

# **THE CHARACTERISATION AND PARTIAL SEQUENCING OF THE GRAPEVINE CHLOROPLAST GENOME.**

by

**B.A. Rose**

Thesis presented in partial fulfillment of the requirements for the degree of  
Master of Science at the University of Stellenbosch.



April 2004

Study leader: Prof J.T. Burger

## **DECLARATION**

I the undersigned hereby declare that the work contained in this thesis is my own original work and has not been previously in its entirety or in part been submitted at any university for a degree.

## ABSTRACT

A number of proteins essential for the survival of a plant are encoded by the chloroplast genome. The characterization and sequencing of a number of algal and plant chloroplast genomes has facilitated researchers understanding of cellular functions and metabolism. Chloroplast DNA (cpDNA) has also been used to determine inter- and intraspecies evolutionary relationships and this organelle offers an alternative means of expressing foreign genes. Although a number of species' chloroplast genomes have been characterized and sequenced, no previous attempts of this kind have been made for a chloroplast genome of the family *Vitaceae*.

In this study, attempts were made to characterize and partially sequence the chloroplast genome of *Vitis vinifera*. Chloroplast DNA was isolated from the Sultana and Sogra 1 cultivars and digested with restriction enzymes that produced cpDNA fragments of a suitable size for cloning. The fragments were shotgun-cloned into a plasmid vector and white colonies were screened by means of PCR and colony blotting. Three *EcoRI*-digested clones and one *PstI*-digested clone were obtained in this manner. Walking outwards from a previously sequenced grapevine *rrn16* gene region by means of PCR also allowed us to sequence a further ~3310 bp region of the Sultana chloroplast genome.

BAC clones containing *V. vinifera* cv L. Cabernet Sauvignon cpDNA inserts became available later in the project. It was decided to use these clones for further library construction instead of isolated cpDNA. The 5' and 3' end sequences of seven of the 24 BAC clones were obtained. These were compared to sequences found in the NCBI database to find homologous chloroplast regions and determine the size of each BAC insert. One clone appeared to contain the entire grapevine chloroplast genome, apart from a 500 bp region. This clone was selected for further analysis. The BAC clone DNA was isolated and restriction-digested fragments were shotgun-cloned into a plasmid vector. White colonies were screened by isolating the plasmid DNA and digesting it with appropriate restriction enzymes. The 5' and 3' ends of putative positive clones were sequenced and mapped onto the *Atropa belladonna* chloroplast genome.

A total of 15 clones were obtained in this project. Five of these contain cpDNA isolated from grapevine leaves and 10 contain fragments sub-cloned from the BAC clone. The biggest

problem encountered with both methods used for library construction was genomic DNA contamination. Genomic DNA either originated from the plant nuclear genome or from the bacterial host cells in which the BAC clones were maintained. Many of the clones screened contained genomic DNA, and these could only be identified and removed once the clones had been sequenced. Even when a commercial kit was used for BAC clone isolation, 31% of the clones screened contained genomic DNA. This kit was specifically designed for the isolation of genomic DNA-free large constructs.

The clones obtained from the two strategies provided a good representation of the grapevine chloroplast genome. The only region not represented was the Small Single Copy (SSC) region. Approximately 40% of the grapevine chloroplast genome was covered by these clones. This provides a basis for further genome characterization, physical mapping and sequencing of the grapevine chloroplast genome.

## OPSOMMING

Die chloroplasgenoom kodeer vir 'n hele aantal proteïene wat essensieel is vir die voortbestaan van 'n plant. Die karakterisering en volgorde bepaling van 'n aantal alg en plant chloroplasgenome het dit vir navorsers moontlik gemaak om sellulêre funksies en metabolisme van plante te ontrafel. Chloroplas DNA (cpDNA) is ook gebruik om intra- en interspesies evolusionêre verwantskappe vas te stel. Dié organel verskaf ook 'n alternatiewe manier vir die uitdrukking van transgene. Alhoewel die chloroplasgenome van 'n hele aantal spesies al gekarakteriseer is en die DNA volgorde daarvan bepaal is, is daar nog geen navorsing van bogenoemde aard op die chloroplasgenoom van die Vitaceae familie gedoen nie.

In hierdie studie is beoog om die chloroplasgenoom van *Vitis vinifera* te karakteriseer en gedeeltelike volgordebepaling daarvan te doen. Chloroplas DNA is geïsoleer vanaf Sultana en Suga 1 kultivars en restriksie-ensiem vertering is gedoen met ensieme wat cpDNA fragmente, met geskikte grootte vir klonering, produseer. Dié fragmente is in 'n plasmiedvektor gekloneer met die haelgeweer-metode en wit kolonies is gesif deur middel van PKR en die kolonieklaar metode. Op hierdie manier is drie *EcoRI*-verteerde klone en een *PstI*-verteerde kloon verkry. Deur uitwaarts te loop, deur middel van PKR, vanaf 'n druif *rrn16* geenstreek, waarvan die volgorde vooraf bepaal is, was dit vir ons moontlik om ook die volgorde te bepaal van 'n verdere ~3310 bp streek van die Sultana chloroplasgenoom.

BAC klone wat *V. vinifera* cv L. Cabernet Sauvignon cpDNA fragmente bevat, het later in die projek beskikbaar geraak. Daar is besluit om hierdie klone, i.p.v. die geïsoleerde cpDNA, te gebruik vir verdere biblioteek konstruksie. Die 5' en 3' entpuntvolgordes van sewe uit die 24 BAC klone is verkry. Hierdie volgordes is vergelyk met volgordes in die NCBI databasis om homoloë chloroplas streke te identifiseer, en die grootte van elke BAC fragment te bepaal. Dit het geblyk dat die hele druif chloroplasgenoom in een van die klone vervat is, behalwe vir 'n 500 bp streek. Die BAC-kloon DNA is geïsoleer en die restriksie-verteerde fragmente is in 'n plasmiedvektor gekloon d.m.v. die haelgeweer-metode. Wit kolonies is gesif deur die isolering van plasmied DNA en die vertering daarvan met geskikte restriksie-ensieme. Die volgorde van die 5' en 3' entpunte van skynbare positiewe klone is bepaal en gekarteer op die *Atropa belladonna* chloroplasgenoom.

In hierdie studie is 'n totaal van 15 klone verkry. Vyf hiervan bevat cpDNA wat vanaf druifblare geïsoleer is, en 10 bevat fragmente wat vanaf die BAC-klone gesubkloneer is. Genomiese DNA kontaminasie was die grootste probleem wat ondervind is tydens beide metodes wat gebruik is vir biblioteek konstruksie. Genomiese DNA was afkomstig vanaf óf die plant nukleêre genoom óf die bakteriële gasheerselle waarin die BAC-klone gehou is. Baie van die klone wat gesif is, het genomiese DNA bevat, en dit kon eers geïdentifiseer en verwyder word nadat die volgorde van die klone bepaal is. Selfs al is 'n kommersiële produk vir BAC-kloon isolasie gebruik, het 31% van die gesifde klone steeds genomiese DNA bevat. Dié kommersiële produk is spesifiek vir die isolasie van groot konstruksie, wat genomiese DNA vry is, ontwerp.

Die klone wat deur die twee strategieë verkry is, het 'n goeie verteenwoordiging van die druif chloroplasgenoom gegee. Die enigste streek wat die verteenwoordig is nie, was die Klein Enkelkopie (SSC) streek. Ongeveer 40% van die druif chloroplasgenoom is deur hierdie klone gedek. Dit verskaf 'n basis vir verdere genoomkarakterisering, fisiese kartering en volgordebepaling van die druif chloroplasgenoom.

## **ACKNOWLEDGEMENTS**

I would like to express my gratitude and appreciation to the following people:

- My supervisor, Prof. J.T. Burger, for your advice and guidance throughout this project.
- Michael-John for your constant enthusiasm and encouraging words.
- Shelly Deane for her advice on colony blotting.
- Winetech, THRIP, and Stellenbosch 2000 for financial support.
- The Vitis lab (Helena, Chris, Helen, Julia, Jacques, Mandi) for their companionship and interesting discussions.
- Last, but not least, to my friends and family for their constant support and encouragement at all times, throughout this project.

## LIST OF ABBREVIATIONS AND SYMBOLS

$\alpha$	Alpha
$\lambda$	lambda
$A_{600}$	Absorption value
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
bp	Base pairs
BSA	Bovine serum albumin
$^{\circ}\text{C}$	Degrees Celsius
$\text{CaCl}_2$	Calcium chloride
CDP- <i>Star</i> <sup>TM</sup>	Disodium 4-chloro-3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'-chloro)tricyclo[3.3.1.1.3,7]decan}-4-yl) phenyl phosphate
$\text{cm}^2$	Centimeters squared
$\text{CO}_2$	Carbon dioxide
cpDNA	Chloroplast DNA
CsCl	Cesium chloride
CTAB	N-cetyl-N,N,N-trimethyl ammonium bromide
ddH <sub>2</sub> O	Double distilled water
DIG	Digoxigenin-dUTP
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
EDTA	Ethylene diamine tetra-acetic acid di-sodium salt
h	Hours
IPTG	Isopropylthiogalactoside
Kb	Kilobases
KOH	Potassium hydroxide
LB	Luria Bertani
M	Molar
mg	Milligrams
$\text{MgCl}_2$	Magnesium chloride
$\text{MgSO}_4$	Magnesium sulphate
min	Minutes



ml	Millilitres
mm	Millimeters
mM	Millimolar
MnCl <sub>2</sub>	Manganese chloride
MOPS	(3-[N-morpholino]propanesulfonic acid)
mRNA	Messenger ribonucleic acid
NaCl	Sodium chloride
NaOAc	Sodium acetate
NaOH	Sodium hydroxide
ng	Nanograms
NH <sub>4</sub>	Ammonium
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	Ammonium sulphate
nm	Nanomoles
PAC	P1-derived artificial chromosome
PCR	Polymerase chain reaction
PIPES	1,4-Piperazinediethanesulfonic acid
pm	Picomoles
PVP	Polyvinylpyrrolidone
RbCl	Rubidium chloride
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulphate
SSC	Sodium chloride Sodium Citrate
TAE	Tris-Acetate-EDTA
TBE	Tris-Borate-EDTA
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol
tRNA	Transfer ribonucleic acid
U	Units
V	Volts
v/v	Volume per volume
w/v	Weight per volume
X-gal	5-bromo-4-chloro-3-indocyl-β-D-galactoside
YAC	Yeast artificial chromosome

# CONTENTS

<b>ABSTRACT</b>	<b>III</b>
<b>OPSOMMING</b>	<b>V</b>
<b>ACKNOWLEDGEMENTS</b>	<b>VII</b>
<b>ABBREVIATIONS</b>	<b>VIII</b>
<b>CONTENTS</b>	<b>X</b>
<b>1 INTRODUCTION</b>	<b>1</b>
<b>2 LITERATURE REVIEW</b>	<b>4</b>
<b>2.1 THE CHLOROPLAST</b>	<b>4</b>
<b>2.1.1 Structure and Function</b>	<b>4</b>
<b>2.1.2 Plastid Types</b>	<b>6</b>
<b>2.1.3 Plastid Origin and Evolution</b>	<b>7</b>
<b>2.1.3.1 Origin</b>	<b>7</b>
<b>2.1.3.2 Evolution</b>	<b>7</b>
2.1.3.2.1 Cyanobacterial Ancestor	<b>7</b>
2.1.3.2.2 Primary Endosymbiosis	<b>9</b>
2.1.3.2.2.1 <i>Green algae</i>	<b>12</b>
2.1.3.2.2.2 <i>Red algae</i>	<b>16</b>
2.1.3.2.2.3 <i>Glaucophytes</i>	<b>16</b>
2.1.3.2.3 Secondary Endosymbionts	<b>16</b>
<b>2.1.4 The Chloroplast Genome</b>	<b>20</b>
<b>2.1.4.1 Size and Structure</b>	<b>21</b>
<b>2.1.4.2 Gene Arrangement, Function</b>	
<b>and Expression</b>	<b>23</b>
<b>2.1.4.3 Protein Import</b>	<b>24</b>
<b>2.1.4.4 Inheritance</b>	<b>25</b>
<b>2.1.4.5 Genetic Mutations</b>	<b>26</b>
<b>2.1.5 The <i>Vitis vinifera</i> Chloroplast Genome</b>	<b>27</b>
<b>2.1.6 Applications of Sequenced Chloroplast</b>	
<b>Genomes</b>	<b>27</b>
<b>2.1.6.1 Phylogenetic Studies</b>	<b>27</b>
<b>2.1.6.2 Plastid Transformation</b>	<b>28</b>

2.1.6.3	Drug Targets	31
2.2	<b>CHLOROPLAST GENOME</b>	
	<b>CHARACTERISATION AND SEQUENCING</b>	31
2.2.1	Chloroplast DNA Isolation	32
2.2.2	Restriction Digestion, Physical and Genetic Map Construction	35
2.2.3	Clone Library Construction and Genome Sequencing	36
2.3	<b>LARGE-CONSTRUCT VECTORS</b>	37
2.3.1	Cosmids	38
2.3.2	Yeast Artificial Chromosomes (YACs)	38
2.3.3	Bacterial Artificial Chromosomes (BACs)	39
2.4	<b>CONCLUSION</b>	40
3	<b>MATERIALS AND METHODS</b>	42
3.1	<b>ISOLATION AND CHARACTERISATION OF <i>VITIS VINIFERA</i> CHLOROPLAST DNA</b>	42
3.1.1	Plant Material	42
3.1.2	Isolation of <i>V. vinifera</i> Chloroplast DNA	42
3.1.3	Restriction Enzyme Digestion	43
3.1.4	Library Construction	43
3.1.4.1	Shotgun Cloning	43
3.1.4.2	Preparation of Competent Cells	44
3.1.4.3	Transformation	45
3.1.5	Library Screening by PCR	45
3.1.6	Library Screening by Colony Blotting	45
3.1.6.1	Colony Lifts	45
3.1.6.2	Probe labeling	46
3.1.6.3	Prehybridisation and Hybridisation	46
3.1.6.4	Detection	47
3.1.7	Plasmid DNA Isolation	47
3.1.8	Clone Sequencing and Sequence Analysis	48
3.1.9	PCR of the <i>rrn16</i> Flanking Region	48
3.2	<b>BAC CLONE CHARACTERISATION AND SEQUENCING</b>	51

3.2.1	BAC Clone Storage	51
3.2.2	BAC-End Analysis	51
3.2.3	Isolation of BAC Clone DNA	51
3.2.4	Restriction Digestion of BAC clone DNA	52
3.2.5	Library Construction	53
3.2.6	Colony Screening	53
3.2.7	Sub-Clone Sequencing and Sequence Analysis	54
4	RESULTS AND DISCUSSION	55
4.1	ISOLATION AND CHARACTERISATION OF <i>VITIS VINIFERA</i> CHLOROPLAST DNA	55
4.1.1	Chloroplast DNA Isolation	55
4.1.2	Restriction Digestion	57
4.1.3	Library Construction and Library Screening using PCR	58
4.1.4	Library Screening by Colony Blotting	59
4.1.5	Clone DNA Isolation and Sequencing	60
4.1.6	PCR of the <i>rrn16</i> Flanking Regions	65
4.2	BAC CLONE CHARACTERIZATION AND SEQUENCING	66
4.2.1	BAC-End Analysis	66
4.2.2	Isolation of BAC Clone DNA	66
4.2.3	Restriction Digestion	67
4.2.4	Subcloning and Colony Screening	68
4.2.5	Sequencing and Sequence Analysis	69
4.3	GENERAL DISCUSSION	78
5	CONCLUSION	81
	APPENDIX 1	84
	APPENDIX 2	87
	APPENDIX 3	89
	APPENDIX 4	96
	REFERENCES	99

*Picture a palm tree growing peacefully on the shore of a spring, and a lion, lying hidden beside the palm, all its muscles tense, blood lust in its eyes, ready to pounce on an antelope and slaughter it. In order to understand fully the inner secret of this picture with its two such drastically different manifestations of life, a palm tree and a lion, it is essential to appreciate the theory of endosymbiosis. The life of the palm tree is so calm and peaceful because it is a symbiosis, it contains a legion of workers, green slaves (plastids) that work for it and nourish it. The lion has to feed itself.*

*Imagine that every cell of the lion's body was filled with plastids, and I have no doubt that it would immediately lay peacefully by the palm, feeling replete with nothing more than some water and a few nutrient salts.' Mereschkowsky, 1905*

# 1. INTRODUCTION

## 1.1 General Introduction

The chloroplast is essential for the survival of all photosynthetic algae and plants. This organelle is believed to have originated from an endosymbiotic event between a primitive eukaryotic host cell and a cyanobacterial progenitor, and has since become integrated into the host cell's metabolism. Existing chloroplasts have lost the majority of their DNA, but still encode for a number of proteins involved in photosynthesis. The eukaryotic cell has also become dependant on the energy produced during photosynthesis by the chloroplast.

The chloroplast genome is relatively small (120 to 160 kb in plants) compared to the nuclear genome, making it ideal for inter- and intraspecies phylogenetic studies. Restriction-site mutations have traditionally been used for these types of studies (Palmer et al. 1985). However, rapid improvements in cloning and sequencing technology have facilitated the process of chloroplast genome characterization and sequencing. Phylogenetic studies involving whole genome sequences are now being undertaken, offering more reliable results that can assist in clarifying evolutionary relationships between photosynthetic eukaryotes (O'Kane 1995).

The use of chloroplast genomes for foreign gene expression has become a viable alternative to nuclear transgene expression (McFadden 2001a). The chloroplast genome is predominantly maternally inherited and, therefore, foreign genes are less likely to be transferred to weedy relatives or non-transgenic relatives. The high ploidy level of chloroplast genomes per plant cell can allow for higher levels of foreign gene and protein expression in the chloroplast than those encountered in nuclear transgene expression. Chloroplast genomes have a prokaryotic nature and, as a result, allow the simultaneous co-transcription of a number of genes into a polycistronic mRNA. The genes for entire metabolic pathways could therefore be integrated into and expressed in the chloroplast genome with more ease than in the nuclear genome, where genes are transcribed monocistronically. The biggest advantage of chloroplast transformation is that genes are integrated into the chloroplast genome by means of homologous recombination. Consequently, transgenes can be targeted to a specific region of the chloroplast genome, avoiding the position effect that often occurs in nuclear transformants (Hager and Bock 2000, Daniell et al. 2002).

The characterization and sequencing of chloroplast genomes provides valuable information that can be applied in a number of areas. As already mentioned, this information facilitates evolutionary studies. Chloroplast DNA (cpDNA) sequences are also used to design transgene flanking regions, to allow targeted integration of the foreign DNA into the chloroplast genome. The chloroplast genomes of a number of plant species have been characterized and sequenced. However, no complete chloroplast genome sequences are available for fruit-producing species and only the chloroplast genomes of two woody species have been completely sequenced. The characterization and sequencing of the chloroplast genome of a species such as *Vitis vinifera* could therefore provide novel information.

All grapes belong to the family Vitaceae and the genus *Vitis*, which has been divided into two sections; namely *Vitis (Euvitis)* and *Muscadinia*. These are further subdivided into 11 series and more than 60 species, containing over 14 000 cultivars. True grapevines are found in the *Vitis (Euvitis)* group (Alleweldt et al. 1991, Lodhi and Reisch 1995) and *Vitis* species are regarded as one of the most important agronomic plant species in the world (Alleweldt et al. 1991, Lodhi and Reisch 1995). *V. vinifera*, in particular, plays an important role in the production of wine. In South Africa alone 108 000 hectares under vine are utilised for wine production and 9.7% of the Western Cape's gross domestic product is provided by the wine industry (SA Wine Industry Information & Systems 1999, <http://www.wosa.co.za>).

The *Vitis* species has become quite varied over the years, making it difficult to determine its history. Domesticated *V. vinifera* is thought to have first been cultivated in the Middle East somewhere at the foothills of the Caucasus Mountains. This was concluded from evidence of 5000 year-old seeds found at Jericho. From there, this species was thought to have spread via trade channels around the Mediterranean Basin to the Far East and Europe and later to the New World (Alleweldt et al. 1991). However, recent studies performed on the cpDNA of more than 500 varieties of grapevine grown around the Mediterranean Sea, as well as several wild isolates, have come up with a different theory for the origin of grapevine domestication. Unlike nuclear DNA, cpDNA is maternally inherited and has been relatively well conserved through evolution. Chloroplast-specific molecular markers were therefore used to compare the above-mentioned cultivars and determine their similarity to each other. The studies concluded that wild vines and domesticated vines have strong homology within their cpDNA. This implies that they are closely related and that the wild vines could have been independently domesticated in a number of areas, rather than one line of domesticated vine

having spread from one specific region (Garcia et al. 2002b). However, more phylogenetic studies using *V. vinifera* cpDNA sequences are needed to determine the correct theory.

*V. vinifera* is at a high risk of contracting a number of pests as well as fungal, bacterial and viral diseases, causing large-scale losses of grape yields worldwide. Methods of combating these diseases include spraying the crops with insecticides and pesticides (Boubals 2000). However, these methods are expensive and cause environmental pollution. The incorporation of foreign resistance genes into the grapevine could serve as an alternative method of fighting pests and diseases, as well as helping the grapevine withstand other tough environmental conditions. This will naturally lead to improved grape quality and yields. Grapevine chloroplasts are maternally inherited (Garcia et al. 2002a) and, therefore, chloroplast-specific transgene expression could offer an environmentally safe method of controlling pests and diseases.

## 1.2 Project Proposal

The initial goal of this study was to generate a library of overlapping clones covering most of the grapevine chloroplast genome. These clones would then be characterized by partial sequencing and mapped against completely sequenced chloroplast genomes to determine their order.

The more specific aims of this study are:

- The construction of a clone library of overlapping cpDNA fragments using cpDNA isolated from either grapevine leaves or BAC clones containing grapevine cpDNA inserts.
- The characterization and partial sequencing of the clone library.
- The organization of the clones in the correct order by mapping them against completely sequenced plant chloroplast genomes.
- Physical mapping and partial genetic mapping of the grapevine chloroplast genome.



## 2. LITERATURE REVIEW

### 2.1 THE CHLOROPLAST

The chloroplast is an organelle found in all phototrophic eukaryotes. Its most important function of the chloroplast is photosynthesis, for which it contains the entire enzymatic machinery (Sugiura 1992, Kato et al. 2000). The metabolic energy generated during photosynthesis drives all photoautotrophic plant growth that forms an indispensable part of the ecosystem (De Las Rivas et al. 2002).

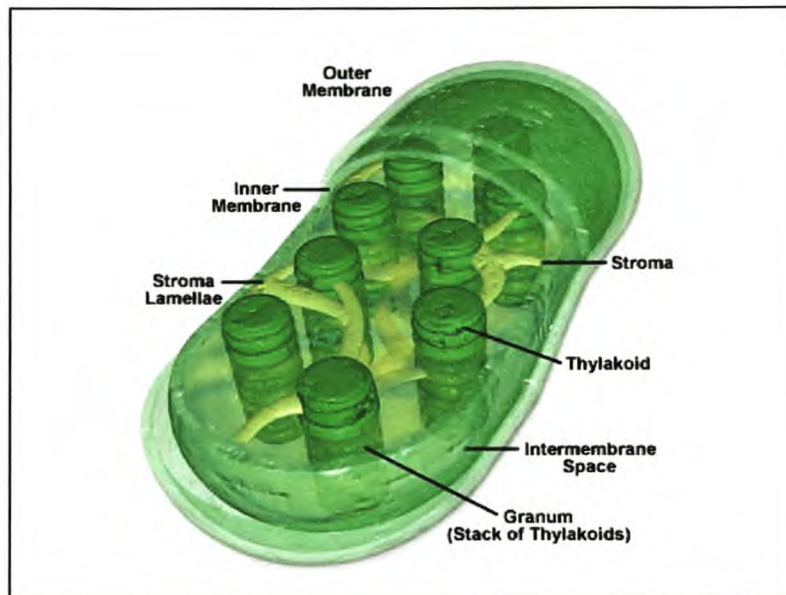
The chloroplast is derived from an endosymbiotic event in which a cyanobacterial progenitor was engulfed by a primitive eukaryotic host cell and integrated into the cell's metabolism. This primary endosymbiotic event gave rise to red algae, green algae, and glaucophytes. Two or more secondary endosymbiotic events have since taken place. These involved the engulfment of an already existing red or green alga by an eukaryotic host cell and gave rise to most of the algal diversity found today (Cavalier-Smith 2000, Archibald and Keeling 2002).

A number of plastid types can be found in algae or plants. Each of these contains different pigments but the same genome. The chloroplast genome of higher plants is relatively conserved in genome size and gene arrangement, but those of algae show a higher degree of variation. Higher plant chloroplast DNA (cpDNA) is divided into four regions. Two identical inverted repeats (IRs) separate the large single copy (LSC) and small single copy (SSC) regions. Only about 10% of the genes required by the chloroplast are encoded by its own genome. The remaining proteins are transcribed by the host nuclear genome and imported into the chloroplast.

#### 2.1.1 Structure and Function

Chloroplasts are oval to lenticular in shape with dimensions of 2 to 4  $\mu\text{m}$  by 5 to 10  $\mu\text{m}$ . They are surrounded by a double membrane, enclosing a more complex third membrane system. The outer chloroplast membrane is continuous and contains porins, making it completely permeable to small molecules. The inner membrane, on the other hand, is differentially permeable, allowing molecules to enter only through specific membrane transporters. The cavity enclosed by the inner membrane is called the stroma and contains DNA, ribosomes, lipid droplets and starch granules. The third membrane system is found in

the stroma where it forms a series of flattened discs called thylakoids. Most thylakoids are orientated parallel to the length of the plastid and are often arranged in stacks called grana. These are interconnected by stromal lamellae and together they form a complex photosynthetic system of which chlorophyll molecules form an integral part. The membrane of each granum encloses a thylakoid lumen (Bowes 1996, Cooper 2000).



**Figure2.1.** Green algae and higher plant chloroplast structure ([www.angelfire.com/on2/daviddarling/chloroplast.jpg](http://www.angelfire.com/on2/daviddarling/chloroplast.jpg)).

More than 100 chloroplasts may be found per leaf mesophyll cell, but numbers vary in different plant tissues and plant species (Hager and Bock 2000). Each chloroplast contains a number of circular DNA molecules that are unique to this organelle. A single leaf cell can contain more than a thousand genetically identical copies of the chloroplast genome, while cells found in other organs of the plant have 10 times fewer copies (Heifetz 2000). The genomes in a chloroplast are clustered together in groups called nucleoids. Proplastids and chromoplasts mostly contain only one nucleoid, while chloroplasts have many nucleoids are found throughout the stroma (Palmer 1987).

During photosynthesis, light energy is captured by the chloroplast's photosynthetic machinery and used to power the conversion of  $\text{CO}_2$  into carbohydrates in the stroma (Smith and Ma 1985, Cooper 2000). The thylakoid membrane plays a very important part in the functioning of the chloroplast since it contains the electron transport system. This system allows protons to be transported into the thylakoid lumen, creating an electrochemical gradient. The gradient, in turn, causes protons to move back across the membrane into the stroma and in the

process ATP is synthesized (Cooper 2000). The energy produced during photosynthesis is made available for a number of processes in the cell. These include a number of metabolic pathways, such as nitrate reduction and sulphur assimilations, as well as the synthesis of amino acids, nucleotides, fatty acids, starch, vitamins, plant hormones and pigments (Palmer 1987, Sugiura 1992, Kunnimalaiyaan and Nielsen 1997, Leister 2003).

### **2.1.2 Plastid Types**

A number of different plastid types are found in a plant, each having different morphologies and functions in the cell. However, only one type of plastid is usually found in a cell at a given time (Watson and Murphy 1999). Two membranes envelop all plastids and chloroplasts contain a third, thylakoid membrane system (Bowes 1996, Cooper 2000).

The pigments found in plastids often play a role in their classification. Chloroplasts all contain chlorophyll. Chromoplasts contain carotenoids and give a yellow, orange or red colour to some fruits and flowers. Leucoplasts do not contain any pigments and are responsible for storing a number of energy sources, such as starch and lipids. Proplastids are the precursors of all other plastid types, and are found in plant roots and shoots. Mature plastids also have the ability to convert from one type to another. For example, chloroplasts change into chromoplasts during fruit ripening. During this process the chlorophyll and thylakoid membranes are degraded, while carotenoids are synthesized (Bowes 1996, Cooper 2000).

Certain environmental signals and built-in cell differentiation programs regulate plastid development. In a photosynthetic leaf cell protoplasts differentiate into chloroplasts in the presence of light. The thylakoid membrane is formed by invagination of the inner membrane and the machinery needed for photosynthesis are synthesized and assembled. However, if a plant is stored in the dark, the development of the chloroplast is halted and only half-formed thylakoid membranes are present with no chlorophyll having been produced. A plastid in this intermediate stage is called an etioplast. If the plant is again exposed to light, the etioplasts will complete their metamorphosis into mature chloroplasts (Cooper 2000).

## 2.1.3 Plastid Origin and Evolution

### 2.1.3.1 Origin

The important role the chloroplast plays in the establishment of phototrophic eukaryotes makes its origin one of the most important events in eukaryotic cell evolution (Archibald and Keeling 2002). The endosymbiotic theory describes the most likely manner in which these organelles originated. According to this theory the plastid was originally a free-living cyanobacterial progenitor that was engulfed by a primitive (non-photosynthetic) eukaryotic host cell and enclosed in a vacuole. Instead of being digested, the cyanobacterium managed to escape the vacuole into the cytosol and was gradually integrated into the metabolism of the host cell. An endosymbiotic relationship developed between the host cell and the organelle, with the photosynthetic plastid providing the host with some of its produced carbohydrates and photosynthetic energy and the host providing the plastid with other compounds and protection from the environment (Cavalier-Smith 2000, Archibald and Keeling 2002). The mitochondrion originated in a similar way, but with an  $\alpha$ -Proteobacteria (purple sulphur bacteria) as its ancestor (Gray 1999).

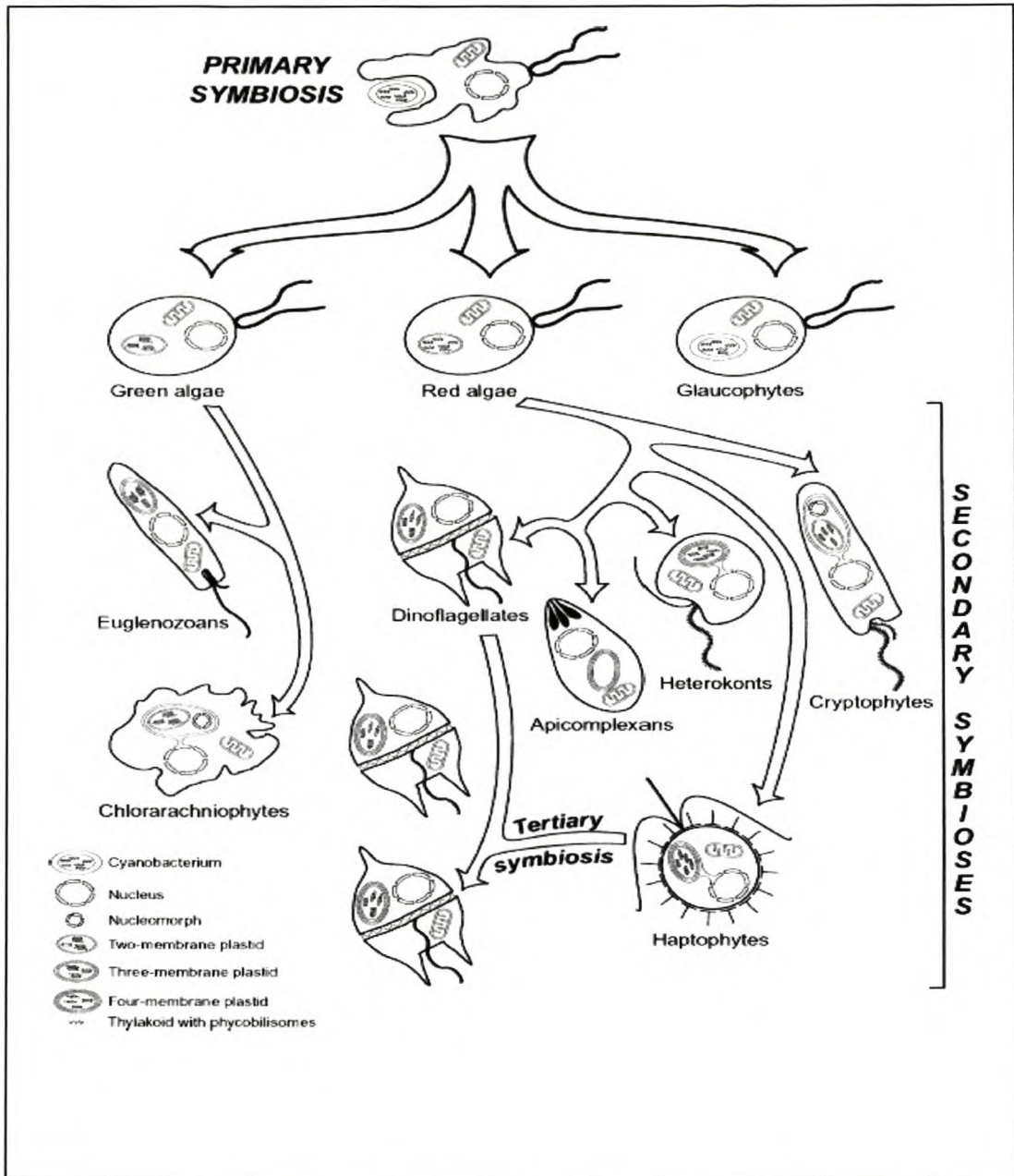
Primary endosymbiosis is believed to have taken place at least two billion years ago (Raven et al. 2002), giving rise to all plastids. There is, however, evidence of a second event that gave rise to most of the plastid diversity found today. This involved the engulfment of an already existing phototrophic eukaryotic alga by another eukaryote and is known as secondary endosymbiosis (Bhattacharya and Medlin 1998, Archibald and Keeling 2002).

