

# **Das Nucleomorph-Genom der Cryptomonade *Guillardia theta***

DISSERTATION  
zur Erlangung des Doktorgrades  
der Naturwissenschaften  
(Dr. rer. nat.)

dem Fachbereich Biologie  
der Philipps-Universität Marburg  
vorgelegt von

Stefan Zauner  
aus Ludwigsburg

Marburg/Lahn (2001)

Vom Fachbereich Biologie der  
Philipps-Universität Marburg

als Dissertation am 30.03.2001 angenommen.

Erstgutachter: Prof. Dr. Uwe-G. Maier

Zweitgutachter: Prof. Dr. Klaus Lingelbach

Tag der Disputation am: 25.04.2001

**Für meine Eltern**

**Ingrid und Johann Zauner**

---

Aus der vorliegenden Arbeit sind folgende Publikation entstanden:

Fraunholz, M. J., Wastl, J., Zauner, S., Rensing, S. A., Scherzinger, M.M., Maier, U.-G. (1997). The evolution of cryptophytes. *Plant Syst. Evol.* [Suppl.] **11**:163-174.

Wastl, J., Fraunholz, M. J., Zauner, S., Maier, U.-G. (1999). Lessons from nucleomorphs and chloroplasts. *Endocytobiology VII, Proceedings.*

Wastl, J., Fraunholz, M. J., Zauner, S., Douglas, S., Maier, U.-G. (1999). Ancient gene duplication and differential gene flow in plastid lineages: The GroEL/Cpn60 example. *J. Mol. Evol.* **48**: 112-117.

Maier, U. G., Fraunholz, M., Zauner, S., Penny, S., Douglas, S. (2000). A Nucleomorph-Encoded CbbX and the Phylogeny of RuBisCo Regulators. *Mol Biol Evol* Apr;17(4):576-583

Zauner, S., Fraunholz, M., Wastl, J., Penny, S., Beaton, M., Cavalier-Smith, T., Maier, U. G., Douglas, S. (2000) Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph. *Proc Natl Acad Sci (USA)* **97**: 200-205.

Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L.-T., Wu, X., Reith, M., Cavalier-Smith, T., Uwe-G Maier, U.-G. (2001). The highly reduced genome of an enslaved algal nucleus. *Nature* **410**: 1091 - 1096

Weitere Publikationen :

McLenachan, P. A., Stöckler, K., Winkworth, R. C., McBreen, K., Zauner, S., Lockhart, P.,J. (2000). Sequence specific markers derived from AFLP gels for plant ecology and evolution studies. *Molecular Ecology* **9**: 1899-19

# I. INHALTSVERZEICHNIS

<b>I. INHALTSVERZEICHNIS .....</b>	<b>1</b>
<b>II. ABKÜRZUNGSVERZEICHNIS .....</b>	<b>8</b>
<b>1. EINLEITUNG.....</b>	<b>9</b>
<b>1.1. Die Entstehung der Plastide.....</b>	<b>9</b>
1.1.1. Die primäre Endocytobiose.....	9
1.1.2. Die sekundäre Endocytobiose.....	11
<b>1.2. Die Cryptomonaden (<i>Guillardia theta</i>).....</b>	<b>15</b>
1.2.1. Das Nucleomorph .....	17
<b>1.3. Das Genomprojekt .....</b>	<b>19</b>
<b>1.4. Zielsetzung .....</b>	<b>20</b>
<b>2. MATERIAL UND METHODEN .....</b>	<b>21</b>
<b>2.1. Material.....</b>	<b>21</b>
2.1.1. Chemikalien .....	21
2.1.2. Puffer.....	22
2.1.3. Filter und Membranen.....	23
2.1.4. Fotomaterialien .....	23
2.1.5. Molekularbiologische Kits und Reagenzien .....	24
2.1.6. Restriktionsendonucleasen, Polymerasen und DNA-modifizierende Enzyme....	24
2.1.7. DNA-Längenstandards .....	25
2.1.8. Oligonucleotide.....	25
2.1.9. Vektoren.....	25
2.1.10. Bakterienstämme.....	26
2.1.11. Kultur-Medien.....	27
2.1.11.1. Medien für <i>E. coli</i> .....	27
2.1.11.2. Kultur-Medium f/2 für die Algen .....	27
2.1.12. Geräte.....	28
2.1.13. Soft- und Hardware.....	29
<b>2.2. Methoden .....</b>	<b>30</b>
2.2.1. Algenkultur und Ernte.....	30
2.2.2. Einbetten der Algenzellen in Agarose .....	30
2.2.3. Pulsfeld-Gelelektrophorese (PFGE) der Nucleomorph Chromosomen.....	31
2.2.4. Generelle molekularbiologische Methoden .....	32
2.2.4.1. Plasmidisolierung aus <i>E. coli</i> .....	32
2.2.4.2. Isolierung genomischer DNA aus <i>G. theta</i> .....	32

2.2.4.3. Isolierung von DNA aus Bakteriophagen $\lambda$ .....	33
2.2.5. Herstellung Nucleomorph Chromosom II angereicherter Plasmidbibliotheken..	33
2.2.6. Polymerase-Kettenreaktion (PCR) .....	33
2.2.7. DNA-Sequenzierung.....	34
2.2.7.1. Das ALFexpress System.....	34
2.2.7.2. Das LI-COR System .....	35
2.2.8. Analyse fluoreszenzmarkierter DNA-Stränge durch die Sequenzierautomaten..	35
2.2.9. Alignment der Sequenzen.....	36
2.2.10. Southern-Transfer von hochmolekularen Nukleinsäuren.....	36
2.2.11. Markierung von DNA.....	36
2.2.12. Hybridisierung von Nukleinsäuren.....	37
<b>3. ERGEBNISSE.....</b>	<b>38</b>
<b>3.1. Die Klonierungsstrategie.....</b>	<b>38</b>
3.1.1. Das Alignment der Sequenzen von Chromosom II .....	38
3.1.2. Arbeiten mit einer genomischen $\lambda$ EMBL3 Bibliothek.....	39
3.1.3. Plasmidbibliotheken.....	39
3.1.3.1. Die <i>Xba</i> I Plasmidbibliothek.....	41
3.1.3.2. Die <i>Eco</i> RI Plasmidbibliothek .....	42
3.1.3.3. Die <i>Spe</i> I Plasmidbibliothek .....	42
3.1.3.4. Die <i>Bam</i> HI Plasmidbibliothek.....	42
3.1.3.5. Die <i>Bgl</i> II und <i>Hind</i> III Plasmidbibliotheken.....	43
3.1.3.6. PCR zur Überprüfung der Alignments und Schließen verbliebener Sequenzlücken.....	43
<b>3.2. Die Sequenz des Nucleomorph Chromosoms II.....</b>	<b>46</b>
3.2.1. Die subtelomerischen <i>inverted repeats</i> .....	46
3.2.2. Die <i>single-copy</i> Region.....	47
3.2.2.1. Identifikation offener Leserahmen.....	47
3.2.2.1.1. Überlappende Protein-codierende Leserahmen .....	57
3.2.2.1.2. Introns in Protein-codierenden Leserahmen .....	59
3.2.2.1.3. Codon-Nutzung.....	61
3.2.2.2. Das Nucleomorph ist transkriptionell aktiv .....	62
3.2.2.3. Identifikation tRNA-codierender Sequenzabschnitte .....	63
3.2.2.3.1. Überlappungen tRNA-codierender Sequenzabschnitte mit Protein- codierenden Leserahmen .....	65
3.2.2.4. Identifikation weiterer RNA-Spezies.....	66
<b>3.3. Nucleomorph Chromosomen I und III .....</b>	<b>68</b>
3.3.1. Nucleomorph Chromosom I .....	68
3.3.2. Nucleomorph Chromosom III.....	70
<b>3.4. Zusammenfassende Darstellung der strukturellen Daten der Nucleomorph Chromosomen von <i>G. theta</i>.....</b>	<b>72</b>

<b>4. DISKUSSION .....</b>	<b>73</b>
<b>4.1. Der strukturelle Aufbau des Nucleomorph-Genoms von <i>Guillardia theta</i> .....</b>	<b>73</b>
4.1.1. Die <i>inverted repeats</i> sind für die strukturelle Stabilität und die Erhaltung der Telomere der Nucleomorph Chromosomen verantwortlich .....	73
4.1.2. A/T-Gehalt und der Einfluss auf die Codon-Nutzung .....	76
4.1.3. Das Nucleomorph-Genom ist extrem kompakt .....	82
4.1.3.1. Die intergenischen Regionen sind sehr kurz, codierende Bereiche können überlappen .....	83
4.1.3.2. Die Transkripte Nucleomorph-codierter, offener Leserahmen überlappen und die 3'-Prozessierung ist inakurat.....	86
4.1.3.3. Repetitive Strukturen und Pseudogene .....	87
4.1.4. Introns .....	89
4.1.4.1. Spliceosomale Introns in Protein-codierenden offenen Leserahmen sind sehr klein und selten.....	89
4.1.4.2. tRNAs besitzen teilweise zwei Introns und Introns in der D- bzw. der Pseudouridin-Schleife.....	90
4.1.5. Centromere und Mitose.....	90
4.1.6. Gegenüberstellung der Nucleomorph-Genome von Cryptomonaden und Chlorarachniophyceae .....	92
<b>4.2. Die Genausstattung des Nucleomorph-Genoms von <i>G. theta</i> .....</b>	<b>94</b>
4.2.1. Das Nucleomorph-Genom codiert hauptsächlich für <i>house-keeping</i> Funktionen	94
4.2.2. Das Nucleomorph ist das Genom eines semiautonomen eukaryoten Organells..	95
<b>4.3. Ausblick .....</b>	<b>98</b>
<b>5. ZUSAMMENFASSUNG .....</b>	<b>99</b>
<b>6. LITERATURVERZEICHNIS.....</b>	<b>101</b>
<b>7. ANHANG.....</b>	<b>110</b>

## II. ABKÜRZUNGSVERZEICHNIS

A	Adenin	RT	Raumtemperatur
AS	Aminosäurereste	SDS	Natriumdodecylsulfat
APS	Ammoniumpersulfat	SLS	Natriumlauroylsulfat
bp	Basenpaar(e)	SSU	<i>small subunit</i> , kleine ribosomale Untereinheit
C	Cytosin	T	Thymin
cDNA	<i>copy</i> -DNA	TEMED	N,N,N',N'-Tetramethylethylen-diamin
CHEF	<i>clamped homogenous electric field</i>	Tris	Tris (hydroxymethyl)-amino-methan
CTAB	Cetyltrimethylammonium Bromid	U	<i>units</i> (als Enzymaktivität), Uracil
dATP	2'-Desoxyadenosintriphosphat	X-Gal	5-Brom-4-chlor-3-indolyl- $\beta$ -D-galactopyranosid
dCTP	Desoxycytidintriphosphat		
DEAE	Diethylaminoethyl		
dGTP	2'-Desoxyguanosintriphosphat		
dNTP	2'-Desoxynucleosidtriphosphat		
DTT	Dithiothreitol		
dTTP	2'-Desoxythymidintriphosphat		
EDTA	Ethylendiamintetraacetat		
ER	Endoplasmatisches Reticulum		
EtBr	Ethidiumbromid		
G	Guanosin		
IPTG	Isopropyl- $\beta$ -D-thiogalactosid		
kb	Kilobasenpaar(e)		
LMP	<i>low melting point</i>		
LSU	<i>large subunit</i> , große Untereinheit		
mRNA	messenger-Ribonucleinsäure		
nt	Nucleotide (als Längenangabe für Einzelstrang-Nucleinsäuren)		
PFGE	Pulsfeldgel Elektrophorese		
rDNA	Gen für ribosomale Ribonucleinsäure		
rRNA	ribosomale Ribonucleinsäure		

# 1. EINLEITUNG

## 1.1. Die Entstehung der Plastide

Plastiden sind die charakteristischen Organellen aller photoautotrophen Eukaryoten. Bereits vor mehr als 100 Jahren bemerkte Schimper (1883), dass „Chlorophyllkörner“ (die Plastiden) nie *de novo* entstehen, sondern vielmehr immer aus ihresgleichen hervorgehen. Er verglich ihre Beziehung zu den sie enthaltenden Organismen mit einer Symbiose. Zu Beginn des letzten Jahrhunderts erkannte Mereschkowsky (1905) die morphologische und physiologische Ähnlichkeit der Plastiden zu freilebenden Cyanobakterien und formulierte die Endosymbionten-Theorie in ihren Grundzügen. Aber erst die Entdeckung Plastiden eigener DNA (Ris und Plaut, 1962; Chun et al., 1963) führte letztendlich zum Durchbruch der Endosymbionten-Theorie, die 1970 von Lynn Margulis in ihrer heutigen Form formuliert wurde (Margulis, 1970). Die Endosymbionten-Theorie ist heute generell akzeptiert und durch ultrastrukturelle, biochemische und molekularbiologische Untersuchungen belegt (z.B. Gray, 1992; Bhattacharya und Medlin, 1995; Cavalier-Smith, 2000).

### 1.1.1. Die primäre Endocytobiose

Nach der Endosymbionten-Theorie geht die Entstehung photoautotropher Eukaryoten auf eine primäre Endocytobiose zurück. Hierbei war das Initialereignis die phagocytotische Aufnahme eines freilebenden, photoautotrophen Prokaryoten durch einen heterotrophen Eukaryoten. Das phagocytote Cyanobakterium wurde nicht katabolisiert, vielmehr führte die Coevolution von Wirt und Endocytobiont zu einem, von zwei Membranen umgebenen semiautonomen Organell, die Plastide (Douglas, 1994; Palmer und Delwiche, 1996; Martin et al., 1998; Douglas, 1998; Gray, 1999; Cavalier-Smith, 2000).

Anhand morphologischer, ultrastruktureller und phylogenetischer Untersuchungen können drei evolutionäre Linien identifiziert werden, die durch eine primäre Endocytobiose entstanden sind: Die Glaucocystophyta, die Rhodophyta und die Chlorophyta (Abb. 1.1.). Sie alle besitzen Plastiden, die von zwei Membranen umgeben sind (Helmchen et al., 1995; Melkonian et al., 1995; Bhattacharya and Medlin, 1995;

Delwiche 1999). Gencluster-Analysen weisen ebenso auf eine monophyletische Entstehung dieser Plastiden hin, wie phylogenetische Berechnungen anhand von rRNA Sequenzen bzw. aneinandergereihten Protein-codierenden Sequenzen (Douglas, 1998; Martin et al., 1998; Stöbe und Kowallik, 1999; Maier et al., 2000).

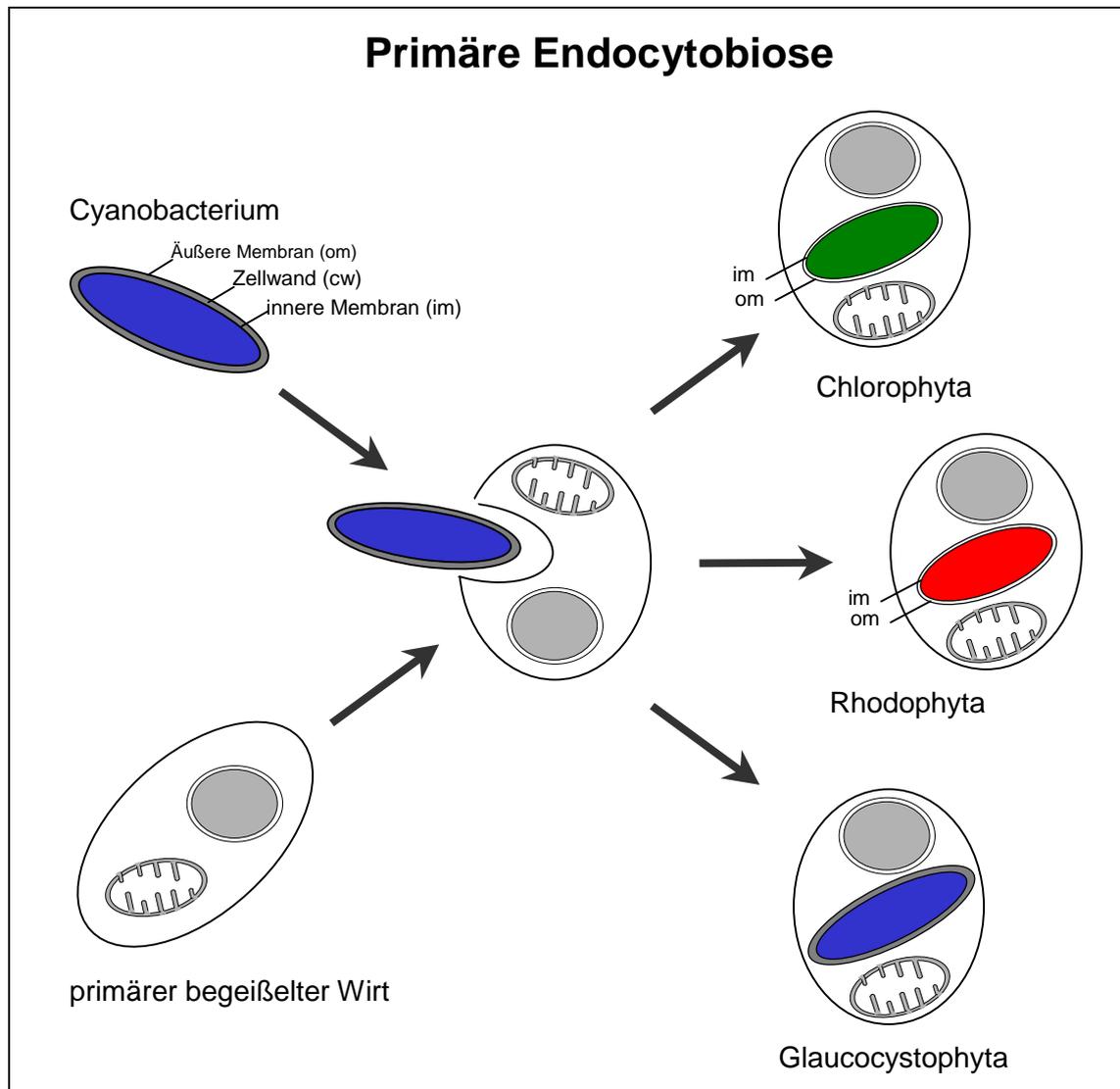


Abb. 1.1.: **Schema der primären Endocytobiose.** Dargestellt ist die monophyletische Entwicklung der Chlorophyta, der Rhodophyta und der Glaucocystophyta. Näheres siehe Text; nach Stöbe und Maier (2001).

Im Laufe der Coevolution erfolgte, bei dem vormals freilebenden Endocytobionten eine drastische Reduzierung des Genoms. Die Reduktion der codierenden Bereiche der Plastome, bei gleichzeitiger Aufrechterhaltung der Funktion der Plastide, erfasste in erster Linie redundante Funktionen, die vom Endosymbiont nicht mehr gebraucht wurden. Hierzu zählen zum Beispiel die Proteine die für die Synthese der Lipoproteine der

Membran zuständig sind. Diese konnten, ohne die Funktion des Symbionten zu beeinflussen eliminiert werden (Maier et al., 2000). Eine weitere Möglichkeit der Reduktion stellt der Gentransfer vom Symbionten zum Nucleus des Wirtes dar (z.B. Delwiche und Palmer, 1997; Cavalier-Smith, 2000;). Erste Hinweise auf einen DNA-Transfer von Plastom zu Wirtsgenom lieferte die Identifizierung plastidärer DNA im Nucleus von *Spinacia oleracea* (Timmis und Scott, 1983; 1984), heutzutage, in der Ära der Genomprojekte konnten im Genom von *Arabidopsis thaliana* Insertionen plastidärer DNA nachgewiesen werden (‘The Arabidopsis Genom Initiative’, 2000). Unter der Voraussetzung, dass Gene im Genom des Wirtes integrieren, diese mit einer N-terminalen Extension ausgestattet und deren Genprodukte synthetisiert werden, können diese im photosynthetisch aktiven Endosymbionten die Plastom-codierten Proteine funktionell ersetzen (Baldauf et al., 1990).

### **1.1.2. Die sekundäre Endocytobiose**

Mit der Endosymbionten-Theorie in obiger Form kann die Entstehung von Plastiden erklärt werden, die von zwei Membranen umgeben sind (Rhodophyta, Glaucocystophyta und Chlorophyta). Die meisten Algengruppen besitzen jedoch Plastiden, die von mehr als zwei Hüllmembranen umgeben sind, sogenannte komplexe Plastiden (Sitte, 1983), und somit nicht mit der primären Endocytobiose zu erklären sind. Die Entstehung komplexer Plastiden ist eleganter durch die Phagocytose und Reduktion eines photoautotrophen Eukaryoten anstelle eines Cyanobakteriums zu erklären (Gibbs, 1981; Douglas, 1994; McFadden und Gilson, 1995; Palmer und Delwiche, 1996; Cavalier-Smith, 2000). Nach dieser Hypothese gehen Algen mit komplexen Plastiden phylogenetisch auf phagotrophe Eukaryoten mit phototrophen, eukaryotischen Endosymbionten zurück. Zur Entstehung komplexer Plastiden waren demzufolge zwei aufeinanderfolgende, unabhängige Endocytobiose-Ereignisse nötig: Die phyletisch primäre Endocytobiose (Kap. 1.1.2.), gefolgt von einer phyletisch sekundären Endocytobiose. Diese ist charakterisiert durch die Aufnahme eines, aus einer primären Endocytobiose hervorgegangenen, photoautotrophen Eukaryoten durch eine phagotrophe eukaryotische Wirtszelle. Der Verlauf eines sekundären Endocytobiose-Ereignisses ist in Abbildung 1.2. schematisch dargestellt. Das Ergebnis dieser intertaxonischen Kombination (Sitte, 1993) war eine evolutionäre Chimäre aus zwei Zellen. Die resultierende Superzelle besaß ein redundantes Cytosol mit

Mitochondrium, Ribosomen, Endomembransystem und den Nucleus des Symbionten (Abb. 1.2.b). Diese wurden in der Evolution sukzessive eliminiert (Abb. 1.2.c), wobei schließlich nur noch eine Membran und die Plastide auf den ursprünglichen Endosymbionten zurückgehen (Abb. 1.2.d; Cavalier-Smith, 2000).

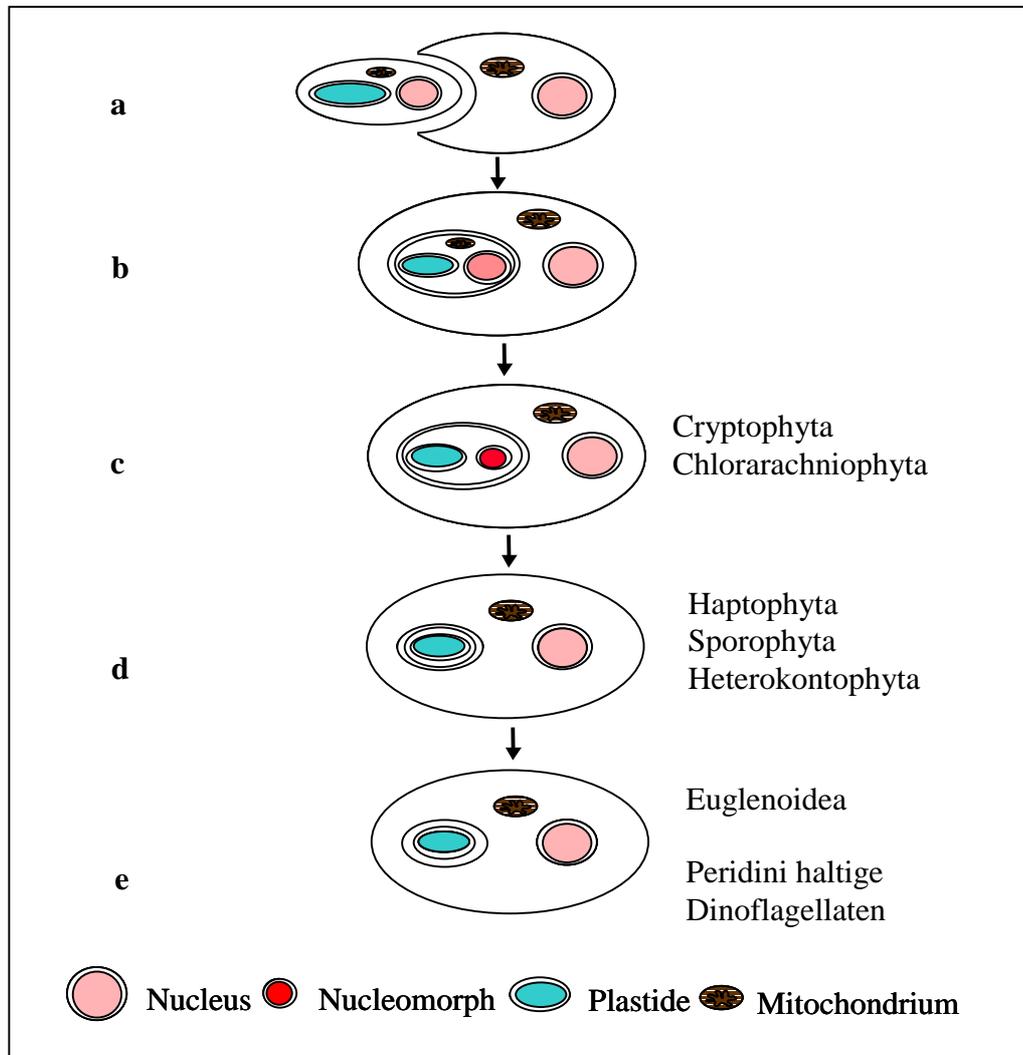


Abb. 1.2.: **Schematische Darstellung der sekundären Endocytobiose.** **a + b** Eine phototrophe eukaryote Zelle wird von einer eukaryoten Zelle phagozytiert. **c**, Reduktion des Endosymbiontencytosols zu einem periplastidären Kompartiment, gleichzeitige Reduktion des Nucleus zum Nucleomorph. **d**, Die vollständige Reduktion des periplastidären Kompartimentes führt zu einer komplexen Plastide, die von vier Membranen umgeben ist. **e**, Reduktion einer der beiden äußeren Membranen aus **d**, führt zu komplexen Plastiden mit drei Membranen (Maier et al., 2000).

Abbildung 1.2. stellt die Änderungen in der Anzahl der Plastidenmembranen dar, die im Laufe der sekundären Endocytobiose auftreten können. Sie zeigt nicht die phylogenetischen Beziehungen der durch Endocytobiose entstandenen Organismen.

Einen Beweis für die Richtigkeit der Annahme der sekundären Endocytobiose liefert die Existenz zweier rezenter Algengruppen, die eine intermediäre Stellung zwischen der eigentlichen Phagocytose und der Reduktion des Symbionten zu der mit vier Membranen umgebenen, komplexen Plastide darstellen (Abb. 1.2.c). Diese beiden Algengruppen, die Cryptophyta (Gillot, 1990; Zauner et al., 2000) und die Chlorarachniophyta (Ludwig und Gibbs, 1989; Gilson und McFadden, 1996), besitzen zwischen den beiden Membranpaaren einen Restzellkern, das Nucleomorph, das auf den Nucleus des Symbionten zurückgeht, und ein Restcytoplasma, das periplastidäre Kompartiment. Phylogenetische Untersuchungen der 18S rDNA Sequenzen der beiden Nucleomorph-haltigen Algengruppen (Abb. 1.2.c) belegen einen gemeinsamen Vorfahren des Symbionten der Cryptomonade *Guillardia theta* und der Rhodophyta (Van de Peer et al., 1996). Weitere phylogenetische Berechnungen an Nucleomorph-codierten Tubulinen in *G. theta* stützen diesen Befund (Keeling et al., 1999). Dem gegenüber stehen die Chlorarachniophyta, deren Symbiont einen gemeinsamen Vorläufer mit den Chlorophyta besitzt (Van de Peer et al., 1996). Die sekundäre Endocytobiose ist demnach polyphyletisch (Abb. 1.3.).

Auf der Stufe der Reduktion des Symbionten auf seine ehemalige Plasmamembran und die durch primäre Endocytobiose und daher von zwei Membranen umgebene Plastide (Abb. 1.2.d), befinden sich die komplexen Plastiden der Heterokontophyta und die Haptophyta (Cavalier-Smith, 2000). Neuere Befunde belegen die Existenz komplexer, jedoch photosynthetisch inaktiver Plastiden mit vier Membranen in den Apicomplexa (McFadden und Waller, 1997; McFadden et al., 1996). Hierzu gehören die humanpathogenen, obligat intrazellulären Parasiten *Plasmodium falciparum* und *Toxoplasma gondii*. Die komplexen Plastiden dieser drei Gruppen stellen Vertreter der „roten“ Linie dar, d.h. sie besitzen einen gemeinsamen Vorfahren mit den Rhodophyta. Weitere Vertreter der „roten“ Linie sind die Nucleomorph-besitzenden Cryptomonaden und die Peridinin haltigen Dinoflagellaten mit ihren von drei Membranen umgebenen komplexen Plastiden (Abb. 1.2.e; Maier et al., 2000). Einen gemeinsamen Vorläufer mit den Chlorophyta besitzen die komplexen Plastiden der Euglenoidea und der Chlorarachniophyta. Die komplexen Plastiden der Euglenoidea sind, ebenso wie die der Dinoflagellaten, durch die Reduktion einer der beiden äußeren Membran aus den komplexen Plastiden mit vier Membranen der Stufe 1.2.d hervorgegangen und besitzen demnach nur drei Plastidenmembranen (Cavalier-Smith, 2000). Eine Übersicht über die phylogenetischen Zusammenhänge der Plastiden-Evolution gibt Abbildung 1.3.

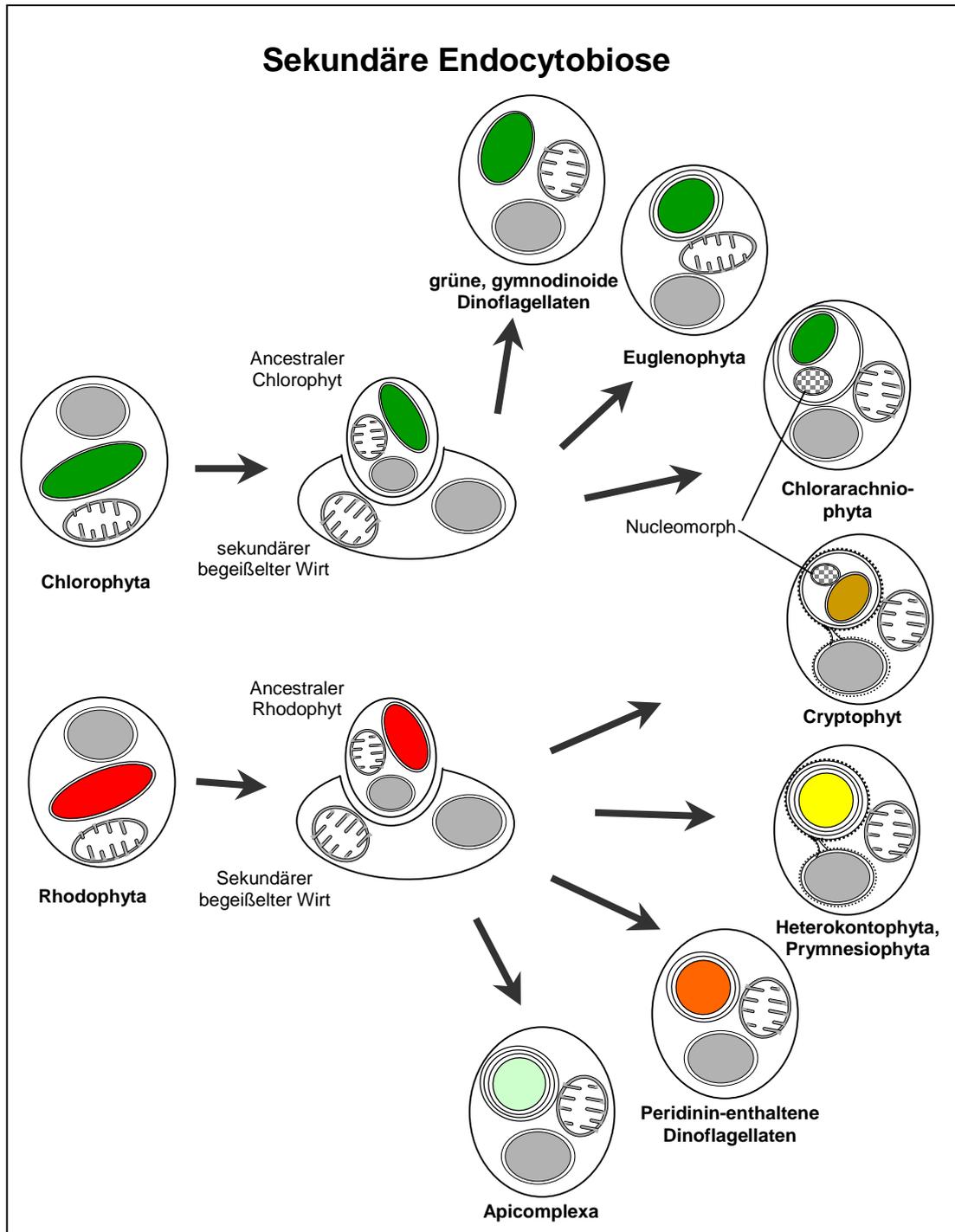


Abb. 1.3.: **Überblick über die Phylogenie der sekundären Endocytobiose.** Alle Organismen mit komplexen Plastiden gehen auf mindestens zwei unabhängige sekundäre Endocytobiose-Ereignisse zurück. Näheres siehe Text, nach Stöbe und Maier (2001).

## 1.2. Die Cryptomonaden (*Guillardia theta*)

Cryptomonaden sind eine kleine Gruppe mariner oder limnischer Phytoflagellaten, die eine charakteristische, abgeflacht-asymmetrische Form aufweisen. Ihre Fortpflanzung erfolgt durch Zweiteilung (Klaveness, 1981). Ihr ventrales Ende wird von einer Einbuchtung der Zelloberfläche (*gullet*) eingenommen. In dieser Einbuchtung sind zahlreiche größere Ejektosomen angeordnet und an ihrer Basis entspringen zwei ungleich lange Flagellen (Gillot und Gibbs, 1980). Die Zelle besitzt keine echte Zellwand, sondern einen Periplast aus Proteinplatten, die der Plasmamembran innen anliegen (Hibberd et al., 1971; Abb. 1.4.). Im Cytoplasma finden sich der Zellkern, die komplexe(n) Plastide(n), Dictyosomen und ein einziges netzartig verzweigtes Mitochondrium mit lamellären Cristae (Santore, 1984; Gillott, 1990). Cryptomonaden nutzen Amylopektin als Speicherstärke, welche im periplastidären Kompartiment in Form von Stärkekörnern oder haubenähnlich um das Pyrenoid der Plastide abgelagert wird (Gillot und Gibbs, 1980).

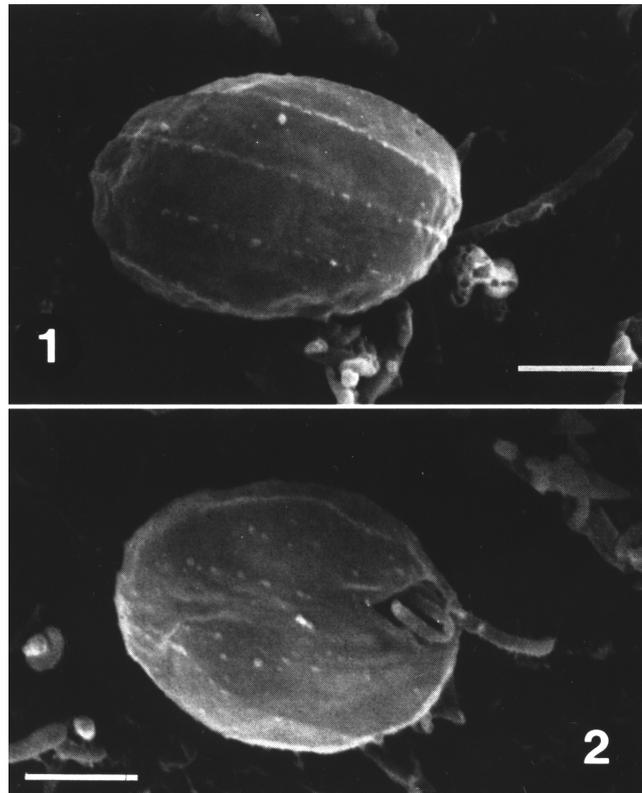


Abb. 1.4.: Rasterelektronenmikroskopische Aufnahme von *Guillardia theta*. 1) Deutlich zu erkennen ist die longitudinale Anordnung der Proteinplatten des Periplasten. 2) Aus einer Einbuchtung der Zelloberfläche entspringen zwei Flagellen. Balken = 2  $\mu\text{m}$  (aus Deane et al., 1998).

Die meisten Arten der Cryptomonaden besitzen eine, wenige zwei komplexe Plastiden, deren Gestalt von bootförmig bis H-förmig reicht. Die Pigmentausstattung besteht aus Chlorophyll a und c<sub>2</sub> und, je nach Art, eines der Phycobiliproteine Phycoerythrin oder Phycocyanin (Gantt, 1979). Letztere sind außer in Cryptomonaden und Organismen mit Cryptomonaden als Endosymbionten nur in Rhodophyten und Cyanobakterien identifiziert worden. Die Thylakoid-Membranen liegen paarweise aneinander, sind relativ dick und mit elektronendichtem Material gefüllt. Hierbei handelt es sich um die Phycobiliproteine, die dicht gepackt im Thylakoidlumen lokalisiert sind (Spear-Bernstein und Miller, 1989). Dies steht im Gegensatz zu den Rotalgen und den Cyanobakterien, deren Phycobiliproteine in Multiprotein-Komplexen aus Phycoerythrin, Phycocyanin und Allophycocyanin (Phycobilisomen), die sich auf der dem Stroma zugewandten Seite der Thylakoide befinden, angeordnet sind.

