zwischen orthologen Genen bestimmt. Die errechneten Werte wurden nach einem statistischen Modell korrigiert, um das Auftreten multipler Basenaustausche während der Evolution zu berücksichtigen (Jukes und Cantor, 1969). Die Tabelle 4.7 zeigt die evolutionären Raten der Chironomiden-Hämoglobine. Sie wurden für 16 orthologe „single copy“-Gene und die 7B-Gen-Subfamilie getrennt betrachtet.

Unter der Voraussetzung, daß die synonymen Substitutionen weitgehend neutral sind, d.h. nicht den Einflüssen der Selektion unterliegen (Wolfe et al., 1989; Li, 1997), sollten sie sich proportional zur Zeit anhäufen. Die Höhe des  $K_S$ -Wertes sollte demnach durch den Zeitpunkt der Trennung der Gene und ihrer Evolutionsrate bestimmt werden. Synonyme Substitutionen werden daher häufig als Maß für eine molekulare Uhr verwendet und können Auskunft über den Zeitpunkt geben, an dem Arten sich voneinander getrennt haben (Überblick in Ayala (1997)). Auch die Hämoglobin-Gene der Chironomiden wurden als molekulare Marker für die Speziation verwendet. Die Beispiele zeigen jedoch, daß die auf Grundlage der  $K_S$  Werte ermittelten evolutionärer Abstände zwischen den untersuchten Genen stark schwanken. So wurde anhand der Hämoglobin-Gene 2 $\beta$  und 9 eine getrennte Entwicklung

A	KA	KA Variation um Faktor	KA Durchschnitts-Wert
single copy-Gene <i>C. tentans</i> / <i>C. thummi</i>	0,119-0,139 W: 0,063	0,7 incl. W: 2,2	0,131 incl. W: 0,225
single copy-Gene <i>C. tentans</i> / <i>C. melanotus</i>	0,094-0,196 8:00,1	2,09 incl. 8: 3,37	0,147 incl. 8: 0,137
7B-Subfamilie <i>C. tentans</i> / <i>C. thummi</i>	0,06-0,104	1,73	0,08
7B-Subfamilie paraloge Gene (homogenisiert)	0,003-0,016	5,33	0,007

B	KS	KS Variation um Faktor	KS Durchschnitts-Wert
single copy-Gene <i>C. tentans</i> / <i>C. thummi</i>	0,381-0,856	2,25	0,644
single copy-Gene <i>C. tentans</i> / <i>C. melanotus</i>	0,511-1,019	2,02	0,743
7B-Subfamilie <i>C. tentans</i> / <i>C. thummi</i>	0,360-0,672	1,87	0,44
7B-Subfamilie paraloge Gene (homogenisiert)	0,008-0,05	6,25	0,027

C	KA/KS incl. der Gene W und 8	KA/KS Variation um Faktor	KA/KS Durchschnitts-Wert
single copy-Gene <i>C. tentans</i> / <i>C. thummi</i>	0,094-0,336	3,57	0,214 ohne Gen W (incl. W um Faktor 2,3 kleiner)
single copy-Gene <i>C. tentans</i> / <i>C. melanotus</i>	0,087-0,275	3,16	0,202 ohne Gen 8 (incl. 8 um Faktor 2,3 kleiner)
7B-Subfamilie <i>C. tentans</i> / <i>C. thummi</i>	0,097-0,245	2,53	0,192
7B-Subfamilie paraloge Gene (homogenisiert)	0,088-0,375	4,26	0,284

Tabelle 4.7: Evolutionäre Raten der Hämoglobin-Gene des Clusters D

Die Anzahl der synonymen-, sowie der nicht-synonymen Austausche wurde aus dem Vergleich von 8 orthologen single copy-Genen der Arten *C. tentans* und *C. thummi*, sowie 8 orthologen single copy-Genen der Arten *C. tentans* und *C. melanotus* ermittelt. Innerhalb der 7B-Subfamilie konnten 7 orthologe Gene von *C. tentans* und *C. thummi* verglichen werden (A und B). Die KA-Werte der Hämoglobin-Gene W und 8 sind gesondert aufgeführt, da sie signifikant von dem Durchschnittswert aller anderen Gene abweichen. Der Grad der Konservierung der Hämoglobine auf Proteinebene läßt sich aus dem Verhältnis KA/KS ersehen (C). Werte  $\ll 1$  deuten darauf hin, daß die Gene unter dem Einfluß der reinigenden Selektion stehen und an ihre Funktion sehr gut angepaßt sind.



der Arten *C. thummi* und *C. tentans* über einen Zeitraum von 60 Millionen Jahren postuliert (Kao et al., 1994). Hankeln und Mitarbeiter postulierten hingegen eine Trennung dieser Verwandtschafts-Gruppen vor nur 30 Millionen Jahren, auf Grundlage der Hämoglobin-Genfamilie E. Die hier vorliegenden Daten resultieren aus dem Vergleich von acht orthologen „single copy“-Genen der Arten *C. tentans* und *C. thummi*, sowie acht orthologen „single copy“-Genen der Arten *C. tentans* und *C. melanotus*. Innerhalb der 7B-Subfamilie können 7 orthologe Gene von *C. tentans* und *C. thummi* verglichen werden. Sowohl innerhalb der „single copy“-Gene, als auch innerhalb der 7B-Subfamilie variieren die errechneten KS-Werte um den Faktor 2 (siehe Tab. 4.7a). Die synonymen Substitutionen scheinen sich entgegen der Voraussage nicht streng proportional zur Zeit anzuhäufen.

Die paralogen 7B-Gene von *C. tentans* weisen in der Regel eine größere Übereinstimmung zueinander auf als zu ihren orthologen Partnern in *C. thummi*. Hierfür verantwortlich sind vermutlich Sequenz-Konversionen innerhalb der 7B-Genfamilie von *C. tentans*, die zu einer Homogenisierung der Gene 7B5/7B6, 7B9/7B10 und 7B4/7B8 geführt haben. Die Tatsache, daß die für diese Gene ermittelten KS-Werte um den Faktor 6 schwanken, läßt vermuten, daß diese Ereignisse zu unterschiedlichen Zeiten während der Evolution der 7B-Subfamilie in *C. tentans* stattfanden. Dies zeigt, daß interchromosomale Rekombinations-Ereignisse, wie beispielsweise Genkonversionen, den tatsächlichen evolutionären Zeitpunkt der Trennung zweier Gene überdecken können. Jedoch nicht nur Mitglieder von Gen-Subfamilien, sondern auch „single copy“-Gene können ganz oder teilweise von innerartlichen Konversions-Ereignissen betroffen sein. So wurde voraussichtlich ein Intron zwischen den Hämoglobin-Genen 2 $\beta$  und 9 von *C. thummi* durch Genkonversion übertragen (Kao et al., 1994; Hankeln et al., 1997). Neben dem Intronbereich sind zusätzlich auch Bereiche der kodierenden Region von diesem Ereignis betroffen. Im Vergleich dieser Gene mit ihren orthologen Partnern aus anderen Arten können somit synonyme Substitutionen vorgetäuscht werden, die nicht im Laufe der Zeit kontinuierlich durch einzelne Mutationsereignisse angesammelt wurden, sondern ihren Ursprung möglicherweise in einem singulären Rekombinations-Ereignis haben. Da Mitglieder von Genfamilien den hier ermittelten Daten zufolge besonders häufig Rekombinations-Ereignissen unterliegen, die zu einer nachträglichen Veränderung ihrer Sequenzen führen, sind sie möglicherweise schlechte Kandidaten für die Ermittlung evolutionärer Distanzen zwischen verschiedenen Arten.

Auch der Anteil nicht-synonymer Basensubstitutionen und damit der Grad an Sequenz-Konservierung auf Aminosäure-Ebene schwankt bei den verglichenen orthologen Genen stark. Zwischen den Arten *C. tentans* und *C. thummi* variieren die KA-Werte um den Faktor 2, zwischen den Arten *C. tentans* und *C. melanotus* sogar um den Faktor von 3,4. Das Verhältnis nicht-synonymer Substitutionen zu synonymen Substitutionen (KA/KS) spiegelt den unterschiedlichen Grad der Konservierung der Protein-Sequenzen und somit den unterschiedlichen Selektionsdruck wider, der auf den verschiedenen Varianten liegt. Im Interspezies-Vergleich übert-

riff die synonyme Substitutionsrate zumeist deutlich die nicht-synonyme Substitutionsrate ( $KA/KS \ll 1$ ). Sowohl bei den Gen-Einzelkopien, als auch bei den Gen-Subfamilien treten synonyme Substitutionen im Durchschnitt 5mal häufiger auf als nicht-synonyme Austausche. Dies zeigt, daß die Gene unter dem Einfluß der stabilisierenden Selektion stehen und an ihre Funktionen sehr gut angepaßt sind. Eine mögliche Beeinflussung der Funktion durch eine Sequenzänderung auf Aminosäure-Ebene wird nicht toleriert. Den höchsten Grad der Sequenz-Konservierung weisen die Hämoglobin-Gene W und 8 auf. Bei einer durchschnittlichen Anzahl synonyme Basenaustausche besitzen sie eine nur geringe Menge nicht-synonymer Substitutionen. Das Verhältnis von KA zu KS von 0,09 liegt daher bei den Hämoglobinen W und 8 um mehr als 50% unter dem Durchschnittswert von 0,2 für „single copy“-Varianten. Möglicherweise deutet dies auf eine spezifische, physiologisch bedeutsame Funktion der Gene W und 8 hin.

Die Gene Cte 12-1 und Cte 12-2 sind in *C. tentans* durch eine rezente Duplikation entstanden. Im Vergleich miteinander weisen sie als einziges Genpaar ein erhöhtes Verhältnis an nicht-synonymen Substitutionen zu synonymen Substitutionen auf ( $KA/KS=1,1$ ). Dies deutet darauf hin, daß diese Gene im kurzen Zeitraum ihrer getrennten Evolution keine größere Anzahl synonyme Substitutionen ansammeln konnten. Stark erhöhte KA/KS-Verhältnisse sind häufig zwischen Genabschnitten zu beobachten, die auf eine bestimmte Funktion hin selektiert werden (z.B. variable Regionen von Immunglobulinen, Li (1997)). Auch die Inaktivierung eines Gens läßt den KA-Wert ansteigen. In völlig frei mutierenden Kodons könnte man ein theoretisches Ratenverhältnis von 3 beobachten (Goodman et al., 1987). Wie dieses Beispiel zeigt, sind KA/KS Werte größer 1 nicht immer ein Indiz für die Funktionsänderung oder den Funktionsverlust eines Gens, sie werden auch als Folge eines rezente Duplikations-Ereignisses beobachtet.

Die Abbildung 4.28 zeigt graphisch die ermittelten KA- und KS-Werte, sowie ihr Verhältnis zueinander für alle verglichenen Hämoglobin-Gene. Orthologe Gene der 7B-Subfamilie wurden zwischen den Spezies *C. tentans* und *C. thummi* verglichen. Des weiteren ist der Einfluß der Genkonversion auf synonyme- und nicht-synonyme Substitutionsraten an Beispielen aus den Gen-Subfamilien 7A und 7B von *C. tentans* gezeigt. Die Auswirkungen von Genduplikationen auf die Anzahl synonyme- und nicht-synonymer Substitutionsraten läßt sich anhand der Gene Cte12-1/Cte12-2, sowie Cte7A4/Cte 7A6 beschreiben.

Ein Vergleich von Nukleotid-Substitutionsraten zwischen Genen kann somit helfen zu erkennen, ob Gen-Duplikate der stabilisierenden oder adaptiven Selektion unterliegen, ob sie konzertiert evolvieren oder ob sie als funktionslose Pseudogene im Genom existieren. In Unkenntnis der detaillierten Genfunktion kann man so indirekt auf die Bedeutung eines Gens schließen. Dies ist umso notwendiger, wenn bislang keine Daten über die RNA oder korrespondierende Proteine vorliegen. Die Abb. 4.29 zeigt zusammengefaßt das mögliche Schicksal von Genkopien.

## 4.5 Das evolutionäre „Schicksal“ von Genduplikaten

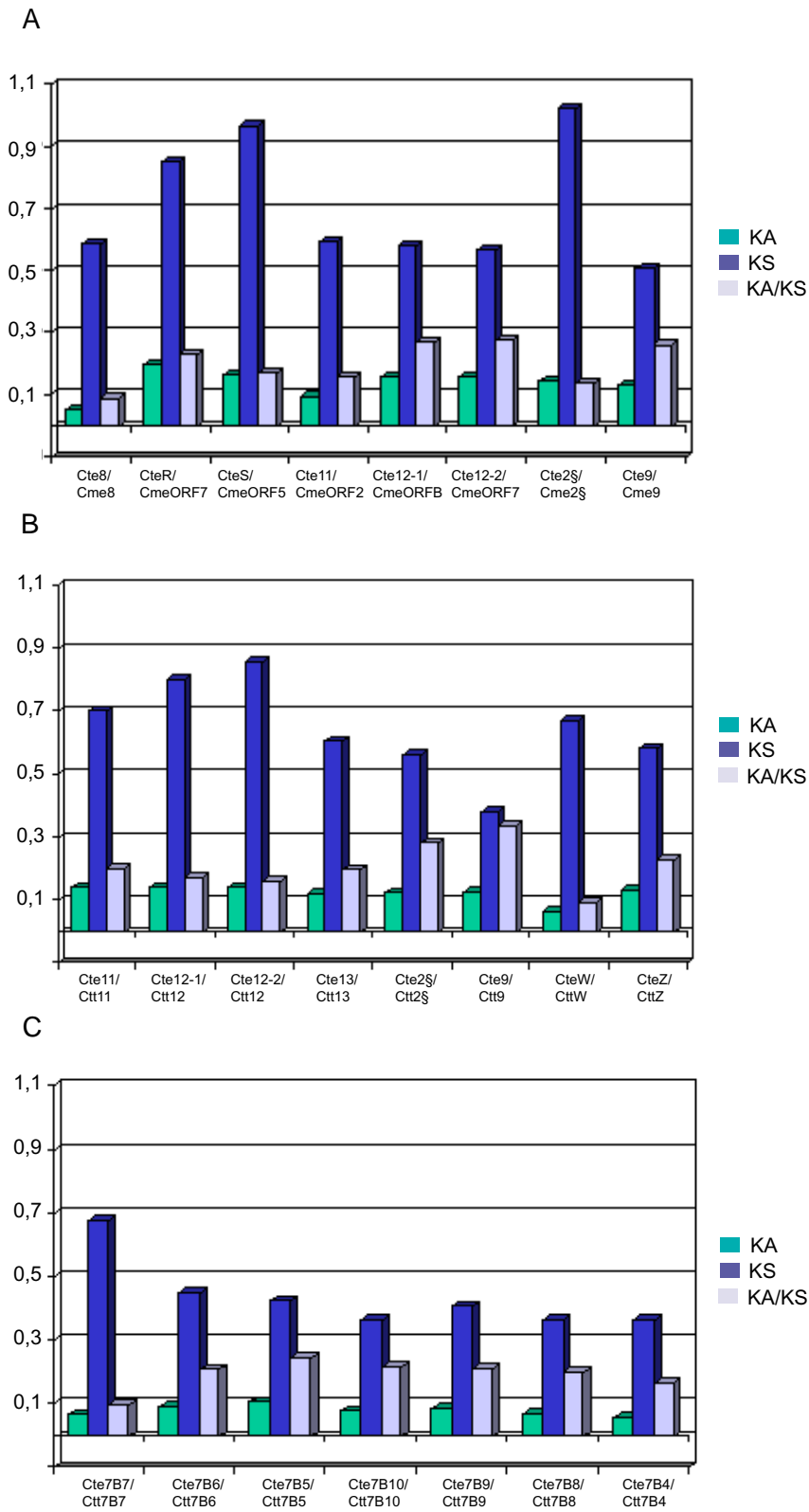


Abbildung 4.28 Legende siehe folgende Seite

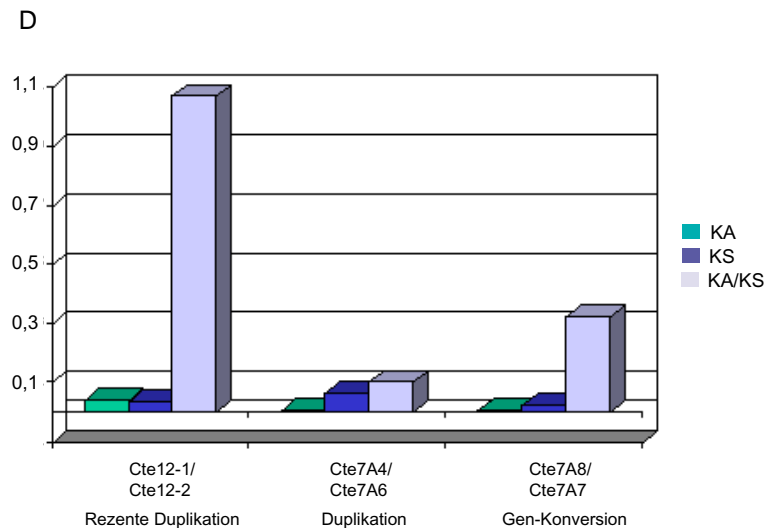


Abbildung 4.28: Synonyme- und nicht-synonyme Substitutionsraten im Vergleich

Zwischen den Chironomus-Arten *C. tentans* und *C. thummi*, sowie zwischen *C. tentans* und *C. melanotus* wurden je 8 orthologe „single copy“-Gene verglichen (A und B). Das Balkendiagramm C zeigt den Vergleich der 7B-Subfamilie in den Arten *C. tentans* und *C. thummi*. Die korrigierten Werte für synonyme Substitutionen (KS), sowie für nicht-synonyme Substitutionen (KA) sind an der Y-Achse aufgetragen. Die verglichenen Genpaare an der X-Achse. Der KS-Wert übersteigt den KA-Wert häufig um ein Vielfaches. Das KA/KS-Verhältnis spricht somit bei allen Vergleichen für eine funktionelle Konservierung der Hämoglobin-Gene. Eine auffällig geringe Anzahl synonymen Austausche findet sich bei den Genen Cte9 und Ctt9. Dies führt, bei einem durchschnittlichen KA-Wert, zu einem leicht erhöhten Verhältnis von KA/KS. Die Genpaare Cte 8/Cme 8 und Cte W/Ctt W weisen eine sehr geringe Anzahl nicht-synonymer Substitutionen ( $KA \leq 0,06$ ) bei durchschnittlichen KS-Werten auf. Dies resultiert in einem geringen KA/KS-Verhältnis und signalisiert den hohen Konservierungsgrad der Gene 8 und W.

Durch Gen-Duplikation und Genkonversion wird die Anzahl von synonymen und nicht-synonymen Austauschen aneinander angeglichen (D). Das paraloge Gen-Paar Cte 12-1/Cte 12-2 weist möglicherweise aufgrund seiner rezenten Duplikation ein KA/KS-Verhältnis von 1 auf.

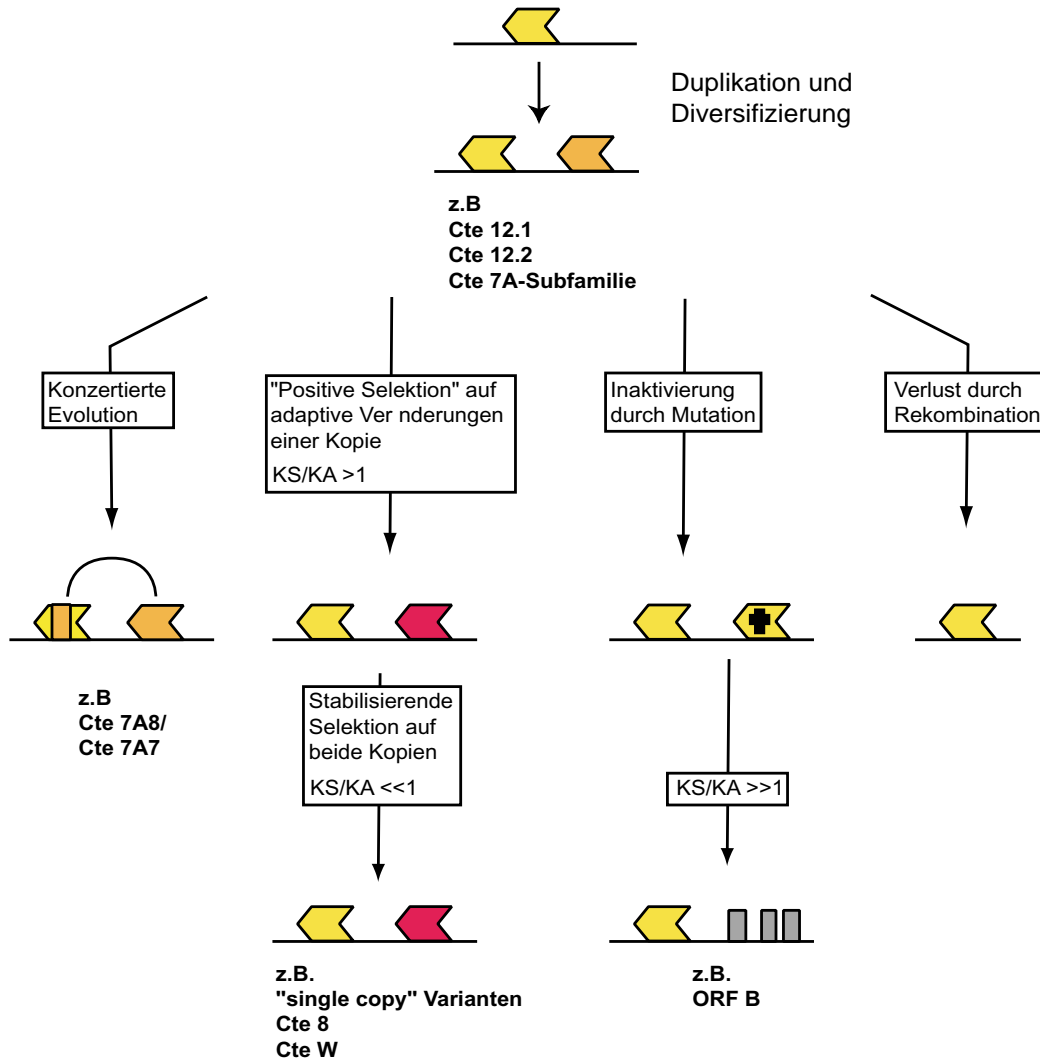


Abbildung 4.29: Das mögliche Schicksal von Genkopien nach einem Duplikations-Ereignis

Genfamilien entstehen durch schrittweise Duplikation und anschließender Modifizierung der Genkopien. Durch Deletion oder Inaktivierung können Gen-Duplikate wieder verloren gehen. Die stabilisierende Selektion stellt sicher, daß funktionell bedeutsame Gene in ausreichender Menge zur Verfügung stehen. Durch adaptive Veränderungen einer Genkopie können Varianten mit veränderter Funktion entstehen.

## 4.6 Aussichten und offene Fragen

Umfassende phylogenetische Analysen der Chironomiden-Hämoglobine zeigen die Dynamik der Entwicklung dieser Multigenfamilie auf. Die Sequenz-Vergleiche paraloger, sowie orthologer Gene ermöglichen es, die Mechanismen der Evolution aufzuzeigen und das Schicksal einzelner Gen-Varianten zu rekonstruieren. Dennoch ist die funktionelle Bedeutung der ungewöhnlichen Multiplizität der Chironomiden-Hämoglobine bislang ungeklärt. Dient die große Zahl verschiedener Hämoglobine ausschließlich dazu, eine möglichst hohe Dosis an Hämoglobin-Molekülen bereitzustellen (Trewitt et al., 1995; Gruhl et al., 1997), oder besitzen einzelne Varianten eine spezifische, für das Überleben der Larven bedeutsame Funktion? Die überdurchschnittliche Konservierung einzelner Gene, insbesondere der „single copy“-Varianten über evolutionär lange Zeiträume hinweg, scheint diese Hypothese zu unterstützen. Des weiteren scheint der Funktionsverlust einzelner Varianten von erheblichem Nachteil zu sein, denn trotz des großen Vorrats an Genen wurde nur der ORF B als einziges Pseudogen in *C. tentans* sicher identifiziert.

Manche Hämoglobin-Varianten werden während der Ontogenese der Chironomiden-Larven differentiell, d.h. nur in bestimmten Stadien exprimiert (Thompson und English, 1966; Tichy, 1970; Saffarini et al., 1991). Die Transkripte der Hämoglobin-Gene 1, 4, 7B4 und 7B3 sind ab dem 3. Larvenstadium nachweisbar und nehmen ab dem 3. Tag des 4. Stadiums stark ab. Hämoglobin 3- und 6-mRNA dagegen ist in Stadium 3 nicht nachweisbar und ist in gleichbleibend hohen Mengen ab dem 2. Tag des 4. Stadiums vorhanden. Diese Expressionsmuster weisen darauf hin, daß die Protein-Varianten mit großer Wahrscheinlichkeit stadienspezifische Funktionen erfüllen.

Über die Transkription der zahlreichen in *C. tentans* neu gefundenen Gene ist bislang nichts bekannt. Detaillierte Expressionsstudien dieser Gene in verschiedenen Entwicklungsstufen und unter verschiedenen Umweltbedingungen sind somit erforderlich. Aufgrund der sehr hohen Sequenz-Übereinstimmungen mancher Gen-Varianten haben sich konventionelle Techniken wie S1 Nuklease- oder RNase-Protektionsexperimente als problematisch erwiesen. Mit Hilfe der „Elektrospray-Ionisierungsmassenspektroskopie“ (ESI-MS) ist es jedoch möglich, Proteine in einem Gemisch anhand ihres Molekulargewichts zuverlässig zu identifizieren. Von jedem sequenzierten Gen kann das Molekulargewicht des kodierten Proteins bestimmt werden und seine Expression durch ESI-MS überprüft werden. Erste Ergebnisse lieferten Untersuchungen von Green et al. (1998). Sie identifizierten in der Hämolymphe von *C. thummi* Larven des 4. Stadiums mehr als 20 Protein-Varianten. Interessanterweise wurden keine Proteine nachgewiesen, die den „neuen“ Varianten Y, V, W, Z und E zugeordnet werden konnten. Dies ist auch deshalb erstaunlich, da es sich hier mehrheitlich um Gene handelt, die zwischen den Arten gut konserviert sind. Dies gilt insbesondere für das Hämoglobin W, wie der Vergleich der Nukleotid-Sequenzen von *C. tentans* und *C. thummi* zeigte. Auch weisen sie alle die strukturellen Cha-

rakteristika funktioneller Gene auf. Moderne Methoden wie ESI-MS oder auch die DNA-Chip-Technologie scheinen somit geeignet zu sein, detaillierte Expressionsstudien in *C. tentans* und weiteren Arten durchzuführen.