### 2.1.3.2 Evolution

#### 2.1.3.2.1 Cyanobacterial Ancestor

There is a substantial amount of evidence suggesting that the plastid progenitor originated from a cyanobacterial ancestor. Firstly, all cyanobacteria except *Gloeobacter* have thylakoids. However, no other Gram-negative bacteria contain thylakoids (Cavalier-Smith 2000). Plastid genome organization and gene order also show similarities to that of cyanobacterial genomes, including the conservation of many large operon structures (Ohya et al. 1986, Douglas and Penny 1999, Stoebe and Kowallik 1999, McFadden 2001b). Phylogenetic studies using plastid and cyanobacterial gene sequences have confirmed their

close evolutionary relationship beyond all doubt (Bhattacharya and Medlin 1998, McFadden 2001a).



**Figure 2.2.** The evolutionary pathway of photosynthetic eukaryotes and their plastids. Primary endosymbiosis between a cyanobacterium and a phagotrophic eukaryote gave rise to green algae, red algae and glaucophytes. Euglenozoans and chlorarachniophytes acquired green algae secondarily, while red algae were acquired secondarily by alveolates, heterokonts, cryptophytes and haptophytes. These events gave rise to secondary plastids. Some dinoflagellates have replaced their secondary red algal plastids with green algal- or haptophyte-type plastids through secondary replacement and tertiary symbiotic processes (Moreira and Philippe 2001).

Even though plastids appear to have a cyanobacterial origin, the exact nature of this predecessor is not known. Plastid genomes show no morphological or phylogenetic similarity to any one particular cyanobacterial group, suggesting that modern cyanobacteria have diversified to a large degree since the primary endosymbiotic event occurred. Further characterization of cyanobacteria and their genomes should provide researchers with more information on this ancestor, while comparisons of these genomes to existing plastid genomes could also provide more clues to plastid genome origin and evolution (McFadden 2001b, Archibald and Keeling 2002, Palmer 2003, Stiller et al. 2003).

New information on plastid origin has already been provided by the use of the complete genomes of the cyanobacteria *Synechocystis* (Kaneko and Tabata 1997) and *Nostoc* (Kaneko et al. 2001) in a number of phylogenetic comparisons with plastid genomes (Martin et al. 1998, De Las Rivas et al. 2002, Martin et al. 2002, Maul et al. 2002). For example, the placement of the glaucophyte alga, *Cyanophora paradoxa*, indicates that it is one of the earliest diverging plastid-containing lineages (McFadden 2001a). Glaucophyte plastids have already been found to have a higher degree of similarity in pigments, membranes and cell walls to existing cyanobacteria than other primary plastid lineages (McFadden 2001b).

#### 2.1.3.2.2 Primary Endosymbiosis

It is now widely accepted that only three eukaryotic lineages contain primary plastids: green algae (including land plants), red algae and glaucophytes (cyanelles) (Palmer 2003). Two membranes originating from the cyanobacterial ancestor surround primary plastids. The third phagosomal membrane is believed to have been lost at some point during evolution (Cavalier-Smith 2000). Primary plastids are found in the host cell's cytosol and require only transit peptides to import proteins across the plastid membranes. All photosynthetic eukaryotic plastids and cyanobacteria use chlorophyll *a* as their main light-harvesting pigment (Bhattacharya and Medlin 1998, Archibald and Keeling 2002).

Primary plastid lineages also differ in a number of ways. While red algae and glaucophytes contain the phycobilisomes and unstacked thylakoids found in the cyanobacterial ancestor, green algae contain chlorophyll *b* but no phycobilisomes and have evolved thylakoid stacking. Unlike red and green algae, glaucophytes have a peptidoglycan cell wall similar to that found in cyanobacteria (Bhattacharya and Medlin 1998, Cavalier-Smith 2000, Moreira and Philippe 2001, Archibald and Keeling 2002).

These differences have led researchers to question the monophyletic origin of the primary plastid lineages. Most researchers do, however, believe that all primary plastids originated from a single cyanobacterial progenitor (Martin et al. 1998, Cavalier-Smith 2000, McFadden 2001b, Moreira and Philippe 2001, Archibald and Keeling 2002, Simpson and Stern 2002, Palmer 2003). The high degree of similarity in plastid gene content is thought by many to be convincing evidence for primary plastid monophyly (Bhattacharya and Medlin 1998, Nozaki et al. 2002). However, it has been found that the loss of genes during plastid evolution was not always a random event. Certain genes have been retained by the plastid genome in order for it to remain viable. Other genes produce proteins that are difficult to transfer across multiple membranes (Martin et al. 1998, Race et al. 1999, Cavalier-Smith 2000). Parallel independent losses of plastid genes from multiple lineages have also been found to occur, and actually outnumber unique losses by nearly five to one (Martin et al. 1998, Martin et al. 2002). This implies that convergent evolution could also have led to similarities in gene content in primary plastid genomes and, consequently, that plastid gene content may not always be a reliable indicator of evolutionary relationships (Palmer 2003, Stiller et al. 2003).

Similarities in plastid genome organization are believed to indicate a common origin of primary plastids. Two gene clusters (*psbB/N/H* and *atp/rps/rpo*) have been found in all plastid lineages, but not in cyanobacteria. This appears to be evidence for a single process of reduction and rearrangement having occurred in the original plastid genome before the three lineages diverged (Reith and Munholland 1993, Stoebe and Kowallik 1999). However, there is a possibility that these similarities may also be the result of convergent evolution (Palmer 2003, Stiller et al. 2003). Although, since gene order does differ among cyanobacteria, a considerable convergence would need to have taken place independently in each primary plastid lineage to result in the present gene arrangement (Bhattacharya and Medlin 1998). Further studies need to be undertaken to determine the likelihood of such an event (Stiller et al. 2003).

The majority of phylogenetic analyses performed on plastid gene sequences point towards a single primary endosymbiotic event (Douglas and Penny 1999, Turmel et al. 1999, Cavalier-Smith 2000, Palmer 2003). Inaccurate evolutionary relationships have, however, been obtained in the past when only limited sequence data from a single gene was used or when an inadequate number of taxa were sampled (Bhattacharya and Medlin 1995, Stiller and Hall 1997, Moreira et al. 2000, Nozaki et al. 2002). Variations in the evolutionary rate of different

groups of genes may also result in incorrect phylogenetic positioning (Moreira and Phillippe 2001). However, even studies including a number of different gene sequences and taxa suggest a monophyletic origin of primary plastids (Martin et al. 1998, De Las Rivas et al. 2002, Martin et al. 2002). One consistent result of these studies is the grouping of primary plastids as a monophyletic clade relative to cyanobacteria (Turner et al. 1999, Stiller et al. 2003). This does not necessarily provide proof for a single endosymbiotic event. There is the possibility that each primary plastid lineage originated from related cyanobacteria and/or that other related cyanobacterial lineages became extinct (Delwiche 1999, Palmer 2003, Stiller et al. 2003).

Phylogenetic studies using nuclear gene sequences also give conflicting results. The most convincing evidence for monophyletic primary plastid origin was obtained from a phylogenetic study using the nuclear gene encoding for elongation factor 2 (EF2). This gene has a relatively similar evolutionary rate among different lineages and the resulting tree clearly demonstrates strong support for a sister-relationship between green and red algae (Moreira et al. 2000). There are, however, two factors that could have led to incorrect phylogenetic positioning in this study. Firstly, the cryptomonad nucleomorph was included as the red algal sequence. The nucleomorph genome is the remnant nucleus of the secondary endosymbiont and is highly divergent. Secondly, sequences of mitochondrial species' were included that could have unusual gene substitutions. Studies performed on a wide range of mitochondria-containing eukaryotic taxa have provided more reliable results, but still point towards a monophyletic origin of all plastids (Nozaki et al. 2002).

Similarities in plastid targeting mechanisms and import machinery could help resolve evolutionary relationships among primary plastids (Bhattacharya and Medlin 1998, Cavalier-Smith 2000, Douglas et al. 2001, McFadden 2001b, Gardner et al. 2002, Palmer 2003). Many genes that were found in the original cyanobacterial ancestor(s) have been transferred to the host cell nucleus over time. To facilitate transport of these proteins back into the plastid, an N-terminal leader sequence (transit peptide) is attached to the protein. Translocation complexes found embedded in the double plastid membranes recognize the leader sequence and transport the proteins across the membranes and into the plastid (Bruce 2000). Studies have shown that different transit peptides are able to complement protein translocation function in various primary plastid lineages (Jakowitsch et al. 1996, Lang et al. 1998, McFadden 2002b, Stiller 2003). It is believed that this provides proof of a common origin of the plastid import machinery (Cavalier-Smith 2000, McFadden 2001b) and therefore primary



plastids. However, this could also be a function that was conserved during bacterial evolution as a whole (Stiller et al. 2003). Comparisons between plastid and mitochondrial protein translocation apparatus show a high degree of similarity. Many leader peptides are able to target proteins to both organelles at the same time. It has also been found that plant chloroplast transit peptides can target proteins to fungal mitochondria (Peeters and Small 2001). More studies will need to be performed on transit peptides and especially on the protein translocation machinery of primary plastid types to gain a better understanding of their function and relationships (Palmer 2003, Stiller et al. 2003).

Although further clarification of the majority of evidence is needed for primary plastid monophyly, no significant evidence has been found to refute this hypothesis (Palmer 2003). Considering that, for each endosymbiotic event, thousands of genes need to be transferred and successfully integrated into the nuclear genome, deleted from the chloroplast genome and their transcribed proteins transported back into the chloroplast, it seems unlikely that endosymbiotic events occurred more than the minimum number of times (Cavalier-Smith 2000). Further genomic and proteomic research needs to be performed on primary plastids and their host organisms to conclusively prove or disprove the occurrence of a single primary endosymbiotic event.

#### *2.1.3.2.2.1 Green Algae*

The green lineage is a morphologically diverse group of organisms found in a wide range of different habitats. Land plants, believed to have originated some 480 million year ago, form a large part of this group. Their emergence caused substantial alterations to the earth's environment. By reducing the level of atmospheric CO<sub>2</sub>, and therefore lowering the temperature of the earth's surface, they facilitated the colonization of landmasses by other terrestrial organisms (Qiu and Palmer 1999).

All existing green algae can be classified into one of the two sister clades Streptophyta and Chlorophyta. Streptophyta consists of charophytes (mainly fresh water algae) and land plants (embryophytes), while Chlorophyta includes all the remaining green algae (Bhattacharya and Medlin 1998, Qiu and Palmer 1999, Karol et al. 2001).

There are four main classes of chlorophytes: Prasinophyceae, Ulvophyceae, Trebouxiophyceae, and Chlorophyceae. Prasinophytes do not form a monophyletic group with the other chlorophyte classes and a few members might even be included in

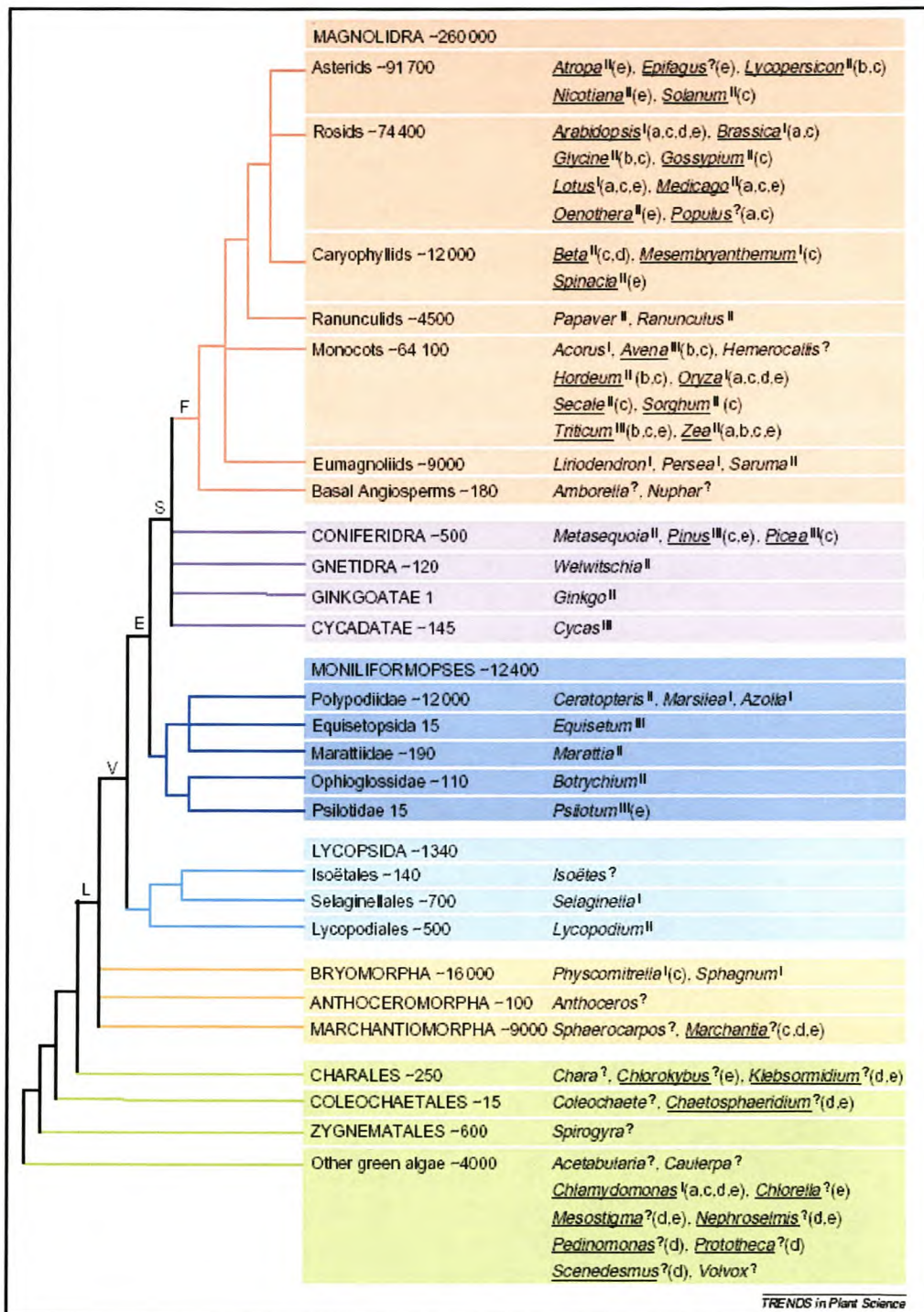
Streptophyta. This group may also represent the earliest branch of the charophytes. The similarities in cpDNA arrangement and the pattern of gene partitioning between the prasinophyte *Nephroselmis olivacea* and land plants (two inverted repeat regions (IRs) separated by a large single copy (LSC) and small single copy (SSC) region) could mean that these characteristics were present in the first green algal ancestor, and most likely also in the cyanobacterial progenitor that gave rise to all chloroplasts. In fact, an IR has already been found in the chloroplast genome of the cyanobacterium *Synechocystis* (Bhattacharya and Medlin 1998, Turmel et al. 1999).

The unicellular, flagellate alga *Mesostigma viride* is a prasinophyte that is very similar to charophytes in cell ultrastructure and small subunit rDNA sequences. The evolutionary positioning of this alga has still not been confirmed. There is evidence to suggest that *M. viride* could be at the base of the split of Chlorophyta and Streptophyta (Lemieux et al. 2000, Turmel et al. 2002a, Turmel et al. 2002b, Grzebyk et al. 2003), but other data place it as a basal lineage of land plants and sister to all green algae (Karol et al. 2001, Delwiche et al. 2002, Martin et al. 2002, Maul et al. 2002). The correct phylogenetic positioning of this alga will provide important information on the early ancestor of land plants (Maul et al. 2002).

Charophytes consist of five orders: Charales, Chlorokybales, Coleochaetales, Klebsormidiales and Zygnematales. The exact relationship between these five orders is not known, but it is believed that, together, they form a paraphyletic group to land plants (Qiu and Palmer 1999, Chapman and Waters 2002). Which of these groups contains the closest living relative to land plants has also not been confirmed. Some phylogenetic studies suggest that Charales is the closest order, with Coleochaetales placed as a sister lineage to the Charales/land plant group (Karol et al. 2001, Delwiche et al. 2002). The completely sequenced chloroplast genome of *Chaetosporidium globosum* (Coleochaetales) shows a high degree of similarity in gene order, gene content and intron composition to that of land plants and phylogenetic analysis also places this species unambiguously at the base of the land plant lineage (Turmel et al. 2002b).

Land plants form a monophyletic group of which bryophytes represent the earliest group (Qiu and Palmer 1999, Kugita et al. 2003). Most molecular and morphological data suggests that bryophytes are a paraphyletic group consisting of three classes: Musci (mosses), Hepaticae (liverworts or hepatics) and Anthocerotae (hornworts) (Nishiyama and Kato 1999). There is still some debate over the evolutionary relationships among the bryophyte classes and

whether hornworts or liverworts are the most basal group (Nishiyama and Kato 1999, Qiu and Palmer 1999, Kugita et al. 2003). Analysis of mitochondrial genes (Qiu et al. 1998), fossil evidence (Edwards et al. 1995) and some sequence data (Qiu and Palmer 1999) point towards



**Figure 2.3.** Phylogenetic relationships among major lineages of green plants. Abbreviations: E = euphyllophytes; F = flowering plants (angiosperms); L = land plants, S = seed plants; V = vascular plants. The genome sizes are indicated where known and were arbitrarily grouped into three classes: I = <1pg of DNA per nucleus; II = 1-10pg; III = >10pg; ? = unknown. See Pryer et al. (2002) for references and web-site addresses where information on the underlined genera can be found.

liverworts being the most basal lineage, but chloroplast and nuclear phylogenetic analysis (Nishiyama and Kato 1999, Qiu and Palmer 1999, Kugita et al. 2003), cladistic analyses based on male gametogenesis (Garbary et al. 1993) and ultrastructural information place hornworts basal to the other bryophyte lineages. An intron of *rrn23*, found in the same location in *Anthoceros* as in the chlorophytes *Chlorella* and *Chlamydomonas*, but not found in any other land plant chloroplast genome or in *Chaetosphaeridium*, supports the latter argument (Kugita et al. 2003). The inclusion of gene data from the recently sequenced *Physcomitrella patens* (moss) chloroplast genome (Sugiura et al. 2003) may provide new insight into the relationship among mosses, hornworts and liverworts.

Pteridophytes and seed plants (angiosperms and four groups of gymnosperms) make up the existing vascular plants. Pteridophytes consist of four lineages: Lycopods, *Equisetum*, Psilotaceae and ferns. This group is paraphyletic to seed plants and played an important role in establishing the early land plants. Very little is presently known about the exact phylogenetic relationships and evolutionary history of this group (Qiu and Palmer 1999, Chaw et al. 2000).

The four gymnosperm seed plant groups are cycads, conifers, *Ginkgo*, and Gnetales. Gymnosperms are believed to be a monophyletic group and sister to angiosperms and most phylogenetic evidence places cycads as the basal seed plant lineage (Bowe et al. 2000, Chaw et al. 2000). Bryophytes, pteridophytes and gymnosperms have enjoyed dominance on land until approximately 90 million years ago when angiosperms replaced them (Pryer et al. 2001). The large number of angiosperm species found on Earth today makes their classification a difficult undertaking. The direct ancestor of angiosperms is also a mystery that still needs to be solved (Bowe et al. 2000). Some phylogenetic studies have placed *Amborella trichopoda* as the most basal angiosperm lineage (Graham and Olmstead 2000, Borsch et al. 2003) with Nymphaeales (water lilies) sometimes placed as the subsequently evolving branch (Soltis et al. 1999, Kuzoff and Gasser 2000). Most extant angiosperm species are grouped as monocots or eudicots (Graham and Olmstead 2000), but an influx of new molecular and morphological data have made it necessary to re-evaluate previous flowering plant classification systems. A consortium called the Angiosperm Phylogeny Group (AGP) has recently been established to enable this re-evaluation process (Angiosperm Phylogeny Group 1998). According to their most recent report (APG II 2003) there are 453 flowering plant families including 27 basal angiosperms, 81 monocots and 345 eudicots (Hanson et al. 2003).

By forming a consortium of plant research specialists to work on a project such as the APG, information may be obtained in a coordinated manner. A similar collaborative effort, known as “The Green Plant Phylogeny Research Coordination Group” (GPPRCG - <http://ucjeps.berkeley.edu/bryolab/GPphylo/>) is enabling the clarification of phylogenetic relationships among all green plants. Studies performed by such groups will combine information obtained by analysis of multigene sequence data, genomic structural characters, morphological traits and fossil data to significantly improve our understanding of the origin and evolution of land plants (Qiu and Lee 2000).

#### 2.1.3.2.2.2 Red Algae

Rhodophytes are found mainly in coastal marine waters, with only a few living in freshwater environments. No red algal lineages are found in terrestrial ecosystems. Rodophyta contains relatively few existing taxa, including *Porphyra purpurea* and *Cyanidium caldarium*. Most eukaryotic algal lineages contain red algal plastids obtained through secondary endosymbiosis (Delwiche 1999, Grzebyk and Schofield 2003).

Red algae contain phycobiliproteins in their chloroplasts but they lack flagellae. These features are also found in cyanobacteria and indicate a close relationship between red algae and cyanobacteria (Nozaki et al. 2002). Most phylogenetic studies indicate that red algae diverged after glaucophytes but before the green algal lineage (Martin et al. 1998, De Las Rivas et al. 2002).

#### 2.1.3.2.2.3 Glaucophytes

As mentioned previously, glaucophyte plastids (cyanelles) have very similar pigments, morphology and cell walls to existing cyanobacteria (McFadden 2001b). These similarities, together with the results of molecular phylogenetic studies, place the glaucophytes as the earliest diverging primary plastid lineage (Martin et al. 1998, De Las Rivas et al. 2002, Grzebyk et al. 2003, Iino and Hashimoto 2003). In the past, *C. paradoxa* has even been mistaken for a type of cyanobacterium (Delwiche 1999). Further investigation of this intriguing lineage could provide further insights into the nature of the cyanobacterial progenitor and early primary plastids.

#### 2.1.3.2.3 Secondary Endosymbionts

The three lineages resulting from primary endosymbiosis account for a small part of the diversity of plastid-containing eukaryotes found today (Moreira and Philippe 2001).

Secondary endosymbiosis involving red and green algae has given rise to a number of major eukaryotic lineages including more than 42 600 species (McFadden 1999). Plastids that developed as a result of secondary endosymbiosis are found in two Kingdoms: Chromista and Protoctista. Cryptomonads and chromobiontes (brown algae, heterokonts, haptophytes) are members of the Kingdom Chromista, while euglenoids, chlorarachniophytes, dinoflagellates and apicomplexans (sporozoa) are members of the kingdom Protoctista (Cavalier-Smith 2000, Yoon et al. 2002).

All secondary plastids are surrounded by three or more membranes and are otherwise known as complex plastids. These extra membranes provide strong evidence for secondary endosymbiosis, because membranes are seldom synthesized *de novo* by the cell and are usually acquired elsewhere. Plastids with four membranes include those of chromobiontes (heterokonts, haptophytes), apicomplexa, cryptomonads and chlorarachniophytes. The outer membrane is the phagosomal membrane derived from the secondary host vacuole, while the periplastid membrane (between the phagosomal membrane and the plastid double membrane) is probably derived from the algal plasma membrane. Three membranes enclose euglenoid and dinoflagellate plastids, although it is still not known why. The most logical explanation is that one membrane has been lost; probably the periplastid membrane (Bhattacharya and Medlin 1998, Cavalier-Smith 2000, Archibald and Keeling 2002).

As a result of the extra membrane(s) surrounding complex plastids, changes had to be made to the protein transport mechanisms used by primary plastid containing organisms. One segment of the bipartite transit peptide is a secretory signal that transports the protein through the outermost membrane of the complex plastid and the other part is a normal transit peptide that transfers the protein across the two inner membranes obtained from the primary plastid (McFadden 1999, Palmer 2003).

There is a large amount of evidence to support the secondary endosymbiotic theory, but there is still a great deal of debate over the number of secondary endosymbiotic events that occurred during evolution. There is, however, a general consensus that at least two events occurred: one giving rise to red algal complex plastids and the other giving rise to green algal complex plastids (Cavalier-Smith 2000).

Euglenoids and chlorarachniophytes are two groups of protist algae that contain green secondary plastids. They have very different ultrastructures, leading some researchers to

believe that they have acquired their green algae by means of two separate secondary endosymbiotic events (Bhattacharya and Medlin 1998, Archibald and Keeling 2002). Other researchers believe that they originated from a single secondary endosymbiotic event (Cavalier-Smith 2000, Archibald and Keeling 2002). Phylogenetic comparisons performed on their nuclear and plastid DNA show no definite proof of a common or separate origin of these two lineages. Further analysis of the nuclear genomes and protein import mechanisms found in these two lineages may help resolve the issue (Palmer 2003).

Cryptomonads, chromobiontes, dinoflagellates, and apicomplexans all contain red algal complex plastids. These lineages have a number of similar features, such as the presence of chlorophyll *c*, but they are found in a wide range of different ecological regions and vary to a large degree in morphology (Harper and Keeling 2003). There are a few researchers who believe that a number of independent events gave rise to each of these groups (Bhattacharya and Medlin 1998). However, it has been argued that the occurrence of independent endosymbiotic events more times than required would be unparsimonious, especially when taking into account the number of genes that are transferred from the plastid to host nuclear genome over evolutionary time as well as the protein-targeting machinery needed to transfer proteins back to the plastid (Cavalier-Smith 2000, Archibald and Keeling 2002). The chromalveolate hypothesis proposed by Cavalier-Smith (2000) suggests that a single secondary endosymbiotic event gave rise to all red algal complex plastids.

The original evidence for the chromalveolate hypothesis was provided by physical characterisation of red algal complex plastid lineages. This includes the presence of chlorophyll *c*<sub>2</sub> and the fact that all chromist complex plastids are found in the lumen of the rough endoplasmic reticulum (RER) (Yoon et al. 2002). The strongest genetic evidence supporting the common origin of chromalveolate plastids is found in studies performed on glyceraldehydes-3-phosphate dehydrogenase (GAPDH) (Fast et al. 2001). Two different types of nuclear-encoded GAPDH genes are found in photosynthetic eukaryotes: one is cytosolic and the other is targeted to the plastid. In the primary plastid lineages, *Euglena* and *Pyrocystis*, the plastid-targeted GAPDH is homologous to the cyanobacterial GAPDH. In the complex plastids of cryptomonads, haptophytes, heterokonts, dinoflagellates and apicomplexans the plastid-targeted GAPDH shows homology to eukaryotic cytosolic GAPDH. It is thought that the latter is a result of the replacement of the original plastid-targeted GAPDH with a copy of a duplicated cytosolic GAPDH. Phylogenetic analysis of plastid-targeted GAPDH sequences of these secondary endosymbiotic groups also indicate a

close relationship between them, but no significant relationship to the cytosolic GAPDH found in these groups (Archibald and Keeling 2002, Harper and Keeling 2003). This data as well as results of another study performed on cryptophytes, heterokonts and haptophytes provide convincing support for the chromalveolate hypothesis (Yoon et al. 2002).

Complex plastids have a variety of unique features. One example of this is the discovery of the remnants of the secondary endosymbiont nuclear genome found in the periplastidial compartment of cryptophytes and chlorarchniophytes. This small double-membraned, eukaryotic nucleus is called a nucleomorph and contains a highly reduced genome (551 kb in cryptophytes and 380 kb in chlorarchniophytes). The nucleomorph genome of a cryptomonad, *Guillardia theta*, has been completely sequenced and, of the 511 genes identified, only 30 code for proteins that are targeted to the plastid. As found in all other secondary plastid-containing algae without nucleomorphs, most genes coding for plastid-targeted proteins have been transferred to the host nuclear genome. It is not known why cryptophytes and chlorarchniophyte algae still contain these nucleomorphs. It may be that they are in the process of reduction or that, for some unknown reason, this process has stopped. Analysis of the sequences of this nucleomorph DNA has already provided further evidence for the origin of these plastids and may help us to gain a better understanding in the future of how genome reduction occurs (Gilson and McFadden 1996, Bhattacharya and Medlin 1998, Cavalier-Smith 2000, Douglas et al. 2001, Archibald and Keeling 2002).

The eukaryotic group Alveolata includes some of the most diverging plastid lineages found to date. The three lineages that make up this group are dinoflagellates, apicomplexans and ciliates. Dinoflagellates contain plastid genomes that are made up of a number of small minicircles, each containing only a single gene (Zhang et al. 1999). Only half of dinoflagellates actually contain plastids that are capable of photosynthesis. Of these, many do not contain the original peridinin-containing red algal complex plastid. Some have replaced it with another complex plastid from a green alga and others with a complex plastid from cryptomonads or chromophytes. This type of plastid replacement is known as tertiary endosymbiosis. In general, dinoflagellate plastids appear to be more prone to plastid loss and replacement than those of other lineages, but the reason for this is not yet known (Delwiche 1999, Archibald and Keeling 2002, Cavalier-Smith 2000).

Apicomplexans (sporozoa) are all obligate intracellular parasites that are non-photosynthetic and therefore contain no pigments (e.g. *Plasmodium*, responsible for causing malaria). Some



apicomplexan species do, however, contain plastids with highly reduced genomes (Köhler et al. 1997). These apicoplasts have been identified in a number of phyla including *Plasmodium*, *Toxoplasma*, *Eimeria*, *Babesia*, *Theileria*, *Sarcocystis* and *Hepatozoon* (McFadden and Roos 1999). Whether the original secondary endosymbiont was of red or green algal origin is still unclear. A few apicomplexans appear to have plastids of green algal origin (Köhler et al. 1997, Bhattacharya and Medlin 1998, Cavalier-Smith 2000, Archibald and Keeling 2002) while others contain red algal complex plastids (Blanchard and Hicks 1999, Stoebe and Kowallik 1999, Cavalier-Smith 2000). The most likely explanation for this is that the apicomplexan ancestor initially engulfed a red alga that was later replaced by a green alga in some members (Palmer 2003). Ciliates also form a non-photosynthetic alveolate group. Members of this group are known to acquire plastids (“kleptochloroplasts”) temporarily from their prey (Delwiche 1999). However, very little research has been done on the plastids found in this group.

#### **2.1.4 The Chloroplast Genome**

In the early 1900s studies carried out on variegation in higher plants revealed that some variegated leaf patterns were caused by factors inherited in a non-mendelian manner. After further analysis of these plants it was suspected that the chloroplast was involved. This implied that the chloroplast had its own genetic system, separate from that of the cell nucleus (Shinozaki et al. 1986, Sugiura 1992). This discovery led to a greater interest in, and therefore more intensive study of, chloroplasts and their DNA.

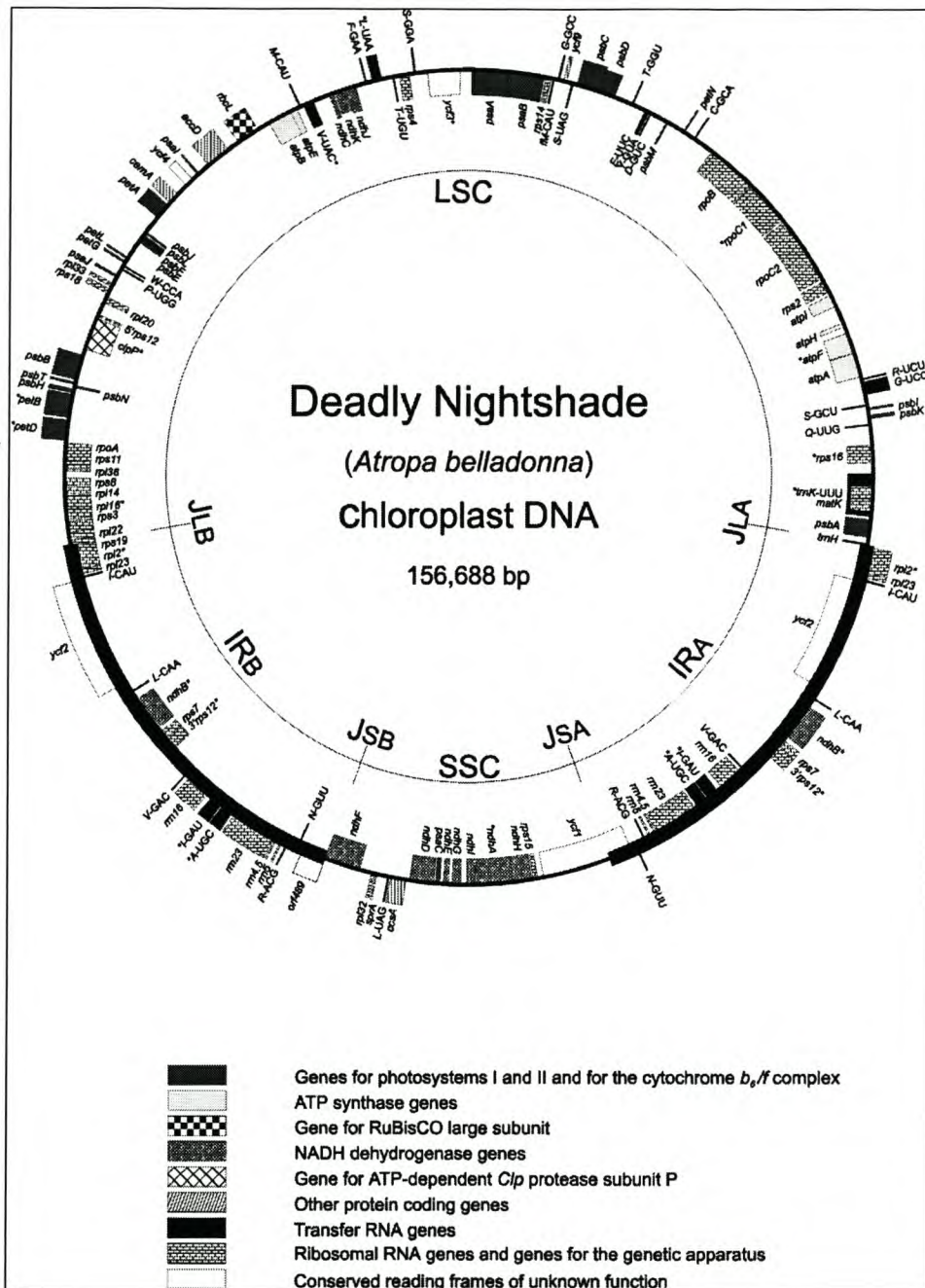
The use of CsCl gradient centrifugation demonstrated that the chloroplast contains its own characteristic circular DNA with a specific buoyant density (Chun et al. 1963, Sager and Ishida 1963, Manning et al. 1971). In fact, the chloroplast genomes of all algae and higher plants studied thus far have been found to be circular DNA molecules, separate from the nucleus. With the development of methods such as gene cloning and DNA sequencing chloroplast genome research has advanced at an increasing rate (Sugiura 1992). In 1976 the first physical map of cpDNA (maize) was constructed (Bedbrook and Bogorad 1976). The first chloroplast gene was cloned in 1977 (Bedbrook et al. 1977), also from maize. Nine years later the first chloroplast genomes were completely sequenced, namely those of tobacco (Shinozaki et al. 1986) and the liverwort (Ohshima et al. 1986). These were also the first plant genomes to be completely sequenced. The chloroplast genomes of 32 algae and plant species ([http://megasun.bch.umontreal.ca/ogmp/projects/other/cp\\_list.html](http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list.html)) have now been

completely sequenced. Examples include the chloroplast genomes of *Oryza sativa* (rice) (Hiratsuka et al. 1989), the *Lotus japonicus* (legume) (Kato et al. 2000), *Zea mays* (maize) (Maier et al. 1995), *Arabidopsis thaliana* (Arabidopsis) (Sato et al. 1999), *Nicotiana tabacum* (tobacco) (Shinozaki et al. 1986) and black pine (*Pinus thunbergii*) (Wakasugi et al. 1994).