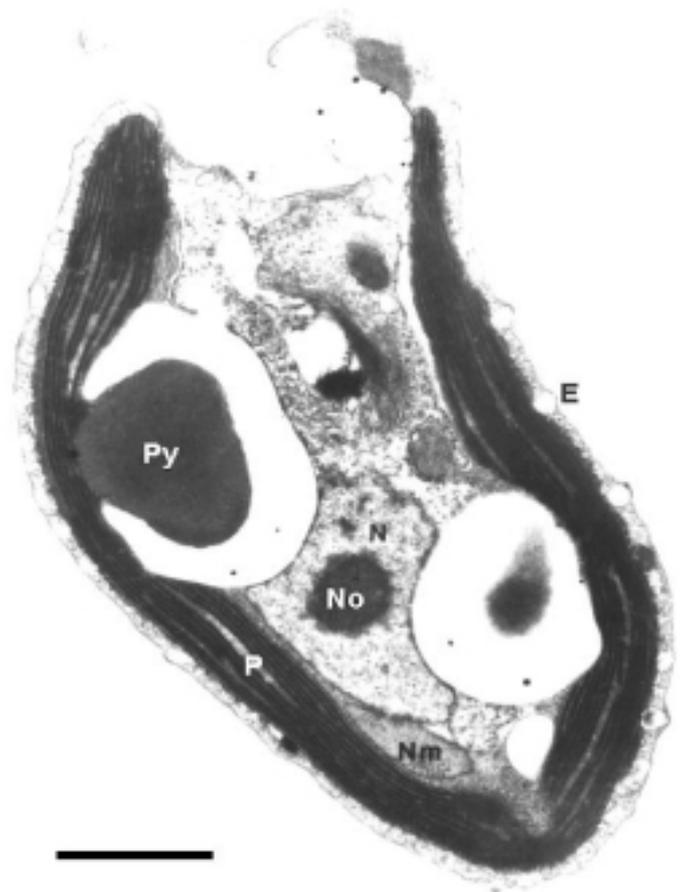


Abb. 1.5.: **Transmissionselektronen-**  
**mikroskopische Aufnahme von *G.***  
***theta*.** N, Nucleus; No, Nucleolus; Nm,  
Nucleomorph; P, Plastide; Py, Pyrenoid;  
E, Ejektosom. Balken = 1  $\mu\text{m}$  (aus  
Fraunholz et al., 1997)

Bei den Plastiden handelt es sich - wie bereits erwähnt - um komplexe Plastiden, die von vier Membranen umgeben sind. Das äußere Paar wird, da es mit der Kernhülle, und somit auch mit dem ER, in Verbindung steht als periplastidäres-ER oder Chloroplasten-ER (CER; Gibbs, 1981; McFadden et al., 1994) bezeichnet. Die äußere Membran ist zudem mit 80S Ribosomen besetzt. Zwischen dem äußeren und inneren Membranpaar liegt das periplastidäre Kompartiment, bei dem es sich um die cytoplasmatischen Reste des Endocytobionten handelt. Im periplastidären Kompartiment können keine Mitochondrien oder Dictyosomen identifiziert werden (Sitte und Baltes, 1990), wohl aber Ribosomen vom 80S-Typ (McFadden, 1990; McFadden et al., 1994), Stärkekörner, einige Vesikel und das Nucleomorph (Abb. 1.5.; Douglas et al., 1991; Maier et al., 1991).

### 1.2.1. Das Nucleomorph

Das Nucleomorph ist im periplastidären Kompartiment zwischen dem äußeren und inneren Plastiden-Membranpaar lokalisiert. Es ist von einer Doppelmembran mit Poren umgeben (Greenwood et al., 1977). Diese Poren sind, im Gegensatz zu anderen Nuclei, nicht kreisrund, sondern ähneln länglichen Schlitzten (Gillott and Gibbs, 1980). Das Nucleomorph enthält einen fibrillogranulären Bereich (Gillott and Gibbs, 1980) und teilt sich „amitotisch“ vor der Teilung des Zellkerns (McKerracher and Gibbs, 1982). Bei manchen Cryptomonaden, wie beispielsweise *Rhodomonas salina*, ist das Nucleomorph in das innerhalb der Plastide liegende Pyrenoid eingesenkt. Diese für einige Cryptomonaden typische Struktur besteht aus massiv akkumulierten Proteinen, für die Ribulose-Bisphosphat-Carboxylase-Aktivität nachgewiesen wurde (Okabe and Okada, 1988). Im Nucleomorph wurde sowohl DNA (Ludwig and Gibbs, 1985; Hansmann et al., 1986) als auch RNA (Hansmann, 1988) nachgewiesen. Durch *in situ* Hybridisierung auf elektronenmikroskopischer Ebene konnte gezeigt werden, dass es sich bei der rRNA der periplastidären Ribosomen und im fibrillogranulären Bereich des Nucleomorphs um eukaryotische rRNA handelt (McFadden, 1991), wohingegen die rRNA, die sich innerhalb der Plastide befindet, vom prokaryotischen-Typ ist (McFadden et al., 1994).

Hansmann und Eschbach (1990) isolierten aus der Cryptomonade *Rhodomonas salina* Nucleomorph-Pyrenoid-Komplexe. Im Nucleomorph von *Rhodomonas salina* konnten durch Pulsfeldgel-Elektrophorese (PFGE; Eschbach et al., 1987) drei lineare Chromosomen (240 kb, 225 kb und 195 kb) nachgewiesen werden (Eschbach et al., 1991).

Kurz darauf konnten auch in einem Nucleomorph der Chlorarachniophyten drei lineare Chromosomen nachgewiesen werden (McFadden et al., 1994). Rensing et al. (1994) konnten zeigen, dass alle bisher untersuchten Cryptomonaden drei Chromosomen in ihren Nucleomorphen besitzen, auf denen jeweils Gene für ribosomale RNA lokalisiert sind.

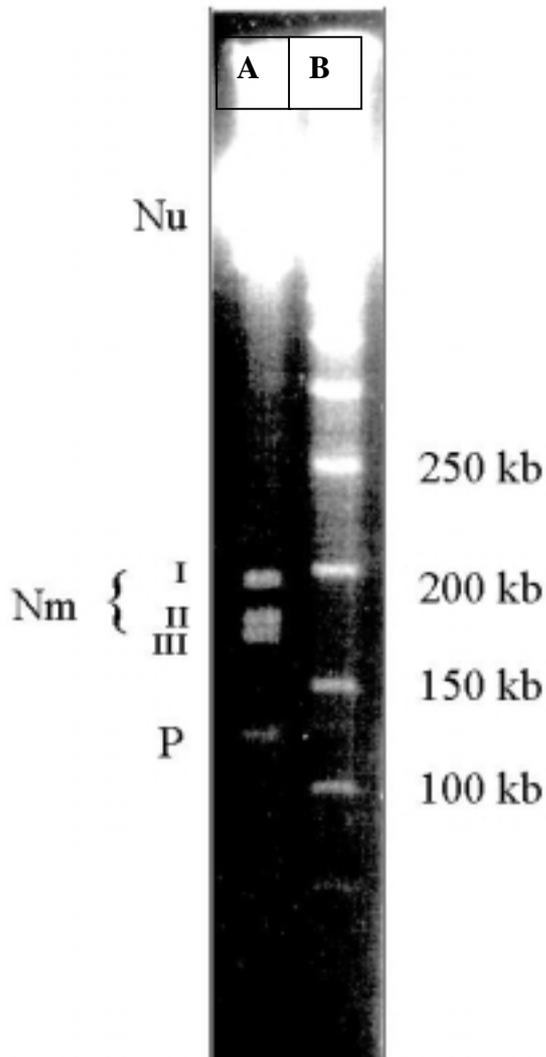


Abb. 1.6.: **Das Nucleomorph-Genom der Cryptomonade *Guillardia theta*.**

A) *G. theta* DNA. B)  $\lambda$ -Oligomer Standard. **Nm**, die drei Chromosomen des Nucleomorph-Genoms. Chromosom I ca. 200 kb, Chromosom II ca. 180 kb und Chromosom III ca. 175 kb. **P**, Plastidengenom (121 kb). Aus Fraunholz et al. (1997) verändert.

Die in dieser Arbeit bearbeitete Cryptomonade *Guillardia theta* besitzt nur eine Plastide mit dem Phycobiliprotein Phycoerythrin. Ihr Nucleomorph ist nicht in das Pyrenoid eingebettet und somit einer direkten Isolierung nur schwer zugänglich. Durch PFGE genomischer DNA konnten drei lineare Chromosome (200 kb, 180 kb und 175 kb) nachgewiesen werden (Abb. 1.6.; Rensing et al., 1994; Zauner et al., 2000).

### 1.3. Das Genomprojekt

Häuber et al. (1994) postulierten, dass Nucleomorphe existieren, um einen funktionellen Transkriptions – und Translationsapparat aufrechtzuerhalten. Dieser diene dazu wenige plastidäre Funktionen, die nicht durch Genprodukte des jeweiligen Nucleus ersetzt werden können, bereitzustellen. Um diese Hypothese und die gesamte Endosymbionten-Theorie zu überprüfen, wurde ein internationales Genom-Projekt ins Leben gerufen, dessen Ziel die vollständige Sequenzierung der Nucleomorphgenome einer Cryptomonade, *Guillardia theta*, und einer Chlorarachniophyte, *Pedinomonas minutissima*, ist. Für die Klärung weitreichender Fragen bezüglich der biologischen Funktion des Nucleomorphes sind insbesondere Nucleomorph-codierte Gene mit plastidärer Funktion, die Organisation des Nucleomorph-Karyotyps und Transportmechanismen von Interesse. Die Daten aus beiden Organismen sollen klären, welche Mechanismen in diesen evolutionär unterschiedlichen Organismen zur Etablierung eines Nucleomorphs innerhalb eines periplastidären Kompartiments geführt hat, und welche plastidären Faktoren auf diesen kleinsten bekannten eukaryoten Genomen codiert sind. Ein weiteres Ziel des Nucleomorph-Genomprojektes ist es, die evolutionären Mechanismen der Genom-Reduktion molekular zu untersuchen und die Frage zu klären, auf welchen Ebenen und in welchem Ausmaß die Zusammenarbeit der unterschiedlichen Genome in einer derart komplexen Zelle bewerkstelligt wird.

An diesem Genomprojekt sind Arbeitsgruppen aus Kanada (Susan Douglas, Halifax), England (Thomas Cavalier-Smith, Oxford), Australien (Geoffrey McFadden, Melbourne) und Deutschland (Uwe Maier, Marburg) beteiligt.

## 1.4. Zielsetzung

Das Ziel dieser Arbeit ist es, das Nucleomorph Chromosom II der Cryptomonade *Guillardia theta* vollständig zu sequenzieren. Dabei soll ausgehend von den bereits bekannten subtelomerischen *inverted repeats*, die zentrale *single-copy* Region des Nucleomorph Chromosoms II näher charakterisiert werden. Zu diesem Zweck wird aus einer vorhandenen genomischen  $\lambda$ EMBL3-Bibliothek und aus mehreren zu erstellenden Plasmidbibliotheken Nucleomorph-DNA des Chromosoms II isoliert und anschließend sequenziert. Die Chromosomen I und III sollen auf Basis der im Genomprojekt erhaltenen Sequenzen abschließend bearbeitet und in Zusammenarbeit mit Susan Douglas (Halifax, Kanada) zur Publikationsreife gebracht werden. Anhand der erhaltenen Sequenzinformationen soll das Nucleomorph-Genom in seiner Struktur und Funktion untersucht werden. Dabei steht die Frage nach dem Aufbau der Nucleomorph Chromosomen, speziell Chromosom II, im Vordergrund. Ein besonderes Augenmerk soll dabei auf die Identifizierung codierender Bereiche und deren Organisation gelegt werden.

## 2. MATERIAL UND METHODEN

### 2.1. Material

#### 2.1.1. Chemikalien

Alle nicht aufgeführten Chemikalien stammten von den Firmen Roth (Karlsruhe), Merck (Darmstadt) oder Sigma (München) und waren von Analysenqualität.

• Agarose (NEEO)	Roth, Karlsruhe
• Ampicillin	Boehringer, Mannheim
• Ammoniumpersulfat	Serva, Heidelberg
• ATP	Boehringer, Mannheim
• Bacto Agar	Roth, Karlsruhe
• Bacto Hefeextrakt	Roth, Karlsruhe
• Bacto Trypton	Roth, Karlsruhe
• Binding-Silan	Pharmacia, Freiburg
• Blocking Reagenz	Boehringer, Mannheim
• Bromphenolblau	Serva, Heidelberg
• BSA	Serva, Heidelberg; Sigma, München
• <i>CSPD</i> und <i>CDP-Star</i>	Boehringer, Mannheim
• DMSO	Merck, Darmstadt
• Na <sub>2</sub> EDTA	Sigma, München
• Ethidiumbromid	Roth, Karlsruhe
• IPTG	Roth, Karlsruhe
• Kanamycin	Serva, Heidelberg
• LMP-Agarose (Seaplaque GTG)	Biozym, Hameln
• <i>Long Ranger-Gel Solution</i>	FMC Bioproducts, Rockland (USA)
• N-Lauroylsarcosin	Sigma, München
• Pepton (tryptisch verdaut)	Roth, Karlsruhe
• Phenol (TE-gepuffert)	Roth, Karlsruhe
• Proteinase K	Boehringer, Mannheim
• Pronase	Boehringer, Mannheim
• <i>Rapid gel XL (40 %)</i>	Amersham Pharmacia, Freiburg
• SDS	Roth, Karlsruhe
• SLS	Roth, Karlsruhe
• Spermidin	Sigma, München
• TEMED	Serva, Heidelberg
• Tetracyclin	Sigma, München
• Triton X-100	Sigma, München
• X-Gal	Roth, Karlsruhe
• Xylencyanol FF	Fluka, Schweiz

### 2.1.2. Puffer

Hier aufgeführte Puffer und solche, die unmittelbar aus diesen hervorgehen, werden im Kapitel „Methoden“ nicht nochmals gesondert aufgelistet.

- 10 x TE-Puffer
 

100	mM	Tris/HCl, pH 8,0
10	mM	EDTA, pH 8,0
  
- 10 x TBE
 

1	M	Tris/HCl, pH 8,8
0,83	M	Borsäure
10	mM	EDTA
  
- SM-Puffer
 

50	mM	Tris/HCl, pH 7,5
100	mM	NaCl
10	mM	MgSO <sub>4</sub>
0,1	%(w/v)	Gelatine
  
- Top-Agarose:
 

0,7	%(w/v)	Agarose in 10 mM MgSO <sub>4</sub>
-----	--------	------------------------------------
  
- 20 x SSC:
 

3,0	M	NaCl
0,3	M	Na-Citrat
		pH 7,0
  
- STET
 

100	mM	NaCl
10	mM	Tris/HCl, pH 8,0
1	mM	EDTA
5	%(v/v)	Triton X-100
  
- 5 x *random priming* Puffer
 

1	M	HEPES, pH 6,6
250	mM	Tris/HCl, pH 8,0
25	mM	MgCl <sub>2</sub>
50	mM	β-Mercaptoethanol
  
- Prähybridisierungs- und Hybridisierungslösung
 

		Roti-Hybri-Quick (Roth, Karlsruhe)
--	--	------------------------------------
  
- Waschpuffer
 

		2 x SSC bis 0,2 x SSC; mit je 0,1 % (w/v) SDS
--	--	---



### 2.1.5. Molekularbiologische Kits und Reagenzien

- ‘*Double-stranded nested Deletions-Kit*’ Pharmacia, Freiburg
- ‘*Prep-a-gene DNA Purification-Kit*’ BioRad, München
- ‘*Thermo sequenase fluorescent labelled primer cycle sequencing Kit with 7-deaza-dGTP*’ Amersham Pharmacia, Freiburg
- ‘*Nucleobond-AX Plasmid-Purification*’ Macherey-Nagel, Düren
- ‘*Plasmid-Midi-Kit*’ Qiagen, Hilden
- ‘*Plasmid-Mini-Kit*’ Qiagen, Hilden
- ‘*DIG-labelling and Detection-Kit*’ Boehringer Mannheim
- ‘*JETsorb DNA Extraction-Kit*’ GENOMED, Bad Oeynhausen
- ‘*TransformAID™ Bacterial Transformation-Kit*’ MBI Fermentas, St. Leon-Rot
- ‘*InsT/Aclone™ PCR Product Cloning Kit*’ MBI Fermentas, St. Leon-Rot

### 2.1.6. Restriktionsendonucleasen, Polymerasen und DNA-modifizierende Enzyme

Die verwendeten Restriktionsendonucleasen, Polymerasen und DNA-modifizierenden Enzyme wurden von den Firmen Boehringer (Mannheim), Pharmacia (Freiburg), MBI Fermentas (St. Leon-Rot) und Promega (Mannheim) bezogen.

- Alkalische Phosphatase (1000 U/ml)
- ‘*AgarACE™*’ (0,15-0,3 U/μl) Promega, Mannheim
- DNase I (10 U/ml)
- Klenow-Polymerase (2000 U/ml)
- Lysozym (20 mg/ml)
- Pronase (10 mg/ml)
- Proteinase K (10 mg/ml)
- RNase A (5 mg/ml)
- Taq DNA Polymerase (*low cons.*) (1000 U/ml)
- Taq DNA Polymerase (5000 U/ml)
- T4 DNA Ligase (*low cons.*) (1000 U/ml)
- T4 DNA Ligase (6000 U/ml)

### 2.1.7. DNA-Längenstandards

Der DNA-Längenstandard wurde durch dreistündige Hydrolyse von 12,5 µg λ-DNA mit den Endonucleasen *EcoR* I und *Hind*III selbst hergestellt (EK 100 ng/µl).

Der λ-DNA-PFGE Marker bestehend aus λ-Oligomeren wurde von Pharmacia (Freiburg) bezogen.

### 2.1.8. Oligonucleotide

Oligonucleotide wurden von der Firma MWG (Ebersberg) bezogen. Die Sequenzierprimer für den ALFexpress waren am 5'-Ende mit dem Fluoreszenzfarbstoff CY5 markiert, Primer für den Einsatz mit dem LI-COR 4200 mit IRD800- (800 nm-Kanal) bzw. IRD700-Farbstoff (700 nm-Kanal).

Name	Sequenz	Modifikation
M13 <i>universal</i>	5'-TGT AAA ACG ACG GCC AGT-3'	CY5/IRD700/IRD800
M13 <i>reverse</i>	5'-CAG GAA ACA GCT ATG ACC-3'	CY5/IRD700/IRD800

Alle Oligonucleotide für spezifische Amplifikationen wurden ohne zusätzliche Modifikationen ebenfalls von der Firma MWG (Ebersberg) bezogen. Eine Liste der Oligonucleotide, die im Rahmen der spezifischen Amplifikation mittels PCR genutzt wurden, ist im Anhang zu finden.

### 2.1.9. Vektoren

- pBluescript II KS (+) Amp<sup>r</sup>, *lacZ'*; f1ori; P<sub>T7</sub>; P<sub>T3</sub> (Stratagene, Heidelberg).
- pBluescript II SK (+) Amp<sup>r</sup>, *lacZ'*; f1ori; P<sub>T7</sub>; P<sub>T3</sub> (Stratagene, Heidelberg);
- pGEM-T *Vector System I* Amp<sup>r</sup>, *lacZ'*, P<sub>T7</sub>, P<sub>SP6</sub>, (Promega, Mannheim).pUC57/T Amp<sup>r</sup>, *lacZ'*; f1ori (MBI Fermentas, St. Leon-Rot).
- pUC19 Amp<sup>r</sup>, *lacZ'*; f1ori; geöffnet und dephosphoryliert (MBI Fermentas, St. Leon-Rot und Appligene, Heidelberg)
- λZAPII (Stratagene, Heidelberg; Short et al., 1988).
- λEMBL3 (Stratagene, Heidelberg; Frischauf et al., 1987).

### 2.1.10. Bakterienstämme

- *E. coli* XL1-Blue MRF<sup>1</sup> :  $\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac$  [F' *proAB lacI<sup>q</sup>Z*  $\Delta M15$  Tn5 (Kan<sup>r</sup>)] Stratagene, Heidelberg
- *E. coli* XL1-Blue MRA(P2):  $\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac(P2 lysogen)^c$  Stratagene, Heidelberg
- *E. coli* SOLR  $e14^-(mcrA^-) \Delta(mcrCB-hsdSMR-mrr)171 endA1 supE44 thi-1 sbcC uvrC umuC::Tn5(Kan^r)recJ gyrA96 relA1 lac$  [F' *proAB lacI<sup>q</sup>Z*  $\Delta M15$ ]<sup>c</sup> Stratagene, Heidelberg
- *E. coli* XL10-gold  $tet^f \Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte$  [F' *proAB lacI<sup>q</sup>Z*  $\Delta M15$  Tn10(Tet<sup>f</sup>) Amy Cam<sup>r</sup>]<sup>a</sup> Stratagene, Heidelberg

Die Bakterien wurden gemäß den Anweisungen der Firma Stratagene angezogen.

Kompetente *E. coli* Zellen wurden unter Verwendung des TransformAID-Kits der Firma MBI Fermentas (St. Leon-Rot) oder nach Sambrook et al. (1989) hergestellt.

Bei der Erstellung der Plasmidbibliotheken wurden *Epicurian Coli*<sup>®</sup> *SoloPack*<sup>™</sup> *supercompetent Cells* basierend auf dem *E. coli* XL10-gold Stamm der Firma Stratagene (Heidelberg) verwendet.

Für die Anzucht von  $\lambda$ ZAPII wurde *E. coli* XL1 Blue MRF<sup>1</sup> verwendet, für  $\lambda$ EMBL3 der P2-lysogene Stamm *E. coli* XL1 Blue MRA (P2).

### 2.1.11. Kultur-Medien

#### 2.1.11.1. Medien für *E. coli*

LB-Medium (pro l):	10 g Bacto-Trypton
	5 g Bacto-Hefeextrakt
	10 g NaCl

Für die Herstellung von LB-Platten wurden 15 g Bacto-Agar pro Liter LB-Medium eingewogen. Für selektive Ampicillin-Platten bzw. -Medien wurde nach dem Autoklavieren bei einer Temperatur von  $< 50^{\circ}\text{C}$  Ampicillin auf eine Endkonzentration von  $50\ \mu\text{g/ml}$  zum LB-Medium bzw. LB-Platten gegeben.

Für selektive X-Gal-IPTG-Ampicillin-Platten wurde zusätzlich X-Gal und IPTG auf eine Endkonzentration von  $40\ \mu\text{g/ml}$  zugegeben.

#### 2.1.11.2. Kultur-Medium f/2 für die Algen

(Guillard and Ryther, 1962; Guillard, 1975)

##### f/2-Grundlösung

Substanz	MW	mmol/l	g/l
Meersalz ( <i>Tropic Marine</i> )			30,0
$\text{NaNO}_3$	85,0	0,883	0,075
$\text{NaH}_2\text{PO}_4 \times 2\ \text{H}_2\text{O}$	156,0	0,036	0,006
Tris	121,1	5,0	0,602

Der pH-Wert wird mit 5 N HCl auf 7,2 eingestellt und das Medium anschließend autoklaviert.

##### f/2 Spurenelement-Stammlösung

Substanz	MW	mmol/l
$\text{FeCl}_3 \times 6\ \text{H}_2\text{O}$	270,30	11,65
$\text{Na}_2\text{EDTA}$	372,30	11,71
$\text{CuSO}_4 \times 5\ \text{H}_2\text{O}$	249,68	0,039
$\text{ZnSO}_4 \times 7\ \text{H}_2\text{O}$	287,54	0,077
$\text{CoCl}_2 \times 6\ \text{H}_2\text{O}$	237,93	0,042
$\text{MnCl}_2 \times 4\ \text{H}_2\text{O}$	197,91	0,91
$\text{Na}_2\text{MoO}_4 \times 2\ \text{H}_2\text{O}$	241,90	0,026

Jede Substanz muss gelöst sein, bevor die nächste addiert wird. Nach dem Autoklavieren wird die Lösung lichtgeschützt bei 4 °C aufbewahrt. Der pH-Wert beträgt etwa 2,0.

#### **f/2 Vitamin-Stammlösung**

<u>Substanz:</u>	<u>MW</u>	<u>mmol/l</u>
Biotin	244,31	0,0020
Cyanocobalamin	135,4	0,00037
Thiamin-HCl	337,27	0,297

Die Vitamin-Stammlösung wird sterilfiltriert und in aliquoten Mengen bei -20 °C gelagert.

### **2.1.12. Geräte**

#### **PCR-Thermocycler:**

MasterCyclerGradient (Eppendorf, Hamburg)

Biomed Thermocycler (Bachofer, Reutlingen)

Primus gold (MWG, Ebersberg)

#### **Sequenzierautomaten:**

ALFexpress (Amersham Pharmacia, Freiburg)

LI-COR 4200 (MWG, Ebersberg)

#### **Zentrifugen:**

RC-5B Refrigerated Superspeed (Sorvall, Bad Homburg)

GP Centrifuge (Beckman)

Sepatech Biofuge 17RS (Heraeus Instruments, Hanau)

Labofuge 400R (Heraeus Instruments, Hanau)

Centrifuge 5417R und 5417C (Eppendorf, Hamburg)

#### **Sonstige Geräte:**

GelDoc 1000 (BioRad, München)

Land Camera MP4 (Polaroid, Offenbach)

Repro-Anlage Leitz Focomat IIc (Leica)

Spektralphotometer Ultrospec III (Pharmacia, Freiburg)

UV-Lampe MinUVIS 254 nm/366 nm (DESAGA, Heidelberg)

BioRAD GelDoc (BioRad, München)

Pulsfeld Gelelektrophoresesystem CHEF DRIII (BioRad, München)

Transmissions Elektronenmikroskop EM301 (Philips, Hamburg)

Mikrotom Ultracut (Reichert-Jung, Wien)

Expositionskassetten (Appligene, Heidelberg)

Hybridisierungsöfen (Bachofer, Reutlingen) und OV1 (Biometra, Göttingen)

### 2.1.13. Soft- und Hardware

Es wurde an IBM kompatiblen PC's mit Standart Software gearbeitet. Das Computerprogramm „Sequencher“ (PC-Version 4.05; GeneCodes, USA) wurde eingesetzt, um genomische Sequenzen (Rohdaten) zu alignen, Korrektur zu lesen und um Restriktionsanalysen durchzuführen. Des weiteren diente „Sequencher“ zur Darstellung der Sequenzen. Zur Berechnung der Codon-Nutzung wurden *DNATools* Versio 5.1 revision 6.44 verwendet. Dieses steht im Internet frei zum Herunterladen bereit (<http://www.crc.dk/phys/dnatools.htm>).

Über das Internet wurde weiterhin auf folgende Programme und Dienste zugegriffen:

BlastX und BlastN (Altschul et al., 1997)	<a href="http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST">http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST</a>
FastA (Pearson und Lipman, 1988)	<a href="http://www.fasta.genome.ad.jp/">http://www.fasta.genome.ad.jp/</a>
tRNAscan-SE (Lowe und Eddy, 1997)	<a href="http://genome.wustl.edu/eddy/tRNAscan-SE">http://genome.wustl.edu/eddy/tRNAscan-SE</a>
Pol3Scan (Pavesi et al., 1994)	<a href="http://irisbioc.bio.unipr.it/pol3scan.html">http://irisbioc.bio.unipr.it/pol3scan.html</a>
ORF-Finder	<a href="http://www.ncbi.nlm.nih.gov/gorf/gorf.html">http://www.ncbi.nlm.nih.gov/gorf/gorf.html</a>
virtual Genomes	<a href="http://alces.med.umn.edu/VGC.html">http://alces.med.umn.edu/VGC.html</a>
CyanoBase	<a href="http://www.kazusa.or.jp:8080/cyano/index.html">http://www.kazusa.or.jp:8080/cyano/index.html</a>
PSORT	<a href="http://psort.nibb.ac.jp:8800/">http://psort.nibb.ac.jp:8800/</a>
SOSUI	<a href="http://azusa.proteome.bio.tuat.ac.jp/sosui/">http://azusa.proteome.bio.tuat.ac.jp/sosui/</a>
smallRNA database	<a href="http://mbr.bcm.tmc.edu/smallRNA/">http://mbr.bcm.tmc.edu/smallRNA/</a>

Weiterhin wurde im Rahmen der internationalen Zusammenarbeit innerhalb des Nucleomorph Genomprojektes eine Passwort-geschützte Datenbank zum Zwecke der Zusammenführung der Sequenzierdaten aufgebaut. Diese Datenbank diente auch dem Transfer der Rohdaten zwischen den beteiligten Arbeitsgruppen. Für eine Schnellanalyse der Sequenzdaten wurde das „*Magpie project*“ Nucleomorph (<http://niji.imb.nrc.ca/magpie/nucleomorph/private/>) genutzt.

## 2.2. Methoden

### 2.2.1. Algenkultur und Ernte

*Guillardia theta* wurde in Meersalzlösung f/2 (Kap. 2.1.11.2.) bei 14°C mit einem 12 h Dunkel-/12 h Hell-Rhythmus kultiviert. Zu diesem Zweck wurden jeweils 1000 ml der Algenkultur in 2 l-Erlenmeyerkolben mit 'warm white' Fluoreszenzlampen beleuchtet. Das Animpfen erfolgte durch Zugabe von 100 ml Stammkultur zu 900 ml f/2-Medium. Die Stammkulturen von *Guillardia theta* wurden wöchentlich überimpft, indem 20 ml Stammkultur zu 80 ml frischem Medium gegeben wurden. Alle Maßnahmen der Kultivierung erfolgten unter sterilen Bedingungen.

Nach zwei bis drei Wochen (in der späten log-Phase) wurden die Algen durch Zentrifugation bei 3500 g (10 min, 14°C) in einer Beckman GP Centrifuge geerntet. Das Algen-Pellet wurde einmal in f/2-Medium gewaschen und anschließend weiterverarbeitet

Zur Bestimmung der Zell-Dichte wurde die Menge an Chlorophyll *a* in einem Aliquot (3 ml) der Algen-Kultur nach der Methode von Jeffrey und Humphrey (1975) bestimmt. Bei *Guillardia theta* ist 1 µg Chlorophyll *a* äquivalent zu  $2 \times 10^5$  Zellen (Hansmann und Eschbach, 1990).

### 2.2.2. Einbetten der Algenzellen in Agarose

Nach Eschbach et al. (1991). Intakte hochmolekulare chromosomale DNA für die Pulsfeld-Gelelektrophorese wurde gewonnen, indem ganze Algen-Zellen in *low-melting-point* Agarose eingebettet, und anschließend die DNA-Moleküle durch Detergenzien und Enzyme von Proteinen befreit wurden.

Für die Einbettung und DNA-Präparation wurden folgende, nicht in Kapitel 2.1.2. aufgeführte Lösungen benötigt:

Lösung 1:	10	mM	Tris/HCl, pH 8,0
	100	mM	Na <sub>2</sub> EDTA
	100	mM	NaCl

Lösung 2:	1	% (w/v)	LMP-Agarose
	2	mg/ml	Proteinase K in Lösung 1

Lösung 3:	10	mM	Tris/HCl, pH 8,0
	400	mM	Na <sub>2</sub> EDTA, pH 8,0
	1	%(w/v)	SLS
	0,25	mg/ml	Pronase
Lösung 4:	10	mM	Tris/HCl, pH 8,0
	400	mM	Na <sub>2</sub> EDTA, pH 8,0
	1	%(w/v)	SLS
	1	mg/ml	Pronase
Lösung 5:	10	mM	Tris/HCl, pH 8,0
	400	mM	Na <sub>2</sub> EDTA

Lösungen 3 und 4 wurden vor dem Gebrauch 30 Minuten bei 50 °C präinkubiert.

*G. theta* wurde gemäß Kapitel 2.2.1. geerntet, in Lösung 1 zu einer resultierenden Zelldichte von  $2 \times 10^8$ /ml resuspendiert und bei 37 °C sofort mit einem Volumen Lösung 2 vermischt. Aliquote Mengen von je 150 µl dieser Suspension wurden in Giesformen (je 10 x 8 x 2 mm) verteilt und fünf Minuten bei Raumtemperatur inkubiert, wobei die LMP-Agarose erstarrt. Die sogenannten "Inserts" wurden nach dem Herauslösen aus der Giesform zehn Minuten bei 50 °C in Lösung 3 inkubiert. Darauf schloss sich eine 48-stündige Inkubation in Lösung 4 an (50 °C), wobei die Lösung nach 24 h erneuert wurde. Die Inserts wurden anschließend in der Pulsfeld-Gelelektrophorese (Kap. 2.2.3.) zur Separation der Nucleomorph Chromosomen weiterverwendet. Alternativ konnte die so präparierte, hochmolekulare DNA mehrere Wochen in Lösung 5 bei 4 °C aufbewahrt werden.

### 2.2.3. Pulsfeld-Gelelektrophorese (PFGE) der Nucleomorph Chromosomen

Nach Eschbach et al., (1991). Nachdem die Algen-Inserts (Kap. 2.2.2.) zweimal je 15 Minuten in 0.5 x TBE-Puffer äquilibriert worden waren, wurden sie vorsichtig in die Geltaschen eines 1 % (w/v) Agarosegel (0.5 x TBE) überführt und mit flüssiger LMP-Agarose fixiert. Anschließend wurden die Nucleomorph Chromosomen in 0.5 x TBE-Laufpuffer mittels CHEF-Elektrophorese (CHEF DRIII, BioRAD) aufgetrennt. Die im

folgenden angegebenen Laufbedingungen erwiesen sich für die Trennung der drei Chromosomen als optimal. Die Elektrophorese wurde bei einer konstanten Laufpuffertemperatur von 14 °C, einem Winkel von 120° und bei einer Spannung von 5.4 V/cm in zwei Phasen durchgeführt. In der ersten Phase (25h) betrug die Pulsfrequenz konstant 15 s, in der zweiten 17-stündigen Phase wurde ein linearer Gradient der Pulsfrequenzen von 15 s bis 6 s eingestellt.

Nach der Elektrophorese wurde das Gel in 0.5 µg/ml Ethidiumbromidlösung 30 Minuten gefärbt, danach zur Kontrastierung für 30 Minuten in demineralisiertem Wasser entfärbt und anschließend dokumentiert.

Für die angereicherte Isolation der Nucleomorph Chromosomen wurde das Pulsfeld-Gel in der Weise präpariert, dass die aufgetrennten Chromosomen nach der Elektrophorese in LMP-Agarose zu liegen kamen. Diese wurden anschließend an die Ethidiumbromidfärbung aus dem LMP-Agarose mit einem sterilen Skalpell ausgeschnitten.

#### **2.2.4. Generelle molekularbiologische Methoden**

Grundlegende molekularbiologische Methoden wurden, wenn nicht explizit angegeben, nach Sambrook et al. (1989) durchgeführt.

##### **2.2.4.1. Plasmid isolation aus *E. coli***

Für die Isolierung von Plasmiden aus den verwendeten *E. coli*-Stämmen wurden die Methoden der Alkalischen Lyse (Birnboim und Doly, 1979) oder der *boiling preparation* (Holmes und Quigley, 1981) genutzt. Für die Isolierung hochreiner Plasmid-DNA wurden Kits der Firmen Macherey & Nagel (Düren) und Qiagen (Hilden) nach Angaben der Hersteller verwendet.

##### **2.2.4.2. Isolierung genomischer DNA aus *G. theta***

Die Präparation von DNA aus Cryptomonaden wurde durch die CTAB-Methode (Doyle und Doyle, 1990) bewerkstelligt. Auf die Verwendung von Glaskügelchen beim Zellaufschluss wurde verzichtet. Gesamt-DNA Präparationen wurden bei 4 °C aufbewahrt.

### 2.2.4.3. Isolierung von DNA aus Bakteriophagen $\lambda$

In einem 1,5 ml-Reaktionsgefäß wurden *E. coli* MRA(P2) Zellen (Kap. 2.1.10.) und  $\lambda$ EMBL3 Bakteriophagen gemischt und 20 Minuten bei 37 °C präinkubiert. Das optimale Verhältnis der Titer von Phagen und Zellen wurde empirisch ermittelt. Die infizierten Zellen wurden zu 5 ml LB in ein steriles Reagenzglas gegeben und unter Schütteln über Nacht angezogen. Zur Präparation der Phagen-DNA wurden die Minilysate wie beschrieben aufgeschlossen.

Rekombinante  $\lambda$ ZAPII-DNA wurde durch *in vitro*-Excision unter Verwendung des VCSM13 Helfer-Phagen gewonnen. Die Prozeduren sind vom Hersteller dieses Systems, Stratagene (Heidelberg), genau beschrieben.

### 2.2.5. Herstellung Nucleomorph Chromosom II angereicherter Plasmidbibliotheken

Pro Ansatz wurden je zwei der aus einem Pulsfeld-Gel ausgeschnittenen LMP-Agarose Blöcke (Kap. 2.2.3.) des Chromosoms II vereinigt. Die LMP-Agarose wurde bei 65°C geschmolzen, nach Abkühlen auf 45°C wurde 1U AgarACE (Promega, Mannheim) pro 100mg LMP-Agarose zugegeben und zwei Stunden bei 45°C inkubiert. Anschließend wurde mit den spezifischen Puffern für die entsprechenden Restriktionsendonucleasen aufgesalzt und mit 10U Enzym über Nacht bei 37°C restringiert. Im folgenden wurde die DNA gefällt, gewaschen, getrocknet und in ein entsprechend geöffnetes und dephosphoryliertes pUC19 Plasmid (MBI Fermentas, St. Leon-Rot und Appligene, Heidelberg) einligiert. Nach Transformation in kompetente *E. coli* MRF'-Zellen und anschließender Isolierung der rekombinanten Plasmide durch in Kapitel 2.2.4.1. beschriebene Methoden, wurden die Amplifikate sequenziert (Kap. 2.2.7.).