Alle bislang untersuchten Chironomiden-Arten besitzen ein individuelles Sortiment von Hämoglobin-Varianten. Es stellt sich daher die Frage, inwiefern dieses speziesspezifische Hämoglobin-Repertoire die unterschiedlichen Lebensbedingungen der Larven ermöglicht. Die extrem hohe Redundanz der Gene bietet zweifellos die Voraussetzung für die adaptive Evolution einzelner Varianten. Besonders die Gen-Subfamilien erweisen sich im Interspezies-Vergleich als extrem variabel in ihrer Gen-Anzahl und -Anordnung. Selbst die Schwester-Arten *C. thummi* und *C. piger* weisen große Unterschiede innerhalb der 7B-Subfamilie auf. Gen-Subfamilien stellen somit speziesspezifische Gen-Gruppen dar und könnten bei der Einnischung der Arten in ihre unterschiedlichen Lebensräume eine wichtige Rolle gespielt haben. Die Hämoglobin-Multiplizität könnte demnach auch eine Bedeutung für die Speziation der Chironomiden haben. Um Erkenntnisse über den Zusammenhang der Gen-Evolution und der Spezies-Evolution zu erhalten, ist es erforderlich, die einzelnen Varianten physiologisch zu charakterisieren. Die bislang vorliegenden Ergebnisse bestätigen, daß sich einzelne Hämoglobin-Komponenten z.B. in ihren Sauerstoffbindungs-Eigenschaften unterscheiden können (Weber et al. (1985); Hankeln, Ungerechts und Schmidt in Vorbereitung). Röntgenstruktur-Analysen können zusätzlich dazu beitragen die Beziehung zwischen der Protein-Struktur und der Funktion für die einzelnen Varianten getrennt zu ermitteln. Der Austausch einiger weniger Aminosäuren der Hämtasche kann bereits eine Variation der Sauerstoff-Affinität um das 27 000-fache bewirken.

Hämoglobine und ihre Verwandten weisen im Tier- und Pflanzenreich die unterschiedlichsten physiologischen Funktionen auf. So wirkt beispielsweise das Leguminosen-Hämoglobin als Sauerstoff-Falle. Es gewährleistet so die Stickstoff-Fixierung der symbiontischen Bakterien (Powell und Gannon, 1988). Auch das Hämoglobin des Nematoden *Ascaris lumbricoides* ist aufgrund seiner extrem hohen Sauerstoff-Affinität (25000 x höher als die des menschlichen Hämoglobins) als Sauerstofftransporter ungeeignet. Es fungiert hingegen als NO-abhängige Deoxygenase und ist so in der Lage, den gebundenen Sauerstoff enzymatisch umzuwandeln und somit zu entgiften (Minning et al., 1999). Einige marine Würmer (*Amphitrite ornata*) besitzen das Enzym Dehaloperoxydase. Diese spaltet halogenierte aromatische Verbindungen, die von anderen Polychaeten als chemische Abwehrstoffe ausgeschieden werden. Wie Lebioda und Mitarbeiter durch Röntgenstruktur-Analysen zeigen konnten, weist dieses Enzym die typische Globinfaltung auf und besitzt eine Aminosäure-Übereinstimmung von 20% zu dem Myoglobin des Seehasen (*Aplysia*). Es ist daher wahrscheinlich, daß es sich von diesem ableitet. Fraglich ist somit auch, ob die Funktion der Chironomiden-Hämoglobine möglicherweise über die reiner O<sub>2</sub>-Transport- und Speicher-Moleküle hinausgeht.





## 5 Zusammenfassung

In der vorliegenden Arbeit wurden 74736 bp genomischer DNA-Sequenz der Hämoglobingen-Gruppe D aus der Chironomiden Art *C. tentans* entschlüsselt und analysiert. Durch Datenbankrecherchen und Sequenz-Vergleiche wurden 29 vollständige Hämoglobin-Gene identifiziert und klassifiziert. Es zeigt sich, daß alle derzeit bekannten Hämoglobin-Gene der Chironomiden auch in *C. tentans* vorhanden sind. Zusätzlich konnten in *C. tentans* sechs neue Hämoglobin-Varianten identifiziert werden, die bislang weder auf Protein- noch auf Gen-Ebene in anderen Spezies nachgewiesen wurden. Die Hämoglobin-Gene liegen in dichter Abfolge innerhalb des Clusters, wobei durchschnittlich etwa alle 2 kb ein Gen zu finden ist. Die Abfolge der Hämoglobin-Gene innerhalb der Gengruppe wird nur an einer Stelle durch ein interspergiertes Gen aus der Familie der Glukosetransporter unterbrochen. Desweiteren konnten zwei retrotransponierbare Elemente der SINE-Klasse (CP1) innerhalb des Hämoglobingen-Clusters identifiziert werden. Alle Gene besitzen die für ihre Expression erforderlichen Signalsequenzen, so daß es sich höchstwahrscheinlich um aktive Gene handelt. Die abgeleiteten Aminosäure-Sequenzen weisen alle Charakteristika sauerstofftransportierender Moleküle auf.

Da es sich bei den Hämoglobinen um eine sehr alte Genfamilie handelt, kann die vergleichende Analyse der Hämoglobin-Genstruktur bei Vertebraten, Invertebraten, Pflanzen und Protozoen zur Rekonstruktion der Intron-Evolution genutzt werden. Die Konservierung von Intronpositionen in homologen Genen verschiedener Taxa gilt dabei als Maß für das relative stammesgeschichtliche Alter der Introns. Eine Vielzahl der Hämoglobin-Gene von Invertebraten weisen ein Intron im zentralen Genbereich auf. Auch bei einigen Chironomiden-Arten konnten diese „zentralen Introns“ nachgewiesen werden. Die Hämoglobin-Gene von *C. tentans* galten hingegen bislang als intronlos. Die vorliegende Untersuchung zeigt, daß auch zwei Hämoglobin-Gene dieser Spezies je ein kurzes Intron aufweisen. Der Vergleich der Intronverteilung in den Hämoglobin-Genen der Chironomiden führt zu dem Ergebnis, daß alle vorhandenen Introns am sparsamsten (im Sinne des „maximum parsimony“-Prinzips) durch unabhängige Insertionen in ein intronloses Vorläufer-Hämoglobin-Gen erklärt werden können. Alle bislang in Chironomiden beschriebenen Introns sind mit großer Wahrscheinlichkeit nicht ortholog (Hankeln et al., 1997; diese Arbeit). Das Vorläufer-Hämoglobin-Gen in Chironomiden besaß daher vermutlich kein „zentrales Intron“. Die in Chironomiden gefundenen Verhältnisse stellen somit die von Go (1981) formulierte Hypothese der Ursprünglichkeit des „zentralen Introns“ in Hämoglobin-Genen in Frage. Die in Invertebraten und Pflanzen beschriebenen „zentralen Introns“ sind

vermutlich nicht homolog und dementsprechend auch nicht auf ein Intron im anze-stralen Globin zurückzuführen. Vielmehr implizieren die in hohem Maße variablen Positionen der „zentralen Introns“ bei Pflanzen und Invertebraten ihre unabhängige Insertion in die jeweiligen Globin-Gene nach der Aufspaltung der Taxa.

Grundsätzlich können zwei Klassen von Hämoglobin-Genen in Chironomiden unterschieden werden. Die überwiegende Mehrzahl der Hämoglobine wird von Genen kodiert, die nur in einer Kopie im Genom vorliegen. Sie werden dementsprechend als „single copy“ Varianten bezeichnet. Für andere Hämoglobin-Varianten konnte hingegen eine Vielzahl leicht unterschiedlicher Gene beschrieben werden. Diese bilden sogenannte Gen-Subfamilien. In *C. tentans* konnte gezeigt werden, daß neben den 7B-Genen auch die 7A-Gene eine eigene Subfamilie bilden. Die „single copy“ Varianten zeichnen sich im Interspezies-Vergleich durch ihre konservierte Nukleotid-Sequenz aus: Sie unterliegen während ihrer Evolution offenbar einer stabilisierenden Selektion, d.h. Veränderungen ihrer Protein-Sequenzen werden nur in geringem Maße toleriert. Auch ihre räumliche Anordnung innerhalb der Gengruppe ist zwischenartlich konserviert. Der Vergleich der „single copy“ Varianten innerhalb einer Art zeigt, daß diese sehr deutliche Sequenz-Unterschiede zueinander aufweisen. Sie bilden somit ein konserviertes Sortiment an Hämoglobin-Genen, das weitgehend vor der Radiation der Arten entstanden ist und eine über die Artgrenzen hinweg unveränderte „Hämoglobin-Grundausstattung“ gewährleistet.

Im Gegensatz hierzu zeichnen sich die Mitglieder von Hämoglobin-Gen-Subfamilien durch eine hohe Variabilität aus: Nukleotid-Sequenz, Anzahl und Organisation der Gene innerhalb der Genfamilie weisen im zwischenartlichen Vergleich zahlreiche Unterschiede auf. Es ist daher nur selten möglich allein auf Grundlage der Nukleotid-Sequenzen orthologe Genpaare zu identifizieren. Die orthologen Gene der 7B-Subfamilie aus *C. tentans* und *C. thummi* konnten ausschließlich anhand korrespondierender Intergen-Sequenzen einander zugeordnet werden. Somit sind die Gen-Subfamilien präferenziell an der Entstehung eines speziesspezifischen Gen-Repertoires beteiligt. Variationen der Nukleotid-Sequenz, Gen-Anzahl und Gen-Organisation innerhalb der Subfamilie werden im Gegensatz zu den „single copy“ Varianten in einem hohen Maße toleriert.

Aufgrund der hohen Sequenz-Übereinstimmungen zwischen den Mitgliedern der Gen-Subfamilien unterliegen diese einer Vielzahl von Rearrangements, die in Gen-Duplikationen, Deletionen und Sequenz-Homogenisierungen resultieren. So führten beispielsweise Genduplikationen durch ungleiches, homologes Crossing-over mit großer Wahrscheinlichkeit zur Entstehung und Expansion der 7A-Subfamilie. Auch die Gene Cte12-1 und Cte 12-2 sind vermutlich das Ergebnis eines rezenten Duplikations-Ereignisses. Der Mechanismus der Retrotransposition, der zu einer Duplikation eines 3'-untranslatierten Bereichs innerhalb der 7A-Subfamilie führte, scheint für die Entstehung der Hämoglobin-Multiplizität in Chironomiden hingegen weniger bedeutsam zu sein. Innerhalb der 7A-Subfamilie ist eine Angleichung der Gene durch konzertierte Sequenz-Evolution zu beobachten. Der nukleotidweise Vergleich von

---

Gen-Sequenzen zeigt am Beispiel der Gene 7A7 und 7A8, daß die konzertierte Evolution dieser Gen-Varianten auf dem Mechanismus der Genkonversion beruht. Auch die Gen-Subfamilie 7B unterliegt offenbar in hohem Maße einer solchen Sequenz-Homogenisierung.

Im Sinne einer molekularen Uhr sollten synonyme Basenaustausche weitgehend neutral sein und sich proportional zur Zeit in den Genen anhäufen. Der Vergleich der Hämoglobin-Gen-Sequenzen zeigt, daß große Unterschiede in der Anzahl der synonymen Basenaustausche zwischen orthologen Genen nicht zwangsläufig das Ergebnis einer frühen Trennung dieser Gene sind. Die Übertragung von Sequenzen zwischen paralogen Genen kann die Anzahl der synonymen Basenaustausche orthologer Gene in kürzester Zeit verändern und den tatsächlichen Zeitpunkt der Trennung zweier orthologer Gene überdecken. Werden Genkonversionen nicht erkannt, weil beispielsweise nicht alle Gene der Gruppe vollständig erfaßt werden konnten, führt der Vergleich orthologer Gen-Sequenzen zwangsläufig zu falschen evolutionären Gendistanzen. Da die Mitglieder der Hämoglobin-Genfamilie besonders häufig Rekombinations-Prozessen unterliegen, sind sie daher möglicherweise weniger nützliche Kandidaten für die Ermittlung evolutionärer Distanzen zwischen den verschiedenen Chironomiden-Arten.

Insgesamt zeigen die Ergebnisse dieser Arbeit, daß anhand detaillierter phylogenetischer Analysen sich die Evolution der Hämoglobin-Multigenfamilie von Chironomiden umfassend beschreiben läßt. Ob einzelne, besonders gut konservierte Gen-Varianten (wie z. B. die Gene Cte 8 und Cte W) eine spezifische physiologische Funktion erfüllen oder ob die Gen-Subfamilien, die ein speziesspezifisches Genrepertoire bilden, an der Einnischung der verschiedenen Arten beteiligt sind, sollte durch weiterführende Untersuchungen (z. B. der Genexpression sowie der physiologischen Eigenschaften einzelner Varianten) ermittelt werden können.



## Literaturverzeichnis

- Altschul S.F., Gish W., Miller W., Myers E.W. und Lipman D.J. (1990). "Basic local alignment search tool". *J. Mol. Biol.*, 215:403–410.
- Antoine M., Erbil C., Münch E., Schnell S. und Niessing J. (1987). "Genomic organization and primary structure of five homologous pairs of intron-less genes encoding secretory globins from the insect *Chironomus thummi thummi*." *Gene*, 56:41–51.
- Antoine M. und Niessing J. (1984). "Intron-less globin genes in the insect *Chironomus thummi thummi*." *Nature*, 310:795–798.
- Aronson H., Royer W. und Hendrickson W. (1994). "Quantification of tertiary structural conservation despite primary sequence drift in the globin fold." *Protein Sci.*, 3:1706–1711.
- Aschauer M.D., Kleinschmidt T., Steer W. und Braunitzer G. (1982). "Haemoglobin polymorphism in *Chironomus* (Diptera). Primary structure. Allergenic determinants and monomeric- dimeric equilibrium." *Chem. Peptide Prot.*, 1:61– 70.
- Ayala F.J. (1997). "Vagaries of the molecular clock." *Proc. Natl. Acad. Sci. USA*, 94:7776–7783.
- Bailey G.S., Poulter R.T. und Stockwell P.A. (1978). "Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci." *Proc. Natl. Acad. Sci.*, 75(11):5575–5579.
- Baldwin A.S., Bergtrom G. und Haralson M.A. (1981). "Cell-free synthesis of *Chironomus thummi* (Diptera) globins." *Arch. Biochem. Biophys.*, 211:390–395.
- Baltimore D. (1981). "Gene conversion: some implications for immunoglobulin genes." *Cell*, 24:592–594.
- Behlke J., Müller K. und Scheler W. (1972). "Assoziation- Dissoziations- Gleichgewichte bei einigen Fraktionen des Methämoglobins von *Chironomus thummi thummi*." *Acta. biol. med. germ.*, 28:1069–1072.
- Benton W. und Davis R. (1977). "Screening gt recombinant clones by hybridization to single plaques 'in situ'." *Science*, 196:180–182.
- Bergtrom G., Laufer H. und Rogers R. (1976). "Fat body: A site of hemoglobin synthesis in *Chironomus thummi* (Diptera)." *J. Cell Biol.*, 69:264–274.

- Birnboim H.C. und Doly J. (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." *Nucleic Acids Res.*, 24:1513–1523.
- Blake C.C.F. (1978). "Do genes-in-pieces imply proteins-in-pieces?" *Nature*, 273:267.
- Blanchetot A., Price M. und Jeffreys A.J. (1986). "The mouse myoglobin gene. Characterisation and sequence comparison with other mammalian myoglobin genes." *Eur. Biochem.*, 159:469–474.
- Blanchetot A., Wilson V., Wood D. und Jeffreys A.J. (1983). "The seal myoglobin gene: an unusually long globin gene." *Nature*, 301:732–734.
- Blaxter M.L., Ingram L. und Tweedie S. (1994). "Sequence, expression and evolution of the globins of the parasitic nematode *Nippostrongylus brasiliensis*." *Mol. Biochem. Parasitol.*, 68:1–14.
- Boedtker H., Finer M. und Aho S. (1985). "The structure of the chicken alpha 2 collagen gene." *Ann. N. Y. Acad. Sci.*, 460:85–116.
- Boer P.H., Adra C.N., Lau Y.F. und McBurney M.W. (1987). "The testis-specific phosphoglycerate kinase gene pgk-2 is a recruited retroposon." *Mol. Cell. Biol.*, 7(9):3107–3112.
- Bollag R.J., Elwood D.R., Tobin E.D., Godwin A.R. und Liskay R.M. (1992). "Formation of heteroduplex DNA during mammalian intrachromosomal gene conversion." *Mol. Cell. Biol.*, 12:1546–1552.
- Bolognesi M., Bordo D., Rizzi M., Torricone C. und Ascenzi P. (1997). "Nonvertebrate hemoglobins: structural bases for reactivity." *Prog. Biophys. Mol. Biol.*, 68:68, 29–68.
- Bonner J.J. und Pardue M.L. (1976). "Ecdysone-stimulated RNA synthesis in imaginal discs of *Drosophila melanogaster*." *Chromosoma (Berl.)*, 58:87–99.
- Braun V., Crichton R.R. und Barunitzer G. (1968). "Hemoglobins. XV. On monomeric and dimeric insect hemoglobins." *Hoppe Seylers Z. Physiol. Chem.*, 349:197–210.
- Braunitzer G., Buse G. und Braig S. (1968). "On the stability of monomeric and dimeric hemoglobins (erythrocrucorins) of the midge *Chironomus thummi*." *Hoppe Seylers Z. Physiol. Chem.*, 349:263–264.
- Brisson N. und Verma D.P. (1982). "Soybean leghemoglobin gene family: normal, pseudo, and truncated genes." *Proc. Natl. Acad. Sci. Jul.*, 79:4055–4059.
- Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D., Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D., Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I., Oberbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A., Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D., Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C., Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M., Klenk H.P., Fraser C.M.,

- 
- Smith H.O., Woese C.R. und Venter J.C. (1996). "Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*." *Science*, 273:1058–1072.
- Burmester T. und Hankeln T. (1999). "A globin gene of *Drosophila melanogaster*." *Mol. Biol. Evol.*, 16(12):1809–1811.
- Burmester T., Weich B., Reinhardt S. und Hankeln T. (2000). "A Vertebrate globin expressed in the brain". *Nature*, 401:520–522.
- Cerff R. (1995). "The chimeric nature of nuclear genomes and the antiquity of introns as demonstrated by the GAPDH gene system." In G.M. und Schimmel P. (ed.) (Hg.), "Tracing Biological Evolution in Protein and Gene Structures. Proc. of the 20th Taniguchi Int. Symposium, Nagoya, Japan", Elsevier Science, Amsterdam; 205-227.
- Chen E. und Seeburg P. (1985). "Supercoil sequencing: a fast and simple method for sequencing plasmid DNA." *DNA*, 4:165–170.
- Clewell D. und Helinski D. (1969). "A supercoiled circular DNA-protein-complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form." *Proc. Natl. Acad. Sci. USA*, 62:1159–1166.
- Cooper S.J. und Hope R.M. (1993). "Evolution and expression of a beta-like globin gene of the Australian marsupial *Sminthopsis crassicaudata*." *Proc. Natl. Acad. Sci.*, 15:11777–11781.
- Couture M., Chamberland H., St-Pierre B., Lafontaine J. und Guertin M. (1994). "Nuclear genes encoding chloroplast hemoglobins in the unicellular green alga *Chlamydomonas eugametos*." *Mol. Gen. Genet.*, 243:185–197.
- Craik C.S., Buchman S.R. und Beychok S. (1980). "Characterization of globin domains: heme binding to the central exon product." *Proc. Natl. Acad. Sci.*, 77:1384–1388.
- Craik C.S., Buchman S.R. und Beychok S. (1981). "O<sub>2</sub> binding properties of the product of the central exon of beta-globin gene." *Nature*, 291:87–90.
- Czeczuga B. (1960). "Haemoglobin content of the larvae of *Tendipes f. l. Plumosus* in the Rajgradzkie Lake in different year seasons." *Polskie Archiwum Hydrobiol.*, 7:93–102.
- Dayhoff M.O. (1972). *Atlas of Protein Sequence and Structure*. Natl. Biomed. Res. Found. 5, Washington DC.
- Denhardt D. (1966). "A membrane filter technique for the detection of complementary DNA." *Biochem. Biophys. Res. Com.*, 23:641–645.
- Derr L.K. und Strathern J.N. (1992). "A role for reverse transcripts in gene conversion." *Nature*, 361:170–173.

- DeSanctis G., Falcioni G., Giardina B., Ascoli F. und Brunori M. (1986). "Mini-myoglobin: preparation and reaction with oxygen and carbon monoxide." *J. Mol. Biol.*, 188(1):73–76.
- DeSanctis G., Falcioni G., Giardina B., Ascoli F. und Brunori M. (1988). "Mini-myoglobin. The structural significance of haem-ligand interactions." *J Mol Biol.*, 200(4):725–733.
- Dewey D.R. (1980). "Some applications and misapplications of induced polyploidy to plant breeding." In W.e. Lewis (Hg.), "Polyploidy: Biological Relevance.", Plenum Press, New York; 445-470.
- Dewilde S. (1999). *Nonvertebrate globins and globin genes: structure, funktion and evolution*. Doktorarbeit, Universität Antwerpen (Belgien).
- Dickerson E.R. und Geis E. (1983). *Hemoglobin: structure, function, evolution and pathologie*. Benjamin/Cummings Publ. Comp. Inc., Menlo Park California.
- Dickerson R.E. und Geis I. (1969). *The Structure and Action of Proteins*. Harper and Row, New York.
- Dixon B. und Pohajdak B.. (1992). "Did the ancestral globin gene of plants and animals contain only two introns?" *Trends Biochem. Sci.*, 17(12):486–488.
- Dixon B., Walker B., Kimmins W. und Pohajdak B.. (1992). "A nematode hemoglobin gene contains an intron previously thought to be unique to plants." *J. Mol. Evol.*, 35(2):131–136.
- Doolittle W.F. (1987). "The evolution of the vertebrate plasma proteins." *Biol. Bull.*, 172:269–283.
- Duret L., Gasteiger E. und Perriere G. (1996). "LALNVIEW: a graphical viewer for pairwise sequence alignments." *CABIOS*, 12:507–510.
- Eck R.V. und Dayhoff M.O. (1966). *Atlas of Protein Sequence and Structure*. Natl. Biomed. Res., Found., Washington, DC.
- Edelman G.M. und Gally J.A. (1970). "Arrangement and evolution of eukaryotic genes." In F.e. Schmitt (Hg.), "The Neurosciences: Second Study Program.", Rockefeller University Press, New York; 962-972.
- Efstratiadis A., Posakony J.W., Maniatis T., Lawn R.M., O'Connell C., Spritz R.A., Derrin J.K., Forget B.G., Weissman S.M., Slightom J.L., Blechl A.E., Smithies O., Baralle F.E., Shoulders C.C. und Proudfoot N.J. (1980). "The structure and evolution of the human beta-globin gene family." *Cell.*, 21:653–668.
- Elgin S.C. und Weintraub H. (1975). "Chromosomal proteins and chromatin structure." *Annu. Rev. Biochem.*, 44:725–774.



- 
- English D.S. (1969). "Ontogenetic changes in hemoglobin synthesis of two strains of *Chironomus tentans*." *J. Embryol. Exp. Morphol.*, 22:465–476.
- Feinberg A.P. und Vogelstein B. (1983). "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity." *Anal. Biochem.*, 132:6–13.
- Felsenstein J. (1985). "Confidence limits on phylogenies: an approach using the bootstrap." *Evolution*, 39:783–791.
- Fisher R.A. (1935). "The sheltering of lethals." *Am. Nat.*, 69:446–455.
- Fitch D.H., Mainone C., Goodman M. und Slightom J.L. (1990). "Molecular History of gene konversion in the primate fetal gamma globin genes. Nucleotide sequences from the common gibbon *Hylobates lar*." *J. Biol. Chem.*, 132:132, 6–13.
- Fitch W.M. (1977). "On the problem of discovering the most parsimonious tree." *Am. Nat.*, 111:223–257.
- Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R., Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M., Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D., Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M., van V.R., Palmer N., Adams M.D., Gocayne J., Weidmann J., Utterback T., Watthey L., McDonald L., Artiach P., Bowman C., Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B., Smith H.O. und Venter J.C. (1997). "Genomic sequence of a Lyme disease spirochaete *Borrelia burgdorferi*." *Nature*, 390:580–586.
- Frischauf A.M., Lehrach H., Poustka A. und Murray N. (1983). "Lambda replacement vectors carrying polylinker sequences." *J. Mol. Biol.*, 170:827–842.
- Gerbi S.A. (1985). "Evolution of Ribosomal DNA." In M.R.J. (ed.) (Hg.), "Molecular evolutionary genetics.", Plenum Publ. Corp.
- Gilbert W. (1987). "Cold Spring Harbor Symp." *Quant. Biol.*, 52:901–905.
- Gilbert W., de S.S.J. und Long M. (1997). "Origin of genes". *Proc Natl Acad Sci*, 94:7698–7703.
- Gilbert W. und Glynias M. (1993). "On the ancient nature of introns." *Gene*, 135:137–144.
- Go M. (1981). "Correlation of DNA exonic regions with protein structural units in haemoglobin." *Nature*, 291:90–92.
- Goodman M., Braunitzer G., Kleinschmidt T. und Aschauer H. (1983). "The Analysis of a Protein-Polymorphism. Evolution of Monomeric and Homodimeric Haemoglobins (Erythrocrucorins) of *Chironomus thummi thummi* (Insecta, Diptera)." *Hoppe-Seyler's Z. Physiol. Chem.*, 364:205–217.
- Goodman M., Czelusniak J., Koop B.F., Tagle D.A. und Slightom J.L. (1987). "Globins: a case study in molecular phylogeny." *Cold Spring Harb. Symp. Quant. Biol.*, 52:875–890.