#### 2.1.4.1 Size and Structure

Most higher plant chloroplast genomes are between 120 and 160 kb in size (Hager and Bock 2000). The genomes of monocotyledonous plants are generally smaller than those of dicotyledonous plants (Ogihara et al. 1988). The cpDNA molecule of higher plants is divided structurally into four regions, namely: a large single copy (LSC - about 80 to 100 kb) and small single copy region (SSC - about 15 to 25 kb), separated from each other by two copies of an inverted repeat region (IR - 10-85 kb) (Palmer et al. 1985, Harris et al. 1994, Cosner et al. 1997) (figure 2.4). The two IRs are identical to each other in genetic sequence as a result of an active copy correction system (Harris et al. 1994). This region is also relatively well conserved in higher plants, especially in angiosperms (De Las Rivas et al. 2002). However, most variation in the size of the chloroplast genome in different plant species is a result of changes in the size of the IR (Palmer 1987, Sugiura 1992, Harris et al. 1994).

Chloroplast genomes can be divided into three groups according to their inverted repeats: those missing IRs (group I), those containing IRs (group II) and those with tandem repeats (group III). The pea, broad bean and alfalfa are examples of chloroplast genomes lacking an IR (Sugiura 1992). The black pine chloroplast genome has a short IR that has a tRNA gene and part of the 3' end of the *psbA* gene, but lacks the rRNA genes (Wakasugi et al. 1994). Six tribes of the Fabaceae (legume) family have lost one segment of the IR and therefore only have one copy of each of the rRNA genes (Palmer and Thompson 1981). The chloroplast genome of *Euglena gracilis* has no IRs, but instead contains three tandem repeats, each of these having an rRNA gene cluster (Sugiura 1992). Intramolecular recombination often takes place at a high frequency between the two parts of the IR, causing inversion isomers to form in species containing these repeats. This allows chloroplast molecules to be divided into two physically dissimilar, but genetically identical, groups according to the relative orientation of the single copy region. A small proportion of the molecules also exist in dimer, trimer and tetramer forms (Palmer 1987, Sugiura 1992).



**Figure 2.4.** Gene organization of the *Atropa belladonna* chloroplast genome. The identical inverted repeat regions (IR<sub>A</sub> and IR<sub>B</sub>) separate the large single copy region (LSC) from the small single copy region (SSC). Genes transcribed clockwise are drawn inside the circle, and those outside are transcribed anticlockwise. The different gray-scales or patterns denote different functional groups. See Stoebe, Martin and Kowallik (1998) for abbreviations and nomenclature of protein-coding genes (Schmitz-Linneweber et al. 2002)

Studies performed on chloroplast genomes have revealed that the chloroplast genomes of higher plants have an overall genetic structure that is highly conserved. This is mainly as a result of its uniparental inheritance, which excludes any recombination, and the stabilizing action of long inverted repeats, resulting in a lower mutation rate than found in other cell genomes (Wolfe et al. 1987, Maier et al. 1995, Cipriani et al. 1998, Ogihara et al. 2000). However, the sequencing of algal plastid genomes has revealed a much larger degree of

variation in chloroplast gene organization and genome size among different algae. This variation is most evident in green algae. The siphonous green alga *Codium fragile* has a plastid genome of 85 kb in size, while on the other end of the scale the giant green alga *Acetabularia* has a very complex plastid genome that is 1500 kb in size (Sugiura 1992, Harris et al. 1994, Simpson and Stern 2002).

#### **2.1.4.2 Gene Arrangement, Function and Expression**

The majority of genes present in the original cyanobacterial endosymbiont have been lost or transferred to the host nucleus. Only 60 to 200 genes still remain in the chloroplast genome of higher plants and green algae (Martin et al. 2002). The chloroplast genome codes for all of its own tRNA and rRNA molecules and some the proteins that it requires (Hiratsuka et al. 1989). Two main groups of protein-encoding genes are found in chloroplast genomes. The first group is involved in expression and translation of chloroplast genes and the second group plays a role in bioenergetics and photosynthesis (Leister 2003). A number of hypothetical open reading frames (*ycfs*) have been found of which the functions are not yet known (Watson and Murphy 1999, Leister 2003). The remainder of the proteins required by the chloroplast (about 90 %) are encoded in the nucleus and transported through the cytoplasm to the chloroplast (Hiratsuka et al. 1989, Cooper 2000). Plastids are therefore only semi-autonomous (Sato 2001).

The average chloroplast genome contains the genes for the transcription of four rRNA (23S, 16S, 5S and 4.5S) and 30 tRNA species. This means that, in contrast to the mitochondrial genome, enough tRNA types are encoded by the chloroplast genome to allow the translation of all mRNA codons according to the universal genetic code (Cooper 2000). The chloroplast genome also codes for approximately 20 ribosomal proteins and a few RNA polymerase subunits. About 30 proteins encoded by cpDNA are involved in photosynthesis. These form part of photosystem I and II, the cytochrome *bf* complex, and ATP synthase. The large subunit of Rubisco (Ribulose-1,5-bisphosphate Carboxylase/Oxygenase) is also encoded in the chloroplast (Schopfer 1995). This enzyme plays a critical role in photosynthesis and is believed to be the 'single most abundant protein on earth' (Cooper 2000). However, the nuclear genome codes for at least one of the subunits of each of the above mentioned complexes, giving the nucleus strict control over the expression of chloroplast genes (Watson and Murphy 1999). Light and other external factors also play a role in chloroplast gene

regulation, but these methods of control are still under investigation (Schopfer 1995, Gray et al. 2003, Pfannschmidt 2003)

Chloroplast gene organization and expression has retained many of its prokaryotic features. Many genes found in plastids are organized into operons and produce polycistronic mRNAs by co-transcription. The gene order in several chloroplast operons is highly conserved and very similar to that of Cyanobacteria and other Eubacteria. For example, the ribosomal RNA operon is identical in chloroplast and eubacterial genomes. Another prokaryotic feature of cpDNA is the presence of a number of overlapping genes that often code for proteins that form part of the same complex (Palmer 1987, Stoebe and Kowallik 1999, Watson and Murphy 1999, Hager and Bock 2000).

A number of eukaryotic features can also be observed in the chloroplast genome. Certain proteins, e.g. those used in photosystem II, are encoded by genes found throughout the genome and are transcribed as monocistronic mRNAs. A distinctly eukaryotic feature is the presence of introns, suggesting that an RNA-splicing mechanism is also present (Palmer 1987, Watson and Murphy 1999).

The gene regulation mechanisms used by chloroplasts have retained many of their prokaryotic characteristics, but have made evolutionary adaptations to facilitate their integration into the eukaryotic cell. The chloroplast has moved from the typically prokaryotic transcriptional control mechanism to a primarily translational system of control. The latter method allows the plastid to respond more rapidly to environmental changes, since the completed transcripts are readily available for protein synthesis (Kunnimalaiyaan and Nielsen 1997, Hager and Bock 2000).

#### **2.1.4.3 Protein Import**

Recent research on the *Arabidopsis thaliana* chloroplast genome suggested that between 2100 and 3600 proteins could be present in the chloroplast (Abdallah et al. 2000, The Arabidopsis Genome Initiative 2000). This means that approximately 90% of proteins found in this organelle are encoded by the nuclear genome, transported into the chloroplast and distributed internally.

The transport of proteins from the cytosol into the chloroplast is facilitated by chloroplast transit peptides (cTPs). This N-terminal sequence of 30 to 100 amino acids is attached to the end of the completed polypeptide in the cytosol. The translocation complex found in the outer chloroplast membrane (the Toc complex) recognizes the cTP and assists in the transport of the protein across this membrane. The protein is then transported in the same fashion through a second translocation complex (the Tic complex) found in the inner membrane. Molecular chaperones found in the cytosol and stroma also facilitate protein import, using ATP in the process (Bruce 2000, Cooper 2000).

Once the protein is in the stroma the cTP is cleaved by a stromal processing peptidase. From here the protein is transported to its destination in the chloroplast. Higher plants possess four different pathways that the protein may follow to travel into or through the thylakoid membrane. Three of these pathways rely on a second hydrophobic signal sequence found on the C-terminal end of the cTP. Once the protein is in the thylakoid lumen this sequence is cleaved by the thylakoidal processing peptidase. No extra cTPs are required to transport a protein from the stroma to the inner membrane. Proteins targeted to the outer membrane seem to be directly inserted. However, very little is known about the targeting sequences or pathways used to transport proteins to the intermembrane space (Cooper 2000, Leister 2003).

#### **2.1.4.4 Inheritance**

Chloroplast genes are inherited differently to nuclear genes and do not obey a number of Mendel's laws. While nuclear genes only undergo replication during mitosis (and therefore not during asexual reproduction), chloroplast alleles undergo vegetative segregation as well. This means that chloroplasts and their DNA segregate during mitosis as well as meiosis and that the replication of chloroplast genomes occurs randomly (Birky 1995, Birky 2001). In this way any heteroplasmy (where more than one type of chloroplast occurs in an individual) may be sorted out in one or a few generations (Frey 1998).

Chloroplast genomes are usually inherited from one parent, i.e. uniparental inheritance, while inheritance of nuclear genes is biparental. Some plants have been found to produce progeny of maternal, paternal, and biparental origin. However, even where both parents donate chloroplasts (in some flowering plants) no fusion or recombination of the respective genomes takes place. Their inheritance is therefore still asexual. Both uniparental inheritance and lack

of recombination could play an important role in the conservation of higher plant chloroplast genomes (Palmer 1987, Birky 1995, Birky 2001).

Intramolecular and intermolecular recombination occurs in the chloroplast genome. These kinds of recombination help to preserve the sequence homology between the two IR regions. They can, however, also lead to changes in the single-copy regions' gene order and produce dimeric genomes (Birky 2001).

#### **2.1.4.5 Genetic Mutations**

Chloroplast genomes share a large degree of similarity in size, structure and nucleotide sequence, especially in the IR regions (Palmer 1987, Wolfe et al. 1987). Chloroplast genes seldom undergo rearrangement and are rarely lost all together (Cosner 1997). Non-coding regions, on the other hand, have a high evolutionary rate (Fujii et al. 1997) and the amount of non-coding DNA appears to have actually increased with an increase in biological complexity (De Las Rivas et al. 2002).

Most changes in sequence complexity are as a result of length mutations. This involves the addition or deletion of fragments, usually 1 bp to a few dozen bp in length, found in non-coding regions. Because these changes often occur near regions of short direct repeats, they are thought to be a result of slipped-strand mispairing (Palmer 1987, Kelchner and Wendel 1996, Ogiwara et al. 2000), intramolecular recombination between flanking or nearby repeats (Palmer et al. 1985, Ogiwara et al. 1988), and stem-loop secondary structures (vanHam et al. 1994, Ferris et al. 1995). Certain "hot spots" have also been discovered, where length mutations occur more often than in the rest of the genome (Ogiwara et al. 1988).

Other small structural changes occurring in higher plants are inversions (Howe 1985, Hiratsuka et al. 1989) and translocations (Ogiwara et al. 1988). Inversions are often larger than deletions or insertions (1-62 kb) (Kelchner and Wendel 1996). One possible way in which these inversions could occur is when IR regions undergo intramolecular recombination (Ogiwara et al. 1988, Kelchner and Wendel 1996). For example, the chloroplast genome of *Trachelium caeruleum* has a high level of rearrangement in comparison to other higher plant chloroplast genomes. This divergence from the ancestral gene arrangement would require contraction/expansion of the IR, at least seven inversions, two large deletions, several insertions, and one or two possible transposition (Cosner 1997).

Any other length mutations found in the chloroplast genome occur at a much lower frequency than those already mentioned. Certain mutations of between 50 and 1200 bp have been observed in restriction-fragment comparisons and are most likely a result of nucleotide substitutions. These substitutions occur mainly as silent changes in the third codon position and they are usually accumulated at a very slow rate in chloroplast genomes (Palmer 1987).

### **2.1.5 The *Vitis vinifera* Chloroplast Genome**

Very little information is currently available on the *Vitis* chloroplast genome. Most studies done up to this point use chloroplast-specific microsatellite markers to screen different grapevine cultivars. The results obtained give researchers an idea of the level of polymorphism found among different cultivars, and allow them to divide grapevine varieties into haplotypes. This information, in turn, aids in determining the distribution and evolution of different grapevine cultivars (Garcia et al. 2002b, [www.biology.uoc.gr/gvd/contents/general-info/06b.htm](http://www.biology.uoc.gr/gvd/contents/general-info/06b.htm)). These microsatellite polymorphism studies have also been used to prove the maternal inheritance of *V. vinifera* chloroplast genomes (Garcia et al. 2002a).

### **2.1.6 Applications of Sequenced Chloroplast Genomes**

#### **2.1.6.1 Phylogenetic Studies**

Plastid genomes have an important application in research into eukaryotic evolutionary relationships. They are ideal for these studies because of their small genome size and their relatively well-conserved nature. Chloroplast genomes have a relatively low nucleotide substitution rate found between individuals of the same species or between related species (Matsuoka et al. 2002).

Chloroplast genome restriction-site mutations have been used by many researchers to study the evolution of populations or closely related species (Palmer et al. 1985, Gounaris et al. 1986, Huang and Sun 2000, Katayama and Uematsu 2003). The restriction fragment patterns obtained allow researchers to discern variation at an interspecies and, to a certain extent, intraspecies level (Palmer 1987). With the development of improved cloning and sequencing methods it has become possible to sequence entire chloroplast genomes with relative ease. The chloroplast genomes of a wide variety of higher plant and algae species have already



been completely sequenced. Phylogenetic studies using chloroplast sequences are presently being used as an alternative to restriction site analysis because they allow evolutionary relationships to be determined on a finer scale. Specific genes or gene regions can now be compared among a wide variety of taxa and lineages. However, variations in the evolutionary rates of different genes or gene regions do cause incorrect phylogenetic relationships to arise at times. Therefore, the analysis of restriction sites still provides researchers with a good overall idea of variation occurring on the genome level (O'Kane 1995).

Comparison of the nuclear-, chloroplast- and the related protein coding sequences of different species will give researchers clearer insight, first, into the functioning of chloroplasts in the cell (Sato et al. 1999) and second, into mechanisms and reasons for chloroplast gene migration to the nucleus (Stoebe and Kowallik 1999). Phylogenetic studies using a number of plastid, mitochondrial and nuclear chloroplast gene sequences from a variety of taxa will provide a robust picture of evolutionary relationships among plastid-containing organisms. This data, together with studies of plastid morphology, ultrastructure and genome arrangement could help clear up many of the present discrepancies in evolutionary relationships.

#### **2.1.6.2 Plastid Transformation**

The genetic engineering of plant nuclear genomes to express foreign proteins is now a routine procedure that has had many successes (Bogorad 2000). However, nuclear transformation does have a number of limitations. The development of technologies to integrate foreign DNA into plant plastid genomes has recently developed as a promising alternative to nuclear transformation (McFadden 2001a).

The most successful method of delivering foreign genes into the plastid involves biolistic bombardment of the appropriate plant tissue with the cloned DNA (Hager and Bock 2000). Other delivery systems that have been used with limited success are polyethylene glycol (PEG) treatment and microinjection. Once inside the plastid the foreign DNA is integrated into the organellar genome via homologous recombination. This is made possible by the presence of transgene flanking regions with gene sequences homologous to the plastid genome. Only one or a few genome copies found in a plastid will be successfully transformed and, therefore, wild-type genomes will still be present. Selection of transformed cells is carried out through repeated cell division on a selective medium, with the help of plastid-

specific selectable markers (e.g. antibiotic resistance genes). The homoplasmic cells are regenerated to form stable transplastomic plants (Hager and Bock 2000, Heifetz 2000, Maliga 2002, Reece 2004).

Plastid transformation offers a number of advantages over nuclear transformation. (i) Plastids are maternally inherited in most plants. Transgenes are thus efficiently contained and foreign DNA is not usually spread to related wild-type relatives via pollen, as is the case with nuclear transformants (Hager and Bock 2000, Daniell et al. 2002). (ii) The high ploidy level of plastid genomes per cell results in high levels of foreign gene expression and foreign protein accumulation in plastids (Daniell and McFadden 1987, Hager and Bock 2000, McFadden 2001a). (iii) Foreign genes can be targeted to a specific site in the plastid genome, giving more predictable levels of transgene expression and eliminating the position effect often observed in nuclear transformants (Hager and Bock 2000, Daniell et al. 2002). (iv) Epigenetic gene silencing is an effect often observed when foreign genes are incorporated into plant nuclear genomes. These effects have not been observed in transplastomic plants (Bock and Hagemann 2000, Daniell et al. 2002). (v) Many plastid chloroplast genes are organized into operons and therefore allow the simultaneous co-transcription of several genes into polycistronic mRNAs. There is the possibility that a number of foreign genes may be incorporated into the plastid genome at one time, using a single transformation vector and driven by one promoter. In this way a trait that is transcribed by multiple genes or even an entire biochemical pathway can be expressed in plants. The nuclear genome, on the other hand, is only capable of monocistronic transcription. The expression of multiple foreign genes in nuclear genomes is therefore difficult to achieve (Bogorad 2000, Hager and Bock 2000, Daniell et al. 2002). (vi) Most selectable markers used at present are antibiotic resistance genes. There is some concern that these genes could be transferred to other organisms in the environment, offering them the same resistance. One alternative is to use the plant-derived betaine aldehyde dehydrogenase (BADH) gene as a selectable marker. Selection involves the conversion of toxic betaine aldehyde (BA) to non-toxic glycine betaine by the chloroplast-specific BADH enzyme. This method has proven successful, with 25-fold higher plastid transformation efficiency than obtained through spectinomycin selection (Daniell et al. 2001, Daniell and Dhingra 2002). Another option is to remove the selectable marker gene after transformation (Iamtham and Day 2000, Corneille et al. 2001, Reece 2004).

There are a few concerns that still need to be addressed before plastid transformation can be successfully and routinely used in the future. The first drawback of plastid transformation is

the limitations on current methods of foreign DNA delivery. Particle bombardment remains the most effective manner in which to incorporate foreign DNA into the chloroplast genome. Transient foreign gene expression and difficulties in transferring the cloned DNA across the double plastid membrane have been encountered with other methods. Second, successful plastid transformation systems have been successfully produced for only a few higher plants. Most major crop plants, especially monocotyledons, do not have a working plastid transformation system. Third, the plastid genomes of a large number of commercial crops have not yet been fully sequenced. These sequences are needed to design foreign DNA flanking regions to facilitate successful integration into the targeted plastid genome. However, the conserved nature of the chloroplast genome in higher plants could allow targeting sequences from one species to be used to target genes to chloroplast genomes of a related species with an unknown plastid sequence. Lastly, the process of obtaining genetically stable (homoplasmic) transplastomic plants is difficult and time-consuming. Therefore, only a limited number of transformants can be produced at a time, making large-scale transplastomic plant production a problem (Sidorov et al. 1999, Hager and Bock 2000, Heifetz 2000, Daniell et al. 2002, Maliga 2002).

The first chloroplast genome to be stably transformed was that of *Chlamydomonas reinhardtii* in 1988 (Boynton et al. 1988). Two years later a successful method for tobacco plastid transformation was developed (Svab et al. 1990). Most subsequent plastid transformations have been performed in tobacco (Gewolb 2002, Maliga 2002). Plastid transformation technology has, however, not been successfully extended to many other plants. The development of methods to transform crop plants such as maize will be of particular importance for the success of plastid transgenics in the future.

Plastid transformation technology has a number of present and possible future applications in scientific research, agronomy and biotechnology (Maliga 2002). Reverse genetics involves the targeted deletion or mutation of a gene or its product in vivo with the help of plastid transformation technology. This method is proving useful for the characterization of plastid genes and for ascertaining the functions of hypothetical open reading frames (*ycfs*) (Rochaix 1997, Hager and Bock 2000, Heifetz 2000). Attempts to optimize metabolic functions and photosynthesis by means of plastid transformation are already underway. Experiments performed to improve the efficiency of Rubisco (Ribulose-1,5-bisphosphate Carboxylase/Oxygenase) have already been successful (Whitney and Andrews 2001, Maliga 2002). A number of traits with potential agronomic value have been stably transformed into

plastid genomes. These include antibiotic resistance (Svab and Maliga 1993), herbicide resistance (Daniell et al. 1998), insect resistance (Kota et al. 1999) and bacterial and fungal disease resistance (DeGray et al. 2001). The biodegradable protein-based polymer GVGVP became the first pharmaceutical protein to be stably expressed in tobacco chloroplasts (Guda et al. 2000). Since then, a number of human therapeutic proteins have been successfully expressed in plastids. For example: human somatropin (hST) (Staub et al. 2000), cholera toxin fragments (Daniell et al. 2001), tetanus toxin fragments (Tregoning 2003), and “thermostatable xylanase” (Leelavathi et al. 2003). These experiments demonstrate that plastid-expressed proteins are biochemically active, correctly folded, and produce the appropriate disulfide bonds (Maliga 2003, Reece 2004). The production of human antibiotic and antigen proteins at high levels in the edible tissue of plants could lead to the creation of edible antibiotics and vaccines (Daniell et al. 2002). Transformation experiments performed on tomato (Ruf et al. 2001) and potato plastids (Sidorov et al. 1999) have already shown promising results. However, expression levels are much lower in non-green tissue than in chloroplasts. The creation of more advanced expression tools could enhance expression levels in future experiments (Maliga 2002).

Although foreign proteins have been successfully expressed and inserted into the chloroplast genome, it is not known if these proteins can be exported from the chloroplast into the cytosol or into other cellular compartments. Mechanisms of protein export from the chloroplast have been studied (Weber et al. 2000, Weber and Flügge 2002) but further research still needs to be performed.

### **2.1.6.3 Drug Targets**

A number of apicomplexan parasites contain a reduced plastid genome of between 27 and 35 kb in size. The apicoplast is essential for the survival of these parasites, and is known to be involved in fatty acid and isoprenoid metabolism. These metabolic pathways are cyanobacterial in nature and therefore differ from the equivalent, cytosolic pathways found in animal hosts. The apicoplast therefore has great potential as a drug target aimed at destroying these parasites (McFadden and Roos 1999, Gardner et al. 2002).

## 2.2 CHLOROPLAST GENOME CHARACTERISATION AND SEQUENCING

Since the construction of the first physical map of cpDNA in 1976 (Bedbrook and Bogorad 1976) and the cloning of the first chloroplast genome a year later (Bedbrook et al. 1977) the field of chloroplast genetics has progressed at an ever increasing rate.

A number of methods for cpDNA isolation have been developed and optimized for different plant species. Physical and genetic maps have also been produced for many plant species and have been used extensively for interspecies and intraspecies chloroplast genome comparisons. Finally, the improvement in molecular cloning and sequencing techniques has allowed several novel chloroplast genomes to be revealed, providing important information for further genetic and proteomic studies (table 2.1).

### 2.2.1 Chloroplast DNA Isolation

“The problems inherent in working with plants are largely manifested in the DNA isolation stage” (Milligan 1992). This is especially true when attempting to isolate cpDNA from plant material. A number of methods are currently available for extracting cpDNA, but there are a few general principles that apply to all of them.

The quality of the starting material and the way in which it is stored has played a critical role in successful cpDNA extractions. Very fresh, young and healthy material has given the best results. Storing the leaves for 48 h in a cool, dark environment gave the chloroplasts time to deplete their stored starch grains. The presence of starch during cpDNA extraction makes the chloroplast membranes more prone to rupturing and therefore lowers the yield of cpDNA obtained (Palmer 1988, Mourad 1998, Milligan 1992, Lodhi et al. 1994, Mariac et al. 2000). Most protocols that have been used to isolate cpDNA make use of aqueous buffers and have two steps: First, the isolation of intact plastids, separate from the remaining cell contents and second, the lyses of the plastid and purification of the cpDNA. These methods are often simple, inexpensive and use non-toxic compounds (Milligan 1992).

Extraction steps carried out as swiftly as possible and at 0 to 4°C retard the action of hydrolytic enzymes and reduce chloroplast membrane shearing. Problems in extracting pure,

**Table 2.1** Completely Sequenced Chloroplast Genomes (Last updated 12 Nov 2003)  
([http://megasun.bch.umontreal.ca/ogmp/projects/other/cp\\_list.html](http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list.html)).

<b>Locus</b>	<b>Accession</b>	<b>Last updated</b>	<b>Organism</b>
<a href="#">AY178864</a>	AY178864	02-May-2003	<i>Adiantum capillus-veneris</i> [Embryophyta]
<a href="#">ATR506156</a>	AJ506156	22-Sep-2003	<i>Amborella trichopoda</i> [Embryophyta]
<a href="#">AB086179</a>	AB086179	10-Mar-2003	<i>Anthoceros formosae</i> [Embryophyta]
<a href="#">AP000423</a>	AP000423	08-Apr-2000	<i>Arabidopsis thaliana</i> [Embryophyte]
<a href="#">ALO294725</a>	AJ294725	12-Apr-2001	<i>Astasia longa</i> [Euglenozoa]
<a href="#">ABE316582</a>	AJ316582	06-Sept-2002	<i>Atropa belladonna</i> [Embryophyta]
<a href="#">CFE421413</a>	AJ428413	07-Aug-2003	<i>Calycanthus fertilis</i> var. <i>ferax</i> [Embryophyta]
<a href="#">AF494278</a>	AF494278	22-Aug-2002	<i>Chaetosphaeridium globosum</i> [Embryophyta]
<a href="#">AB001684</a>	AB001684	14-Apr-2000	<i>Chlorella vulgaris</i> [Chlorophyta]
<a href="#">AF022186</a>	AF022186	19-Jan-2000	<i>Cyanidium caldarium</i> [Rhodophyta]
<a href="#">CPU30821</a>	U30821	13-Nov-1995	<i>Cyanophora paradoxa</i> [Glaucocystophyceae]
<a href="#">AY217738</a>	AY217738	02-Jun-2003	<i>Eimeria tenella</i> [Alveolata]
<a href="#">EPFCPCG</a>	M81884	14-Apr-2000	<i>Epifagus virginiana</i> [Embryophyta]
<a href="#">CLEGCGA</a>	X70810	01-Feb-2003	<i>Euglena gracilis</i> [Euglenozoa]
<a href="#">AF041468</a>	AF041468	03-Mar-1999	<i>Guillardia theta</i> [Cryptophyta]
<a href="#">AP002983</a>	AP002983	22-Jul-2003	<i>Lotus corniculatus</i> var. <i>japonicus</i> [Embryophyta]
<a href="#">CHMPXX</a>	X04465	16-Feb-2000	<i>Marchantia polymorpha</i> [Embryophyta]
<a href="#">AC093544</a>	AC093544	11-Mar-2003	<i>Medicago truncatula</i> [Embryophyta]
<a href="#">AF166114</a>	AF166114	17-Mar-2000	<i>Mesostigma viride</i> [Chlorophyta]
<a href="#">AF137379</a>	AF137379	08-Aug-2003	<i>Nephroselmis olivacea</i> [Chlorophyta]
<a href="#">CHNTXX</a>	Z00044	27-Apr-1998	<i>Nicotiana tabacum</i> [Charophyta/Embryophyta group]
<a href="#">OXCHLPLXX</a>	Z67753	21-Jan-1998	<i>Odontella sinensis</i> [Stramenopiles]
<a href="#">OEL271079</a>	AJ271079	10-Mar-2001	<i>Oenothera elata</i> [Embryophyta]
<a href="#">X15901</a>	X15901	29-Mar-2001	<i>Oryza sativa</i> ( <i>japonica</i> cultivar-group) [Embryophyta]
<a href="#">AP005672</a>	AP005672	02-Oct-2003	<i>Physcomitrella patens</i> [Embryophyta]
<a href="#">PINCPTRPG</a>	D17510	14-Apr-2000	<i>Pinus thumbergii</i> [Embryophyta]
<a href="#">PPU38804</a>	U38804	27-Mar-1998	<i>Porphyra purpurea</i> [Rhodophyta]
<a href="#">AP004638</a>	AP004638	15-Nov-2002	<i>Psilotum nudum</i> [Embryophyta]
<a href="#">SOL400848</a>	AJ400848	20-Feb-2003	<i>Spinacia oleracea</i> [Embryophyta]
<a href="#">U87145</a>	U87145	29-Jun-1999	<i>Toxoplasma gondii</i> [Alveolata]
<a href="#">AB042240</a>	AB042240	13-Apr-2002	<i>Triticum aestivum</i> [Embryophyta]
<a href="#">ZMA86563</a>	X86563	25-Jan-2003	<i>Zea mays</i> [Embryophyta]

intact cpDNA have often been a result of high levels of secondary metabolites present in certain plants. These contaminants include tannins and polyphenolics, and cause DNA degradation, a viscous end sample and difficulties in PCR and restriction digestions. The extraction buffer therefore usually needs to be optimised for a specific species. Most extraction buffers have a pH of 8.0, but some have a pH as high as 9.0. The ingredients of the extraction buffer often had a considerable influence on the success of the isolation. Most extraction buffers include sucrose/mannitol/sorbitol, Tris-Cl (pH 8.0) and EDTA (pH 8.0). EDTA inhibits the action of cellular enzymes that could degrade the cpDNA. A number of other components have also been included in extraction buffers that protect the cpDNA from degradation by native enzymes or secondary metabolites. PVP helps to absorb tannins and other secondary plant compounds. NaCl serves as an osmoticum, minimising nuclear DNA contamination and removing polysaccharides. High levels of 2-mercaptoethanol, glutathione, cysteine, dithiothreitol (DTT), and other thiols prevent phenolic oxidation. Bovine serum albumin (BSA) causes surface denaturation of enzymes that degrade cpDNA (Herrmann et al. 1980, Bookjans et al. 1984, Smith and Ma 1985, Palmer 1988, Schuler and Zielinski 1989, Mariac et al. 2000, Milligan 1992, Lodhi et al. 1994, Porebski et al. 1997).

Special care needs to be taken during the extraction process to disrupt only the cell walls and not the chloroplast membranes. For this reason the use of a Waring blender is the preferred method of leaf maceration. The homogenate is filtered through cheesecloth followed by Miracloth to remove large pieces of tissue and cells. This mixture is then centrifuged to separate chloroplasts from large cell debris (Palmer 1988, Schuler and Zielinski 1989, Mariac et al. 2001).

Nuclear DNA contamination is the most common problem in aqueous cpDNA isolation procedures. One of three methods is usually used to eliminate nuclear DNA. The first method involves treatment with DNase, which digests nuclear DNA but is unable to digest the cpDNA contained in the intact chloroplast membranes. This method gives very pure cpDNA, but only at low yields. Degradation of the cpDNA also occurs due to the partial rupturing of some chloroplast membranes (Herrmann et al. 1975, Palmer 1988, Triboush et al. 1998). The second method involves the use of discontinuous (step) sucrose or percoll gradients, in which the intact chloroplasts are banded separately from other cellular components according to their buoyant densities. This method gives pure cpDNA, but is time-consuming (Hirai et al. 1985, Mourad 1998, Palmer 1988). The third method involves the differential lyses of chloroplasts and nuclei and involves extraction of cpDNA with CTAB. This protocol has the

advantage of being simple and inexpensive, but nuclear DNA contamination may still be present (Palmer 1988, Milligan 1992).