### 2.2.6. Polymerase-Kettenreaktion (PCR)

Für die Polymerase-Kettenreaktion wurden Enzyme und zugehörige Puffersysteme von Eppendorf (Hamburg), Boehringer (Mannheim), MBI Fermentas (St. Leon-Rot) oder Amersham Pharmacia (Freiburg) verwendet. Annealingtemperaturen, Elongationszeiten und MgCl<sub>2</sub>-Konzentrationen wurden nach Anforderungen der Primer, der eingesetzten DNA und der Größe des zu erwartenden Amplifikats verändert.

Ein Standard PCR-Ansatz wurde in einem Endvolumen von 25  $\mu$ l mit Endkonzentrationen von 1,5-7 mM MgCl<sub>2</sub>, 1 x Reaktionspuffer, je 0,25  $\mu$ M Primer und 1 U *Taq*-DNA-Polymerase durchgeführt. Es wurde eine DNA-Menge von 100 ng bis 250 ng *G. theta*-DNA eingesetzt.

PCR-Produkte wurden durch Fällung oder Elution aus Agarosegelen (*JETSORB*, GENOMED, Bad Oeynhausen) gereinigt, in das Plasmid pGEM-T einligiert (Promega, Mannheim) und anschließend in kompetente *E. coli* MRF'-Zellen transformiert. Alternativ wurde der 'InsT/Aclone™ PCR Product Cloning Kit' der Firma MBI Fermentas (St. Leon-Rot) verwendet. Es folgte die Isolierung der rekombinanten Plasmide durch in Kapitel 2.2.4.1. beschriebene Methoden, anschließend wurden die Amplifikate sequenziert (Kap. 2.2.7.).

### 2.2.7. DNA-Sequenzierung

In Sequenziervektoren klonierte DNA wurde mit dem Kit „*Thermosequenasae fluorescent labelled primer cycle sequencing Kit with 7-deaza-dGTP*“ der Firma Amersham Pharmacia (Freiburg) und Fluoreszenz gelabelten Oligonucleotiden (MWG, Ebersberg) sequenziert.

#### 2.2.7.1. Das ALFexpress System

Für den ALFexpress (Amersham Pharmacia, Freiburg) wurden CY5-markierte Oligonucleotide verwendet. Standardmäßig wurde pro Sequenzierreaktion 1  $\mu$ g Plasmid-DNA (Kap. 2.2.4.1.) mit destilliertem Wasser auf ein Volumen von 21  $\mu$ l verdünnt und mit 4  $\mu$ l Sequenzierprimer (2,5  $\mu$ M) versetzt. Je 6  $\mu$ l dieses Gemisches wurden auf vier PCR-Gefäße verteilt, in denen je 2  $\mu$ l der einzelnen Nucleotid-spezifischen Terminationslösungen vorgelegt waren. Die Reaktionslösungen wurden mit Öl überschichtet und 25 Zyklen aufeinanderfolgender Denaturierungs- (45 s), Annealings- (45 s) und Extensionsphasen (60 s) unterzogen. Nach der Sequenzierreaktion wurde jedem PCR-Gefäß 4  $\mu$ l formamidhaltigen Ladepuffers (*formamide loading dye*, Amersham Pharmacia, Freiburg) zugegeben.

### 2.2.7.2. Das LI-COR System

Für den LI-COR-Sequenzierautomaten (MWG, Ebersberg) wurden IRD700- bzw. IRD800-markierte Sequenzierprimer verwendet. Pro 1000 bp zu sequenzierender DNA (einschl. Plasmid) wurden 150 ng eingesetzt, mit destilliertem Wasser auf 11  $\mu$ l verdünnt und mit 1  $\mu$ l IRD700-markiertem Sequenzierprimer (2  $\mu$ M), 1  $\mu$ l IRD800-markiertem Sequenzierprimer (4  $\mu$ M) und 1  $\mu$ l DMSO versetzt. Je 3  $\mu$ l dieses Ansatzes wurden auf vier PCR-Gefäße verteilt, in denen je 1  $\mu$ l der einzelnen Nucleotid-spezifischen Terminationslösungen vorgelegt waren. Die Reaktionslösungen wurden mit Öl überschichtet und 25 Zyklen aufeinanderfolgender Denaturierungs- (30 s), Annealings- (30 s) und Extensionsphasen (30 s) unterzogen. Nach der Sequenzierreaktion wurde jedem PCR-Gefäß 4  $\mu$ l eines formamidhaltigen Ladepuffers (*formamide loading dye*, Amersham Pharmacia, Freiburg) zugesetzt.

### 2.2.8. Analyse fluoreszenzmarkierter DNA-Stränge durch die Sequenzierautomaten

Die Analyse der in Kapitel 2.2.7. erstellten Sequenzierreaktionen erfolgte auf dem LI-COR4200 Sequenzierautomaten mit Hilfe eines Acrylamid-Gels basierend auf einer 40 %igen *Rapidgel XL* Stocklösung der Firma Amersham Pharmacia (Freiburg). Die Gelmatrix für die Separation der DNA-Fragmente setzte sich aus 12,6 g Harnstoff; 3,5 ml *Rapid Gel XL* (40 %; Amersham Pharmacia, Freiburg) und 3 ml 10 x TBE-*Long-run*-Puffer (1,34 M Tris/HCL pH 8,8; 0,45 M Borsäure; 25 mM EDTA) gelöst in 18 ml ddH<sub>2</sub>O zusammen. Nach dem Entgasen, wurden 210  $\mu$ l APS (10 %) und 45  $\mu$ l TEMED zugegeben und das 0,2 mm dicke Gel im Gelgiesstand der Firma MWG (Ebersberg) nach deren Anleitung gegossen.

DNA-Fragmente, die auf dem ALFexpress-Sequenzierautomaten analysiert werden sollten, wurden über ein Long Ranger-Gel (FMC Bioproducts, Rockland USA) separiert. Zur Herstellung der Gelmatrix wurden 21,7 g Harnstoff mit 6,6 ml Long Ranger-Gel und 9 ml 10 x TBE-Puffer versetzt und mit destilliertem Wasser auf 60 ml aufgefüllt. Nach Entgasen der Matrix wurden 300  $\mu$ l 10 % APS und 30  $\mu$ l TEMED zugegeben, der Ansatz gemischt und das Gel nach Angaben der Firma Amersham Pharmacia (Freiburg) gegossen. Der Betrieb der Sequenzierautomaten erfolgte entsprechend den Anweisungen des Herstellers.

### 2.2.9. Alignment der Sequenzen

Alle Alignments wurden mit dem Computerprogramm ‚Sequencher‘ der Firma GeneCodes (USA) durchgeführt. Mit diesem Programm können die *output-files* aller gängigen Sequenzierautomaten geladen und ein Alignment erstellt werden. Parallel zum Alignment können die Chromatogramme der einzelnen Sequenzen angezeigt werden, so dass es möglich ist, die Sequenzen im Alignment Korrektur zu lesen. Diese Alignments dienen auch der Auffindung einzelner Restriktionsschnittstellen.

### 2.2.10. Southern-Transfer von hochmolekularen Nukleinsäuren

Elektrophoretisch separierte *Guillardia theta*-DNA wurde durch Southern-Transfer (Southern, 1975) auf positiven Nylon-Membranen immobilisiert. Durch CTAB-Extraktion gewonnene DNA (Kap. 2.2.2.2.) wurde durch 0.25 N HCl depuriniert (Sambrook et al., 1989), in durch Pulsfeld-Gelelektrophorese getrennte Nucleomorph Chromosomen (Kap. 2.2.3.) wurden durch UV-Bestrahlung (240 nm, 30 sec) Einzelstrangbrüche eingeführt. Die weitere Behandlung der Agarosegele erfolgte nach Sambrook et al. (1989). Als Transferpuffer wurde 20 x SSC verwendet. Die irreversible Bindung der DNA an die Membran erfolgte durch einstündige Hitzefixierung bei 80 °C.

### 2.2.11. Markierung von DNA

Die Markierung der Sonden erfolgte nach der *random priming* Methode von Feinberg und Vogelstein (1983). Zur Markierung der Sonden verwendet wurde DIG-11-dUTP (Boehringer, Mannheim).

Zur DIG-Markierung der DNA wurden 4 µl 5 x *random priming*-Puffer (Kap. 2.1.2.), 200-500 ng Sonden-DNA und 1 µl randomisierter Hexanucleotide ( $A_{260}=1$ ) in einem Endvolumen von 17 µl gemischt, zwei Minuten bei 100 °C denaturiert und auf Raumtemperatur abgekühlt. Anschließend wurden 2 µl 10 x *DIG-labelling mix* (Boehringer, Mannheim) und 2 U Klenow-Enzym zugegeben und 30 Minuten bei 37 °C inkubiert. Zum Entfernen nicht eingebauter Nucleotide wurde die DNA gefällt, gewaschen und in Wasser aufgenommen. Die Konzentration der Sonden-DNA in der Hybridisierungslösung wurde auf etwa 25 ng/ml eingestellt.

### 2.2.12. Hybridisierung von Nukleinsäuren

Das *screening* in der vorhandenen cDNA-Bibliothek und der  $\lambda$ EMBL3 gesamt-DNA Bibliothek (Fraunholz, 1999) erfolgte nach Sambrook et al. (1989). Plaque-Lifts auf Nylon-Membranen (Hybond N<sup>+</sup>, Amersham Pharmacia, Freiburg) wurden zur Detektion mit den Digoxigenin-markierten Sonden verwendet. Die Detektion von DNA auf Plaque-Lifts und auf Southern-Blots erfolgte nach folgendem Schema: Die Hybridisierung wurde über Nacht bei Sonden-spezifischer Temperatur durchgeführt. Nach dem Abgießen der Sondenflüssigkeit wurde durch zweimaliges Waschen mit 2 x SSC/0,1 % SDS nicht gebundene Sonden-DNA entfernt. Die Digoxigenin-markierten Sonden wurden durch mit Alkalischer-Phosphatase konjugierten anti-Digoxigenin F(ab)<sub>2</sub>-Fragmenten und dem Chemilumineszenzsubstrat CSPD nach dem von Boehringer (Mannheim) beschriebenen Protokoll auf Röntgenfilmen (Fuji) detektiert. Für die Wiederverwendung der Membranen wurde die Digoxigenin-markierte Sonden-DNA nach der Anweisung im "*DIG-Labeling and Detection Manual*" (Boehringer, Mannheim) durch Alkali-Behandlung (2 x 15 min bei 37 °C mit 0,2 N NaOH, 1 % SDS) entfernt. Die Membranen wurden anschließend in 2 x SSC gewaschen, feucht in Haushaltsfolie eingeschweißt und aufbewahrt. Diese Membranen standen somit für erneute Detektionen zu Verfügung.

### 3. ERGEBNISSE

Ziel dieser Arbeit war es, ausgehend von den bereits bekannten subtelomerischen Bereichen (Fraunholz, 1999), das Nucleomorph Chromosom II der Cryptomonade *Guillardia theta* (im weiteren als Chromosom II oder Chr II bezeichnet) im Rahmen eines internationalen Genomprojektes, vollständig zu sequenzieren und anschließend strukturell sowie funktionell zu charakterisieren.

Bezüglich der Sequenzierungs- und Klonierungsstrategie wurden drei Varianten verfolgt:

- 1) Screening einer genomischen  $\lambda$ EMBL3 Bibliothek (*chromosome walking*),
- 2) Erstellung mehrerer, voneinander unabhängiger Plasmidbibliotheken (*shotgun* Sequenzierung),
- 3) spezifische Amplifikation mittels PCR.

Die strukturelle und funktionelle Analyse wurde mit Hilfe verschiedener Computerprogramme „*in silico*“ durchgeführt.

Ein weiterer Teil dieser Arbeit bestand in der Klonierung und Sequenzierung letzter Lücken in der Sequenz und die Doppelstrang-Sequenzierung der Nucleomorph Chromosomen I und III von *Guillardia theta* (Chromosom I bzw. III). Hierzu wurden die in den Arbeitsgruppen von Susan Douglas und Tom Cavalier-Smith ermittelten Alignments und Konsensussequenzen der Chromosomen I und III herangezogen und, analog zu den Arbeiten an Chromosom II, durch spezifische Amplifikation mittels PCR ihre relative Lage zueinander bestimmt.

#### 3.1. Die Klonierungsstrategie

##### 3.1.1. Das Alignment der Sequenzen von Chromosom II

Alle Sequenzen der rekombinanten Klone mit Chromosom II spezifischen Insertionen wurden aktuell in das Alignment (Kap. 2.2.9.) eingerechnet, um auf diese Weise sich überlappende Sequenzen der Klone unterschiedlicher Plasmidbibliotheken und rekombinanter  $\lambda$ EMBL3 Bakteriophagen zu erhalten (Abb. 3.1b.). Es wurde jeweils im Einzelfall entschieden, welcher Klon durch Subklonierung und anschließende Plasmidisolation (Kap. 2.2.4.1.) vollständig sequenziert werden sollte.

### 3.1.2. Arbeiten mit einer genomischen $\lambda$ EMBL3 Bibliothek

Um rekombinante Bakteriophagen mit neuer Sequenzinformation zu erhalten, wurden von den nichttelomerischen Enden der bekannten Sequenzen der invertierten Sequenzwiederholungen mit den flankierenden Bereichen (Fraunholz, 1999) Sonden abgeleitet (Kap. 2.2.11.), um mit diesen eine vorhandene genomische  $\lambda$ EMBL3 Bibliothek zu *screenen* (Kap. 2.2.12.). Die Inserts detektierter rekombinanter Phagen wurden sequenziert und auf diese Weise ein *chromosome walking* durchgeführt. Es konnten ausschließlich rekombinante Bakteriophagen isoliert werden, deren Inserts subtelomerisch lokalisiert waren und nur zu einem sehr geringen Anteil über die bereits bekannte Sequenz hinaus reichten (Daten nicht gezeigt). Allen gemeinsam war das Vorhandensein großer Teile der rRNA Transkriptionseinheit.

Durch den alternativen Ansatz der direkten Sequenzierung der Enden rekombinanter  $\lambda$ EMBL3 Bakteriophageninserts (in der Arbeitsgruppe von Tom Cavalier-Smith) konnte ein Chromosom II-spezifischer, rekombinanter Bakteriophage (LT89) isoliert werden. Dieser rekombinante Bakteriophage wurde im Rahmen dieser Arbeit vollständig sequenziert. Die vollständige Sequenzierung des Inserts von LT89 ergab 12028 bp kontinuierliche Sequenzinformation. Anhand dieser Sequenzinformationen konnten einige Klone der Plasmidbibliotheken (Kap. 3.1.3.) in ihrer räumlichen Lage zueinander bestätigt werden. Die relative Lage bezüglich der Gesamtsequenz von Chr II ist der Abbildung 3.1a. zu entnehmen.

### 3.1.3. Plasmidbibliotheken

Im Rahmen dieser Arbeit wurden insgesamt vier mit Nucleomorph Chromosom II DNA angereicherte Plasmidbibliotheken mit den Restriktionsendonucleasen *Bam*HI, *Eco*RI, *Spe*I und *Xba*I gemäß Kapitel 2.2.6. erstellt.

Die Auswahl der Restriktionsendonucleasen erfolgte unter der Prämisse, dass die nach der Restriktion entstehenden Fragmente größer als 300 bp sein sollten. Die obere Grenze setzte die Klonierungskapazität des verwendeten pUC19 Plasmid Vektors. Aus diesem Grund wurden nur Restriktionsendonucleasen mit einer hexameren Erkennungssequenz in Betracht gezogen. Restriktionskarten der bereits bekannten Bereiche der *single-copy* Region (Fraunholz, 1999; Zauner et al., 2000) ergaben, dass, aufgrund des hohen A/T-Gehaltes dieser Regionen, Restriktionsenzyme deren Erkennungssequenz ausschließlich

aus Adenin und Thymin aufgebaut sind, sehr häufig schneiden und somit die erhaltenen Fragmente statistisch gesehen sehr kurz sind. Restriktionsendonucleasen mit einer ausschließlich aus Guanin und Cytosin aufgebauten Erkennungssequenz wurden aufgrund der zu erwartenden sehr großen Fragmente nicht benutzt, da es sehr schwierig wäre, diese in den pUC19 Plasmid Vektor zu klonieren. Die Restriktionsendonucleasen *EcoRI*, *SpeI* und *XbaI* zeichnen sich durch je zwei Adenine und Thymine in ihrer Erkennungssequenz aus und sollten somit hauptsächlich Fragmente liefern, die in ihrer Größe den oben gestellten Bedingungen entsprechen. Zur Generierung größerer Fragmente wurde die Restriktionsendonuclease *BamHI* eingesetzt. Die Erkennungssequenz des Restriktionsenzym *BamHI* besteht aus dem Palindrom GGATCC. Aufgrund des hohen A/T-Gehaltes der bekannten Bereiche der *single-copy* Region sind *BamHI* Erkennungssequenzen statistisch betrachtet seltener zu erwarten und somit die Fragmente größer.

Des Weiteren wurde im Rahmen der Zusammenarbeit innerhalb des Nucleomorph Genomprojektes zwei mit Nucleomorph-DNA (Chromosomen I, II und III) angereicherte Plasmidbibliotheken mit Fragmenten der Restriktionsendonucleasen *HindIII* und *BglII* in der Arbeitsgruppe von Susan Douglas generiert und sequenziert. Die zur Herstellung dieser Plasmidbibliotheken eingesetzte DNA wurde mittels CsCl/Hoechst33258 Dichtegradienten-zentrifugation gewonnen (S. Douglas, pers. Mitteilung). Die Klone dieser Plasmidbibliotheken wurden durch Sequenzvergleiche den einzelnen Chromosomen zugeordnet.

Die Sequenzdaten der rekombinanten Klone aller Plasmidbibliotheken und  $\lambda$ EMBL3 Bakteriophagen wurden in einer speziell für diesen Zweck eingerichteten Datenbank verwaltet, miteinander verglichen und waren allen beteiligten Arbeitsgruppen zugänglich.

Größe, Anzahl und Lokalisation der aus den Plasmidbibliotheken hervorgegangenen rekombinanten Klone sind in Abbildung. 3.1b. und Tabelle 3.1. zusammengestellt.

	Plasmidbibliotheken					
	<i>Bam</i> HI	<i>Bgl</i> II	<i>Eco</i> RI	<i>Hind</i> III	<i>Spe</i> I	<i>Xba</i> I
Gesamtzahl der rekombinanten Klone	19	n.b.	92	n.b.	28	105
Anzahl rekombinanter Klone auf Chr II	15	56	79	39	19	100
Länge der klonierten Bereiche innerhalb der <i>single-copy</i> Region [bp]	24060	51146	64432	66321	34082	96984
% bezogen auf <i>single-copy</i> Region [%]	15,5	33	41,6	42,8	22	62,6

Tab. 3.1.: Zusammenfassung der verwendeten Plasmidbibliotheken und der Gesamtlängen der klonierten Bereiche rekombinanter Klone. n.b.: Die Gesamtzahl der Klone der Nucleomorph-DNA angereicherten Plasmidbibliotheken wurde im Rahmen dieser Arbeit nicht bestimmt. Näheres siehe Text.

### 3.1.3.1. Die *Xba*I Plasmidbibliothek

Insgesamt wurden 105 rekombinante Plasmide der *Xba*I Plasmidbibliothek beidseitig mit den Sequenzierprimern M13 *universal* und M13 *reverse* (Kap. 2.1.8.) sequenziert. Sequenzvergleiche erbrachten 100 Chromosom II spezifische Klone, wobei 54 Plasmide mit Insertgrößen zwischen 68 bp und 5795 bp weiter bearbeitet wurden. 46 Plasmide stellten in ihrer Sequenz bereits bekannte rekombinante Plasmide dar und wurden nicht weiter bearbeitet. Vier der 54 unabhängigen Klone entfielen auf die bereits bekannten subtelomerischen *inverted repeats*, wohingegen 50 Klone neue Sequenzabschnitte innerhalb der *single-copy* Region erschlossen. Parallel zur Arbeit an weiteren Plasmidbibliotheken wurden diese 50 rekombinanten Plasmide sukzessive vollständig sequenziert. Diese 50 rekombinanten Plasmide repräsentieren 96984 bp der *single-copy* Region, dies entspricht 62,6 % (Tab. 3.1. und Abb. 3.1b.).

Fünf rekombinante Plasmide der insgesamt 105 sequenzierten Plasmide ergaben Sequenzen, die durch Sequenzvergleiche nicht mit der Sequenz von Chromosom II zur Übereinstimmung zu bringen waren. Sie repräsentieren Abschnitte der Sequenzen des Chromosoms III (3 Klone); des Chromosoms I (1 Klon) und der nucleären DNA von *Guillardia theta* (1 Klon).

### 3.1.3.2. Die *EcoRI* Plasmidbibliothek

Die *EcoRI* Plasmidbibliothek erbrachte 92 rekombinante Klone, die von beiden Seiten sequenziert wurden. Es konnten 79 Klone ermittelt werden, die Chr II zuzurechnen sind, somit repräsentierten 13 der Klone nicht Teile des Chr II, sondern Abschnitte des Chromosoms III (5 Klone), der nucleären DNA von *Guillardia theta* (7 Klone) und des Plastoms (1 Klon).

Die 41 unterschiedlichen, rekombinanten Plasmide der Chr II spezifischen Klone wurden, soweit sich ihre Gesamtsequenz nicht bereits durch das Alignment bis dato bekannter Sequenzen ergab, vollständig sequenziert. Die Größe ihrer Inserts variierte von 44 bp bis 8223 bp. Sequenzübereinstimmungen mit den subtelomerischen *inverted repeats* konnte bei vier dieser Klone festgestellt werden, so dass 37 Klone Abschnitte innerhalb der *single-copy* Region des Chr II mit einer Gesamtlänge von 64432 bp (41,6 %) repräsentierten (Tab. 3.1. und Abb. 3.1b.).

### 3.1.3.3. Die *SpeI* Plasmidbibliothek

Es wurden 28 rekombinante Klone der *SpeI* Plasmidbibliothek beidseitig sequenziert. Sequenzvergleiche ergaben 19 Chromosom II spezifische Klone. Von zehn Klonen wurde die Gesamtsequenz durch das Alignment (Kap. 3.1.1.) mit bereits bekannten Sequenzen bzw. durch Subklonierung und anschließende Sequenzierung (Kap. 2.2.7.) ermittelt. Die Insertgröße variierte von 567 bp bis 6338 bp. Die klonierten Bereiche mit einer Länge von 34082 bp repräsentieren 22 % der *single-copy* Region (Tab. 3.1. und Abb. 3.1b.).

Acht Klone, die nicht Chr II zugeordnet werden konnten, repräsentieren Sequenzen der nucleären DNA von *Guillardia theta*, ein Klon konnte dem Plastom von *Guillardia theta* zugeordnet werden.

### 3.1.3.4. Die *BamHI* Plasmidbibliothek

Von der *BamHI* Plasmidbibliothek wurden 19 rekombinante Klone sequenziert. Es konnten 15 Klone ermittelt werden die Chr II zuzurechnen sind. Insgesamt wurden 11 unabhängige Klone, sofern ihre Gesamtsequenz nicht bereits durch das Alignment (Kap. 3.1.1.) bekannt war, vollständig sequenziert (Kap. 2.2.7.). Fünf dieser 11 unabhängigen Klone entfielen auf die bereits bekannten subtelomerischen *inverted repeats*, während sechs Klone Sequenzabschnitte innerhalb der *single-copy* Region erschlossen. Ihre Größe

umfasste eine Spanne von 696 bp bis 7059 bp. Sequenzübereinstimmung mit den subtelomerischen *inverted repeats* (Fraunholz, 1999) konnte bei vier dieser Klone beobachtet werden, so dass sieben Klone Abschnitte innerhalb der *single-copy* Region des Chr II mit einer Gesamtlänge von 20060 bp (15,5 %) repräsentierten (Tab. 3.1. und Abb. 3.1b.).

Die vier Klone die nicht Chr II zugeordnet werden konnten, repräsentierten Sequenzen aus dem nucleären Genom von *Guillardia theta*.

### 3.1.3.5. Die *Bgl*III und *Hind*III Plasmidbibliotheken

Dem Chromosom II konnten durch Sequenzvergleiche insgesamt 56 Klone der *Bgl*III Plasmidbibliothek und 39 Klone der *Hind*III Bibliothek zugeordnet werden, wobei die Anzahl der unterschiedlichen Klone bei 14 respektive 21 lag. Zwei Klone der *Bgl*III und keiner (trotz vorhandener Restriktionsschnittstellen) der *Hind*III Bibliothek repräsentierten Bereiche der subtelomerischen *inverted repeats* (Fraunholz, 1999). Die 12 Klone der *Bgl*III Bibliothek repräsentierten, bei einer Insertgröße zwischen 101 bp und 6506 bp, 51146 bp (33 %) der *single-copy* Region des Chr II. Die klonierten Bereiche der 21 rekombinanten Klone der *Hind*III Plasmidbibliothek variierte von 234 bp bis 6991 bp. Insgesamt repräsentierten diese 21 *Hind*III Klone 42,8 % (66321bp) der *single-copy* Region (Tab. 3.1. und Abb. 3.1b.).

### 3.1.3.6. PCR zur Überprüfung der Alignments und Schließen verbliebener Sequenzlücken

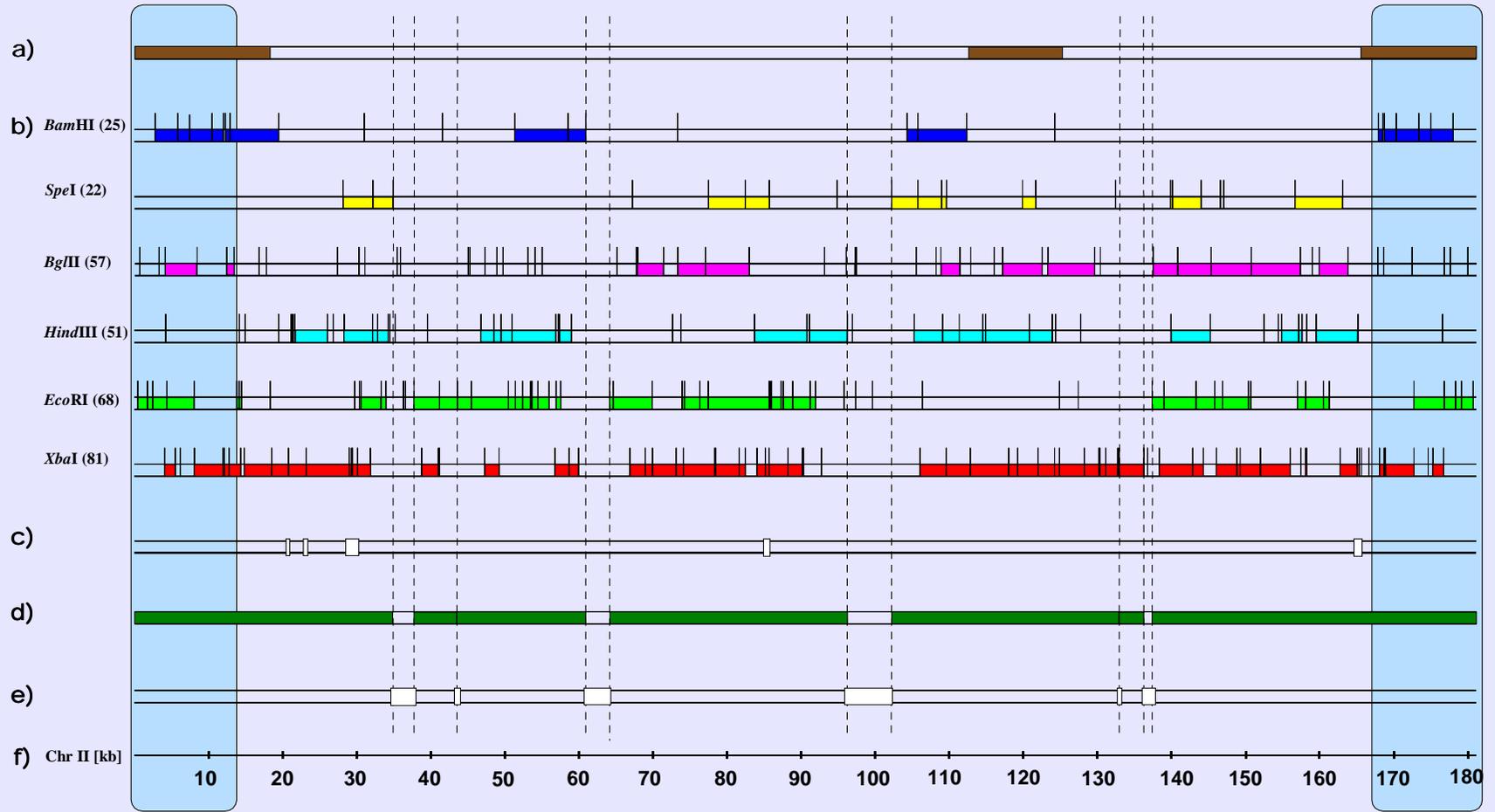
Nichtidentische Klone derselben Plasmidbibliothek, die Homologien zu den gleichen Datenbankeinträgen aufwiesen, wurden mittels PCR (Kap. 2.2.6.) bezüglich ihrer räumlichen Lage zueinander überprüft (Abb. 3.1c.). Hierzu wurde an den Enden der beiden Klone bzw. Konsensussequenzen Oligonucleotide abgeleitet, deren 5'-Ende näher zur Mitte der bekannten Sequenz liegt als deren 3'-Ende. Diese Oligonucleotide wurden in einer PCR als Primer eingesetzt, die entstandenen PCR-Produkte wurden kloniert (Kap. 2.2.6.) und nach der Isolierung der rekombinanten Plasmide (Kap. 2.2.4.1.), sequenziert (Kap. 2.2.7.). Um PCR-Artefakte ausschließen zu können, wurde von PCR-Produkten, deren Sequenz nicht durch einen Klon der Plasmidbibliotheken oder einen  $\lambda$ EMBL3 Bakteriophagen abgedeckt wurden, mindestens drei verschiedene Klone sequenziert. Auf diese Weise konnte sichergestellt werden, dass Fehler während der Inkorporation der

Nucleotide in den wachsenden DNA-Strang, verursacht durch die verwendeten Polymerasen, nicht in die Gesamtsequenz mit einfließen.

Mit zunehmender Anzahl sequenzierter Klone der Plasmidbibliotheken stieg die Anzahl der Klone mit bereits bekannten rekombinanten Plasmiden stark an, so dass, beim Stand des Alignments wie ihn Abbildung 3.1d. repräsentiert, keine weiteren Plasmidbibliotheken generiert wurden. Aus dem Alignment ergaben sich sieben Konsensussequenzen mit einer Gesamtlänge von 170842 bp, dies entspricht ca. 94,4 % bezogen auf die durch Pulsfeld-Gelelektrophorese (Kap. 2.2.3.) abgeschätzte Größe des Chr II von 180 kb (Zauner et al., 2000).

Abgeleitet von der Überprüfung der räumlichen Lage verschiedener Klone mit Homologien zu den gleichen Datenbankeinträgen durch PCR (siehe oben), wurde eine Klonierungsstrategie angewandt, die auf den Einsatz der PCR zum Schließen der verbleibenden Lücken setzte. Hierzu wurde, ausgehend von den sieben entstandenen Konsensussequenzen (Abb. 3.1d), terminal jeweils mit dem 3'-Ende nach außen gerichtete Oligonucleotide abgeleitet. Alle möglichen Kombinationen wurden unter unterschiedlichen PCR-Bedingungen getestet. Die PCR-Bedingungen variierten in der Länge der Elongation, der Annealingtemperatur und der MgCl<sub>2</sub>-Konzentration. Als Ausgangs-DNA wurde *G. theta* gesamt-DNA eingesetzt (Kap. 2.2.4.2.). Entstandene PCR-Produkte wurden in pGEM-T bzw. pUC57/T ligiert, in den *E. coli* Stamm XL1 blue MRF' transformiert und nach der Isolierung der Plasmid-DNA (Kap. 2.2.4.1.) sequenziert (Kap. 2.2.7.). Die erhaltenen Sequenzen wurden in das Alignment miteinbezogen und im positiven Fall (Verbindung zweier Konsensussequenzen) die Primer-Kombinationen neu angepasst. Auch hier wurden von allen positiven PCR-Produkten mindestens drei Klone sequenziert (siehe oben). Auf diese Weise wurden 11397 bp kloniert und sukzessive alle Contigs in die richtige räumliche Beziehung zueinander gebracht (Abb. 3.1e.). Es entstand eine 180915 bp lange Konsensussequenz (Abb. 3.1f.), die der Gesamtsequenz des Nucleomorph Chromosoms II entspricht.

### Übersicht der Klonierungsstrategie



### Vorangegangene Seite

Abb. 3.1.: Schematische Darstellung der durch die Klonierungsstrategie erhaltenen Bereiche des Nucleomorph Chromosoms II aus *Guillardia theta* und deren relative Lage zueinander. Blau unterlegt sind die subtelomerischen *inverted repeats*, die die *single-copy* Region flankieren. **a)** Regionen (braun) abgedeckt durch Bakteriophagen einer genomischen  $\lambda$ -EMBL3 Bibliothek; **b)** relative Lage der Klone der sechs Plasmidbibliotheken untereinander und bzgl. Chromosom II. Die Zahlen in Klammern geben die Anzahl der jeweiligen Schnittstellen an. Die Schnittstellen selbst sind durch kurze senkrechte Markierungen gekennzeichnet; **c)** spezifische Amplifikationen mittels PCR (weiß) zur Absicherung kritischer Regionen des Alignments; **d)** nach Abschluss des Alignments der Klone aus a)-c) entstandene Konsensussequenzen; **e)** spezifische Amplifikationen mittels PCR (weiß) zum Lückenschluss und zur räumlichen Orientierung der Konsensussequenzen aus d); **f)** *G. theta* Nucleomorph Chromosom II skaliert in 10 kb Schritten. Näheres siehe Text.

## 3.2. Die Sequenz des Nucleomorph Chromosoms II

Das Alignment aller Nucleomorph Chromosom II spezifischer Sequenzen ergab eine 180915 bp lange Konsensussequenz, diese stellt die Sequenz des Nucleomorph Chromosoms II der Cryptomonade *Guillardia theta* dar (Abb. 3.1f.). Das entstandene Alignment setzt sich aus 1603 Einzelsequenzen mit einer Gesamtlänge von 959477 bp zusammen (in diesen Angaben sind die Sequenzen der Bakteriophagen der subtelomerischen *inverted repeats* (Fraunholz, 1999) nicht berücksichtigt). Dies entspricht der 5,3-fachen Länge des Chromosoms II. Die Sequenz ist zu ca. 90 % doppelsträngig sequenziert, wobei die verbleibenden ca. 10 % durch mindestens drei verschiedene Klone der Plasmidbibliotheken repräsentiert werden. Die Sequenz ist in den Datenbanken unter der Accession Nr. AJ010592 hinterlegt.

Die Sequenz des Nucleomorph Chromosoms II von *Guillardia theta* lässt sich in zwei große Bereiche unterteilen: Die subtelomerisch lokalisierten *inverted repeats* und die zentral gelegene *single-copy* Region.

### 3.2.1. Die subtelomerischen *inverted repeats*

Die 13030 bp langen subtelomerischen *inverted repeats* des Nucleomorph Chromosoms II sind in ihrer Sequenz identisch und bereits durch Fraunholz (1999) charakterisiert. Die Basenzusammensetzung zeichnet sich durch einen A/T-Gehalt von 51,1 % aus (Kap. 3.4.).

Sie setzen sich aus je einem Gen der 5S rRNA, das zwischen den ungewöhnlichen, aus  $(AA(GA)_7AG_6)_n$  Repetitionen bestehenden Telomeren und den 18S-ITS1-5,8S-ITS2-28S rDNA Transkriptionseinheiten gelegen ist, und einem 4050 bp großen nicht rRNA codierenden Bereich zusammen. Die rDNA Transkriptionseinheiten werden in Richtung

der Telomere transkribiert, während das Gen für die 5S rRNA entgegengesetzt transkribiert wird.

Evolutionäre Aspekte der 18S rRNA (Van de Peer et al., 1996) und der 28S rRNA Sequenzen (Van der Auwera et al., 1998) sind beschrieben.

In den nicht für rRNA codierenden Bereichen konnten sechs offene Leserahmen (*open reading frames*; *orfs*) mit Codierungskapazitäten von 50 bis 210 Aminosäuren stromaufwärts der rDNA-Transkriptionseinheit identifiziert werden. Diese *orfs* unterschieden sich hinsichtlich ihrer Basenkomposition und Codon-Nutzung drastisch von offenen Leserahmen innerhalb der *single-copy* Region (Fraunholz, 1999). Fünf dieser *orfs* zeigen keinerlei Homologien zu Datenbankeinträgen. Aus diesen beiden Gründen wurden sie nicht als potentiell Protein-codierende Leserahmen berücksichtigt (Zauner et al., 2000). Aufgeführt wird lediglich ein 441 bp langer, offener Leserahmen mit Homologien zu einem Ubiquitin-konjugierenden Enzym *ubc4* (Zauner et al., 2000).

### 3.2.2. Die *single-copy* Region

Die zentral gelegene, 154855 bp lange *single-copy* Region wird von den beiden subtelomerischen, invertierten Sequenzwiederholungen flankiert. Im Gegensatz zu diesen können in der *single-copy* Region zahlreiche offene Leserahmen mit Homologien zu bestehenden Datenbankeinträgen, mehrere tRNAs und zwei snoRNAs identifiziert werden. Auch hinsichtlich der Basenkomposition bestehen erhebliche Unterschiede; so ist der A/T-Gehalt mit 77 % signifikant höher als in den *inverted repeats* (51,1 %; Abb. 3.8.).