- Goodman M., Pedwaydon J., Czelusniak J., Suzuki T., Gotoh T., Moens L., Shishikura F., Walz D. und Vinogradov S. (1988). "An evolutionary tree for invertebrate globin sequences." *J. Mol. Evol.*, 27:236–249.
- Green B.N., Kuchumov A.R., Hankeln T., Schmidt E.R., Bergtrom G. und Vinogradov S.N. (1998). "An electrospray ionization mass spectrometric study of the extracellular hemoglobins from *Chironomus thummi thummi*." *Biochim. Biophys. Acta*, 1383:143–150.
- Gruhl M., Kao W.Y. und Bergtrom G. (1997). "Evolution of orthologous intronless and intron-bearing globin genes in two insect species." *J. Mol. Evol.*, 45:499–508.
- Grunstein M. und Hogness D. (1975). "Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc." *Natl. Acad. Sci. USA*, 72:3961–3965.
- Hankeln T., Amid C., Weich B., Niessing J. und Schmidt E. (1998). "Molecular evolution of the globin gene cluster E in two distantly related midges, *Chironomus pallidivittatus* and *C. thummi thummi*." *J. Mol. Evol.*, 46:589–601.
- Hankeln T., Friedl H., Ebersberger I., Martin J. und Schmidt E.R. (1997). "A variable intron distribution in globin genes of *Chironomus*: evidence for recent intron gain." *Gene*, 205:151–160.
- Hankeln T., Rozynek P. und Schmidt E.R. (1988). "The nucleotide sequence and in situ localization of a gene for a dimeric haemoglobin from the midge *Chironomus thummi piger*." *Gene*, 64:297–304.
- Hankeln T. und Schmidt E.R. (1991). "The organization, localization and nucleotide sequence of the histone genes of the midge *Chironomus thummi*." *Chromosoma*, 101:25–31.
- Hardison R.C. (1991). "A brief history of hemoglobins : Plant, animal, protist and bacteria." *Proc. Nat. Acad. Sci.*, 93:5675–5679.
- Hardison R.C. (1996). "A brief history of hemoglobins: plant, animal, protist, and bacteria." *Proc. Natl Acad. Sci.*, 93:5675–5679.
- He M., Wilde A. und Kaderbhai M.A. (1989). "A simple single-step procedure for small-scale preparation of *Escherichia coli* plasmids." *Nucl. Acids Res.*, 18:1660.
- Henikoff J.G., Henikoff S. und Pietrovski S. (1999). "New features of the Blocks Database servers,". *Nucl. Acids Res.*, 27:226–228.
- Henikoff S. und Henikoff J.G. (1994). "Protein family classification based on searching a database of blocks." *Genomics*, 19:97–107.
- Hickey D.A., Bally-Cuif L., Abukashawa S., Payant V. und Benkel B.F. (1991). "Concerted evolution of duplicated protein-coding genes in *Drosophila*." *Proc. Natl. Acad. Sci.*, 88:1611–1615.

- 
- Higgins D.G. und Sharp P.M. (1988). "CLUSTAL: a package for performing multiple sequence alignment on a microcomputer." *Gene.*, 73:237–244.
- Higgins D.G. und Sharp P.M. (1989). "Fast and sensitive multiple sequence alignments on a microcomputer." *Comput. Appl. Biosci.*, 5(2):151–153.
- Hillis D.M. und Bull J.J. (1993). "An empirical test of bootstrapping as a method for assessing confidence on phylogenetic analysis." *Syst. Biol.*, 42:182–192.
- Hirokawa T., Boon-Chieng S. und Mitaku S. (1998). "SOSUI: Classification and Secondary Structure Prediction System for Membrane Proteins". *Bioinformatics*, 14(4):378–379.
- Holland P.W.H. und Garcia-Fernandez J.G. (1996). "Hox genes and chordate evolution." *Dev. Biol.*, 173:382–395.
- Holliday R. (1964). "A mechanism for gene conversion in fungi." *Genet. Res.*, 5:282–304.
- Hong H., Rovira C., Recco-Pimentel S., Ching L. und Edstöm J.E. (1995). "Polymorphic SINEs in Chironomids with DNA derived from the R2 insertion site." *J. Mol. Biol.*, 245:34–42.
- Hu X.Y., Ray P.N. und Worton R.G. (1991). "Mechanisms of tandem duplication in the Duchenne muscular dystrophy gene include both homologous and nonhomologous intrachromosomal recombination." *EMBO J.*, 10(9):2471–2477.
- Huang H., Crews S. und Hood L. (1981). "An immunoglobulin VH pseudogene." *J. Mol. Appl. Genet.*, 1(2):93–101.
- Huang X. und Webb M.W. (1991). "Ä Time-Efficient, Linear-Space Local Similarity Algorithm". *Advances in Applied Mathematics*, 12:337–357.
- Hurst L.D. (1994). "Molecular evolution. The uncertain origin of introns." *Nature*, 371:381–382.
- Hurst L.D. und McVean G.T. (1996). "A difficult phase for introns-early. Molecular evolution." *Curr Biol.*, 6:533–536.
- Hyldig-Nielsen J.J., Jensen E.O., Paludan K., Wiborg O., Garrett R., Jorgensen P. und Marcker K.A. (1982). "The primary structures of two leghemoglobin genes from soybean." *Nucleic Acids Res.*, 10:689–701.
- Innis M.A., Gelfand D.H., Sinsky J.J. und White T.J. (1990). "PCR Protocols. A guide to methods and applications." *Academic Press London Ann. Rev. Genet.*, 20:593–624.
- Jacq C., Alt-Mörbe J., Andre B., Arnold W., Bahr A., Ballesta J.P.G., Bargues M., Baron L., Becker A., Biteau N., Blöcker H., Blugeon C., Boskovic J., Brandt P., Brückner M., Buitrago M.J., Coster F., Delaveau T., del F. Rey, Dujon B., Eide L.G., Garcia-Canalejo J.M., Goffeau A., Gomez-Peris A., Granotier C., Hanemann V., Hankeln T., Hoheisel

- J.D., Jäger W., Jimenez A., Jonniaux J.L., Krämer C., Küster H., Laamanen P., Legros Y., Louis E., Möller-Rieker S., Monnet A., Moro M., Müller-Auer S., Nußbaumer B., Paricio N., Paulin L., Perea J., Perez-Alonso M., Perez-Ortin J.E., Pohl T.M., Prydz H., Purnelle B., Rasmussen S.W., Remacha M., Revuelta J.L., Rieger M., Salom D., Saluz H.P., Saiz J.E., Saren A.M., Schäfer M., Scharfe M., Schmidt E.R., Schneider C., Scholler P., Schwarz S., Soler-Mira A., Urrestarazu L.A., Verhasselt P., Vissers S., Voet M., Volckaert G., Wagner G., Wambutt R., Wedler E., Wedler H., Wölfl W., Harris D.E., Bowman S., Brown (1997). "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IV." *Nature*, 387 Supp.:75–78.
- Jacq C., Miller J.R. und Brownlee G.G. (1977). "A pseudogene structure in 5S DNA of *Xenopus laevis*." *Cell*, 12:109–120.
- Jeffreys A., s S.H., e P.B., Wood D., Blanchetot A. und Adams S. (1983). In D.S.B. (Hrsg.) (Hg.), "Evolution from Molecules to Men", 43–50. Cambridge, Cambridge University Press.
- Jeffreys A.J. und Flavell R.A. (1977). "The rabbit beta-globin gene contains a large large insert in the coding sequence." *Cell*, 12(4):1097–1108.
- Jensen E., Paludan K., Hyldig-Nielsen J.J., Jorgensen P. und Marcker K.A. (1981). "The strukture of a chromosomal leghaemoglobin gene from soy bean." *Nature*, 291:677–679.
- Jhiang S.M. und Riggs A.F. (1989). "The Structure of the Gene Encoding Chain C of the Hemoglobin of the earthworm, *Lumbricus terrestris*." *J. Biol. Chem.*, 264:19002–19008.
- Jukes T.H. und Cantor C.R. (1969). "Evolution of protein molecules." In e. H. N. Munro (Hg.), "Mammalian protein metabolism.", Academic Press, New York and London; 21–132.
- Kao W.Y. und Bergtrom G. (1995). "Sequences of globin 6 gene alleles and linkage of globin 6 and 7 B genes in the insect *Chironomus thummi thummi*." *Gene*, 153:209–213.
- Kao W.Y., Hankeln T., Schmidt E.R. und Bergtrom G. (1995). "Sequence and evolution of the gene for the monomeric globin I and its linkage to genes coding for dimeric globins in the insect *Chironomus thummi*." *J. Mol. Evol.*, 40:354–361.
- Kao W.Y., Trewitt P.M. und Bergtrom G. (1994). "Intron-containing globin genes in the insect *Chironomus thummi*." *J. Mol. Evol.*, 38:241–249.
- Karlson P. (1994). *Kurzes Lehrbuch der Biochemie*. Thieme Verlag, Stuttgart, New York.
- Karlsson S. und Nienhuis A.W. (1985). "Developmental regulation of human globin genes." *Annu. Rev. Biochem.*, 54:1071–1108.
- Keyl H.G. (1962). "Chromosomenevolution bei *Chironomus*. II. Chromosomenumbauten und phylogenetische Beziehungen der Arten." *Chromosoma*, 13:464–514.

- 
- Keyl H.G. und Strenzke K. (1956). "Taxonomie und Cytologie von zwei Subspezies der Art *Chironomus thummi*." *Z. Naturforschg.*, 11b:727–735.
- Kimura M. und King J.L. (1979). "Fixation of a deleterious allele at one of two duplicate loci by mutation pressure and random drift." *Proc. Natl. Acad. Sci.*, 76(6):2858–2861.
- Kleinschmidt T., Aschauer H., Steer W., Pfletschinger J. und Braunitzer G. (1981). "Polymorphism of hemoglobins (erythrocrucorins) of *Chironomus* (Insecta): primary structure and monomeric- dimeric equilibrium." In L.J. und Lamy J. (Editors) (Hg.), "Invertebrate Oxygen- Binding Proteins. Structure, Active Site and Function.", Marcel Dekker New York and Basel; 267-273.
- Kleinschmidt T. und Braunitzer G. (1976). "The sequence of a dimetric hemoglobin." *Hoppe Seylers Z. Physiol. Chem.*, 357(12):1805–1808.
- Kloek A.P., McCarter J.P., Setterquist R.A., Schedl T. und Goldberg D.E. (1996). "Caenorhabditis globin genes: rapid intronic divergence contrasts with conservation of silent exonic sites." *J. Mol. Evol.*, 43(2):101–108.
- Kloek A.P., Sherman D.R. und Goldberg D.E. (1993). "Novel gene structure and evolutionary context of *Caenorhabditis elegans* globin." *Gene*, 129(2):215–221.
- Koop B.F. und Goodman M. (1988). "Evolutionary and developmental aspects of two hemoglobin beta-chain genes (epsilon M and beta M) of opossum." *Proc. Natl. Acad. Sci.*, 85(11):3893–3897.
- Kumar S., Tamura K. und Nei M. (1993). *MEGA: Molecular Evolutionary Genetics Analysis, version 1.01*. The Pennsylvania State University, University Park, PA 16802.
- Kunst N., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessires P., Bolotin A., Borchert S., Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Düsterhöft A., Ehrlich S.D., Emmerson P.T., Entian K.D., ngton J.E., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Hnaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi, Koetter P., Koningstein, Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Mdiq C., Medina N., Mellado R. (1997). "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*." *Nature*, 390:249–256.
- Langer-Safer P., Levine M. und Ward D.C. (1982). "Immunological method for mapping genes on *Drosophila* polytene chromosomes." *Proc. Natl. Acad. Sci.*, 79:4381–4385.
- Langley K.E., Villarejo M.R., Fowler A.V., Zamenhof P.J. und Zabin I. (1975). "Molecular basis of  $\beta$ -galactosidase alpha-complementation." *Proc. Natl. Acad. Sci.*, 72:1254–1257.

- Laufer H. und J.P. (1971). "A factor controlling the concentration of hemoglobin in Chironomus during metamorphosis." *Limnologica*, 8:125–126.
- Laufer H., Vafopoulou-Mandalos X., Kuliawat R., Gundling G. und Ramirez F. (1982). "Tissue-specific and gene-specific sites of hemoglobin synthesis in Chironomus." *Prog. Clin. Biol. Res.*, 85 Pt. A:327–335.
- Lewin R. (1984). "Surprise Finding with Insect Globin Genes." *Science*, 226:328.
- Lewis W.H. (1980). *Polyploidy: Biological Relevance*. Plenum Press, New York.
- Leyko W. und Osmulski P.A. (1985). "Seasonal variability of hemoglobin content and component composition of Chironomus thummi thummi larvae." *Comp. Biochem. Physiol.*, 80B:613–616.
- Li W.H. (1997). *Molecular Evolution*. Sinauer Associates, Sunderland, M. A.
- Linnenbach A.J., Seng B.A., Wu S., Robbins S., Scollon M., Pyrc J.J., Druck T. und Huebner K. (1993). "Retroposition in a family of carcinoma-associated antigen genes." *Mol. Cell. Biol.*, 13:1507–1515.
- Little P.F. (1982). "Globin pseudogenes." *Cell.*, 28(4):683–684.
- Logsdon J.M.J., Tyshenko M.G., Dixon C., D-Jafari J., Walker V.K. und Palmer J.D. (1995). "Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory." *Proc. Natl. Acad. Sci.*, 92(18):8507–8511.
- Long M., de S.S.J. und Gilbert W. (1995). "Evolution of the intron-exon structure of eukaryotic genes." *Curr. Opin. Genet. Dev.*, 5(6):774–778.
- Lundin L.G. (1993). "Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse." *Genomics*, 16:1–19.
- MacGregor H. und Mizuno S. (1976). "In-situ hybridization of 'nick-translated' 3H-ribosomal DNA to chromosomes of salamanders." *Chromosoma (Berl.)*, 54:15–25.
- Maeda N. und Smithies O. (1986). "The evolution of multigene families: human hapto-globin genes." *Annu Rev Genet.*, 20:81–108.
- Mandel M. und Higa A. (1970). "Calcium-dependent bacteriophage DNA infection". *J. Mol. Biol.*, 53:158–162.
- Maniatis T., Fritsch E. und Sambrook J. (1982). *Molecular cloning. A laboratory manual*. Cold spring Harbor, New York.
- Mattick J.S. (1994). "Introns: evolution and function." *Current Opin. Genet.*, 5:1613–1620.

- 
- McDonnell M.W., Simon M.N. und Studier F.W. (1977). "Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels." *J. Mol. Biol.*, 110:119–146.
- Melamed C., Nevo Y. und Kupiec M. (1992). "Involvement of cDNA in homologous recombination between Ty elements in *Saccharomyces cerevisiae*." *Mol. Cell. Biol.*, 12:1613–1620.
- Metzenberg A.B., Wurzer G., Huisman T.H.J. und Smithies O. (1991). "Homology requirements for unequal crossing over in humans." *Genetics*, 128:143–161.
- Mewes H.W., Albermann K., Bähr M., Frishman D., Gleissner A., Hani J., Heumann K., Kleine K., Maierl A., Oliver S.G., Pfeiffer F. und Zollner A. (1997). "Overview of the yeast genome." *Nature*, 387:7–9.
- Minning D.M., Gow A.J., Bonaventura J., Braun R., Dewhirst M., Goldberg D.E. und Stamler J.S. (1999). "Ascaris haemoglobin is a nitric oxide-activated „deoxygenase“." *Nature*, 401:479–502.
- Moens L., Vanfleteren J., De B.K., Jellie A.M., Tate W. und Trotman C.N.A. (1992). "Unexpected intron location in non-vertebrate globin genes." *FEBS Lett.*, 312:105–109.
- Moens L., Vanfleteren J., de Peer Y Van, Peeters K., Kapp O., Czeluzniak J., Goodman M., Blaxter M. und Vinogradov S. (1996). "Globins in nonvertebrate species: dispersal by horizontal gene transfer and evolution of the structure-function relationships." *Mol Biol Evol.*, 13(2):324–333.
- Morl M. und Schmelzer C. (1990). "Integration of group II intron bII into a foreign RNA by reversal of the self-splicing reaction in vitro." *Cell*, 60(4):629–636.
- Mouches C., Pasteur N., Berge J., Hyrien O., Raymond M., De Saint Vincent B., De Silvestri M. und Georghiou G. (1986). "Amplification of an esterase gene is responsible for insecticide resistance in a California culex mosquito." *Science*, 233:778–780.
- Mueckler M. (1993). "Facilitative glucose transporters". *Eur. J. Biochem.*, 219:713–725.
- Mueckler M., Caruso C., Baldwin S.A., Panico M., Blench I., Morris H.R., Allard W.H., Lienhard G.E. und Lodish H.F. (1985). "Sequence and structure of a human glucose transporter." *Science*, 229:941–945.
- Müntzing A. (1936). "The evolutionary significance of autopolyploidy." *Hereditas*, 21:263–378.
- Nathans J., Thomas D. und Hogness D.S. (1986). "Molecular genetics of human color vision: the genes encoding blue, green, and red pigments." *Science.*, 232(4747):193–202.
- Nei M. und Roychoudhury A.K. (1973). "Probability of fixation and mean fixation time of an overdominant mutation." *Genetics.*, 74(2):371–380.

- Neitz M., Neitz J. und Jacobs G.H. (1991). "Spectral tuning of pigments underlying red-green color vision." *Science.*, 252(5008):971–974.
- Nishioka Y., Leder A. und Leder P. (1980). "Unusual alpha-globin-like gene that has cleanly lost both globin intervening sequences." *Proc. Natl. Acad. Sci.*, 77(5):2806–2809.
- Ohno S. (1970). *Evolution by gene duplication*. Springer-Verlag, Berlin.
- Osmulski P.A. und Leyko W. (1986). "Structure, function and physiological role of Chironomus haemoglobin." *Comp. Biochem. Physiol.*, 85B(4):701–722.
- Parniske M., Hammond-Kosack K.E., Golstein C., Thomas C.M., Jones D.A., Harrison K., Wulff B.B.H. und Jones D.G. (1997). "Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato." *Cell*, 91:821–832.
- Powell R. und Gannon F. (1988). "The leghaemoglobins." *Bioessays*, 9(4):117–121.
- Proudfoot N.J. und Maniatis T. (1980). "The structure of a human alpha-globin pseudo-gene and its relationship to alpha-globin gene duplication." *Cell*, 21(2):537–554.
- Rogers J.H. (1985). "The origin and evolution of retroposons." *Int. Rev. Cytol.*, 93:187–279.
- Rogers J.H. (1986). "Introns between protein domains: selective insertion or frameshifting?" *Trend. Genet.*, 2:223.
- Rozynek P., Broecker M., Hankeln T. und Schmidt E.R. (1991). "The primary structure of several hemoglobin genes from the genome of Chironomus tentans." Springer-Verlag, New York.
- Rozynek P., Hankeln T. und Schmidt E.R. (1989). "Structure of a hemoglobin gene cluster and nucleotide sequence of three hemoglobin genes from the midge Chironomus thummi piger (Diptera, Insecta)." *Biol. Chem. Hoppe-Seyler*, 370:533–542.
- Ruley H.E. und Fried M. (1983). "Clustered illegitimate recombination events in mammalian cells involving very short sequence homologies." *Nature*, 304:181–184.
- Rüther U. (1982). "pUC 250 allows rapid chemical sequencing of both DNA strands of its inserts." *Nucl. Acids Res.*, 10:5765–5772.
- Saffarini D.A., Trewitt P.M., Luhm R.A. und Bergtrom G. (1991). "Differential regulation of insect globin and actin mRNAs during larval development in Chironomus thummi." *Gene*, 101:215–222.
- Saiki R., Gyllensten U. und Erlich H. (1988). "The polymerase chain reaction." IRL Press, Oxford.



- 
- Saitou N. und Nei M. (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." *Mol. Biol. Evol.*, 4(4):406–425.
- Sambrook J., Fritsch E.F. und Maniatis (1989). "Molecular cloning: A laboratory manual. 2nd edition." Cold Spring Harbor, New York.
- Sanger F., Nicklen S. und Coulson A.R. (1977). "DNA sequencing with chain-terminating inhibitors." *Proc. Natl. Acad. Sci. USA*, 74:5463–5467.
- Sargent C.A., Young C., Marsh S., Ferguson-Smith M.A. und Affara N.A. (1994). "The glycerol kinase gene family: structure of the Xp gene, and related intronless retroposons." *Hum. Mol. Genet.*, 3:1317–1324.
- Schin K.S., Poluhowich J., Gamo T. und Laufer H. (1974). "Degredation of himoglobin in Chironomus during metamorphosis." *J. Insect Physiol.*, 20:561–571.
- Schmidt E., Godwin E., Keyl H.G. und Israelewski N. (1982). "Cloning and analysis of ribosomal DNA of Chironomus thummi thummi and Chironomus thummi piger." *Chromosoma*, 87:389–407.
- Schmidt E., Vistorin G. und Keyl H.G. (1980). "An AT-rich DNA component in the genomes of Chironomus thummi thummi and Chironomus thummi piger." *Chromosoma*, 76:35–45.
- Schmidt E.R. (1992). *Nonradioactive labeling and detection of biomolecules*, Kap. Mapping of polytene chromosomes. C. Kessler. Springer Verlag Berlin, Heidelberg.
- Schmidt E.R., Keyl H.G. und Hankeln T. (1988). "In situ localization of two haemoglobin gene clusters in the chromosomes of 13 species of Chironomus." *Chromosoma*, 96:353–359.
- Sherman D.R., Kloek A.P., Krishnan B.R., Guinn B. und Goldberg D.E. (1992). "Ascaris hemoglobin gene: plant-like structure reflects the ancestral globin gene." *Proc. Natl. Acad. Sci.*, 89(24):11696–11700.
- Shilova K.S. (1958). "Biologie of insect pupae from fish ponds." *Trudy biol. Sta. Borok*, 3:250–258.
- Sidow A. (1996). "Gen(om)e duplications in the evolution of early vertebrates." *Curr. Opin. Genet and Dev.*, 6:715–722.
- Singh L., Purdom J.F. und Jones K.W. (1977). "Effect of denaturing agents on the detectability of specific DNA sequences of various base compositions by in situ hybridization." *Chromosoma*, 60:377–389.
- Slightom J., Blechl A.E. und Smithies O. (1980). "Human fetal G<sup>γ</sup>- and A<sup>γ</sup>-globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes." *Cell*, 21:627–638.

- Smith G.P. (1976). "Evolution of repeated DNA sequences by unequal crossover." *Science*, 191:528–535.
- Smithies O. und Powers P.A. (1986). "Gene conversions and their relation to homologous chromosome pairing." *Phil. Trans. R. Soc. Lond. B*, 312:291–302.
- Soares M.B., Schon E., Henderson A., Karathanasis S.K., Cate R., Zeitlin S., Chirgwin J. und Efstratiadis A. (1985). "RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon." *Mol Cell Biol.*, 5(8):2090–2103.
- Southern E.M. (1975). "Detection of specific sequences among DNA fragments separated by gel electrophoresis." *J. Mol. Biol.*, 98:503–517.
- Sparrow A.H. und Naumann A.F. (1976). "Evolution of genome size by DNA doublings." *Science*, 192:524–529.
- Stebbins G. (1971). *Chromosomal Evolution in Higher Plants*. Edward Arnold, London.
- Steigemann W. und Weber E. (1979). "Structure of erythrocyruorin in different ligand states refined at 1.4 Å resolution." *J. Mol. Biol.*, 127(3):309–338.
- Stoltzfus A. und Doolittle W.F. (1993). "Slippery introns and globin gene evolution." *Curr. Biol.*, 3:215–217.
- Stoltzfus A., Logsdon J.M.J., Palmer J.D. und Doolittle W.F. (1997). "Intron sliding and the diversity of intron positions." *Proc. Natl. Acad. Sci.*, 94(20):10739–10744.
- Stryer L. (1997). *Biochemie*. Spektrum der Wissenschaft Verlag mbH.
- Sullivan S.L., Adamson M.C., Ressler K.J., Kozak C.A. und Buck L.B. (1996). "The chromosomal distribution of mouse odorant receptor genes." *Proc. Natl. Acad. Sci.*, 93:884–888.
- Swofford D.L. (1993). "PAUP: Phylogenetic analysis using parsimony. Version 3.1.1. Computer program." *Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts*.
- Tani T. und Ohshima Y. (1991). "mRNA-type introns in U6 small nuclear RNA genes: implications for the catalysis in premRNA splicing." *Genes Dev.*, 5:1022–1031.
- Thompson P.E. und English D.S. (1966). "Multiplicity of hemoglobin in the genus *Chironomus* (Tendipes)." *Science*, 152:75–76.
- Tichy H. (1970). "A biochemical and cytogenetic study on the nature of the hemoglobin-polymorphism in *Chironomus tentans* and *Chironomus pallidivittatus*." *Chromosoma*, 29(2):131–188.
- Tichy H. (1975). "Nature genetic basis and evolution of the haemoglobin polymorphism in *Chironomus*." *J. Mol. Evol.*, 6(1):39–50.