The first step in cpDNA isolation is the lyses of the chloroplast membrane. CTAB, SDS (sodium dodecyl sulphate), sodium sarkosinate and/or proteinase K (Bookjans et al. 1984, Mariac et al. 2000, Mpoloka 2001) have been used for this purpose. Subsequent organic (phenol/chloroform/isoamyl alcohol) extraction steps have been performed in some cases to remove proteins that may still have been present. Further removal of proteins and RNA has also been carried out through the use of CsCl density gradients. Chloroplast DNA and nuclear DNA band at different densities and therefore very pure cpDNA can be recovered. However, this is a time-consuming and costly process that gives relatively low yields (Herrmann et al. 1975, Palmer 1988).

For certain plant species, especially those with narrow, fibrous leaves, aqueous isolation methods do not give satisfactory results. Non-aqueous methods of cpDNA extraction have been utilized as an alternative. These methods do not require the isolation of intact plastids and therefore plant material may be stored for a long time before isolations need to be done. Non-aqueous organic solvents, such as hexane-carbon tetrachloride and trichlorobenzene-naphthalene, are used to isolate the cpDNA in these protocols (Milligan 1992).

### **2.2.2 Restriction Digestion, Physical and Genetic Map Construction**

Once the cpDNA has been successfully isolated from a plant it is usually digested with a number of different restriction enzymes and separated on an agarose gel. The banding patterns created by the separation of the digested fragments on the gel aids in determining which restriction enzymes are most appropriate for creating a physical map of the chloroplast genome. These are usually enzymes that cut the chloroplast genome rarely (about 10-20 times) (Palmer 1988).

Physical (restriction site) maps give an indication of the number, order, and distance between restriction enzyme cutting sites along the chloroplast genome (Klug and Cummings 1997). They have allowed researchers to determine the size of the chloroplast genome and have confirmed its circular nature. Restriction site maps have been created for a number of species including *N. tabacum*, *Spinacea oleracea*, and *Z. mays* (Bovenberg et al. 1981).



A number of methods can be used to create a physical map of a chloroplast genome. One method uses the sequential (reciprocal double digest) technique. The cpDNA is first digested with a single enzyme and run on low-gelling-temperature agarose. The individual fragments are excised from the gel and digested with a second enzyme. These digested fragments are run separately on a second low-gelling-temperature agarose gel. By comparing the fragment sizes obtained from both single and double digests, the fragments can be placed in the correct order and a restriction site map can be created (Herrmann et al. 1980, Gounaris et al. 1986, Palmer 1988, Lim et al. 1990).

A second method for creating a physical map makes use of overlap hybridization. A number of restriction fragments, together covering the entire chloroplast genome, are used as probes. These probes are hybridized to Southern blots of single- and double-digested cpDNA separated on agarose gels. The resulting autorads are compared to the digested cpDNA run on agarose gels and the order of the restriction fragments can be determined in this way (Smith and Ma 1985). An alternative is to use the cpDNA probes in Southern hybridization with restriction-digested total DNA (Palmer 1988, Lee et al. 1996, Cosner et al. 1997).

Genetic mapping allows genes to be placed in the correct place on a genome. Chloroplast genes that have been isolated from another plant species are often used as probes (Ma and Smith 1985). These heterologous, labeled probes are hybridized to Southern blots of restriction-digested cpDNA run on agarose gels. The photographs are compared to ethidium bromide-stained agarose gel photos of the digested cpDNA and a genetic map is constructed of the entire chloroplast genome (Lee et al. 1996, Cosner et al. 1997, Katayama and Uematsu 2003).

### **2.2.3 Clone Library Construction and Genome Sequencing**

The chloroplast genome is relatively small (usually between 120 and 160 kb) and is therefore easier and faster to clone and sequence than plant nuclear genomes. A clone library that covers an entire chloroplast genome usually consists of overlapping fragments to ensure that small fragments are not overlooked (Shinozaki et al. 1986). A number of chloroplast genomes of both plants and algae species have already been cloned and sequenced.

Restriction enzymes that cut the chloroplast genome between 10 and 20 times are ideal for use in clone library construction. The restriction endonucleases *Pst*I, *Sac*I, *Sal*I, *Kpn*I, *Xho*I, *Sph*I,

*Sma*I and *Nru*I have proven to be most successful for this purpose. The cpDNA fragments are usually cloned into plasmid vectors, although cosmid vectors and  $\lambda$  phage vectors have been used to clone larger fragments (Palmer 1988, Sugiura et al. 1986, Lee et al. 1996). The entire maize chloroplast genome has even been cloned into a yeast artificial chromosome (YAC) vector (Gupta and Hoo 1991).

The commonly used strategy for library construction is to first shotgun-clone restriction digested cpDNA fragments into a plasmid vector. Recombinant colonies are then isolated and screened. The screening process involves isolating the clone DNA by means of an alkaline miniprep procedure (e.g. Sambrook et al. 1989) and digestion of the clones with the appropriate restriction enzyme(s). These clones are electrophoresed on an agarose gel to determine their sizes. Any fragments that have not been successfully cloned can be isolated from agarose gels and ligated in separate reactions (Palmer and Thompson 1981, Ohshima et al. 1986, Palmer 1988, O’Kane 1995, Cosner et al. 1997). The Polymerase Chain Reaction (PCR) has also been used to close gaps between restriction fragments (Schmitz-Linneweber et al. 2001).

As mentioned previously, it is important to obtain overlapping DNA fragments that will cover the entire genome. Therefore, more than one library is usually constructed; each consisting of fragments digested with different restriction enzymes. The next step is to determine the sequence of each clone and then compile and analyze the information using computer software programs (Shinozaki et al. 1986, Hiratsuka et al. 1989, Sugiura et al. 2003).

## 2.3 LARGE-CONSTRUCT VECTORS

With the improvement in sequencing technology the sequencing of the entire genomes of higher organisms became a viable option. However, most conventional plasmid vectors cannot incorporate inserts larger than 20 kb and were therefore not ideal for the physical mapping and large-scale sequencing of large genomes. For this purpose vectors that could stably maintain very large inserts, but which would still be relatively easy to manipulate, were required.

Several types of large-insert DNA vectors have been created. These can be classified into two categories, namely bacteria- and yeast-based cloning systems. Yeast artificial chromosomes

(YACs) are an example of a yeast-based cloning system while cosmids, fosmids, P1, bacterial artificial chromosomes (BACs) and P1-derived artificial chromosomes (PACs) are bacteria-based cloning systems. The latter are often favored because insert DNA purification is much easier for bacteria-based cloning systems than for those involving yeasts. Bacteria-based systems are also all circular constructs and therefore protect their DNA from shearing during purification (Zhang and Wu 2001).

### 2.3.1 Cosmids

Cosmids were one of the first vectors to be used for cloning large constructs and for the study of large, complex genomes. Cosmids are hybrid vectors that include parts of both the lambda chromosome and plasmid DNA. The *cos* sequence obtained from the lambda phage is required for the packaging of the phage DNA into the phage protein coat. The plasmid sequences for the antibiotic resistance gene and for replication are also included in the cosmid vector. Fragments of up to 50 kb can be cloned into and maintained in a cosmid vector by using the standard bacteriophage-based transfection system. Although this system was once widely used and the cosmid DNA is easy to purify, it has limited use for large-scale physical mapping and genome sequencing because of its limited cloning capacity and the instability of insert DNA (Klug and Cummings 1997, Zhang and Wing 1997, Zhang and Wu 2001).

### 2.3.2 Yeast Artificial Chromosomes (YACs)

The YAC system were created in 1987 by Burke et al. (Burke et al. 1987) and have played an important role in the initial physical mapping and genome sequencing of large genomes such as the human chromosomes. YACs are linear constructs that have the typical features of a chromosome. These include an autonomous replication sequence (ARS), a yeast centromere (CEN), and two yeast telomeres (TEL). Selectable markers (TRP1 and URA3) are found on each side of the centromere and a multiple cloning site is also present. Very large fragments of up to one million base pairs (1 megabase) of DNA can be inserted into a YAC, a 25-fold advance over cosmid vectors (Klug and Cummings 1997, Zhang and Wu 2001).

A number of problems have, however, been encountered when using YACs. YAC DNA is difficult to purify from the yeast host genomic DNA. This can only be done via pulsed-field gel electrophoresis (PFGE) and gel isolation (Zhang and Wing 1997, Venter et al. 1996). Approximately half of YAC clones are chimeric, possessing two or more inserts in one clone.

Many also contain inserts that are structurally unstable, with deletions or rearrangements occurring. Such clones cannot be used for further mapping and sequencing and much time is spent trying to identify and remove these clones. The presence of a large number of tandem repeat regions in some genomes makes this task even more difficult.

### **2.3.3 Bacterial Artificial Chromosomes (BACs)**

The development of BACs (Shizuya et al. 1992) provided an alternative method of constructing large-insert based genome libraries and has the advantages of both cosmid and YAC systems. The BAC vector is based on the *Escherichia coli* fertility (F-factor) plasmid and contains the F-factor genes for replication and copy number. The F-factor strictly regulates its own transcription and prevents more than one or two copies of the BAC from being present in a bacterial cell at one time. Insert DNA is therefore stably integrated, with only a small percentage of sequence artefacts present. BAC vectors usually also possess an antibiotic resistance gene and a polycloning site situated within a reporter gene (Yang et al. 1997).

There are a number of advantages in using the BAC system. BACs are circular and supercoiled and therefore easier to manipulate and less prone to shearing than the linear YAC. Since BACs are transformed into normal *E. coli* bacterial cells they are easy to grow up and isolate, with high transformation efficiencies. Conventional bacterial colony lifts and hybridization methods can also be used to screen colonies in a library. Inserts of up to 500 kb, with most averaging around 150 kb, can be stably integrated and maintained in a BAC vector (Yang et al. 1997, Zhang and Wu 2001).

Although the BAC system is relatively new, it has been used for the creation of high-resolution physical maps and map-based gene cloning of a number of large genomes. It has also facilitated the process of sequencing and subsequent sequence compilation. One example is the crucial role BAC clones played in the clone-by-clone shotgun (CBCS) approach used in the Human Genome Project (HGP). YACs and cosmid vectors were initially utilized for the construction of a physical map of the genome, with more than 75% of genome being covered by YAC contigs. However, difficulties in manipulating the large YAC inserts and eliminating yeast host DNA during YAC purification were serious obstacles. Subcloning of large YAC inserts into BACs was therefore an attractive option for further physical mapping and sequencing (Zhang and Wing 1997, Zhang and Wu 2001).

The CBCS system that was subsequently used involved a number of steps. Firstly, high molecular weight genomic DNA was partially digested with restriction enzymes and fragments of the correct size (40 to 400 kb) were cloned into a BAC vector. The next step was to construct a physical map of the entire genome. This step was very important in ensuring that only minimally overlapping clones were chosen for sequencing. A number of different methods were used to obtain a physical map. One such strategy, originally proposed by Venter et al. (1996), combined physical mapping and genome sequencing and has also been used in the *Arabidopsis* (Mozo et al. 1999, Marra et al. 1999) and *Drosophila* (Adams et al. 2000) sequencing projects. This method involved the sequencing of both ends of each BAC, the so-called 'sequence-tagged connectors' (STCs). Fingerprint analysis, involving the digestion of each BAC clone with a single restriction enzyme, was then performed. The fragments were separated on electrophoretic gels, allowing insert sizes to be determined and overlapping clones to be identified. 'Seed' BACs of interest were chosen and completely sequenced. These were checked against the STC database to find overlapping clones. This method allowed researchers to walk out from the seed BACs in a manner that would include overlapping clones (Venter et al. 1996, Olson 2001, Zhang and Wu 2001).

BAC vectors offer a simple, fast and economic alternative to other large-construct vectors. They have already become a popular choice for the physical mapping and sequencing of large genomes and could provide a powerful tool in accelerating future genomic research of these genomes.

## 2.4 CONCLUSION

Our understanding of the chloroplast has improved dramatically over the past few decades. There are, however, many questions regarding both primary and secondary plastids that remain unanswered. Although the chloroplast originated from a cyanobacterial progenitor, no modern cyanobacterial group shows any specific similarity to chloroplasts and their genomes. The number of primary and secondary endosymbiotic events also remains the subjects of intense debate. Most evidence points towards a single primary endosymbiotic origin and at least two secondary endosymbiotic events. However, a great deal of research still needs to be performed to clarify these issues.

The marked improvement in methods of cpDNA isolation, cloning technology, sequencing methods, computer power and software programs have meant that complete chloroplast genome sequences can be obtained at a faster rate. Large-construct vectors such as BACs may also prove useful. Once isolated and cloned, the entire chloroplast genome could be contained in only one or two BAC clones. Further manipulation of the genome would be facilitated by the relative ease of bacterial DNA isolation and maintenance, in comparison to the isolation procedures required for the isolation of cpDNA from plant material.

The information made available through chloroplast genome sequencing can be used in further genomic and proteomic studies. Phylogenetic analysis including multiple genes and taxa are now possible. Phylogenetic data, together with morphological observations and fossil records, should provide us with a more comprehensive view of plastids and their evolution. By understanding the evolution of plastids, researchers will have further insight into early eukaryotic evolution.

Studies on gene expression and the regulation of gene expression, as well as studies of protein transport mechanisms used to and from the nucleus will also be possible. This information, together with improvements in plastid transformation technology, will contribute to more successful foreign gene expression in the plastid.

Plastid transformation technology could provide researchers with a valuable alternative to current methods of nuclear transformation. The former offers a number of advantages over the latter, including containment of transgenes and higher expression levels. This technology is already being applied for research and agricultural purposes and has a promising future for pharmaceutical and edible vaccine production. Transplastomic plants have the potential to provide a cost-effective, simpler and more effective method of foreign gene expression than those currently available.

### 3 MATERIALS AND METHODS

#### 3.1 ISOLATION AND CHARACTERISATION OF *VITIS VINIFERA* CHLOROPLAST DNA

*V. vinifera* grapevine chloroplast DNA, isolated from the leaves of the cultivar Sultana and Sogra 1, was digested with appropriate restriction enzymes and shotgun-cloned into a plasmid vector. The clones obtained in this way were characterised, partially sequenced and mapped against the *Atropa belladonna* chloroplast genome.

##### 3.1.1 Plant Material

Young, green *Vitis vinifera* cv. L. Sultana leaves were collected from the Agricultural Research Council (ARC) Infruitec-Nietvoorbij during spring and summer. During autumn and winter *Vitis vinifera* cv L. Sogra 1 leaves were obtained from the glasshouse at the Institute of Wine Biotechnology (IWBT), University of Stellenbosch. The leaves were stored in the dark at 4°C for 48 hours before being utilised for chloroplast DNA (cpDNA) extractions.

##### 3.1.2 Isolation of *V. vinifera* Chloroplast DNA

A variation on the method of Yeoh and Joseph (1995) was used to isolate grapevine cpDNA from both Sogra 1 and Sultana leaves. Leaves were washed in distilled water and cut into pieces of 1 to 2 cm<sup>2</sup>. All further steps were performed at 0 to 4°C unless otherwise indicated. Fifty grams of leaf material was added to 500 ml of extraction buffer (1.75 M NaCl; 20 mM EDTA; 0.1 M Tris-Cl, pH 8.0; 1% BSA (w/v); 0.2% β-mercaptoethanol (v/v); 5% PVP-10 (w/v)) and homogenised in a Waring blender with 5 pulses of 5 seconds each. The homogenate was filtered through 4 layers of cheesecloth (with squeezing) and through 4 layers of Miracloth (Calbiochem; without squeezing). The filtrate was centrifuged (1800 x g, 10 min, 4°C) and the resulting pellets were resuspended in 10 ml of extraction buffer. The centrifugation step was repeated and the pellet was again resuspended in 10 ml of extraction buffer. A 1/10 volume of 10% CTAB (w/v) was added and the mixture was incubated at 60°C for one hour. An equal volume of phenol:chloroform:isoamylalcohol (25:24:1 v/v) was added, the solution was mixed by inversion and centrifuged (11 000 x g, 10 min, 4°C). The

aqueous phase was removed to new microfuge tubes and an equal volume of chloroform:isoamylalcohol (24:1) was added and vortexed. The mixture was centrifuged (11 000 x g, 10 min, 4°C). The aqueous phase was removed to new tubes containing a 1/10 volume of 3 M sodium acetate and 0.8 volumes of isopropanol and precipitated overnight at -20°C. The DNA was concentrated by centrifugation (11 000 x g, 10 min, 4°C). The resulting pellet was washed in 100% ethanol (11 000 x g, 10 min, 4°C) followed by a 70% ethanol wash (11 000 x g, 5 min, 4°C). The pellet was air-dried for 5 to 10 minutes and resuspended in 500 µl of double distilled water containing RNase A (30 µg/ml). The resuspended cpDNA was incubated at 37°C for 20 minutes to allow RNA degradation. To get an estimate of the cpDNA concentration, the cpDNA was run on an agarose gel in parallel with lambda standards. The cpDNA was stored at -20°C until further use.

### 3.1.3 Restriction Enzyme Digestion

A total of 100 ng of isolated cpDNA was digested with a number of restriction endonucleases with hexanucleotide recognition sequences. The enzymes included *EcoRI*, *PstI*, *XhoI*, *Sall*, *BamHI*, *HindIII*, *ClaI*, *BglII* and *XbaI*. All restriction digestions were performed according to the manufacturer's instructions (Roche) with the appropriate buffer and incubated at 37°C, overnight.

Restriction fragments were separated by electrophoresis on a 0.8% (w/v) agarose gel (Hispanagar) in 1X TAE (0.04 M Tris-acetate; 0.001 M EDTA, pH 8.0). The gel was run at 20V for 16 to 20 h and stained in 1X TAE buffer containing 0.5 µg/ml ethidium bromide.

### 3.1.4 Library Construction

#### 3.1.4.1 Shotgun Cloning

Fragments of cpDNA digested with *EcoRI*, *BamHI*, and *PstI*, respectively, were precipitated with 1/10 volumes of NaOAc and 2 volumes of 100% ethanol. The dried pellet was resuspended in 5 µl of ddH<sub>2</sub>O. The plasmid vector pUCBM21 (Roche – appendix 4) was digested with the appropriate restriction enzyme (*EcoRI/BamHI/PstI*), and treated with 1U of SAP (Shrimp Alkaline Phosphatase; Promega) for 15 minutes at 37°C. The reaction was inactivated by incubating for 15 minutes at 65°C. Ligations were



performed according to the manufacturer's instructions (New England Biolabs) and incubated at 16°C, overnight. Refer to table 3.1 for vector:insert ratios

**Table 3.1.** Ligation reactions: Column 1: The order in which ligation reactions were performed. Column 2: Restriction enzymes used to digest the vector and cpDNA for ligation. Column 3: The vector to insert ratios used during each ligation is shown.

Ligation no.	Restriction enzyme	Vector (pUCBM21): Insert (cpDNA)
1	<i>EcoRI</i>	1:2, 1:4, 1:6, 1:12
2	<i>BamHI</i>	1:2
3	<i>PstI</i>	1:12

### 3.1.4.2 Preparation of Competent Cells

The following rubidium chloride competent cell preparation protocol was adapted from Hanahan (1985).

A single colony of *E.coli* DH5 $\alpha$  was inoculated into 2.5 ml of LB medium and incubated at 37°C, overnight (225 rpm). The overnight culture was used to inoculate 250 ml LB medium containing 20 mM MgSO<sub>4</sub>, in a one-liter flask. The cells were incubated at 37°C, with shaking (225rpm). The absorption value ( $A_{600}$ ) was determined after 2 h and every 30 minutes thereafter, until the  $A_{600}$  reached 0.4-0.6. The cells were pelleted by centrifugation (4 500 x g, 5 min, 4°C). The cell pellets were gently resuspended in 100ml of ice-cold TFB1 (30 mM KaO<sub>c</sub>; 10 mM CaCl<sub>2</sub>; 50 mM MnCl<sub>2</sub>; 100 mM RbCl; 15% (v/v) glycerol - pH adjusted to 5.8 with 1 M acetic acid, filter sterilised (0.2 mm)). The resuspended cells were combined and incubated on ice for 5 min at 4°C. For the remainder of the procedure the cells were kept on ice and all pipettes, tips, microfuge tubes and flasks were also chilled. The cells were pelleted by centrifugation (4 500 x g, 5 min, 4°C), and gently resuspended in 1/25 of the original culture volume of ice-cold TFB2 (10 mM MOPS or PIPES, pH 6.5; 75 mM CaCl<sub>2</sub>; 10 mM RbCl; 15% glycerol - pH adjusted to 6.5 with 1 M KOH, filter sterilized (0.2 mm) and stored at room temperature). The competent cells were treated very gently due to their high sensitivity to handling and elevated temperature. The competent cells were incubated on ice for 15-60 minutes and then aliquoted 100  $\mu$ l/tube for storage at -80°C. The cells were subjected to quick-freezing in a dry ice/isopropanol bath before storage.

### 3.1.4.3 Transformation

Competent *Escherichia coli* DH5 $\alpha$  cells were transformed with the ligation mixture according to a variation on the protocol found in the Promega “Protocols and Applications Guide”.

Fifty microlitres of competent cells were thawed on ice. One to ten microlitres of the ligation mixture was added to the competent cells, gently mixed, and incubated for 20 minutes on ice. The tubes were heat-shocked for 45 seconds at 42°C and rapidly placed on ice for 2 minutes. Four hundred and fifty microlitres of LB medium was added and incubated for 90 minutes at 37°C with shaking (150 rpm). One hundred microlitres of the transformation mix was plated out onto plates containing LB medium, 100  $\mu$ g/ml ampicillin, 40  $\mu$ l of a 0.1 M Isopropylthiogalactoside (IPTG) stock and 40  $\mu$ l of a 50 mg/ml 5-bromo-4-chloro-3-indocyl- $\beta$ -D-galactoside (X-gal) stock (Promega). Recombinant colonies were selected using blue/white screening and transferred to fresh LB plates containing ampicillin (100  $\mu$ g/ml). Colonies were arranged in a grid pattern on each plate to facilitate their transfer to new plates by replica plating.

### 3.1.5 Library Screening by PCR

Potentially positive clones were screened by means of PCR using vector specific M13 forward and reverse primers. Each colony was picked using a sterile toothpick and swirled in 10  $\mu$ l of water. The 20  $\mu$ l PCR reaction contained 10  $\mu$ l of clone DNA, 15 pmols of each primer, 1.5 mM MgCl<sub>2</sub>, 50  $\mu$ M dNTPs, 0.25 U/ $\mu$ l BIOTAQ<sup>TM</sup> DNA polymerase (Bioline) and 1 x NH<sub>4</sub> buffer (16 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 670 mM Tris-HCl, pH 8.8; 0.1% Tween-20). The DNA was heat denaturated at 95°C for 5 minutes followed by 30 cycles of 30 seconds at 94°C, 30 seconds at 55°C and 60 seconds at 72°C.

### 3.1.6 Library Screening by Colony Blotting

#### 3.1.6.1 Colony Lifts

The colonies in the *Eco*RI chloroplast DNA library were transferred and fixed onto 82 mm diameter Hybond<sup>TM</sup>-N nylon membrane discs (Amersham LIFE SCIENCE) according to a variation of the protocol found in the “DIG Application Manual for Filter Hybridisation” (Roche 2000). Replica plating was used to transfer colonies onto fresh LB plates

containing ampicillin (100 µg/ml) and these were incubated overnight at 37°C to allow growth. Nylon membranes were placed onto fresh LB plates containing ampicillin (100 µg/ml) and the freshly grown colonies were transferred onto the membranes. Colonies were incubated at 37°C, overnight. Membranes were removed from the plates and each placed onto 1ml drops of Denaturation Solution (1.5 M NaOH; 1.5 M NaCl) on plastic film for 15 minutes. They were transferred onto dry filter paper for one minute and onto one ml drops of Neutralization Solution (1.5 M NaCl; 1.0 M Tris-HCl, pH 7.4) for a further 15 minutes. Membranes were again transferred onto dry filter paper for one minute and submerged in 2X SSC (1.5M Tri-sodium citrate; 3M NaCl, pH 7) for 10 minutes. In order to dry the membranes thoroughly, they were placed on filter paper for approximately 30 minutes and the DNA was cross-linked to the membranes by exposing them to UV light (360 nm) for 4 minutes. To remove any remaining traces of bacterial debris the membranes were washed overnight in 3X SSC and 0.1% SDS (w/v). They were then submerged two times for 5 minutes in 2X SSC. The membranes were stored at 4°C until further use.

### **3.1.6.2 Probe Labelling**

Two *Pst*I-digested grapevine cpDNA fragments of approximately 5500 bp and 2500 bp in size were used as probes to find overlapping clones in the *Eco*RI clone library. These fragments were randomly labelled with DIG-High Prime (Roche) according the manufacturer's instructions. Approximately 500 ng of template DNA was added to sterile ddH<sub>2</sub>O to a final volume of 16 µl. The DNA was denatured in a boiling water bath for 10 minutes and flash cooled on ice. Four microlitres of DIG-High Prime was added to the denatured DNA and centrifuged briefly. The labelling mix was incubated at 37°C for 20 h. The reaction was stopped by adding 0.2 M EDTA and heating to 65°C for 10 minutes. The probe labelling efficiency was checked by means of the direct detection method described in the "DIG Application Manual for Filter Hybridisation" (Roche 2000), using labelled control DNA (Boehringer Mannheim).

### **3.1.6.3 Prehybridisation and Hybridisation**

The protocol found in the Roche "DIG Application Manual for Filter Hybridisation" (Roche 2000) was followed for prehybridisation and hybridisation. Membranes were placed in roller-bottles and prehybridised in DIG Easy Hyb (Roche) at 37°C for a minimum of 1 h. The labelled probe was denatured at 95 to 100°C for 5 minutes. The

probe was flash cooled on ice and diluted in DIG Easy Hyb to a final concentration of 25 ng/ml. This hybridisation solution was added to the roller-tubes after removing the prehybridisation solution. Hybridisation was performed overnight at 37°C. Membranes were washed two times for 5 minutes at room temperature in Low Stringency Wash Buffer (2 x SSC; 0.1% SDS (w/v)) followed by two 15-minute washes in High Stringency Wash Buffer (0.5 x SSC; 0.1% SDS (w/v)) at 65°C.

#### **3.1.6.4 Detection**

After the high stringency washes the membranes were placed in Washing Buffer (0.1 M Maleic acid; 0.15 M NaCl, pH 7.5; 0.3% (v/v) Tween<sup>®</sup> 20) for 5 minutes and then transferred into Blocking Solution (Roche) for 30 minutes. For a further 30 minutes the membranes were placed in Antibody Solution containing a 1:10 000 dilution of anti-digoxigenin-AP Fab fragments (Roche) in Blocking solution. Two 15-minute incubations in Washing Buffer were carried out followed by equilibration of the membranes in Detection Buffer (0.1 M Tris-HCl; 0.1 M NaCl, pH 9.5) for 5 minutes. Each membrane was placed between two sheets of plastic film. CDP-*Star*<sup>™</sup> solution (200 µl; Roche) was spread over each membrane, and the top sheet of plastic was placed over this. The membranes were incubated for 5 minutes and sealed in the plastic film. Exposure of the membrane to X-ray film (hyperfilm - Amersham Pharmacia Biotech) was carried out for 30 minutes to overnight and the film was developed.

#### **3.1.7 Plasmid DNA Isolation**

A single colony of each clone that tested positive through PCR or colony blotting was picked and inoculated into 5 ml of LB medium containing ampicillin (100 µg/ml). The culture was incubated at 37°C for 12 to 16 h, with shaking (~225 rpm). DNA was isolated using the Wizard<sup>®</sup> Plus SV Minipreps DNA Purification System (Promega). The centrifugation protocol accompanying the kit was followed.

The 5 ml culture was pelleted (16 000 x g, 5 min) and the pellet was resuspended in 250 µl of Cell Resuspension Solution (50 mM Tris-Cl, pH 7.5; 10 mM EDTA; 100 µg/ml RNase A). Two hundred and fifty microlitres of Cell Lysis Solution (0.2 M NaOH; 1% SDS) was added and the tube was inverted 4 times. Alkaline Protease solution (10 µl) was added and the tube was inverted four times. The sample was incubated for 5 minutes at room temperature and 350 µl of Neutralization Solution (4.09 M guanidine hydrochloride; 0.759 M potassium

acetate; 2.12 M glacial acetic acid, pH 4.2) was added. The tube was inverted 4 times and centrifuged (16 000 x g, 10 min, 22°C). The cleared lysate was transferred into a spin column placed within a collection tube and centrifuged (16 000 x g, 1 min, 22°C). The flowthrough was discarded and the spin column placed back into the collection tube. Seven hundred and fifty microlitres of Column Wash Solution (60 mM potassium acetate; 8.3 mM Tris-Cl, pH 7.5; 40 µM EDTA; 60% ethanol) was added to the spin column and centrifuged (16 000 x g, 1 min, 22°C). The flow through was discarded; 250 µl of Column Wash Solution was added to the spin column and centrifuged (16 000 x g, 1 min, 22°C). The flow through was discarded; the spin column was placed back in the collection tube and centrifuged (16 000 x g, 1 min, 22°C). The spin column was placed in a sterile 1.5 ml tube and 50 to 100 µl of ddH<sub>2</sub>O was added. The sample was incubated for 5 minutes and centrifuged (16 000 x g, 2 min, 22°C). Plasmid DNA was stored at – 20°C.

### 3.1.8 Clone Sequencing and Sequence analysis

DNA samples were sequenced at the Central Analytical Facility (Genetics Department, Stellenbosch University). Sequencing was carried out on an ABI Prism 3100 Genetic Analyzer using vector-specific M13 primers. The nBLAST 2 program (Altschul et al. 1997) was used to compare sequences to those found in the non-redundant NCBI database (<http://www.ncbi.nlm.nih.gov/>).

### 3.1.9 PCR of the *rrn16* Flanking Regions

In a previous study conducted at the Department of Genetics at the University of Stellenbosch, the sequence of a region of the *rrn16* gene of the grapevine chloroplast genome was determined. The 5' and 3' ends of this sequence were used to design outward-facing primers. Degenerate complementary primers were designed by using completely sequenced chloroplast genomes of a number of other plant species as a template.

The primers were used in an Expand High Fidelity PCR reaction to amplify the *rrn16* flanking regions from isolated grapevine chloroplast DNA. The 50 µl PCR reaction contained 200 µM dNTPs, 300 nmols of each primer, ~1 ng of cpDNA, 1 x buffer (containing 15 mM MgCl<sub>2</sub>), and 2.6 U of Expand High Fidelity PCR System enzyme mix. Reactions were heated at 94°C for 2 minutes. This was followed by 10 cycles of 94°C for

15 seconds, 58°C for 30 seconds, and 68°C for 45 seconds. The 10 cycles were followed by a further 30 cycles with the same steps, except for an increase of the elongation step by 5 seconds for each cycle. After the completion of the last cycle the samples were kept at 68°C for a further 7 minutes for final elongation.

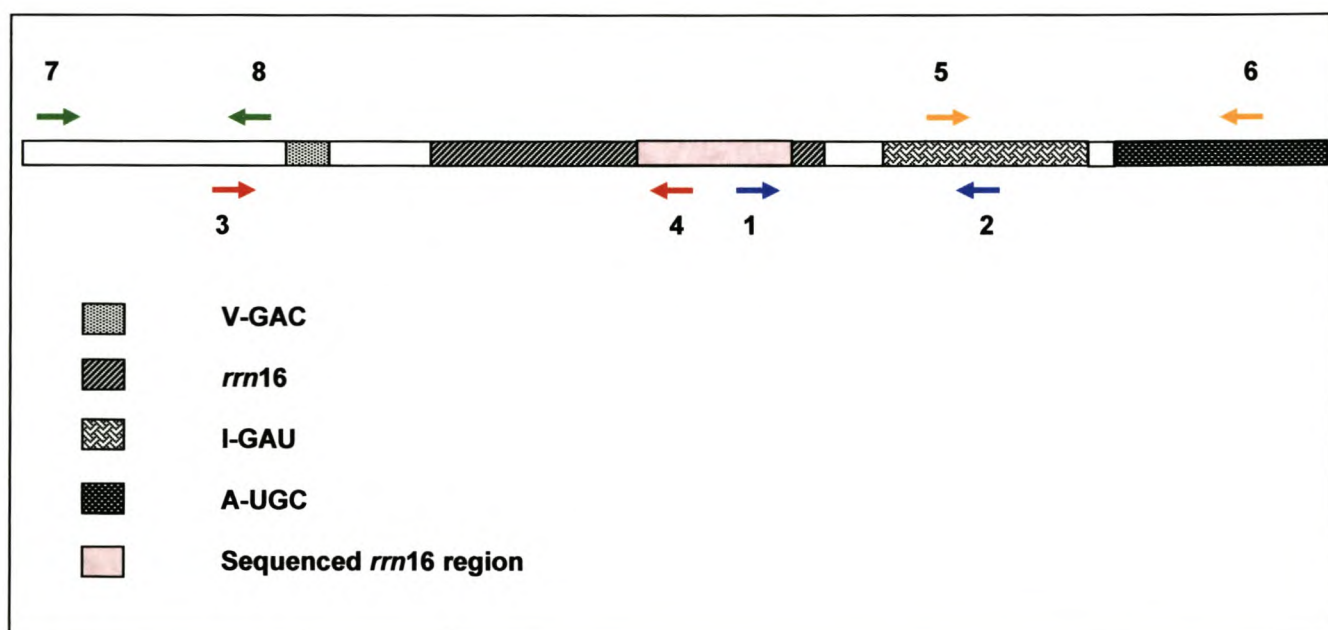
The PCR products were run on a 1% agarose gel in 1X TBE buffer and the resulting band was excised from the gel and purified using a QIAGEN: QIAquick® Gel Extraction Kit. The protocol found in the kit manual was followed.