#### 3.2.2.1. Identifikation offener Leserahmen

Offene Leserahmen mit einer Codierungskapazität von mehr als 50 Aminosäureresten wurden mit „virtual Genomes“ (<http://alces.med.umn.edu/VGC.html>) bzw. „ORF-Finder“ (<http://www3.ncbi.nlm.nih.gov/gorf/gorf.html>) identifiziert. Die Homologien dieser offenen Leserahmen zu Einträgen in den Datenbanken „EMBL“, „GenBank“, „SwissProt“ und „CyanoBase“ wurde mit den Algorithmen blastP und blastX (Altschul et al., 1997) überprüft.

Als potentiell Protein-codierende *orfs* in der *single-copy* Region wurden betrachtet: 1) *orfs* mit Identitäten zu bekannten Proteinen der Datenbanken, die in Bezug auf Größe und Lokalisation der Homologien mit diesen vergleichbar waren; 2) *orfs* ohne signifikante Homologien mit einer Codierungskapazität von mehr als 100 Aminosäureresten oder 3)

*orfs* ohne signifikante Homologien, einer Codierungskapazität von mehr als 50 Aminosäureresten und ohne Überlappungen zu potentiellen offenen Leserahmen der ersten beiden Kategorien.

Die relative Lage der potentiell Protein-codierenden offenen Leserahmen innerhalb Chromosom II sowie deren Homologien zu den Datenbankeinträgen ist in Abbildung 3.2. bzw. Tabelle 3.2. wiedergegeben (siehe auch Karte im Anhang). Die Lokalisation bzgl. der Gesamtsequenz von Chr II wurde ausgehend vom erstmöglichen Start-Codon bis einschließlich Stop-Codon bestimmt. Die Bezeichnung der offenen Leserahmen ohne Homologien zu Datenbankeinträgen bzw. mit Homologien zu anderen hypothetischen Proteinen wurde nach der Anzahl der codierten Aminosäuren gewählt.

Es wurden insgesamt 150 potentiell Protein-codierende Leserahmen und ein Pseudogen (Kap. 3.2.2.1.2.) innerhalb der *single-copy* Region identifiziert (ohne die beiden für *ubc4* codierenden *orfs* der *inverted repeats*). Dies entspricht einer Gendichte von einem Gen pro ca. 1000 bp. Der A/T-Gehalt dieser 150 potentiell Protein-codierenden Leserahmen beträgt 76,6 % (Kap. 3.4.).

Des Weiteren sind die intergenischen Regionen extrem kurz, sie variieren zwischen 0 bp (Kap. 3.2.2.1.1.; auch Abb. 3.5. und Abb. 3.6.) und 338 bp. Der rein rechnerische Durchschnitt beträgt 58 bp. Die zusammen 8750 bp der intergenischen Regionen besitzen einen A/T-Gehalt von 82,28 % und stellen 5,65 % der *single-copy* Region dar. Dies ist Ausdruck der außerordentlichen Kompaktheit des Nucleomorph-Genoms (Zauner et al., 2000).

Tabelle 3.2. zeigt eine Auflistung der codierenden Bereiche und deren Lokalisation auf dem Nucleomorph Chromosom II.

Lokalisation [bp]	Homologie	Organismus	Accession Nr.	Identität+ [%]	Genname in Abb. 3.2.
279-398	5S rRNA	<i>Guillardia theta</i>	AJ010592	100	5S rRNA
936-4934 c	28S rRNA	<i>Guillardia theta</i>	AJ010592	100	28S rRNA
5493-5663 c	5.8S rRNA	<i>Guillardia theta</i>	AJ010592	100	5.8S rRNA
6175-8214 c	18S rRNA	<i>Guillardia theta</i>	AJ010592	100	18S rRNA
12264-12707	ubiquitin conjugating enzyme	<i>Guillardia theta</i>	AJ010592	89	<i>ubc4</i>
13171-13920	TATA binding protein of transcription factor IID	<i>Guillardia theta</i>	AJ010592	95	<i>tflId</i> <sup>§</sup>
14068-16089 c	hypothetical protein	<i>Guillardia theta</i>	AJ010592	94	<i>orf673</i> <sup>§</sup>
16152-16925 c	40S ribosomal protein S4	<i>Guillardia theta</i>	AJ010592	100	<i>rps4</i> <sup>§</sup>
16993-18192	cell division protein FtsZ	<i>Guillardia theta</i>	AJ010592	91	<i>ftsZ</i> <sup>§</sup>
18207-19001 c	translational initiation factor 2 alpha subunit	<i>Guillardia theta</i>	AJ010592	93	<i>tifII211</i> <sup>§</sup>
19044-19757	proteasome IOTA subunit	<i>Guillardia theta</i>	AJ010592	90	<i>prsA1</i>
19886-21394	T-complex protein 1 gamma subunit	<i>Guillardia theta</i>	AJ010592	94	<i>tcp3</i>
21421-21807	transcription initiation factor IIA gamma chain	<i>Guillardia theta</i>	AJ010592	82	<i>tflIA-S</i>
21867-23507		<i>Guillardia theta</i>	AJ010592	85	<i>orf546</i>
23493-24752 c		<i>Guillardia theta</i>	AJ010592	100	<i>orf419</i> <sup>§</sup>
25088-25930 c	RanBPM/hypothetical protein	<i>Guillardia theta</i>	AJ010592	84	<i>ranbpm</i>
25990-26696 Intron: 25993-26039	40S ribosomal protein S3a	<i>Mus musculus</i>	P97351	35	<i>rps3A</i> *
26741-27220	rubredoxin	<i>Guillardia theta</i>	AJ010592	100	<i>rub</i> <sup>§</sup>
27256-27564	histone H2B	<i>Guillardia theta</i>	AJ010592	88	<i>h2B</i>
27588-28229 c	GTP-binding nuclear protein RAN	<i>Guillardia theta</i>	AJ010592	94	<i>gsp2</i>
28257-28580 c	60S ribosomal protein L30	<i>Guillardia theta</i>	AJ010592	82	<i>rpl30</i>
28598-28668	tRNA Gln anticodon (CUG)	<i>Guillardia theta</i>	AJ010592		<i>trnQ</i> (CUG)
28660-28881 c	DNA-directed RNA polymerases I, II and III 7.7 KD polypeptide (ABC10-alpha)	<i>Guillardia theta</i>	AJ010592	100	<i>rpc10</i>
28806-28902 c Intron: 28851-28857 und 28873-28882	tRNA Ser anticodon (AGA)	<i>Guillardia theta</i>	AJ010592		<i>trnS</i> (AGA)*
28913-29152 Intron: 28916-28957	40S ribosomal protein S28	<i>S. pombe</i>	Q10421	75	<i>rps28</i> *
29171-29662 c Intron: 29596-29648	40S ribosomal protein S13	<i>A. thaliana</i>	P49203	58	<i>rps13</i> *
29716-31062	tubulin alpha-1 chain	<i>Guillardia theta</i>	AJ010592	95	<i>tubA</i>
31310-33190 c					<i>orf626</i>
33246-33974	homology to yeast 34kDa subunit of the tetrameric tRNA splicing endonuclease	<i>S. cerevisiae</i>	NP_009405	27	<i>sen34</i>
34034-34417	60S ribosomal protein L13 (BBC1 protein)	<i>Chlamydomonas sp.</i>	O48513	38	<i>rpl13</i>

34493-36340 c	putative ATP-dependent RNA helicase CDC28	<i>S. pombe</i>	Q10752	32	<i>cdc28</i>
36380-39196					<i>orf938</i>
39255-39644 c					<i>orf129</i>
39703-40059	probable 60S ribosomal protein L14 (hydroxyprolin rich glycoprotein HRGP1)	<i>Pisum sativum</i>	P55844	29	<i>rpl14</i>
40109-41293	putative SAR DNA-binding protein-1	<i>A. thaliana</i>	AAF27012	37	<i>nop5</i>
41321-42730	putative RNA-dependent helicase; Has1p	<i>S. cerevisiae</i>	NP_014017	45	<i>has1</i>
42772-43009 Intron: 42783-42825	pseudo 60L ribosomal protein L24	<i>S. pombe</i>	CAA20919	33	$\psi$ <i>rpl24</i>
43013-43085 c	tRNA ARG anticodon (UCG)				<i>trnR</i> (UCG)
43085-43873 c	homologie to Drosophila dribble	<i>Drosophila melanogaster</i>	Z96931	39	<i>rip1</i>
43818-44258	preprotein translocase secE subunit	<i>Synechocystis sp.</i>	P38382	30	<i>secE</i>
44284-45186					<i>orf300</i>
45209-47404					<i>orf731</i>
47417-48907 c	hypothetical protein slr1974	<i>Synechocystis sp.</i>	BAA18206	40	<i>orf496</i>
48941-50233 c					<i>orf430</i>
50262-50774 c					<i>orf170</i>
50834-51433 c	40S ribosomal protein SA (P40)	<i>Drosophila melanogaster</i>	P38979	42	<i>rsp4</i>
51475-51717					<i>orf80</i>
51812-52831					<i>orf339</i>
52845-55412 c	importin beta-1 subunit	<i>Mus musculus</i>	P70168	20	<i>imb1</i>
55528-58740	IAP100	<i>Pisum sativum</i>	AAC49399	18	<i>iap100</i> <sup>§</sup>
58892-59140					<i>orf82</i> <sup>§</sup>
59218-59709 c	hypothetical protein	<i>Pseudoanabaena sp.</i>	CAA04149	42	<i>orf163</i>
59774-60340 c	60S ribosomal protein L9	<i>Mus musculus</i>	P51410	47	<i>rpl9</i>
60378-60463 c Intron: 60433-60441 und 60392-60394	tRNA Val anticodon (AAC)				<i>trnV</i> (AAC)*
60504-61694 c	(putative) nucleolar G-protein; homologs identified in human and <i>Trypanosoma brucei</i> ; <i>Nog1p</i>	<i>S. cerevisiae</i>	NP_015232	26	<i>nog1</i>
61760-62545 c	putative CCR4-associated factor (POP2 homolog <i>S. cerevisiae</i> )	<i>A. thaliana</i>	AAD15397	43	<i>pop2</i>
62569-63075					<i>orf168</i>
63169-63606					<i>orf145</i>
63710-64138					<i>orf142</i>
64233-64664	eukaryotic translation initiation factor 1A, eIF-1A	<i>Mus musculus</i>	AAC63934	56	<i>eif1A</i> <sup>§</sup>

64719-65468 c	ESTs gb R65052, gb AA712146, gb H76533, gb H76282, gb AA650771, gb H76287, gb AA650887, gb N37383, gb Z29721 and gb Z29722 come from this gene	<i>A. thaliana</i>	AAC64889	24	<i>orf249</i>
65746-66261	eukaryotic cap-binding protein	<i>A. thaliana</i>	AAC17220	28	<i>cbp</i>
66253-66552 c					<i>orf99</i>
66563-67327 c	60S ribosomal protein L8	<i>Xenopus laevis</i>	P41116	59	<i>rpl8</i>
67366-67800	ubiquitin-conjugating enzyme E2-21 KD	<i>Pichia angusta</i>	O60015	38	<i>uceE2</i>
67800-67888 Intron: 67820-67829 und 67869-67875	tRNA Phe anticodon (GAA)				<i>trnF</i> (GAA)*
67897-68460 c					<i>orf187</i>
68535-68924	putative 40S ribosomal protein S19(S24)	<i>A. thaliana</i>	AAD56997	43	<i>rps24</i>
68934-69227 c	nuclear cap binding protein subunit 2, 20kD	<i>Homo sapiens</i>	NP_031388	40	<i>ncbP2</i>
69272-69982 c	hypothetical protein	<i>S. pombe</i>	CAA17930	32	<i>orf236</i>
69985-70075 Intron: 70022-70039	tRNA Cys anticodon (ACA)				<i>trnC</i> (ACA)*
70128-71315	photosystem II stability / assembly factor HCF136	<i>A. thaliana</i>	O82660	47	<i>hcf136</i> <sup>§</sup>
71546-73093	chaperonin-containing-TCP1 theta subunit	<i>Tetrahymena pyriformis</i>	AAD34470	27	<i>tcp1T</i> <sup>§</sup>
73109-73522 c	transcription initiation factor IID 28 KD beta SU	<i>Drosophila melanogaster</i>	P49906	27	<i>taf 30</i>
73562-73870 c					<i>orf102</i>
73930-74445 c					<i>orf171</i>
74703-75890	26S proteasome AAA-ATPase subunit RPT4a	<i>A. thaliana</i>	AAF22524	53	<i>prsS10b</i>
75896-76135	probable small nuclear ribonucleoprotein Sm D3	<i>S. pombe</i>	CAB52041	42	<i>snrpd3</i>
76176-76709					<i>orf177</i> <sup>§</sup>
76719-77834 c	CbbX protein homologue	<i>Guillardia theta</i>	AJ251479	94	<i>cbbx</i> <sup>§</sup>
77982-78272 Intron: 77985-78029	40S ribosomal protein S27	<i>Zea mays</i>	AAC97381	50	<i>rps27</i> *
78282-82586 c	DNA-directed RNA polymerase I 190K chain-like protein	<i>A. thaliana</i>	CAB41189	30	<i>rpa1</i>
82605-82982	SNU13 snRNP SU homolog	<i>S. pombe</i>	AAC62085	47	<i>nhp2</i>
82989-83285 c	40S ribosomal protein S26	<i>Pisium sativum</i>	AAD47346	50	<i>rps26</i>
83347-89520 c	splicing factor Prp8	<i>Homo sapiens</i>	AAC61776	39	<i>prp8</i>
89548-91122	T-complex protein 1, zeta SU (TCP-1-ZETA) (CCT-ZETA)	<i>Oryctolagus cuniculus</i>	O77622	35	<i>tcp1Z</i>
91124-91201 Intron: 91183-91188	tRNA Ile anticodon (UAU)				<i>trnI</i> (UAU)*
91211-91894 c	hypothetical protein slr0565	<i>Synechocystis sp.</i>	slr0565	26	<i>orf227</i>
91900-91982 c	tRNA Ser anticodon (GCU)				<i>trnS</i> (GCU)

91986-94067 c	26S proteasome regulatory SU	<i>S. pombe</i>	CAA21078	21	<i>prsS1</i>
94150-95559	TATA box-binding protein-associated factor TAFII 90	<i>S. cerevisiae</i>	NP_009757	24	<i>taf 90</i>
95614-95856					<i>orf80a</i>
96059-96754 c					<i>orf231</i>
96773-97204 c	40S ribosomal protein S19(S16)	<i>S. pombe</i>	P79016	30	<i>rps19</i>
97260-97730	60S ribosomal protein L21	<i>S. pombe</i>	CAB44755	35	<i>rpl21</i>
97734-98135 c Intron: 98088-98132	40S ribosomal protein S17	<i>Neurospora crassa</i>	P27770	65	<i>rps17*</i>
98166-98942 c	26S proteasome regulatory chain 12 - human	<i>Homo sapiens</i>	S65491	25	<i>prsS12</i>
98964-99545	proteasome chain protein	<i>A. thaliana</i>	CAB10259	34	<i>prsB4</i>
99560-100735 c	26S proteasome AAA-ATPase subunit RPT2a	<i>A. thaliana</i>	AAF22522	56	<i>prsS4</i>
100757-100816 c	snoRNA				<i>snR1</i>
100818-101717	DNA-directed RNA polymerase II 33 KD polypeptide (RPB3)	<i>Homo sapiens</i>	P19387	41	<i>rpb3</i>
101688-102308 c					<i>orf206</i>
102383-104200					<i>orf605</i>
104165-104701 c	40S ribosomal protein S8	<i>Zea mays</i>	Q08069	39	<i>rps8</i>
104803-105222					<i>orf139</i>
105236-105571 c					<i>orf111</i>
105618-105995 c	hypothetical protein sll1509	<i>Synechocystis sp.</i>	sll1509	34	<i>orf125</i>
106032-107987 c	DNA primase	<i>Rickettsia prowazekii</i>	CAA15283	28	<i>dnaG</i>
107987-108982 c					<i>orf331</i>
109108-109335	60S ribosomal protein L36	<i>A. thaliana</i>	AAC23630	40	<i>rpl36</i>
109375-109971	multicatalytic endopeptidase complex, proteasome precursor, beta SU	<i>A. thaliana</i>	CAA74028	33	<i>prsB1</i>
110029-111288 c					<i>orf419a</i>
111378-111815 c	40S ribosomal protein S23	<i>Caenorhabditis elegans</i>	Q19877	66	<i>rps23</i>
111861-112475 c	60S ribosomal protein L15	<i>A. thaliana</i>	O23515	49	<i>rpl15</i>
112525-113091	60S ribosomal protein L10 (QM protein homologue)	<i>D. melanogaster</i>	O61231	65	<i>rpl10</i>
113104-115293 c					<i>orf729</i>
115319-116752	Contains Zinc finger, C3HC4 type (RING finger) domain. ESTs gb N96912 and gb AI994359 from this gene	<i>A. thaliana</i>	AAF06038	29	<i>orf477</i>
116789-117634 c	26S proteasome regulatory complex subunit p37B	<i>Drosophila melanogaster</i>	AAF08394	54	<i>prsS13</i>
117644-117716	tRNA Lys anticodon (UUU)				<i>trnK (UUU)</i>
117721-118881 c	RRP3 is a DEAD box gene homologous to eIF-4a which encodes an RNA-dependent ATPase possessing helicase activity which is not specific for RNA; Rrp3p	<i>S. cerevisiae</i>	NP_011932	32	<i>rrp3</i>

118918-119943	N-myristoyltransferase	<i>Plasmodium falciparum</i>	AAF18461	32	<i>nmt1</i>
119942-120005 c	snoRNA				<i>snR4</i>
120058-121827 c	CPN60 protein	<i>Guillardia theta</i>	CAA09304	96	<i>cpn60</i> <sup>§</sup>
121883-122363 c Intr.:122315-122360	60S ribosomal protein L27A	<i>A. thaliana</i>	P49637	51	<i>rpl27A</i> *
122398-123675	seryl-tRNA synthetase (serin-tRNA ligase)	<i>Helianthus annuus</i>	O81983	51	<i>sys1</i> <sup>§</sup>
123683-125632 c	glucose inhibited division protein A	<i>Synechocystis sp.</i>	Q55694	51	<i>gidA</i>
125674-128259 c					<i>orf861</i>
128453-129265	hypothetical protein	<i>S. pombe</i>	CAB11605	37	<i>orf270</i>
129303-131381	component of a nuclear-localized tRNA splicing complex; Sen1p	<i>S. cerevisiae</i>	NP_013534	39	<i>sen1</i>
131401-132288 c	cell division cycle 2 homolog (S. pombe)-like 2.	<i>Mus musculus</i>	NP_031687	38	<i>cdc2</i> <sup>§</sup>
132565-132908 Intr.: 132625-132668	contains similarity to a C3HC4 zinc finger (PS:PS00518)	<i>Caenorhabditis elegans</i>	AAB37826	38	<i>rbp1</i> *
132956-133168					<i>orf64</i>
133187-134284	56 kDa nucleolar snRNP protein that shows homology to beta SU of G-proteins and the splicing factor Prp4; Sof1p	<i>S. cerevisiae</i>	NP_013090	21	<i>sof1</i>
134322-135336 c					<i>orf338</i>
135381-135758					<i>orf125</i>
135770-136693					<i>orf307</i>
136690-137610 c	pleiotropic regulator 1	<i>Mus musculus</i>	AAC04388	22	<i>prl1</i>
137643-137903					<i>orf86</i>
137986-138957 c	hypothetical protein sll2002	<i>Synechocystis sp.</i>	BAA17726	30	<i>orf323</i> <sup>§</sup>
139024-140001					<i>orf325</i>
140025-140219	ubiquitin / ribosomal protein CEP52	<i>A. thaliana</i>		64	<i>rpl40</i>
140224-140295	tRNA Thr anticodon (AGU)				<i>trnT</i> (AGU)
140299-140904 c					<i>orf201</i>
140947-142140	26S protease regulatory SU 6A homolog	<i>Lycopersicon esculentum</i>	P54776	56	<i>prsS6</i>
142170-143354 c	TFIIB related factor hBRF	<i>Homo sapiens</i>	AAB38876	30	<i>tfl1d</i> rbf
143424-143660	putative small nuclear ribonucleoprotein E	<i>S. pombe</i>	CAB59808	40	<i>snrpE</i>
143761-144498	hypothetical 51.4 KD protein C13G1.09 in chromosome II	<i>S. pombe</i>	O60071	25	<i>orf245</i>
144505-148815 c	putative U5 snRNP-specific 200kd protein	<i>S. pombe</i>	CAB57421	24	U5snRNP
148867-151452					<i>orf861</i>
151579-151649 c	tRNA His anticodon (GUG)				<i>trnH</i> (GUG)
151664-152110 c					<i>orf148</i>
152158-152781 c					<i>orf207</i>
152812-153060 c	small nuclear ribonucleoprotein SM D2	<i>Caenorhabditis elegans</i>	Q18786	34	<i>snrpD2</i>
153090-153482 c	40S ribosomal protein S15A	<i>Homo sapiens</i>	P39027	60	<i>rps15A</i>

153528-154712 c	26S protease regulatory SU 7	<i>Spinacia oleracea</i>	Q41365	66	<i>prsS7</i>
154785-155690	beta subunit of G protein coupled to mating factor receptor	<i>S. cerevisiae</i>	NP_014855	30	<i>ste4</i>
155787-156074 c					<i>orf95</i>
156168-156758 c					<i>orf196</i>
156777-156849 c	tRNA Lys anticodon (CUU)				<i>trnK</i> (CUU)
156849-157550 c	hypothetical GTP-binding protein	<i>Sulfolobus solfataricus</i>	CAB57719	32	GTP-bp
157676-162097	DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain	<i>A. thaliana</i>	CAA21466	43	<i>rpb1</i>
162133-162483					<i>orf116</i>
162496-163056 c	CG6712 gene product	<i>D. melanogaster</i>	AAF53162	22	<i>orf186</i>
163110-163721 c					<i>orf203</i>
163796-164071	60S ribosomal protein L37A	<i>Plasmodium falciparum</i>	O96184	45	<i>rpl37A</i>
164106-164804 c	60S ribosomal protein L5	<i>Dunaliella salina</i>	O22608	28	<i>rpl5</i>
164843-165325 c					<i>orf160</i>
165360-166433	mRNA capping enzyme (mRNA guanylyltransferase)	<i>Paramecium bursaria</i> <i>Chlorella virus 1</i>	Q84424	22	<i>mce</i>
166509-167444 c	guanine nucleotide-binding protein beta SU like protein	<i>Nicotiana tabacum</i>	P49026	45	<i>gblp</i> <sup>§</sup>
167528-167848	60S ribosomal protein L24	<i>A. thaliana</i>	AAC31838	26	<i>rpl24</i>
167848-167954 Intr: 167918-167941	tRNA Leu anticodon (CAA)				<i>trnL</i> (CAA)*
168209-168652 c	ubiquitin conjugating enzyme	<i>Guillardia theta</i>	AJ010592	89	<i>ubc4</i>
172702-174741	18S rRNA	<i>Guillardia theta</i>	AJ010592	100	18S rRNA
175253-175423	5.8S rRNA	<i>Guillardia theta</i>	AJ010592	100	5.8S rRNA
175982-179980	28S rRNA	<i>Guillardia theta</i>	AJ010592	100	28S rRNA
180518-180637 c	5S rRNA	<i>Guillardia theta</i>	AJ010592	100	5S rRNA

Tab. 3.2.: **Auflistung der codierenden Bereiche und deren Lokalisation.** Die Homologien beziehen sich bei potentiellen Protein-codierenden *orfs* auf die Ergebnisse einer blast-Suche der jeweiligen Proteinsequenz gegen die nicht redundanten Datenbanken des NCBI-Browsers (blastX bzw. blastP gegen nr), bei rRNA auf die Ergebnisse einer blast-Suche der Nucleotidsequenz ebenfalls gegen die nicht redundanten Datenbanken des NCBI-Browsers (blastN gegen nr). Die Bezeichnung der offenen Leserahmen ohne Homologien bzw. mit Homologien zu hypothetischen Proteinen anderer Organismen wurde dabei nach der Anzahl der codierten Aminosäuren gewählt. Die Identifikation der tRNAs erfolgte am tRNAscan-SE Search Server (Kap. 3.2.2.3.). c, komplementärer DNA-Strang; \*, mit Intron; §, cDNA vorhanden. †, die Identitäten beziehen sich auf die Suchergebnisse der gefilterten Sequenzen (dies entspricht den *default* Einstellungen) in denen Regionen mit geringer Komplexität (z.B. Regionen mit hohem Arginin und/oder Lysin Anteil) nicht berücksichtigt werden um potentiell irreführende Ergebnisse zu vermeiden (Altschul et al., 1997). Die Accession Nr. der angegebenen Homologons beziehen sich auf die Datenbanken „EMBL“, „Genbank“, „SwissProt“ und im Falle von *Synechocystis*-Homologons auf „Cyanobase“. Die Regionen bzw. Gene mit Homologien zu *Guillardia theta* sind bereits publiziert. Die Lokalisation bezieht sich auf die in den Datenbanken hinterlegte Sequenz des Chromosoms II mit der Accession Nr. AJ010592.



### Vorangegangene Seite

Abb. 3.2.: **Karte der codierenden Bereiche des Nucleomorph Chromosoms II von *Guillardia theta*.** \*, mit Intron; §, cDNA vorhanden. Die Bezeichnungen und die Koordinaten der einzelnen codierenden Bereiche sind in Tabelle 3.2. aufgelistet. Hellblau unterlegt sind die invertierten Sequenzwiederholungen. Offene Leserahmen bzw. codierende Bereiche, die oberhalb der als einfache Line schematisch dargestellten Sequenz des Chromosoms II liegen, sind auf dem plus Strang, unterhalb liegende auf dem komplementären Gegenstrang lokalisiert (siehe auch Karte im Anhang).

Die 150 innerhalb der *single-copy* Region von Chromosom II identifizierten potentiellen offenen Leserahmen wurden gemäß ihrer Homologien und der daraus abgeleiteten bekannten bzw. potentiellen Funktion in die Kategorien der Abbildung 3.2. und 3.3. eingeteilt. Etwa 30 % (46) der *orfs* zeigen keine signifikanten Homologien zu Datenbankeinträgen, fünf zeigen Identitäten zu hypothetischen Proteinen anderer eukaryotischer Organismen mit unbekannter Funktion, bzw. zu Sequenzen der *A. thaliana*- bzw. *C. elegans* EST-Projekte. Insgesamt 83 *orfs* zeigen gute Homologien zu Datenbankeinträgen mit bekannter Funktion. Zusammen 65 *orfs* codieren für Bestandteile der Transkriptions- und Translations-Maschinerie, des DNA-Metabolismus (einschließlich Zellzyklus und Signaltransduktion) sowie des RNA-Metabolismus (RNA-Prozessierung und RNA-abhängige Helikasen) des Symbionten. Für Teile des Proteinmetabolismus (Degradation über Proteasom, T-complex Protein und Faltung) codieren 16 *orfs*, weitere zwei *orfs* wurden in der Kategorie Cytoskelett zusammengefasst. Zwei *orfs* konnten in keine der Kategorien eingeteilt werden, da Ihre Funktion im Symbionten nicht eindeutig zugeordnet werden kann. Bekannte plastidäre Funktionen vermitteln neun offene Leserahmen und sechs *orfs* zeigen Homologien zu hypothetischen Proteinen des phototrophen Cyanobakteriums *Synechocystis sp.* (PCC6803). Diese 15 offenen Leserahmen zeichnen sich durch eine N-terminale Extension aus, die *G. theta* Nucleomorph spezifische, aber dennoch typische Merkmale eines Transitpeptides aufweisen (Wastl, 2000). Im Falle des Rubredoxins (*rub*, Lokalisation 26741-27220 bp) konnte eine Funktion der N-terminalen Extension als Transitpeptid auch experimentell nachgewiesen werden (J. Prechtel, pers. Mitteilung; Wastl und Maier, 2000). Keiner der identifizierten *orfs* zeigte Homologien zu Genen des Primärmetabolismus.

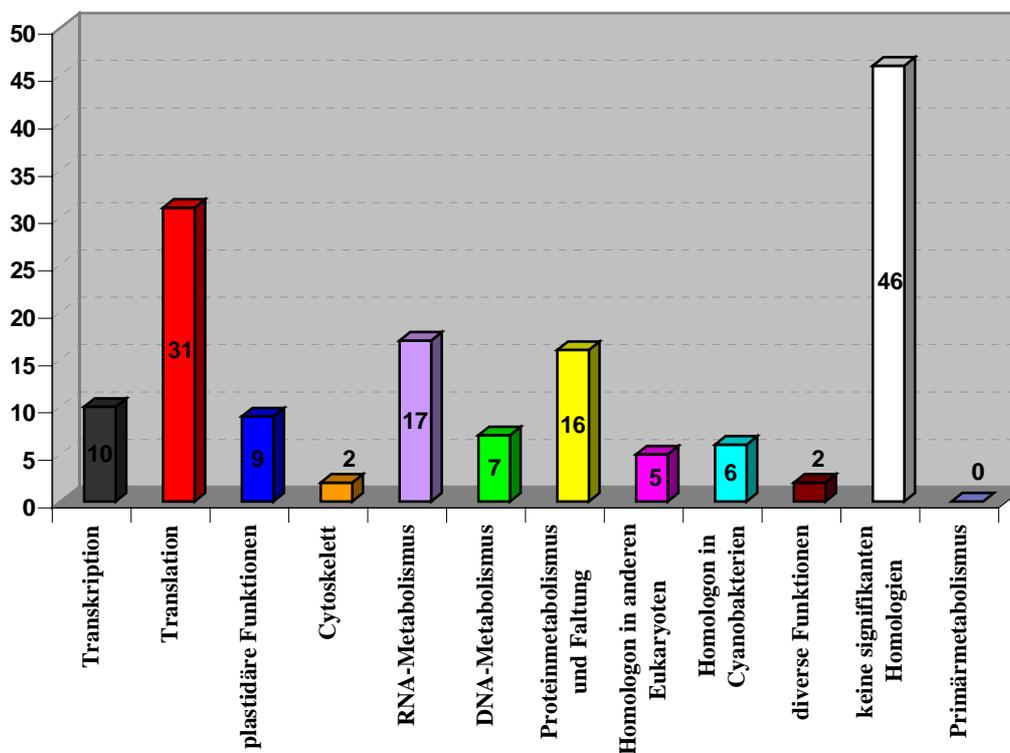


Abb. 3.3.: Funktionen der auf dem Nucleomorph Chromosom II von *Guillardia theta* lokalisierten offenen Leserahmen. Es wurden die selben Kategorien wie in Abbildung 3.2. verwendet, näheres siehe Text.

### 3.2.2.1.1. Überlappende Protein-codierende Leserahmen

Einen weiteren Beleg für die starke Komprimierung des Chromosoms II stellt die Identifizierung von sieben Sequenzbereichen dar, die von je zwei *orfs* gemeinsam genutzt werden. Eine Zusammenfassung zeigt Tabelle 3.3. Mit Ausnahme der Überlappungen von *rip1* mit *secE* und von *dnaG* mit dem *orf331* überlappen die jeweiligen Partner mit ihren 3' codierenden Termini. Diese Überlappungen im 3'-Terminus der codierenden Bereiche werden auch nicht durch die Annahme alternativer Start-Codons in Frage gestellt. Die offenen Leserahmen der DNA Primase (*dnaG*) und des *orf331* überlappen um ein Nucleotid. Da beide offenen Leserahmen die selbe Transkriptionsrichtung aufweisen, bedeutet dies, dass das erste Nucleotid des potentiellen Start-Codons der DNA Primase gleichzeitig das letzte Nucleotid des Stop-Codons von *orf331* darstellt. Die offenen

Leserahmen von *rip1* und *secE* überlappen mit ihren 5'-Bereichen um 56 Nucleotide. Sie sind antiparallel lokalisiert und werden divergent transkribiert. Sowohl die DNA Primase (*dnaG*) als auch *secE* stellen plastidär lokalisierte, Nucleomorph codierende Proteine dar und müssen über die Plastidenmembranen transportiert werden. Zu diesem Zweck besitzen beide eine N-terminale Extension, die ein Transitpeptid darstellt (Wastl, 2000). Da das Transitpeptid nach erfolgter Translokation entfernt wird, ist, im Unterschied zu den Überlappungen im 3'-terminalen Bereich, in diesem Falle keine Überlappung der potentiell Protein-codierenden Bereiche der reifen Proteine gegeben.

Gene	Lokalisation [bp]	gem. genutzte Basenpaare	überlappende Bereiche
<i>orf546</i>	<u>21867-23507</u>	15 bp	3'-terminal
<i>orf419</i>	<u>23493-24752</u> c		3'-terminal
<i>rip1</i>	<u>43085-43873</u> c	56 bp	5'-terminal
<i>secE</i>	<u>43818-44258</u>		5'-terminal
<i>cbp</i>	<u>65746-66261</u>	9 bp	3'-terminal
<i>orf99</i>	<u>66253-66552</u> c		3'-terminal
<i>rpb3</i>	<u>100818-101717</u>	30 bp	3'-terminal
<i>orf206</i>	<u>101688-102308</u> c		3'-terminal
<i>orf605</i>	<u>102383-104200</u>	36 bp	3'-terminal
<i>rps8</i>	<u>104165-104701</u> c		3'-terminal
<i>dnaG</i>	<u>106032-107987</u> c	1 bp	5'-terminal
<i>orf331</i>	<u>107987-108982</u> c		3'-terminal
<i>orf307</i>	<u>135770-136693</u>	4 bp	3'-terminal
<i>prl1</i>	<u>136690-137610</u> c		3'-terminal

Tab. 3.3.: Zusammenfassung der Überlappungen potentiell Protein-codierender Sequenzbereiche des Nucleomorph Chromosoms II von *Guillardia theta*. *rip1*, Rev interagierendes Protein Rip-1; *secE*, Preprotein Translocase *secE* Untereinheit; *cbp*, eukaryotisches *cap*-bindendes Protein; *rpb3*, DNA-abhängige RNA-Polymerase II 33 KD Untereinheit; *rps8*, 40S ribosomales Protein S8; *dnaG*, DNA Primase; *prl1*, PRL1/PRL2 ähnliches Protein; c, komplementär. Die Lokalisation bezieht sich auf die in der Datenbank hinterlegte Sequenz des Chromosoms II mit der Accession Nr. AJ010592.

### 3.2.2.1.2. Introns in Protein-codierenden Leserahmen

Anhand der Homologiesuche in den Datenbanken (blastX bzw. blastP, nr-aa) mit den potentiellen Proteinsequenzen der identifizierten *orfs* und der Nucleotidsequenz der intergenischen Regionen, konnten mehrere Regionen identifiziert werden, die hohe Identitäten zu ribosomalen Proteinen aufwiesen. Diese Homologien bezogen sich nicht auf die gesamte Sequenz der Datenbankeinträge, sondern nur auf deren C-Terminus. Eine nähere Untersuchung der Regionen dieser Sequenzen auf Chr II ergab sieben potentielle spliceosomale Introns mit der Konsensussequenz GTA(A/G)GTN<sub>34-44</sub>AG ( Alignment 3.1.). Nach der Entfernung dieser Sequenzen und nochmaligem Datenbankabgleich zeigten sich Identitäten über die gesamte Länge der Datenbankeinträge. Aus diesem Grund wurden diese Sequenzen als Introns bezeichnet, ein experimenteller Beweis steht noch aus. Einen Sonderfall stellt *rpl24* dar, der 66 AS lange reife ORF zeigt 33 % Identität zu dem N-Terminus von *Schizosaccharomyces pombe rpl24* und ähnliche Werte zu *rpl24* anderer Eukaryoten in den Datenbanken, deren Länge jedoch mindestens 149 AS beträgt. Zudem ist ein weiteres 60S ribosomales Protein L24 auf Chr II lokalisiert (bei den Koordinaten 167528 bp). Aus diesen Gründen wurde der 66 AS lange ORF auf Chr II mit Homologien zu eukaryoten *rpl24* als Pseudogen gewertet ( $\psi$ ). Eine Überprüfung der gesamten Sequenz des Chr II mit den bekannten Parametern der identifizierten Introns ergab ein weiteres potentielles Intron. Eine Homologiesuche des reifen potentiellen Proteins ergab Identitäten von 38 % zu einem C3HC4 *zinc finger* Protein von *Caenorhabditis elegans*.

GENE	5'		3'	PHASE
<i>rps3A</i>	gtaagt	ttacatagtattaacattatTTTTAAAAactaacatttt	ag	0
<i>rps28</i>	gtaggt	attaatttatcattgactcataatcagataatat	ag	0
<i>rps13</i>	gtaagt	tatagtcttaacttaaaaattattatcgattaataaaatt	ag	2
<i>rpl24</i> ( $\psi$ )	gtaagt	atTTTAAAAatataaaaatatTTTTgttgatttc	ag	2
<i>rps27</i>	gtaagt	aaagttaattTTTtagttagacgatctattataatgca	ag	0
<i>rps17</i>	gtaagt	aaataatattTgtttagattgagtttgaaattaataa	ag	0
<i>rpl27A</i>	gtaagt	tttgattTAAAAattTTTTTTacctattaatagattt	ag	0
<i>rbp1</i>	gtaagt	agtaaatattTTTTTTTTtataaataatgttaatat	ag	0
Konsensus	gtaRgt		ag	

Alignment 3.1.: **Alignment der Intronsequenzen und der daraus resultierenden Konsensussequenz.** *rps3A*, 40S ribosomales Protein S3A; *rps28*, 40S ribosomales Protein S28; *rps13*, 40S ribosomales Protein S13; *rpl24*( $\psi$ ), 60S ribosomales Pseudoprotein L24; *rps27*, 40S ribosomales Protein S27; *rps17*, 40S ribosomales Protein S17; *rpl27A*, 60S ribosomales Protein L27A; *rbp1*, C3HC4 *zinc finger* Protein. **R**, Adenin oder Guanin; die Phase stellt die relative Lage des Introns zum Leserahmen dar, näheres siehe Text.