- 
- Tichy H. (1980). "The haemoglobin synthesising tissue of *Chironomus*." In M.D.A. (ed.) (Hg.), "Chironomidae Ecology. Systematics, Cytologie and Physiologie", New York; 43-50.
- Trewitt P.M. und Bergtrom G. (1981). "Immunological characterisation of the haemoglobins of *Chironomus thummi* (Diptera)." *Insect. Biochem.*, 11:635-644.
- Trewitt P.M., Boyer D.R. und Bergtrom G. (1986). "Characterisation of maternal haemoglobins in the eggs and embryos of *Chironomus thummi*." *J. Insect. Physiol.*, 32:963-969.
- Trewitt P.M., Luhm R.A., Samad F., Ramakrishnan S., Kao W.Y. und Bergtrom G. (1995). "Molecular evolutionary analysis of the ywvz/7B globin gene cluster of the insect *Chironomus thummi*." *J. Mol. Evol.*, 41:313-328.
- Trewitt P.M., Saffarini D.A. und Bergtrom G. (1987). "Nucleotide sequence of the intronless gene expressing a member of the globin VIIB subfamily from *Chironomus thummi* (Diptera)." *Nucleic Acids Res.*, 15(13):5494.
- Trewitt P.M., Saffarini D.A. und Bertrom G. (1988). "Nucleotide Sequence of the intronless gene expressing a member of globin VIIB subfamily from *Chironomus thummi* (Diptera)." *Nucleic Acid Res.*, 15:5494.
- Trotman C.N. (1998). "Introns-early: slipping lately?" *Trends Genet.*, 14(4):132-134.
- Tyshenko M.G. und Walker V.K. (1997). "Towards a reconciliation of the introns early or late views: triosephosphate isomerase". *Biochim Biophys Acta*, 1353(2):131-136.
- Vinogradov S.N. (1985). "The structure of invertebrate extracellular hemoglobins (erythrocrucorins and chlorocruorins)." *Comp. Biochem. Physiol.*, 103B:103B, 759-773.
- Vinogradov S.N., Walz D., Pohajdak B., Moens L., Kapp O.H., Suzuki T. und Trotman C.N.A. (1993). "Adventitious variability? The amino acid sequences of nonvertebrate globins." *Comp. Biochem. Physiol.*, 106B(1):1-26.
- Vinogradov S.N., Walz D.A. und Pohajdak B. (1992). "Organization of non-vertebrate globin genes." *Comp. Biochem. Physiol.*, 103(4):759-773.
- Vnencak-Jones C.L., Phillips I.J.A., Chen E.Y. und Seeburg P.H. (1988). "Molecular basis of human growth hormone gene deletions." *Proc. Natl. Acad. Sci.*, 85:5615-5619.
- Watterson G.A. (1983). "On the time for gene silencing at duplicate loci." *Genetics*, 105:745-766.
- Weber R.E. (1980). "Functions of invertebrate hemoglobins with special reference to adaptations to environmental hypoxia." *Am. Zool*, 20:79-101.

- Weber R.E., Braunitzer G. und Kleinschmidt T. (1985). "Functional multiplicity and structural correlations in the hemoglobin system of larvae of *Chironomus thummi thummi* (Insecta, Diptera): Hb components CTT I, CTT II beta, CTT III, CTT IV, CTT VI, CTT VIIb, CTT IX and CTT X." *Comp. Biochem. Physiol. B.*, 80:747–753.
- Weiner A.M., Deininger P.L. und Efstratiadis A. (1986). "Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information." *Ann. Rev. Biochem.*, 55:631–661.
- Weller P., Jeffreys A.J., Wilson V. und Blanchetot A. (1984). "Organization of the human myoglobin gene." *EMBO J.*, 3(2):439–446.
- Wiborg O., Hyldig-Nielsen J.J., Jensen E.O., Paludan K. und Marcker K.A. (1982). "The nucleotide sequences of two leghemoglobin genes from soybean." *Nucleic Acids Res.*, 10(11):3487–3494.
- Wolfe K.H., Sharp P.M. und Li W.H. (1989). "Rates of synonymous substitution in plant nuclear genes." *J. Mol. Evol.*, 29:208–211.
- Wolfe K.H. und Shields D.C. (1997). "Molecular evidence for an ancient duplication of the entire yeast genome." *Nature*, 387:708–713.
- Yamada Y., Avedimento V.E., Mudryj M., Ohkubo H., Vogeli G., Irani M., Pastan I. und Crombrugge B. (1980). "The collagen gene: Evidence for its evolutionary assembly by amplification of a DNA segment containing an exon of 54 bp." *Cell*, 22:887–892.
- Yamauchi K., Ochiai T. und Usuki I. (1992). "The unique structure of *Paramecium caudatum* hemoglobin gene: the presence of one intron in the middle of the coding region." *Biochem. Biophys. Acta*, 1171:81–87.
- Yanisch-Perron C., Vieira J. und Messing J. (1985). "Improved M13 phage cloning vektors and host strains: nucleotide sequences of the M13mp 18 and pUC 19 vectors." *Gene*, 33:103–119.
- Yokoyama S. und Yokoyama R. (1989). "Molecular evolution of human visual pigment genes." *Mol. Biol. Evol.*, 6(2):186–197.
- Zebe E. (1991). "In vivo-studies on the function of hemoglobin in the larvae of *Chironomus thummi* (Insecta, Diptera)." *Comp. Biochem. Physiol.*, 99A(4):525–529.
- Zhang J. und Nei M. (1996). "Evolution of Antennapedia-class homeobox genes." *Genetics*, 142(1):295–303.
- Zimmer E.A., Martin S.L., Beverley S.M., Kan Y.W. und Wilson A.C. (1980). "Rapid duplication and loss of genes coding for the alpha chains of hemoglobin." *Proc. Natl. Acad. Sci.*, 77(4):2158–2162.

# Anhang

Vergleich der Intergenregionen innerhalb der 7B-Subfamilie von *C. tentans* und *C. piger*

<i>C. tentans</i> Intergenregion	<i>C. piger</i> Intergenregion	Lokalisation	Länge Identit t (Bp)	Identit t (%)	Größen- unterschied (%)
2 hbY-hbW	2 hbY-hbW	2123-2280	149	78	10,8
3 hbW-hbV	3 hbW-hbV	3513-3626	114	82,1	7,0
4 hbV-hbZ	4 hbV-hbZ	4120-4930	811	69,5	27,3
5 hbZ-hb7B12	5 hbZ-hb7B7	5427-5559	133	77,0	4,7
6 hb7B12-hb7B13	5 hbZ-hb7B7	/	/	/	/
7 hb7B13-hb7B7	5 hbZ-hb7B7	9119-9246	128	74,8	16,4
8 hb7B7-hb7B6	6 hb7B7-hb7B6	9753-10544	792	79,3	36,6
9 hb7B6-hb7B5	7 hb7B6-hb7B5	11049-11360	312	79,2	6,8
10 hb7B5-hb7B10	8 hb7B5-hb7B8/10	13427-13784 3138-3242	358 104	76,6 84	7,4
11 hb7B10-hb7B9	8 hb7B5-hb7B8/10	17970-18077	108	84,0	1,9
12 hb7B9-hb7B8	8 hb7B5-hb7B8/10	19822-19841 19864-19938	18 75	88,9 88,5	9,0
13 hb7B8-hb7B4	9 hb7B8/10-hb7B4	20446-21260	815	75,8	15,0
14 hb7B4-hb7B11	/	/	/	/	/

Vergleich der Intergenregionen innerhalb der 7B-Subfamilie von *C. tentans* und *C. thummi*

<i>C. tentans</i> Intergenregion	<i>C. thummi</i> Intergenregion	Lokalisation	Länge Identit t (Bp)	Identit t (%)	Größen- unterschied (%)
2 hbY-hbW	1 hbY-hbW	2123-2280	158	79,4	10,8
3 hbW-hbV	2 hbW-hbV	3502-3626	125	82,1	6,4
4 hbV-hbZ	3 hbV-hbZ	4120-4204	85	77,8	1,2
5 hbZ-hb7B12	4 hbZ-hb7B7	5437-5559	123	73,8	4,7
6 hb7B12-hb7B13	4 hbZ-hb7B7	/	/	/	/
7 hb7B13-hb7B7	4 hbZ-hb7B7	9178-9255	78	82,7	6,2
8 hb7B7-hb7B6	5 hb7B7-hb7B6	9753-10544	792	76,1	35
9 hb7B6-hb7B5	6 hb7B6-hb7B5	11049-11332 12778-12907	284 130	78,3 83,3	3,2 10,8
10 hb7B5-hb7B10	7 hb7B5-hb7B10	13427-13787 16554-16658	361 105	80 88,2	7 4,7
11 hb7B10-hb7B9	8 hb7B10-hb7B9	17970-18715	746	75	20
12 hb7B9-hb7B8	9 hb7B9-hb7B8	192214-19298 19770-19933	75 164	86,4 85,3	9 31
13 hb7B8-hb7B4	10 hb7B8-hb7B4	20446-21260	815	76	16,3
14 hb7B4-hb7B11		1-528	528	76,6	21,9

Alignment der Nukleotid-Sequenzen der H moglobine von *C. tentans*,  
*C. thummi*, *C. piger*, *C. melanotus* und *Glyptotendipes barbipes*

	1	60
CteP	ATGAAGTTCCTT--ATTCTTGCTGCCGCTACTAGTTGCTGGTGCCTATGCC-----ACA	
CteQ	ATGAAATTTATTCATTATTATTGCAATGTCATTATGAGTTCAACTTGTGATTTGACACCA	
CteQa	ATGAAATTTATTCATTATTATTGCAATGTCATTATGAGTTCAACTTGTGATTTGACACCA	
Cte8	ATGAAATTCCTC--GTAAGTTCCTTTGTCATTGCGCCGAGCATCAGCGCTGTGTACACCA	
CmeORF6	ATGAAATTTCTC--GTAAGTTCCTTTGTCATTGCGCCGAGCATCAGCGCTGTGTACACCA	
CteR	ATGAAATTTGTTA--ATTCTCGTTTGTGTCATTGCTGCAGCCAAATGTGATCTGGTFTTA	
CmeORF5	ATGAAATTCCTT--ATTCTTGCTTTGTCATGTTTACAACCACTTGTGATTTGGTFTCTT	
CteS	ATGAAATTCCTA--ATTTTGGCTTTGTCATTGTTACAGTTAGTTGTGTCTCGTCCCA	
CmeORF7	ATGAAATTCATA--GTTCTTGCTTTGTCATTGTTGAGCTTACATGTGATCCGGCTCCA	
Cte11	ATGAAATTTATTT--ATTTTATTTTGTGCAATGCTGCTGTTAGTTGCGACCTTATTTCA	
Ctt11	ATGAAGTTTTTA--ATTCTTGCTTTGTCATTGCTGCAGCCAGTTGTGATCTTATTTCCA	
CmeORF2	ATGAAGTTGTTT--ATTTTACTTTGTCATTGTTGAGCCAGTTGTGACCTTATTTCCA	
Cte12.1	ATGAAATTTATTT--ATTTTATTTTGTGCAATGCTGTAACCAAGTTGTGATCTTGTCTAA	
Cte12.2	ATGAAATTTATTT--ATTTTATTTTGTGCAATGCTGTAACCAAGTTGTGAT-----CAA	
Ctt12	ATGAAGTTGTTT--ATATTCACATTTATGATTTGTTGAGCCAGTTGTGACCTTGTCTCAA	
CtehorfA	ATGAAATTCCTTGTGTTGCTTCTGCTCTGTCATTGCGCCGAGCATCAGCTGTGCTGACCA	
CteORFB_ps	TTGAAATCAGTT-----TTAACTTTGTCATTGCTGCTCAGTTGTGACATTTGTTCCA	
CmeORFB	ATGAAATTTGTTT--ATTTTGCATTTGTCATTGTTGAGCTGCTGCTGTAACCTTATTTCA	
Ctt13	ATGAAATTCCTA--ATTTTGGCTTTGTCATTGCGCCGAGCTTACAGCCGCTGTTGTTCCA	
Cte2E	ATGAAATTCCTA--GTTCTTGCTCTGTCATTGTTGAGCCAGTTGTGACCTTATTTCCA	
Cte2E	ATGAAATTCCTA--GTTCTTGCTCTGTCATTGTTGAGCCAGTTGTGACCTTATTTCCA	
Cme2E	ATGAAATTCCTA--GTTCTTGCTCTGTCATTGTTGAGCCAGTTGTGACCTTATTTCCA	
Cte9	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCGTCGCGCAT-----CCA	
Ctt9.1	ATGAAATTCCTTATTTGTTCTGCTCTTGCATGTCGTTGCCATCGCCGAT-----CCA	
Cme9	-----GAT-----CCA	
Cte7A1	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7A2	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7A3	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7A4	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7A5	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7A6	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7A7	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7A8	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctt7A.1	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte1	ATGAAATTCCTG--ATTCTTGCAATGTCGCTGCTGCAAGTTATGCTGCT-----CCA	
Ctt1	ATGAAATTCCTA--ATTCTTGCAATGTCGCTGCTGCAAGTTATGCTGCT-----CCA	
Ctt1A	ATGAAATTCCTA--ATTCTTGCAATGTCGCTGCTGCAAGTTATGCTGCT-----CCA	
CteT	ATGAAATTCCTGCTATTCTGCTTTCTGCTGTTGCTGGTGGCCATCGCCGGA-----CCA	
CteU	-----	
Ctt-6.3	ATGAAATTCCTA--GTTCTTGCAATGTCGCTGCTGCAAGTTATGCTGCT-----GTT	
CteY	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CtpY	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CteW	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CttW	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CtpW	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CteV	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CttV_ps	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CtpV	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CteZ	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CttZ	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CtpZ	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B12	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B13	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B7	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B6	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B5	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B5a	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B10	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B10a	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B9	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B9a	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B8	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B4	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Cte7B11	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctt7B7	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctt7B6	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctt7B5	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctt7B10	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctt7B9	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctt7B8	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctt7B4	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctp7B7	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctp7B6	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctp7B5	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctp7B8_10	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
Ctp7B4	ATGAAATTCCTGCTGTTCTGCTCTGTCATTGCGTGGTGGCCATCGCCACT-----CCA	
CteN	ATGAAGTTTATA--GTAATAGCAATGTCATAGCTTGTGCTACTGTAACCCACATTTGG	
Gly16.4_ORF2	ATGAAATTCCTA--GTTCTTGCAATGTCGCTGTTGCGCCGCTGTAAGCAGATCCACAGTAT	
Gly16.4_ORF3	ATGAAATTCCTT--ATCCTGCAATGTCGCTGCTGCTGCTGATGCGGTGATCCACATTTGG	
Gly5-4.5_ORF2	ATGAAATTTGTTTTAATCACATGTCTGTCATTGTCGAGTGTACATGCGGATATTACACCA	
Gly5-4.5_ORF3	ATGAAGTTCTTT--ATTCTGCTGCTGATATGTCGCTGGTGGCTACGCT-----ACC	
Gly5-4.5_ORF1	ATGAAAGTCTTCGCGTTCTGCTCTGTCATTGCTGCAAGTTGCTGAGCTCACTGCGCCATCAACCA	
Gly5-2.2ORF1	ATGAAAGTCTTCGCGTTCTGCTCTGTCATTGCTGCAAGTTGCTGAGCTCACTGCGCCATCAACCA	

CteP TTAACAGCTGATGAGGCAAG-----CTCGTAAAAGGATCATGGGATAAAGTT  
CteQ CTTAACAGTGAGGAAGCTGAA-----ACTTTAAGAACCCTTATGGAATCGAGTA  
CteQa CTTAACAGTGAGGAAGCTGAA-----ACTTTAAGAACCCTTATGGAATCGAGTA  
Cte8 ATGACAGCTGATCAGCTTGCT-----CTCTCAAGTCATCATGGGCTGAAGTT  
CmeORF6 ATGACAGCTGATCAGCTTGCT-----CTCTCAAGTCATCATGGGCTACAGTT  
CteR TTAACAGCTGATCAAGGAGCC-----ATTGTTCATTCATCATGGAATCAAGTC  
CmeORF5 TTGACATCTGAACAAGGAGCT-----ATCGTTACCTCATCATGGAATCAAGTC  
CteS TTGACTGCTGATCAAGCTAAA-----ATAGTCAAATCCTCATGGGCTGAAGTT  
CmeORF7 TTGTCAGCTGATCAAGTAACA-----CTCGTTCGTAGCTCATGGAATCAAGTC  
Cte11 TTGGTAGATGATGAAGCAACT-----CTTATCAGATCATCATGGGATCAGGTT  
Ctt11 TTAGCTGACGATCAAGCAATC-----TTAATTAGATCATCATGGGATGAAGTC  
CmeORF2 TTAGTAGATGATGAAGAAACT-----ATTATTAGGTCATCATGGGCTCAAGTC  
Cte12.1 TTTGTGGAGGACCAAACTGCT-----CTCATCAGATCATCATGGAATGAAGTT  
Cte12.2 TTTGTGGAGGACCAAACTGCT-----CGCATCAGATCATCATGGAATGAAGTT  
Ctt12 TTCGTTGAAGACCAAACTGAA-----ATCATTAGAGCATCATGGAATCAAGTG  
Ctehorfa TTGTCTGCTGACCAAGCTTCA-----CTTGTCAAGTCATCATGGAACCAAGTC  
CteORFB\_ps TTGGAAGATGATCAAGCAATT-----CTTATTAGATCATCATGGGATCAGGTT  
CmeORFB ACTGTGGATGAACAATCTGCAGCAGATCATCTGCTTTATCAGAGCATCATGGAATCAAGTC  
Cte28 TTATCAGCTGATCAAGCTTCA-----CTTGTCAAATCTCATGGGCTCAAGTT  
Cte28 TTGTCAGCTGATGAAGCTTCA-----TTGTCCTGTCATCATGGAATCAAGTC  
Ctt28 TTGTCAGCTGACGAAGCTTCA-----CTTGTCTGTCATCATGGGCTCAAGTT  
Cme28 TTGTCAGCTGATCAAGCTTCA-----CTTGTCCGTCATCATGGACCAAGTC  
Cte9 ATTTACAGCTGACCAAGCTGAT-----TCATCCGTAGCTCATGGGCTCAAGTT  
Ctt9.1 GTTTCATCTGATCAAGCTAAT-----GCCATCCGTGATCATGGGCTGAAGTT  
Cme9 GTTTCATCTGACCAAGCTGCT-----CTCATCCGTAGCTCATGGGCTGAAGTT  
Cte7A1 TTGACTGCTGACCAAGCTTCA-----CTTGTCAAATCAACATGGGCTCAAGTT  
Cte7A2 TTGTCGCTGACCAAGCTTCA-----CTTGTCAAATCAACATGGGCTCAAGTT  
Cte7A3 TTGACTGCTGACCAAGCTTCA-----CTTGTCAAATCAACATGGGCTCAAGTT  
Cte7A4 TTGACTGCTGACCAAGCTTCA-----CTTGTCAAATCAACATGGGCTCAAGTT  
Cte7A5 TTGACTGCTGACCAAGCTTCA-----CTTGTCAAATCAACATGGGCTCAAGTT  
Cte7A6 TTGACTACTGACCAAGCTTCA-----CTTGTCAAATCAACATGGGCTCAAGTT  
Cte7A7 TTGACTGCTGACCAAGCTTCA-----CTTGTCAAATCAACATGGGCTCAAGTT  
Cte7A8 TTGTCGCTGACCAAGCTTCA-----CTTGTCAAATCAACATGGGCTCAAGTT  
Ctt7A.1 TTGTCGCTGACCAAGCTTCA-----CTTGTCAAATCAACATGGGCTCAAGTT  
Cte1 ---TCGTCGATCAAAATGCT-----GCCGCCAAGGTCATCATGGGCTCAAGTT  
Ctt1 ---TCAGGAGATCAAAATGCT-----GCTGCTAAGGTCATCATGGAACCAAGTT  
Ctt1A ---TCAGGAGATCAAAATGCT-----GCTGCTAAGGTCATCATGGAACCAAGTT  
CteT TTATCTGCTGACGAAGCTGCT-----CTTGTAAATCATCATGGGCTCAAGTT  
CteU -----  
Ctt-6.3 TTAACAACAGAAACAAGCTGAC-----TTGGTCAAAAAGCATGGGACCCGTT  
CteY -----TGTGATAATGCAGGC-----TCTCTAAAGGAAGCATGGGAAACATATT  
CtpY -----TGTGATGACTTCAAA-----ATAATGCAGGAAGCATGGAATCAAAATG  
CteW -----TGTGACAAAGCTTCT-----TTCATCAAGGCATCATGGGATCAGGTT  
CttW -----TGTGACAAAGCTCCG-----TTCATCAAGGCATCATGGAATCAAGTT  
CtpW -----TGTGACAAAGCTCCG-----TTCATCAAGGCATCATGGAATCAAGTT  
CteV TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGGCTTCATGGGACCAAGTT  
CttV\_ps TTGACTTCTGACGAAGCTAAT-----CTTGTAAATCATCATGGAACCAAGTT  
CtpV TTGACTTCTGACGAAGCTAAT-----CTCGTTAAGTCATCATGGAACCAAGTT  
CteZ TTGACTGCTGACCAAGCTTCA-----TTAGTCAAAGTTCATGGGACCAAAAT  
CttZ TTGACTTCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGGCTCAAGTC  
CtpZ TTGACTTCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGGCTCAAGTC  
Cte7B12 TTGACTGCTGACCAAGCTTCA-----CTCGTCAAGTCATCATGGGCTCAAGTT  
Cte7B13 TTGACTGCTGACCAAGCTTCA-----CTCGTCAAGTCATCATGGGCTCAAGTT  
Cte7B7 TTGTCAGCTGACGAAGCTTCA-----CTTGTAAATCATCATGGGATCAAGTC  
Cte7B6 TTGACCGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Cte7B6a TTGACCGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Cte7B5 TTGACCGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Cte7B5a TTGACCGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Cte7B10 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAACCAAGTT  
Cte7B10a TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAACCAAGTT  
Cte7B9 TTGACCGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAACCAAGTT  
Cte7B9a TTGACCGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAACCAAGTT  
Cte7B8 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Cte7B4 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Cte7B11 TTGACTGCTGATGAAGCTTCA-----CTCGTTAAGTCATCATGGAACCAAGTT  
Ctt7B7 TTGTCGCTGACGAAGCTTCA-----ATTGTAAAGTCATCATGGGATCAAGTC  
Ctt7B6 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Ctt7B5 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Ctt7B10 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Ctt7B9 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Ctt7B8 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Ctt7B4 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Ctp7B7 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAATCAAGTC  
Ctp7B6 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Ctp7B5 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Ctp7B8/10 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
Ctp7B4 TTGACTGCTGACGAAGCTTCA-----CTCGTCAAGTCATCATGGAAGGCTGTT  
CteN GATAATCTGGATGAAGTAGAAATGCA----AGCATCAAGGACATGGAATCAAGTC  
Gly16.4\_ORF2\_ GTTACATTTGGATGAAGTGAAGTTGGA----TTTGTCAAATCATCGTGGGATAAGATG  
Gly16.4\_ORF3 CTTATGATGAATGAAGCTGAAGTTGCT----AACATTAACACACATACGACTCAGTC  
Gly5-4.5\_ORF2 ATAAGTAGTGAAGAAGCTGCA-----GTAGTTAAATCAATGTTCTGAAGTA  
Gly5-4.5\_ORF3 TTGACTGCTGATGAAGCTAAA-----CTTGTAAATCATCATGGGATAAAGTT  
Gly5-4.5\_ORF1 TTGAGTCTGACGAAGGAAAC-----CTCGTCAAGTCATCATGGGCTACAGTC  
Gly5-2.2ORF1 TTGACTGCTGACGAAGGAAAC-----CTCGTCAAGTCATCATGGGATCAGTC