The excised band was weighed and placed in a microfuge tube. Three volumes of buffer QG were added for every 1 volume of agarose gel slice (100 mg ~ 100 µl). The sample was incubated at 50 °C for 10 minutes, and vortexed every 2-3 minutes during the incubation. A QIAquick spin column was placed in a collection tube and the completely dissolved gel slice and buffer were loaded onto the spin column. The sample was centrifuged (16 000 x g, 1 min, 22°C), the flow through discarded, and the spin column placed back in the collection tube. Seven hundred and fifty microlitres of Buffer PE was added to the spin column and centrifuged (16 000 x g, 1 min, 22°C). The flow through was discarded; the spin column was placed back in the collection tube and centrifuged (16 000 x g, 1 min, 22°C). The spin column was placed in a sterile 1.5 ml microfuge tube and 50 µl of ddH<sub>2</sub>O was added. The sample was centrifuged (16 000 x g, 1 min, 22°C) and the DNA was sequenced at the Central Analytical Facility as described in 3.1.8.

The *rrn16* flanking region sequences were used to design new outward-facing primers with degenerate reverse primers being designed using completely sequenced chloroplast genomes as a template. These primers were used in a new Expand High Fidelity PCR reaction in order to sequence the next flanking region. See table 3.2 for a complete list of primers and figure 3.1 for the relative binding positions of each of these primers on the grapevine chloroplast genome.

**Table 3.2.** The *rrn16*-flanking region primers: The names and sequences of the primers used to amplify the *rrn16* flanking regions are indicated.

1	16S rRNA right forward	TCG CTA GTA ATC GCC GGT CAG C
2	<i>trnI</i> right reverse	CTG TGA AGA TYC GTT GTT AGG TGC TCC
3	<i>trnV</i> left forward	GGT GGA AGT CAT CAG TTC GAG CC
4	16S rRNA left reverse	AAC CAC ATG CTC CAC CGC TTG T
5	3'- <i>trnI</i> forward	GAT GGG GCG ATT CAG GTG AG
6	3'- <i>trnI</i> reverse	AGA GCC GCC GAC TCC AAC TA
7	5'- <i>trnV</i> forward	TGT TGG CAC CAG TCC TAC AT
8	5'- <i>trnV</i> reverse	CCA CGA GCC TCT TAT CCA CGA



**Figure 3.1** The relative binding positions of each of the 8 primers designed. Primers 1 and 4 were designed using the already sequenced *rrn16* region. The grapevine chloroplast gene regions amplified by the primer pairs are also shown.

## 3.2 BAC CLONE CHARACTERISATION AND SEQUENCING

Midway through the project BAC clones containing *V. vinifera* cv. L. Cabernet Sauvignon chloroplast genome inserts became available. One of these clones appeared to contain most of the grapevine chloroplast genome and was used for further library construction, sequencing and mapping.

### 3.2.1 BAC Clone Storage

Twenty-four pIndigoBAC-5 (*Hind*III-ready) clones transformed into the *Escherichia coli* strain DH10B were obtained from Dr. Anne-Francoise Adam-Blondon (Plant Genomic Research Unit/Unité de Recherche en Génomique Végétale (URGV), France). Each clone contained *V. vinifera* cv. L. Cabernet Sauvignon chloroplast genome fragments. BAC clones were grown overnight at 37°C, with shaking, in 3 ml FM 6% glycerol chloramphenicol (2.5% w/v granulated LB broth; 13 mM KH<sub>2</sub>PO<sub>4</sub>; 36 mM K<sub>2</sub>HPO<sub>4</sub>; 1.7 mM sodium citrate; 6.8 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 6% w/v glycerol; 12.5 µg/ml chloramphenicol; 0.4 mM MgSO<sub>4</sub>). Seven hundred microlitres of culture was added to 700 µl 30% glycerol and stored at -80°C. (See appendix 4 for pIndigoBAC-5 plasmid map.)

### 3.2.2 BAC-End Analysis

The end-sequences of 7 of the BAC clones received were obtained from Dr. Adam-Blondon. The BLAST algorithm was used to compare these sequences to those found in the NCBI database. Homologous chloroplast genome sequences were used to determine the location of each BAC-end on the grapevine chloroplast genome. The BAC clone 001A01 appeared to cover the entire chloroplast genome and was therefore chosen for further analysis.

### 3.2.3 Isolation of BAC Clone DNA

A single BAC colony was inoculated into 5 ml of LB medium containing chloramphenicol (12 µg/ml) and incubated for ~8 h at 37°C with vigorous shaking (~300rpm). The starter culture was diluted 1/500 in 500ml of LB medium containing chloramphenicol (12 µg/ml). The BAC clone DNA was isolated using the QIAGEN<sup>®</sup> Large-construct Kit, according to the handbook protocol accompanying the kit.



The 500 ml culture was pelleted by centrifugation (6000 x g, 15 min, 4 °C). The pellet was resuspended in 20 ml Buffer P1 (50 mM Tris-Cl, pH 8.0; 10 mM EDTA; 100 µg/ml RNase A). Twenty millilitres of Buffer P2 (200 mM NaOH; 1% SDS) was added, the sample was inverted 4-6 times, and incubated for 5 minutes at room temperature. Twenty millilitres of chilled Buffer P3 (3.0 M potassium acetate, pH 5.5) was added, the sample inverted 4-6 times, and incubated on ice for 10 minutes. The sample was centrifuged ( $\geq 20\,000$  x g, 30 min, 4°C) and the supernatant was removed and filtered through a pre-wetted filter. Room temperature isopropanol (0.6 volumes) was added to the cleared lysate, mixed, and centrifuged ( $\geq 15\,000$  x g, 30 min, 4°C). The supernatant was discarded. Five millilitres of room temperature 70% ethanol was added to the pellet and centrifuged ( $\geq 15\,000$  x g, 15 min, 22°C). The supernatant was discarded; the pellet was air-dried and dissolved in 9.5 ml Buffer EX. Two hundred microlitres of ATP-Dependent Exonuclease and 300 µl of 100 mM ATP solution were added to the DNA solution. The sample was mixed gently but thoroughly and incubated in a 37°C water bath for 60 minutes. A QIAGEN-tip 500 was equilibrated by applying 10 ml Buffer QBT (750 mM NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol; 0.15% Triton<sup>®</sup>, X-100) and allowing the column to empty by gravity flow. Ten millilitres of Buffer QS was added to the digested DNA sample, the sample was applied to the QIAGEN-tip and allowed to enter the resin by gravity flow. The QIAGEN-tip was washed with 2 x 30 ml Buffer QC (1.0 M NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol). The DNA was eluted with 15 ml Buffer QF (1.25 M NaCl; 50 mM Tris-Cl, pH 8.5; 15% isopropanol), pre-warmed to 65°C. The DNA was precipitated by adding 0.7 volumes of room temperature isopropanol and centrifuged ( $\geq 15\,000$  x g, 30 min, 4°C). The supernatant was discarded, the pellet was washed in 5 ml of room temperature 70% ethanol and centrifuged ( $\geq 15\,000$  x g, 15 min, 22°C). The supernatant was discarded and the pellet air-dried. The DNA was dissolved in a suitable volume of TE buffer (100 – 200 µl), pH 8.0. BAC DNA concentration was estimated by electrophoresis of the BAC DNA in parallel with lambda standards. The BAC DNA was stored at -20°C until further use.

### 3.2.4 Restriction Digestion of BAC Clone DNA

Approximately 80 ng of the isolated 001A01 BAC DNA was digested with *NotI* and a 6bp cutter (*SalI*, *PstI*, *XhoI* and *EcoRI*, respectively). A digestion of 001A01 with only *NotI*

was also performed. All digestions were performed according to the manufacturer's instructions (Roche) and with the appropriate buffer. The digestions were incubated at 37 °C for more than 3 h.

The digestions were fractionated on a 0.7% (w/v) 1X TAE agarose gel for 16 to 20 h at 20 V. The gel was stained in 1X TAE containing ethidium bromide (0.5 µg/ml) and visualized under UV light.

### 3.2.5 Library Construction

The 001A01 BAC clone DNA was digested with *NotI* in combination with *BamHI*, *Sall*, *PstI*, and *Sall/PstI*, respectively. All restriction digestions were performed according to the manufacturer's instructions (Roche), in the appropriate buffer. The vector, pUCBM21 or pBluescript SK (stratagene – appendix 4), was digested with the same enzyme(s) as the BAC clone. Vectors were, however, not digested with *NotI*. Vector DNA digested with a single enzyme was SAP-treated before ligation (see 2.1.4.1 for method). Ligations were performed according to the manufacturer's instructions (New England Biolabs) and transformed into competent DH5α cells as described in 3.1.4.3. Refer to table 3.3 for the restriction enzymes and vectors used, as well as the vector to insert ratios.

**Table 3.3:** BAC DNA subcloning: The restriction enzymes used to digest the BAC clone DNA are listed in the first column. The vector used during each ligation as well as the ratio of vector to insert in each ligation are also shown.

Restriction enzyme	Vector	Insert :Vector ratio
<i>BamHI</i>	pUCBM21	1:1, 1:5
<i>PstI/Sall</i>	pUCBM21	1:6, 1:2
<i>Sall</i>	pUCBM21	1:1, 1:3, 3:1
<i>PstI/Sall</i>	pBluescript SK	1:1, 1:2, 1:3, 3:1

### 3.2.6 Colony Screening

White colonies were picked and inoculated into 5 ml of LB containing ampicillin (100 µg/ml). Liquid cultures were incubated at 37°C for 12 to 16 h, with shaking (~225 rpm). A variation of a standard alkaline lysis protocol (Sambrook et al. 1989) was used to isolate the plasmid DNA.

One and a half millilitres of the 5 ml culture was pelleted (1 min, 16 000 x g, 22°C). The supernatant was discarded and the pellet was resuspended in 100 µl of cell resuspension solution (50 mM glucose; 25 mM Tris-Cl, pH 8.0; 10 mM EDTA, pH 8.0). A further 200 µl of cell lysis solution (0.2 N NaOH, 1% SDS) was added. The sample was inverted 4 times and incubated for 5 minutes at room temperature. One hundred and fifty microlitres of chilled potassium acetate solution was added, the sample was inverted 4 times and incubated for 5 minutes on ice. The sample was centrifuged (15 min, 16 000 x g, 22°C) and the supernatant transferred to a new microfuge tube. Two volumes of 100% ethanol were added to the supernatant and incubated on ice for 10 minutes. The sample was centrifuged (30 min, 16 000 x g, 22°C) and the supernatant discarded. Two hundred microlitres of 70% ethanol was added to the pellet and centrifuged (5 min, 16 000 x g, 22°C). The supernatant was discarded; the pellet was air-dried and resuspended in 30 to 50 µl of ddH<sub>2</sub>O. RNase A (30 µg/ml) was added to the dissolved DNA and it was stored at -20 °C.

The plasmid insert size was determined by restriction digestion of the isolated plasmid DNA with the appropriate restriction enzyme(s) followed by the visualization of the inserts on 0.8% (w/v) 1X TBE (0.090 M Tris-borate, 0.002 M EDTA) agarose gels.

### **3.2.7 Sub-Clone Sequencing and Sequence Analysis**

A single colony of each potentially positive clone was picked and inoculated into 5 ml of LB containing ampicillin (100 µg/ml). The culture was incubated at 37°C for 12 to 16 h, with shaking (~225 rpm). DNA was isolated using a Wizard<sup>®</sup> Plus SV Minipreps DNA Purification System (Promega), as described in 2.1.8, and handed in for sequencing at the Central Analytical Facility. Sequencing was performed on an ABI Prism 3100 Genetic Analyzer using M13 or T7/T3 primers. The nBLAST 2 program (Altschul et al. 1997) was used to compare sequences to the non-redundant NCBI database.

## **4. RESULTS AND DISCUSSION**

### **4.1 ISOLATION AND CHARACTERISATION OF VITIS VINIFERA CHLOROPLAST DNA**

#### **4.1.1 Chloroplast DNA Isolation**

A number of protocols were followed to isolate chloroplast DNA (cpDNA). These include the isolation methods of Bovenberg et al. (1981), Bookjans (1984), Hirai et al. (1985), Palmer (1988), Edwards et al. (1991), Kim et al. (1997), Mourad (1998), Triboush et al. (1998), Mariac et al. (2000) and Mpoloka (2001). A variation on the method used by Yeoh and Joseph (1995), described in section 3.1.2, proved to be the most successful. This method gave high yields of cpDNA (> 1 µg) that was of suitable quality for restriction digestion and subsequent cloning. However, limited amounts of nuclear DNA contamination and/or sheared cpDNA were still present at the end of the extraction procedure.

Leaves from two different grapevine cultivars were used for cpDNA extractions. Sultana leaves were obtained from the ARC Infruitec-Nietvoorbij in season. Out of season, our only source of leaf material was the Sogra 1 cultivar grown in the glasshouse at the Institute of Wine Biotechnology (IWBT), University of Stellenbosch.

Extracting DNA from grapevine plants was particularly difficult due to the presence of high levels of secondary metabolites. These contaminants include tannins as well as polyphenolics that cause DNA degradation and a viscous DNA pellet. These impurities also impair subsequent analysis of the DNA (Lodhi et al. 1994, Porebski et al. 1997, Kim et al. 1997).

The quality of the starting leaf material played a critical role in determining the quality and quantity of the resulting cpDNA. The best results were obtained when fresh, young, healthy leaves obtained early in the season or from a glasshouse were used. After collection, leaves were stored in the dark for 48 h to allow the chloroplasts to deplete their stored starch grains.

All extraction steps were performed as swiftly as possible at 0 to 4 °C to inhibit the action of hydrolytic enzymes and reduce chloroplast membrane shearing. The ingredients of the extraction buffer had a considerable influence on the success of the isolation. The

incorporation of PVP-10 (5% w/v) and a high molarity of NaCl (1.75 M) resulted in the isolation of higher quantities of more pure grapevine cpDNA. PVP-10 helped to absorb tannins and other secondary plant compounds. Sodium chloride served as an osmoticum, minimising nuclear DNA contamination and removing polysaccharides. A 10:1 buffer to leaf ratio was used to dilute and separate contaminants from the DNA. High levels of 2-mercaptoethanol were also added to prevent phenolic oxidation (Bookjans et al. 1984, Palmer 1988, Schuler and Zielinski 1989, Lodhi 1994, Milligan 1992, Porebski et al. 1997).

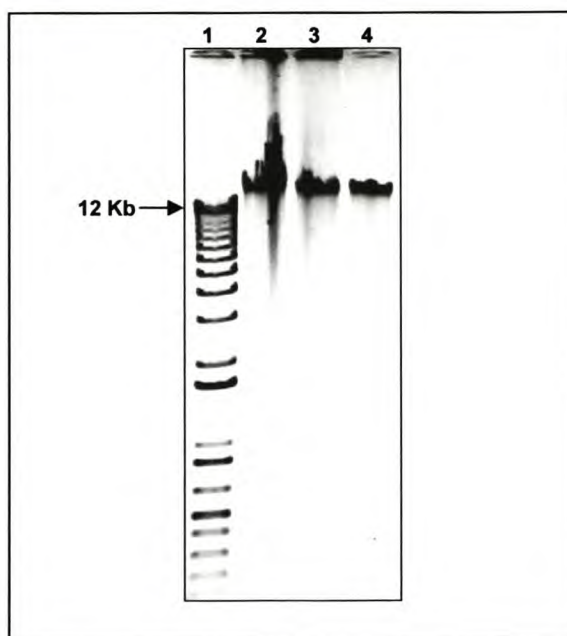
Special care was taken during the extraction process to disrupt only the cell walls and not the chloroplast membranes. For this reason a Waring blender was used for leaf maceration (Palmer 1988, Schuler and Zielinski 1989). The homogenate was filtered through two layers of cheesecloth and then through two layers of Miracloth to remove large pieces of tissue and cells (Mariac et al. 2001).

Three methods were followed in an attempt to eliminate nuclear DNA. The first method involved the treatment of the homogenate with Dnase I (Herrmann et al. 1975, Triboush et al. 1998). However, this method gave very low yields of degraded DNA. Grapevine chloroplasts appear to be quite fragile and, therefore, harsh handling of the chloroplasts before DNase treatment could have disrupted the chloroplast membranes and allowed the cpDNA to be digested (Mourad 1998). The second method involved the use of sucrose gradients (Hirai et al. 1985, Mourad 1998, Palmer 1988, Mpoloka 2001) in which the intact chloroplasts banded separately from other cellular components according to their buoyant densities. However, since only a limited amount of the homogenate could be loaded onto a single sucrose gradient and relatively low yields were obtained, this procedure became very time-consuming when attempting to extract large amounts of cpDNA. The method that proved most successful involved two low-speed centrifugation steps to pellet the chloroplasts, and the resuspension of this pellet in the extraction buffer (Yeoh and Joseph 1995).

Different methods of lysing the chloroplast membrane were also attempted. A high degree of success was experienced when using CTAB for this purpose as opposed to SDS, sodium sarkosinate and/or proteinase K (Bookjans et al. 1984, Mariac et al. 2000, Mpoloka 2001). A possible explanation for this is that SDS used alone, or with sodium sarkosinate, could interfere with organic extractions performed afterwards (Gounaris et al. 1986). Organic extractions (phenol/chloroform/isoamyl alcohol) removed proteins still present and resulted in a much cleaner DNA pellet. CsCl density gradients were also performed in an attempt to

remove contaminating proteins and RNA (Palmer 1988), but this proved to be a time-consuming and costly process, giving low yields of DNA of poor quality.

Genomic DNA contamination has proven to be a common problem in cpDNA isolations (Schuler and Zielinski, 1989). The optimised protocol for *V. vinifera* cpDNA isolation (section 3.1.2) gave high yields of relatively pure cpDNA that could be digested with restriction endonucleases and contained minimal amounts of genomic DNA (figure 4.1).



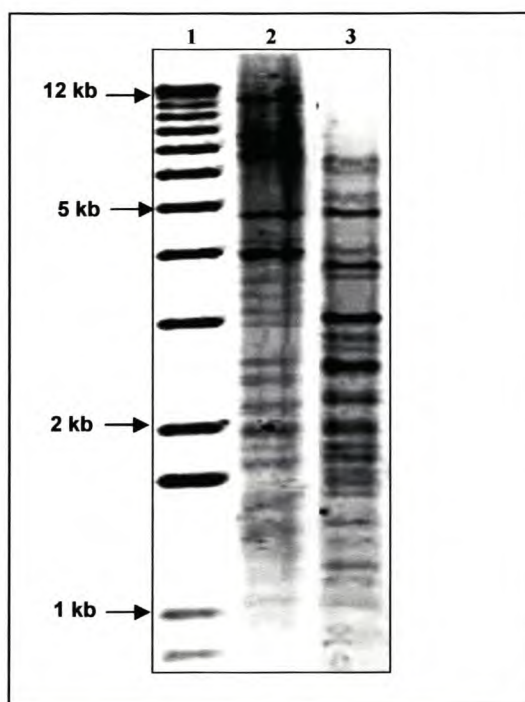
**Figure 4.1** Chloroplast DNA isolated from *Vitis vinifera* cv. L Sugra one. Lane 1: 1Kb<sup>+</sup> molecular marker (Promega) Lane 2: Two microliters of cpDNA isolated according to the method of Mariac et al. (2000). Lane 3: Two microliters of cpDNA isolated according to Mpoloka (2001). Lane 4: Five microliters of cpDNA isolated according to variation on the method of Yeoh and Joseph (1995 (section 3.1.2).

### 4.1.2 Restriction Digestion

The isolated cpDNA was digested with a number of restriction enzymes to determine both the quality of the cpDNA and which enzymes would give fragments of sizes optimal for further cloning. Although difficulties in digesting the cpDNA were encountered at first, cpDNA digested readily with restriction endonucleases once the isolation procedure had been optimised and inhibitors such as tannins and polysaccharides had been removed. The cpDNA isolated was therefore shown to be of suitable quality for further cloning and sequencing.

In figure 4.2 cpDNA isolated according to the optimised protocol (section 3.1.2) was digested with *EcoRI* and run in lane 2. The cpDNA bands are clearly visible, but there is a background

smear of genomic DNA. Digestion with *Bam*HI and *Pst*I yielded comparable results (not shown). This indicates that the grapevine genomic DNA had not been completely removed during the cpDNA isolation process. Spinach cpDNA was used as a positive control and run in lane 3 (figure 4.2). The spinach cpDNA digested more efficiently with less genomic DNA contamination, even when a simple extraction protocol was used. The reason for this is that spinach leaves contain almost no polyphenols or other secondary metabolites that could interfere with both cpDNA extractions and subsequent restriction digestion (Schmitz-Linneweber 2001).



**Figure 4.2** A typical chloroplast DNA restriction digestion: 1: 1Kb<sup>+</sup> molecular marker. 2: *V. vinifera* cv L. Sogra one cpDNA digested with *Eco*RI. 3: *S. oleracea* cpDNA digested with *Eco*RI.

### 4.1.3 Library Construction and Library Screening using PCR

The restriction enzymes *Pst*I, *Eco*RI and *Bam*HI digested the cpDNA into fragments of an optimal size for cloning (1-10kb). *Pst*I digested the cpDNA into large fragments, while *Eco*RI and *Bam*HI produced smaller-sized fragments. Ligations and transformations were carried out as described in 3.1.4. White colonies obtained from ligation 1, 2 and 3 were screened by means of PCR reactions using vector-specific M13 primers (table 4.1).

Ligation 1, 2 and 3 (table 4.1) had very low transformation efficiencies, with a low percentage of white colonies testing positive during PCR. A possible reason for this could be the

presence of impurities in the Sultana cpDNA used in these ligation reactions. This cpDNA was not isolated according to the optimised extraction protocol and may have contained inhibiting elements such as polyphenols.

The highest number of white colonies was obtained in ligation 4 (table 4.1). The quality of the Sugra 1 cpDNA used in ligation 4 could have played a role in the high ligation efficiency obtained. Leaves acquired from the IWBT glasshouse were used to extract the cpDNA according to the optimised isolation protocol (section 3.1.2). The 604 colonies obtained in ligation 4 (table 4.1) were transferred onto new LB plates in a grid pattern, with 48 colonies per plate (6 X 8). These clones were maintained by replica plating them onto fresh plates containing selective media once a month.

**Table 4.1** Ligation efficiency: The number of white colonies obtained for each ligation are indicated, as well as the number of fragments that amplified during the PCR reaction performed on each colony.

Ligation no.	Restriction enzyme	Vector:insert ratio	No. of white colonies	No. amplified in PCR
1	<i>EcoRI</i>	1:6	15	5
2	<i>BamHI</i>	1:2	24	1
3	<i>PstI</i>	1:12	17	6
4	<i>EcoRI</i>	1:2, 1:4	604	No PCR performed

#### 4.1.4 Library Screening by Colony Blotting

Colony blotting was chosen as a method of screening the *EcoRI* cpDNA library (ligation 4) for overlapping clones. It was believed that this would be a more efficient method of screening the large number of colonies obtained in ligation 4 than PCR. Two *PstI*-digested clones that had been sequenced and found to be of chloroplast origin (P3 and P4 - ligation 3, table 4.1) were used as probes to find overlapping clones. The positive *EcoRI* clones could then be used as probes to screen a *PstI* library. The new positive *PstI* clones would then be used to screen the *EcoRI* library again, and in this way we could walk outwards from the first 2 *PstI* clone.

DNA was transferred and fixed to nylon membranes by means of the colony lifting procedure. The arrangement of colonies in a grid pattern facilitated this procedure, allowing 48 colonies to be replica plated at a time. According to the protocol described in the "DIG Application Manual for Filter Hybridisation" (Roche) colonies should be transferred from an LB plate



onto a nylon membrane by placing the membrane face down onto the colonies and then lifting it off, making sure that the colonies adhered to the membrane. We experienced problems when following this method. Some colonies were only partially transferred and some were not transferred at all. Replica plating the bacterial colonies directly onto the membrane and incubating them overnight resulted in an efficient amount of bacterial DNA being transferred onto the membrane (Shelly Deane, Pers. comm., US Microbiology Dept.).

The optimum length of time for UV fixation was determined by exposing a positive control colony to UV light for different lengths of time, ranging from 1 to 5 minutes. Four minutes proved to be the optimal time for UV fixation. Once DNA had been fixed to the membrane all colony debris was removed by washing the membranes in 3 x SSC and 0.1 % SDS (w/v). This was carried out to prevent cell debris from inhibiting the hybridisation of the probe to the positive clone DNA fixed to the membrane.

The efficiency of probe labelling with DIG High Prime was tested by placing a dilution series of the labelled DNA onto a nylon membrane, ranging from 0.01 pg/ $\mu$ l to 1 ng/ $\mu$ l. Labelled control DNA (Roche) was also placed on the membrane using the same dilution series. The signal intensity of the labelled probe was comparable to that of the control DNA, demonstrating the efficiency of probe labelling.

After hybridisation with the *Pst*I probes, only faint signals were observed on the autoradiography film and no conclusions could be drawn from these results. Attempts to optimise the hybridisation reaction by lowering the hybridisation temperature and adjusting the temperature and salt concentration used in the high stringency wash resulted in minor improvements. However, conclusive results were not obtained using colony blotting. This method of colony screening was not optimised further as a result of the availability of an alternative method of constructing a grapevine chloroplast genome library. This method used BAC clones containing cv L. Cabernet Sauvignon chloroplast DNA inserts (see section 4.2).

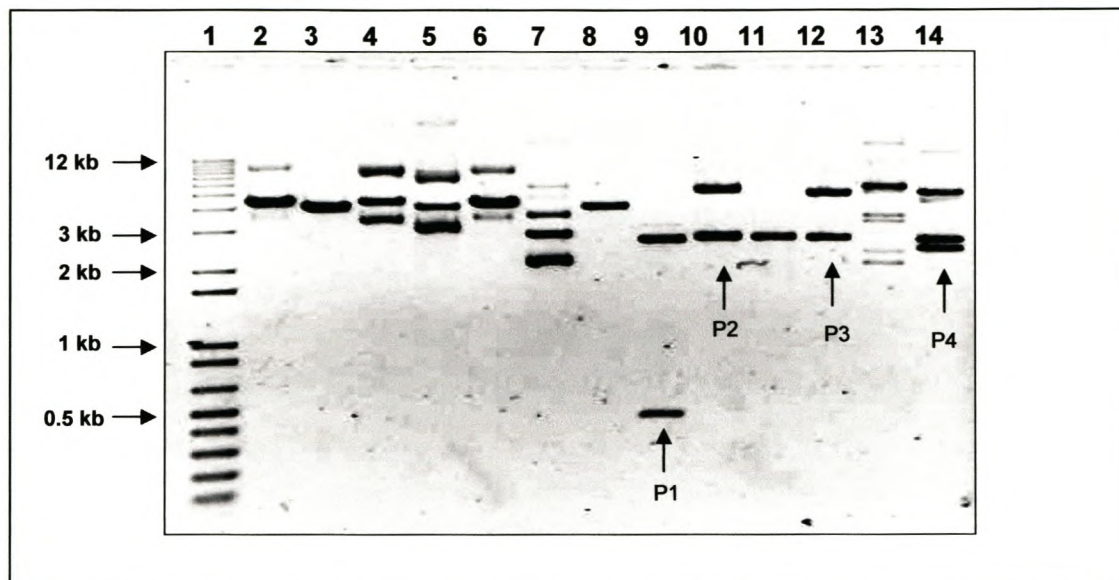
#### **4.1.5 Clone DNA Isolation and Sequencing**

Clones selected for further sequencing were isolated with the Wizard<sup>®</sup> Plus SV Minipreps DNA Purification System (Promega). The isolated DNA was digested with the appropriate restriction enzyme and electrophoresed on an agarose gel to determine the insert sizes. The results of these digests are shown in figure 4.3 and figure 4.4.

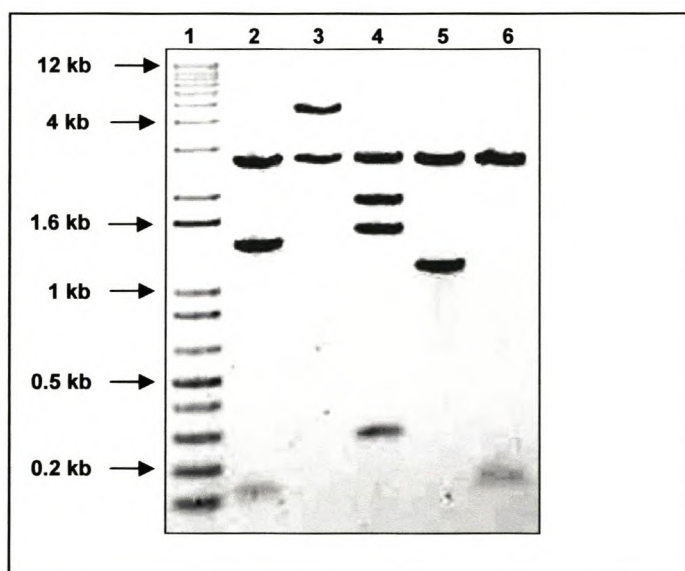
The ligation 1 and ligation 2 clones (table 4.1) gave unexpected banding patterns when digested with the appropriate restriction enzyme, with no vector fragment being visible (figure 4.3, lane 2-8). These results were reproducible and could not be explained. The clones were therefore not analysed further. Clone 1, 2, 4 and 6 of the *Pst*I ligation (P 1, 2, 3, 4) contained insert DNA (figure 4.3, lane 9, 10, 12 and 14 respectively). Clone 3 did not contain an insert and clone 5 gave unexpected banding patterns (figure 4.3, lane 11 and 13 respectively). Of the 5 *Eco*RI clones randomly selected from the *Eco*RI library for sequencing (figure 4.4, lane 2-6), four contained insert DNA and only clone 5 did not. Clone 3 appeared to contain 2 insert fragments.

Four *Pst*I clones (P1, P2, P3, P4 – figure 4.3) and three *Eco*RI clones (E1, E3, E4 – figure 4.4) were sequenced using vector-specific M13 primers. The BLAST algorithm was used to determine sequence homologies between the sequences obtained and those found in the NCBI database (table 4.3, table 4.4). P1 and P2 showed no significant homology to any sequence in the NCBI database. Only a 40 bp fragment of the P2 reverse sequence showed 92% homology to the mitochondrial matK sequence of a number of species. P3 and the reverse sequence of P4 both had a high degree of homology to higher plant chloroplast sequences (sequences in addendum 1). The forward sequence of P4 was not of high enough quality to be used in a BLAST search. All 3 *Eco*RI clones sequenced were of chloroplast origin (sequences in addendum 1). E3 appeared to contain two inserts, with the forward sequence of the one insert and the reverse sequence of the other insert having been sequenced. The results of the BLAST search suggested that the insert was approximately 21 kb in size, which did not correspond with the fragment sizes observed on the gel photo (figure 4.4).

Three of the 4 clones obtained have been mapped against the *Atropa belladonna* chloroplast genome using BioEdit (Hall 1999) (figure 4.5). Clone E3 could not be placed on the map, because two inserts were contained in this clone and only one end of each insert has been sequenced. The forward sequence of clone P4 was not known, but the direction and size of the clone could be determined and P4 could be placed on the map. P3 and E4 are found in the inverted repeat region, while E1 is in the large single copy region. The gene regions where each of these clones are found are shown in table 4.2



**Figure 4.3** Restriction digests of possible positive clones. Lane 1: 1 Kb<sup>+</sup> molecular marker. Lane 2-7: *EcoRI* clones (ligation 1, table 1). Lane 8: *Bam*HI clone (ligation 2, table 1). Lane 9-14: *Pst*I clones (ligation 3, table 1).



**Figure 4.4** Five *EcoRI* library clones (ligation 4, table 4.1): Lane 1: 1Kb<sup>+</sup> molecular marker. Lane 2: *EcoRI* 1 (E1). Lane3: *EcoRI* clone 2 (E2). Lane4: *EcoRI* clone 3 (E3). Lane 5: *EcoRI* clone 4 (E4). Lane 6: *EcoRI* clone 5.

**Table 4.2** The gene locations of the sequenced clones.

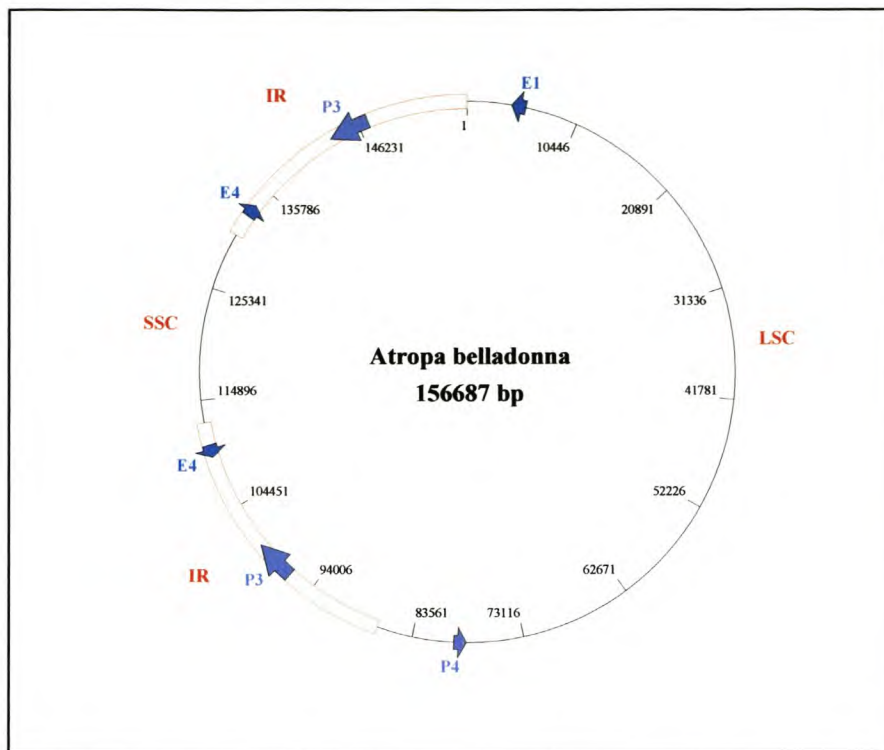
Sequenced Clone	Gene location
E1	<i>rps16 – trnK</i> (tRNA-Lys)
E3	<i>ycf1</i> (insert 1), <i>ycf2</i> (insert2)
E4	<i>rrn5</i> , tRNA-Asn, tRNA-Arg, <i>rrn4.5</i>
P3	<i>ycf2</i> , <i>ycf15</i> , <i>ndhB</i> , <i>rps7</i>
P4	<i>petB</i> , <i>petD</i>

**Table 4.3** The degree of homology found between cpDNA clones E1, E3 and E4 and the chloroplast genomes of *Spinacia oleracia*, *Nicotiana tabacum* and *Atropa belladonna*.