Die relative Lage der Introns bezogen auf die reifen Proteine ist in Abbildung 3.4. graphisch dargestellt. Die Introns befinden sich alle nahe des 5'-terminalen Bereichs der potentiellen offenen Leserahmen. Mit Ausnahme der Introns in *rps13*,  $\Psi rpl24$  und *rbx1* sind die Introns direkt nach dem Start-Codon ATG lokalisiert.

Der A/T-Gehalt der inserierten Sequenzen beträgt 85,5 % (Kap. 3.4.).

Die Phase deutet die relative Lage der Introns bezüglich des Leserahmens an. Phase 0 bedeutet, dass das Intron nach dem dritten Nucleotid, d.h. nach einem Codon, inseriert ist. Dem entsprechend ist ein Phase 1 (2) Intron nach dem ersten (zweiten) Nucleotid eines Codons inseriert

Weitere neun Introns mit der Konsensussequenz GTA(A/G)GTN<sub>34-47</sub>AG konnten auch auf den Chromosomen I und III identifiziert werden (S. Douglas, pers. Mitteilung; Kap. 3.4.).

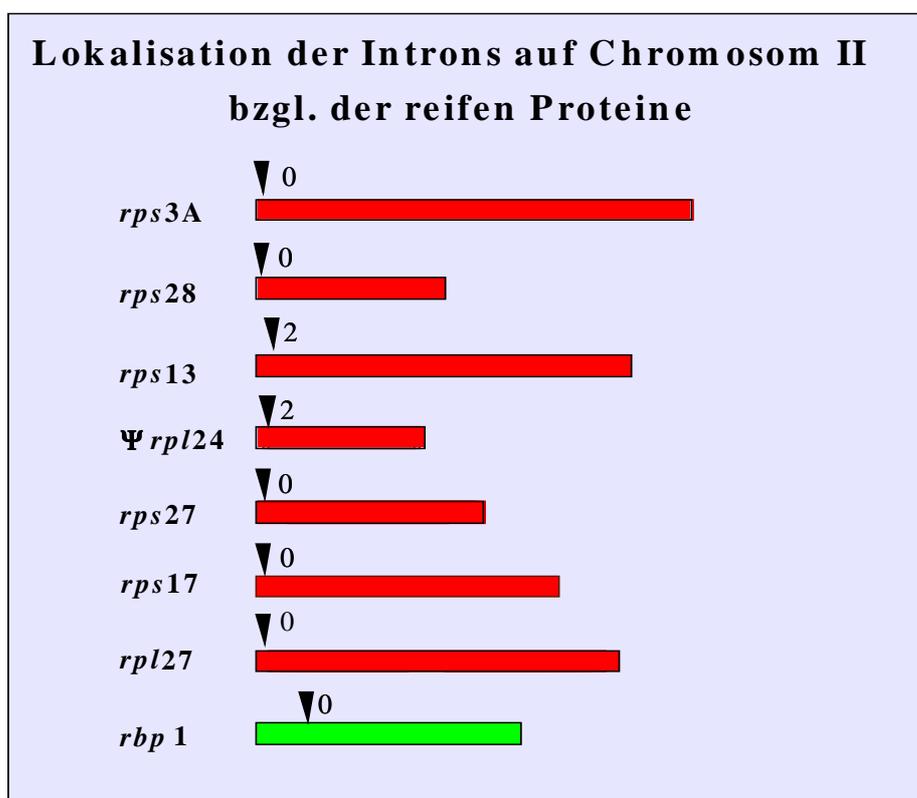


Abb. 3.4.: **Relative Lage der auf Chromosom II identifizierten Introns bezüglich der reifen Proteine.** Die Zahlen neben den Pfeilspitzen stellen die Phase dar. Weitere Erläuterungen siehe Text. Für die Bezeichnung der Gene siehe Alignment 3.1. Die Farben entsprechen den Kategorien aus Abbildung 3.2.

### 3.2.2.1.3. Codon-Nutzung

In der Tabelle 3.4. ist die Codon-Nutzung (*codon usage*) der in dieser Arbeit ermittelten Protein-codierenden Sequenzen der *single-copy* Region des Nucleomorph Chromosoms II dargestellt. Dabei werden die Anteile der genutzten Codons an der Gesamtzahl der jeweilig codierten Aminosäuren angegeben. In der Tabelle zeigt sich, dass die Präferenz zum Einbau von A+T Nucleotiden auch die Codon-Nutzung beeinflusst. Es werden jeweils die Codons verwendet, die A+T reicher sind als ihre Synonyme. Die Codon-Nutzung ist also deutlich gewichtet.

Gut zu erkennen ist die bevorzugte Nutzung A+T reicher Codons an den Aminosäuren, die von zwei Codons codiert werden. In diesen Fällen wird das Codon mit dem höheren A+T Anteil in über 80 % der Fälle genutzt (Tab. 3.4.).

Auch die Verteilung der Nucleotide Adenin und Thymin auf die einzelnen Codon-Positionen ist unterschiedlich. So sind 72,96 % der Nucleotide in der ersten Codon-Position, 75,16 % der zweiten und 82,08 % der dritten Position Adenin oder Thymin.

<b>Ala</b>	GCG	6,7	<b>Gln</b>	CAG	15,3	<b>Leu</b>	TTG	17,6	<b>Ser</b>	AGT	19,1
	GCA	50,4		CAA	84,7		TTA	52,5		AGC	4,8
	GCT	36,3	<b>Glu</b>	GAG	12,3		CTG	2,9		TCG	6,2
	GCC	6,6		GAA	87,7		CTA	10,5		TCA	33,9
<b>Arg</b>	AGG	11,8	<b>Gly</b>	GGG	7,1		CTT	13,7		TCT	30,2
	AGA	73,2		GGA	49,1		CTC	2,7		TCC	5,9
	CGG	0,9		GGT	37,0	<b>Phe</b>	TTT	82,8	<b>Trp</b>	TGG	100,0
	CGA	7,9		GGC	6,8		TTC	17,2	<b>Tyr</b>	TAT	78,4
	CGT	5,0	<b>His</b>	CAT	81,9	<b>Pro</b>	CCG	6,7		TAC	21,6
	CGC	1,2		CAC	18,1		CCA	44,3	<b>Val</b>	GTG	8,3
<b>Asn</b>	AAT	81,5	<b>Ile</b>	ATA	38,3		CCT	41,7		GTA	34,6
	AAC	18,9		ATT	51,7		CCC	7,3		GTT	48,0
<b>Asp</b>	GAT	82,9		ATC	10,0	<b>Thr</b>	ACG	6,3		GTC	9,1
	GAC	17,1	<b>Lys</b>	AAG	12,4		ACA	44,5	<b>End</b>	TGA	16,1
<b>Cys</b>	TGT	80,9		AAA	87,6		ACT	40,5		TAG	18,1
	TGC	19,1	<b>Met</b>	ATG	100,0		ACC	8,7		TAA	65,8

Tab. 3.4.: **Häufigkeit der Nutzung einzelner Codons bezüglich der Gesamtzahl aller Codons einer Aminosäure.** Die Aminosäuren sind fett hervorgehoben, die Codons grau unterlegt, die Häufigkeit [in %] der einzelnen verwendeten Codons bezieht sich auf die Gesamtzahl aller Codons einer Aminosäure (= 100 %). Die Tabelle wurde anhand der potentiell Protein-codierenden offenen Leserahmen aus Tabelle 3.2. der *single-copy* Region des Nucleomorph Chromosoms II erstellt. Es wurden insgesamt 48295 Codons in die Untersuchung mit aufgenommen. Die Berechnungen wurden mit dem Programm *DNA-Tools* 5.1 durchgeführt (Kap. 2.1.13.).

### 3.2.2.2. Das Nucleomorph ist transkriptionell aktiv

Aus einem das Nucleomorph Genomprojekt begleitenden EST-Projekt konnten im Rahmen dieser Arbeit 14 cDNAs isoliert werden, die auf Transkripte des Nucleomorph-Genoms zurück gehen. Ausgehend von einer aus Gesamt-RNA generierten cDNA-Bibliothek von *Guillardia theta* (Fraunholz, 1999) konnten im Rahmen des EST-Projektes bisher ca. 300 cDNAs von *G. theta* isoliert werden. Neun dieser cDNAs konnten eindeutig den Transkripten Nucleomorph Chromosom II codierter offener Leserahmen zugeordnet werden. Zu beachten ist, dass drei der neun isolierten cDNAs auf Transkripte potentiell Protein-codierender offener Leserahmen ohne Homologien zu den in den Datenbanken hinterlegten Sequenzen zurückgehen. Alle Sequenzen der cDNAs waren, mit Ausnahme der 3' poly-Adenylierung, mit der für das Chromosom II ermittelten Gesamtsequenz identisch, so dass keine mRNA Editierung anhand dieser cDNAs nachgewiesen werden konnte. Die Analyse der isolierten cDNAs bestätigte die bereits beschriebene hochgradige Überlappung der Transkripte mit den codierenden Bereichen der benachbarten ORFs (Fraunholz, 1999; Zauner et al., 2000; auch Abb. 3.6.). Weitere fünf cDNAs des EST-Projektes gehen auf die mRNA Nucleomorph Chromosom I bzw. III codierter ORFs zurück.

In vorangegangenen Arbeiten sind weitere cDNAs, die mit den Nucleomorph Chromosoms II codierten ORFs korrespondieren, beschrieben. Sechs cDNAs aus den Randbereichen der *single-copy* Region (Fraunholz, 1999) und fünf weitere, die Transkripten von plastidär-lokalisierten, aber Nucleomorph Chr II codierten offenen Leserahmen zuzuordnen sind (Wastl, 2000), sind beschrieben. All dies beweist die transkriptionelle Aktivität des Nucleomorph-Genoms von *G. theta*.

Alle offenen Leserahmen von denen eine korrespondierende cDNA bekannt ist, sind in Abbildung 3.2. und Tabelle 3.2. mit § gekennzeichnet.

### 3.2.2.3. Identifikation tRNA-codierender Sequenzabschnitte

Die Sequenz des Chr II wurde in überlappenden Teilsequenzen zu je 10000 bp mit Hilfe des Suchalgorithmus tRNAscan-SE (<http://www.genetics.wustl.edu/eddy/tRNAscan-SE>; Lowe und Eddy, 1997) mit den *default* Einstellungen bezüglich tRNA codierender Sequenzabschnitte untersucht. Mit Hilfe dieses Algorithmus konnten zehn tRNAs identifiziert werden. Durch Modifikationen des Algorithmus (Suche nur im *cove* Modus und *cove score cutoff* = 0) konnten drei weitere tRNAs (tRNA<sup>Val</sup>(AAC); tRNA<sup>Phe</sup>(GAA) und tRNA<sup>Leu</sup>(CAA)) identifiziert werden (Tab. 3.5.).

Somit sind insgesamt 13 tRNAs, sechs davon mit Intron(s), auf Chromosom II lokalisiert (Tab. 3.5.). Sieben tRNAs zeigen keine Insertionen (tRNA<sup>Gln</sup>(CUG); tRNA<sup>Arg</sup>(UCG); tRNA<sup>Ser</sup>(GCU); tRNA<sup>Lys</sup>(UUU); tRNA<sup>Thr</sup>(AGU); tRNA<sup>His</sup>(GUG) und tRNA<sup>Lys</sup>(CUU)).

Zwei Insertionen weist die tRNA<sup>Ser</sup>(AGA) auf. Zum einen eine in der Anticodon-Schleife lokalisierte sieben Basenpaare große Insertion und ein weiteres zehn Basenpaare langes Intron in der D-Schleife. Ein Insert in der D-Schleife war bisher nicht bekannt. Mittels RT-PCR konnte die Prozessierung dieser tRNA<sup>Ser</sup>(AGA) nachgewiesen werden, so dass es sich nicht um eine pseudo-tRNA handelt (Zauner et al., 2000).

Ebenso zwei Insertionen, je eine in der Pseudouridin-Schleife und eine in der D-Schleife, weisen die tRNA<sup>Val</sup>(AAC) und tRNA<sup>Phe</sup>(GAA) auf.

Ein „klassisches“ Intron in der Anticodon-Schleife trägt die tRNA<sup>Cys</sup>(ACA). Eine Insertion in der Anticodon-Schleife ist relativ häufig und in unterschiedlichsten Organismengruppen beschrieben. Diese Introns stellen in *S.cerevisiae* eine 14-60 nt lange Insertion direkt 3' zum Anticodon in der Anticodon-Schleife dar (Trotta et al., 1997), die von tRNA-*splicing*-Endonucleasen entfernt werden, wobei die entstehenden tRNA-Schenkel durch tRNA-Ligasen zum funktionellen tRNA-Molekül verbunden werden (Abelson et al., 1998).

Je ein Intron in der Pseudouridin-Schleife besitzen die Sequenzen der tRNA<sup>Ile</sup>(UAU) und tRNA<sup>Leu</sup>(CAA). Die tRNA<sup>Leu</sup>(CAA) ist bereits in Fraunholz (1999) beschrieben und zeichnet sich durch eine ungewöhnliche 24 bp Insertion in der TΨC-Schleife aus

tRNA Typ	tRNA Start – Ende	Strang	Anti-codon	Intron Start – Ende	Lage des Introns	<i>cove A</i> <i>cove B</i>
Gln	28598 – 28668	+	CUG	—	—	66.16
Ser	28902 – 28806	–	AGA	28857 – 28851 28882 – 28873	D Anticodon	47.21 69.84
Arg	43085 – 43013	–	UCG	—	—	70.46
Val	60463 – 60378	–	AAC	60441 – 60433 60394 – 60392	D TΨC	28.77 58.35
Phe	67800 – 67888	+	GAA	67820 – 67829 67869 – 67875	D TΨC	38.11 70.93
Cys	69983 – 70075	+	ACA	70022 – 70039	Anticodon	56.64 55.78
Ile	91124 – 91201	+	UAU	91183 – 91188	TΨC	61.31 77.63
Ser	91982 – 91900	–	GCU	—	—	75.70
Lys	117644 – 117716	+	UUU	—	—	85.97
Thr	140224 – 140295	+	AGU	—	—	72.84
His	151650 – 151580	–	GUG	—	—	68.37
Lys	156860 – 156778	–	CUU	—	—	73.27
Leu	167848 – 167954	+	CAA	167918 – 167941	TΨC	37.76 69.68

Tab. 3.5.: Übersicht der auf Chromosom II lokalisierten tRNA codierenden Bereiche und deren Introns. Die Lage der Introns bezieht sich auf die jeweiligen Schleifen der Sekundärstruktur. *cove A*, *cove B* Wert mit Intron; *cove B*, *cove B* Wert nach der Entfernung der Introns. Die Lokalisationen in bp beziehen sich auf die in der Datenbank hinterlegte Sequenz des Chromosoms II mit der Accession Nr. AJ010592.

Nach der Entfernung der Introns steigt der Wert des *cove* Algorithmus beträchtlich. Dieser Wert ist ein Maß für die Qualität der Vorhersage der tRNA Sekundärstruktur (Lowe und Eddy, 1997). Dies trifft besonders auf die drei durch den modifizierten tRNAscan-SE Algorithmus identifizierten tRNAs zu (Tab. 3.5.). Eine Auflistung der Sequenzen und der potentiellen Sekundärstruktur aller tRNAs des Chr II ist im Anhang zu finden.

### 3.2.2.3.1. Überlappungen tRNA-codierender Sequenzabschnitte mit Protein-codierenden Leserahmen

Die sehr starke Komprimierung des Chromosoms II zeigt sich auch in der gemeinsamen Nutzung codierender Bereiche, die durch unterschiedliche RNA-Polymerasen transkribiert werden (Tab. 3.6.). So wird der 5'-Bereich des offenen Leserahmen mit Homologien zu *rpc10* (eine den RNA-Polymerasen I, II und III gemeinsame 7,7 KD große Untereinheit) auch als codierender Bereich des 3'-Endes der tRNA<sup>Ser</sup>(AGA) genutzt. Darüber hinaus überlappt der 3'-Bereich von *rpc10* mit dem 3'-Ende der tRNA<sup>Gln</sup>(CUG) (Abb. 3.5.).

Weitere Überlappungen finden sich jeweils im 3'-Bereich der RNA Polymerase II transkribierten offenen Leserahmen des hypothetischen GTP bindenden Proteins mit dem 5'-Ende der tRNA<sup>Lys</sup>(CUU), dem Rev interagierenden Protein Rip-1 (*rip1*) mit dem 5'-Bereich der tRNA<sup>Arg</sup>(UGG) und dem ribosomalen Protein L24 (*rpl30*) mit dem 5'-Ende der tRNA<sup>Leu</sup>(CAA) (Tab. 3.6.). Die letztgenannten nutzen je ein Nucleotid gemeinsam.

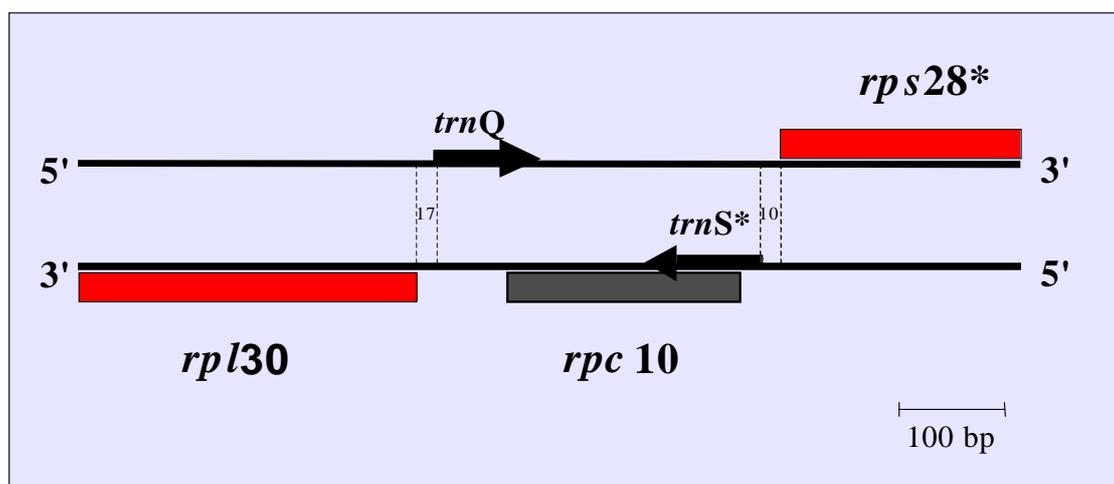


Abb. 3.5.: Schematische Darstellung der überlappenden Bereiche von *rpc10* mit *trnQ* und *trnS\**. Dargestellt ist der DNA-Doppelstrang des Chr II im Bereich zwischen 28260 und 29149 bp. *rps28\**, 40S ribosomales Protein S28; *rpl30*, 60S ribosomales Protein L30; *rpc10*, DNA-abhängige RNA-Polymerase I, II und III 7,7 KD Untereinheit (ABC10-alpha); *trnQ*, tRNA<sup>Gln</sup>(CUG); *trnS\**, tRNA<sup>Ser</sup>(AGA); \*, mit Intron. Die Farben entsprechen den Kategorien aus Abbildung 3.2. Die Zahlen zwischen den senkrechten, gestrichelten Linien stellen die Größe der intergenischen Regionen in bp dar.

codierende Bereiche	Lokalisation [bp]	gem. genutzte Nucleotide	überlappende Bereiche
tRNA <sup>Gln</sup> (CUG) <i>rbc10</i> tRNA <sup>Ser</sup> (AGA)	<u>28596–28668</u> <u>28660–28881</u> c <u>28806–28902</u> (c)	9 nt 76 nt	→ ← ←
tRNA <sup>Arg</sup> (UCG) <i>rip1</i>	<u>43013–43085</u> (c) <u>43085–43873</u> c	1 nt	← ←
tRNA <sup>Lys</sup> (CUU) <i>GTP-bp</i>	<u>156777–156849</u> (c) <u>156849–157550</u> c	1 nt	← ←
<i>rpl24</i> tRNA <sup>Leu</sup> (CAA)	<u>167528–167848</u> <u>167848–167954</u>	1 nt	→ →

Tab. 3.6.: Zusammenfassung der überlappenden tRNA codierenden Sequenzbereiche mit potentiell protein-codierenden offenen Leserahmen des Nucleomorph Chromosoms II von *Guillardia theta*. *rbc10*, DNA-abhängige RNA-Polymerase I, II und III 7,7 KD Untereinheit (ABC10-alpha); *rip1*, Rev interagierendes Protein Rip-1; *GTP-bp*, hypothetisches GTP-bindendes Protein; *rpl24*, 60S ribosomales Protein L24 c, komplementärer Strang; nt, Nucleotide. Die Transkriptionsrichtung wird durch Pfeile wiedergegeben. Die Pfeilspitzen stellen das 3'-Ende dar. Die Lokalisation bezieht sich auf die in der Datenbank hinterlegte Sequenz des Chromosoms II mit der Accession Nr. AJ010592.

#### 3.2.2.4. Identifikation weiterer RNA-Spezies

Es konnten zwei small nucleolar RNAs (snoRNAs) auf Chromosom II identifiziert werden (T. Lowe, pers. Mitteilung). *snR1* (Die Numerierung *snR1* bis *snR5* ist willkürlich festgelegt) ist 60 nt lang und liegt zwischen *prs4* und *rpb3* auf dem komplementären Strang bei 100757–100816 bp in Bezug auf Abbildung 3.2.. Die intergenischen Regionen haben somit eine Größe von 18 bp zu *prs4* und 1 bp zu *rpb3*.

Die zweite, 64nt lange snoRNA, *snR4* liegt zwischen *nmt1* und dem plastidären Chaperon *cpn60*. Auch *snR4* ist auf dem komplementären Strang codiert (119942–120005 bp), wobei die beiden codierenden Bereiche von *snR4* und *nmt1* zwei Nucleotide in ihrem jeweiligen 3'-Terminus gemeinsam nutzen (Abb. 3.6.). Die intergenische Region zwischen dem 3'-Terminus von *cpn60* und dem 5'-Bereich von *snR4* beträgt 52 bp. Eine Untersuchung der vollständigen cDNA von *cpn60* (Wastl et al., 1999) ergab, dass *snR4* auf dem 3'-untranslatierten Bereich der cDNA lokalisiert ist und somit co-transkribiert wird. Der 3'-untranslatierte Bereich der cDNA von *cpn60* überlappt zudem mit dem 3'-Terminus von *nmt1* um 186 bp (Abb. 3.6.).

Auf Chromosom III konnten zwei, auf Chr I eine weitere snoRNAs identifiziert werden (T. Lowe, pers. Mitteilung).

Die Sequenzen der snoRNAs sind im Anhang aufgelistet.

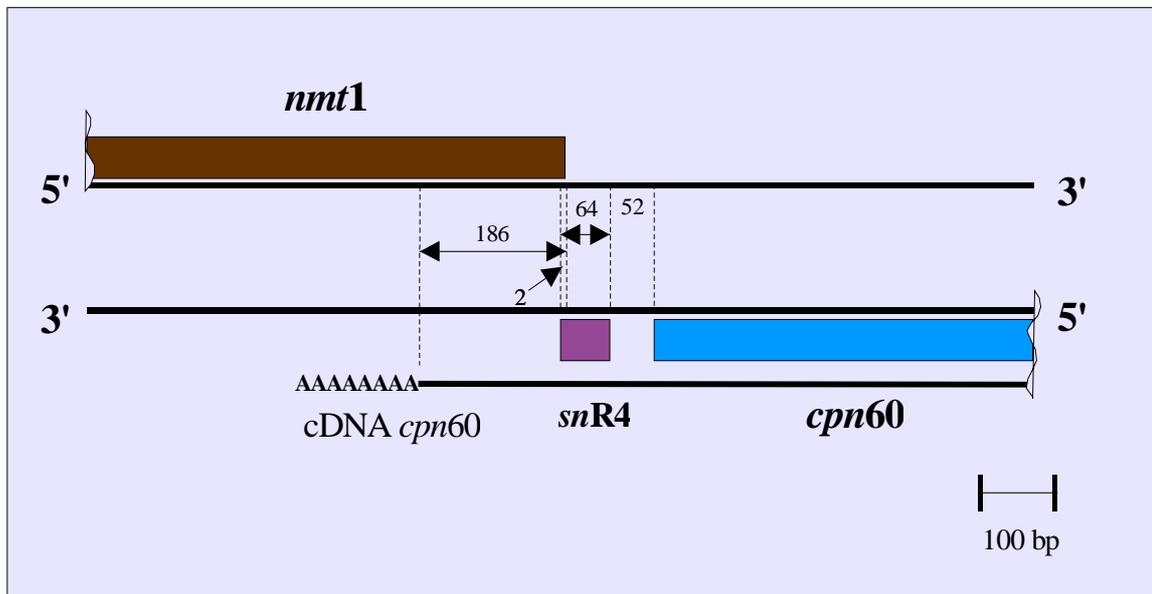


Abb. 3.6.: Lokalisation von *snR4* in der 3'-nichttranslatierten Region der cDNA von *cpn60* und Darstellung der Überlappung des 3'-Bereichs von *snR4* mit dem 3'-Bereich von *nmt1*. Dargestellt ist der DNA Doppelstrang des Chr II im Bereich zwischen circa 119300 und 120550 bp. *nmt1*, N-Myristoyltransferase; *snR4*, small nucleolar RNA 4; *cpn60*, plastidäres Chaperon der 60 KD-Klasse; **cDNA *cpn60***, komplementäre DNA von *cpn60*, die cDNA ist durch eine waagrechte durchgezogene Linie mit poly-A an ihrem 3'-Terminus angedeutet. 5' und 3' geben die Orientierung des DNA Doppelstrangs wieder. Die Farben entsprechen den Kategorien aus Abbildung 3.2. Die Zahlen zwischen den senkrechten gestrichelten Linien stellen die Größe der jeweiligen Regionen in bp dar. Näheres siehe Text.

### 3.3. Nucleomorph Chromosomen I und III

Nach der vollständigen Sequenzierung des Nucleomorph Chromosoms II wurden die frei gewordenen Kapazitäten genutzt, um durch die Schließung letzter Sequenzlücken zwischen den Contigs der Chromosomen I und III diese Contigs in Ihrer räumlichen Lage zueinander zu bestimmen.

Zu diesem Zweck wurden die bereits zu Contigs zusammengefassten Sequenzen einschließlich der Originaldaten der einzelnen Sequenzierreaktionen zur Verfügung gestellt. Verbliebene Sequenzlücken sollten analog zu Chr II mittels spezifischer Amplifikation durch PCR kloniert und anschließend sequenziert werden (Kap. 3.1.4.). Die entstandenen Konsensussequenzen wurden Korrektur gelesen und, soweit erforderlich, zur Doppelstrang-Sequenzierung bzw. zur Absicherung der Gesamt-Sequenz weitere Sequenzierungen durchgeführt.

#### 3.3.1. Nucleomorph Chromosom I

Die Chromosom I spezifischen Sequenzen waren in sieben Contigs zusammengefasst. Ihre Größen lagen zwischen 6212 bp und 77673 bp und wiesen eine Gesamtlänge von 190642 bp auf, dies entspricht ca. 95,3 % bezogen auf die durch Pulsfeld-Gelelektrophorese (Kap. 2.2.3.) abgeschätzte Größe des Chr I von 200 kb (Zauner et al., 2000). Die räumliche Anordnung und Orientierung der sieben Konsensussequenzen konnte durch mehrere PCR-Ansätze unter Verwendung endständiger Primer (Kap. 3.1.4.) aufgeklärt werden. Die PCR-Produkte wurden kloniert (Kap. 2.2.6.) und nach der Isolierung der rekombinanten Plasmide (Kap. 2.2.2.1.) sequenziert (Kap. 2.2.7.). Für das Nucleomorph Chromosom I ergab sich, nach dem Korrekturlesen, eine Gesamtlänge von 196216 bp (siehe Karte im Anhang). Die Sequenz ist in den Datenbanken unter der Accession Nr AF165818 hinterlegt.

Das Chromosom I codiert für 150 potentiell Protein-codierende offene Leserahmen (davon 3 mit Intron) und neun tRNAs, wovon zwei ein Intron aufweisen (S. Douglas, pers. Mitteilung). Gendichte und A/T-Gehalte der verschiedenen Regionen (Abb. 3.7.) sind mit den entsprechenden Werten des Chromosoms II vergleichbar und bewegen sich in derselben Größenordnung (Kap. 3.4.). Die *inverted repeats* besitzen intrachromosomal eine Länge von 11957 bp und codieren nicht für das Ubiquitin konjugierende Enzym (*ubc4*). Die Sequenzübereinstimmung des mit L bezeichneten Chromosomenendes (die

Chromosomen wurden willkürlich orientiert, L bezeichnet das linke Chromosomenende bezogen auf die Karte aller Nucleomorph Chromosomen, welche im Anhang zu finden ist) mit den Chromosomenenden des Chr II beträgt hingegen 12707 bp und beinhaltet eine Kopie von *ubc4* (Abb. 3.8.). Die Sequenzübereinstimmung bricht abrupt nach dem letzten Nucleotid des Stop-Codons von *ubc4* auf dem L-Ende des Chromosoms I ab.

Abbildung 3.7. zeigt den A/T-Gehalt aller Nucleomorph Chromosomen graphisch zusammengefasst.

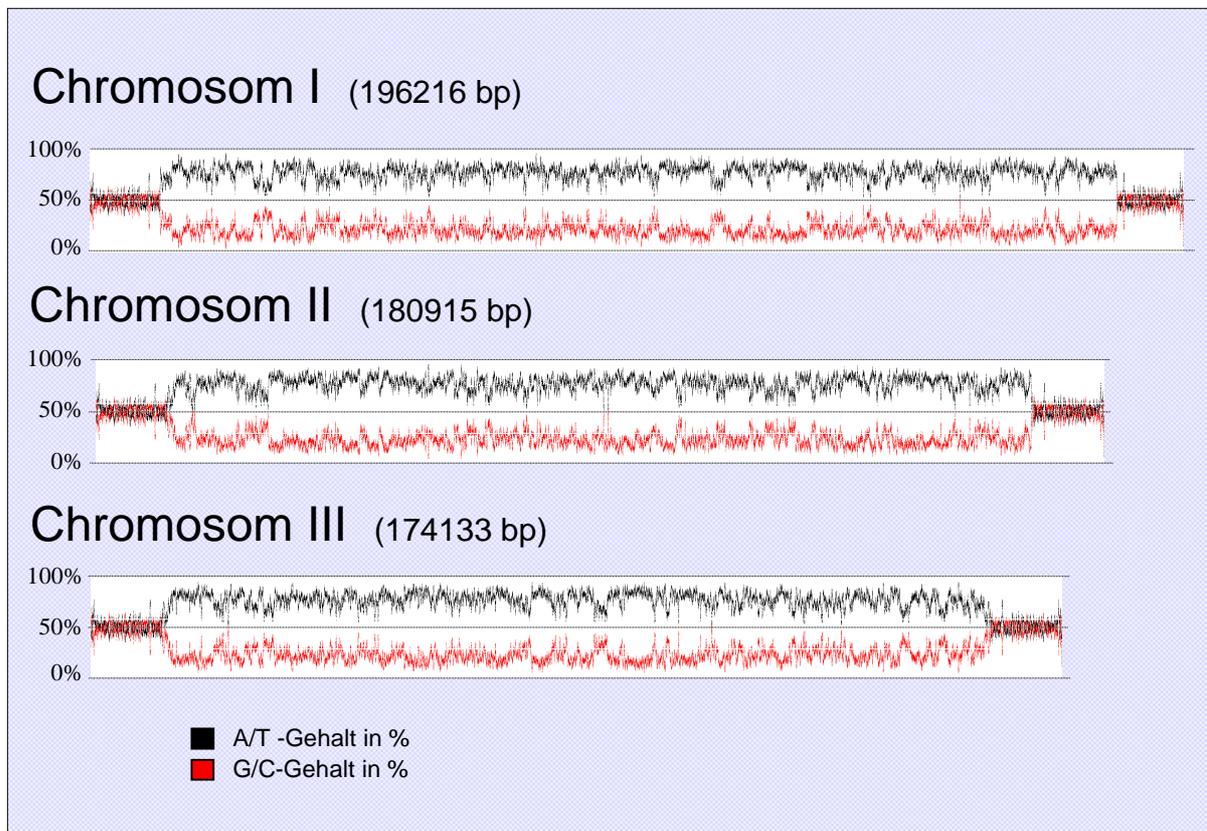


Abb. 3.7.: **Graphische Darstellung des A/T-Gehaltes aller Nucleomorph Chromosomen.** Die Zahlen der Y-Achse beziehen sich auf den A/T-Gehalt in %. Näheres siehe Text.

### 3.3.2. Nucleomorph Chromosom III

Analog zu den Arbeiten an Chromosom I wurde mit Contigs und Konsensussequenzen des Chromosoms III verfahren. Die zehn Contigs des Chromosoms III hatten Größen zwischen 2596 bp und 35877 bp. Die Gesamtlänge betrug 170978 bp, dies entspricht 97,7 % bezogen auf die durch Pulsfeld-Gelelektrophorese (Kap. 2.2.3.) abgeschätzte Größe des Chr III von 175 kb (Zauner et al., 2000). Die räumliche Anordnung und Orientierung der Konsensussequenzen konnte durch PCR (Kap. 3.1.3.6.) aufgeklärt werden. Für das Nucleomorph Chromosom III konnte eine 174133 bp lange Konsensussequenz ermittelt werden (siehe Karte im Anhang). Die Sequenz ist in den Datenbanken unter der Accession Nummer Nr AF083031 hinterlegt.

Das Chromosom III, das kleinste der drei Nucleomorph Chromosomen, codiert für insgesamt 155 potentiell Protein-codierende offene Leserahmen, von denen acht ein Intron besitzen. Weiterhin werden durch das Chr III 15 tRNAs (fünf mit Intron) codiert (S. Douglass, pers. Mitteilung). Auch für Chromosom III gilt, dass die Gendichte und die A/T-Gehalte der einzelnen Regionen (Abb. 3.8.) der Tendenzen der Chromosomen I und III folgt. Die *inverted repeats* besitzen intrachromosomal eine Länge von 14029 bp und beherbergen je einen offenen Leserahmen codierend für *ubc4* und für *tfIIId* (Abb. 3.8.). Interchromosomal hingegen ist die Sequenzidentität zwischen dem L-Ende des Chromosoms III und dem L-Ende des Chromosoms II 14346 bp lang, wobei der Abschnitt der identischen Sequenzen nach *tfIIId* auf Chr II für einen Teil des *orf673* codiert (Abb. 3.8.). Auf Chr III konnte diesem Sequenzabschnitt keine Codierungskapazität zugeordnet werden.

Der A/T-Gehalt der subtelomerischen Sequenzen aus Abbildung 3.8. unterscheidet sich signifikant zwischen Regionen interchromosomaler Sequenzübereinstimmung und intrachromosomaler *single-copy* Region. So weist zum Beispiel der Bereich der intrachromosomalen *single-copy* Region des R-Terminus von Chr I zwischen 11957 bp und 12707 bp einen A/T-Gehalt von 78,6 % auf. Der entsprechende Bereich auf dem L-Ende des Chr I ist Bestandteil der interchromosomalen Sequenzübereinstimmung und weist nur einen A/T-Gehalt von 48 % auf. Gleiches gilt für die Regionen des R-Terminus auf Chr I von 11987 bp bis 14346 bp mit einem A/T-Gehalt von 82,9 % gegenüber einem A/T-Gehalt von 58 % im selben Bereich auf dem L-Ende des Chr III.



### 3.4. Zusammenfassende Darstellung der strukturellen Daten der Nucleomorph Chromosomen von *G. theta*

	Chromosom I	Chromosom II	Chromosom III
Gesamtlänge [bp]	<b>196216</b>	<b>180915</b>	<b>174133</b>
A/T-Gehalt [%]	<b>74,25</b>	<b>73,29</b>	<b>73,05</b>
A [nt]	72745	65994	63570
C [nt]	25391	24247	23316
G [nt]	25134	24081	23612
T [nt]	72945	66593	63635
<i>Single copy</i> [bp]	<b>172242</b>	<b>154855</b>	<b>146075</b>
A/T-Gehalt [%]	<b>77,46</b>	<b>77,01</b>	<b>77,10</b>
intrachromosomale <i>inverted repeat</i> [bp]	<b>11987</b>	<b>13030</b>	<b>14029</b>
A/T-Gehalt [%]	<b>51,08</b>	<b>51,14</b>	<b>52,33</b>
Intergenische Regionen A/T-Gehalt [%]	<b>n.b.</b>	<b>82,28</b>	<b>n.b.</b>
Anzahl identifizierter <i>orfs</i>	<b>150</b>	<b>152</b>	<b>153</b>
Ident. <i>orfs</i> in <i>single-copy</i>	<b>150</b>	<b>150</b>	<b>149</b>
Gendichte bzgl. <i>single copy</i>	<b>1 / 1148 bp</b>	<b>1 / 1032 bp</b>	<b>1 / 980 bp</b>
<i>orfs</i> mit Introns	<b>3</b>	<b>8</b>	<b>6</b>
A/T-Gehalt der Introns [%]	<b>80,71</b>	<b>85,5</b>	<b>81,87</b>
Länge der Introns	<b>43-49</b>	<b>42-48</b>	<b>44-55</b>
Anzahl der überlappenden Protein-codierenden <i>orfs</i>	<b>10</b>	<b>7</b>	<b>4</b>
tRNA's überlappend mit Protein-codierenden <i>orfs</i>	<b>2 tRNAs</b>	<b>5 tRNAs</b>	<b>1 tRNA</b>
Anzahl identifizierter tRNAs	<b>9</b>	<b>13</b>	<b>15</b>
davon mit Introns	<b>2</b>	<b>6</b>	<b>5</b>
Anzahl snoRNAs	<b>1</b>	<b>2</b>	<b>2</b>

Tab. 3.7.: Zusammenfassung der ermittelten Daten über Chromosom II im Kontext aller Nucleomorph Chromosomen. n.b., nicht bestimmt.