	121	180
CteP	AAAGGACAAGAAGACGGAATTCCTACGTAAATTTCAAGGAAAATCCAGACATTCAGGCC	
CteQ	AAGCATAGAGAAGCTGACATACTTTATGTCATTTTTAAGGAGAATCCAGACATTCAGGCT	
CteQa	AAGCATAGAGAAGCTGACATACTTTATGTCATTTTTAAGGAGAATCCAGACATTCAGGCT	
Cte8	AAGCACAACGAAGTTGACATCCTTATGCCATCTTCAAGGATCATCCAGATATCCAAGCT	
CmeORF6	AAGCACAACGAAGTTGACATCCTTATGCCATCTTCAAGGATCATCCAGATATCCAAGCT	
CteR	AAGCATAATGAAGTTGACATTCCTTATGCCGTTTTTAAGCAAACCCAGACATCCAAGCT	
CmeORF5	AAACATAATGAAGTTGACATTCCTTATGCTATCTTCAAGGATCATCCAGATATCCAAGCT	
CteS	AAGCACAACGAAGTTGACATTCCTTATGCCGTTTTTAAGCAAACCCAGACATCCAAGCT	
CmeORF7	AAGCATAACGAAGTTGACATCCTTATGCCATTTTTGCGGCTAATCCAGATATCCAAGCT	
Cte11	AAGCACAATGAAGTTGACATTCCTGCTGCAATTTCAAGGCAATCCAGACATTCAGGCT	
Ctt11	AAACATAATGAAGTTGACATTCCTACGCAATTTCAAGGCAATCCAGACATTCAGGCT	
CmeORF2	AAGCACAACGAAGTTGACATTCCTGCGCTGCTTCAAGGCAACCCAGACATTCAGGCT	
Cte12.1	AAACACAATGAAGTTGACATTCCTGCTCAATTTTTGCTGCTAATCCAGACATTCAGGCT	
Cte12.2	AAACACAATGAAGTTGACATTCCTGCTCAATTTTTGCTGCTAATCCAGACATTCAGGCT	
Ctt12	AAACATAATGAAGTTGACATCCTTATGCCATTTTTGCTGCTAATCCAGACATTCAGGCT	
CtehorfA	AAACACAACGAAGTTGACATTCCTGCTGCAATTTCAAGGCAACCCAGACATTCAGGCT	
CteORFB_ps	AGGCCAATGAAGTTGACATACTTGTGCAATTTCAAGGCAACCCAGACATTCGAA--G	
CmeORFB	AAGCATAATGAAGTTGACATTCCTGCTGCAATTTTTGCTGCTAATCCAGACATTCAGGCT	
Ctt13	AAACATAATGAAGTTGATATTCCTGCTGCTAATCCAGACATTCAGGCT	
Cte28	AAACATAATGAAGTTGATATCCTTACTACATCTTTAAGGCAACCCAGACATTCAGGCT	
Ctt28	AAGCACAATGAAGTTGACATCCTTACTACATCTTCAAGGCAACCCAGACATTCAGGCT	
Cme28	AAGCACAACGAAGTTGACATCCTTACTACATCTTTAAGGCAACCCAGACATTCAGGCT	
Cte9	AAGCACAATGAAGTTGACATCCTTGTCCGCTTCAAGGATCAACCCAGACATTCAGGCT	
Ctt9.1	AAGCACAACGAAGTTGACATTCCTGCTGCGCTTCAAGGATCAACCCAGACATTCAGGCT	
Cme9	AAGCACAACGAAGTTGACATCTAGCTGCGGTTTTGCGGATCAACCCAGACATTCAGGCT	
Cte7A1	AGAAATAACGAAGTTGACATCCTGCGCGCTTCAAGGCAACCCAGACATTCAGGCT	
Cte7A2	AGAAATAACGAAGTTGACATCCTGCGCGCTTCAAGGCAACCCAGACATTCAGGCT	
Cte7A3	AGAAATAACGAAGTTGACATCCTGCGCGCTTCAAGGCAACCCAGACATTCAGGCT	
Cte7A4	AGAAATAACGAAGTTGACATCCTGCGCGCTTCAAGGCAACCCAGACATTCAGGCT	
Cte7A5	AGAAATAACGAAGTTGACATCCTGCGCGCTTCAAGGCAACCCAGACATTCAGGCT	
Cte7A6	AGAAATAACGAAGTTGATATTCCTGCTGCTTCAAGGCAACCCAGACATTCAGGCT	
Cte7A7	AGAAATAACGAAGTTGACATCCTGCGCGCTTCAAGGCAACCCAGACATTCAGGCT	
Cte7A8	AGAAATAACGAAGTTGACATCCTGCGCGCTTCAAGGCAACCCAGACATTCAGGCT	
Ctt7A.1	AGAAATAGCGAAGTTGAAATCCTTGTGCTGCTTCAAGGCAACCCAGACATTCAGGCT	
Cte1	AAGGAAAGCAAGTTGATATCCTGCTGCAATTTCAAGGCAACCCAGACATTCAGGCT	
Ctt1	AAGAACAACCAAGTTGATATCCTTATGCTGCTTCAAGGCAACCCAGACATTCAGGCT	
Ctt1A	AAGAACAACCAAGTTGATATCCTTATGCTGCTTCAAGGCAACCCAGACATTCAGGCT	
CteT	AAGCACAATGAAGTTGACATCCTTATGCTGCTTCAAGGCTATCCAGATATTCAGGCT	
CteU	-----	
Ctt-6.3	AAATTCATGAAGTTGACATCCTTATGCTGTTTTCAAGGCTTACCCAGACATTCAGGCT	
CteY	AAGCAGACGAAGTTGAGATTCCTTATACCGCTTCAAGGCAATCCAGACATTCAGGCT	
CtpY	AAATAATGAAGTTGAAATCCTTACACTGTTTTCAAGGCTTATCCAGACATTCAGGCT	
CteW	AAGCACAATGAAGTTGATATTCCTTACAATGCTTCAAGGCAATCCAGACATTCAGGCT	
CttW	AAGCACAATGAAGTTGACATTCCTTACACTGCTTCAAGGCAATCCAGACATTCAGGCT	
CtpW	AAGCATAATGAAGTTGACATTCCTTACACTGCTTCAAGGCAATCCAGACATTCAGGCT	
CteV	AAGCACAACGAAGTTGATATCCTGCTGCGCTTCAAGGCTTACCCAGACATTCAGGCT	
CttV_ps	AAACACAATGAAGTTGACATTCCTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
CtpV	AAACACAATGAAGTTGACATTCCTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
CteZ	AAGCACAACGAAGTTGATATCCTTATGCTGCTTCAAGGCTTACCCAGACATTCAGGCT	
CttZ	AAACACAATGAAGTTGATATCCTTATACTGTTTTCAAGGCTTACCCAGACATTCAGGCT	
CtpZ	AAGCACAATGAAGTTGATATCCTTATACTGTTTTCAAGGCTTACCCAGACATTCAGGCT	
Cte7B12	AAGAATAATGAAGTTGACATTCCTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B13	AAGAATAATGAAGTTGACATTCCTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B7	AAGCATAATGAAGTTGACATCCTTGTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B6	AGCCACAATGAAGTTGACATCCTGCTGCGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B6a	AGCCACAATGAAGTTGACATCCTGCTGCGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B5	AGCCACAATGAAGTTGACATCCTGCTGCGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B5a	AGCCACAATGAAGTTGACATCCTGCTGCGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B10	AGCCACAATGAAGTTGACATCCTGCTGCGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B10a	AGCCACAATGAAGTTGACATCCTGCTGCGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B9	AGCCACAATGAAGTTGACATCCTGCTGCGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B9a	AGCCACAATGAAGTTGACATCCTGCTGCGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B8	AGCCACAATGAAGTTGACATCCTGCTGCGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B4	AGCCACAATGAAGTTGACATCCTGCTGCGCTTCAAGGCAATCCAGACATTCAGGCT	
Cte7B11	AGCCACAATGAAGTTGACATCCTTATGCTGTTTTCAAGGCTTACCCAGACATTCAGGCT	
Ctt7B7	AAGCACAATGAAGTTGACATTCCTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctt7B6	AGCCACAACGAAGTTGAAATCCTTGTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctt7B5	AGCCACAACGAAGTTGAAATCCTTGTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctt7B10	AGCCACAACGAAGTTGAAATCCTTGTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctt7B9	AGCCACAACGAAGTTGAAATCCTTGTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctt7B8	AGCCACAACGAAGTTGAAATCCTTGTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctt7B4	AGCCACAACGAAGTTGACATCCTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctp7B7	AAGCACAATGAAGTTGACATTCCTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctp7B6	AGCCACAACGAAGTTGAAATCCTTGTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctp7B5	AGCCACAACGAAGTTGAAATCCTTGTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctp7B8/10	AGCCACAACGAAGTTGAAATCCTTGTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
Ctp7B4	AGCCACAATGAAGTTGACATCCTGCTGCTTCAAGGCAATCCAGACATTCAGGCT	
CteN	AAAATGAATGAAATGAAATTCCTTACAATATTTCAAGGCAATCCAGGAAATATGCTCT	
Gly16.4_ORF2	AAGTTTAAACGAAGTTGAAATTCCTACTTTGCTTCAAGGAAATCCAGACATTCAGGAA	
Gly16.4_ORF3	AAGCACAATGAAGTTGACATCCTTACAACATCTTCAAGGCTTATCCAGACATTCAGGCT	
Gly5-4.5_ORF2	AAAAACAATGAAGTTGATATTTTGTACAGAATTTTTAAGGAAAATCCAGACATTCAGGCT	
Gly5-4.5_ORF3	AAGGGAACAGGATGATGCAATGCTCTATGCAATTTTCAAGGAAAATCCAGACATTCAGGCT	
Gly5-4.5_ORF1	AGAAATAGCGAAGTTGACATTCCTTGTGCTTCAAGGCAACCCAGACATTCAGGCT	
Gly5-2.2ORF1	AGAAATAGCGAAGTTGACATTCCTTGTGCTTCAAGGCAACCCAGACATTCAGGCT	



CteP AAATTTCCCTTCATTCGTTGGAAAGAACTTAGAAGAAATTAAGTCAAATGATGAC-----  
CteQ CATTTCCCGCCCTTGTGGAAAGAATCTTGAGGCATTGCGAGGATCAATTGCA-----  
CteQa CATTTCCCGCCCTTGTGGAAAGAATCTTGAGGCATTGCGAGGATCAATTGCA-----  
Cte8 CGTTTCCACAAATTCGCTGGAAAGGACTTGACTCATTGAAGGCACAGCTGCT-----  
CmeORF6 AAATTTCCACAAATTCGCTGGAAAGGACCTCGATTCTGTTAAAGATACCTGCTGCC-----  
CteR AAGTTCACAAATTTCTGGAAAGGACTCGACTCATTGAAGGCACAGCTGCT-----  
CmeORF5 AAATTTCCCAATTTGCTGGAAAGGACCTCGATTCTGAAATCTACTGCTTCA-----  
CteS AAGTTCACAAATTCGCTGGAAAGGACTCGATTCTGAAAGGATACAGCCGCT-----  
CmeORF7 CGTTTCCACAGTTTGTGGAAAGGACTGGCATCATTGAAAATCTACTGCTTCA-----  
Cte11 CGTTTCCACAAATTCGCTGGAAAGGACTCGACTCATTGAAGGCACAGCTGCT-----  
Ctt11 CGTTTCCACAGTTTGTGGAAAGGACTTGACTCAATAAAGACGACTGGTTCAG-----  
CmeORF2 CGTTTCCACAAATTCGCTGGAAAGGACTTGAAATCATTGAAGGATACCTGCTGCC-----  
Cte12.1 CGTTTCCACAGTTTGTGGAAAGGACTTGAAATCATTAAAAATTCGCGCGCT-----  
Cte12.2 CGTTTCCACAGTTTGTGGAAAGGACTTGAATCATTAAAAATTCGCGCGCT-----  
Ctt12 CGTTTCCCTCAGTTTGTGGAAAGGACTAAAAACATTGAAATCGTCTCTTCA-----  
Ctehorfa CGTTTCCACAAATTCGCTGGAAAGGACTTGCTGGATTGAAAGACACAGCTGCC-----  
CteORFB\_ps CACATTTCCACAAATTCGCTGGAAAGGACTTGAATCTTTAAAGGACACAGCTG-----  
CmeORFB CGTTTCCACAAATTCGCTGGAAAGGACTCGACTCATTAAAAACACAGTAGCT-----  
Ctt13 CGTTTCCCTCAATTTGCTGGCAAGGACTCGACTCAATCAAGGATACCTGGTCAACTGGT-----  
Cte28 AAGTTCACAAATTCGCTGGCAAGGACTCGACTCATTGAAGGCACAGCTGCA-----  
Ctt28 AAATTTCCACAAATTCGCTGGCAAGGACTCGAAACACTTAAAGGAACGGACAA-----  
Cme28 AAGTTCACAAATTCGCTGGCAAGGACTCGATTCTGAAATTAAGGATCTGCTTCA-----  
Cte9 CGTTTCCACAAATTCGCTGGAAAGGACTCGACTCATTGAAGGCACAGCTGCT-----  
Ctt9.1 CGTTTCCACAAATTCGCTGGAAAGGACTCGCTCAATCAAGGATACCTGGTGA-----  
Cme9 CGTTTCCACAAATTCGCTGGAAAGGACTCGCTGCTCTTAAAGGACACCGCTGCA-----  
Cte7A1 CGTTTCCACAAATTCGCTGGAAAGGACTTGCTGCAATGAAGGCACAGCTGCT-----  
Cte7A2 CGTTTCCACAAATTCGCTGGAAAGGACTTGCTGCAATGAAGGCACAGCTGCT-----  
Cte7A3 CGTTTCCACAAATTCGCTGGAAAGGACTTGCTGCTTGAAGGCACAGCTGCT-----  
Cte7A4 CGTTTCCACAAATTCGCTGGAAAGGACTTGCTGCTTGAAGGCACAGCTGCT-----  
Cte7A5 CGTTTCCACAAATTCGCTGGAAAGGACTTGCTGCTTGAAGGCACAGCTGCT-----  
Cte7A6 CGTTTCCACAAATTCGCTGGAAAGGACTTGCTGCTTGAAGGCACAGCTGCT-----  
Cte7A7 CGTTTCCACAAATTCGCTGGAAAGGACTTGCTGCTTGAAGGCACAGCTGCT-----  
Cte7A8 CGTTTCCACAAATTCGCTGGAAAGGACTTGCTGCTTGAAGGCACAGCTGCT-----  
Ctt7A.1 CGTTTCCACAAATTCGCTGGAAAGGACTTGCTTCAATCAAGGATACCTGGTGA-----  
Cte1 AAATTTCCACAAATTCGCTGGAAAGGACTCGACTCAATCAAGGAAACCCAGAA-----  
Ctt1 GCATTTCCACAAATTCGCTGGAAAGGACTCGACTCAATCAAGGAAACCCAGAC-----  
Ctt1A GCATTTCCACAAATTCGCTGGAAAGGACTCGACTCAATCAAGGAAACCCAGAC-----  
CteT AAGTTCACAAATTCGCTGGAAAGGACTTGGATGGATTGAAGGCACAGCTGCA-----  
CteU -----GATCTTGAATCAGTGAAGACTCACCCGCT-----  
Ctt-6.3 AAGTTCACAAATTCGCTGGAAAGGACTCGACTCAATCAAGGACTCGCCGCG-----  
CteY CGTTTCCCTAAAATTCGCTGGAAAGGACTTGAATCATTGAAGGACACCGCAGAC-----  
CtpY AAATTTCCACAAATTTGCTGGAAAGGACTTGAAGAAATCAAGGAACTGCTGAG-----  
CteW AAATTTCCACAAATTTGCTGGAAAGGACTCGAAACAAATCAAGGACACTGCTGTT-----  
CttW CGTTTCCCTCAATTTGCTGGAAAGGACTTGAAGAAATCAAGGAAACCGCCGAG-----  
CtpW CGGTTCCTCAATTTGCTGGAAAGGACTTGAAGAAATCAAGGAACTGCTGAG-----  
CteV AAATTTCCACAAATTTGCTGGAAAGGACTCGATTCAATTAAGACGAGTGGTCA-----  
CttV\_ps AAGTTCCTCAATTTTCCGAAAGGACTCGATTCAATTAAGACGAGTGGTCA-----  
CtpV AAGTTCCTCAGTTTGTGGAAAGGACTCGATTCAATTAAGACGAGTGGTCAA-----  
CteZ AAATTTCCACAAATTTGCTGGAAAGGACTCGAATCAGTCAAGGACACAGCTGCT-----  
CttZ CGTTTCCCTCAGTTTGTGGAAAGGACTCGATCAATTAAGACAACTGGTCAA-----  
CtpZ CGTTTCCCTCAGTTTGTGGAAAGGACTCGATTCAATTAAGACAACTGGTCAA-----  
Cte7B12 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B13 AAATTTCCGCAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B7 AAATTTCCACAAATTCGCAAGGAAAGGACTCGCTGCCATCAAGGACACAGCTGCT-----  
Cte7B6 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B6a AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B5 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B5a AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B10 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B10a AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B9 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B9a AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B8 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCT-----  
Cte7B4 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTGCCGCTCAAGGACACAGCTGCC-----  
Cte7B11 AAATTTCCACAAATTCGCTGGAAAGGACTCGACTCCGCTCAAGGAAACAGCTGCT-----  
Ctt7B7 AAGTTCACAGTTTGTGGAAAGGACTCGCTTCAATTAAGGATACCTGCCGCA-----  
Ctt7B6 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTTCAATCAAGGATACCTGGTGA-----  
Ctt7B5 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTTCAATCAAGGATACCTGGTGA-----  
Ctt7B10 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTTCAATCAAGGATACCTGGTGA-----  
Ctt7B9 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTTCAATCAAGGATACCTGGTGA-----  
Ctt7B8 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTTCAATCAAGGATACCTGGTGA-----  
Ctt7B4 AAGTTCACAAATTCGCTGGAAAGGACTCGCTTCAATCAAGGATACCTGGTGA-----  
Ctp7B7 AAGTTCACAAATTCGCTGGAAAGGACTCGCTTCAATTAAGGATACCTGCTGCA-----  
Ctp7B6 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTTCAATCAAGGATACCTGGTGA-----  
Ctp7B5 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTTCAATCAAGGATACCTGGTGA-----  
Ctp7B8/10 AAATTTCCACAAATTCGCTGGAAAGGACTCGCTTCAATCAAGGATACCTGGTGA-----  
Ctp7B4 AAGTTCACAAATTCGCTGGAAAGGACTCGCTTCAATCAAGGATACCTGGTGA-----  
CteN AAATTTCCACAAATTCGCTGGAAAGGACTCGAAGAAATGAAGGAAACGAAACCAA-----  
Gly16.4\_ORF2 AAATTTCCACAAATTCGCTGGAAAGGACTCGATCAAAATTAAGGATCAAGCCGCT-----  
Gly16.4\_ORF3 CGTTTCCAAAATTCGCTGGAAAGGACTTGAACATTAAGGAAACAGCTGAT-----  
Gly5-4.5\_ORF2 CGTTTCCCGCCCTTGTGGAAAGGACTTGAAGAAATTAAGGCTGGAGTTCCCT-----  
Gly5-4.5\_ORF3 AAGTTCACAGCAATTTGTGGAAAGGACTTGAAGGAAATTAAGTCACTGACGAT-----  
Gly5-4.5\_ORF1 CGTTTCCACAAATTCGCTGGAAAGGACTCAATGGAATGAAGGACACAGCTGCT-----  
Gly5-2.2ORF1 CGTTTCCACAAATTCGCTGGAAAGGACTCAATGGATTGAAGGACACAGCTGCT-----

CteP ---TTCACAAAGCATGCTGACAGAAATTGTCGCTGCCGTCACAAAAACGTCGATTTGGT  
CteQ ---TTTGCCATTCATGCAACCAGAAATGTCAGCATCTTCTCAAAAAATGCTTCACTAGCT  
CteQa ---TTTGCCATTCATGCAACCAGAAATGTCAGCATCTTCTCAAAAAATGCTTCACTAGCT  
Cte8 ---TTCGCTACACATGCTGGTAGAAATGTTGGATTCTTCTCAGATGTTATGGACTTATF  
CmeORF6 ---TTTGCCACCCACGCTGGTAGAAATGTCGGATTCTTCTCAGAAATGTTATGGACTTATF  
CteR ---TTCGCTACACATGCCGGAAGAAATGTCGGATTGCTCTCAGAAATGTTATGGCCCTGGCT  
CmeORF5 ---TTTGCAATCCATGCAGGTAGAAATGTCGGATTGCTCTCAGAAATGTTATGCTCTGGCT  
CteS ---TTCGCCACACACGCCGGAAGAAATGTCGGATTGTTTTCAGAAATGTTATTTACTTATF  
CmeORF7 ---TTTGCCACCCATGCCGGAAGAAATGTCGGATTCTTCTCAGAAATGTTATTTACTTATF  
Cte11 ---TTCGCCACACATGCCGGAAGAAATGTTGGATTCTTCTTTCAGAAATGCTAGTCTTTCA  
Ctt11 ---TTCGCAGTTTCATGAGGAAGAAATGTTAGGATCTTCTCAGAAATGTTCCGCTTTCAG  
CmeORF2 ---TTTGCCACCCACGCTGGTAGAAATGTCGGATTCTTCTCAGAAATGCTCAGCATTTCA  
Cte12.1 ---TTTGCTACACATGCAGCCAGAAATGTTGGATTCTTCTCAGAAATGTTATGAAATGAG  
Cte12.2 ---TTTGCTACACATGCAGCCAGAAATGTTGGATTCTTCTCAGAAATGTTATGAAATGAG  
Ctt12 ---TTTGCTACACATGCTGGCAGAAATGTTAGGCTTCTTTCAGAAATGTTACTGAACTTAA  
Ctehorfa ---TTCGCCACACATGCCGGAAGAAATGTCGGATTCTTCTCAGAAATGTTATGGACTTACT  
CteORFB\_ps ---CTTTGCCATACACGCCGCGGCAATGTTGGATTCTTCTTTCAGAAATGCTGGCTTCTT  
CmeORFB ---TTTACTACTCATGCAGTTAGAAATGTTGGCTCTTCTTTCAGAAATGTTACTGAACTACT  
Ctt13 GCATTTGCCACACATGCTGGTAGAAATGTCGGATTCTTCTTTCAGAAATGCTGGACTTATF  
Cte2f ---TTCGCCACACATGCCGGAAGAAATGTCGGATTCTTCTCAGAAATGTTATGCTTATG  
Ctt2f ---TTCGCTACTCATGCTGGAGAAATGTCGGATTGCTCTCAGAAATGTTGCTCTTATG  
Cme2f ---TTTGCCACCCATGCCGGAAGAAATGTCGGATTGCTGAGTGAGATGTTGCTCTTATG  
Cte9 ---TTCGCCACACACGCCGGAAGAAATGTCGGATTCTTCTCAGAAATGTTGCTCTTGT  
Ctt9.1 ---TTCGCCACACACGCCGGAAGAAATGTCGGATTCTTCTCAGAAATGTTGCTCTTGT  
Cme9 ---TTCGCCACCCACGCCGGAAGAAATGTCGGATTGCTTTCAGAAATGTTGCTCTTGT  
Cte7A1 ---TTCGCCACACACGCCGGAAGAAATGTCGGATTGCTTTCAGAAATGTTGCTCTGCT  
Cte7A2 ---TTCGCCACACACGCCGGAAGAAATGTCGGATTGCTTTCAGAAATGTTGCTCTGCT  
Cte7A3 ---TTCGCCACACACGCCGGAAGAAATGTCGGATTGCTTTCAGAAATGTTGCTCTGCT  
Cte7A4 ---TTCGCCACACATGCCGGAAGAAATGTCGGATTGCTTTCAGAAATGTTGCTCTGCT  
Cte7A5 ---TTCGCCACACACGCCGGAAGAAATGTCGGATTGCTTTCAGAAATGTTGCTCTGCT  
Cte7A6 ---TTCGCCACACACGCCGGAAGAAATGTCGGATTGCTTTCAGAAATGTTGCTCTGCT  
Cte7A7 ---TTCGCCACACACGCCGGAAGAAATGTCGGATTGCTTTCAGAAATGTTGCTCTGCT  
Cte7A8 ---TTCGCCACACACGCCGGAAGAAATGTCGGATTGCTTTCAGAAATGTTGCTCTGCT  
Ctt7A.1 ---TTCGCCACACACGCCGGAAGAAATGTCGGATTGCTTTCAGAAATGTTGCTCTGCT  
Cte1 ---TTCGCCACACACGCTGGTAGAAATGTTGGATTCTTTCAGAAATGTTGCTCTGCT  
Ctt1 ---TTTAGCAAGCATGCTGGTAGAGTAGTCGGATTATTTCTGAAAGTTATGGACTCTCC  
Ctt1A ---TTTAGCAAGCATGCTGGTAGAGTAGTCGGATTATTTCTGAAAGTTATGGACTCTCC  
CteF ---TTTGCCACACACGCCAGAAATGTTTTCATTTCTGAAAGTTATGGACTCTCC  
CteU ---TTCGCCACACATGCCACTAGAAATTTTCATTTCTTGTCTCAGAAATGTTGCTCTGCT  
Ctt-6.3 ---TTTGCAACTCATGCAACACGATAGTATCATTTCTTGTCTCAGAAATGTTGCTCTGCT  
CteY ---TTCGCCATCCATGCAACAGAAATTTATGGCTTCATGACTGAAAGTTATGGACTCTCT  
CtpY ---TTTGCACTCCATGCAACTAGAAATGTCAGTTTATGACAGAAATTTACTCTCTCT  
CteW ---TTTGCCATACACGCCACAGAAATGCTTTCATTCATGTCGAAATGTTAGCTTGT  
CttW ---TTTGCACTCCATTCACAGAAATGTTTTCATTTATGCTGAAATGTTAAGCTTGT  
CtpW ---TTTGCACTTCATTCACAGAAATGTTTTCATTTATGCTGAAATGTTAAGCTTGT  
CteV ---TTCGCCACACACGCTTACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTACTCTCT  
CttV\_ps ---TTTGCAACTCATGCAACTGCAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
CtpV ---TTCGCCAECTCATGCAACTGCAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
CteZ ---TTTGCCACACACGCCACAGAAATGCTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
CttZ ---TTCGCCAECTCATGCAACTGCAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
CtpZ ---TTTGCAACTCATGCAACTGCAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B12 ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B13 ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B7 ---TTCGCCACACACGCCACAGAAATGCTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B6 ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B6a ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B5 ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B5a ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B10 ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B10a ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B9 ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B9a ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B8 ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B4 ---TTCGCCACACACGCCACAGAAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Cte7B11 ---TTCGCCACACATGCCACAGAAATGCTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctt7B7 ---TTCGCCAECTCATGCAACTGCAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctt7B6 ---TTCGCCACACACGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctt7B5 ---TTCGCCACACACGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctt7B10 ---TTCGCCACACACGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctt7B9 ---TTCGCCACACACGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctt7B8 ---TTCGCCACACACGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctt7B4 ---TTCGCCACACACGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctp7B7 ---TTCGCCAECTCATGCAACTGCAATGCTCTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctp7B6 ---TTCGCCACACACGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctp7B5 ---TTCGCCACACACGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctp7B8/10 ---TTCGCCACACACGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Ctp7B4 ---TTCGCCACACACGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
CteN ---TTTGCAATTCATGCTACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTGCTCTCT  
Gly16.4\_ORF2\_ ---TTCGCGTTCATGCCACAGAAATGTTTTCATTTCTTGTCTCAGAAATGTTACTCTTA  
Gly16.4\_ORF3 ---TTCGCCCTTCATGCTGGAAGAAATGTTGGTTTGTCTCACTGAAATACATTTACTCTCT  
Gly5-4.5\_ORF2 ---TTTGCACTTCATGCTACAGAAATGTCAGTTTCTTTCAGAAATGTTATCAGATTAAGT  
Gly5-4.5\_ORF3\_ ---TTCACAAACATGCTGACAGAAATGTTGCTGCAAGTTACAAAAACGTCGATTAAGT  
Gly5-4.5\_ORF1 ---TTCGCTACACACGCTGGAAGAAATGTTGGTTTCTTGTCTCAGAAATGTTGCTCTCT  
Gly5-2.2ORF1 ---TTCGCTACACACGCTGGAAGAAATGTTGGTTTCTTGTCTCAGAAATGTTGCTCTCT