Clone	Primer	Plant species	% Homology	Clone region	E value
P3	M13 Forward	<i>S. oleracia</i>	94	37-778	0.0
		<i>N. tabacum</i>	96	157-778	0.0
		<i>A. belladonna</i>	96	157-778	0.0
	M13 Reverse	<i>S. oleracia</i>	96	1-559	0.0
			90	586-668	3e-14
			85	683-779	0.11
		<i>N. tabacum</i>	96	5-467	0.0
			88	496-710	2e-46
		<i>A. belladonna</i>	96	5-467	0.0
			90	496-710	8e-61
P4	M13 Reverse	<i>S. oleracia</i>	94	366-422	6e-15
			93	438-610	1e-62
		<i>N. tabacum</i>	91	367-422	1e-09
			94	438-610	2e-67
		<i>A. belladonna</i>	91	367-422	1e-09
			94	438-610	2e-67

**Table 4.4** The degree of homology found between cpDNA clones P3 and P4, and the chloroplast genomes of *Spinacia oleracia*, *Nicotiana tabacum* and *Atropa belladonna*.

Clone	Primer	Plant species	% Homology	Clone region	E value
E1	M13 Forward	<i>S. oleracia</i>	86	124-441	2e-68
		<i>N. tabacum</i>	95	138-181	2e-09
		<i>A. belladonna</i>	94	195-430	e-100
	M13 Reverse	<i>S. oleracia</i>	95	138-181	2e-09
		<i>N. tabacum</i>	94	195-441	e-102
		<i>A. belladonna</i>	89	18-183	8e-52
E3	M13 Forward	<i>S. oleracia</i>	89	273-311	0.11
		<i>N. tabacum</i>	92	335-376	7e-06
		<i>A. belladonna</i>	88	399-469	3e-11
	M13 Reverse	<i>S. oleracia</i>	80	495-686	2e-21
		<i>N. tabacum</i>	85	414-470	0.007
		<i>A. belladonna</i>	91	537-585	1e-07
E4	M13 Forward	<i>S. oleracia</i>	84	399-470	1e-04
		<i>N. tabacum</i>	91	537-585	1e-07
		<i>A. belladonna</i>	96	1-63	8e-21
	M13 Reverse	<i>S. oleracia</i>	93	220-688	0.0
		<i>N. tabacum</i>	94	1-690	0.0
		<i>A. belladonna</i>	94	1-690	0.0
E4	M13 Forward	<i>S. oleracia</i>	96	713-740	1.5
		<i>N. tabacum</i>	88	22-153	2e-34
		<i>A. belladonna</i>	93	260-490	8e-86
		<i>S. oleracia</i>	92	502-641	1e-47
		<i>N. tabacum</i>	94	680-765	2e-21
		<i>A. belladonna</i>	89	22-490	e-158
		<i>S. oleracia</i>	92	502-641	8e-46
		<i>N. tabacum</i>	94	678-765	1e-22
		<i>A. belladonna</i>	92	30-179	2e-49
		<i>S. oleracia</i>	93	229-490	e-105
		<i>N. tabacum</i>	91	502-641	2e-43
		<i>A. belladonna</i>	94	678-765	1e-22
	M13 Reverse	<i>S. oleracia</i>	95	6-184	5e-71
		<i>N. tabacum</i>	92	166-206	1.4
		<i>A. belladonna</i>	94	227-491	e-144
		<i>S. oleracia</i>	95	474-497	5.7
		<i>N. tabacum</i>	96	672-701	0.001
		<i>A. belladonna</i>	97	1-184	1e-83
		<i>S. oleracia</i>	95	166-210	2e-05
		<i>N. tabacum</i>	98	226-497	e-144
		<i>A. belladonna</i>	90	519-656	5e-40
		<i>S. oleracia</i>	100	672-701	6e-06
		<i>N. tabacum</i>	97	1-184	1e-83
		<i>A. belladonna</i>	95	166-210	2e-05
<i>S. oleracia</i>	98	226-497	e-144		
<i>N. tabacum</i>	89	519-656	1e-37		
<i>A. belladonna</i>	100	672-701	6e-06		



**Figure 4.5** Four *V. vinifera* cpDNA clones mapped against the *Atropa belladonna* chloroplast genome by using their forward and reverse sequences. The direction of each arrow indicates the orientation of the clone in the plasmid vector pUCBM21 (M13 forward to M13 reverse). P3 and P4 contain Sultana cpDNA inserts, while E4 contains a Suga I cpDNA insert. IR: Inverted repeat region. LSC: Large single copy region. SSC: Small single copy region.

#### 4.1.6 PCR of the *rrn16* Flanking Regions

At one point in the project leaf material was not available to allow further isolation of cpDNA. At this time, walking outwards from the *rrn16* gene region was chosen as an alternative method of chloroplast genome sequencing.

Enough cpDNA of high enough quality had previously been isolated to allow for a number of PCR reactions to be carried out. The designed primers were successfully used in an Expand High Fidelity PCR (table 3.2). Julia Robson performed the amplification of the region immediately flanking the left side of the *rrn16* region (using primers “16S rRNA left reverse” and “trnV left forward”). This region’s sequence was added to those that we sequenced. A total region of approximately 3320 bp was successfully amplified from the *V. vinifera* cv L. Sultana chloroplast genome (sequences in addendum 2). However, this method is expensive and time-consuming and was not continued once leaf material became available again.

## 4.2 BAC CLONE SEQUENCING

### 4.2.1 BAC-End Analysis

The BLAST algorithm was used to compare the BAC-end sequences received from Dr. Adam-Blondon (URGV) to the NCBI database. One clone, 001A01, appeared to cover almost the entire genome, except a region of approximately 500bp. This BAC clone was chosen for further sub-cloning and analysis.

### 4.2.2 Isolation of BAC Clone DNA

Four different protocols were followed in an attempt to isolate BAC clone DNA. Protocol 1 was obtained from Dr. Adam-Blondon (URGV) and was a variation on a standard alkaline lysis miniprep procedure. This method gave relatively pure BAC DNA, but only small culture volumes were used (5 ml/reaction) and therefore low concentrations of DNA were obtained (~60 ng/5 ml culture).

Protocol 2 (<http://bioprotocols.bio.com/protocolstools/protocol.jhtml?id=1338>) was similar to protocol 1, but used much larger culture volumes (250 to 500 ml). A higher yield of BAC DNA was therefore obtained, but a background smear of bacterial host-cell DNA was present. This BAC DNA was not used in further analyses. When using traditional methods of BAC DNA isolation, up to 30% of the resulting DNA can be genomic DNA contamination (QIAGEN<sup>®</sup> Large-Construct Kit handbook). Only one or two copies of the BAC plasmid are present per bacterial cell. Therefore, yields of BAC DNA are considerably lower than obtained when isolating plasmid DNA. Large culture volumes are required for sufficient BAC DNA to be obtained for further experimentation. However, the number of bacterial cells also increases with larger culture volumes and, thus, the amount of cell debris and host nuclear DNA present. The average *E. coli* genome is approximately 5 Mb (<http://www.ncbi.nlm.nih.gov/>), and therefore makes up a large percentage of the total DNA isolated.

Protocol 3 involved the purification of the BAC DNA according to the 'very low-copy plasmid purification protocol' found in the QIAGEN<sup>®</sup> Plasmid purification handbook. The QIAGEN<sup>®</sup> Plasmid Midi Prep Kit (100) was used to carry out the protocol. This method was attempted in the hope that the column would prevent the genomic DNA from being eluted.

However, the isolated DNA had a large background smear and a large percentage of bacterial host cell DNA was present. It is possible that the bacterial lysate could have been mixed too vigorously or that the column allowed sheared bacterial DNA to be eluted with the BAC DNA.

The QIAGEN<sup>®</sup> Large-Construct Kit (10) was used to carry out protocol 4, according to the protocol accompanying the kit. Both the protocol and the kit have been specifically adapted for the isolation of BAC, PAC, P1, and cosmid DNA without genomic DNA contamination. The protocol involved the digestion of the lysate with ATP-Dependent Exonuclease, prior to loading the DNA onto a column. The Exonuclease only digested linear DNA, such as genomic DNA and sheared or nicked BAC DNA, leaving the circular, supercoiled BAC DNA intact. This method gave a very low yield of ~200 ng of BAC DNA per 500ml culture. Although the kit protocol was followed precisely, genomic DNA was still present in the eluate. There are a few possible reasons for this contamination. The Exonuclease digestion may have been insufficient. Large fragments of genomic DNA, not digested efficiently with the exonuclease, may then have been eluted with the BAC DNA. Another reason could be that the ATP, not supplied with the kit but obtained independently, may not have worked efficiently.

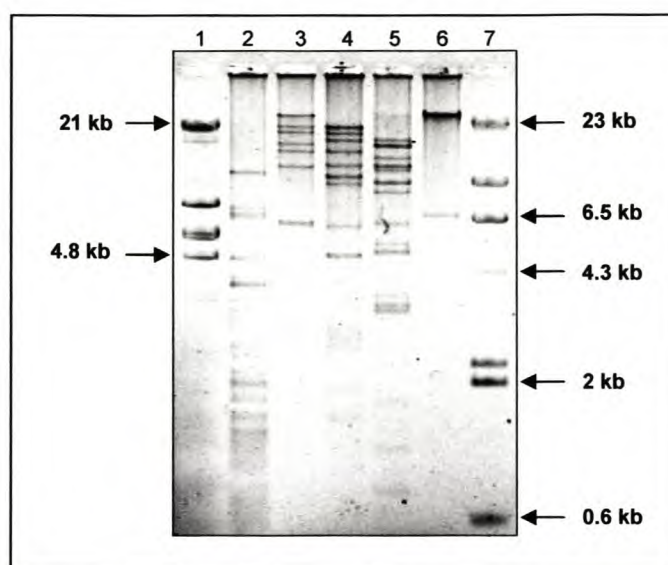
### 4.2.3 Restriction digestion

The isolated BAC clone DNA was efficiently digested with a variety of restriction enzymes. *NotI* was used to cut the insert out of the vector. The digested BAC DNA was run on a 0.7% agarose gel overnight to ensure that the large fragments separated properly (figure 4.5).

The photo of the digested 001A01 BAC DNA (figure 4.5, lane 2-6) was used to calculate the size of the chloroplast genome. The software programs Scion image (Scion corporation 2000) and dnafrag V3.03 (Schaffer and Sederoff 1981) were used to calculate the size of each fragment in lane 3, 4 and 5 (figure 4.5). The totals were added together and this value was divided by three to obtain an average value. The pIndigo-5 BAC vector size was subtracted from this to get a final size of ~115 Kb. However, this is only a minimum size value. Some fragments may actually be doublets, containing two identical IR fragments. Very large and very small fragments may also not be clearly visible on a 0.7% gel. The digested BAC clone DNA would need to be separated on both lower and higher percentage agarose gels to see these fragments clearly. The size of the entire BAC insert run in lane 6 (figure 4.5) was



calculated as being ~28 Kb, but this is inaccurate. The fragment in lane 6 runs higher on the gel than the largest molecular marker fragment, making it difficult to determine an accurate size value.



**Figure 4.5** Restriction digestion of BAC clone 001A01: Lane 1:  $\lambda$ /*EcoRI* molecular marker. Lane 2: 001A01 isolated with protocol 1 and digested with *NotI* and *EcoRI*. Lane 3: 001A01 digested with *NotI* and *SalI*. Lane 4: 001A01 digested with *NotI* and *PstI*. Lane 5: 001A01 digested with *NotI* and *XhoI*. Lane 6: 001A01 digested with *NotI*. Lane 7:  $\lambda$ /*HindIII* molecular marker.

#### 4.2.4 Subcloning and Colony Screening

Restriction digestion of the 001A01 BAC clone DNA with *BamHI*, *SalI*, and a combination of *PstI* and *SalI*, respectively, gave fragments of a suitable size for subcloning (1-10 kb). Digested BAC DNA was ligated into either the pUCBM21 or pBluescript SK plasmid vector. White colonies were screened by comparing clone insert sizes on agarose gels.

See table 3 for the results of the ligations performed. The limited success of the *BamHI* and *SalI* ligations (ligation 1 and 2, table 4.5) could have been a result of incomplete SAP treatment and the consequent re-ligation of the pUCBM21 vector onto itself. The double digestion of the vector with *PstI* and *SalI* eliminated this problem. However, the *PstI* and *SalI* restriction sites are only 11 bp apart on the pUCBM21 vector's multiple cloning site. The close proximity of these sites could have hindered the restriction digestion of the vector with both enzymes (Sambrook and Russel 2001) and could explain the limited success obtained for ligation 4 (table 4.5). The *PstI* and *SalI* restriction sites are 29 bp apart from each other on the plasmid vector pBluescript SK. This vector was therefore chosen for further *PstI/SalI* ligations (ligation 5- 7, table 4.5).

**Table 4.5** Ligation efficiency: The results of the ligation of restriction-digested BAC clone fragments into appropriate plasmid vectors, as well as the number of possible positive clones obtained after screening of the white colonies.

Ligation no.	BAC isolation protocol used	Restriction enzyme(s)	Vector	Insert:Vector ratio	No. of white colonies (total)	Possible positive clones
1	1	<i>Bam</i> HI	pUCBM21	1:1	28	5
2	1	<i>Sal</i> I	pUCBM21	1:1, 1:3, 3:1	6	1
3	1	<i>Pst</i> I/ <i>Sal</i> I	pUCBM21	1:6, 1:2	22	2
4	1	<i>Pst</i> I/ <i>Sal</i> I	pSK	1:1	16	9
5	1	<i>Pst</i> I/ <i>Sal</i> I	pSK	1:3, 3:1	200	5
6	3	<i>Pst</i> I/ <i>Sal</i> I	pSK	1:2	200	16
7	4	<i>Pst</i> I/ <i>Sal</i> I	pSK	1:1, 1:2, 1:3	>400	13

All 56 colonies obtained from ligation 1, 2 and 3, and approximately 360 colonies obtained from ligations 5, 6 and 7 (table 4.5), were screened. The plasmid DNA was isolated from each colony and digested with the restriction enzyme(s) used for the ligation reaction. The digested DNA was electrophoresed on agarose gels and the insert sizes were determined by using a molecular marker ( $\lambda$ /*Sty*I). Fragments of similar size were digested with a range of restriction enzymes (e.g. *Eco*RI, *Bam*HI and *Sty*I) and separated on an agarose gel to determine the banding patterns produced. Since the entire BAC clone was included in the ligation, some clones may have contained BAC vector insert. To determine the expected sizes of these fragments, the pIndigoBAC-5 vector was virtually digested with the appropriate enzyme(s) by using the pDRAW32 1.0 (Olesen 1998) software program. Cloned fragments of the same size as the expected BAC vector fragments were eliminated. The clones that appeared to be novel after being subjected to the above-mentioned tests were selected for sequencing.

#### 4.2.5 Sequencing and sequence analysis

Selected clones were isolated with the Wizard<sup>®</sup> Plus SV Minipreps DNA Purification System (Promega) and sequenced. The BLAST algorithm was used to compare sequences to those found in the NCBI database in order to find homologous chloroplast sequences (table 4.7).

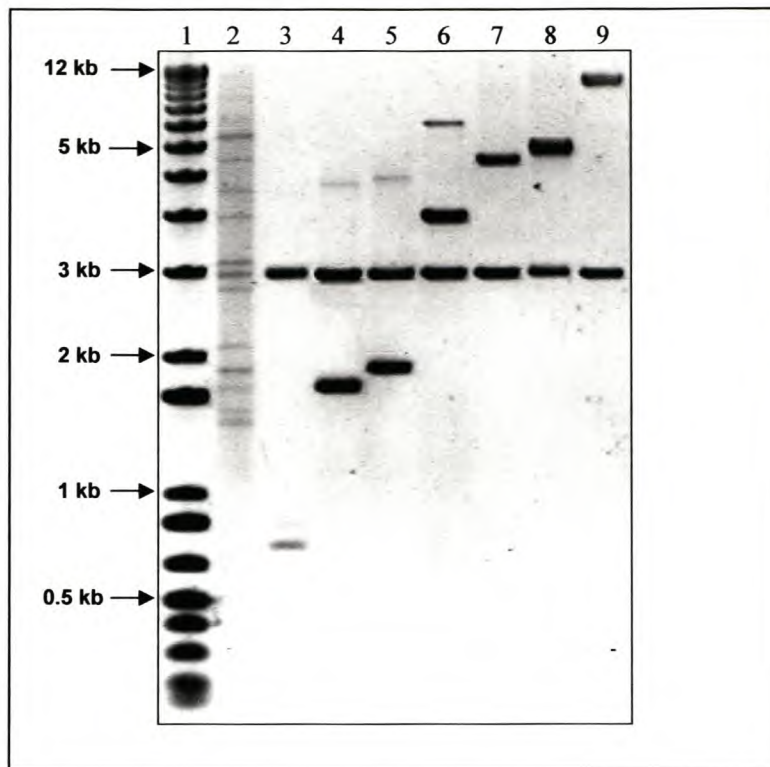
The 3 clones selected from ligation 1 were all positive, but 2 contained the same fragment (table 4.6). Partial digestion of the clones could have lead to them appearing different during gel photo analysis. The ligation 2 clone contained a fragment of the BAC vector (table 4.6).

One of the ligation 3 clones was of chloroplast origin, but the second was of BAC vector origin (table 4.6).

Of the nine ligation 4 clones (table 4.6), two were of chloroplast origin (PS1, PS2). However, one of these contained the same insert as the positive ligation 3 clone. Bacterial DNA contamination was detected for the first time in ligation 4. The two BAC vector fragments that would be produced during a *PstI/SalI*-digest had both been cloned and sequenced in ligation 4, and could be used for subsequent restriction fragment analysis to minimise the chance of sequencing another BAC vector fragment. Genomic DNA contamination was found in three of the five ligation 5 clones (table 4.6). The remaining two were of chloroplast origin (PS3, PS4). BAC isolation protocol 3 was followed in an attempt to eliminate the genomic DNA contamination. However, this method was unsuccessful, since all 16 clones sequenced contained bacterial genomic DNA. Ligation 7 (table 4.6) gave seven clones of chloroplast origin, of which three sets of two were identical and one contained an insert that had already been obtained in ligation 5 (PS5, PS6, PS7). Four of the 13 sequenced clones contained bacterial DNA inserts. This was a vast improvement over the results obtained from the third protocol, but 31% (4/13) of the clones still contained bacterial DNA contamination. In figure 4.6 the 001A01 BAC clone digested with *PstI*, *SalI* and *NotI* was run in lane 2. A background smear was visible in the digested BAC DNA, but this could be either genomic DNA contamination or other impurities. Clone PS1 to PS7 were run in lane 3 to 9 (figure 4.6).

**Table 4.6** Sequencing results

<b>Ligation no.</b>	<b>Possible positive clones</b>	<b>Actual positive clones (Name)</b>	<b>BAC vector</b>	<b>Bacterial DNA clones</b>
1	3	3 (B1, B2, B3)	0	0
2	1	0	1	0
3	2	1	1	0
4	9	2 (PS 1, PS 2)	4	4
5	5	2 (PS 3, PS 4)	0	3
6	16	0	0	16
7	13	7 (PS 5, PS 6, PS 7)	2	4



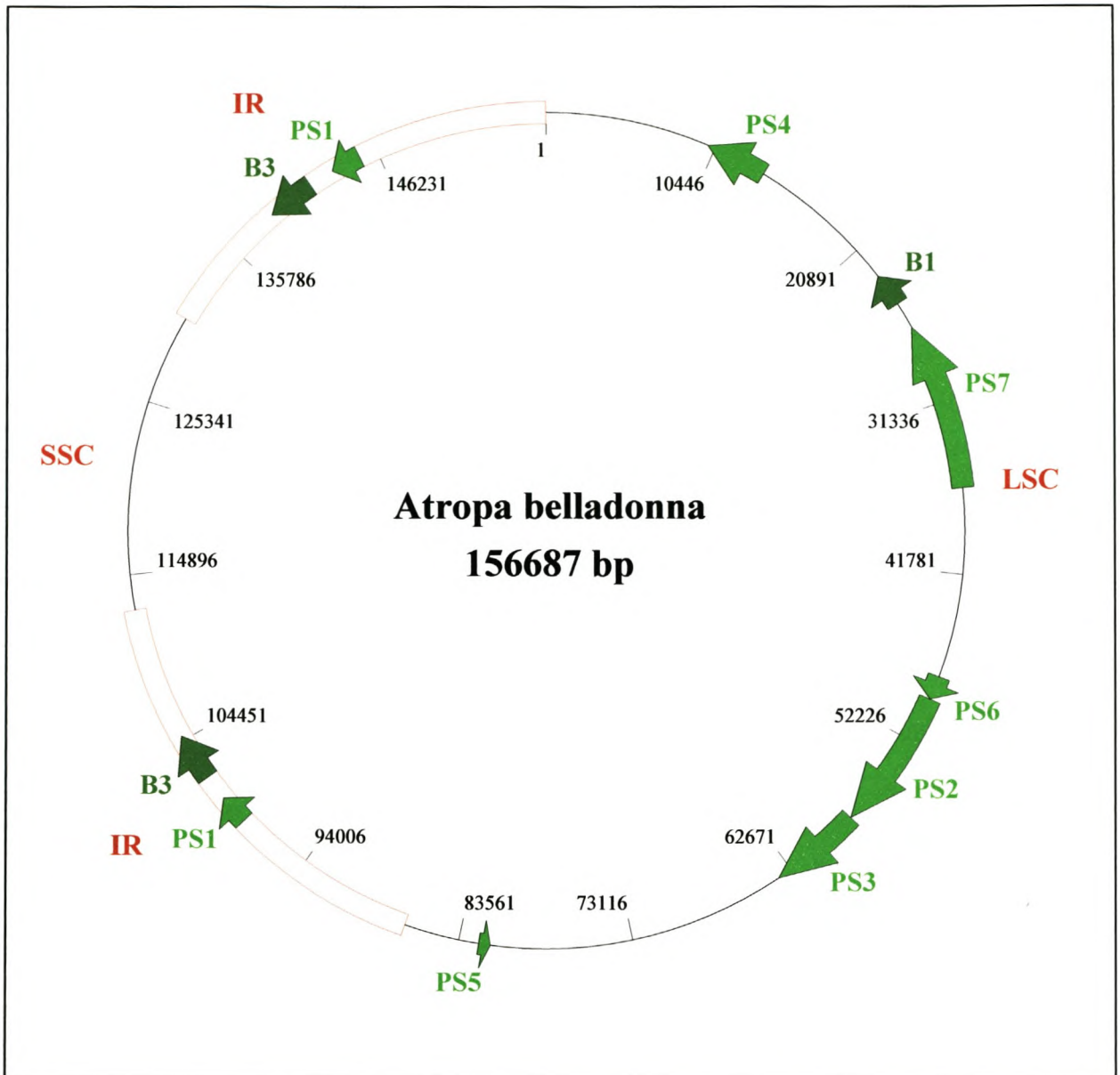
**Figure 4.6** A 0.7% agarose gel containing digested 001A01 DNA and *V. vinifera* chloroplast DNA sub-clones. Lane 1: 1Kb<sup>+</sup> molecular marker. Lane 2: 001A01 DNA isolated by means of protocol 4 and digested with *Pst*I, *Sal*I and *Not*I. Lane 3: Clone PS5. Lane 4: Clone PS6. Lane 5: Clone PS1. Lane 6: Clone PS4. Lane 7: Clone PS3. Lane 8: Clone PS7. Lane 9: Clone PS2.

Figure 4.7 indicates the mapped positions of the 001A01 BAC sub-clones against the *A. belladonna* chloroplast genome. Clone B3 was not included on the map, because it contained two inserts and only one end of each insert was sequenced. Seven of the clones were found in the LSC region, with PS3, PS2 and PS6 lying next to each other. Two clones mapped to the IR region, but no fragments from the SSC region were successfully cloned. The gene regions covered by each of the clones on the map are given in table 4.8.

The location of all the clones sequenced in this project, except for B3 and E3, are mapped against the *Atropa belladonna* chloroplast genome. The clones are well distributed across the IR and LSC regions, with only the SSC region not having been cloned.

In table 4.9 the size of each clone obtained, apart from B3 and E3, has been determined. By adding the values together a total of 44 434 bp is obtained. When taking into account that the IR repeat regions are identical to each other, the values of clones falling into this region can be doubled. This increases the total to 54263 bp. The grapevine chloroplast genome is

probably between 120 and 160 kb in size. Therefore, the clones obtained in this project cover roughly 40% of the genome.



**Figure 4.7** The *V. vinifera* cv L. Cabernet Sauvignon sub-clones from BAC clone 001A01 were mapped onto the *Atropa belladonna* (deadly nightshade) chloroplast genome. The forward and reverse sequences of each clone were used to map the clones. The direction of each arrow indicates the orientation of the clone in the plasmid vector used (M13 forward to M13 reverse, or T3 to T7). SSC: Small single copy region. LSC: Large single copy region. IR: Inverted repeat.

**Table 4.7** The percent homology found between the sequenced BAC sub-clones and the chloroplast sequences of *Atropa belladonna*, *Nicotiana tabacum* and *Spinacia oleracia*.

Clone	Primer	Plant Species	% Homology	Clone region	Homologous area	E-value	
<b>B1</b>	M13 forw	<i>A. belladonna</i>	94	12-839	24897-24067	0.0	
		<i>N. tabacum</i>	94	12-839	25130-24300	0.0	
		<i>S. oleracia</i>	94	8-812	23241-22437	0.0	
		<i>A. thaliana</i>	92	8-811	23946-23143	0.0	
	M13 rev	<i>A. belladonna</i>	86	532-788	23382-23645	3e-63	
			84	240-463	23113-23330	6e-50	
			83	1-187	22881-23060	3e-37	
		<i>N. tabacum</i>	85	532-788	23614-23878	3e-59	
			86	240-463	23345-23562	4e-55	
			83	1-187	23113-23299	3e-37	
		<i>S. oleracia</i>	82	322-788	21517-21986	2e-93	
			84	2-147	21209-21361	2e-27	
			89	244-282	21447-21485	0.001	
			90	27-59	80179-80211	0.063	
		<i>A. thaliana</i>	86	601-788	22500-22690	2e-45	
			83	240-457	22133-22342	6e-39	
			92	5-81	21878-21953	3e-19	
	90	130-181	22012-22063	2e-09			
<b>B2</b>	M13 forw	<i>A. belladonna</i>	94	1-239	101805-102038;141752-141519	e-102	
			90	225-523	102090-102383;141467-141174	e-100	
			93	519-618	102484-102577;141073-140980	2e-31	
		<i>N. tabacum</i>	94	1-239	101537-101770;141089-140856	e-102	
			90	276-523	101842-102084;140784-140542	1e-81	
			93	519-618	102185-102278;140441-140348	2e-31	
		<i>S. oleracia</i>	92	225-263	101801-101839;140825-140787	5e-04	
			97	1-239	96480-96718;136965-136727	e-114	
			88	237-523	96753-97044;136692-136401	2e-86	
			88	519-815	97139-97441;136306-136004	2e-80	
		<i>A. thaliana</i>	95	1-206	99557-99762;139092-138887	2e-89	
			93	521-693	100208-100380;138441-138269	6e-65	
			89	220-384	99809-99969;138840-138680	3e-42	
		87	714-840	100404-100542;138245-138107	3e-20		
		93	418-462	99988-100032;138661-138617	1e-07		
		94	490-523	100070-100103;138579-138546	0.002		
	M13 rev	<i>A. belladonna</i>	98	3-831	104738-103903;138819-139654	0.0	
		<i>N. tabacum</i>	98	3-831	104428-103594;138198-139032	0.0	
		<i>S. oleracia</i>	98	91-831	99522-98781;133923-134664	0.0	
			96	13-98	99612-99527;133833-133918	4e-32	
<i>A. thaliana</i>		98	1-831	102677-101846;135972-136803	0.0		
<b>PS1</b>		M13 forw	<i>A. belladonna</i>	98	9-981	98427-99399;145130-144158	0.0
			<i>N. tabacum</i>	98	9-981	98161-99133;144465-143493	0.0
	<i>S. oleracia</i>		96	13-981	93411-94373;140034-139072	0.0	
	<i>A. thaliana</i>		96	9-981	96060-97036;142589-141613	0.0	
	M13 rev	<i>A. belladonna</i>	95	199-409	100066-99858;143491-143699	6e-91	
			94	5-157	100255-100103;143302-143454	3e-62	
		<i>N. tabacum</i>	95	199-409	99793-99585;142833-143041	6e-91	
			94	5-157	99982-99830;142644-142796	3e-62	
		<i>S. oleracia</i>	96	199-409	95406-94838;138399-138607	3e-93	
			94	1-157	95239-95083;138206-138362	1e-64	
	95	199-409	97731-97523;140918-141126	6e-91			
	94	1-157	97924-97768;140725-140881	1e-64			

PS2	M13 forw	<i>A. belladonna</i>	92	312-454	49830-49973	4e-45	
			93	1-76	49482-49556	7e-19	
		<i>N. tabacum</i>	85	109-230	49643-49765	1e-14	
			90	892-946	50379-50433	9e-09	
			92	315-454	50186-50326	2e-43	
			93	1-76	49841-49915	7e-19	
			84	109-230	50002-50119	1e-14	
			92	892-946	50738-50792	4e-11	
		<i>S. oleracia</i>	97	363-444	46932-47013	7e-32	
			100	1-46	46558-46603	2e-14	
			77	109-225	46675-46787	7e-06	
			87	504-536	47041-47073	4.5	
			94	368-451	48175-48258	7e-28	
		M13 rev	<i>A. thaliana</i>	93	391-419	48018-48046	0.31
	97			4-43	47451-47490	2e-09	
	93			906-968	48597-48658	2e-14	
	91			1-749	57975-57228	0.0	
92	1-749			58322-57575	0.0		
90	1-747			54552-53801	0.0		
92	1-749			55685-54938	0.0		
PS3	T3	<i>A. belladonna</i>	91	1-663	57980-58642	0.0	
			90	710-758	58755-58805	0.004	
		<i>N. tabacum</i>	91	1-686	58327-59012	0.0	
			90	710-758	59102-59152	0.004	
			92	1-689	54557-55245	0.0	
		T7	<i>A. thaliana</i>	92	1-688	55690-56377	0.0
				93	97-159	63493-63431	5e-16
	<i>A. belladonna</i>		92	256-293	63343-63306	0.002	
			92	802-852	62856-62803	0.002	
			82	503-588	63115-63027	0.11	
			92	43-70	63538-63511	6.9	
			87	97-159	63292-63230	2e-06	
	<i>N. tabacum</i>	92	802-852	63189-63136	0.002		
		89	802-856	59104-59047	0.44		
		100	105-128	60581-60558	0.028		
	PS4	T3	<i>A. belladonna</i>	94	545-749	13268-13063	1e-83
				98	8-73	13724-13659	5e-25
<i>N. tabacum</i>			89	238-284	13513-13467	4e-04	
			91	494-527	13304-13271	0.38	
			95	545-749	13480-13275	6e-86	
			100	8-73	13918-13853	2e-27	
			89	238-284	13714-13668	4e-04	
			91	494-527	13516-13483	0.38	
<i>S. oleracia</i>			96	553-733	12109-11929	1e-80	
			96	8-73	12549-12484	1e-22	
			87	179-233	12387-12332	0.024	
			92	560-748	12811-12622	7e-67	
<i>A. thaliana</i>			96	8-72	13326-13262	5e-22	
			93	239-284	13156-13111	3e-08	
		86	187-237	13191-13141	0.095		
		90	485-774	10387-10680	5e-87		
		98	213-310	10192-10289	5e-44		
T7	<i>A. belladonna</i>	98	1-66	9982-10047	6e-25		
		96	33-59	37840-37866	0.11		
		90	485-774	10619-10912	2e-89		
<i>N. tabacum</i>	98	213-309	10421-10517	2e-43			