## 4. DISKUSSION

Am Ende dieser Arbeit steht die vollständige Sequenzierung des Nucleomorph Chromosoms II der Cryptomonade *Guillardia theta*. Des weiteren konnten an den Chromosomen I und III abschließende Arbeiten zur Ermittlung und Absicherung der Gesamtsequenz durchgeführt werden.

### 4.1. Der strukturelle Aufbau des Nucleomorph-Genoms von *Guillardia theta*

Das Nucleomorph-Genom von *Guillardia theta* besteht aus drei linearen Chromosomen mit einer Gesamtgröße von 551264 bp, wobei auf das Chromosom I 196216 bp, auf das Chromosom II 180915 bp und auf das Chromosom III 174133 bp entfallen.

Nucleomorphe sind bisher nur bei Cryptomonaden und Chlorarachniophyten beschrieben (Ludwig und Gibbs, 1987). Dabei fällt auf, dass alle bisher untersuchten Nucleomorphe der Cryptomonaden sowie die Nucleomorphe der Chlorarachniophyte *Pedinomonas minutissima* genau drei Chromosomen besitzen (Rensing et al., 1994; Gilson und McFadden, 1996; McFadden et al., 1997). Es konnte gezeigt werden, dass jeweils alle drei Nucleomorph Chromosomen rDNA codieren (Rensing et al., 1994; Gilson et al., 1997).

#### 4.1.1. Die *inverted repeats* sind für die strukturelle Stabilität und die Erhaltung der Telomere der Nucleomorph Chromosomen verantwortlich

Der repetitive, subtelomerische Bereich der Nucleomorph Chromosomen von *G. theta* codiert 5S rRNA und die vollständige rDNA-Transkriptionseinheit aus 18S, 5.8S und 28S rDNA. Diese identifizierten subtelomerischen Regionen sind auf allen drei Chromosomen des Nucleomorph-Genoms existent. Zudem ist auf fünf von sechs Chromosomenenden ein Gen für das Ubiquitin-konjugierende Enzym (*ubc4*) und auf drei Chromosomenenden zusätzlich je ein Gen für das TATA-Box bindende Protein (*tfIIId*) lokalisiert (Abb. 3.8.).

Für die Nucleomorphe der Chlorarachniophyte *Pedinomonas minutissima* ist gezeigt, dass auf dem Chromosom III eine 8.5 kb große, terminale Sequenzwiederholung existiert (Gilson und McFadden, 1996). Diese codiert ebenso jeweils eine rDNA-Transkriptionseinheit, die jedoch in "centromerischer" Transkriptionsrichtung und somit entgegengesetzt zu den Transkriptionseinheiten der Nucleomorph Chromosomen der Cryptomonade *G. theta* orientiert sind (Gilson und McFadden, 1996; Gilson et al., 1997). Diesen subtelomerischen *inverted repeats* der Chlorarachniophyte *P. minutissima* fehlt, im Gegensatz zu denen des Nucleomorphs von *G. theta*, der für die 5S rRNA codierende Bereich. Ob sich die invertierten Sequenzwiederholungen im Falle von *P. minutissima* inter- bzw. intrachromosomal auch auf Protein-codierende offene Leserahmen erstrecken, ist zur Zeit noch nicht bekannt. Dies wird sich bei weiterem Fortschreiten der Sequenzierung dieses Nucleomorph-Genoms zeigen.

Auch von dem eukaryoten intrazellulären Parasiten *Encephalitozoon cuniculi* (Phylum Microspora) ist beschrieben, dass alle 11 Chromosomen des Nucleus palindromisch angeordnete rDNA-Transkriptionseinheiten in subtelomerischen, circa 15 kb großen *inverted repeats* besitzen (Brugere et al., 2000). *Encephalitozoon cuniculi* zeichnet sich als intrazellulärer Parasit durch ein nur 2,8 Mb großes Genom aus, dessen 11 Chromosomen eine Größe zwischen 210 und 304 kb aufweisen.

Die Aufteilung der linearen Nucleomorph Chromosomen von *G. theta*, *P. minutissima* und der nucleären Chromosomen von *E. cuniculi* in eine *inverted repeat* und eine *single-copy* Region erinnert an die ringförmigen Genome der Chloroplasten, die Plastome. Dort unterteilt die repetierte Sequenz das Genom in eine große und eine kleine *single-copy* Region. Die *inverted repeats* in Plastomen codieren ebenfalls rDNA-Transkriptionseinheiten. Möglicherweise stellt die Anordnung der DNA in repetierte und *single-copy* Bereiche ein Stabilitätskriterium besonders kleiner Chromosomen dar.

Der Aufbau der Nucleomorph Chromosomen in subtelomerische *inverted repeats*, die eine *single-copy* Region einschließen, könnte auch in bezug auf die Erhaltung der Telomere und somit dem Schutz der DNA-Termini eine wichtige Rolle spielen. Die ungewöhnlichen Telomere der Nucleomorph Chromosomen von *G. theta* mit mindestens 14 Repetitionen der Sequenz [(AG)<sub>7</sub>AAG<sub>6</sub>A] (G-reicher Strang in 5'-3'-Richtung notiert) (Zauner et al, 2000), werden vermutlich nicht durch eine Telomeraseaktivität aufrechterhalten. So konnte auf den Nucleomorph Chromosomen weder ein offener

Leserahmen für die reverse Transkriptase-Untereinheit noch eine für den RNA-Anteil der aktiven Telomerase codierende Region identifiziert werden. In einigen anderen Organismen konnte eine Telomerase unabhängige Aufrechterhaltung der Anzahl an Telomerwiederholungen und der Telomersequenz beobachtet werden (Biessmann und Mason, 1997). Für *S. cerevisiae* konnte gezeigt werden, dass im Falle der Inaktivierung der Telomerase Rekombinationsvorgänge die Telomere aufrechterhalten und auf diese Weise die Telomeraseaktivität ersetzt werden kann. Voraussetzung für derartige Rekombinationsvorgänge ist die Präsenz von subtelomerischen Sequenzwiederholungen. Solche subtelomerischen *repeats* auf unterschiedlichen Chromosomen konnten unter anderem in *Drosophila melanogaster* (Mason und Biessmann, 1995), *S. cerevisiae* (Bussey, 1995) und *Plasmodium falciparum* (Bowman et al, 1999; Gardner et al, 1998) nachgewiesen werden. Ähnliches wird für die linearen mitochondrialen Genome, wie sie in manchen Ciliaten vorkommen, postuliert (Nosek et al., 1998).

Demnach könnten auch die Telomere der Nucleomorph Chromosomen von *G. theta* ohne Telomeraseaktivität in Anzahl der Wiederholungen und ihrer Nucleotidsequenz durch Rekombinationsvorgänge aufrechterhalten werden. Diese rekombinatorischen Ereignisse müssten nicht auf die Telomere beschränkt sein, sondern könnten vielmehr auch zur Aufrechterhaltung der gesamten subtelomerischen *inverted repeats* führen. Durch intra- bzw. intermolekulare Rekombination homologer Sequenzen würden sich alle Chromosomenenden untereinander austauschen und somit ihre Sequenz gegenseitig fixieren. Dies könnte die extreme Konservierung der Sequenz der subtelomerischen DNA-Regionen plausibel erklären. Der Mechanismus dieser homologen Rekombination ist noch nicht aufgeklärt, könnte aber mit dem Prinzip der *holiday junction* (Holiday, 1964) erklärt werden. Holiday postulierte einen intermediären Zwischenschritt während der homologen Rekombination in Form einer „vier Wege Kreuzung“. Diese Struktur konnte mit Hilfe eines Palindrom-artigen Oligonucleotids als Kristall Struktur verifiziert werden (Eichmann, 1999).

Ein weiterer Aspekt, der die subtelomerischen Sequenzwiederholungen mit der Stabilität der Nucleomorph Chromosomen in Verbindung bringt, ist der unterschiedliche A/T-Gehalt der *inverted repeats* und der *single-copy* Region. Der A/T-Gehalt ist in der *single-copy* Region der Nucleomorph Chromosomen mit 77,19 % signifikant höher als in den subtelomerischen Sequenzwiederholungen einschließlich der rDNA-Transkriptionseinheiten mit 51,51 % Adenin und Thymin. Dieser durchschnittliche A/T-

Gehalt der subtelomerischen Sequenzwiederholungen spiegelt die Tatsache wider, dass die rDNA keine Präferenz zeigt Adenine und Thymin anzureichern und sich in ihrer Basenzusammensetzung meist mehr oder weniger deutlich vom restlichen Genom unterscheidet (Muto und Osawa, 1987). Dies ist auf die Struktur und Funktion der rRNA-Moleküle und ihre damit verbundenen intra- und intermolekularen Interaktionen zurückzuführen. Zudem wird dieser signifikante Unterschied im A/T-Gehalt durch die bereits erwähnten möglichen inter- und intrachromosomalen Rekombinationsereignisse fixiert. Die relativ (bezogen auf die *single-copy* Region) G/C reichen *inverted repeats* bilden somit eine G/C Klammer um die A/T reiche *single-copy* Region (Abb. 3.7.). In den Nucleomorphen der Chlorarachniophyte *P. minutissima* stellen sich die Verhältnisse ähnlich dar. Hier zeichnen sich die subtelomerisch lokalisierten invertierten Sequenzwiederholungen durch einen A/T-Gehalt von 46 % aus während die *single-copy* Region des Chromosoms III von *P. minutissima* einen A/T-Gehalt von circa 73 % aufweisen (P. R. Gilson, pers. Mitteilung).

Die besondere Struktur der *inverted repeats* könnte somit erheblichen Einfluss auf die strukturelle Stabilität besonders kleiner linearer Chromosomen haben und gleichzeitig durch rekombinatorische Ereignisse für die Aufrechterhaltung der Telomere verantwortlich sein.

#### 4.1.2. A/T-Gehalt und der Einfluss auf die Codon-Nutzung

Die Analyse der ermittelten Nucleotidsequenzen der Nucleomorph Chromosomen I, II und III von *Guillardia theta* ergibt einen durchschnittlichen A/T-Gehalt von 73,53 %. Betrachtet man die beiden großen DNA-Regionen, erhält man mit 51,51 % in der *inverted repeat* Struktur einen relativ niedrigeren Anteil an Adenosin und Thymin, als in der *single-copy* Region, deren Sequenzen im Durchschnitt zu 77,19 % aus A und T aufgebaut sind. Die genannten prozentualen Werte sind Durchschnittswerte aller drei Nucleomorph Chromosomen. Wie der Tabelle 3.7. zu entnehmen ist, liegen die Werte für die einzelnen Chromosomen in allen Bereichen innerhalb eines sehr engen Rahmens. Einzelne, kleinere Unterschiede des A/T-Gehaltes in vergleichbaren Regionen unterschiedlicher Chromosomen lassen sich auf das unterschiedliche Verhältnis der *inverted repeats* zu der *single-copy* Region in Bezug auf die Gesamtlänge des Chromosoms zurückführen. So nimmt zum Beispiel die Länge der intrachromosomalen Sequenzwiederholungen bei gleichzeitig abnehmender Chromosomenlänge (Chromosomen I → III) prozentual zu, und

somit ihr Anteil am gesamten A/T-Gehalt. Folglich sinkt der A/T-Gehalt von 74,25 % im Chr I bis auf 73,05 % im Chr III.

Der Übergang der *inverted repeats* zur *single-copy* Region und die damit einhergehende unterschiedliche Verteilung von A/T bzw. G/C ist in der graphischen Gegenüberstellung von A/T-Gehalt bzw. G/C-Gehalt (Y-Achse) gegen die Nucleotidposition (X-Achse) in Abbildung 3.7. verdeutlicht.

Die potentiell Protein-codierenden offenen Leserahmen innerhalb der *single-copy* Region des Chromosoms II zeigen einen durchschnittlichen A/T-Gehalt von 76,80 %, dem gegenüber bestehen die intergenischen Regionen (bezogen auf die Initiations-Codons der Leserahmen) zu 82,28 % aus A und T. Die ermittelten Werte für das Nucleomorph der Cryptomonade *G. theta* stehen in Übereinstimmung mit Ergebnissen aus dem Nucleomorph Genomprojekt des Chlorarachniophyten *P. minutissima* (Gilson und McFadden, 1996).

Der Unterschied des A/T-Gehaltes in codierenden und intergenischen Bereichen ist darauf zurückzuführen, dass die Proteine auf Funktionalität selektiert werden und ihre Gene somit nicht beliebig mutieren, d.h. in diesem Zusammenhang nicht beliebige Codons nutzen können, sondern innerhalb hochkonservierter, funktioneller Proteinbereiche synonyme Codons genutzt werden müssen. Diese können nur in einem gewissen Rahmen einen höheren Anteil an A und T besitzen. Ist eine Aminosäure in einer funktionellen Position unabdingbar, kann ihr Codon nur gegen ein synonymes ausgetauscht werden, um die Funktion des Proteins nicht zu beeinflussen (z.B. Cystein: 80,9 % TGT gegenüber 19,1 % TGC). In einem Genom, das generell auf einen hohen Gehalt an A und T selektiert wird, wird in der *wobble*-Position bevorzugt A oder T eingebaut werden (z.B. Andersson und Sharp, 1996).

Das Nucleomorph-Genom von *G. theta* hat eindeutig die Tendenz, A und T zu akkumulieren. Nicht stringent selektierte Regionen von Proteinen werden gemäß dieser gewichteten, kompositionellen Zusammensetzung (*compositional bias*) des Genoms Codons anhäufen, die A/T reich sind. In den nicht konservierten Regionen sollten also Aminosäuren (Asn, Ile, Leu, Lys, Phe, Tyr) vermehrt auftreten, die auf diese Codons zurückgehen, sofern nicht Ladung oder Größe der codierten Aminosäure im Widerspruch zur Funktion oder Struktur der Proteine stehen. Ein Beispiel hierfür könnten die N-terminalen Extensionen der Nucleomorph-codierten, plastidär lokalisierten ORFs

darstellen (Wastl; 2000). Generell sollten die codierenden Bereiche hochkonservierter Moleküle weniger frei evolvieren können, da Funktion, Struktur oder Interaktion mit anderen Molekülen davon beeinflusst würde. Die codierenden Nucleotidsequenzen dieser Bereiche sollten demnach weniger A und T akkumulieren als DNA-Sequenzen mit „evolutionärem Spielraum“. Da der A/T-Gehalt der Nucleomorph Chromosomen sich regional unterscheidet (erstens die invertierten Sequenzwiederholungen mit einem A/T-Gehalt von durchschnittlich 51,51 %, zweitens die für die Erhaltung des Nucleomorphs codierenden offenen Leserahmen der *single-copy* Region mit ca. 78 % und drittens die tRNA Gene und die *orfs* plastidär lokalisierter Proteine mit einem intermediärem A/T-Gehalt von ca. 65 %), müssen verschiedene Evolutionsdrücke bezüglich der Basen Komposition dieser drei DNA-Populationen angenommen werden. So lassen sich die in der Abbildung 3.7. erkennbaren kurzen Abfälle im A/T-Gehalt der *single-copy* Region gut mit tRNA Genen, offenen Leserahmen cyanobakteriellen Ursprungs mit potentiellen N-terminalen Extensionen oder einzelnen hochkonservierten Proteindomänen korrelieren (Daten nicht gezeigt).

In der Tabelle 3.4. ist die Codon Nutzung (*codon usage*) der in dieser Arbeit ermittelten potentiell Protein-codierenden offenen Leserahmen der *single-copy* Region des Nucleomorph Chromosoms II dargestellt. Darin werden die Anteile der genutzten Codons an der Gesamtzahl der jeweilig codierten Aminosäure in Prozent angegeben. Aus der Tabelle geht hervor, dass die Präferenz des Einbaus von A und T Nucleotiden auch die Codon Nutzung beeinflusst. Es werden jeweils bevorzugt die Codons verwendet, die A und/oder T reicher sind als ihre Synonyme. Die Codon Nutzung ist also deutlich gewichtet. So werden in Protein-codierenden Bereichen des Nucleomorph Chromosoms II von *G. theta* 86,1 % der Glycine (synonyme Codons: GGN) durch GGA oder GGT und 86,7 % der Alanine (GCN) durch GCA bzw. GCT codiert. Bei Aminosäuren mit zwei synonymen Codons werden die A bzw. T reicheren Varianten in mehr als 80 % der Fälle gegenüber den G bzw. C reichen Varianten bevorzugt. Lysin wird zum Beispiel in 87,6 % aller Fälle durch das Codon AAA repräsentiert und nur zu 12,4 % durch AAG. Gleiches gilt für die Aminosäuren mit sechs synonymen Codons, so wird zum Beispiel Arginin zu 73,2 % durch das Codon AGA codiert.

Analysiert man die verwendeten Codons der potentiell Protein-codierenden *orfs* der *single-copy* Region auf A/T-Gehalt an den einzelnen Codonpositionen, zeigt sich, dass 72,96 % in der ersten Codon Position A oder T enthalten, 75,16 % in der zweiten und

82,08 % in der dritten (*wobble*) Position. Dies entspricht nach Muto und Osawa (1987) den erwarteten Werten für A/T reiche Genome.

Für das Gen des Ubiquitin-konjugierenden Enzyms (*ubc*) und das Gen des TATA-Box bindenden Proteins (*tfIIId*) zeichnet sich ein anderes Bild ab. In diesem Fall werden Codons genutzt die nur zu 56,22 % (erste Codon Position), zu 61,04 % (zweite Codon Position) und zu 63,45 % (*wobble*-Position) A oder T Nucleotide enthalten. Dies erklärt sich wiederum aus der Tatsache, dass diese Gene Bestandteile der *inverted repeats* darstellen.

Ein Vergleich des A/T-Gehaltes in den verschiedenen Codonpositionen der codierenden Regionen von einigen Vertretern der Rot- und Grünalgen, Organellen, Endosymbionten und obligat intrazellulären Parasiten ist in Tabelle 4.1. wiedergegeben.

Organismus	A/T [%] cod. Regionen	A/T [%] Codon Position		
		1.	2.	3.
<i>Porphyra purpurea</i> (NU)	36,75	41,83	57,14	11,29
<i>Gracilaria gracilis</i> (NU)	48,82	44,30	59,98	42,18
<i>Chlamydomonas reinhardtii</i> (NU)	34,91	36,88	54,57	13,38
<i>Synechocystis sp.</i>	51,34	43,92	60,08	50,01
<b><i>Guillardia theta</i> (NM)</b>	<b>76,80</b>	<b>72,96</b>	<b>75,16</b>	<b>82,08</b>
<i>Buchnera aphidicola</i>	72,11	61,9	68,65	85,78
<i>Rickettsia prowazekii</i>	69,40	58,61	68,06	81,53
<i>Plasmodium falciparum</i> (NU)	74,07	66,05	74,35	82,82
<i>Guillardia theta</i> (PL)	66,83	54,84	64,15	81,49
<i>Porphyra purpurea</i> (PL)	66,25	55,22	63,96	79,56
<i>Chlamydomonas reinhardtii</i> (PL)	65,77	55,31	62,02	79,99
<i>Odontella sinensis</i> (PL)	67,45	55,30	64,29	82,77

Tab. 4.1.: A/T-Gehalt der verschiedenen Codon Positionen in codierenden Regionen der Genome von Rot- bzw. Grünalgen, Cyanobakterien, Chloroplasten, Endosymbionten und obligat intrazellulären Parasiten. *P. purpurea* und *G. gracilis* sind freilebende Rotalgen und als solche mit dem evolutionären Vorläufer des Symbionten der Cryptomonaden verwandt. Als weitere Beispiele freilebender Organismen wurde die Grünalge *C. reinhardtii* (NU) und das Cyanobakterium *Synechocystis sp.* mit in die Tabelle aufgenommen. Plastiden-Genome (*G. theta* (PL), *P. purpurea* (PL), *C. reinhardtii* (PL) und *O. sinensis* (PL)), intrazelluläre Symbionten (*G. theta* (NM), *B. aphidicola*) und obligat intrazelluläre Parasiten (*P. falciparum* (NU), *R. prowazekii*) nutzen zu ca. 80 % Codons mit A oder T in der *wobble*-Position. Alle Werte mit Ausnahme von *G. theta* (NM) sind aus <http://www.kazusa.or.jp/codon/> entnommen. **PL**, Plastide; **NU**, Nucleus; **NM**, Nucleomorph.

Phylogenetische Studien anhand von 18S rRNA Daten konnten zeigen, dass die rezenten, freilebenden Rhodophyta einen gemeinsamen Vorfahren mit dem Endosymbiont von *G. theta* aufweisen (Van de Peer et al., 1996). Des Weiteren kann aus diesen Untersuchungen ein gemeinsamer Vorfahre der Chlorophyta mit den Endosymbionten der Chlorarachniophyten abgeleitet werden. Die Daten für die nucleären codierenden Bereiche der Rotalgen in der Tabelle 4.1. sind nicht statistisch abgesichert (die Werte für *P. purpurea* (NU) beruhen auf 1939 Codons, die von *G. gracilis* (NU) auf 4490), es ist jedoch die Tendenz zu erkennen, dass die codierenden Bereiche rezenter, freilebender Rot- und Grünalgen nicht generell A/T reich sind. Die Basen-Komposition der Nucleomorphe der Cryptomonaden sowie der Chlorarachniophyten ist demnach unabhängig voneinander sekundär entstanden: Durch selektiven Einbau von A oder T in der *wobble*-Position der Codons, bzw. durch die Nutzung von Aminosäuren mit generell A/T reichen Codons in den nicht stringent selektierten Bereichen der reifen Proteine.

Generell ist die Akkumulation von A und T in Genomen intrazellulärer Symbionten und Parasiten (Moran und Baumann, 2000) oder in Organellen-Genomen zu beobachten. Bakterielle intrazelluläre Symbionten der Gattung *Buchnera* besitzen ebenso A/T reiche Genome (Moran, 1996) wie das Bakterium *Rickettsia prowazekii* (Andersson et al., 1998) oder der Eukaryot *Plasmodium falciparum* (Gardner et al., 1998), ein obligat intrazellulärer Parasit des Menschen. Auch mitochondriale (Gray et al., 1998) und plastidäre Genome zeichnen sich durch einen hohen A/T-Gehalt aus. So besitzt das Plastiden-Genom von *G. theta* einen A/T-Gehalt von 67 % (Douglas und Penny, 1999), das von *Chlorella vulgaris* 68.4 % (Wakasugi et al., 1997). In derselben Größenordnung bewegen sich auch die Plastome von *Odontella sinensis* mit 68.8 % (Kowallik et al., 1995), *Porphyra purpurea* mit 67 % (Reith und Munholland, 1995), das Cyanellen-Genom von *Cyanophora paradoxa* 69.6 % (Stirewalt et al., 1995) und die Plastome höherer Pflanzen.

Da Chlorarachniophyten und Cryptomonaden aus Symbiosen unterschiedlicher Wirte mit unterschiedlichen Symbionten entstanden sind (Cavalier-Smith, 2000) und dieselbe Tendenz zur Akkumulation von A und T in den Genomen intrazellulär lebender Organismen (obligat oder fakultativ) und der Organellen festzustellen ist, scheinen ähnliche Mechanismen für eine Selektion A/T reicher Genome zu sorgen. Die Ursache der Unterschiede in der kompositionellen Basenzusammensetzung nucleärer und endocytobiotischer Genome ist noch nicht verstanden. Eine mögliche Ursache dieses

Phänomens könnte sein, dass A/T reiche Sequenzen schneller und effizienter repliziert werden können und somit die Koordination nucleärer und symbiontischer Replikation erleichtern. Eine weitere Möglichkeit wäre, dass sich A/T reiche Genome leichter „Muller's ratchet“, zum Beispiel durch erleichterte Eliminierung auftretender Mutationen, entziehen können.

H. J. Muller postulierte 1964 für sich asexuell reproduzierende Populationen eine Anhäufung von Mutationen im Genom, da die Wahrscheinlichkeit für eine Rückmutation geringer ist als die Wahrscheinlichkeit einer weiteren Mutation an anderer Stelle des Genoms („Muller's ratchet“). Die daraus folgende Anhäufung und Fixierung von Mutationen in den oben erwähnten Populationen führt zu einer geringeren „Fitness“ der betroffenen Organismen und letztendlich zur Letalität. „Muller's ratchet“ kann für bakterielle Endosymbionten wie zum Beispiel *Buchnera* (Moran, 1996) ebenso wie für die Nucleomorphe der Cryptomonaden und Chlorarachniophyten und für Organellen angenommen werden. Das Fehlen sexueller Rekombinationsvorgänge wird für die Anhäufung von nichtsynonymen Mutationen verantwortlich gemacht. Das Cytoplasma mit seiner konstanten Zusammensetzung sorgt allerdings für einen geringeren Selektionsdruck auf intrazellulär lebende Organismen oder Organellen, als die wechselnden Lebensbedingungen, denen freilebende Organismen ausgesetzt sind. Somit wirken diese Mutationen an neutralen oder nur schwach selektierten Positionen der Nucleotidsequenz nicht zwingend letal. Im Falle von *Buchnera* konnte zudem eine erhöhte Expressionsrate molekularer Chaperone gezeigt werden (Moran, 1996), die ebenfalls die Funktionalität der mutierten Genprodukte aufrechterhalten könnte. Das Chaperon HSP70 ist auf den Nucleomorph-Genomen von Chlorarachniophyten und Cryptomonaden, auch *G. theta*, codiert (Hofmann et al., 1994; Rensing et al., 1994) und auch andere Chaperonine wie etwa Vertreter der HSP90 und HSP60 Familien konnten auf den Nucleomorph Chromosomen von *G. theta* detektiert werden. Diese Faltungshelfer könnten bei der Aufrechterhaltung der Funktion von Proteinen assistieren, deren Gene durch den evolutionären Druck A und T Nucleotide anhäufen und Mutationen akkumulieren.

Generell können die angesprochenen Mechanismen aber nicht verhindern, dass letztendlich die zunehmende Zahl von Mutationen in Protein-codierenden Bereichen zur Zerstörung ihrer Funktion führen. Dies gilt ebenfalls für die *inverted repeats* mit ihren rDNA-Transkriptionseinheiten, auch hier sollten sich Baseninsertionen und – deletionen ebenso wie Basenaustausche anhäufen und zu einem höheren A/T-Gehalt dieser Region

führen. Unter Berücksichtigung der extremen Konservierung der Nucleotidsequenz in diesem Bereich, muss von einer aktiven Aufrechterhaltung des *Status quo* ausgegangen werden. Wird für die Konservierung der subtelomerischen Bereiche ein Rekombinationsmechanismus postuliert, ist dieser eventuell auch unter homologen Chromosomen aktiv; dies könnte zur Reversierung ansonsten letaler Mutationen führen. So wird das Nucleomorph der Cryptomonaden in der Literatur als mindestens diploid bezeichnet (Hansmann und Eschbach, 1990). Auch für Organellen, speziell tierische und menschliche Mitochondrien, werden in der Literatur Rekombinationsereignisse diskutiert (Lunt und Hyman, 1997; Hagelberg et al., 1999; Eyre-Walker, 1999).

Nicht zuletzt könnten Mutationen, die zum „Ausfall“ eines Proteins führen auch durch ein nucleär-codiertes Homolog, bei anschließender Eliminierung des mutierten Nucleomorph-Gens funktionell ersetzt werden. Damit würden sukzessive Wirtszellkomponenten die Funktion des periplastidären Kompartiments aufrechterhalten.

#### **4.1.3. Das Nucleomorph-Genom ist extrem kompakt**

Das Nucleomorph-Genom von *Guillardia theta* codiert - trotz der begrenzten Kapazität von 551 kb – für insgesamt 455 potentiell Protein-codierende offene Leserahmen, die strukturellen rRNA Moleküle, fünf *small nucleolar RNAs* (snoRNA), eine *small nuclear RNA* (snRNA) und 37 tRNAs. Anhand der Karte der codierenden Bereiche des Nucleomorph Chromosoms II in der Abbildung 3.2. und der Karte aller drei Nucleomorph Chromosomen im Anhang ist diese extreme Kompaktheit des Nucleomorph-Genoms deutlich zu ersehen.

Betrachtet man nur die 449 RNA-Polymerase II transkribierten offenen Leserahmen der *single-copy* Region, so ergibt sich eine durchschnittliche Gendichte von einem Gen pro 1054 Basenpaaren. Dieser Wert übertrifft die Gendichte von *S. cerevisiae* um das doppelte und die von *A. thaliana* um mindestens das vierfache. In *S. cerevisiae* beträgt die Gendichte etwa ein Gen pro 2 kb (Dujon, 1996), in *A. thaliana* etwa ein Gen pro 4.5 kb (Kotani et al., 1998; Kotani et al., 1997; Bevan et al., 1998). Die Berücksichtigung der 37 tRNAs, der fünf snoRNA und der snRNA steigert die Kompaktheit beträchtlich.

Die extreme Kompaktheit des Nucleomorph-Genoms lässt sich auf mehrere Faktoren zurückführen, bzw. äußert sich in für eukaryote Genome untypischen Erscheinungsformen.

#### 4.1.3.1. Die intergenischen Regionen sind sehr kurz, codierende Bereiche können überlappen

Die außerordentlich hohe Gendichte der Nucleomorph Chromosomen von *Guillardia theta* kommt vor allem in den extrem kurzen intergenischen Regionen zum Ausdruck, die in ihrer Länge zwischen 0 bp (Kap. 3.2.2.1.1.; auch Abb. 3.5. und Abb. 3.6.) und 338 bp variieren. Der rechnerische Durchschnitt beträgt 58 bp. Die 8750 bp der intergenischen Regionen stellen nur 5,65 % der *single-copy* Region dar. Der A/T-Gehalt der intergenischen Regionen ist mit 82,28 % gemäß der kompositionellen Basenzusammensetzung (*compositional bias*) extrem hoch. Unter Berücksichtigung der *inverted repeats* mit ihren rDNA-Clustern und den offenen Leserahmen für *ubc4* ergibt sich ein Anteil von 88,6 % codierender Bereiche in Bezug auf die Gesamtsequenz des Nucleomorph Chromosoms II.

Tabelle 4.2. zeigt eine Auflistung der prozentualen Anteile der codierenden Bereiche an den bisher veröffentlichten genomischen Regionen unterschiedlicher Organismen. Es ist zu erkennen, dass die Genome eukaryoter Organismen einen deutlich geringeren prozentualen Anteil an codierenden Bereichen aufweisen als prokaryote Organismen. Eine Ausnahme hierzu stellen *S. cerevisiae* und das Nucleomorph Chr II von *G. theta* dar. Ähnliches gilt für die Chromosomen I und III. Im direkten Vergleich zwischen Hefe und dem Nucleomorph Chr II von *G. theta* zeigt sich der höhere Prozentsatz codierender Regionen dennoch deutlich. Dies steht in Übereinstimmung mit der um das doppelte höheren Gendichte des Nucleomorph Chromosoms II gegenüber der Gendichte von *S. cerevisiae* (Kap. 4.1.3.). Der für ein eukaryotes Genom außerordentlich hohe Anteil codierender Bereiche ist Ausdruck der starken evolutionären Kraft, die die Komprimierung des Nucleomorph-Genoms bewirkt. Die starke Reduktion des Nucleomorph-Genoms geht in erster Linie auf die Eliminierung redundanter codierender Bereiche für Gene des Primär- und Sekundär-Metabolismus und die Eliminierung aller repetitiven Strukturen außerhalb der *inverted repeats* zurück. Einher ging eine extreme Reduktion der intergenischen Regionen.

<b>Organismus</b>	<b>codierender Bereich [%]</b>	<b>Bemerkung</b>
<i>Escherichia coli</i>	87	bezogen auf das ges. Genom
<i>Rickettsia prowazekii</i>	75	bezogen auf das ges. Genom
<i>Synechocystis sp.</i>	87	bezogen auf das ges. Genom
<b><i>Guillardia theta</i> (NM)</b>	<b>88,6</b>	<b>bezogen auf das Nucleomorph Chr II</b>
<i>Saccharomyces cerevisiae</i>	72	bezogen auf das ges. Genom
<i>Plasmodium falciparum</i>	ca. 50	bezogen auf die Chr. 2 und 3
<i>Caenorhabditis elegans</i>	21	bezogen auf das ges. Genom
<i>Drosophila melanogaster</i>	ca. 20	bezogen auf die bisher veröffentlichten Sequenzen
<i>Homo sapiens</i>	ca. 1	bezogen auf die Chr. 21 und 22

Tab. 4.2.: **Prozentualer Anteil der codierenden Bereiche innerhalb der Genome verschiedener Organismen.** Alle Werte mit Ausnahme *G. theta* sind <http://www.cbs.dtu.dk/services/GenomeAtlas/> entnommen.

Als extremste Form der Minimierung der intergenischen Regionen können die Bereiche des Genoms aufgefasst werden, in denen zwei potentiell Protein-codierende Bereiche dieselbe Nucleotidsequenz gemeinsam nutzen. Auf dem Nucleomorph Chromosom II von *G. theta* konnten sieben Regionen identifiziert werden, die von zwei potentiell Protein-codierenden Bereichen genutzt werden (Kap. 3.2.2.1.1.). Insgesamt wurden auf dem Nucleomorph-Genom 24 potentielle Überlappungen von RNA-Polymerase II transkribierten Genen nachgewiesen (S. Douglas, pers. Mitteilung). Die Anzahl der gemeinsam genutzten Nucleotide variiert zwischen einem und 56.

Bei insgesamt 14 Bereichen (bezogen auf das gesamt Nucleomorph-Genom, fünf davon auf Chr II) besteht die Überlappung im 3'-codierenden Terminus beider Beteiligten, wodurch die Annahmen alternativer Start-Codons ohne Auswirkung auf die gemeinsame Nutzung derselben Sequenzabschnitte bleibt. In sechs Fällen besteht die Überlappung im 5'-Bereich eines Partners (vier Fälle) bzw. beider Partner (zwei). Hier könnten alternative Start-Codons angenommen werden, so dass es sich nicht um echte Überlappungen der codierenden Bereiche handeln muss. Keine Überlappung der potentiell Protein-codierenden Bereiche der reifen Proteine stellen vier Fälle dar, in denen der 5'-Bereich eines potentiell plastidär lokalisierten, aber Nucleomorph-codierten Proteins mit dem 3'-

Terminus (zwei Fälle) bzw. dem 5'-Bereich (zwei) eines Protein-codierenden offenen Leserahmens, dessen Produkt nicht plastidär lokalisiert ist, überlappt. Plastidär lokalisierte Präproteine zeichnen sich durch den Besitz einer N-terminalen Extension aus. Diese N-terminale Extension ist nicht Bestandteil des reifen Proteins. Die ‚*in silico*‘ ermittelten potentiellen Überlappungen bestehen sechsmal zwischen tandemartig angeordneten- und achtzehnmal zwischen antiparallelen, konvergent bzw. divergent transkribierten Protein-codierenden Bereichen.

Überlappende Protein-codierende Bereiche sind in eukaryoten Genomen nicht bekannt und bisher nur bei Prokaryoten, Mitochondrien und viralen Genomen beschrieben (Krakauer, 2000).

Des Weiteren konnten auf dem Nucleomorph Chromosom II von *G. theta* fünf Regionen identifiziert werden (16 im Nucleomorph-Genom), die Überlappungen zwischen Genen darstellen, die von zwei unterschiedlichen RNA-Polymerasen (Pol II und Pol III) transkribiert werden (Kap. 3.2.2.2.1.). Die Anzahl der von einer tRNA und einem Protein-codierenden offenen Leserahmen gemeinsam genutzten Nucleotide reicht von einem bis 76. Von diesen 16 Überlappungen sind zehn zwischen Genen mit gleicher, fünf zwischen Genen konvergenter und eine zwischen Genen divergenter Transkriptionsrichtung lokalisiert. In zwei Regionen (beide Chr I) überlappt je eine tRNA an ihren Termini mit jeweils einem Protein-codierenden offenen Leserahmen. Die umgekehrte Situation stellt sich in je einer Region des Chr III und des Chr II dar. In diesen Regionen überlappen zwei tRNAs mit je einem Terminus eines RNA-Polymerase II transkribierten offenen Leserahmens (Abb. 3.5.).

Die Überlappungen scheinen nicht nur von den evolutionären Anforderungen an die codierenden Sequenzen abhängig zu sein, sondern auch von den Anforderungen an die Promotor-Regionen. In den Fällen einer Überlappung zweier RNA-Polymerase II transkribierter offener Leserahmen besteht diese zu 75 % zwischen antiparallelen offenen Leserahmen und hier in 14 von 18 Fällen zwischen konvergent transkribierten Genen. Das bedeutet, dass die beteiligten Partner in ihren jeweiligen 3'-Termini überlappen und somit die Promotor-Regionen nicht an dieser Überlappung beteiligt sind. In den Regionen in denen tandemartig angeordnete oder antiparallele und divergent transkribierte offene Leserahmen überlappen, sollte die Promotor-Region mindestens eines beteiligten Genes innerhalb des Protein-codierenden Bereiches des Zweiten zu finden sein. Die Promotor-

Region des mit dem 5'-Terminus überlappenden Genes wäre während der Transkription des zweiten beteiligten Genes maskiert und somit die Promotor-Bindung erschwert.

Die überlappenden codierenden Bereiche sind nur ein Sonderfall der insgesamt sehr kurzen intergenischen Regionen. So wird die Transkription der Nucleomorph-codierten offenen Leserahmen generell durch die extreme Kompaktheit der Chromosomen beeinflusst.