CteP GGC AATGAAGCTAACACACCAGCTATCAAGACCCCTCCTTAAACGAATTAGGACAAAAGCCAC  
CteQ GGAGATTC AACAATATTC CAGCTTCAA AAAACGTTAATCAATGAATTGGCAGCAAGCCAT  
CteOa GGAGATTC AACAATATTC CAGCTTCAA AAAACGTTAATCAATGAATTGGCAGCAAGCCAT  
Cte8 GAAACTCAGAAAACCGTCCAGCACTTAAGACCCTCATCGATGGATTGGCATCAAACCCAC  
CmeORF6 GAAATCCAGAAAACCGTCCAGCCCTTAAGACTCTCATTTGATGCATTGGCATCAAGCCAC  
CteR GATAGTGAATCGAATGTTGCCGCCCTCAAAAATCCTTATCAACGGCTTAGCCGAAAACCCAC  
CmeORF5 GACAGTGAATCAAAATGTTGCTGCATCGAAAATCCTTATCAATGATATGGCAACTAGACAT  
CteS GAAGGTGAATCAAATGTTGCCGCCCTCAAAAATCAATTAATCAATGACATAGCCGCAAGCCAC  
CmeORF7 GAAGGCCAATCAAATATTCCTGCCTCGAAAATATCTCATTAACGATTTGGCCGCAACTCAT  
Cte11 ACAGACGAATCAAGTATTTCCAGCTGCAAAAACCTCATTAATCAATTTGGGTGCTAGTCAAT  
Ctt11 CCAGACGAATCAGGTATCTCAGCAGCAAAGACTCTCATAAATGCAATTTGGGAGCTAGTCAAT  
CmeORF2 ACAGACGCATCAAGTGTTCAGCTGCAAAAAGACTATCATTAATGAATTTGGGTGCTAATCAC  
Cte12.1 CCAAACGATTCAGGTGTTTCCGCCGCAAAAACCTCATCAATCAGTTAGCTGTTAGTCAAC  
Cte12.2 CCAAACGATTCAGGTGTTTCCGCCGCAAAAACCTCATCAATCAGTTAGCTGTTAGTCAAC  
Ctt12 CCAAATGATTCAGGTGTTTCCAGCTGCAAAAACCTTAATAAATGAAGTAGCTGCTAGTCAAT  
Ctehorfa GAAACCGCAGCTAATGCCCCAGCCCTCAAACACTCGTCGGACAACCTCGCTGCCAGCCAC  
CteORFB\_ps CAAATGATGAATCAGGTGTTCCAGCTGCAATGACTCTC-AAATCAATTTGGGTGCTAGTCAAT  
CmeORFB CCAAATGACTCTGGTGTTCAGCAGCAAAGACCCTTAATCAACAGTTAGCTGTTAGTCAAC  
Ctt13 GAAATCAATCAAATGTCAGCACTTTACACTCTTACCAGAACACTTGCTGCCAGCCAT  
Cte2B GGAATGAAGCAAACAGACCAGCCATGGTAACTTTATCAACGAAAATGGCTGCCAACCCAC  
Ctt2B GGTAAATCTGCTAACATGCCAGCTATGGAAAACCTCATCAAAGACATGGCTGCCAACCCAC  
Cme2B GGC AATTC TCGTAAACATGCCAGCAATGGAAAACCTTTGACTCACTGAAAATGGCCGGTAAACCCAC  
Cte9 GAAACGAAAGCTAACCGCCAGCCATGAACACCCTTACC AAGCAATTTGGCAACTTAACCAT  
Ctt9.1 GAAACGAAATCAAATGCCCCAGCCATGGCTACACTCATCAAATGAATTTATCAACAAAGCCAT  
Cme9 GGAATGAAGCAAACAGCAACAGCCATGGCTACTTTAATCAACGAAAATGGCCACAACCCAC  
Cte7A1 GAAACGAAATCAAATGCCCCAGCTGTCAAACACTCGTCGGACAACCTCGCTGCCAGCCAC  
Cte7A2 GGAACCAATCAAATGCCCCAGCTGTCAAACACTCGTCGGACAACCTCGCTGCCAGCCAC  
Cte7A3 GAAACGAAATCAAATGCCCCAGCTGTCAAACACTCGTCGGACAACCTCGCTGCCAGCCAC  
Cte7A4 GAAACGAAATCAAATGCCCCAGCCACCAAACACTCGTCGGACAACCTCGCTGCCAGCCAC  
Cte7A5 GAAACCAATCAAATGCCCCAGCTGTCAAACACTCGTCGGACAACCTCGCTGCCAGCCAC  
Cte7A6 GAAACGAAATCAAATGCCCCAGCCACCAAACACTCGTCGGACAACCTCGCTGCCAGCCAC  
Cte7A7 GAAACCAATCAAATGCCCCAGCTGTCAAACACTCGTCGGACAACCTCGCTGCCAGCCAC  
Cte7A8 GAAACCAATCAAATGCCCCAGCTGTCAAACACTCGTCGGACAACCTCGCTGCCAGCCAC  
Ctt7A.1 GAAACGAAATCAAATGCCCCAGCACTTCAAACCTTAGTCGGACAACCTCGCAGCTAGCCAC  
Cte1 GGAATGATGCCAACACAGCAACAATTTGATCAAAGGCCAAGGACTTAGGAAAAGACACAC  
Ctt1 GGAACGATGCCAACACACCAACAATCTCGCAAAGGCCAAGGACTTAGGAAAAGTCTCAC  
Ctt1A GGAACGATGCCAACACACCAACAATCTCGCAAAGGCCAAGGACTTAGGAAAAGTCTCAC  
CteT GGAAGTGA AACAATTTGGCTGCAGTTCATACACTCGTCACAAAATTTAGGAAAAGGATCAT  
CteU GGAAGTGAATGATACCAATTCCTGCAATCAA AATCTTTGCTAAGGATTTGGCCACATCACAC  
Ctt-6.3 GGAAGTGAATGATACCAATTCCTGCAATCAA AATCTTTGCTAAGGATTTGGCCACATCACAC  
CteY GGTAAATGACGCGAATACGCCAGCAATCAA AACTCTCCTCAATCAGTTGGCCAAAGAACAC  
CtpY GGTAAATCTGACAATCTTCCAGCAATCATGTCTCTCCTCAATTAATTTGGGAAAAGATCAT  
CteW GGAATTCATCAGTTC A A A A C T A A T C G A C T C A C T T C T G T A A A A T T G G C T A A T G A T C A T  
CttW GGAATCCAGCAGTTCAATCATCTATCGACTTACTTCTTGTGAAAATGGCAAATGATCAT  
CtpW GGAATCCAGCAGTTCAATCATCTATCGACTTACTTCTTGTGAAAATGGCAAATGATCAT  
CteV GGAATGAAGGAAAACCTTTCCAGCTGTCTATGGACTCGTCAGAAAATTTAGGAGTTGACCAC  
CttV\_ps GGAATGAAGCAAACCTTTCTGCTGTTTATGGACTTGTCAA AAAATTTAGGAGTTGATCAC  
CtpV GGAATGAAGCAAACCTTTCCAGCTGTCTATGGACTCGTC AAAAAATTTAGGAGTTGATCAC  
CteZ GGAAGCGAAGCAATCTTACAGCAATCTATGGACTCGTCA ACAAATTTGGGAGAGACCAC  
CttZ GGAAGCGAAGCAATCTTTCCAGCCATCTATGGACTTATCTCAAAGATGGGAACAGACCAC  
CtpZ GGAAGCGAAGCAATCTTTCCAGCCATCTATGGACTTATCTCAAAGATGGGAACAGACCAC  
Cte7B12 GGAACCTCGCCAAATGCTGCTGCCATCAACTCACTCGTCACAAAAGTTGGGATCAGATCAT  
Cte7B13 GGAACCTCGCCAAATGCTGCTGCCATCAACTCACTCGTCACAAAAGTTGGGATCAGATCAT  
Cte7B7 GGTGATCAAGCAATCTTTCTGCCGCTCAACGCCCTCGTCTCAA AATTTGGGAGCTGATCAC  
Cte7B6 GGAACGAAATCAAACCAAGCCGCTAATTAATGGACTCGTCTCAA AATTTGGGATCAGACCAC  
Cte7B6a GGAACGAAATCAAACCAAGCCGCTAATTAATGGACTCGTCTCAA AATTTGGGATCAGACCAC  
Cte7B5 GGAACGAAATCAAACCAAGCCGCTAATTAATGGACTCGTCTCAA AATTTGGGATCAGACCAC  
Cte7B5a GGAACGAAATCAAACCAAGCCGCTAATTAATGGACTCGTCTCAA AATTTGGGATCAGACCAC  
Cte7B10 GGAATGAATCAAACCAAGCCGCTCGCCTCACTCGTCGCCAAGTTGGGAGATGACCAC  
Cte7B10a GGAATGAATCAAACCAAGCCGCTCGCCTCACTCGTCGCCAAGTTGGGAGATGACCAC  
Cte7B9 GGAACGAAATCAAACCAAGCCGCTCGCCTCACTCGTCGCCAAGTTGGGAGATGACCAC  
Cte7B9a GGAACGAAATCAAACCAAGCCGCTCGCCTCACTCGTCGCCAAGTTGGGAGATGACCAC  
Cte7B8 GGAACGAAATCAAACCAAGCCGCTCAAATCACTCGTCTCAA AAGTTGGGAGATGACCAC  
Cte7B4 GGAACGAAATCAAACCAAGCCGCTCAAATCACTCGTCTCAA AAGTTGGGAGATGACCAC  
Cte7B11 GGAACGAAATCAAACGCTGCCGCTATCAACTCACTTGTCA A AAGTTGGGAAACAGACCAC  
Ctt7B7 GGAATCAAGCCAACTTTCCAGCCGCTTACGCACTCGTCTCAA AATTTGGGAGTTGATCAC  
Ctt7B6 GGAACGACTCAAACGCTGCCGAGTCAACTCACTCGTCTCAA AAGTTGGGAGATGACCAC  
Cte7B5 GGAACCACTCAAACGCTGCCGAGTCAACTCACTCGTCTCAA AAGTTGGGAGATGACCAC  
Ctt7B10 GGAACGAAATCAAACGCTTCTGCTGTCAA AACTCACTCGTCTCAA AAGTTGGGAGATGACCAC  
Ctt7B9 GGAACCACTCAAACGCTGCCGAGTCAACTCACTCGTCTCAA AAGTTGGGAGATGACCAC  
Cte7B8 GGAACGAAATCAAACGCTTCTGCTGTCAA AACTCACTCGTCTCAA AAGTTGGGAGATGACCAC  
Cte7B4 GGAACGCTCAAACGCTGCCGAGTGAAGGACTCTCAA AATTTGGGATCAGACCAC  
Ctp7B7 GGAATCAAGCCAACTTTCCAGCCGCTTACGCACTCGTCTCAA AATTTGGGAGTTGATCAC  
Ctp7B6 GGAACCACTCAAACGCTGCCGAGTCAACTCACTCGTCTCAA AAGTTGGGAGATGACCAC  
Ctp7B5 GGAACCACTCAAACGCTGCCGAGTCAACTCACTCGTCTCAA AAGTTGGGAGATGACCAC  
Ctp7B8/10 GGAACGAAATCAAATGCTTCCGCTGTCAA AACTCACTCGTCTCAA AATTTGGGAGATGACCAC  
Ctp7B4 GGAACGCTCAAACGCTGCCGCTTGAAGGACTCTCAA AATTTGGGATCAGACCAC  
CteN GGAAGGAAATCAAGTAAACCATGCCTTAGAACAATTAATAATGATATGGCGAAGTGCAC  
Gly16.4\_ORF2\_ GGA A A A A A A A T C T A G C C A A C C A T G C A T T A A G A C C A T C T T A T A T G C A T G G G T C A A A C A C A T  
Gly16.4\_ORF3 GGA A A G G A A T C A A C A C A C C A G C C A T C A A A C C A T C T T G A A C G A A A T G G G A C A A A A T C A C  
Gly5-4.5\_ORF2 GGA A A T C C T G A A A A C T T T C A G C T A C A A A A C C C T C A T C A A T G A A T T G G G A A C A G T C A T  
Gly5-4.5\_ORF3\_ G T A A T A T G C A A A C A C A C C A G C C A T C A A G A C C C T C T T A A T G A A T T A G G T C A A A G C C A T  
Gly5-4.5\_ORF1 GGA A A C G A A T C A A A T G C C C A G C T C T T A A G A C A C T T G T C A C C A A A T T G G G T G C T T C A C A C  
Gly5-2.2ORF1 GGA A C G A A T C A A A T A C C C A G C T C T T A A G A C A C T T G T C A C C A A A T T G G G T G C T T C A C A C

361 420

CteP AGAACCGTGGTGCACAAAGGAACAATTCGAAAAATTCAGGCATCAGTTGCTAAATAC  
CteQ AAAAGTCGTGGAATACAAAAAGAATTTTTCAACAAATTCGTCATCACTCGATGGATTT  
CteQa AAAAGTCGTGGAATACAAAAAGAATTTTTCAACAAATTCGTCATCACTCGATGGATTT  
Cte8 AAGGCTCGTGGATTTGAAAAAGCCAATTCGAAGAATTCGTCGCTCACTTGTGACTAC  
CmeORF6 AAAGCCCGTGGATTTGAGAAAGCTCAATTCGAAGAATTCGTCATCACTTGTGACTAC  
CteR AAAGCCCGTGGATTTGAGAAAGCTCAATTCGAAGAATTCGTCGCTCACTTGTGACTAC  
CmeORF5 AAAAATCGTGGGTTACAAAAGCTCAATTTAATCCCTCCGTCGCTCAATTTGTAATAT  
CteS AAAGTCGTGGAATACAAAAGCTCAATTTAATGATTCGTCGATCAATCGTCAATAC  
CmeORF7 AAGAAATCGTGGATTCAAAAGCTCAATTTAATGATTCGTCGATCAATCGTCAATAC  
Cte11 AGAGGTCGTGGATTCAAAAGCTCAATTCGAAGAATTCGTCGCTCACTTGTGACTAC  
Ctt11 AGAGGTCGTGGATTCAAAAGCTCAATTTAATGATTCGTCGCTCACTTGTGACTAC  
CmeORF2 AGGGCACGTGGAATTCAAAAGCTCTTTTAAACCAATTCGTCGCAACATTAATGACTTAT  
Cte12.1 AAAGAACGTGGATTTCAAAGGCCAATTCAGTCATTCGAGTTTCGATGGTAAATAC  
Cte12.2 AAAGAACGTGGATTTCAAAGGCCAATTCAGTCATTCGAGTTTCGATGGTAAATAC  
Ctt12 AAAGGACGTGGAGTTCAAAAGCTCAATTTAATGCTTTCGAGTCACTAATGCTTAT  
CtehorfA AAAGCCCGTGGATTCGCAACTGCTCAATTCGGTGAATTCAGAACCTCACTCGTCGATAC  
CteORF\_ps AGAGGACGTGGAATTCAAAAGCTCAATTTAACGAATTC-----  
CmeORFB AAAGGACGTGGAATTCAAAAGCAGATTCAGTCAATTCGAGTTTCGTTGATAAATAC  
Cte13 AAAGTCGTGGAATTCACAAGCTCAATTCGGTGAATTCGTCGATCACTGTCCTACAC  
Cte2f AAAGTCGTGGAATTCACAAGCTCAATTCACAAGATTCGTCGATCACTGTCCTACAC  
Ctt2f AAAGTCGTGGAATTCAAAAGCTCAGTTCAGTCAATTCGTCGCTCACTCGTTTCTAC  
Cme2f AAAGTCGTGGAATTCAAAAGGCAATTTAATGAGTTCGAGCTTCCCTAGTTTCTTAT  
Cte9 CACAACCGTGGATTCAAAAGGCAATTCAGTCATTCGAGTTTCGATGGTAAATAC  
Ctt9.1 CACAACCGTGGATTCAAAAGGCAATTTAAGCAATTCGTCGCTCACTGTCCTACAC  
Cme9 CACAACCGTGGATTCAAAAGGCAATTTAAGCAATTCGTCGCTCACTGTCCTACAC  
Cte7A1 AAAGCCCGTGGATTCACAAGCTCAATTCAGTCAATTCGTCGCTCACTGTCCTACAC  
Cte7A2 AAAGCCCGTGGATTCACAAGCTCAATTCAGTCAATTCGTCGCTCACTGTCCTACAC  
Cte7A3 AAAGCCCGTGGATTCACAAGCTCAATTCAGTCAATTCGTCGCTCACTGTCCTACAC  
Cte7A4 AAAGTCGTGGAATTCACAAGCTCAATTCACAAGATTCGTCGATCACTGTCCTACAC  
Cte7A5 AAAGCCCGTGGATTCACAAGCTCAATTCACAAGATTCGTCGATCACTGTCCTACAC  
Cte7A6 AAAGCCCGTGGATTCACAAGCTCAATTCACAAGATTCGTCGATCACTGTCCTACAC  
Cte7A7 AAAGCCCGTGGATTCACAAGCTCAATTCAGTCAATTCGTCGCTCACTGTCCTACAC  
Cte7A8 AAAGCCCGTGGATTCACAAGCTCAATTCACAAGATTCGTCGCTCACTGTCCTACAC  
Ctt7A.1 AAAGTCGTGGAATTCACAAGCTCAATTCAGTCAATTCGTCGCTCACTGTCCTACAC  
Cte1 AAAGCCCGT--GCCACAGCTCAATTCGATGCTTCCGTAATCACTCGTTGCTGG  
Ctt1 AAATCTCGT--GCATCACCAGCTCACTCGATTAATTCGTCGCTCACTGTCCTACAC  
Ctt1A AAATCTCGT--GCATCACCAGCTCACTCGATTAATTCGTCGCTCACTGTCCTACAC  
CteT AAAGTCGTGGAATTCACAAGCTCAATTCACAAGATTCGTCGATCACTGTCCTACAC  
CteU AAAGCCCGTGGAGTTCAAAAGGATCAATTCAGTCAATTCGTCGATCACTGTCCTACAC  
Ctt-6.3 AAAGCCCGTGGAGTTCAAAAGGATCAATTCAGTCAATTCGTCGATCACTGTCCTACAC  
CteY AAAGAACGTGGATTCAGTTCAGGAGTTCAGTCAATTCGTCGATCACTGTCCTACAC  
CtpY AAAGTCGTGGAATTCACAAGGATTAATTCGATGAATTCAGTCAAGCTTCCATAAATTC  
CteW AAAGTCGTGGAATTCACAAGGATTAATTCGATGAATTCAGTCAAGCTTCCATAAATTC  
CttW AAAGTCGTGGAATTCACAAGGATTAATTCGATGAATTCAGTCAAGCTTCCATAAATTC  
CteV AAAGAACCGTGGAGTTCACAAGGATTAATTCGATGAATTCAGTCAAGCTTCCATAAATTC  
CttV\_ps CAGAACCCTGGATTCACAAGGATTAATTCGATGAATTCAGTCAAGCTTCCATAAATTC  
CtpV AAAGAACCGTGGATTCACAAGGATTAATTCGATGAATTCAGTCAAGCTTCCATAAATTC  
CteZ AAAGCCCGTGGAGTTCACAAGGATTAATTCGATGAATTCAGTCAAGCTTCCATAAATTC  
CttZ AAAGAACCGTGGATTCACAAGGATTAATTCGATGAATTCAGTCAAGCTTCCATAAATTC  
CtpZ AAAGAACCCTGGATTCACAAGGATTAATTCGATGAATTCAGTCAAGCTTCCATAAATTC  
Cte7B12 CAAGTCGTGGAGTTTCACAAGCTCAATTCAGTCAATTCGTCGCTCACTGTCCTACAC  
Cte7B13 CAAGTCGTGGAGTTTCACAAGCTCAATTCAGTCAATTCGTCGCTCACTGTCCTACAC  
Cte7B7 AAGGCACGTGGATTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B6 AAAGCCCGTGGAGTTTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B6a AAAGCCCGTGGAGTTTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B5 AAAGCCCGTGGAGTTTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B5a AAAGCCCGTGGAGTTTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B10 AAAGCCCGTGGAGTTTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B10a AAAGCCCGTGGAGTTTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B9 AAAGCCCGTGGAGTTTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B9a AAAGCCCGTGGAGTTTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B8 AAAGCCCGTGGAGTTTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B4 AAAGCCCGTGGAGTTTCAGCCGCTCAATTCGAGAAATTCAGAAACCGCTTGTCTCATAAC  
Cte7B11 AACGTCGTGGTTCACAAGGCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctt7B7 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctt7B6 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctt7B5 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctt7B10 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctt7B9 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctt7B8 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctt7B4 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctp7B7 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctp7B6 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctp7B5 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctp7B8/10 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Ctp7B4 AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
CteN AAAGTCGTGGAGTTTCAGCTGCTCAATTCGTCGATTAATTCAGAAACCGCTTGTCTCATAAC  
Gly16.4\_ORF2\_ AGAAATCGTGGATTCACAAGGATTCCTTCAACAATTCGCAAACTCCGTCAGCTTCAAGATAT  
Gly16.4\_ORF3 AGAAATCGTGGATTCACAAGGATTCCTTCAACAATTCGCAAACTCCGTCAGCTTCAAGATAT  
Gly5-4.5\_ORF2 AGAAATCGAGGATTTGTAAGGAACAATTTAATGGATTCGCTACACTTCTGTCATAT  
Gly5-4.5\_ORF3\_ AGAAATCGTGGTGCACAAAGGAACAATTCGAGAAATTCAGTCACTAGTACGCAAAATAT  
Gly5-4.5\_ORF1 AAGAACCCTGGAGTTTCACAAGCTCAATTCACAAGATTCGTCGATCACTGTCCTACAC  
Gly5-2.2ORF1 AAGAACCCTGGAGTTTCACAAGCTCAATTCACAAGATTCGTCGATCACTGTCCTACAC