			98	1-66	10201-10266	6e-25		
		<i>S. oleracia</i>	96	33-59	38101-38127	0.11		
			89	474-720	9164-9411	4e-69		
			97	210-310	9030-9130	5e-41		
			98	1-61	8875-8935	6e-22		
		<i>A. thaliana</i>	96	36-63	35205-35232	0.028		
			86	501-721	9948-10169	2e-46		
			98	217-294	9585-9662	4e-32		
			96	1-61	9372-9432	1e-19		
			95	36-59	36536-36559	6.9		
<b>PS5</b>	Compl. Sequ	<i>A. belladonna</i>	95	1-64	82358-82295	5e-19		
			93	129-588	82245-81786	0.0		
			91	616-724	81773-81665	1e-31		
				<i>N. tabacum</i>	95	1-64	82125-82062	5e-19
		93	129-588		82012-81553	0.0		
		91	616-724		81540-81432	1e-31		
				<i>S. oleracia</i>	91	1-62	78172-78111	4e-13
		91	129-589		78011-77551	e-167		
		88	631-724		77524-77425	1e-19		
				<i>A. thaliana</i>	90	1-64	79562-79499	7e-12
90	181-585	79386-78982	e-136					
92	641-724	78937-78854	8e-24					
<b>PS6</b>	T3	<i>A. belladonna</i>	97	125-207	48163-48245	5e-33		
		<i>N. tabacum</i>	97	125-207	48503-48585	5e-33		
		<i>S. oleracia</i>	95	124-228	45492-45595	1e-36		
			96	29-57	45377-45405	0.003		
		<i>A. thaliana</i>	98	131-210	46209-46288	1e-33		
	T7	<i>A. belladonna</i>	85	1-389	49487-49112	3e-96		
			87	404-670	49108-48840	2e-75		
			83	740-794	48785-48730	0.064		
		<i>N. tabacum</i>	84	1-389	49846-49465	5e-91		
			87	404-670	49461-49198	8e-75		
			82	740-816	49138-49062	0.004		
		<i>S. oleracia</i>	91	429-566	46352-46212	7e-46		
			88	1-145	46563-46411	2e-39		
			97	351-389	46412-46374	7e-09		
			84	739-790	46052-46000	0.064		
		<i>A. thaliana</i>	84	271-566	47179-46894	1e-73		
			89	1-131	47453-47322	3e-37		
<b>PS 7</b>	T3	<i>A. belladonna</i>	93	1-802	36585-35777	0.0		
		<i>N. tabacum</i>	93	1-802	36830-36022	0.0		
		<i>S. oleracia</i>	91	1-802	33867-33059	0.0		
		<i>A. thaliana</i>	91	1-802	35071-34263	0.0		
	T7	<i>A. belladonna</i>	92	1-779	26489-27267	0.0		
		<i>N. tabacum</i>	92	1-779	26722-27500	0.0		
		<i>S. oleracia</i>	90	1-779	24829-25607	0.0		
			90	186-772	25725-26311	0.0		
		<i>A. thaliana</i>	92	3-139	25542-25678	8e-46		

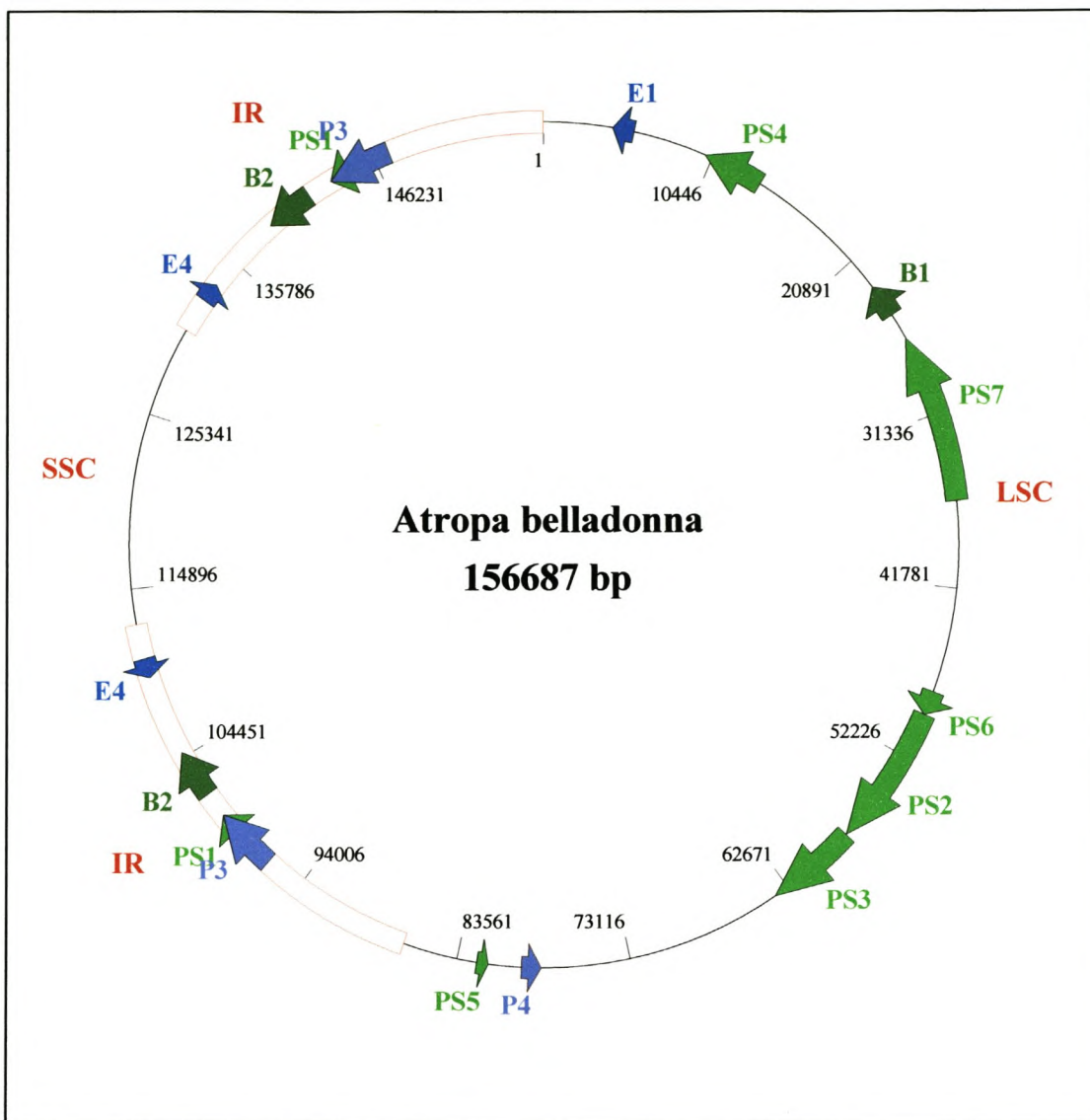


**Table 4.8** The gene locations of the 001A01 sub-clones.

BAC 001A01 sub-clone	Gene location
B1	<i>rpoB - rpoC</i>
B2	<i>rps7</i> , tRNA-Val (GAC) - <i>rrn16</i>
PS1	<i>rps7 - ndhB</i>
PS2	tRNA-Leu (UAA), tRNA-Phe(GAA) - <i>rbcL</i>
PS3	<i>rbcL - ycf4</i>
PS4	<i>atpF</i> , <i>atpH</i> – tRNA-Gly(UCC), tRNA-Arg(UCU), <i>atpA</i>
PS5	<i>rpoA</i> , <i>rps11</i> , <i>rpl36</i>
PS6	tRNA-Thr(UGU) – tRNA-Leu(UAA)
PS7	<i>psbC - rpoB</i>

**Table 4.9** The insert size of each cpDNA clone.

CpDNA Clone	Size of insert (bp)
E1	1304
E4	1186 (x2)
P3	3866 (x2)
P4	1200
B1	2005
B2	2935 (x2)
PS1	1842 (x2)
PS2	8493
PS3	5601
PS4	3750
PS5	703
PS6	1453
PS7	10096



**Figure 4.8** Sultana and Sugra 1 cpDNA clones, as well as BAC 001A01 sub-clones, mapped onto the *Atropa belladonna* (deadly nightshade) chloroplast genome. The forward and reverse sequences of each clone were used for mapping. The orientation of each clone in the appropriate plasmid vector is indicated by the direction of the arrows. SSC: Small single copy region. LSC: Large single copy region. IR: Inverted repeat.

### 4.3 GENERAL DISCUSSION

The primary goal of this project was to create a library of overlapping clones covering most of the *V. vinifera* chloroplast genome and to characterise these clones. The first step towards accomplishing this goal was to isolate pure grapevine cpDNA. The cpDNA would then be digested with appropriate restriction enzymes and the fragments shotgun-cloned into a plasmid vector. Clones would be partially sequenced and arranged correctly on a map by comparing them to already sequenced higher plant chloroplast genomes.

The isolation of pure cpDNA from grapevine proved to be a challenging task. High levels of secondary metabolites present in grapevine leaves hindered the isolation procedure and made it difficult to eliminate contaminating genomic DNA. A number of methods were attempted for grapevine cpDNA isolation and chemicals such as PVP-10 and high levels of NaCl were included to counter the actions of secondary metabolites. The optimised protocol did not involve the use of sucrose gradients or DNase I treatment. Cell debris and genomic DNA were removed by means of a simple centrifugation step and CTAB was utilized for chloroplast lysis. Proteins and other impurities were removed by means of organic extraction. The optimised protocol gave high yields of relatively pure cpDNA, with low levels of genomic DNA contamination.

The second step involved the digestion of the isolated cpDNA with a number of restriction enzymes, to determine which of these would be suitable for further library construction. The restriction enzymes *Bam*HI, *Pst*I and *Eco*RI were selected and the digested cpDNA was shotgun-cloned into the plasmid vector pUCBM21. White colonies obtained from ligation 1, 2 and 3 (table 4.1) were screened by means of PCR amplification. PCR is a relatively quick and inexpensive method of determining the actual number of positive clones. Clones that may have contained identical inserts could have been eliminated by subsequent restriction digestion of the PCR products with frequent-cutter restriction enzymes. Thirteen of the 56 clones screened were amplified during PCR. It was, however, later realised that clones that appeared negative after PCR may have contained large fragments that were not optimally amplified during PCR. We did, however, manage to amplify fragments of up to 5 500 bp.

Due to the seasonal nature of grapes, leaves were not always available for cpDNA extractions and another strategy had to be followed. A region of the grapevine *rrn16* gene had been sequenced by a previous student and we decided to use this sequence as a starting

point from which to walk out. This was done by designing outward facing primers using the *rrn16* region and designing complementary primers using sequences of other higher plant chloroplast genomes. Using this approach an area of 3320 bp was covered. However, the expense and time consuming nature of this method meant that it was not continued once leaf material became available again.

An alternative method was attempted for screening the 604 clones obtained from ligation 4 (table 4.1). Since two *Pst*I clones had already been sequenced and found to be of chloroplast origin these were used to find overlapping clones in the *Eco*RI library. The strategy was to then construct a separate *Pst*I clone library and use the positive *Eco*RI clones to find overlapping clones in the *Pst*I library. The positive *Pst*I clones would again be used to screen the *Eco*RI library. In this way we would be able to walk outwards from the two original *Pst*I clones. Although the positive control signal was observed on the autoradiography film, no conclusive results were obtained from the hybridisation of the *Pst*I clones to the *Eco*RI library. One possible reason could have been that the *Pst*I clone was too large (~5 500bp) and homologous regions may have been very small, resulting in only a feint signal being visible on the autoradiography film. Further optimisation of this method was not carried out due to the availability of BAC clones containing chloroplast DNA.

The BAC clones offered an alternative method of sub-cloning chloroplast DNA that appeared to be less arduous than the methods already used. While the isolation of pure chloroplast DNA is made difficult by secondary metabolites and genomic DNA contamination, cloning the chloroplast genome into BACs could make further restriction digestion and sub-cloning easier. BAC clone isolation only requires a variation on a standard alkaline lysis procedure. It is therefore cheaper and easier to obtain high enough yields to carry out further analysis than obtained from cpDNA isolation procedures.

The end sequences of seven of the BAC clones received were compared to the sequences available in the NCBI database, using BLAST 2. Analysis of the search results allowed us to estimate the size of each chloroplast insert, as well as the percentage of the genome covered by each clone. BAC clone 001A01 was found to contain most of the genome, except for a ~500 bp region, and was therefore chosen for further analysis.

Four methods were used for the isolation of the BAC clone 001A01 DNA. Although the variation on a plasmid miniprep procedure (protocol 1) gave relatively clean DNA, it was not

enough to use for further analysis. Large culture volumes were therefore required. Protocol 2 and 3 did not give much success, with large amounts of genomic DNA being present. Protocol 4 gave low yields of clean BAC DNA, but ~31% genomic DNA contamination was still present.

The 001A01 BAC clone DNA was digested with a number of restriction enzymes and fragments were separated by agarose gel electrophoresis. The size of the fragments as represented on the agarose gel were used to calculate the size of the grapevine genome and to determine which restriction enzymes were suitable for library construction. The minimum size of the grapevine chloroplast genome was calculated as ~115 Mb. Possible doublet bands containing IR fragments, as well as very small and very large fragments, could not be detected on a gel photo. The sizes of these fragments would need to be added to the value obtained. However, the value is consistent with the size of the average higher plant genome (120 to 160 kb). The restriction enzymes *Bam*HI, *Sal*I and *Pst*I were selected for further sub-cloning of the BAC clone DNA into a plasmid vector. The insert DNA was not separated from the BAC vector backbone prior to digestion and cloning because of the difficulty of purifying such a large insert fragment from a gel. The DNA could be sheared during the gel purification process and yields would be lowered.

Plasmid DNA isolation and restriction digestion of this DNA with the appropriate enzyme(s), proved to be a more successful method of screening white colonies for positives than PCR or colony blotting. Enough plasmid DNA was isolated from each clone to allow for a number of subsequent restriction digestions to be performed. Clones containing BAC vector fragments or identical inserts were eliminated. Putative positive clones were sequenced. A total of three *Bam*HI clones (B1, B2, B3) and seven *Pst*I/*Sal*I clones (PS1-PS7) were obtained in this manner.

The use of BAC clones to create a sub-genomic library of grapevine cpDNA proved to be more effective than isolating cpDNA from grapevine leaves. While only 5 clones (E1, E3, E4, P3, P4) were obtained from extracted cpDNA, 10 clones were obtained from ligations using the BAC clone DNA. All the clones obtained in this project were mapped against the *A. belladonna* chloroplast genome. They represented most of the chloroplast genome well, apart from the SSC region, and they covered approximately 40% of the grapevine chloroplast genome.

## 5. CONCLUSION

The chloroplast genome encodes for a number of genes that are vital for plant survival. The characterization and sequencing of cpDNA therefore provides researchers with information that will help them understand the functions of the chloroplast in the cell as well as its evolution. This understanding will allow them to manipulate the chloroplast genome to improve cellular and plant functioning.

The aim of this project was to create a library of overlapping clones that would cover most of the *V. vinifera* chloroplast genome. Chloroplast DNA was initially isolated from grapevine leaves for library construction, but the use of BAC clones containing grapevine chloroplast DNA proved to be more successful. However, even when a commercial BAC DNA isolation kit was used, ~31% of clones still contained genomic DNA inserts. In fact, problems with genomic DNA contamination were experienced throughout this project. Clones containing genomic DNA could also only be identified and removed once putative positive clones had been sequenced.

Plant nuclear DNA as well as secondary metabolites hindered the isolation of cpDNA from leaves, but the addition of compounds such as PVP-10 and NaCl did help in eliminating these contaminants. Further optimization of the cpDNA isolation procedure could help in purifying the grapevine cpDNA even further. Sodium chloride levels have been found to play an important role in minimizing nuclear DNA contamination, and could be added to the chloroplast suspension during chloroplast lyses. The addition of NaCl (0.7 M) to the chloroplast suspension with CTAB has had positive results (Milligan, 1994). The pH of the isolation buffer as well as the components of the isolation buffer used to protect the DNA from harmful enzymes or secondary compounds could also be further adapted (Milligan, 1994). A pH of 8.0 to 9.0 for the isolation buffer is preferable. The inclusion of PEG (10–25%), as well as higher concentrations of PVP-10 (up to 10%) or 2-mercaptoethanol (up to 25 mM), could minimise the degradation of cpDNA by secondary compounds (Milligan, 1992). An option that came to light when using the QIAGEN® Big Construct Kit is the use of ATP-Dependent Exonuclease for the digestion of genomic DNA contamination. This may prove to be a very useful tool for obtaining pure cpDNA in future isolations.

The isolation of high yields of clean BAC DNA also proved to be difficult. The low copy number of BACs (1-2 per cell) meant that large culture volumes were needed; with the added implication of an increase in host cell genomic DNA. Further optimization of the BAC clone DNA isolation procedure will need to be performed to eliminate the genomic DNA completely. Protocol 1 gave the purest BAC DNA and attempts to carry out this method on a large-scale, followed by further purification steps could eliminate genomic DNA further. Purification of isolated BAC DNA by means of Cs-Cl gradient centrifugation has proven successful (Dr. Adam-Blondon, Pers. comm.). However, this method is very time consuming and gives low yields of BAC DNA. The digestion of linear genomic DNA fragments with ATP-Dependent Exonuclease may provide cleaner BAC DNA. Success has been obtained with this type of protocol (<http://wheat.pw.usda.gov/~lazo/methods/caltech/caltech.html>).

Although attempts at cloning were hindered by the above-mentioned contamination, a total of 15 grapevine chloroplast genome fragments have been cloned and partially sequenced in this project. Ten of the clones obtained were BAC sub-clones and the remaining five were obtained from the isolated Sultana and Sugra 1 cpDNA. The use of BAC clones for cpDNA library construction therefore gave better results than when using cpDNA extracted from grapevine leaves. This could also partially be a result of the screening technique used. The isolation and digestion of putative positive clones gave more consistent results than PCR, and was less expensive and time-consuming than colony blotting. Further library construction would involve the isolation of the remaining *PstI/SalI* clones, either by means of additional shotgun-cloning or by excising individual fragments from a gel and cloning each into a plasmid vector. The restriction enzyme *XhoI* also produced fragments of optimal sizes for library construction. Digesting the BAC DNA with different combinations of *XhoI*, *PstI* and *SalI* could assist in generating overlapping clones. A second BAC clone could be used to generate clones of the region not covered by clone 001A01.

Towards the end of this project we were approached by Dr. J. Martinez-Zapater (Centro Nacional de Biotecnología, Madrid, Spain) who was interested in entering into a collaboration with us to complete the sequencing of the *V. vinifera* cv L. Cabernet Sauvignon chloroplast genome. We agreed to this and have since sent him the BAC clones we obtained from Dr. Adam-Blondon. His group will also use these clones to construct grapevine chloroplast DNA libraries. The clones produced by both the Spanish group and ourselves will be combined to provide overlapping clones of the entire Cabernet Sauvignon chloroplast genome. Further sequencing of this genome could then be carried out.

The clones and information produced by the project we have undertaken will also contribute to a larger project aiming to sequence the entire grapevine genome (470 Mb). The 'International Grape Genome Project' (IGGP) involves labs in many wine-producing countries around the world. The goal of the IGGP is to analyze genetic traits that have not previously been researched by other genome sequencing projects. Examples include grape quality, pathogen resistance and tendril development traits. This information will give researchers, plant biologists, viticulturists and oenologists a better understanding of *Vitis* and help them to improve both grapevine resistance and grape quality (Ablett et al. 2001, <http://www.vitaceae.org/>).

In conclusion, even though we encountered problems with genomic DNA we still managed to clone 15 fragments, covering approximately 40% of the grapevine chloroplast genome. These clones represent the genome well, with only the SSC region not having clones mapped to it. These clones provide a valuable tool for further characterization and sequencing of the grapevine chloroplast genome. The knowledge gained from this research will, in turn, give us a better understanding of how the chloroplast genome of a woody, fruit-producing plant functions and novel characteristics, not yet observed in chloroplast genomes, may even be discovered.



## Addendum 1

### Sugra 1 and Sultana cpDNA Sequences

#### E1 (M13 forward)

1 ATTCCTTTATT CAATTCCGAC ATAAATGAAA TATTATTGAG TAGTCTACTT CCCTTCGAAT GATGAATCCC CTTAACTTA  
 81 AATTAAAAAT TAAAGGAGTA CCTTAGAATT CTATACATCG ACATAACTTT GAATCATTFTT TTCTCGAGCC GTACGAGGAG  
 161 AAAACTTCTT ATACGTTTCT AGGGGGGGGG GGTATTGTTC ACCTACATCT ATCCCAATGA GCCGTCTATC GAATCGTTGC  
 241 AATTGATGCT CGATCCCGAA GAGAAGGAAG AGATCTTCGG AAAGTGGGTT TTTATGATCC GATAAAGAAT CAAACTTATT  
 321 CAAATGTTCC TGCTATTCTA TATTCCTTG AAAAAGGAGC TCAACCTACA GGAAGTGTTC ATGATATTTC AAAGAAAGCG  
 401 GAGGTTTTTA CGGAAGTTCG TCTTAATCAA ACGAAATCA AATTCATCA ATGAAATACA AAAACGAGG GGGGGATGAC  
 481 TCGTATATAG CTTTGTATAA CTTTCTCTAC TACTGCCCTT TAATCTATCT TATTGATATA AGATGGATAG AAAAAAGATC  
 561 AAGTGNATAG ATGAAGGAAT TATATCTAGA AATATAAAAT TCTGACATTA CCTTCAATCA GCGGGTTTTT CTTAGAACTC  
 641 TACTAACAAA CACGGGATCA AGTTTGCTTA TTATACATTC TCTTTATATT GATTGAATAG ATNATGGAAT AGACCGCCGA  
 721 TTTTCTCTAC TTTTCTCTTA TTATTCGCCA TCTAACCTTT TTGTATCTAT GTAGATTCTT ATATAGGGCC ATCAC

#### E1 (M13 reverse)

1 AATTCTCCAT TGgTAagTTG ATACGACATG CTGTTTTTTC CATTTCATCC CTTTCAGGGT ATCAGTCGCG GTCTTACAAA  
 81 CTATACCAAT GGTATGGACG AATCCGTTCC TTCATCCAAA TGTGTAAGG ATTCTAGCCG CACTTAAAG CCGAGTACTC  
 161 TACCATTGAG TTAGCAACCC GAATAAGATA AAATTCATTA GAAAATAATT AGGGTATGTA GATACAACCG GAATCAAAAC  
 241 AAATGAAAGT AAAGAGATTG GGTACACGA CGCAATCgga aCATTGAACT AACaATAAGA TAAATAAAAT AAAAAAGAAA  
 321 ACACACaTTT TCTAATCGAT TCTGAAGATA AAAATGAACA GCTCAGATGA AAATACAAGA AATCTTAgT TATTAGATAG  
 401 ATAAATGTAA AAATTATAG ACCAACTCTT ATCTTTTCCA TTTTTTGTAT TGAAAAAACT GCTTATCTTA TGGCACAGCG  
 481 AATACAACGA AACCATCATT TGAATGAGGT AgAgTAAAA CCAAACCTAT GGAATGGAGA AAAGAAGATT CATTATCTA  
 561 CGATCAAATT ATAGTTGGTT TCTACACAGT GTCAATATGA ATGAATGTTG AGAAAAAGAA TACAATGTAG AAAACAAAAA  
 641 AA

#### E3 (M13 forward)

1 TGAATYCTTT GAAGGCTNCA TCATGAAGAG CAAGTCAACC ACTCGAGCTA GTGCATGCCG ACATATGTGG TCCAATGCAA  
 81 ACGTTGTCTC TCAACAAAAG TAGTTCTCAT CTTTGTGAT GATTTCATA GGGGTATATT TTATTA AAAA CAAAGAAGAT  
 161 GCATTTCTCG TTTTCTTCA ATTCAAAGCT TATGTAGAAA AACATATTGG TCATAATTGG AAGGTCCTTC GAACTGCTTA  
 241 TAGTCTAATA TAGCCGGTGG TGAGTTCACA TCGAATTCOA CTTTAAAGAG ACATGCTACA AAAATAGCCC AGTTTATGAA  
 321 ACTTCTTATC TGGATGGGAA TCAAGAAAAT TCGAAGTTAG AAATACTTAA AGAAGATGAA GATAATAAAA ACAAAAAATG  
 401 GTTTGAAAAA CCCCTGTGA CTCTTCTTTT TGAATAAAA CGATGGAATC GTCCATTGCG ATATATAAAG GCTCGATTTC  
 481 AAATTA AAAA AGCTGTAAGA AATGAAATGT CACAATATTT TTTTATACA TGCCTCAGTG ATGGAAAACA AAGAATATCT  
 561 TTTACATATC CCCCCAGTTT GTCAATTTTT TTGAAATGA TACAAAGAAA GATGTCTTTG TCCACAACAG AAAGACTCTC  
 641 CTATGACGAA CTATATAATC ATTGGATTTA TACCAATGAA CAAAAGAAA CACACTAAGA AAGGGTTTAT ACGAGAATCG  
 721 AGGCTCTAGA CAATGGATCG CTTACTCTGG ATGTACTGGA AAAAGACTAG ATGTGTATGA TGCACTAAAA AGATACTTGC  
 801 CTAATATAT GATCTTCTG ACGACCTATC GTGAACATCA AAATGTTTCC CTCATCTATG AACTC

**E3 (M13 reverse)**

1 AATTCATAT TCGAAGTTGC GATCGGATCT ATTCATTAAA AAGAATCGAT TCAATACATT TCTTATGTAC CCATAGGTGC  
 81 TATATTGGAT TTGAATCAGA TTTCGGATCA ATATATATTG ATTGACTGCC TCCATTATGT TGTGCTAGC AAATACCACT  
 161 ATTTTTGGTT TTGGATCTTC CAAATCATTT CCGCAGGAGA TCCGGACCCA TTTTTTCTG ATCCTTCGAT AAAAAGATTC  
 241 ATTCTCTTCA TAAAACTAG GAGGTAGAAC CAATAAGAT TTCTTGTTTCG ATTCATCCCT GGAGTTAAAT ACCTCATTCA  
 321 AGAATTGTTT TTGATCCAAT CCGTAGGAAT CAATAGAAAA GGCAATCCCT TTATGATACA CCAGATCCGG CTCGGTTATT  
 401 GATAGAGTGA ATCGATCTGC CATTCTTGA AATCTCTCTT CTGATTCAAA ATCGTGGTGT AACGTGTATC CCCCCTGT  
 481 CCGGTCATGG AATAGATGAA ATAAATCAAA AAATGGATTT TTGTTCAAGA ATGAAATCTT ATTGGAACGTGCCATATCCG  
 561 GTTCATCCTT CGGAACCATA TCACATCCCG GATCTGATGA AATAGGATGA ATTGAGACGG TATTTTGTAATACGTAATW  
 641 ATCTTGAATA TATTACCATT TCTTTTATTT TCCGATCGCC TGAAGGGAC AAAAAACAT CTTGGTGTTC TCACAATTT  
 721 TGATCTCTAG GGACCTCTCA TAGGATCGAC CCGAGAAGTT CTGACATCTG TC

**E4 (M13 forward)**

1 GATGACTCT TACTAGCAGC AGCATCAAAG ATGCAGTCAT CGATTCTCCC GAGAGGCCAC AATTACCGCG AGCAAACATA  
 81 TTAATGACGA GGAACGCATT TTTGCTATGC TACTAATACT TGTACTTGCT CTGCTATTCT GCCCAAGCCT GGCTGAGGAA  
 161 GAGTTACGGG GCGTAAAACA AAAAAATATG CTGATTCGGG CCGGGCATA TATAAGTAAT GATTATATCA TTCGCGATAA  
 241 ATATAAATAA AAAGTAAGGC CATTCCATTT CGACAAAAGA CCCACACCA AGCCAAGTTC CATAGCTTTG GGTCCGCTAT  
 321 CCCGATCAGG ATTTTCTTAC CCTCAGAGG AAAGGTCCTT CCCTTTTGGG CCGGTTGTGG GCGAGGAGG ATTCGAACCC  
 401 CCGACACCGT GGTTCTGAGC CACGTGCTCT AATCTCTGA GCTACAGGCC CCACCCGTC TCCACTGGAT CTGTCCCGG  
 481 GAGTACCCTC AAAAAAAAAA AGGAACCTTT CCTCTCCCA GACATTTGCG GTTAAGAAGA TGTGAAAGCG CATTCTCTC  
 561 TATAACTATA AGAAGGGTGC GTTCCGAGGT GTGAAGTGGG AGAGAAGGA TGTCATAAAT GGGGTTTGA ATAAAACGAC  
 641 CTTTCTTTTCT ATTTTTTCT TTTTCATATT TAAAAGTATA AGAATGAGAG GTGTAGCTTT TTATCATCCT GGCCTCGAGC  
 721 TATTTTCCGC AGGACTCCCC TACAGTATCG TCACGAGTA GAGTTACACA GTTCGGGATG GATGGGGGGG GTCCTCTACC  
 801 CTAGGACCCG AATTCGAACC TGAACGAAAA AGGCTGAAAA AAAGCTTTTG GCTAG

**E4 (M13 reverse)**

1 AATTCAAGAG AAGGTGCACG GCGAGACGAG CCGTGTATC ATTACGATAG GTGTCAAGTG GAAGTGCAGT GATGTATGCA  
 81 GCTGAGGCAT CCTAACAGAC CGGTAGACTT GAACCTTGT CCTACATGAC CCGATCAATT CGATCAGGCA CTCGCCATCT  
 161 ATTTTCATTG TTCAACTCTT TGACGAACAT GTAAAAAACC AAAAGCTCTG CCCCCCTCT CTATCGGATG GAAGGCAGA  
 241 GGCCTTTGGT GKCCCTTCCA GTCAAGAATT GGGGCCCTCAC AATCACTAGC CAATATGCTT TTCTCTCATG CCTTCTCTG  
 321 TTCATGGTTC GATATCTG TGTCCTAGGC GTAGAGGAAC CACACCAATC CATCCCGAAC TTGGTGGTTA AACTCTACTG  
 401 CCGTGACGAT ACTGTAGGGG AGGTCCTGCG GAAAAATAGC TCGACGCCAG GATGATAAAA AGCTTAACAC CTCTCATTCT  
 481 TATTACTTTT TAAATATGAA AAAGAAAAA ATGAAAAGAA AGGTCGTCTT ATTCAAGACC CCAATATGA CATCCCTTCT  
 561 CTCCCACTTC ACACCTCGGA ACGCACCTT CTTATAGTTA TAGAGAGAAA TGCGCTTTC CATCTTCTTA ACCCGAAATG  
 641 TCTGGGGAGA GGAAAGTTC TTTTTTTTTT TGAGGGTACT CCCGGGAACA GATCCAGTGG AACGGGAGGG GCCTGTAATC  
 721 CAAAGATTAA A

**P3 (M13 forward)**

1 GACGAAGGGA AGGTCTTCCA TTTATTAGTA TTCAGTAACC CATGATTCGT TATGGAGCAG GTAGCAGCAA CCATTCATCA  
 81 GACATGGGTA TTTTGTGATT TCCAATGGAT TACATCTTTC ATTAATGGAA ATTTTTGATG TAGGGAGTAA TAGGCTCTGG  
 161 TGGTTCGCTG TTCAAGAATT CTTGTTTAGG CAGTTTWTAC CATCCATACA TAGTGTTTTG ATCTRAGATT TCAATTCTTC  
 241 CCATGTTTCA GCAGTAGCAT ATTGTTCCCA TGGAGCTAAG GTCCAAAATA TGAAGAAAC AAGTGTTC ACGACTCTAC  
 321 CACACAGTCA ATTCTGTTC ACTTAATCCC TATTTTCATGG CCACATATCT TTCCGGCTAA GGAATGGGAA ATCTTTCTCC  
 401 TGTTACATGA ATCCAATTTT CATTTCATCC GGGAAAAGCC ATCTTTTTCT CAACAATGTC TTTGTCAATT GATCCAATAG  
 481 CCTTGCCTTA GATAGGAACA GATTTGATAA ATACTGATAA CTCTCGGATG GAGTATTAGA ACGGAAAGAT CCATTAGATA  
 561 ATGAACTATT GGTTCTAAGC CATCTCTGGC GATGAATCAA CAATTCGAAG TGCTTTTCTT GCGTATTCTT GATAAACCAAG  
 641 CGTTTATATA TAGATGTAGG AGGATCTGTT TGGGAAGTAA GAAGCCCTT TGACATCTCT TCATCTGCAA AGAATTCTCG  
 721 ATGTGAAAAC ACAGAGACAA AGGGCTGATC TTTGAATAGG AAAAAGAGTG GATCTGCA

**P3 (M13 reverse)**

1 GAAGAAAAA CTGCAAAATC CGATCCAATT TATCGTAATC GATTAGTTAA CATGTTGGTT AACCGTATTC TGAAACACGG  
 81 AAAAAAATCA TTGGCTTATC AAATTATCTA TCGATCCGTG AAAAAGATTC AACAAAAGAC AGAAACAAAT CCACTATCTG  
 161 TTTTACGTCA AGCAATACGT AGGAGTAACT CCCGATATAG CAGTAAAAGC AAGACGTGTA GCGGATCGA CTCATCAAGT  
 241 TCCCATGAA ATAGGATCCA CACAAGGAAA AGCACTTGCC ATTCGTTGGT TATTAGGGGC ATCCCGAAAA CGTCCGGGTC  
 321 GAAATATGGC TTTCAAATTA AGTTCCGAAT TAGTGGATGC TGCCAAAGGG AGTGCGGATG CCATACGCAA AAAGGAAGAG  
 401 ACTCATAGAA TGGCAGAGGC AAATAGAGCT TTTGCACATT TTCGTTAATC CATGAACAGG ATCTATATAG ACACATAGAT  
 481 CCATGGATCC AATCCATACA TCTCGATCGG AAAAGAATCA ATAGAAAAAG AAAGAATCGG AATTGATCGA TATATTTCTC  
 561 GAAACAAACG AAAAGGAAAC GAAAGATGAA ACATAAATCA TGGGATCCAC TAAGTCTCT CGGGGACTTG CTTAAGAATA  
 641 AGAAAGAGGA TCTCATGTAA ATACCATGGA ATAAGGTTGA TCCTATTCAT GCGGATTCGT AAATATTCCA TTCCAAAATA  
 721 GAACATTTGG GATTTTTTGG AGATTGGWTG CAGTACTATC ATGATCTGGC ATATACAGAT GAACTTCATC TCGATCTCGA  
 801 GATTTATGAA GCTTCATTGC TCTCTCGATG AGTTTTTTC AGATGATCTA TTTGCAATCT CTCGAGACGA TCACTCTGWC  
 881 AAAGATACTG GGTATTCTCC TCAA