#### **4.1.3.2. Die Transkripte Nucleomorph-codierter, offener Leserahmen überlappen und die 3'-Prozessierung ist inakurat**

Unter Berücksichtigung der sehr kurzen intergenischen Regionen, mit Überlappungen als Extremfall, fällt auf, dass diese meist zu kurz sind, um alle für die Transkription nötigen Signale aufzunehmen (einschließlich die der *up-* bzw. *downstream* Gene). Ein Vergleich der 5'-Regionen der potentiell protein-codierenden offenen Leserahmen ergibt keine bekannten Konsensus-Motive für die Promotoren der RNA-Polymerase II abhängigen Transkription im Nucleomorph, die einen konstantem Abstand zur Initiation der Translation aufweisen. Ebenso konnten in den 3'-Bereichen keine Konsensussequenzen für Signale der posttranskriptionalen Prozessierung der prä-mRNAs (Termination der Transkription, Polyadenylierung) identifiziert werden (Daten nicht gezeigt). Auch zusätzliche Nucleomorph spezifische Konsensus-Motive konnten nicht identifiziert werden. Betrachtet man das Konsensus-Motiv der TATA-Box (TATAA/TA), das *core* Promotor Element der RNA-Polymerase II abhängigen Transkription, so fällt ihr hoher A/T-Gehalt auf. Gleiches gilt für das Polyadenylierungssignal mit der Konsensussequenz TTATTT im Template DNA-Strang bzw. AAUAAA in der mRNA. Unter Berücksichtigung des hohen A/T-Gehaltes (Kap. 4.1.2.) der single-copy Region, und im Besonderen der intergenischen Bereiche, konnten durch Computeranalysen diverse dieser Konsensus-Motive identifiziert werden. Allerdings waren die Motive entweder unterschiedlich weit von den jeweiligen Initiations-Codons entfernt oder in hohem Maße degeneriert, so dass eine genaue Zuordnung der Signale zu bestimmten Genen nicht möglich ist (Daten nicht gezeigt). Dies steht in Kongruenz zu bereits publizierten Daten (Fraunholz, 1999; Zauner et al., 2000), die von einer inakkuraten und somit auch ineffizienten 3'-Prozessierung der mRNA berichten. Ob auch die beobachteten, unterschiedlich langen 5'-Bereiche verschiedener cDNAs desselben Genes auf verschiedene Promotor-Bindungsstellen oder auf eine unvollständige Synthese der cDNAs

bei der Herstellung der cDNA-Bibliothek zurückzuführen sind, kann nur über die Analyse weiterer cDNAs beantwortet werden.

Die Analyse der Gen-Ausstattung des Nucleomorph-Genoms zeigt, dass alle drei eukaryotischen RNA-Polymerasen sowie die Transkriptions-Faktoren TfiID und TfiIB Nucleomorph-codiert sind. TfiID und TfiIB stellen zusammen mit TfiIA den Präinitiations-Komplex der Transkription dar. TfiID (TATA-Box bindendes Protein) innerhalb der interchromosomalen *inverted repeats* der Chromosomen II und III codiert, ist verantwortlich für die Erkennung der Promotor-Region und bindet an diese. TfiIB, codiert auf Chr III, ist für die anschließende Bindung der RNA Polymerase II an den Präinitiations-Komplex nötig. TfiIA konnte nicht auf dem Nucleomorph-Genom identifiziert werden. Auch spricht das Vorhandensein offener Leserahmen auf dem Nucleomorph-Genom, die für ein mRNA-*capping* Enzyms (*mce*) und ein *cap*-bindenden Enzyms (*cbp*) codieren, für eine *cap* Bildung am 5'-Ende der mRNA. Ebenso konnte ein offener Leserahmen mit Homologien zu einem PolyA-bindenden Protein (*pab*) identifiziert werden. Die eigentliche Polyadenylierung des 3'-Terminus der mRNA konnte durch diverse cDNAs nachgewiesen werden (Kap. 3.2.2.2.). All dies spricht dafür, dass die RNA-Polymerase II abhängige Transkription der offenen Leserahmen des Nucleomorph-Genoms im Laufe der Symbiogenese nicht radikal vereinfacht wurde.

Somit ist anzunehmen, dass die inakkurate 3'-Prozessierung der mRNA und die eventuell ungenaue Promotor-Bindung von TfiID als Reminiszenz der Selektion hin zu einem extrem reduzierten eukaryotischen Genom mit hohem A/T-Gehalt und kurzen intergenischen Bereichen anzusehen sind.

#### 4.1.3.3. Repetitive Strukturen und Pseudogene

Ein großer Teil der genomischen DNA eukaryoter Organismen besteht aus repetitiven Elementen einschließlich Retroelementen, Transposons und Mikrosatelliten. Eine Überprüfung des Nucleomorph-Genoms bezüglich dieser Strukturen ergibt intra- und interchromosomal keine tandemartigen Sequenzwiederholungen. Ebenso können, außer den bereits erwähnten terminalen *inverted repeats*, keine weiteren invertierten Sequenzwiederholungen identifiziert werden. Anhand einer interchromosomalen „DotPlot“-Analyse kann lediglich ein etwa 800 bp langer Sequenzabschnitt, der für einen *orf160* mit Homologien zu einem möglichen Membranprotein von *S. cerevisiae* codiert, identifiziert werden. Dieser potentiell Protein-codierende Sequenzabschnitt ist identisch

einmal auf Chr I und zweimal auf Chr III wiederholt, und wird zur Zeit in unserer Arbeitsgruppe weitergehend untersucht. Es scheint demnach möglich zu sein in einem derart minimierten eukaryoten Genom alle Arten von repetitiven Strukturen zu eliminieren, ohne dass dadurch die Struktur oder die Funktion der Chromosomen Schaden nimmt.

Ein weiterer, wichtiger Aspekt ist das beinahe vollständige Fehlen von Pseudogenen. Insgesamt konnte auf dem Nucleomorph-Genom nur ein Pseudogen identifiziert werden: Ein C-terminal stark verkürzter offener Leserahmen, der N-terminal Homologien (33 % Identität) zu einem ribosomalen Protein L24 (*rpl24*) liefert. Da zudem ein weiteres 60S ribosomales Protein L24, das über die gesamte Länge Homologien zu den Datenbankeinträgen aufweist, auf dem Nucleomorph Chromosom II lokalisiert ist, wurde die verkürzte Variante als Pseudogen gewertet. Weitere korrumpierte Gene konnten nicht identifiziert werden.

In *Rickettsia prowazekii* konnte gezeigt werden, dass in Pseudogenen, die durch natürliche Mutationen einen Wechsel des Leserahmens oder/und ein Stop-Codon enthielten, die Anzahl der Mutationen, die einen Aminosäureaustausch bewirken sprunghaft anstieg (Andersson und Andersson, 1999). Dies zeigt, dass Pseudogene keinen DNA-Reparaturmechanismen, die Punktmutationen reversieren könnten, unterliegen. Die Pseudogene können demnach frei evolvieren (Andersson und Andersson, 1999). Das heißt, dass das inaktivierte Pseudogen Nucleotid-Substitutionen, kleinere Deletionen und/oder Insertionen anhäuft bis es entweder nicht mehr als Pseudogen zu erkennen ist oder, falls die Deletionen die Insertionen überwiegen, ganz eliminiert wird. In dem zweifelsohne auf Kompaktheit selektierten Nucleomorph-Genom sollten Deletionen die Insertionen überwiegen. Unter Berücksichtigung des geringen Anteils nicht-codierender DNA einschließlich der sehr geringen Anzahl an Pseudogenen im Nucleomorph-Genom bedeutet dies, dass Reparaturmechanismen den *Status quo* aufrechterhalten; als ideale Komprimierung bei gleichzeitiger Funktionalität. Das Nucleomorph-Genom ist somit ein stabiler Endzustand.

Sollte die Reduktion des Nucleomorph-Genoms und die Verlagerung seines genetischen Materials in den Zellkern, bei gleichzeitigem Reimport der dort codierten Proteine, noch nicht abgeschlossen sein, wären größere nicht-codierende Bereiche und eine höhere Anzahl an Pseudogenen zu erwarten. Alternativ müssten komplette *orfs* deletiert werden.

#### 4.1.4. Introns

##### 4.1.4.1. Spliceosomale Introns in Protein-codierenden offenen Leserahmen sind sehr klein und selten

Auf dem Nucleomorph Chromosom II sind insgesamt acht potentiell Protein-codierende offene Leserahmen lokalisiert, die je ein kleines spliceosomales Intron beinhalten (Kap. 3.2.2.1.2.). Im Gesamtgenom des Nucleomorphs können insgesamt 17 Introns nachgewiesen werden, die alle in Struktur und Lokalisation denen des Chromosoms II vergleichbar sind (M. Reith, pers. Mitteilung). Die Konsensussequenz der Introns besteht aus GTA(A/G)GTN<sub>34-44</sub>AG, ihre Länge variiert zwischen 42 und 52 bp und sie besitzen die Standardgrenzen spliceosomaler Introns GT/AG. Mit einer Länge von 42-52 bp sind die Introns im Vergleich zu den spliceosomalen Introns anderer Eukaryoten sehr kurz, aber immer noch etwa doppelt so lang wie die Introns, die aus dem Nucleomorph-Genom der Chlorarachniophyte *P. minutissima* beschrieben wurden (Gilson und McFadden, 1996). Auch in dem Ciliat *Paramecium tetraurelia* wurden Introns, die mit einer Größe von 23 und 24 bp nur halb groß sind, wie die in dieser Arbeit vorgestellten Introns des Nucleomorph-Genoms der Cryptomonade *G. theta*, identifiziert (Zhang et al., 1999). Die große Differenz in der Anzahl der Introns zwischen den Nucleomorphen der Cryptomonaden und der Chlorarachniophyten ist auf ihre unterschiedliche phylogenetische Abstammung zurückzuführen. So stammt der Endosymbiont der Cryptomonaden von Vorläufern der heutigen Rotalgen ab, deren Genom als Intron-arm beschrieben wird, während die Chlorarachniophyten einen gemeinsamen Vorfahren mit den Intron-reichen Grünalgen besitzen (Maier et al.; 2000). Alle Introns sind im 5'-Bereich der jeweiligen offenen Leserahmen lokalisiert, oftmals direkt nach dem Initiations-Codon. Zudem sind 11 der 17 identifizierten Introns innerhalb von Genen lokalisiert, die für ribosomale Proteine codieren. Auch in *S. cerevisiae* wurden Introns identifiziert, die direkt nach dem Initiations-Codon ribosomaler Proteine inserierten. Für *S. cerevisiae* wurde beschrieben, dass das Herausschneiden dieser Introns das mRNA Level der betroffenen Proteine negativ reguliert (Spingola et al, 1999). Ob diese regulatorische Aktivität auch auf die spliceosomalen Introns des Nucleomorphs von *G. theta* zutrifft, muss durch zusätzliche Daten über Expressionsraten und die Lebensdauer verschiedener mRNA-Populationen eindeutig geklärt werden.

#### 4.1.4.2. tRNAs besitzen teilweise zwei Introns und Introns in der D- bzw. der Pseudouridin-Schleife

Von den auf Chromosom II identifizierten 13 tRNAs (37 auf allen drei Chromosomen) enthalten sechs (13) tRNAs mindestens ein Intron. Neben den auch in anderen Organismen beschriebenen Introns in der Anticodon-Schleife (Abelson et al., 1998), konnten fünf tRNAs ermittelt werden, die ein Intron in der D- bzw. der Pseudouridin-Schleife besitzen. Drei tRNAs enthalten zwei Introns in unterschiedlichen Schleifen. Introns in der D- bzw. der Pseudouridin-Schleife wurden bislang bei keinem anderen Organismus beschrieben, hingegen konnte in *Methanobacterium thermoautotrophicum* eine tRNA identifiziert werden, die sich durch zwei Introns in der Anticodon-Schleife auszeichnet (Smith et al., 1997). Die tRNAs sind, abgesehen von ihren ungewöhnlichen Introns, in ihrer durch Computeranalysen vorhergesagten Sekundärstruktur äußerst stabil. Dies wird durch den gestiegenen *cove*-Wert der tRNAs nach der Entfernung der Introns dokumentiert. Hieraus ergeben sich zwei Möglichkeiten: a) Die Insertionen sind relativ „jung“, da sich keine weiteren Mutationen in den pre-tRNAs angereichert haben, oder b) es gibt einen spezifischen Editierungs-Mechanismus, der diese ungewöhnlichen Introns aus den pre-tRNAs entfernt. Im Fall a) würde es sich um pseudo-tRNAs handeln, hingegen im Falle der Editierung der pre-tRNAs wären die tRNAs funktional. Die Isolierung eines Transkriptes der reifen tRNA<sup>Ser</sup>(AGA) und einiger Zwischenstufen der Editierung (Zauner et al., 2000) legt die Existenz eines spezifischen Mechanismus, der am *splicing* der pre-tRNAs des Nucleomorph-Genoms beteiligt ist, nahe.

#### 4.1.5. Centromere und Mitose

Die Vermehrung der Cryptomonaden erfolgt asexuell durch Zellteilung. Für eine vollständige Zellteilung sind mindestens sechs einzelne Schritte bzw. Organellenteilungen notwendig: Mitochondrien, Plastide, Nucleomorph, Teilung des Symbionten, Nucleus und zuletzt die Teilung der Zelle in zwei genetisch identische Tochterzellen. Die gesamte Zellteilung und somit auch die Teilung des Nucleomorphs steht ohne Zweifel unter Kontrolle des Nucleus von *G. theta*. Weiterhin muss, trotz des Fehlens ultrastruktureller und cytologischer Hinweise, eine Mitose für das Nucleomorph-Genom angenommen werden. Wichtige strukturelle Komponenten hierfür sind Centromere und ein Spindelapparat, die zusammen die Segregation der Schwesterchromatiden bewerkstelligen. Die Identifikation eines Centromer-spezifischen H3-ähnlichen Histons (Cenp-A)

(Glowczewski et al., 2000) im Nucleomorph-Genom liefert einen ersten Hinweis auf die Existenz von Centromeren, obwohl ihre genaue Lokalisation vorerst nicht angegeben werden kann. Einerseits besitzen nur die terminalen, invertierten Sequenzwiederholungen nicht-codierende DNA-Regionen mit signifikanter Länge, die auf allen drei Nucleomorph Chromosomen vorhanden sind. Andererseits ist es auf Grund ihrer Lage an beiden Chromosomenenden sehr unwahrscheinlich, dass die *inverted repeats* die Centromere ausbilden, da zwei Centromere die Chromatiden-Segregation zumindest erheblich erschweren würden.

Ausgehend von der Annahme, dass die Nucleomorph Centromere in ihrer Größe denen von *Saccharomyces cerevisiae* (Hegemann und Fleig, 1993; Bechert et al., 1999) ähnlich sind, können fünf Regionen identifiziert werden, die lang genug sind, um als Centromere zu fungieren. Zwei dieser Regionen sind auf Chromosom I (520 nt zwischen *orf147* und *orf471*, und 540 nt zwischen *orf244* und *orf446*), eine auf Chromosom II (203 nt zwischen *orf80a* und *orf231*) und weitere zwei sind auf Chromosom III (374 nt zwischen *orf141* und *orf180*, und etwa 300 nt zwischen *orf62* und *ebi*) lokalisiert (S. Douglas, pers. Mitteilung). Da alle fünf möglichen Centromer-Regionen nahe der Mitte der jeweiligen Chromosomen liegen, und alle anderen intergenischen Regionen zu kurz sind, würde sich ergeben, dass alle drei Nucleomorph Chromosomen metazentrisch sind. Versuche diese Regionen miteinander bzw. mit den Centromer-Regionen von *S. cerevisiae* zu alignen, scheiterten am hohen A/T-Gehalt der Nucleomorph Sequenzen, so dass eine Aussage über die Lokalisation der Centromere im Nucleomorph-Genom anhand dieser Daten derzeit nicht getroffen werden kann.

Zusätzliche Hinweise auf die Natur der Nucleomorph Teilung liefert die Identifikation weiterer am Aufbau der Kinetochore (*ranbpm*) oder am Aufbau des Spindelapparates ( $\alpha$ -,  $\beta$ -, and  $\gamma$ -Tubulin) beteiligter Komponenten (Zauner et al., 2000). Des weiteren sind die Chaperone Hsp82 (ein Chaperon der 90 kD Klasse) und Hsp70 Nucleomorph-codiert (S. Douglas, pers. Mitteilung). Hsp82 ist bekannt für Interaktionen mit centrosomalen Proteinen (Lange et al., 2000) wie  $\gamma$ -Tubulin mit dem es ko-lokalisiert ist (Zarsov et al., 1997). Zudem assoziiert Hsp90 mit Hsp70 zu einem Multichaperon-Komplex (Scheufler et al., 2000). Die Anwesenheit eines *nuclear localisation signal* (NLS) bei beiden Chaperonen legt die Vermutung nahe, dass das Centrosom innerhalb des Nucleomorphs lokalisiert ist und es sich bei der Nucleomorph Teilung möglicherweise um eine

vereinfachte, geschlossene Mitose des Rotalgen-Typs handelt. Eine weitere Möglichkeit wäre der Import fehlender Komponenten wie z.B. weitere Kinetochor-bildende Proteine (cenp-B, cenp-C) aus dem Cytoplasma des Wirtes.

#### 4.1.6. Gegenüberstellung der Nucleomorph-Genome von Cryptomonaden und Chlorarachniophyceae

Gegenstand des internationalen Genomprojektes war bzw. ist die vollständige Sequenzierung der Nucleomorphe je einer Cryptomonade (*Guillardia theta*) und einer Chlorarachniophyceae (*Pedinomonas minutissima*). Das Nucleomorph-Genom von *G. theta* ist vollständig sequenziert (diese Arbeit), während bei *P. minutissima* zur Zeit nur die Sequenz des Nucleomorph Chromosoms III vollständig bekannt ist (P. R. Gilson, pers. Mitteilung).

Beide Algengruppen sind in zwei voneinander unabhängigen Ereignissen durch sekundäre Endocytobiose entstanden. Während der Symbiont der Cryptomonaden einen gemeinsamen Vorfahren mit den Rotalge teilt, geht der Symbiont der Chlorarachniophyta auf eine Grünalge, die den heutigen Ulvophyceen nahe stehen soll, zurück. Ebenso gehen die Wirte auf verschiedene Organismengruppen zurück (Van de Peer et al., 1996).

Die Unterschiede in der Lokalisation der Nucleomorphe bzw. im Genomaufbau (Tab. 4.3.) beruhen auf der unterschiedlichen phylogenetischen Ausgangsposition und die daraus resultierende verschiedenartige Co-Evolution in beiden Algengruppen. So gehen die unterschiedliche Lokalisation und Anzahl der Nucleomorphe in beiden Algengruppen wahrscheinlich auf die unterschiedliche phylogenetische Stellung der Wirte zurück. Wohingegen Abweichungen in Anzahl und Länge der Introns, Orientierung der rDNA-Cluster und Telomersequenz auf die ungleichen Symbionten zurückzuführen sind.

Trotz der polyphyletischen Entstehung der Nucleomorphe zeigen sie in beiden Algengruppen starke Ähnlichkeiten in ihrer Struktur und dem Genomaufbau (Tab. 4.3.). Diese Gemeinsamkeiten sind sicherlich unabhängig voneinander entstanden und spiegeln den Evolutionsdruck wider, der die Genome intrazellulär lebender Organismen und Organellen auf Kompaktheit und Gendichte trimmt. Hierzu zählen die Anzahl der linearen Chromosomen, der hohe A/T-Gehalt, die für eukaryote Genome hohe Gendichte, überlappende Transkripte und die Organisation der Chromosomen in terminale *inverted repeats* mit je einem rDNA-Cluster und eine *single-copy* Region. Auch in der Genausstattung bestehen große Übereinstimmungen; so codieren beide Nucleomorphe

plastidär lokalisierte Proteine. Des weiteren codiert weder das Nucleomorph von *G. theta* noch das von *P. minutissima* für Gene des Primärmetabolismus. Dies trägt der Tatsache Rechnung, dass der Nucleomorph nicht das Genom eines unabhängigen Organismus darstellt, sondern vielmehr als eukaryotes Organell ohne eigenen Primärmetabolismus anzusprechen ist.

	<i>Guillardia theta</i>	<i>P. minutissima</i>
Lokalisation des Endosymbionten	Lumen des ER	ohne Verbindung zum ER; in einem cytoplas. Vesikel
Anzahl der Endosymbionte pro Zelle	1	>1
Anzahl der Chromosomen	3 (196, 181 und 174 kb)	3 (ca. 145, 140 und 98 kb)
A/T-Gehalt (codierende Bereiche)	77 %	70 %
Anzahl der Gene	455	~ 300
Telomere	[(AG) <sub>7</sub> AAG <sub>6</sub> A]	[TCTAGGG]
Terminale <i>inverted repeats</i>	mit rDNA Cluster und 5S rDNA	mit rDNA Cluster ohne 5S rDNA
Überlappende Transkripte	ja	ja
Co-Transkription Proteincodierter <i>orfs</i>	nicht bekannt	ja
Überlappende codierende Bereiche	tRNA mit RNA-Pol II <i>orfs</i> Zweimal RNA-Pol II <i>orfs</i>	nicht bekannt nicht bekannt
Introns	17, 42 – 52 bp lang Introns in tRNAs bekannt	sehr viele, 17 – 20 bp lang nicht bekannt
Gene für Enzyme des Primär-Metabolismus	keine	bisher keine identifiziert
Gene für plastidäre Funktionen	30	bisher 2 identifiziert

Tab. 4.3.: **Vergleich der Nucleomorph-Genome von Cryptomonaden und Chlorarachniophyta.** Die Cryptomonaden werden repräsentiert durch *G. theta*, die Chlorarachniophyta durch *Pedinomonas minutissima*. Alle Daten bezüglich *P. minutissima* aus Gilson und McFadden (1996); McFadden et al. (1997) bzw. pers. Mitteilung P. R. Gilson. Näheres siehe Text. Aus Maier et al. (2000) modifiziert.

## 4.2. Die Genausstattung des Nucleomorph-Genoms von *G. theta*

Im Rahmen dieser Arbeit wurden die potentiell Protein-codierenden offenen Leserahmen und die für tRNAs codierenden Sequenzabschnitte des Nucleomorph Chromosoms II identifiziert. Die entsprechenden Bereiche der Chromosomen I und III wurden innerhalb der internationalen Zusammenarbeit des Genomprojektes hauptsächlich in der Arbeitsgruppe von Susan Douglas in Halifax (Kanada) bearbeitet. Die Ergebnisse wurden anschließend vereint und untereinander abgeglichen.

### 4.2.1. Das Nucleomorph-Genom codiert hauptsächlich für *house-keeping* Funktionen

Das Nucleomorph-Genom codiert in seiner Gesamtheit für 455 (Chr II: 152) potentiell Protein-codierende offene Leserahmen. 173 dieser *orfs* (Chr II: 46) zeigen keine signifikanten Homologien zu Datenbankeinträgen, weitere 31 *orfs* zeigen Identitäten zu hypothetischen Proteinen anderer Organismen mit unbekannter Funktion, davon 20 (Chr II: fünf) zu anderen Eukaryoten und 11 (Chr II: sechs) zu Cyanobakterien. Da diese 31 *orfs* in einem außergewöhnlich kompakten Genom, wie es das Nucleomorph-Genom darstellt, identifiziert werden können, muss ihnen eine fundamentale, wenn auch noch unbekannte Bedeutung zugesprochen werden. Die verbleibenden 251 *orfs* (Chr II: 95) besitzen homologe Gegenstücke in anderen eukaryoten Organismen mit bekannter Funktion (siehe auch Genkarte im Anhang). Diese Gene lassen sich in die Hauptkategorien Chromosomen-Replikation (Replikation, Zellteilung), Expression der genetischen Information (Transkription, Translation und RNA-Prozessierung) und cytoplasmatische Funktionen (Proteinmetabolismus und –faltung, Signaltransduktion bzw. –regulation, Zellzyklus-Kontrolle und Membrantransport) einteilen.

Des Weiteren konnten insgesamt 47 (Chr II: 19) unterschiedliche RNA-codierende Bereiche identifiziert werden (37 tRNAs (Chr II: 13), fünf snoRNAs (Chr II: zwei), vier rDNAs (Chr II: vier) und eine snRNA (Chr II: keine)).

Das Nucleomorph-Genom beinhaltet 30 *orfs* (Chr II: 15) für plastidär lokalisierte Proteine, wovon 19 (Chr II: 9) eine bekannte Funktion aufweisen und 11 *orfs* (Chr II: sechs) Homologien zu hypothetischen Proteinen des Cyanobakteriums *Synechocystis sp.* zeigen. Diese 30 Nucleomorph-codierten *orfs* weisen eine N-terminale Extension auf, die

Nucleomorph spezifische, aber dennoch typische Merkmale eines Transitpeptides zeigt (Wastl, 2000). Die Funktion der N-terminalen Extension als Transitpeptid konnte auch experimentell nachgewiesen werden (J. Prechtel, pers. Mitteilung; Wastl und Maier, 2000).

Die große Mehrheit der identifizierten Nucleomorph codierten Gene wird dazu benötigt das Nucleomorph und das periplastidäre Kompartiment funktionell aufrechtzuerhalten, um die Expression 30 plastidär lokalisierter Proteine zu gewährleisten.

#### **4.2.2. Das Nucleomorph ist das Genom eines semiautonomen eukaryoten Organells**

Bei der Betrachtung der tRNAs fällt die geringe Anzahl an Nucleomorph-codierten tRNAs auf. Da im Nucleomorph-Genom von *G. theta* alle Codons genutzt werden (Kap. 3.2.2.1.3.) muss von 61 benötigten tRNAs ausgegangen werden. Eine Untersuchung der Anticodons der codierten tRNAs ergibt, dass keine tRNAs für die beiden Codons der Glutaminsäure auf dem Nucleomorph-codiert sind. Dies bedeutet, dass mindestens eine nucleär-codierte tRNA<sup>Glu</sup> in das periplastidäre Kompartiment importiert werden muss. Da im Nucleomorph keine Hinweise auf eine „Wobbel-Basenpaarung“ zwischen Codon und Anticodon bekannt sind, muss auch für die verbleibenden nicht detektierbaren tRNAs ein Import cytosolischer tRNAs analog zu dem bei vielen anderen Organismen beschriebenen Import von tRNAs in Mitochondrien postuliert werden (reviewed in Schneider und Maréchal-Drouard, 2000). Eine weitere Möglichkeit die fehlenden tRNAs zu erklären geht auf die teilweise ungewöhnlichen Introns einiger identifizierten tRNAs zurück. So zeichnet sich die tRNA<sup>Ser</sup>(AGA) durch je eine Insertion in der Anticodon-Schleife und in der D-Schleife aus (Kap. 3.2.2.3.). Auch konnte die Prozessierung dieser tRNA<sup>Ser</sup>(AGA) gezeigt werden (Zauner et al., 2000). Es wäre demnach denkbar, dass die fehlenden tRNAs mit den benutzten Suchalgorithmen nicht identifiziert werden konnten, obwohl sie Nucleomorph-codiert sind.

Im Wirtszellkern codierte und anschließend importierte cytosolische tRNAs könnten, wie im Falle des tRNA Transportes ins Mitochondrium von *S. cerevisiae* beschrieben (Tarassov und Martin, 1996), bereits aminoacetyliert in das periplastidäre Kompartiment importiert werden oder, wie im Falle von *Tetrahymena pyriformis*, unbeladen transportiert werden (Rusconi und Cech, 1996). Das Nucleomorph-Genom codiert nur für eine tRNA-Synthetase, die Seryl-Synthetase. Demzufolge müssen die für die Aminoacetylierung der periplastidär-lokalisierten tRNAs notwendigen tRNA-Synthetasen (mit Ausnahme der

Seryl-Synthetase) nucleär codiert sein und in das periplastidäre Kompartiment importiert werden.

Die Analyse der Genausstattung des Nucleomorph-Genoms ergibt, dass das Nucleomorph nicht für Gene der Untereinheiten der DNA-Polymerase, keine Kernporen-Komponenten und nur eine von 20 nötigen tRNA-Synthetasen codiert. Ein Vergleich von Multiproteinkomplexen mit den Homologons anderer Eukaryoten zeigt, dass auf dem Nucleomorph weniger Untereinheiten codiert sind als in anderen Eukaryoten beschrieben sind. Eine Ausnahme stellt das T-Komplex Protein 1 dar, dessen acht Untereinheiten ausnahmslos Nucleomorph-codiert sind (Archibald et al., 2001).

Es gibt drei Möglichkeiten, zu erklären wie die nicht Nucleomorph-codierten Komponenten funktionell ersetzt werden können, bzw. die biologischen Funktionen dieser Komponenten aufrecht erhalten werden kann:

- a) Die fehlenden Komponenten sind im Nucleus des Wirtes codiert und werden ins periplastidäre Kompartiment transportiert. Dies trifft sicherlich auf die Untereinheiten der DNA-Polymerase und die tRNA Synthetasen zu. Diese Proteine sind relativ groß und hochkonserviert und sollten aus diesem Grund durch eine Homologiesuche gegen die Einträge der Datenbanken des NCBI-Browsers auf dem Nucleomorph-Genom identifizierbar sein. Auch müssen die fehlenden Untereinheiten der Multiproteinkomplexe durch die kerncodierten Homologons komplementiert werden.
- b) Die fehlenden Gene oder Komponenten sind so divergent, dass die Identifizierung über eine Homologiesuche in den Datenbanken nicht zum Ziel führt, d.h. die Funktionen sind durch Computeranalysen nicht zu erfassen, sind aber auf dem Nucleomorph-Genom vorhanden. Somit könnten die Genprodukte potentiell Protein-codierender offener Leserahmen die Funktion der fehlenden Komponenten ausführen bzw. übernehmen. Diese Möglichkeit ist vor allem für kleinere, wenig konservierte Proteine in Betracht zu ziehen.
- c) Eine dritte Möglichkeit stellt die Vereinfachung biologischer Prozesse dar, so dass die vermeintlich fehlenden Komponenten redundant sind und ihr Fehlen auf andere, noch unbekannt Weise ausgeglichen wird (z.B. posttranskriptionelle Editierung; Kap. 4.1.3.2.).

Auf dem Nucleomorph konnten keine Gene identifiziert werden, die für Enzyme des Primärmetabolismus codieren. Der Genverlust im Nucleomorph führt zu einem semiautonomen Genom, das nur für seine eigene Aufrechterhaltung und einige cytoplasmatische Funktionen codiert, um 30 Nucleomorph-codierte plastidäre Proteine zu exprimieren. Im Gegensatz dazu werden in anderen sehr stark reduzierten Genomen mit ähnlich geringer Codierungskapazität, z.B. *Mycoplasma genitalium* (Fraser et al., 1995) oder *Rickettsia prowazekii* (Andersson et al., 1998) Enzyme für metabolische Reaktionswege codiert. Dies ist auf die Tatsache zurückzuführen, dass es sich bei diesen parasitischen, prokaryoten Organismen um echte Einzeller und nicht, wie im Falle des Nucleomorphs, um den integrierten Bestandteil einer komplexen, chimären Zelle handelt (Zauner et al., 2000; Maier et al., 2000).

Andererseits weisen die Genome intrazellulärer, prokaryoter Parasiten oder Symbionten ebenso wie die Organellen einige Gemeinsamkeiten mit dem des Nucleomorphs auf. So verfügen alle über einen extrem hohen A/T-Gehalt, auch überlappende Gene sind eine Besonderheit der Prokaryoten und der Organellen. Auch der humanpathogene, obligat intrazelluläre Parasit *Plasmodium falciparum* zeigt einen extrem hohen A/T-Gehalt (Gardner et al., 1998). Diese Übereinstimmungen sind ohne Zweifel konvergent entstanden, weisen aber auf ähnliche evolutionäre Mechanismen und eine ähnliche Selektion hin, die zur Entstehung kompakter Genome führen.

Wie allen prokaryoten semiautonomen Organellen (Plastiden und Mitochondrien) fehlen auch dem Nucleomorph wichtige Komponenten für seine eigene Replikation (z.B. DNA-Polymerase). Dies zeigt, dass auch im Falle der sehr komplexen Zelle mit vier phylogenetisch unterschiedlichen Genomen, wie sie in *G. theta* vorliegt, die Evolution zu einem „Master-Genom“ führt, dem die Kontrolle und das Management der anderen Genome unterliegt.

### 4.3. Ausblick

Die vollständige Entschlüsselung eines eukaryoten Genoms wirft naturgemäß mehr Fragen auf als *ad hoc* beantwortet werden können. Die Untersuchungen einzelner Nucleomorph-codierter Gene bezüglich ihrer Funktion sind teilweise bereits angelaufen. Ebenso sind phylogenetische Untersuchungen anhand der Untereinheiten einzelner Multiproteinkomplexe initiiert. Besonders erfolgversprechend sind Fragestellungen, die sich an den Besonderheiten des Nucleomorph-Genoms orientieren.

Die Nucleomorphe der Cryptomonaden und Chlorarachniophyta sind ein Paradebeispiel für die Komprimierung eukaryoter Genome. Es ist davon auszugehen, dass Teile der Protein-codierenden offenen Leserahmen, die keine Homologie zu Datenbankeinträgen aufweisen, auch in anderen Genomen eine wichtige Funktion besitzen. Diese *orfs* können kloniert und durch *knockout*-Experimente in geeigneten Organismen (*S. cerevisiae* oder *Synechocystis sp.*) charakterisiert werden. Auch Fragen bezüglich der Koordination der vier Genome, z.B. während des Zellzyklus, sind von außerordentlichem Interesse.

Ein weiterer wichtiger Aspekt wird die Isolierung nucleär-codierter und anschließend in das periplastidäre Kompartiment transportierter Proteine sein. Anhand der in dieser Arbeit vorgestellten Daten kann gezielt nach solchen Genen ‚gescreent‘ werden. Von Interesse ist das Vorhandensein bzw. Fehlen topogener Signale, die den bisher noch nicht charakterisierten Transport über die periplastidäre Membran initiieren.

## 5. ZUSAMMENFASSUNG

Das Nucleomorph von *Guillardia theta* ist das zweite vollständig sequenzierte eukaryote Genom neben *Saccharomyces cerevisiae*. Die drei linearen Nucleomorph Chromosomen besitzen eine Gesamtgröße von 551263 Basenpaaren (I: 196215 bp, II: 180915 bp und III: 174133 bp). Je eine rDNA-Transkriptionseinheit ist Teil der auf allen drei Chromosomen lokalisierten, bis zu 14 kb großen, subtelomerischen *inverted repeats*. Diese schließen eine zentrale *single-copy* Region ein. Der A/T-Gehalt der *single-copy* Bereiche ist mit 77,2 % in den codierenden und 82,3 % in den intergenischen Regionen äußerst hoch.

Es konnte gezeigt werden, dass die *single-copy* Regionen für 455 potentiell Protein-codierende offene Leserahmen, 37 tRNAs, vier kleine nukleoläre RNAs (snoRNA) und eine kleine nucleare RNA (snRNA) codieren. Hieraus ergibt sich eine rechnerische Gendichte von etwa einem Gen pro 950 bp. In Protein-codierenden offenen Leserahmen konnten nur 17 Introns, mit der Konsensussequenz GTARGTN<sub>(34-44)</sub>AG, identifiziert werden. Die Introns sind alle im 5'-Bereich der offenen Leserahmen lokalisiert und treten gehäuft in ribosomalen Proteinen auf. Die innerhalb von tRNAs identifizierten Introns können in allen drei Schleifen lokalisiert sein, zudem konnten tRNAs mit mehr als einem Intron identifiziert werden. Es wurde gezeigt, dass die intergenischen Regionen äußerst kurz sind und in einigen Fällen RNA-Polymerase II transkribierte Protein-codierende Bereiche a) mit einander oder b) mit RNA-Polymerase III transkribierten tRNA codierenden Bereichen überlappen. All diese Befunde lassen sich auf den evolutionären Druck zur Komprimierung dieses zweiten eukaryoten Genoms der einzelligen Cryptomonade *G. theta* zurückführen.

Die Genausstattung des Nucleomorph-Genoms lässt sich in mehrere Kategorien unterteilen. Die meisten offenen Leserahmen stellen Komponenten der Transkriptions- und Translations-Maschinerie dar oder repräsentieren Teile des DNA- bzw. RNA-Metabolismus. Weiterhin sind Komponenten der Proteindegradation und der Proteinfaltung, Bestandteile des Zellzyklus sowie Bestandteile des Cytoskeletts codiert. Es konnten zudem 19 *orfs* mit Homologien zu plastidären Funktionen und weitere 11 *orfs* mit Homologien zu hypothetischen Proteinen des phototrophen Cyanobakteriums *Synechocystis sp.* (PCC6803) identifiziert werden. Etwa 30 % aller identifizierten offenen

Leserahmen zeigen keine Homologie zu den Einträgen in den Datenbanken. Komponenten des Primärmetabolismus konnten nicht identifiziert werden. Weitere Anzeichen, dass das Nucleomorph nicht als autonom anzusehen ist, sind zum Beispiel das komplette Fehlen von DNA-Polymerasen und das Fehlen wichtiger Komponenten einiger Multiproteinkomplexe (z.B. Proteasom, Ribosom, Spliceosom). Zudem codiert das Nucleomorph-Genom nicht für alle erforderlichen tRNAs und deren Synthetasen. Die fehlenden Komponenten bzw. Untereinheiten werden entweder vom Wirts-Genom codiert und in das periplastidäre Kompartiment transportiert, sind äußerst divergent in ihrer Sequenz und deshalb einer Homologiesuche nicht zugänglich oder werden durch andere Mechanismen funktionell ersetzt.

Das Fehlen von Komponenten des Primärmetabolismus und die Komplementierung essentieller Funktionen durch den Wirt zeigen, dass das Nucleomorph nicht mehr der Zellkern einer autonomen Zelle ist und deshalb sein Genom nicht als Minimalgenom anzusehen ist, sondern als Genom eines semiautonomen Organells.