CteP CTTAAGGATCACTCTGGAAGCTGGTCTGATGCCACTGGTCCGCATGGAACAAGGCATTC  
 ATGCAAAGACAATCA---TCATGGAATGAAAAATACAGCTGCTGTTTGGAAATAAGGCATCA  
 CteQ ATGCAAAGACAATCA---TCATGGAATGAAAAATACAGCTGCTGTTTGGAAATAAGGCATCA  
 CteQa ATGCAAAGACAATCA---TCATGGAATGAAAAATACAGCTGCTGTTTGGAAATAAGGCATCA  
 Cte8 CTTTCAAAGCACCTT---GACTGGAATGACACCATGAAATCAACATGGGATCTTGCATG  
 CmeORF6 CTCTCTAAGAATCTC---GATTGGAACGACACCATGAAATCAACATGGGATCTTGCATG  
 CteR CTTCAAAATCATGTC---TCATGGGGTGATAATGTTTCATCAGCATGGACAAGAGCCCTT  
 CmeORF5 CTGCAAAAATTAATTTCT---TCATGGGGTGATAATGTTTCATCAGCATGGACAAGAGCCCTT  
 CteS CTTCAAAATCATGTC---TCATGGGGTGACAATGTTTCAGAAGCATGGACAAAAAGCTCTT  
 CmeORF7 CTTCAGCTAATGTG---TCATGGGGGACAATGTTTCAGTTGCTTGGACAAAAGCTATT  
 Cte11 CTTTCAAATCATGTA---TCATGGGGTGATAATGTAGCTAAGGCTTGGGAAAAAGGTTTT  
 Ctt11 CTATCACAAAATGTA---TCATGGGGTGATAACGTAGCTCAAGCTTGGGAAAAAGGCTTT  
 CmeORF2 CTTTTCAGAACATGTA---TCATGGGGTGAAAAATGTAGCTAAGGCTTGGGAAAAAGGATTT  
 Cte12.1 CTATCAGAACATGTC---TCATGGGGAGATAATGTAGCTCAGGCTTTTGA AAAAGGCTCTT  
 Cte12.2 CTCTCAGAACATGTC---TCATGGGAGATAATGTAGCTCAGGCTTTTGA AAAAGGCTTTT  
 Ctt12 CTTCAGATCATGTG---ACATGGAATGATAATGTAGCTCAAGCTTGGGAAAAAGGACTT  
 Ctehorfa CTCCAAGCTAATGTC---TCATGGGGTGACAATGTGCTGCTGCTGGAACCAAGCTCTT  
 CteORF\_ps -----TGATAACGTTGCTAAGGCTGGGAAAGAGGTTTGAATAAT  
 CmeORFB CTATCACAACATGTC---ACATGGGGAGATAATGTAGCTCAGGCTTGGGAAAAAGGCTCTT  
 Ctt13 CTCTCAAACCATGTC---TCATGGGGTGAAAAATGTGCTGCTTGGAACCTGCTCTT  
 Cte2S CTCCAAGGCCAGT---ACATGGAACGATAAATCGGAGCTGCCTGGACACAAGGCTCTT  
 Ctt2S CTCCAAGTAAAGT---TCATGGAACGATTTCTTGGAGCTGCCTGGACACAAGGCTCTT  
 Cme2S CTACAACAAAATGTA---GCTTTGGTGACAATGTGCTGCTGCTGGAACCAAGGCTCTT  
 Cte9 CTTTCAACCACTACT---ACATGGAATGATGCCACAGCCGCTGCCTGGACTCATGGCTCTT  
 Cte9.1 CTCTCAAGCCACGCT---TCATGGAATGATGCTACCGCTGATGCTGGACCAATGGACTT  
 Cme9 TTGCTGCCCACGCA---GCTTGGTGCAGCCACCGCTGATGCTGGACCAATGGACTT  
 Cte7A1 CTCCAAGCTAATGTT---GCATGGGGCACAATGTGCTGCTGGACACAAGGACTT  
 Cte7A2 CTCCAAGCTAATGTT---GCATGGAGTGACGCTACTGCTGCTGGACACAAGGACTT  
 Cte7A3 CTCCAAGCTAATGTT---GCATGGGGCACAATGTGCTGCTGGACACAAGGACTT  
 Cte7A4 CTCCAAGCTAATGTT---TCATGGGGCACAATGTGCTGCTGGACACAAGGACTT  
 Cte7A5 CTCCAAGCTAATGTT---GCATGGGGCACAATGTGCTGCTGGACACAAGGACTT  
 Cte7A6 CTCCAAGCTAATGTT---TCATGGGGCACAATGTGCTGCTGGACACAAGGACTT  
 Cte7A7 CTCCAAGCTAATGTT---GCATGGAGTGACGCTACTGCTGCTGGACACAAGGACTT  
 Cte7A8 CTTCAAGCTAATGTT---GCATGGAGTGACGCTACTGCTGCTGGACACAAGGACTT  
 Ctt7A.1 GTCTCAAGCAATGTT---GCATGGAACGCTGCTGCTGATCAGCATGGACTGCTGGCTCTT  
 Cte1 CTTAAGGGAGCTACC---GCATGGGACGCAACCTGTAATCAAGCTGGAACGCTGCTCCT  
 Ctt1 CTTAAGGGAGCACT-----AAATGGGACTCTGCTGCTGCAATCAAGCTGGGCTCCAGTT  
 Ctt1A CTTAAGGGAGCACT-----AAATGGGACTCTGCTGCTGCAATCAAGCTGGGCTCCAGTT  
 CteT CTCTCAAGCCAGT---ACATGGGGTGACAATGTGCTGCTGGAACCAAGCTCTT  
 CteU CTTAAGGCCATATC---ACCTTCGATGAAACACAGAACTGCATGGACTCTTGCCTCTT  
 Ctt-6.3 CTTAAAGCACATATC---AACTTCGATGAAACCACTGAACTGCATGGACTCTTGCCTCTT  
 CteY TTCCAACACACACA---AAATGGAATGAAAACCTGAGGATCATGATGCTGCAATGAA  
 CtpY TTGACACACACTCA---GTATGGAATGACAATGTGATGCTGCTGGCACTGCAATGAA  
 CteW CTCAAAAGCACATACA---ACATGGGATGAGAAAGCTGAAAATGCAATGGAAAAGTTGTGGT  
 CttW CTTAAATCACATACA---ACATGGGATGAGAAAGCTGAAAATGCAATGGAAAAGTTGTGGT  
 CtpW CTCTCAAGCCATGTC---TCATGGGGTGATAACGTTGCTGCTGGAAATCATGCCCTC  
 CteV CTCTCAAGCCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTC  
 CttV\_ps CTCTCAAGCCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTC  
 CtpV CTCTCAAGCCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTC  
 CteZ ATTTCAATCACATGTA---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTC  
 CttZ ATCTCATCAAACGTC---GCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTC  
 CtpZ ATCTCATCAAACGTC---GCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTC  
 Cte7B12 CTTCAAGCTAATGTT---TCATGGGGTGAAAAATGTGCAAGCTGCAATGGAATCAAGCTCTT  
 Cte7B13 CTTCAAGCTAATGTT---TCATGGGGTGAAAAATGTGCAAGCTGCAATGGAATCAAGCTCTT  
 Cte7B7 CTTTCAATCAAGCTC---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B6 CTCTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7Ba CTCTCAAACCATGTA---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B5 CTCTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B5a CTCTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B10 CTCTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B10a CTCTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B9 CTTTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B9a CTCTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B8 CTCTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B4 CTCTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B11 CTCTCAATCAAGCTC---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Cte7B7 CTACAAGCTCATGTC---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Ctt7B6 CTCCAAGCTAATGTT---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Ctt7B5 CTCCAAGCTAATGTT---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Ctt7B10 CTCCAAGCTAATGTT---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Ctt7B9 CTCCAAGCTAATGTT---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Ctt7B8 CTCTCAAACCATGTC---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Ctt7B4 CTCTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 Ctp7B7 CTTCAAGCTCATGTT---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Ctp7B6 CTCCAAGCTAATGTT---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Ctp7B5 CTTCAAGCTAATGTT---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Ctp7B5/10 CTCCAAGCTAATGTT---TCATGGGGTGACAATGTGCTGCTGGAAATCATGCCCTT  
 Ctp7B4 CTCTCAAACCATGTC---TCATGGGGTGATAATGTGCTGCTGGAAATCATGCCCTT  
 CteN TTAAGGCTCATGTT---GAAATGGACGAAAAGGCTTACACGCTTGGAAATGATGCTCTT  
 Gly16\_4\_ORF2\_ ATGAGGCGCCATTTCT---GATTTCACTGCGACTATGAGCAGCCTGGAATGATGCTCTT  
 Gly16\_4\_ORF3 CTTAGSACACATGGC-----GAGGGATACAATGCTCAATACTGGGACGATASATCT  
 Gly5-4.5\_ORF2 CTGAAACACACACT---TCTTGGAAATGATGATGCTGCAAAATTTCTGGAACAAGGCGCTC  
 Gly5-4.5\_ORF3 CTCAAAGAACATGCTGGAGCTTGGACTGATGCCACAGCAGCCGATGGAACAAGGCATTTT  
 Gly5-4.5\_ORF1 CTTCAACAAAATGTT---TCATGGAATGACGACACCGCTGCTGCTGGACACAAGGCTCTT  
 Gly5-2.2ORF1 CTTCAACAAAATGTT---TCATGGAATGACGACACCGCTGCTGCTGGACACAAGGCTCTT

	481	537
CteP	GATGAAATGTATGCCATCGTCTTCAGCAATCTCGATGGAAATCCAGTTCACATAA---	
CteQ	AATAATTTTTATTTTGTCTTTTTCAGCTCTCGATGGAAATCCAGTGAACATAA---	
CteQa	AATAATTTTTATTTTGTCTTTTTCAGCTCTCGATGGAAATCCAGTGAACATAA---	
Cte8	AACAACATGTCTTCTACATCTACATGCTTTGGAATCTGCTCATATAA-----	
CmeORF6	AACAACATGTCTTCTACATCTACATGCTTTGGAATCTGCTCATATAA-----	
CteR	GACAATGCATACTTTATCTCTTTTCGGCCCTTGATGGACATCCATGTATAA---	
CmeORF5	GACAATGCCTTCTTCTCTTTTCTCAGCTCTTGATGGAAATCCATGTATAA---	
CteS	GACGATGCTTACTTTTATCTCTTCTCAGCCCTTGATGGAAATCCATGTATAA---	
CmeORF7	GGCGATGCCATTTATTTATCTCTTCTCAGCTCTTGATGGAAATCCATGTATAA---	
Cte11	AATAATGTTTATTCACATCTTTTCTTAGCCTTGATGGAAATCCATGTATAA---	
Ctt11	AATAATGTTTATTTCAATCTTTTCAATGCTCTTGACGGCACTCCTATATAA---	
CmeORF2	AATAATGTTTTTTTCTCTTCTCAGCTCTTGATGGAAATCCATGTATAA---	
Cte12.1	GATAATGTTTATTTAGTCTTTTCTCAGCTTTGATGGAAATCCGATATAA---	
Cte12.2	GATAATGTTTATTTGTTCTTTTCTCAGCTTTGATGGAAATCCGATATAA---	
Ctt12	GATAATGTTTATTTGTTCTTTTCTCAGCTTTGATGGAAATCCATGTATAA---	
Ctehorfa	GACAACCTTATTTCTGTTCTTACCTCAAACCTATAA-----	
CteORFB_ps	ATGTATCACATTTATTTCTTAGCCTTGATGGAAATCCGCTGTAA-----	
CmeORFB	GAAAATATTAATACGGTCTTATTTCTCAGCACTTGATGGAAATCCGATG-----	
Ctt13	GACAACCTTCTTCTGTTCTTACCTCAAACCTATAA-----	
Cte2E	GACAACGGCTTTCACAATGATCTTCTGCTAAATTTGTAA-----	
Ctt2E	GACAACGCTTTCACAATGATCTTCTCATATTTGTAA-----	
Cme2E	GACAATACCTTTCACAATGATCTTCTGCTAAATTTGTAA-----	
Cte9	GACAACATCTTTCAGCCTATCTTCTGCCCTTGTAA-----	
Ctt9.1	GACAACATCTTTCAGCCTATCTTCTGCCCTTGTAA-----	
Cme9	GACAACATCTTTCAGCCTATCTTCTGCCCTTGTAA-----	
Cte7A1	GATAACATCTACGGACTCTTCTTCTGCCCTTGTAA-----	
Cte7A2	GATAACATCTACGGACTCTTCTTCTGCCCTTGTAA-----	
Cte7A3	GACAACATCTACGGACTCTTCTTCTGCCCTTGTAA-----	
Cte7A4	GACAACATCTACGGACTCTTCTTCTGCCCTTGTAA-----	
Cte7A5	GACAACATCTACGGACTCTTCTTCTGCCCTTGTAA-----	
Cte7A6	GACAACATCTACGGACTCTTCTTCTGCCCTTGTAA-----	
Cte7A7	GACAACATCTACGGACTCTTCTTCTGCCCTTGTAA-----	
Cte7A8	GACAACATCTACGGACTCTTCTTCTGCCCTTGTAA-----	
Ctt7A.1	GACAACATCTACGGACTCTTCTTCTGCCCTTGTAA-----	
Cte1	GACTTTGTCTTTCGGAGTATTGAAAGAACAACTTGCAGCAGCAAAATAA---	
Ctt1	CTTGACTTTGTCTTCTCAACATTTGAAGATGAACCTTAA-----	
Ctt1A	CTTGACTTTGTCTTCTCAACATTTGAAGATGAACCTTAA-----	
CteT	GACAACACAATGGCCATCGCTTTCAAACAACCTTGATGATATAA-----	
CteU	GAAAACAATACAGATGCTCTTCTGCCATGGACTCATATAA-----	
Ctt-6.3	GACAACAATACAGATGCTTTTCTTCTGCCATGGACTCATATAA-----	
CteY	AAAGAAGCTCGTAAACTATTAACGAGTCTTGGAGGACATCCAATTTGAATAA---	
CtpY	AAGGAATTCGTAAGTAACTAACGCTAACCTTTGAATAA-----	
CteW	GATGAACATCATGCTATTGTTCTACTCAATTTTGAATAA-----	
CttW	GATGAACATCATGCTATTGTTCTACTCAATTTTGAATAA-----	
CtpW	GATGAACATCATGCTATTGTTCTACTCAATTTTGAATAA-----	
CteV	GAAAATACATACACAGTTGCCCTTGAAGTCAATCCAGCCTAA-----	
CttV_ps	GAAAACACATATACAGTTGCCCTTGAAGTCAATCCAGCCTAA-----	
CtpV	GAAAACACATATACAGTTGCCCTTGAAGTCAATCCAGCCTAA-----	
CteZ	GAAAATGTATACACCGCTGCCCTTCAAATCATCACAGATAA-----	
CttZ	GACAAATGTATACACCGCTGCCCTTCAAATCATCACAGATAA-----	
CtpZ	GACAATGTATACACCGCTGCCCTTCAAATCATCACAGATAA-----	
Cte7B12	GACAATACATATGCTATTGCCCTTCAAACCTTTACTGCCTAA-----	
Cte7B13	GACAATACATATGCTATTGCCCTTCAAACCTTTACTGCCTAA-----	
Cte7B7	GATAACACTTATGCCCTTGTAAATTAAGCTTTGCTTAA-----	
Cte7B6	GACAACACCTACGCCATCGTCTTGGCCAACTTTTAA-----	
Cte7B6a	GACAACACCTACGCCATCGTCTTGGCCAACTTTTAA-----	
Cte7B5	GACAACACCTACGCCATCGTCTTGGCCAACTTTTAA-----	
Cte7B5a	GACAACACCTACGCCATCGTCTTGGCCAACTTTTAA-----	
Cte7B10	GACAACACATTCGCCATCGCTTGGCCAACTTTTAA-----	
Cte7B10a	GACAACACATTCGCCATCGCTTGGCCAACTTTTAA-----	
Cte7B9	GACAACACATTCGCCATCGCTTGGCCAACTTTTAA-----	
Cte7B9a	GACAACACATTCGCCATCGCTTGGCCAACTTTTAA-----	
Cte7B8	GATAACACAATGGACATCGCTTGGCCAACTTTTAA-----	
Cte7B4	GATAACACAATGGACATCGCTTGGCCAACTTTTAA-----	
Cte7B11	GACAACACAATGGATGTTGCCCTTCAAGGTTATGGAACACTAA-----	
Ctt7B7	GATAACACTTATGCAGTTGCACCTCAAGTCTCTTGAATAA-----	
Ctt7B6	GACAACACATTCGCCATCGTCTTGGCCAACTTTTAA-----	
Ctt7B5	GACAACACCTTTGCCATCGTCTTGGCCAACTTTTAA-----	
Ctt7B10	GACAACACATTCGCCATCGTCTTGGCCAACTTTTAA-----	
Ctt7B9	GACAACACCTTTGCCATCGTCTTGGCCAACTTTTAA-----	
Ctt7B8	GATAACACCTTATGCCATCGTCTTGGCCAACTTTTAA-----	
Ctt7B4	GATAACACAATGGCTGCTGCTTGGCTCACTTTTAA-----	
Ctp7B7	GACAACACTTATGCCCTTGCACCTCAAGTCTCTTGAATAA-----	
Ctp7B6	GACAACACCTTGCCTATCGTCTTGGCCAACTTTTAA-----	
Ctp7B5	GACAACACCTTGCCTATCGTCTTGGCCAACTTTTAA-----	
Ctp7B8/10	GACAACACCTTGCCTATCGTCTTGGCCAACTTTTAA-----	
Ctp7B4	GATAACACCAATGGCTGCTGCTTGGCTCACTTTTAA-----	
CteN	GATAGTATGTATACATCATTTTTCGCAATTTGATGGAAAACAGTTAGCTAA---	
Gly16.4_ORF2	GATAAGATGTACTTTGTATCTTTCAGCACTCTTGATGGAAAATCTTTAATATAA---	
Gly16.4_ORF3	GACAATAATGTACTATGTATCTTTCAGTGTCTTGGATGGAGGACAGT-----	
Gly5-4.5_ORF2	GACAATTTCTACGCAATCTTTTTCAGGCTCTCAATGGAAACCCATGACATAA---	
Gly5-4.5_ORF3	GAAGAATGTATCAAAATCATCTTTCAGCACTCTGACGGAAACCCAGTGCACAAATAA---	
Gly5-4.5_ORF1	GACAATCTCTACTTTGTCTCTTCTCAGCTCTCGATGGAAACCCATCAACTAA---	
Gly5-2.2ORF1	GACAATCTCTACTTTGTCTCTTCTCAGCTCTCGATGGAAACCCATCAACTAA---	

Alignment der Aminos ure-Sequenzen der H moglobine von *C. tentans*,  
*C. thummi*, *C. piger*, *C. melanotus* und *Glyptotendipes barbipes*

```

1                               60
CTEP MKFL-ILAAVLVAGAYA--TLTADAEK---LVKGSWDKVKQEDGILYVIFKENPDIQA
CTEQ MKFLIIIALCIMSSTCDLTPLNSEAE---TLRTLWNRVKHREADILYVIFKENPDIQA
CTEQa MKFLIIIALCIMSSTCDLTPLNSEAE---TLRTLWNRVKHREADILYVIFKENPDIQA
CTEVIII MKFL-VLALCIAAASAAVTPMTADQLA---LFRSSWAEVKHNEVDILYAVFKANPDIQA
CTVII MKFL-VLALCIAAASAAVTPMTADQLA---LFRSSWAEVKHNEVDILYAVFKANPDIQA
CMEORF6 MKFL-VLALCIAAASAAVTPMTADQLA---LFRSSWAEVKHNEVDILYAVFKANPDIQA
CMEORF5 MKLL-ILVLCIAAANCDLVLLTADQGA---LVHSSWQVKHNEVDILYAVFKANPDIQA
CTES MEFL-IFALCIVTVSCVPAPLTADQAK---IVKSSWAEVKHNEVDILYAVFKANPDIQA
CMEORF7 MKFL-VLALCIVAASCDPAPLSADQVT---LVRSSWQVKHNEVDILYAVFKANPDIQA
CTEXI MKFL-IFILCIAAASCDLPLVDDDEAT---LIRSSWDQVKHNEVDILYAVFKANPDIQA
CTXI MKFL-ILALCIAAASCDLPLVDDQAI---LIRSSWAEVKHNEVDILYAVFKANPDIQA
CMEORF2 MKFL-IFTLCIAAASCDLPLVDDDEET---LIRSSWQVKHNEVDILYAVFKANPDIQA
CTEII1 MKFL-IFILCIAVTSCLDAQFVEDQTA---LIRSSWQVKHNEVDILYAVFKANPDIQA
CTEII2 MKFL-IFILCIAVTSCLDAQFVEDQTA---LIRSSWQVKHNEVDILYAVFKANPDIQA
CTEII MKFL-IFTLCIAAATCDLAQFVEDQTE---LIRSSWQVKHNEVDILYAVFKANPDIQA
CTEORFA MKFFVVLALCIAAASAAVPLSADQAS---LVKSSWQVKHNEVDILYAVFKANPDIQA
CTEORFB ps LKS--VLTLCIVAVSCDIVPLEDDQAI---LIRSSWQVKHNEVDILYAVFKANPDIQA
CMEORFB MKFL-IFALCIVAVCCELISTVDEQSAADHLLIRASWQVKHNEVDILYAVFKANPDIQA
CTVII MKFL-ILALCIAAASAAVPLSADQAS---LVKSSWQVKHNEVDILYAVFKANPDIQA
CTEII2 MKFL-VLALCIVGAI--PLSADQAS---LVRASWQAKHSEVDILYVIFKANPDIQA
CTEII2 MKFL-VLALCIAAAV--PLSADQAS---LVRSSWQVKHNEVDILYVIFKANPDIQA
CMEII2 MKFL-VLALCIAAIA--PLSADQAS---LVRSSWQVKHNEVDILYVIFKANPDIQA
CTEIX MKFLAVLALCIVGAIAD--PLSADQAD---LIRSSWQVKHNEVDILYAVFKANPDIQA
CTEIX MKFFIVLALCIVGAIAD--PVSSDQAN---AIRASWAGVKHNEVDILYAVFKANPDIQA
CTEIX.1 MKFFIVLALCIVGAIAD--PVSSDQAN---AIRASWAGVKHNEVDILYAVFKANPDIQA
CMEIX -----D--PVSSDQAG---LIRSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIA1 MKFFAVLALCIVGAIAT--PLTADQAS---LVQSTWQVRNNEVDILYAVFKANPDIQA
CTEVIIA2 MKFFAVLALCIVGAIAT--PLSADQAN---LVQSTWQVRNNEVDILYAVFKANPDIQA
CTEVIIA3 MKFFAVLALCIVGAIAT--PLTADQAS---LVQSTWQVRNNEVDILYAVFKANPDIQA
CTEVIIA4 MKFFAVLALCIVGAIAS--PLTADQAN---LVKSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIA5 MKFFAVLALCIVGAIAT--PLTADQAS---LVQSTWQVRNNEVDILYAVFKANPDIQA
CTEVIIA6 MKFFAVLALCIVGAIAT--PLTADQAN---LVKSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIA7 MKFFAVLALCIVGAIAT--PLTADQAS---LVQSTWQVRNNEVDILYAVFKANPDIQA
CTEVIIA8 MKFFAVLALCIVGAIAT--PLSADQAS---LVQSTWQVRNNEVDILYAVFKANPDIQA
CTEVIIA -----A--PLSADQAS---LVKSTWQVRNNEVDILYAVFKANPDIQA
CTEVIIA.1 MKFFAVLALCIVGAIAS--PLSADQAA---LVKSTWQVRNNEVDILYAVFKANPDIQA
CTEI MKFL-ILALCVAYTMAA--P--SADQIA---AAKASWATVKGQVVDILYAVFKANPDIQA
CTEI MKFL-ILALCVAAAMAG--P--SGDQIA---AAKASWNTVKNQVVDILYAVFKANPDIQA
CTEII MKFL-ILALCVAAAMAG--P--SGDQIA---AAKASWNTVKNQVVDILYAVFKANPDIQA
CTET MKFFAILALPCVVGAIAG--PLSADQAA---LVKSSWDQVKHNEVDILYAVFKANPDIQA
CTEU -----
CTVII.3 MKFL-VLALCIAAASAA--VLTTEQAD---LVKKTWSTVKFNEVDILYAVFKANPDIQA
CTTY MKFLVVLALCIVGAIAS---CDNAG---SLKEAWEHIKHDEVEIILYVIFKANPDIQA
CTPY MKVLAIFALCIIIGALAT--P--CDDFK---IMQEAWNTMKNEVEIILYVIFKANPDIQA
CTEW MKFIVLTLCLVAGAIAS---CDKAP---FIKASWQVKHNEVDILYVIFKANPDIQA
CTTW MKFLVLTLCIAGAIAS---CDKAP---FIKASWQVKHNEVDILYVIFKANPDIQA
CTPW MKFLVLTLCIAGAIAS---CDKAP---FIKASWQVKHNEVDILYVIFKANPDIQA
CTEV MKFFAVLALCIVGAIAS--PLTADQAS---LVKSSWQVKHNEVDILYAVFKANPDIQA
CTTV ps MKFFAVLALCIVGAIAS--PLTADQAS---LVKSSWQVKHNEVDILYVIFKANPDIQA
CTPV MKFFAVLALCIIIGAIAS--PLTADQAS---LVKSSWQVKHNEVDILYVIFKANPDIQA
CTEZ MKFFAVLALCIVGAIAS--PLTADQAT---LVKSSWQVKHNEVDILYVIFKANPDIQA
CTTZ MKFFAVLALCIVGAIAS--PLTADQAS---LVKSSWQVKHNEVDILYVIFKANPDIQA
CTPZ MKFFAVLALCIVGAIAS--PLTADQAS---LVKSSWQVKHNEVDILYVIFKANPDIQA
CTEVIIB2 MKFFAVLALCIVGAIAT--PLTADQAN---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB3 MKFFAVLALCIVGAIAT--PLTADQAN---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB7 MKFFAVLALCIVGAIAS--PLSADQAA---LVKSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB6 MKFFAVLALCIVGAIAS--PLTADQAN---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB6a MKFFAVLALCIVGAIAS--PLTADQAN---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB5 MKFFAVLALCIVGAIAS--PLTADQAN---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB5a MKFFAVLALCIVGAIAS--PLTADQAN---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB10 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB10a MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB9 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB9a MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB8 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB4 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEVIIB11 MKFFAVLALCIVGAIAS--PLTADQAS---LVKSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB7 MKFFAVLALCIVGAIAS--PLSADQAA---LVKSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB6 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB5 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB10 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB9 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB8 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB4 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB9/5 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB7 MKFFAVLALCIVGAIAS--PLSADQAA---LVKSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB6 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB5 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB10 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTVIIB4 MKFFAVLALCIVGAIAS--PLTADQAS---LVQSSWQVKHNEVDILYAVFKANPDIQA
CTEN MKFI-VIALCIACVTANPHWDNLDEVEIA---SIRKRTWQVKHNEVDILYVIFKANPDIQA
GLYP16_4_ORF2 MKFL-VLALCVAAVADPOVYVTLDEAVG---FVKSSWQVKHNEVDILYVIFKANPDIQA
GLYP16_4_ORF3 MKFL-ILALCAAVVMADPHWLMNNEAEVA---NIKHTYVSKHNEVDILYVIFKANPDIQA
GLYP5-4_5_ORF2 MKLPLITCLCIVSVCITDTPISSEEA---VVKSSWQVKHNEVDILYVIFKANPDIQA
GLYP5-4_5_ORF3 MKFL-ILAAVLVAGAYA--TLTADAEK---LVKSSWQVKHNEVDILYVIFKANPDIQA
GLYP5-4_5_ORF1 MKFFAVLALCIVGAIAS--PLSADQAN---LVRSSWQVKHNEVDILYVIFKANPDIQA
GLYP5-1.2_ORF1 MKFFAVLALCIVGAIAS--PLSADQAN---LVRSSWQVKHNEVDILYVIFKANPDIQA