## 6. LITERATURVERZEICHNIS

- Abelson, J., Trotta, C. R., Li, H. J. (1998). tRNA splicing. *Biol Chem* **273**: 12685-12688.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. (1997). Gapped-BLAST and PSI-BLAST: a new generation of protein database search programs. *Nuc Acids Res* **25**: 3389-3402.
- Andersson, S. G. E., Sharp, P. M. (1996). Codon usage and base composition in *Rickettsia prowazekii*. *J Mol Evol* **42**: 525-536.
- Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C., Podowski, R. M., Naslund, A. K., Eriksson, A. S., Winkler, H. H., Kurland, C. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**: 133-40
- Andersson, J. O., Andersson, S. G. E. (1999). Genom degradation is an ongoing process in *Rickettsia*. *Mol Biol Evol* **16**: 1178-1191.
- Archibald, J., Cavalier-Smith, T., Maier, U.-G., Douglas, S. E. (2001). Molecular chaperones from a reduced eukaryotic nucleus – the cryptomonad nucleomorph. *J Mol Evol* in press.
- Baldauf, S., Manhart, J., Palmer, J. D. (1990). Different fates of the chloroplast *tufA* gene following its transfer to the nucleus in green algae. *Proc Natl Acad Sci (USA)* **87**: 5317-5321.
- Bechert, T., Heck, S., Fleig, U., Diekmann, S., Hegemann, J. H. (1999). All 16 centromere DNAs from *Saccharomyces cerevisiae* show DNA curvature. *Nucleic Acids Res* **27**: 1444-9.
- Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., Van, Staveren, M., Stiekema, W., Drost, L., Ridley, P., Hudson, S. A., Patel, K., Murphy, G., Piffanelli, P., Wedler, H., Wedler, E., Wambutt, R., Weitzenegger, T., Pohl, T. M., Terryn, N., Gielen, J., Villarroel, R., Chalwatzis, N., et al. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **391**: 485-488.
- Bhattacharya, D., Medlin, L. (1995). The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. *J Phycol* **31**: 489-498.
- Bhattacharya, D., Medlin, L. (1998). Algal phylogeny and the origin of land plants. *Plant Physiol* **116**: 9-15.
- Biessmann, H., Mason, J. M. (1997). Telomere maintenance without telomerase. *Chromosoma* **106**: 63-69.
- Birnboim, H. C., Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc Acids Res* **7**: 1513-1522.

Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Barrell, B. G., et al. (1999). The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**: 532-8.

Brugere, J. F., Cornillot, E., Metenier, G., Vivares, C. P. (2000). Occurrence of subtelomeric rearrangements in the genome of the microsporidian parasite *Encephalitozoon cuniculi*, as revealed by a new fingerprinting procedure based on two-dimensional pulsed field gel electrophoresis. *Electrophoresis* **21**: 2576-81.

Bussey, H., Kaback, D. B., Zhong, W., Vo, D. T., Clark, M. W., Fortin, N., Hall, J., Ouellette, B. F., Keng, T., Barton, A. B., et al. (1995) The nucleotide sequence of chromosome I from *Saccharomyces cerevisiae*. *Proc Natl Acad Sci (USA)* **92**: 3809-3813.

Cavalier-Smith, T. (2000). Membrane heredity and early chloroplast evolution. *Trends Plant Sci* **5**: 174-182.

Chun, E.H.L, Vaughan, M.H, Rich, A. (1963). The isolation and characterization of DNA associated with chloroplast preparations. *J Mol Biol* **7**:130-141.

Deane, J. A., Hill, D. R. A., Brett, S. J., McFadden, G. I. (1998). *Hanusia phi* gen. et sp. nov. (Cryptophyceae): characterization of "*Cryptomonas* sp.  $\Phi$ ". *Eur J Phycol* **33**: 149-154.

Delwiche, C. F., Palmer, J. D. (1997). The origin of plastids and their spread via secondary endosymbiosis. In: Bhattacharya D. (ed.): *Origins of algae and their plastids*. Vienna, New York: *Springer Verlag*. 53-86.

Delwiche, C. F. (1999). Tracing the thread of plastid diversity through the tapestry of life. *Am Nat* **154**: 164-177.

Douglas, S. E., Murphy, C. A., Spencer, D. F., Gray, M. W. (1991). Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* **350**: 148-151.

Douglas, S. E. (1994). Chloroplast origins and evolution. In: *The molecular biology of cyanobacteria* (D.A. Bryant ed.) *Kluwer Academic Press, (the Netherlands)* : 91-118.

Douglas, S. E. (1998). Plastid evolution: origins, diversity, *trends Curr Opin Genet Develop* **8**: 655-661.

Douglas, S. E., Penny, S. L. (1999). The plastid genome of the cryptophyte alga, *Guillardia theta* : complete sequence and conserved synteny groups confirm its common ancestry with red algae. *J Mol Evol* **48**: 236-244.

Doyle, J. L., Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* **12**: 13-15.

Dujon, B. (1996). The yeast genome project: what did we learn? *Trends Genet* **12**: 263-270.

- Eschbach, S., Hofmann, C. J. B., Maier, U.-G., Sitte, P., Hansmann, P. (1991). A eukaryotic genome of 660 kb: electrophoretic karyotype of nucleomorph and cell nucleus of the cryptomonad alga *Pyrenomonas salina*. *Nuc Acids Res* **19**: 1779-1781.
- Eyre-Walker, A. (1999). Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. *Genetics* **152**: 675-83.
- Feinberg, A. P., Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analyt Biochem.* **132**: 6-13.
- Fraunholz, M. J., Wastl, J., Zauner, S., Rensing, S. A., Scherzinger, M. M., Maier, U.-G. (1997). The evolution of cryptophytes. *Plant Syst Evol* [Suppl.] **11**: 163-174.
- Fraunholz, M. (1999) Das Nucleomorph Chromosom II der Cryptomonade *Guillardia theta*: Struktur, Sequenzanalyse und evolutionäre Implikation. Dissertation, Philipps-Universität Marburg.
- Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., et al. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397-403.
- Gantt, E. (1979). Phycobiliproteins of Cryptophyceae. in *Biochemistry and Physiology of Protozoa*. New York, Academic Press. 121-137.
- Gardner, M. J., Tettlin, H., Carucc, D. J., Cummings, L. M., Adams, M. D., Smith, H. O., Venter, J. C., Hoffman, S. L. (1998). The malaria genome sequencing project. *Protist* **149**: 109-112.
- Gibbs, S. P. (1981). The chloroplast of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Ann NY Acad Sci* **361**: 193-207.
- Gibbs, S. P. (1983). The cryptomonad nucleomorph: Is it the vestigial nucleus of a eukaryotic endosymbiont? in *Endocytobiology II. Intracellular space as an oligogenetic system. Proc. 2nd Int. Colloquium Endocytobiology*. Berlin, New York, de Gruyter. 987-992.
- Gillot, M. A., Gibbs, S. P. (1980). The cryptomonad nucleomorph: its ultrastructure and evolutionary significance. *J Phycol* **16**: 558-568.
- Gillot, M. A. (1990). Phylum *Cryptophyta* (Cryptomonads). in *Handbook of Protoctista*. Boston, Jones & Bartlett. 139-151.
- Gilson, P. R., McFadden, G. I. (1996). The miniaturized genome of a eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and the smallest known spliceosomal introns. *Proc Natl Acad Sci (USA)* **93**: 7737-7742.
- Gilson, P. R., Maier, U.-G., McFadden, G. I. (1997). Size isn't everything: lessons in genetic miniaturisation from nucleomorphs. *Curr Opin Genet Develop* **7**: 800-806.

- Glowczewski, L., Yang, P., Kalashnikova, T., Santisteban, M. S., Smith, M. M. (2000). Histone-histone interactions and centromere function. *Mol Cell Biol* **20**: 5700-11
- Gray, M. W. (1992). The Endosymbiont Hypothesis Revisited. *Int Rev Cyt* **141**: 233-357.
- Gray, M. W., Lang, F. B., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T. G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y., Burger, G. (1998). Genome structure and gene content in protist mitochondrial DNAs. *Nuc Acids Res* **26**: 865-78.
- Gray, M. W. (1999). Evolution of organellar genomes. *Curr Opin Genet Dev* **9**: 678-687.
- Greenwood, A. D., Griffiths, H. B., Santore, U. J. (1977). Chloroplasts and cell compartments in cryptophyceae. *Brit Phycol J* **12**: 119.
- Guillard, R. R. L., Ryther, J. H. (1962). Studies on Marine Planktonic Diatoms. I. *Cyclonella nana*, Hustedt and *Detomula confervacea* (Cleve). *Gran Can J Microbiol* **8**: 229-239.
- Guillard, R. R. L. (1975). Culture of Phytoplankton for Feeding Marine Invertebrates. in *Culture of Marine Invertebrate Animals*. New York, Plenum Press. 29-60.
- Hagelberg, E., Goldman, N., Lio, P., Whelan, S., Schiefenhovel, W., Clegg, J. B., Bowden, D. K. (1999). Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc R Soc Lond B Biol Sci* **266**: 485-92.
- Hansmann, P., Falk, H., Scheer, U., Sitte, P. (1986). Ultrastructural localization of DNA in two *Cryptomonas* species by use of monoclonal DNA antibody. *Endocyt Cell Res* **4**: 289-295.
- Hansmann, P. (1988). Ultrastructural localization of RNA in cryptomonads. *Protoplasma* **146**: 81-88.
- Hansmann, P., Eschbach, S. (1990). Isolation and preliminary characterization of the nucleus and the nucleomorph of a cryptomonad, *Pyrenomonas salina*. *Eur J Cell Biol* **52**: 373-378.
- Häuber, M. M., Müller, S. B., Speth, V., Maier, U.-G. (1994). How to evolve a complex plastid? - A hypothesis. *Bot Acta* **107**: 369-472.
- Helmchen, T. A., Bhattacharya, D., Melkonian, M. (1995). Analyses of ribosomal RNA sequences from glaucocystophyte cyanelles provide new insights into the evolutionary relationships of plastids. *J Mol Evol* **41**: 203-10.
- Hegemann, J. H., Fleig, U. N. (1993). The centromere of budding yeast. *BioEssays* **15**: 451-460.
- Hibberd, D. J., Greenwood, A. D., Griffiths, A. H. (1971). Observations on the ultrastructure of the flagella and periplast in the *Cryptophyceae*. *Brit Phycol J* **6**: 61-71.

- Hibberd, D. J., Norris, R. E. (1984). Cytology and ultrastructure of *Chlorarachnion reptans* (*Chlorarachniophyta* divisio nova, *Chlorarachniophyceae* classis nova). *J Phycol* **20**: 310-330.
- Hofmann, C. J. B., Rensing, S. A., Häuber, M. M., Martin, W. F., Müller, S. B., Couch, J., McFadden, G. I., Igloi, G. L., Maier, U.-G. (1994). The smallest known eukaryotic genomes encode a protein gene: towards an understanding of nucleomorph functions. *Mol Gen Genet* **243**: 600-604.
- Holmes, D. S., Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. *Analyt Biochem* **114**: 193-197.
- Jeffrey, S. W., Humphrey, G. F. (1975). New spectrophotometric equations for determining chlorophyll *a*, *b*, *c*<sub>1</sub> and *c*<sub>2</sub> in higher plants, algae and natural phytoplankton. *Biochem Physiol Pflanzen* **167**: 191-194.
- Keeling, P. J., Deane, J. A., Hink-Schauer, C., Douglas, S. E., Maier, U.-G., McFadden, G. I. (1999) The secondary endosymbiont of the cryptomonad *Guillardia theta* contains alpha, beta, and gamma-tubulin genes. *Mol Biol Evol.* **9**: 1308-1313
- Klaveness, D. (1981). *Rhodomonas lacustria* (Pastner & Ruttner) Javornicky (*Cryptomonadia*): ultrastructure of the vegetative cell. *J Protozool* **28**:
- Kotani, H., Nakamura, Y., Sato, S., Kaneko, T., Asamizu, E., Miyajima, N., Tabata, S. (1997). Structural analysis of *Arabidopsis thaliana* chromosome 5. II. Sequence features of the regions of 1,044,062 bp covered by thirteen physically assigned P1 clones. *DNA Res* **4**: 291-300.
- Kotani, H., Nakamura, Y., Sato, S., Asamizu, E., Kaneko, T., Miyajima, N., Tabata, S. (1998). Structural analysis of *Arabidopsis thaliana* chromosome 5. VI. Sequence features of the regions of 1,367,185 bp covered by 19 physically assigned P1 and TAC clones. *DNA Res* **5**: 203-216.
- Kowallik, K. V., Stoebe, B., Schaffran, I., Kroth-Pancic, P., Freier, U. (1995). The chloroplast genome of a chlorophyll *a+c*-containing alga, *Odontella sinensis*. *Plant Mol Biol Rep* **13**: 336-342.
- Krakauer, D. C. (2000). Stability and evolution of overlapping genes. *Evol Int J Org* **54**: 731-9.
- Lange, B. H. M., Bachi, A., Wilm, M., Gonzalez, C. (2000). Hsp90 is a core centrosomal component and is required at different stages in the centrosome cycle in *Drosophila* and vertebrates. *EMBO J* **19**: 1252-1262.
- Lowe, T.M., Eddy, S.R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955-964.
- Ludwig, M., Gibbs, S. P. (1985). DNA is present in the nucleomorph of cryptomonads: further evidence that the chloroplast evolved from a eukaryotic endosymbiont. *Protoplasma* **127**: 9-20.

- Ludwig, M., Gibbs, S. P. (1987). Are the nucleomorphs of cryptomonads and *Chlorarachnion* the vestigial nuclei of eukaryotic endosymbionts? *Ann New York Acad Sci* **501**: 198-211.
- Ludwig, M., Gibbs, S. P. (1989). Evidence that the nucleomorphs of *Chlorarachnion reptans* (Chlorarachniophyceae) are vestigial nuclei: morphology, division and DNA-DAPI fluorescence. *J Phycol* **25**: 385-394.
- Lunt, D. H., Hyman, B. C. (1997). Animal mitochondrial DNA recombination. *Nature* **387**: 247.
- Maier, U.-G., Hofmann, C. J. B., Eschbach, S., Wolters, J., Igloi, G. I. (1991). Demonstration of nucleomorph-encoded eukaryotic small subunit ribosomal RNA in cryptomonads. *Mol Gen Genet* **230**: 155-160.
- Maier, U.-G., Douglas, S. E., Cavalier-Smith, T. (2000). The nucleomorph genomes of cryptophytes and chlorarachniophytes. *Protist* **151**: 103-109.
- Margulis, L. (1970). Origin of eukaryotic cells. Yale University Press, New Haven, Connecticut.
- Martin, W., Stöbe, B., Goremykin, V., Hansmann, S., Hasegawa, M., Kowallik, K. V. (1998 b). Gene transfer to the nucleus and evolution of chloroplasts. *Nature* **393**: 162-165.
- Mason, J. M., Biessmann, H. (1995). The unusual telomeres of *Drosophila*. *Trends Genet* **11**: 58-62.
- McFadden, G. I. (1991). In situ hybridization techniques: Molecular cytology goes ultrastructural. In: Hall JL, Hawes CR (eds) *Electron microscopy of plant cells*. Academic Press, London, pp 219-255.
- McFadden, G. I., Gilson, P. R., Hofmann, C. J. B., Adcock, G. J., Maier, U.-G. (1994). Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. *Proc Natl Acad Sci (USA)* **91**: 3690-3694.
- McFadden, G. I., Gilson, P. R. (1995) Something borrowed, something green: lateral transfer of chloroplasts by secondary endosymbiosis. *Trends Ecol Evol* **10**: 12-17.
- McFadden, G. I., Reith, M. E., Lang-Unnasch, N. (1996). Plastids in human parasites. *Nature* **381**:
- McFadden, G. I., Gilson, P. R., Hofmann, C. J. B. (1997). Division Chlorarachniophyta. *Plant Syst Evol [Suppl.]* **11**: 175-185.
- McFadden, G. I., Waller, R. F. (1997). Plastids in parasites of humans. *Bioessays* **19**: 1033-40.
- McKerracher, L., Gibbs, S. P. (1982). Cell and nucleomorph division in the alga *Cryptomonas*. *Canad J Bot* **60**: 2440-2452.

- Melkonian, M., Marin, B., Surek, B. (1995) Phylogeny and evolution of algae. In: Arai R, Kato M, Doi Y (eds) Biodiversity and evolution. *National Science Foundation, (Tokyo)*: 153-176.
- Mereschkowsky, V. C. (1905). Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biologisches Centralblatt* **25**: 593-604.
- Moran, N. A. (1996). Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci (USA)* **93**: 2873-2878.
- Moran, N. A., Baumann, P. (2000). Bacterial endosymbionts in animals. *Curr Opin Microbiol* **3**: 270-5.
- Muller, J. J. (1964). The relation of recombination to mutational advance. *Mutat Res* **1**: 2-9
- Muto, A., Osawa, S. (1987). The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci (USA)* **84**: 166-169.
- Nosek, J., Tomaska, L., Fukuhara, H., Suyama, Y., Kovac, L. (1998). Linear mitochondrial genomes: 30 years down the line. *Trends Genet* **14**: 184-188.
- Okabe, Y., Okada, M. (1988). Isolation of native core matrix having Ribulose 1,5-bisphosphate Carboxylase activity from the green alga *Bryopsis maxima*. *Plant Cell Physiol* **29**: 89-96.
- Palmer, J. D., Delwiche, C. F. (1996). Second-hand chloroplasts and the case of the disappearing nucleus. *Proc Natl Acad Sci (USA)* **93**: 7432-7435.
- Reith, M. E., Munholland, J. (1995). Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol Biol Rep* **13**: 333-335.
- Rensing, S. A., Goddemeier, M., Hofmann, C. J. B., Maier, U.-G. (1994). The presence of a nucleomorph *hsp70* gene is a common feature of Cryptophyta and Chlorarachniophyta. *Curr Genet* **26**: 451-455.
- Ris, H., Plaut, W. (1962). The ultrastructure of DNA-containing areas in the chloroplast of *Chlamydomonas*. *J Cell Biol* **13**: 383-391.
- Rusconi, C. P., Cech, T. R. (1996). The anticodon is the signal sequence for mitochondrial import of glutamine tRNA in *Tetrahymena*. *Genes Dev* **10**: 2870-80
- Sambrook, J., Fritsch, E. F., Maniatis, T. (1989). Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press.
- Santore, U.J. (1984). Some aspects of taxonomy in the cryptophyceae. *New Phytol* **98**: 627-646.
- Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F U., Moarefi, I. et al. (2000) Structure of TPR Domain–Peptide Complexes: Critical Elements in the Assembly of the Hsp70–Hsp90 Multichaperone Machine. *Cell* **101**: 199-210

- Schimper, A. F. W. (1883). Ueber die Entwicklung der Chlorophyllkörner und Farbkörper. *Botanische Zeitung* **41**: 106-114.
- Schneider, A., Maréchal-Drouard, L. (2000). Mitochondrial tRNA import: are there distinct mechanisms? *Trends Cell Biol* **10**: 509-513
- Sitte, P. (1983). General organization of the eucyte and its bearings on cytosymbiosis and cell evolution. In: Schenk HEA, Schwemmler W (eds) *Endocytobiology II*. De Gruyter, Berlin, pp 101-119.
- Sitte, P., Baltés, S. (1990). Morphometric analysis of two cryptomonads. Quantitative evaluation of fine-structural changes in an endocytobiotic system. In: Nardon P, Gianinazzi-Pearson V, Grenier AM, Margulis L, Smith DC (eds) *Endocytobiology IV* INRA, Paris, pp 229-233.
- Sitte, P. (1993). Symbiogenetic evolution of complex cells and complex plastids. *Eur J Protistol* **29**: 131-143.
- Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., et al. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum*  $\Delta$ H: Functional analysis and comparative genomics. *Journal of Bacteriology* **179**: 7135-7155
- Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**: 503-517.
- Spear-Bernstein, L., Miller, K. R. (1989). Unique localization of the phycobiliprotein light harvesting pigment in the *Cryptophyceae* *J Phycol* **25**: 412-419.
- Spingola, M., Grate, L., Haussler, D., Ares, M. (1999). Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* **5**: 221-234.
- Stirewalt, V. L., Michalowski, C. B., Löffelhardt, W., Bohnert, H. J., Bryant, D. A. (1995). Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Mol Biol Rep* **13**: 327-332.
- Stöbe, B., Kowallik, K. V. (1999). Gene-cluster analysis in chloroplast genomics. *Trends Genet* **15**: 344-347.
- Stöbe, B., Maier, U. G. (2001). One, two three: How to get a Plastid?
- Tarassov, I. A., Martin, R. P. (1996). Mechanisms of tRNA import into yeast mitochondria: an overview. *Biochimie* **78**: 502-10.
- ,The Arabidopsis Genom Initiative', (2000). Analysis of the genom sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796-815.
- Timmis, J. N., Scott, N. S. (1983). Spinach nuclear and chloroplast DNAs have homologous sequences. *Nature* **305**: 65-67.
- Timmis, J. N., Scott, N. S. (1984). Promiscuous DNA: sequence homologies between DNA of separate organelles. *Trends Biochem Sci* **9**: 271-273.

- Trotta, C. R., Miao, F., Arn, E. A., Stevens, S. W., Ho, C. K., Rauhut, R., and Abelson, J. N. (1997). The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell* **89**: 849-858
- Van de Peer, Y., Rensing, S. A., Maier, U.-G., De Wachter, R. (1996). Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc Natl Acad Sci (USA)* **93**: 7732-7736.
- Van der Auwera, G., Hofmann, C. J. B., De Rijk, P., De Wachter, R. (1998). The origin of red algae and cryptomonad nucleomorphs: a comparative phylogeny based on small and large subunit rRNA sequences of *Palmaria palmata*, *Gracilaria verrucosa* and the *Guillardia theta* Nucleomorph. *Mol Phylogenet Evol* **10**: 333-342.
- Wakasugi, T., Nagai, T., Kapoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J., Nakashima, K., Tsudzuki, T., Suzuki, Y. et al. (1997). Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: The existence of genes possibly involved in chloroplast division. *Proc Natl Acad Sci (USA)* **94**: 5967-5972.
- Wastl, J., Fraunholz, M. J., Zauner, S., Douglas, S., Maier, U.-G. (1999). Ancient gene duplication and differential gene flow in plastid lineages: The GroEL/Cpn60 example. *J Mol Evol* **48**: 112-117.
- Wastl, J., Maier, U.-G. (2000). Transport of proteins into Cryptomonads complex plastids. *JBC* **275**: 23194-23198
- Wastl, J. (2000). Proteintranslokation in der komplexen Plastide von *Guillardia theta*: Import Kern- und Nucleomorph-kodierter Plastidenproteine und Identifizierung und Charakterisierung neuer plastidärer Funktionen Dissertation, Philipps-Universität Marburg.
- Zhang, H., Berger, J. D. (1999). Two Distinct Classes of Mitotic CyclinHomologues, Cyc1 and Cyc2, are Involved in Cell Cycle Regulation in the Ciliate *Paramecium tetraurelia*.
- Zarzov, P., Boucherie, H., Mann, C. (1997). A yeast heat shock transcription factor (Hsf1) mutant is defective in both Hsc82/Hsp82 synthesis and spindle pole body duplication. *J Cell Sci* **110**: 1879-1891
- Zauner, S., Fraunholz, M., Wastl, J., Penny, S., Beaton, M., Cavalier-Smith, T., Maier, U. G., Douglas, S. (2000) Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph. *Proc Natl Acad Sci (USA)* **97**: 200-205.

## **7. ANHANG**

A: Genomkarte des Nucleomorph-Genoms

B: Sequenz und Struktur der auf Chr II identifizierten tRNAs

C: Sequenzen der snoRNAs

D: Primerliste

### A: Genomkarte des Nucleomorph-Genoms

Die Karte entstand in Zusammenarbeit mit S. Douglas.

Die Kategorien entsprechen denen aus Kapitel 3.2.2.1.





## C: Sequenzen der snoRNAs

### snR1:

Seq: >Nucleomorph-chrII 20.51 (100816-100757) Cmpl: Gt-nm-16S-Am981

Len: 60

Seq:

GCUAAUGAUGAAAAUUUAAUCCGGUUCUGCUUUUGAUUAAUUGAUAAAAUUUCUCUGACA

### snR2:

Seq: >Nucleomorph-chrIII 19.68 (37039-36986) Cmpl: Gt-nm-28S-Um2692

Len: 54

Seq: UAAAAUGAUGCAGACUAGAGUCUUUGAUUAUUAUUGACAAUAAUUAUUACUGAAA

### snR3:

Seq: >Nucleomorph-chrIII 16.18 (140215-140271) Cmpl: Gt-nm-28S-Um30

Len: 57

Seq:

AUAAAUGAUGAUAGUUUGACGGUCUAAUGAUUAUCAAUGAAAAUCAAUUCACCUGAAC

### snR4:

Seq: >Nucleomorph-chrII 9.68 (120005-119942) Cmpl: Gt-nm-28S-Um2562

Len: 64

Seq:

UCACAUGAUGACAACUAAUAAAUGAAAUAUUAUUGACACUUAACCUCUAUAGUUACACUG  
AUU

### snR5:

Seq: >Nucleomorph-chrI 12.52 (32835-32928) Cmpl: Gt-nm-16S-Cm357

Len: 94

Seq:

UGCUAUGAUGAUUUGCUCCUCAAUCUGUUCGCUUAUAAAUUCUUGAAAAUUCUACUAUAAG  
AAUUAUUCGUAAGUAAUCCUUGGAUGUCUGAGC

## D: Primerliste

$T_m = 69,3^{\circ}\text{C} + 0,41 * (\text{GC in } \%) - 650 / \text{Primerlänge}$

### Primer abgeleitet von Chromosom II

Name	Sequenz	T <sub>m</sub> [°C]	Label
2B128R	GTC AGT TTT TCC AAG GAT C	52,4	
2B1U	CAT CGT TCT AAT AAC TTG G	50,2	
2C1R	CGC TCA AGA TTC AAT AG	47,9	IRD 800
2C1U	GTC ATA ATG TCC AAC G	46,6	IRD 700
2E101R	CGA CAT ACT TGT CTT G	46,6	
2E10R	GAG AAT TAG AAC GGA CG	50,4	
2E10U	CCA AAT TCT TTT ACT G	41,5	
2E10U2	GTA ATA TTT GCT AGA CTC	46,9	
2E11	GCC CCT TTT TTA AAC G	46,6	
2E115U	GGC ATC TAT GCA CTC TTG TTG	57,9	
2E123R	CAA ACC AGC TCG TAT TCC	53,7	
2E123U	CTA ATT CTG TAC CCT CG	50,4	
2E125R	CAA TAT GTG GAT TTA CCA GC	53,2	
2E125U	GGT TTG GGT CTT TTC GAT TG	55,3	
2E164_2	GTG TCA CAT CTT CCG TTG CC	59,4	
2E31	CTT CCA TGA TAT AAG ACC C	52,4	
2E47U	GTT GTC AAA TGA ATC AG	45,5	
2E62R	CGA CTA CCA ATA AAG CTG C	54,5	
2E62U	CTC TCG AAT CAC CTA AC	50,4	
2E87R	CAT TGA GCT CAT TAC CTC	51,4	
2ET278R	GTA CAA TCC CAT CCT GC	52,9	
2ET278U	CAA AGT GGT ATG AGA GGT GC	57,3	
2ET290R	GAA GCT GCA AGG GTT AAC G	56,7	
2ET290U	GTG TCA GAC TTC CTT CTG G	56,7	
2G_5	GAT TGG AAT AAT TAT GGG G	50,2	IRD 700
2G_6	CAA CTC CAC TCG GGT CTG	58,2	IRD 800
2G_7	GAA TCG GAG CCA TTT GGA GC	59,4	IRD 700
2G1700	GTC CAA AAA CCA GCT TG	50,4	IRD 700
2G2700	GTA AAA ATC GTC TCT ATT C	48,0	IRD 700
2G3700	GGA TTC CAT TTT GAA TGC TC	53,2	IRD 700
2G4700(335)	GCT TGA AAA CGT TGA GAA TG	53,2	IRD 700
2L3	GGG AAG CAT CTT TAC TCG	53,7	
2LT8X5	GAG ATT ACC AAA CTT CC	47,9	
2R2129_2	GGT TTT TTA TTT CAA AAC TTG GC	53,5	IRD 800
2R2129_3	CCA TTC ATT TTG TCT CTA AG	51,2	IRD 800
2R2129_5	CCA AAA AGG GTT ATA ATT CTC C	54,7	
2RE113F	CAT TCG ATT TAA TCA TCC	46,9	

<b>2RT1200</b>	CAG GTT TTA TCA ACT TAG	50,2
<b>2RT1608</b>	CTC AAT CGA AGA ACG AAA TCG	55,9
<b>2RTH100F</b>	GTA GCA TTT TGA TTG CG	47,9
<b>2RTS2129</b>	GTC AAT AAA CTC GCA GAG	51,4
<b>2RTS327</b>	GAC CTG GAA GAA ATA CC	50,4
<b>2RTV436_1</b>	GAC ATT TAT GAG TGG TGG	51,4
<b>2RTV436f</b>	CGG CAT AGT AAT TGA TG	47,9
<b>2S1608</b>	CGG TAA GAA CCA GTT G	49,2 <i>IRD 800</i>
<b>2S208</b>	GTC GGA ACC TCG TAT AAC C	56,7
<b>2S208_1</b>	CCT CTT AAA GAT GTA CTT G	50,2
<b>2S208_2</b>	CTT ATA TCG CAG ACT GCC	53,7
<b>2SP63R</b>	GCT TTA AAA ACT GCT TGG	49,1
<b>2SP70</b>	GGA AGA TTT ACG CAT TTA G	50,2 <i>IRD 700</i>
<b>2SP78_1a</b>	CAG AGC ATT ACT GCT GCA GCC	61,8
<b>2SP78_1b</b>	CTG TAG CTT CAC TAG GAG G	56,7
<b>2SP78_2a</b>	CCT CCT AGT GAA GCT ACA G	56,7
<b>2SP78_2b</b>	CGT ATA ATT GAA CCC TC	47,9
<b>2UWX19R</b>	CAG TTG AAA ATT GTG GCG	53,7
<b>2UWX19U</b>	GGT AAC CAG TTT TGT TC	47,9
<b>2UWX22R</b>	CGA AAA AAG CAT GAC TC	47,9
<b>2UWX22U</b>	GCT CTT ATG TTG AAT GC	47,9
<b>2X109R</b>	CGA GTA TGG TAA TAT CG	47,9
<b>2X109U</b>	CCC TTT TCA AAT TAG GAA G	50,2
<b>2X10R</b>	CGA TAT CGA CTC TGA GTT G	54,5
<b>2X10U</b>	CTC TCT CCG TTA C	40,0
<b>2X10U2</b>	GAT TTA TAA TAG TCC TAG AG	49,1
<b>2X176R</b>	GCG ACT AAT CAT TTC AG	47,9
<b>2X176R1</b>	CAA TAC TTC ATC GTA ATA G	48,0
<b>2X245U</b>	GCA ATG TAG TTA GAA CTT G	50,2
<b>2X298R</b>	CCT TAC GAT TAC GGA TGG	53,7
<b>2X298U</b>	GTA GTG AAT TTG CCA GCT G	54,5
<b>2X303R</b>	GTA GCA ACT GAT CTT GC	50,4
<b>2X303U</b>	CGC TCA GAT AGC TAA AC	50,4
<b>2X304</b>	CAA GTC ATA ATT TGT GCA CC	52,2
<b>2X308R</b>	GCA TTG GAC TTA AGC G	49,2
<b>2X314</b>	CTA CTT CTC CGT CGA AAG	53,7
<b>2X318R</b>	GAA GGT AGT ACC CAA GAG	53,7
<b>2X318SEQ</b>	GAA TAC AAA CAT GGG TGT C	52,4 <i>Cy5</i>
<b>2X318U</b>	GTG TGA AAA AGG GCA TG	50,4
<b>2X320</b>	GAC AGA AAG ATA CCA TC	47,9
<b>2X327R</b>	GTC TAT CGA ACT TGA C	46,6
<b>2X327U</b>	GAA TCG AAA TCT CCC CAT C	54,5
<b>2X335_1</b>	CTA TAG AAT ATT ATA GTG G	45,9
<b>2X335_2</b>	GAT AAC AGA TTT GTG TTG G	50,2
<b>2X3351</b>	CTA TAG AAT ATT ATA GTG G	45,9
<b>2X3352</b>	GAT AAC AGA TTT GTG TGG G	50,2
<b>2X335R</b>	CAT AAT CAC ATT TCT CG	45,5

2X335U	CAT ATT CTA GAA CAT AAA GGG C	54,7
2X340U	GAA AAG GCA AAG CAA G	46,6
2X341U	GCA AGC GGT TTG ACA G	51,7
2X343_1	CGA GAA TCA ATA CAC AAA CC	53,2
2X343_2	GAC AGC CAT CTT TGC AC	52,8
2X345R	GGC TGA TGG TTT GAT TTG	51,4
2X345U	CCC CTG CAT ACT AAA G	49,2
2X51	CAT CAA GAA CTC CTC TC	50,4
2X67R	CGA GGC AAT GTT TTT GGG ATC	57,9
2X67U	CCC AAA AGA AGC CTT GCC GC	57,3
2X93i	CGA TAA ACT TTC ATT CTG C	50,2
2X93R	CAA GGG TGG CCA ACA AAT C	56,7
2X93U	GCT ATA ATC ATC TCG G	46,6
2UWX16R	CCC ACA TAC CAT TAA AG	47,9
2UWX16U	GTT TAA AGT TCG ATG GCT CAC C	58,4
2UWX31R	CCT CTA TTG AGG TTA CC	50,4
2UWX31U	CGG ATC CAT GAC ACG TTT TAT G	58,4

### Primer abgeleitet von Chr III

Name	Sequenz	Tm [°C]	Label
3L008R1	GGC GAA CAA TTT CAA GC	50,4	
3P123_7	GCT ACA TAG TCC TTA TGT TG	53,2	<i>IRD 700</i>
3P123_8	GAA ATC TTA AAG ATC CCG C	52,4	<i>IRD 800</i>
3RT2S76	CAA GGT AGA AAT CAG GAG	51,4	
3RTH630	CCT TCC ACA TAT CAA ACG	51,4	
3RTS390	GGA ACA CTG GGA TTC	47,8	
3RTX880	GAT TGT TGA AGA CCA CAG ATT AG	57,1	
3X020	GGT TCA TTT AGA CTC TAC	49,1	
3X046	CGT CAA TCA TAG AAT TCT TG	51,2	
3X1183	CAC GTA GTT TAA GTT GCC	51,4	
3X124_3'	CAC TCG ATA TCC TTG GTG TC	57,3	
3X1394	GCA TGA TGG ATA CTT CTA GC	55,3	
3X192	GTA AGT TTT AAC GGT TAT C	48,0	
3X302R	CTT CCG GTT TTC CCT AG	52,8	
3P_X7	CTT ATT CCT TTT CAG TTT CCC	54,0	
3P_X8	CAA TTA ATT CTA GCC TAG AGC	54,0	
3_7r	CCC ATT AGT TCT TCT TCG	51,4	
3_7l	CCT TCC ACT ATC ATC AAT AAC C	56,5	
3_9e1	GAG ATA GGA GAT GGG ACA AC	57,3	
3_9e2	CGA AAT CGA ATC CCA TAA ACC	55,9	
3RTH486F2	GCC TGA TGA TAC CAG TAG	53,7	
3RTH486F3	GAA TCT GAC AAA TTA GTA TC	49,1	

## Primer abgeleitet von Chromosom I

Name	Sequenz	Tm [°C]	Label
<b>1E204</b>	CTT TGT CGG ACT CGA ATG G	56,7	
<b>1P_D1_7</b>	CGG AGA CGT TGC GTG TG	57,6	<i>IRD 700</i>
<b>1P_D1_8</b>	CGT ACC ATA AAG TCC ATG G	54,5	<i>IRD 800</i>
<b>1P_D72</b>	GAT TCC AGA AGC ACT CCA AG	57,2	
<b>1V459_3'</b>	GCT GTA TTA ATG CTG CAT AC	53,2	
<b>1V459-5'</b>	CAT GGC TGA CTA AGG CCA G	58,8	
<b>1P_D12</b>	GCC CTT CGA TGT AGG TTT G	56,7	
<b>1M128_1</b>	CGA TAT TAG TAG AAG ACC AC	53,2	
<b>1M128_2</b>	GAA AGG GTA GCC AAA GCC	56,0	
<b>1M128_3</b>	GAG AAC GTT AAC TAA GTT ATC G	54,7	
<b>1E218_1</b>	CGG GTC TGA TTT CGA AGA AC	57,3	
<b>1X212_1</b>	GTG GAT CAT ACT TTA TAC CC	53,2	
<b>1X006_1</b>	GAT CCA AAC CAA TTG AGT C	52,4	
<b>1X006_2</b>	GTA TTA CCC GAA ACT GTG G	54,5	
<b>1X182F2</b>	CCA CAG TTT CGG GTA ATA C	54,5	
<b>1X006_3</b>	GCC TTG AGT AAC AAA GAA TG	53,2	
<b>1V219_1</b>	GAT GTT TAA ACA AGG TGT GC	53,2	
<b>1E218_2</b>	GAT GGT AGT CTA GGA TCA C	56,7	
<b>1X006_3</b>	CGA GTA GCA TTG CTA ATA GAC	55,9	
<b>1X212_5</b>	GTT AAT TCC TGC AAA GAG GG	55,3	
<b>1Dc</b>	GTG CAA CAC CAT CTA TTG CC	57,3	
<b>1X182F3</b>	CAA TAA AAT TGA TTT CCG TGA G	52,8	
<b>EIF2_1</b>	CCT AGC TGT TAC CAA TCG C	56,7	
<b>1_new3'</b>	GAT ACA TGA TTA TGC CCC	51,4	
<b>1_new5'</b>	GAT TCA CAC GCA ACG TCT CCG	61,0	