```

CTEP KFPFSVGNLEELIKSNDD---FTKHADRIVAAVTKYVDLVGNEANTPAIKTLLNELQGSH  
CTEQ HFPFAVGNLEALRGSIA---FAIHATRIVSIFSKIASLAGDSNNI PASKTLINELAASH  
CTEQa HFPFAVGNLEALRGSIA---FAIHATRIVSIFSKIASLAGDSNNI PASKTLINELAASH  
CTVIIII RFPQFAGKDLDSLKDTAA---FATHAGRIVGFFSDVIGLIGNENRPAKTLIDGLASNH  
CTVIIII KFPQFAGKDLDSLKDSAD---FAVHSGRIVGFFSEVIGLIGNENRPAKTLIDGLASSH  
CMEORF6 KFPQFAGKDLDSLKDTAA---FATHAGRIVGFFSEVIGLIGNENRPAKTLIDALASSH  
CTER KFSQFSGKDLDSLKDTAA---FATHAGRIVGFFSEIIALADSESNVAASKFLINGLAENH  
CMEORF5 KFPQFAGKDLDSLKSTAS---FAIHAGRIVGFFSEIIALADSESNVAASKFLINDMATRH  
CTES KFPQFARKNLASLKSTAA---FATHAGRIVGFFSEIILLIEGESNVAASKFIINDIAASH  
CMEORF7 RFPQFVGKDLASLKTTAS---FATHAGRIVGFFSEIITLIEGESNI PASKYLINDLAATH  
CTEXI RFPQFAGKDVESLKDTAA---FATHAGRIVGFFSEIASLSTDESSISAAKTLINQLGASH  
CTTXI RFPQFAGKDLSEIKTTGQ---FAVHAGRIVGFFSEIASLQPDSEGI SAAKTLINLQASH  
CMEORF2 RFPQFAGKDLSELKDTAA---FATHAGRIVGFFSEIASLSTDASSVSAKTI INELGANH  
CTEXII1 RFPQFAGKDLSELKNSAA---FATHAARIVGFFSQLEIEFPNDSNVSAKTLINQLAVSH  
CTEXII2 RFPQFAGKDLKTLKSSSS---FASHAGRIVGFFSKITELNPNDSGVSAKTLINELAASH  
CTTXII RFPQFAGKDLVAGLKDTAA---FATHAGRIVGFFSEIIGLIGNAANA PALQTLVQGLAASH  
CTEORFB\_ps HFPQFAGKDVESLKDTAA---FAIHAGRIVGFFSEMAGLSNDESGVPAAMTL-NQLGASH  
CMEORFB RFPQFAGKDLRLSKNTVA---FTTHAVRIVGFFSEITELPPNDSGVSAKTLINQLAVSH  
CTTXIII RFPQFAGKDLISIKDTGATGAFATHAGRIVGFFSEIVGLIGNQSNVPAVQTLVQGLAASH  
CTEII6 KFPQFVGKDLDSLKDTAA---FATHAGRIVGFFMSQI IALMGNEANRPAMVTLINEMAANH  
CTII6 KFPQFAGKDLLETLKGTGQ---FATHAGRIVGFFSEIIVALMGNSANMPAMETLIDKMAANH  
CMEII6 KFPQFAGKDLDSLKNTAS---FATHAGRIVGFFSEIIVALMGNSANMPAMETLITEMAGNH  
CTEIX RFPQFAGKDLDSLKDTAA---FATHAGRIVGFFSEIIVALVGNENRPAAMNTLINELATNH  
CTTX RFPQFAGKDLASIKDTGA---FATHAGRIVGFFSEIIVALVGNESNA PAMATLINELSTSH  
CTTX.1 RFPQFAGKDLASIKDTGA---FATHAGRIVGFFSEIIVALVGNESNA PAMATLINELSTSH  
CMEIX RFPQFAGKDLAALKDTAA---FATHAGRIVGFFSEIIVALVGNESNA PAMATLINELATNH  
CTVIIIA1 RFPQFAGKDLAALKDTAA---FATHAGRIVGFFSEIIVALVGNESNA PAVQTLVQGLAASH  
CTVIIIA2 RFPQFAGKDLAALKDTAA---FATHAGRIVGFFSEIIVALIGNQSNAPAVQTLVQGLAASH  
CTVIIIA3 RFPQFAGKDLAALKDTAA---FATHAGRIVGFFSEIIVALVGNESNA PAVQTLVQGLAASH  
CTVIIIA4 RFPQFAGKDLAALKDTAA---FATHAGRIVGFFSEIIVALVGNESNAPATQTLVQGLAASH  
CTVIIIA5 RFPQFAGKDLAALKDTAA---FATHAGRIVGFFSEIIVALIGNQSNAPAVQTLVQGLAASH  
CTVIIIA6 RFPQFAGKDLAALKDTAA---FATHAGRIVGFFSEIIVALVGNESNAPATQTLVQGLAASH  
CTVIIIA7 RFPQFAGKDLAALKDTAA---FATHAGRIVGFFSEIIVALIGNQSNAPAVQTLVQGLAASH  
CTVIIIA8 RFPQFAGKDLAALKDTAA---FATHAGRIVGFFSEIIVALIGNQSNAPAVQTLVQGLAASH  
CTVIIIA RFPQFAGKDLASIKDTGA---FATHAGRIVGFFSEIIVALIGNESNAPAVQTLVQGLAASH  
CTVIIIA.1 RFPQFAGKDLASIKDTGA---FATHAGRIVGFFSEIIVALIGNESNAPAVQTLVQGLAASH  
CTEI KFSQFAGKDLVDSIKGTPE---FATHAGRIVGFFGDVMDLLGNDANTATIVSKAKDLKTH  
CTTI AFSQFAGKDLDSIKGTPD---FSKHAGRVVGLFSEVMDLLGNDANTPTILAKAKDFGKSH  
CTTIA AFSQFAGKDLDSIKGTPD---FSKHAGRVVGLFSEVMDLLGNDANTPTILAKAKDFGKSH  
CTET KFPQFVGKDLGLKDTAA---FATHATRIVSPLTEVVSLSGSENNLAHVTLVTKLGDH  
CTEU -----DLESVKDSPA---FATHATRIVSPLSQVSLAGSDDTI PAIQNLAKDLATSH  
CTVII.3 KFPQFAGKDLDSIKDSAA---FATHATRIVSPLSEVSLAGSDANI PAIQNLAKELATSH  
CTTY RFPQFAGKDLSEIKGTAD---FAIHATRIVGFFSEIIEGNDANTPAIKTLLNQLAKSH  
CTPY KFPQFVGKDLLETKGTAE---FAVHATRIVSFFSEIISLGNPDNLPAIMSLSLKLGKDH  
CTEW KFPQFAGKDLLETKGTAV---FAIHATRIVSFFSEIISLGNPDNLPAIMSLSLKLGKDH  
CTTW RFPQFAGKDLLETKGTAE---FAVHSTRIVSFFSEIISLGNPDNLPAIMSLSLKLGKDH  
CTPW RFPQFAGKDLLETKGTAE---FAVHSTRIVSFFSEIISLGNPDNLPAIMSLSLKLGKDH  
CTEV KFPQFAGKDLDSVKETA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTTV\_ps RFPQFAGKDLDTIKTSGQ---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTPV KFPQFAGKDLDSIKTSGQ---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTEZ KFPQFAGKDLSEVKTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTTZ RFPQFAGKDLDTIKTSGQ---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTPZ RFPQFAGKDLDSIKTSGQ---FATHATRIVSFFSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB12 KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB13 KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB7 KFPQFAGKDLAALKDTAA---FATHATRIVSFFSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB6 KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB6a KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB5 KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB5a KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB10 KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB10a KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB9 KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB9a KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB8 KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB4 KFPQFAGKDLAALKDTAA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB11 KFPQFAGKDLDSVKETA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB7 KFPQFAGKDLASIKDTAA---FATHATRIVSFFSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB6 KFSQFAGKDLASIKDTGA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB5 KFSQFAGKDLASIKDTGA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB10 KFSQFAGKDLASIKDTGA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB9 KFSQFAGKDLASIKDTGA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB8 KFPQFAGKDLASIKDTGA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB4 KFPQFAGKDLASIKDTGA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB7/5 KFPQFAGKDLASIKDTGA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB5 KFPQFAGKDLASIKDTGA---FATHATRIVSFFSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB5 KFSQFAGKDLASIKDTGA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB8/10 KFPQFAGKDLASIKDTGA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTVIIIB4 KFPQFAGKDLASIKDTGA---FATHATRIVSPLSEIISLGNPDNLPAIMSLSLKLGKDH  
CTEN KFPQFAGKDLLEELKGTNQ---FAIHATRIVSFFSEIISLGNPDNLPAIMSLSLKLGKDH  
GLYP16.4\_ORF2 KFPQFAGKDLLEELKGTNQ---FAIHATRIVSFFSEIISLGNPDNLPAIMSLSLKLGKDH  
GLYP16.4\_ORF3 RFPKFGKDIETLKGTA---FALHAGRIIGLLTEYITLLGKESQPAIKTLINEMGQNH  
GLYP5-4.5\_ORF2 RFPFAVGNLEELIKSNDD---FTKHADRIVAAVTKYVDLVGNDANTPAIKTLLNELQGSH  
GLYP5-4.5\_ORF3 RFPQFAGKDLNGLKDTAA---FATHAGRIVGFFSEIIGLIGNESNAPAKTLLVQGLGASH  
GLYP5-4.5\_ORF1 RFPQFAGKDLNGLKDTAA---FATHAGRIVGFFSEIIGLIGNESNAPAKTLLVQGLGASH  
GLYP5-1.2\_ORF1 RFPQFAGKDLNGLKDTAA---FATHAGRIVGFFSEIIGLIGNESNAPAKTLLVQGLGASH



CTEP RNRGATKEQFEKFKASVAKYLDKDHSGSWSDATGAAWNKAFFDEMYAIVFSNLDGNPVH--  
CTEQ KSRGIQKEFFNKFRTSLDGFMRQS--SWNENTAAVWNKASNNFYVFLFSALDGNPVN--  
CTEQa KSRGIQKEFFNKFRTSLDGFMRQS--SWNENTAAVWNKASNNFYVFLFSALDGNPVN--  
CTEVIII KARGIEKAQFEFRASLVLYLQSHL--DWNDTMKSTWDLALNNMFFYIHLHALESAH---  
CTTVIII KARGIEKAQFEFRASLVLYLQSHL--DWNDTMKSTWDLALNNMFFYIHLHALESAH---  
CMEORF6 KARGIEKAQFEFRASLVLYLQSHL--DWNDTMKSTWDLALNNMFFYIHLHALESAH---  
CTER KARGITKAQFNEFRASLVLYLQSHV--SWGDNVSAWTRALDNYAIFLFSALDGNPL--  
CMEORF5 KNRGVTKAQFNSFRASLVLYLQSHV--SWGDNVSAWTRALDNYAIFLFSALDGNPL--  
CTES KARGITKAQFNDPRDSLVLYLQTHV--SWGDNVSAWTRALDNYAIFLFSALDGNPL--  
CMEORF7 KKCGITKANFNFRASLVLYLQSHV--SWGDNVSAWTRALDNYAIFLFSALDGNPL--  
CTEXI RGRGISKAQFNEFRATFITYLSNHV--SWGDNVAKAWKGFNNVYHILFLALDGNPMQ--  
CTTXI RGRGISKAQFNEFRASLVLYLQSHV--SWGDNVAKAWKGFNNVYHILFLALDGNPMQ--  
CMEORF2 RARGISKALFNQFRATLMTYLSHV--SWGENVAKAWKGFNNVYHILFLALDGNPIL--  
CTEXII1 KERGISKAQFNAFRVSMVKYLSHV--SWGDNVAQAFKGLDNVYLVLSAFDGNPI--  
CTEXII2 KERGISKAQFNAFRVSMVKYLSHV--SWEDNVAQAFKGLDNVYLVLSAFDGNPI--  
CTXYII KRGVSKAQFNAFRVSLTAYLADHV--TWNDNVAQAWKGLDNVYLVLSAFDGNPM--  
CTEORFA KARGIPTAQFGEFRASLVLYLQSHV--SWGDNVAAAWNQLDNLFFVLTSTNY-----  
CTEORFB\_ps RGRGISKAQFNEFR-----DNVAKAWERGLNNMFFYIHLFLALDGNLSL--  
CMEORFB KGRGISKAQFNAFRVSMVKYLSHV--TWGDNVAKAWKGFNNVYHILFLALDGNPM--  
CTXYIII KARGISTAQFGEFRASLVLYLQSHV--SWGENVAAAWNVALDNFFVLTSTNY-----  
CTEII8 KARKIPQAQFNEFRASLVLYLQSHV--TWNDKLGAAWTQGLDNVYHILFLALDGNPIL--  
CTTII8 KARGIPKAQFNEFRASLVLYLQSHV--TWNDKLGAAWTQGLDNVYHILFLALDGNPIL--  
CMEII8 KARGITKQGFNEFRASLVLYLQSHV--AFGDNVAAAWNQLDNLFFVLTSTNY-----  
CTEXI HNRGISKAQFNEFRASLVLYLQSHV--TWNDATAAAWTHGLDNVYHILFLALDGNPIL--  
CTTX HNRGITKQGFNEFRASLVLYLQSHV--SWNDATAAAWTHGLDNVYHILFLALDGNPIL--  
CTTX.1 HNRGITKQGFNEFRASLVLYLQSHV--SWNDATAAAWTHGLDNVYHILFLALDGNPIL--  
CMEXI HNRGITKQGFNEFRASLVLYLQSHV--AWSDATADAWTHGLDNVYHILFLALDGNPIL--  
CTEVIA1 KARGISQAQFNEFRASLVLYLQSHV--AWGDNVAAAWTQGLDNVYHILFLALDGNPIL--  
CTEVIA2 KARGISQAQFNEFRASLVLYLQSHV--AWSDATGAAWTQGLDNVYHILFLALDGNPIL--  
CTEVIA3 KARGISQAQFNEFRASLVLYLQSHV--AWGDNVAAAWTQGLDNVYHILFLALDGNPIL--  
CTEVIA4 KARGISQAQFNEFRASLVLYLQSHV--SWGDNVAAAWTQGLDNVYHILFLALDGNPIL--  
CTEVIA5 KARGISQAQFNEFRASLVLYLQSHV--AWGDNVAAAWTQGLDNVYHILFLALDGNPIL--  
CTEVIA6 KARGISQAQFNEFRASLVLYLQSHV--SWGDNVAAAWTQGLDNVYHILFLALDGNPIL--  
CTEVIA7 KARGISQAQFNEFRASLVLYLQSHV--AWSDATGAAWTQGLDNVYHILFLALDGNPIL--  
CTEVIA8 KARGISQAQFNEFRASLVLYLQSHV--AWSDATGAAWTQGLDNVYHILFLALDGNPIL--  
CTVIA KARGISQAQFNEFRASLVLYLQSHV--AWNAAEASAWTAGLNDVYHILFLALDGNPIL--  
CTVIA.1 KARGISQAQFNEFRASLVLYLQSHV--AWNAAEASAWTAGLNDVYHILFLALDGNPIL--  
CTEI KTR--ASPAQDFSRKSLVVMKLGAT--AWDATVESWNAVLDVDFVGLKEQLAAAK---  
CTTI KSR--ASPAQLDNFRKSLVVMKLGAT--KWDSAVESWAPVLDVDFVSTLKNEL-----  
CTTIA KSR--ASPAQLDNFRKSLVVMKLGAT--KWDSAVESWAPVLDVDFVSTLKNEL-----  
CTET KARGVTKAQFNTFRALVLYLQSHV--TWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEU KPRGVSQDFEFTALVLYLQSHV--TFDEPTEAWTALDNTYAVVVKALAA---  
CTTVI.3 KPRGVSQDFEFTALVLYLQSHV--TFDEPTEAWTALDNTYAVVVKALAA---  
CTTY KERGITVKEFDDFHAALHTYPTHT--KWNENVEASWMCIEKEARKTINAVLEGHPIE--  
CTPY KGRGITVKEFDDFHAALHTYPTHT--KWNENVEASWMCIEKEARKTINAVLEGHPIE--  
CTEW KARGVTKELFEKFNIAFMGYLKSHT--TWDEKTEAWKVVGDHHAIVYSILE-----  
CTTW KARGVTKELFEKFNIAFMGYLKSHT--TWDEKTEAWKVVGDHHAIVYSILE-----  
CTPW KARGVTKELFEKFNIAFMGYLKSHT--TWDEKTEAWKVVGDHHAIVYSILE-----  
CTEV KNRGVTQGFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTTV\_ps KNRGITQTFNKFRTALVLYLQSHV--AWGDNVAAAWNHALDNTYAVVVKALAA---  
CTPV KNRGVTQGFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEZ KARGVTQGFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTTZ KNRGITQTFNKFRTALVLYLQSHV--AWGDNVAAAWNHALDNTYAVVVKALAA---  
CTPZ KNRGITQTFNKFRTALVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB12 QARGVSAQFNEFRATLIVLYLQSHV--SWGENVAAAWNHALDNTYAVVVKALAA---  
CTEVIB13 QARGVSAQFNEFRATLIVLYLQSHV--SWGENVAAAWNHALDNTYAVVVKALAA---  
CTEVIB7 KARGITAAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB6 KARGVTAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB6a KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB5 KARGVTAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB5a KARGVTAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB10 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB10a KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB9 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB9a KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB8 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB4 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEVIB11 NRGVTKAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB7 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB6 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB5 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB10 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB9 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB8 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB4 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB9/5 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB7 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB6 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB5 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB10 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB9 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB8 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTVIB4 KARGVSAQFNEFRATLIVLYLQSHV--SWGDNVAAAWNHALDNTYAVVVKALAA---  
CTEN KARGVTKDEFIKFKDSIFKYLKAHV--EMDEKVLHAWNDPDSMYIIFANFDGKPVS--  
GLYP16\_4\_ORF2 KTRGVTKDQFNEFRATLIVLYLQSHV--DPTADYEHAWNDPDKMYVIFSTLDGKSFN--  
GLYP16\_4\_ORF3 RNRGIPKDLFNNFRATLIVLYLQSHV--EGYNAQYWDKSDKMYVIFSTLDGKSFN--  
GLYP5-4.5\_ORF2 KNRGIVKEQFNGFRATLIVLYLQSHV--SWNDAAANFNKALDNTYAVVVKALAA---  
GLYP5-4.5\_ORF3 RNRGATKEQFEKFKSSVAKYLDKDHSGSWSDATGAAWNKAFFDEMYAIVFSNLDGNPVH--  
GLYP5-4.5\_ORF1 KNRGVSQAQFNEFRATLIVLYLQSHV--SWNDATAAAWTHGLDNVYHILFLALDGNPIL--  
GLYP5-1.2\_ORF1 KNRGVSQAQFNEFRATLIVLYLQSHV--SWNDATAAAWTHGLDNVYHILFLALDGNPIL--

Alignment der H-moglobin-Sequenzen von *C. tentans*, *C. t. thummi* und *D. melanogaster*

	1	60
CTEU	-----	-----
CTEVII	MKFLVLALCIAAASAAV--TPMTADQLALFKSSWAEVK---HNEVDILYAFKDHDPDIQA	
CTEIX	MKFLAVLALCIVGAVAD--PISADQADLRSSWAQVK---HNEVDILAAVFKDHDPDIQA	
CTEIIβ	MKFL-VLALCIVGAIIA--PLSADEASLVRASWVQAK---HSEVDILYYIFKANPDIMA	
CTEVIIA-01	MKFFAVLALCIVGAIAT--PLTADQASLVQSTWAQVR---NNEVDILAAVFTANPDIQA	
CTEVII B-01	MKFFAVLAVCIVGAIAT--PLTADQANLVQSSWAQVK---NNEVDILAAVFAAYPDIQA	
CTEI	MKFL-ILALCVAYTMAA--P-SADQIAAAKASWATVK---GKQVDILAAIFKANPDIKD	
CTEE	MKFI-ILALCVAASA---LSADQIGLVQTTYGKVK---GDSTGILYAVFKADSSIQ	
CTEIII-1	MKFL-ILALCFAASA---LSADQISTVQASFQVK---GDSVGLLYAVFKADPSIMA	
CTEIV-1	MKFI-VLALCFAASA---LTADQISTVQASFQVK---GDSVGLLYAVFKADPSIQA	
CTTVI	-----A---VLTTEQADLVKKTWSTVK---FNEVDILYAVFKAYPDIMA	
CTTIIβ	-----A---PLSADEASLVRGSAQVK---HSEVDILYYIFKANPDIMA	
CTTIX	-----D---PVSSDEANIRASWAGVK---HNEVDILAAVFSHDPDIQA	
CTTVIII	-----AV--TPMSADQLALFKSSWNTVK---HNEVDILYAVFKANPDIQA	
CTTVIIA	-----A---PLSADQASLVKSTWAQVR---NSEVEILAAVFTAYPDIQA	
CTTVIIB	-----SPLTAEASLVQSSWKAVS---HNEVDILAAVFAAYPDIQA	
CTTIII	-----LSADQISTVQASFQVK---GDPVGLLYAVFKADPSIMA	
CTTIV	-----LTADQISTVQSSFAGVK---GDAVGLLYAVFKADPSIQA	
CTTIIa	-----GP-SGDQIAAAKASWNTVK---NNQVDILYAVFKANPDIQT	
CTTI	-----GP-SGDQIAAAKASWNTVK---NNQVDILYAVFKANPDIQT	
CTTX	-----DPEWHTLDAHEVEQVQATWKAWS---HDEVEILYTVFKAHPDIMA	
Dme glob1	-----NSDEVQLIKKTWEI PVATPTD SGAAIITQFFNRFPSNLE	
		120
CTEU	-----DLESVKDSPAFATHATRIISFLSQVISLAGSDDTIPAIQNLAKDLATSHKPR	
CTEVII	RFPQFAGKDLDSLKDTAAAFATHAGRIVGFFSDVIGLIGNENRPAKTLIDGLASNHKAR	
CTEIX	RFPQFAGKDVDSLKDTAAAFATHAGRIVGFISEIVALVGNENRPAAMTLTINELATNHHR	
CTEIIβ	KFPQFVKGDLDSLKDTAAAFATHAGRIVGFMSQI IALMGNEANRPAAMVTLINEMAANHKAR	
CTEVIIa-01	RFPQFAGKDVAALKDTAAAFATHAGRIVGFVSEI IALVGNENRPAVQTLVGLAASHKAR	
CTEVII B-01	KFPQFAGKDLAAVKDTAAAFATHATRIVSFLSELISLSGNSANAAAINSLVTKLGSDDHQR	
CTEI	KFSQFAGKDVDSIKGTFEPATHAGRIVGFFGDMVLLGNDANTATIVSKAKDLGKTHKTR	
CTEE	AFPQFVKGDLDAIKGGAEPATHAGRIVGFLSGV---IDDLPNIGKH---VDALVATHKPR	
CTEIII-1	KFTQFAGKDLSEIKGTAPPESHANRIVGFFSKI---IGDLPSIEGD---VDTFVASHKPR	
CTEIV-1	KFSQFAGKDLSEIKGSAEPASHAGKIVGFFSKI---VGLDPNISGD---VDTFVASHKPR	
CTTVI	KFPQFAGKDLDSIKDSAAFATHATRIVSFLSEVISLAGSDANIPAIQNLAKELATSHKPR	
CTTIIβ	KFPQFAGKDLSEIKGTGQFATHAGRIVGFVSEI IALMGNSANMPAMETLIDMAANHKAR	
CTTIX	RFPQFAGKDLASIKDTGAFATHAGRIVGFISEI IALVGNENRPAAMATLINELSTSHHR	
CTTVIII	KFPQFAGKDLDSIKDSADFAVHSGRIVGFFSEVIGLIGNENRPAKTLIDGLASSHKAR	
CTTVIIA	RFPQFAGKDVASIKDTGAFATHAGRIVGFVSEI IALIGNENRPAVQTLVGLAASHKAR	
CTTVIIB	KFPQFAGKDLASIKDTGAFATHATRIVSFLSEI IALVGNENRPAAMVQTLVGLAASHKAR	
CTTIII	KFTQFAGKDLSEIKGTAPPETHANRIVGFFSKI---GELPNIEAD---VNTFVASHKPR	
CTTIV	KFTQFAGKDLSEIKGSADFAHANKIVGFFSKI---GDLPNIDGD---VNTFVASHKPR	
CTTIIa	AFSQFAGKDLSEIKGTPDFSKHAGRVVGLFSEVMDLLGNDANTPTILAKAKDFGKSHKSR	
CTTI	AFSQFAGKDLSEIKGTPDFSKHAGRVVGLFSEVMDLLGNDANTPTILAKAKDFGKSHKSR	
CTTX	KFPQFAGKDLSEIKDTADFAVHASRIIGFFGEYVTLGSSGNQAAIRTLHDLGVFHKTR	
Dme glob1	KFP-PRDVPLEELSGNARFRAHAGRIIRVFDESIVLQGDGDLKLEIWTKIAVSHIPR	
		174
CTEU	GVSKDQFTEFRTALVSYLKSHITFDEPTEATAWTLALETYKMLFAAMDS-----	
CTEVII	GIEKAQFEEFRASLVSYLKHLDWNTMKSTWDLALNNMFFYILHALESAH---	
CTEIX	GISKAQFNEFRASMTSYLSHHTWINDATAAAWTHGLDNI FDAIFAHL-----	
CTEIIβ	KIPQAQFNEFRASLVSYLQSHVTWNDKLGAAWTQGLDNATMI FAKL-----	
CTEVIIA-01	GISQAQFNEFRASLVSYLQANVAWGDVNAAAWTQGLDNIYGLLFAAL-----	
CTEVII B-01	GVSQAQFNEFRALVAYLQANVSWGENVAANWQALDNTYAIAPTTLTA-----	
CTEI	-ASPAQFDSFRKSLVWVWLGATAWDATVSSWNAVLDVFGVGLKEQLAAAK---	
CTEE	GVTHAQFNNFRAAFIGYLGKGVVDTAAVEAAWAGATFDAFFGAVFAKM-----	
CTEIII-1	GVTHDQLNNFRAGFVSYMKAHTDY-AASESAWAGATLDTFFGMIFAKM-----	
CTEIV-1	GATHDQLNNFRSAFVSYMKAHTDF-AGSESAWAGATLDAFFGAVFAKM-----	
CTTVI	GVSKDQFTEFRTALFTYLYLKAHINFDGPTETAWTLALDNTYAMLFSAMDS----	
CTTIIβ	GIPKAQFNEFRASLVSYLQSKVSWNDLSGAAWTQGLDNI FVNMFFSYL-----	
CTTIX	GITKQFNEFRSLSVSYLSHASWINDATAADWTHGLDNI FGMIFAHL-----	
CTTVIII	GIEKAQFEEFRASLVSYLHLLDWNMTMKSTWDLALNNMFFYILHALEVAQ---	
CTTVIIA	GISQAQFNEFRALVSYVSNVAWNAAAESAWTAGLDNI FGLLFAAL-----	
CTTVIIB	GVSAQFGEFRLALVAYLQAHVSWGNVNAANWASKALDNTYAIIVVPRL-----	
CTTIII	GVTHDQLNNFRAGFVSYMKAHTDF-AGAEAAWAGATLDTFFGMIFSKM-----	
CTTIV	GVTHDQLNNFRAGFVSYMKAHTDF-AGAEAAWAGATLDAFFGMVFAKM-----	
CTTIIa	-TSPAQLDNFRKSLVYVYLGKATKWSAVESSWAPVLDVFFVSTLKNEL-----	
CTTI	-ASPAQLDNFRKSLVYVYLGKATKWSAVESSWAPVLDVFFVSTLKNEL-----	
CTTX	GITKAQFGEFRETMTAYLKGHNKWNADISHSWDDAFDRAFSVIFEVLES-----	
Dme glob1	TVSKESYNQLKGVILDVLTAAACSLDESQAATWAKLVVHVVYAIIFKAIIDDDGNAK	

Ich versichere hiermit, die vorliegende Arbeit selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt zu haben.

Solingen den 22. September 2001  
Heike Friedl