**P4 (M13 reverse)**

1 GGGAGGGATC CGCGAACCT ATACACATAC TATGCCATCG ACCTATCCGG GTCATACCAT CATTGACCAA AAAACCTATC  
 81 CTATCCATAC CAGATGAGCC ATTCCTTCTC CGTCTGGCCC AAGGTATCAC CATCGTTACA AAGACCGGAT AAGGCGGCAG  
 161 ATGCTTCCTT GAAGCGAACG AGTTGTCCA ATCGGGCAGC TAGAACAGAT ATCGGAAAGG CAAGGCAGTT CTTGCAGGGC  
 241 AGAGGGATAC ATAGAGATAG ATCTCCATCT ATGCGCATGG AAATAGCTAC GGGAGGGTTT GATCGGGATA AGTAGCCGT  
 321 TATATCTAGT GCCTGACCTC GCCTGCCCAA AATCCAGTAG TTGGCCTATG ATTGGTTCGA AGAACGTCTC GAGATTCAGG  
 401 CGATTGCAGA TTCTATAACT AGCTGATCTA CGACCAACCT CCTCATGTCA ACATATTTTA TTGTCTAGGA GGAATTACGC  
 481 TTACTIONTTT TTTAGTACAA GTAGCTACGG GGTGCTAT GACTTTTGAC TATCGTCCGA CGGTTACTGA GGCTTTTGCT  
 561 TCTGTTCAAT ACATAATGAC TGAAGCTAAC TTCGGTGGT TAATCCGATC

## Addendum 2

### *rrn16* Flanking Region Sequences

#### Right flanking region 1

```

1   CTACGGCGGT GAATTCGTTT CCGGGCCTTG TACACACCGC CCGTCACACT ATGGGAGCTG GCCATGCCCG AAGTCGTTAC
81  CTTAACCGCA AGGAGGGGGA TGCCGAAGGC AGGGCTAGTG ACTGGAGTGA AGTCGTAACA AGGTAGCCGT ACTGGAAGGT
161 GCGGCTGGAT CACCTCCTTT TCAGGGAGAG CTAATGCTTG TTGGGTATTT TGGTTTGATA CTGCTTCACA CCCAAAAAGA
241 AGCGAGCTAC GTCTGAGTTA AACTTGGAGA TGGAAGTCTT CTTTCGTTTC TCGACGGTGA AGTAAGACCA AGCTCATGAG
321 CTTATTATCC TAGGTCGGAA CAAGTTGATA GGATCCCCTT TTTTACGTCC CCATGTCCCT CCCGTGTGGC GACATGGGGG
401 CGAAAAAAGG AAAGAGAGGG ATGGGGTTTC TCTCGCTTTT GGCATAGCGG GCCCCAGTG GGAGGCCCGC ACCGCACGAC
481 GGGCTATTAG CTCAGTGTA GAGCGCGCCC CTGATAATTG CGTCGTTGTG CCTGGGCTGT GAGGGCTCTC AGCCACATGG
561 ATAGTTCAAT GTGCTCATCA GCGCCTGACC CTGAGATGTG GATCATCAA GGCACATTAG CATGGCGTAC TCCTCTGT
641 CGAACCGGGG TTTGAAACCA AACTTCTCCT CAGGAGGATA GATGGGGCGA TTCAGGTGAG ATCCAATGTA GATCCAACCT
721 TCTATTCACT C

```

#### Right flanking region 2

```

1   CATGTAAATC ACTTCTATT CCTCGTGGGA TCCGGGCGGT CCGGGGGGGA CCACCACGGC TCCTCTCTTC TCGAGAATCC
81  ATACATCCCT TATCAGTGA TGGACAGCTA TCTCTCGAGC ACAGGTTTAG GTTCGGCCTC AATGGGAAAA TAAATGGAG
161 CACCTAACAA CGCATCTTCA CAGACCAAGA ACTACGAGAT CACCCCTTTC ATTCTGGGGT GACGGAGGGA TCGTACCATT
241 CGAGCCTTTT TTTTTTTCAT GCTTTTCCCG GAGGTCTGGA GAAAGCTGCA ATCAATAGGA TTTTCTAAT CCTCCCTTCC
321 CGAAAGGAAN AACGTGAAAT TCTTTTCTCCT TTCCACAGGG ACCAGGAGAT TGGATCTAGC CGTAAGAAGA ATGCTTGGTA
401 TAAATAACTC ACTTCTTGGT CTTTCGACCC CTCAGTCACT ACGAACGCC CCGATCAGTG CAATGGGATG GGTCTATTTA
481 TCTATCTCTT GACTCGAAAT GGGAGCAGGT TTGAAAAAGG ATCTTANAGN GTCTAGGGTT GGGCCAGGAG GGTCTCTTAA
561 CGCCTTCTTT TTTCTTCTCA TCGGAGTTAT TTCACAAAGA CTTGCCATGG TAAGGAAGA GGGGGGAACA AGCACACTTG
641 GAGAGCGCAG TACAACGGAG AGTTGTATGC TCGTTCGGG AAGGATGAAT CGCTCCGAA AAGGAATCTA TTGATTCTCT
721 CCAATTGGT TGGACCGTAG GTGCGATGAT TTACTTCACG GCGAGGTCT CTGGTTCAAG TCCAGGATGG CCCAGCTGTG
801 CCAGGGAAAA GAATAGAAGA AGCATCTAAC TCCTTCATGC ATGCTCCACT TGGCTCGGG GATATAGCTC AGTTGGTAGA
881 GCTCCGCTCT TGCAATTGGG TCGTTGCGAT TACGGGTTGG ATGTCTAATT GTCCAGGCGG TAATGATAGT ATCTTGTACC
961 TGAACCGGTG GCTCACTTTT TCTAAGTAAT GGGGAAGAGG ACCGAAACAT GCCACTGAAA GACTCTACTG AGACAAAGAT
1041 GGGCTGTCAA GAACGTAGAG GAGGTAGGAT GGCAGTTGG TCAGATCCTA GTATG

```

#### Left flanking region

```

1   GAGAAGGGG CTCAGCGGA AAGAGGATTG TACCATGAGA GAAGCAAGGA GGTCACCTC TTCTCTTCA AATATACAAC
81  ATGGATTCTG ACAATGCAAT GTAGTTGGAC TCTCATGTCG ATCCGAATGA ATCATCTTTT CAACGGAGGT AATGCCTGCT
161 AGGTAAGAGG ATAGCAAGTT ACAAATTCTG TCTCGGAAGG AATTTGTCCA TTTTTCGGG TCTCAAAGG GCGTGAAAC
241 ACATAAGAAC TCTTGAATGG AAATGGAAAA GAGATGTAAC TCCAGTTCCT TCGGAAATGG TAAGATCTTT GGCACAAGAA
321 GAAGGGGTTG ATCCGTATCA TCTTGACTTG GTTCTGATTC CTCTATTTTT TTAAGAATAC CCGAGTCGGG TTCTTCTCCT

```

```
401 ACCCGTATCG AATAGAACAT GCTGAGCCAA ATCTTCTTCA CTGCTTGCTT TAGATCGGGA AAATCGTACG GTTTTATGAA
481 ACCATGTGCT ATGGCTCGAA TCCGTAGTCA ATCCTATTTC CGATAGGAGC AGTTGACAAT TGAATCCCAT TTTTCCCATT
561 ATTTTCGTAT CCGTAATAGT GCGAAAAGAA GGCCCGGCTC CAAGTTGTTC AAGAATAGTG GCGTTGAGTT TCTCGACCCT
641 TTGCCTTACT TAGGATTAGT CAGTCTATAT TCTCGATAGG GGCAGGGAAG GGATATAACT CAGCGGTAGA GTGTCACCTT
721 GACGTGGTGG AAGTCATCAG TTCGAGCCTG ATTATCCCTA AACCCAATGT GAGTTTCTA TTTGACTTGC TCCCCGCGG
801 TGATCGTGGA TAAGAGGCTC GTGGGATTGA CGTGAGGGG TAGGGATGAC TATATTTCTG GGAGCGAACT CCAGGCGAAT
881 ATGAAGCGCA TGGATACAAG TTATGCCTTG GAATGAAAGA CAATTCGAA TCCGCTTTGT CTACGAACAA GGAAGCTATA
961 AGTAATGCAA CTATGAATCT CATGGAGAGT TCGATCCTGG CTCAGGATGA ACGCTGGCGG CATGCTTAAC ACATGCAAGT
1041 CGGACGGGAA GTGGTGT TTCAGTGGCGA CGGGTGAGTA ACGCGTAAGA ACCTGCCCTT GGGAGGGGAA CAACAGCTGG
1121 AAACGGCTGC TAATACCCCG TAGGCTGAGG AGCAAAGGA GGAATCCGCC CGAGGAGGGG CTCGCGTCTG ATTAGCTAGT
1201 TGGTGAGGCA ATAGCTTACC AAGCGATGA TCAGTAGCTG GTCCGAGAGG ATGATCAGCC AACTGGGAC TGAGACACGG
1281 CCCAGACTCC TACGGGAGGC AGCAGTGGG AATTTCCGC AATGGGCGAA AGCCTGACGG AGCAATGCCG CGTGGAGGTA
1361 GAAGGCCAC GGGTCGTGAA CTTCTTTCC CGGAGAAGAA GCAATGACGG TATCTGGGA ATAAGCATCG GCTAACTCTG
1441 TGCCAGCAGC CGCGTAATA CAGAGGATGC AAGCGTTATC CGGAATGATT GGGCGTAAAG CGTCTGTAGG TGGCTTTTAA
1521 AGTCCGCCGT CAAATCCCAG GGCTCAACCC TGGACAGGCG GTGAAACTA CCAAGCTGGA GTACGGTAGG GGCAGAGGGA
1601 ATTTCCGGTG GAGCGGTGAA ATGCGTAGAG ATCGGAAAGA ACACCAACGG CGAAAGCACT CTGCTGGGCC GACTGACA
1681 CTGAGAGACG AAAGCTAGGG GAGCGAATGG GATTAGATAC CCCAGTAGTC CTAGCCGTAA ACGATGGATA CTAGGCGCTG
1761 TCGGTATCGA CCCGTGCAGT GCTGTAGCTA ACGCGTTAAG TATCCGCCT GGGGAGTACG TTCGCAAGA
```

## Addendum 3

### BAC Sub-Clone Sequences

#### B1 (M13 forward)

1 ATCCTCCAAA AGGGTGGTTC TAGTTATAAT CCAGAAACGA TTCGTGTATA TATTTACAG AACGTGAAA TCAAAGTAGG  
 81 TGATAAAGTA GCTGGAAGAC ATGGAAATAA GGGTATCATT TCCAAAATTT TGCCTAGACA AGATATGCCT TATTTGCAAG  
 161 ATGGAAGACC TGTTGATATG GTCTTTAACC CATTAGGAGT ACCGTCACGA ATGAATGTAG GACAGATATT TGAATGCTCG  
 241 CTGGGGTTAG CGGGGGTCT GCTAGACAGA CATTATCGAA TAGCACCTTT TGATGAGAAA TATGAACAAG AGGCTTCGAG  
 321 AAAACTAGTG TTTTCTGAAT TATATGAAGC CAGTAAGCAA ACAGCGAATC CATGGGTATT TGAACCCGAG TATCCGGGAA  
 401 AAAGCAGAAT ATTTGATGGA AGAACGGGAG ATCCTTTTGA ACAACCTGTT ATAATAGGAA AGCCTTATAT CTTGAAATTA  
 481 ATTCATCAAG TTGATGATAA AATCCATGGA CGTCCAGTG GACATTATGC ACTTGTTACA CAACAACCCC TTAGAGGAAG  
 561 GGCCAAGCAA GGGGGACAAC GGGTAGGAGA AATGGAGGTT TGGGCTCTAG AGGGATTGG AGTTGCTCAT ATTTTACAAG  
 641 AGATGCTTAC TTATAAATCT GATCATATTA GAGCTCGCCA GGAAGTACTT GGTACTACGA TCATTGGAGG AACAAATCCT  
 721 AACCTGGAG GATGCTCCAG AATCTTTTCG ATTGCTCGTT CGAGAATAC GATCTTTGGC TCTRGATGA ATCATTTCCT  
 801 TGTATCTGAG AGAACTCCG ATTAATGGAA GGAAGCTTAT CGAATCGGAC TGAATCAAAA TTTTT

#### B1 (M13 reverse)

1 GATCCTATAG GATCTCCCC CTTCAAAATC GGACGTGAAA GTTTCCTCTC GTCCGGCTCA AGTAGTTACA CCAAATAAAG  
 81 ATAAAGAAAG GGGTCTGGC TTTCAAATC TAGAAAATCC TCAAACAAC TACTCCTTAC TCAAGTTCCC AGTGAAGACC  
 161 AAGCAACATT TCATTGATTC ATTCTTCCTT ATTATTTAGA TTTTCTGAAT TCTTTATTCA ATTACGACAT AAATGAAATA  
 241 TTATTGAGTA GTCTACTTCC CTTTCAATGA TGAATCCCT TAAACTTAAA TTAATAATTA AAGGAGTACC TTAGAATTCA  
 321 TAAGGGATTT ACTTGCTAT GTATCGTTT ATTTCGATCTT TTTAGGTCCT GACTTCGCCT CGACGGTTAT GCCACGATGC  
 401 CCTTAAAGCC TATATGCGAT GGATAGACTC CTGTAACCAT GACATATTTG TTTACTTGAA CATAAAAAA ATTCTTTCTA  
 481 AAAGAAAGAG AAAAGAGAAT GGTTAATTCC ACAAAGCAA AAGAAAGAAG TCTTTTTTCA CGAGGTACAA CTAAACATTC  
 561 CTATTTATTT GTTACGGAAT CGACCATAGA TCAATTACCC TTTTATTGG TAGTATAGAA TACACCCCAT AATTCTGAGC  
 641 TTCATGTTAC TCCTCTTAAG AGACATGTCA GAGCCGAGGC ATCCATTGGA TTGAATGGGA TGACAGTTTC TCATCTGAA  
 721 TCTGTAAATC AGAATTCGAT CAAATCCCAC ATCGCAATAT ACTAGGCCTT CTAATTCTTT AAGAGTTAT CTAAGATTT  
 801 GGGATTA

#### B2 (M13forward)

1 ATCCCAAATT GACGGGTTAG TGTGAGCTTA TCCATGCGGT TATGCACTCT TCGAATAGGA ATCCATTTTC TGAAAGATCC  
 81 TGACTTTCGT GCTTTGGTGG GTCTCCGAGA CCCTTTCGAT GACCTATGTT GTGTTGAAGG GATATCTATA TGATCCGATC  
 161 GATTGCGTAA AGCCCGCGT AGCAACGGAA CCGGGGAAAG TATACAGAAA AGACAGTTCT TTTCTATTAT ATTAGTATTA  
 241 GTTAGTGATC CCGGCTCAGT GAGCCCTTTC TTACGTGATG AACTGTTGGC ACCAGTCTTA CATTTTGTCT CTGTGGACCG  
 321 AGGAGAAAGG GGGCTCAGCG GGAAGAGGAT TGTACCATGA GAGAAGCAAG GAGGTCAACC TCTTCTCTTT CAAATATACA  
 401 ACATGGATTC TGACAATGCA ATGTAGTTGG ACTCTCATGT CGATCCGAAT GAATCATCTT TTCAACGGAG GTAATGCCTG  
 481 CTAGGTAAGA GGATAGCAAG TTACAAATTC TGTCTCGGAA GGAATTTGTC CATTTTTCGG GGTCTCAAAG GGGCGTGAA  
 561 ACACATAAGA ACTCTGAAT GGAAATGGAA AAGAGATGTA ACTCCAGTTC CTTCGGAAAT GGTAAGATCT TTGGCGCAAG  
 641 AAGAAGGGT TGATCCGTAT CATCTTGACT TGGTTCTGAT TCCTCTATTT TTAAGAATA CCGAGTCGGG TTCTTCTCCT

721 ACCCGTATCG AATAGAACAT GCTGAGCCAA TCTTCTTCAC TGCTTGCTTT AGATCGGGAA AATCGTACGG TTTTGAAC  
 801 CATGTGCTAT GGTCGATCC GTAGTCATCC TATTTCGAT GGGaCAG

**B2 (M13 reverse)**

1 GATCCTATCA ACTTGTTCCG ACCTAGGATA ATAAGCTCAT GAGCTTGGTC TFACTTCACC GTCGAGAAAC GAAAGAAGAC  
 81 TTCCATCTCC AAGTTTAACT CAGACGTAGC TCGCTTCTTT TTGGGTGTGA AGCAGTATCA AACCAAATA CCCAACAAGC  
 161 ATTAGCTCTC CCTGAAAAGG AGGTGATCCA GCCGCACCTT CCAGTACGGC TACCTTGTTA CGACTTCACT CCAGTCACTA  
 241 GCCCTGCCTT CGGCATCCCC CTCCTTGGCG TTAAGGTAAC GACTTCGGGC ATGGCCAGCT CCCATAGTGT GACGGGCGGT  
 321 GTGTACAAGG CCCGGGAACG AATTCACCGC CGTATGGCTG ACCGGCGATT ACTAGCGATT CCGGCTTCAT GCAGGCGAGT  
 401 TGCAGCCTGC AATCCGAAC T GAGGACGGGT TTTTGGAGTT AGCTCACCTT CGCGGGATCG CGACCCTTTG TCCCGGCCAT  
 481 TGTAGCACGT GTGTCGCCCA GGGCATAAGG GGCATGATGA CTTGACGTCA TCCTCACCTT CCTCCGGCTT ATCACCGGCA  
 561 GTCTGCTCAG GGTTCCAAC TCAACGGTGG CAACTAAACA CGAGGGTTCG GCTCGTTGCG GGACTTAAC CAACACCTTA  
 641 CGGCACGAGC TGACGACAGC CATGCACCAC CTGTGTCCGC GTTCCCGAAG GCACCCCTCT CTTTCAAGAG GATTTCGGGC  
 721 ATGTCAATCC CTGGTAAGT TCTTCGCTTT GCATCGAATT AACCACATGC TCCACCCTT GTGCGGGCCC CcGTCAAtTC  
 801 CTTTGATTT CATTCTTGG AAcGTACTCC CAGGGGGATC TTACGCGTTA CTAAGCACTG C

**B3 (M13 forward)**

1 GTAATCGACT CaCTATAGGG CGAATTGGGT ACCGGGCCCC CCCTCGAGGT CGACGGTATC GATAAGCTTG ATTACGATTT  
 81 AGGTGACACT ATAGAATACT CAAGCTATGC ATCCAACCGG TTGGGAGCTC TCCCATATGG TCGACCTGCA GCGGCGCGG  
 161 AATTCACTAG GGAATTCGA TTGATACTCA ATCATAAACC AACCTATGAG AATACTTTTT ATTACTATTT TAGTTTTAGT  
 241 TGAATAAAG AAATAAAAA AACTTAGAAA AGTAAAAGAC TACTAAGATA AATAAAAGAC TACTAATAA AGGAGCAATA  
 321 CCATCCCCCT TGATAAAGA GGAAATGGT TATTGCTCCT TTACATTCAA AAACCTCGTAA AACTAAGAC GGAAATCAAT  
 401 CACTAGTTCT AGAGCGGCCG CTTTACTTGT ACAGCTCGTC CATGCCGTGA GTGATCCCGG CGGCGGTCAC GAACTCCAGC  
 481 AGGACCATGT GATCGCGCTT CTCGTTGGGG TCTTTGCTCA GGGCGGACTG GGTGCTCAGG TAGTGGTTGT CGGGCAGCAG  
 561 CACGGGGCCG TCGCCGATGG GGGTGTCTG CTGGTAGTGG TCGGCGAGCT GCACGCTGCC GTCCTCGATG TTGTGGCGGA  
 641 TCTTGAAGTT CACCTTGATG CCGTCTTCT GCTTGTCCGC CATGATATAG ACGTTGTGGC TGTTGTAGTT GTACTCCAGC  
 721 TTGTGCCCCA GGATGTTGCC GTCCTCCTTG AAGTCGATGC CCTTCAGCTC GATGCGGTTT ACCAGGGtGT CGCCCTCGAA  
 801 CTTACCTCG GCGCGGGTCT TGTAATTGCC

**B3 (M13 reverse)**

1 AGCTCCACCG CGGAATGATA CAAGTCTTGG ATAAGAATCT ACAACGCACT AGAACGCCCT TGTTAGCAT CCTTTACTCC  
 81 GACAGCGTCT AGAGTTCCTC GAACAATGTG ATATCTCACA CCGGGTAAAT CCTTAACCTT TCCCCCTCTT ACTAAGACTA  
 161 CAGAATGTTC TTGTGAATTA TGGCCAATAC CAGGTATATA AGCAGTGATT TCAAATCCAG AGGTTAATCG TACTCTGGCA  
 241 ACTTTACGTA AGGCAGAGTT TGGTTTTTTG GGGGTGATAG TGGAAAAGTT GACAGATAAG TCACCCTTAC TGCCACTCTA  
 321 CAGAACCATA CATGAGATTT TCACCTCATA CGGCTCCTCG TTCAATTCTT TTGAAGTCAT TGGATCCTTT TCCTCGTTCCG  
 401 AGAATCTCCT CCCCTCTCC ACTCCGTCCT GAAGAGTAAC TAGGACCAAT TCAGTCACGT TTTTCATGTT CAATTGAACA  
 481 CTTTCGATTT TTGATTATTC TCAAATCAAA GGAGAAGATT ATTCTTTTTA CCAAACATAT GCGGATCAAA TCACGATCTT  
 561 ATAATAAGAA CAAGAGATCT TTCTCGATCA ATCCCTTTCG CCGAGAATCA GAAAGATCCT TTTCAAGTTT GAATTTGCTC  
 641 ATTTGGAATC TGGGTTCTTC TACTTCATTT TTATTTACTT ATTTATTtAT tATTTTGATT TCCCTCTCTT TTCTTTTTTt  
 721 ATCCCTTCCA TCATTCCTTA AGTCCCATAG GgTTGATCct GGTAGAATCT GACCCATTTT CTCATTGAGC GAaGGGTACG  
 801 AAATAATCaA ATGATTTTCG ATCAAAGT

**PS1 (M13 forward)**

1 TCGACGGAAT GCTCCTATTA CACTCGTAGT CTCTGAAGGA TGAGAACCAA CTATGTAGCA TCTACATCGA GAATTC AAGT  
81 ATTGTATACG CCATTAGTCC GATCCTTTGT AGGAACTACC CGTAATAACG AACTTGCAAA ATGGATCTGT TTATCATAAA  
161 GAGATTGCGT GTTCTTGACC CTGCTTCACC TTAATTGTTA TTTGAACAAG TAAAAGTTAT GTCTTGGTCC GAGTGGGGAT  
241 AGCATTCTC TTCTGCATGT CCATGGAGTT TTGAAAAATC CAAACATCTC AGAGATAGAT AGAGAGGTAG GAATTTATCG  
321 AACGAACCGC ACTCCTTCGT ATACGTCAGG AGTCCATTGA TGAGAAGGGG CTGGGGAAAG CTTGAACCCA ATTCTACAG  
401 TGATGAATAT AAGCGCAATT GAAATTCCTG GGGAGTTATA CATTTGTGTA TTGATAAGAC CATTCACTAT TTCTTGAAGC  
481 TCGATCTCTC CCCC GGATGA ACCATATAGC CAAGAGAAAC CATGAACCAG AATAGAAGAA CTTGCCCCAC CCATGAGTAA  
561 ATATTTGCGT GTAGCTCAT TAGACCGTAC ATCTTTCTTG GTATATCCAG ATAATAGGTA GGAGCATAAA CTGAAACATT  
641 CTGGAGCTAC AAAGATAGTT ATTAATCGT TAGCACC GCA TAAAAACATT CCTCCTAGAG TAGCTGT TAA TACGAATAAG  
721 AGAACTCTG TTATAGCCAT TTCTGTACAT TCAATGTACT CTACGGATAG AGGAATACAT AGAGTTGAAC ATAGTAAAAT  
801 AAGAAATTGA AAGATTCGT TGAAATTGTT CGTTTGAAA TTTCCGAAA GCTAATCATA GGTTCTTCTC TCCATCGGAA  
881 CAATAGGGCC GTTATGCTCA TTAATAACT tGTTGAAGAG ATGAAATATA ACCAAGGTAT ATCTTTTTTG ATCAGAGGtT  
961 GAATCGATCA TCAGAAGAAG ATTAGGCAA TAGGATACTT CTGGGAAATA AACTtCcTc GAAAGAGCAA TGAAGcTTTC  
1041 ATAAATCTC GTAgATCGAg ATGAGTTTca TT

**PS1 (M13 reverse)**

1 GAAGAAAAA CTGCAAAATC CGATCCAATT TATCGTAATC GATTAGTTAA CATGTTGGTT AACCGTATTC TGAAACACGG  
81 AAAAAATCA TTGGCTTATC AAATTATCTA TCGATCCGTG AAAAAGATTC AACAAAAGAC AGAAACAAAT CCaCtaTATC  
161 TgGTTTcaCG CCAAGCATGG GGAACCCCC TCCCCGATAT ATAGCAGTAA AAGCAAGACg TGTGGGGCG GATCGACTCA  
241 TCAAGTCCC ATTGAAATAG GATCCACACA AGGAAAAGCA CTGCCATTC GTTGGTTATT AGGGGCATCC CGAAAACGTC  
321 CGGGTCGAAA TATGGCTTTC AAATTAAGTT CCGAATTAGT GGATGCTGCC CAAGGGAGtG GcGATGCCCT TCGCAAAAAG  
401 GAAGAGACT

**PS2 (M13 forward)**

1 GTCGACTTTA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAAGC CTA CTGACT CCCTAACTAT TTATCCTATC  
81 CTATCCTCTC TTTGCAGAAA TTTTTTTTCT CTTATCACAA GTCTTGTTGGT ATATATATAT GATACGCGTA CAAACGAACA  
161 TCTTTGAGCA AGGAATCTCC ATTTGAATGA TTCACAGTAC ATATCATTAT TCGTACTGAA ACTTACAAAG TTTTTTTTTT  
241 TGAAGATCCA AGAAATCCA GGTCTGGAT AATACTTGT TTTGTAATAC AATACCCTT CGTCTTTTTA TTTTAAATG  
321 ACATAGACCA AGTCATCTAG TAAAATGAGG ATGATGTATC GGGAAATAGCC GGGATAGCTC AGCTGGTAGA GCAGAGGACT  
401 GAAAATCCTC GTGTACCAG TTCAAATCTG GTTCTGGCA CATGATTAAT TTGTCTGAGT ATCTATTCTA CAAATTCATT  
481 GATGTTTAGA TGAATCGATA TACATATTCA TTAATAGTCT AGATCATGAT ATATACTTAT CCCTCTAGGT GTCTGGAGAT  
561 ATACTCCTCC ATCTTTTAGA TAAGTAAAA GTATATAAAA TATGTAAGG AAAAGAATTC GATTATGTTT CCTCTTCTTT  
641 TTTATTTTT TTGTTTATAT TGTGTCTTGT CTA CTCTCAT TCAAAAAGAA TGTTAATACT TCATACATAC ATATTCGAAA  
721 GATTAATTA GTTAGTTGAA AGACCCAAAA GTATAGTCTA GAGTAGTTGA AGGATGGGAA TAGACAAGGT TCATCTCAGA  
801 TATAGTACAA ATATAATCCG ATCCCCTTTC ATTTCTTTG GATTTTCTTT CATATCTATT TTCTATTTCC TTCATTTCCC  
881 CCTACgTGA TTTTCTAGAG CCCATCTAAG TTATGTGCGC GGGACAAAGT TCATGATGCA gAACTCTTTT TCTTTTAGTT  
961 CATCCTATGG CTAGGCTCTA AGTATCTTCA AATTGaGAAT CTCAAtGaGC CTTTCTTTTT TTTTATTaTT TTA GCCatC  
1041 CTAAtCAA



**PS2 (M13 reverse)**

1 GTAGCATTCA AGTAATGCC TTTGATTTCA CCTGTTTCAG CCTGTGATTT AAAAATGGCT TCGGCACAAA ATAAGAAACG  
 81 GTCTCTCCAA CGCATAAATG GTTGGGAGTT CACGTTCTCA TCATCTTTAG TAAAATCAAG TCCACCGCGG AGACATTCAT  
 161 AAAGTCTCT ACCATAGTTC TTAGCGGATA ACCCCAATTT AGGTTTAATA GTACATCCCA ATAGGGGACG ACCATACTTG  
 241 TTCAATTTAT CTCTCTCAAC TTGGATGCCA TGAGGCGGGC CTTGGAAAGT TTTAGAATAA GCAGGGGGGA TTCGCAGATC  
 321 CTCTAGACGT AGAGCGCGCA GAGCTTTGAA CCCAAACACA TTACCCACAA TGGAAGTAAA CATGTTAGTA ACAGAGCCTT  
 401 CTTCAAAAAG GTCTAAAGGG TAAGCTACAT AAGCAATAAA TTGACTTTCT TCTCCAGCAA CGGGCTCGAT GTGGTAGCAT  
 481 CGTCCTTTGT AACGATCAAG GCTGGTAAGT CCATCAGTCC ACACAGTTGT CCATGTACCA GTAGAAGATT CAGCAGCTAC  
 561 CGCGGCCCT GCTTCTCGG GTGGAAGTCC AGGTTGAGGA GTTACTCGGA ATGCTGCCAA GATATCAGTA GGTTTGGTCT  
 641 CATATTCAGG AGTATAATAA GTCAATTTGT AATCTTTAAC ACCGGCTTTT GAATCCAACA CTTGCTTTAG TCTCTGGTTG  
 721 GGGGGGACAT AGTCCCTCCC TACAACCTCAC GA

**PS3 (T3)**

1 GGTACATGCG AAGAAATGAT CAAAAGGGCT GTATTTGCCA GAGAATTGGG AGTTCCTATC GTAATGCATG ACTACTTAAC  
 81 AGGGGGATTC ACCGCAAATA CTAGCTTGGC TCATTATTGC CGAGATAATG GCCTACTTCT TCACATCCAT CGTGCAATGC  
 161 ATGCAGTTAT TGATAGACAG AAGAATCATG GTATGCACCT TCGTGTACTA GCTAAAGCCT TACGCTGTGC TGGAGGAGAT  
 241 CATATTCACG CCGGTACCGT AGTAGGTAAG CTTGAAGGAG AAAGAGAGAT CACTTTGGGC TTTGTTGATT TATTACGTGA  
 321 TGATTTTGT GAAAAAGACC GAAGTCGCGG TATTTATTTT ACTCAAGATT GGGTCTCTCT ACCAGGTGTT CTGCCAGTGG  
 401 CTTCTGGGGG TATTCACGTT TGGCATATGC CTGCTCTGAC CGAGATCTTT GGAGATGATT CCGTACTACA GTTCGGTGGG  
 481 GGAACCTTAG GACACCTTGG GGGAAATGCA CCGGGTGGCG TAGCTAATCG AGTAGCTCTT GAAGCATGTG TACAAGCTCG  
 561 TAATGAGGGA CGTGATCTTG CTCGTGAGGG TAATGAAATT ATCCGTGCAG CTAGCAAATG GAGTCCTGAA CTAGCTGCTG  
 641 CTTGTGAAGT ATGGAAGGAA ATCAAATTTG AATTCCCAGC AATGGATACT TTGTAATCCA GTAATTATGT TCGTTTCTTA  
 721 AATTGAATTG CAATTAACCTC GGCCCATCTT TTACTAAA

**PS3 (T7)**

1 GTCGACTCTC GATCTAATAG TTCTATCAAG CATGAGTTCT ATTACAAGGT ATCGAGCCTA TGAATCTATT CTATGTTTTT  
 81 TATTTGCGAT TCAGCAGTTA GTCCTTGAAT CTAGAAAGAA TACAGAAATA GAATAAGAGT CCACTTCGAA TGAAGAAATA  
 161 ATAAAAAATA TTGTTTCCAG ATATCTCCGC AATTGGTCAA ATTCTAAGCA CCCCTTTCGA TGAATGCCCA AGAATGCCCG  
 241 AAAGAACCGC TTTTTTAATG AATATTATTC GGATTTCAAG ACGAAAGAAC TCCCTTTCGA CCATTTCTAT CAATATATTT  
 321 AATATTTTCGA AAAACTTTTG CTGGCTACCG GAATAATGTA GAGAAATTTT TGAAGAATAT TGTGATGGGC TGAAGTGGCA  
 401 AAACAAGATA GAAATTTGAT AGTTATCATT ATAAGAGGTT TAGTCATTAA GGAACACAAA AGAAAGGATA AAAAGAAACG  
 481 CCGATTGACA AAAAAGAGAG AGAATTTACA AGATATGATC TATCTGCATC TAGCGTATCA ATTCCAAATA TTGGACCTAA  
 561 AATAAATTCT TTATTTTCGAA ATGTTTTGAT ATATGGGAAT TTTTCGATTT GTTACTTGCT TTGTTACTGT GAGTAGATTT  
 641 CCATACGCAT AAAAGACCAT TTTTGGCGAC AAAAACAGTT TCCTTTTATG ATCGTATACG TCTGCTTTGG CTCAAATGAG  
 721 CGAACCTTCA GTTAGGTTAT GCTATAGAAA AgGAAGGGCG GATTCTTGAA TTTTCTCTCC TGCTAAGAAA TCAGTTCATT  
 801 CAAATACTtC ATCGGTACaC GCAAGAAATA GGCTATTTCAG CAGCTTTTGT TCCATT

**PS4 (T3)**

1 GGCATGCAAG CTTTAAACAAT TTATGGACTG GTTGTAGCAT TAGCACTTTT ATTTGCGAAT CCTTTTGTTT AATTGTTTAA  
 81 TCCTAGAAAT ATGAAAAATA CGTATTTTTC ATATTTTCTT ATTTTATTC CTGGACCTG TCGCTTGCTT TTTAGAATTA  
 161 TATCAAGATT GAACTACGAC AATTACTTAT TCGTTGAGAT AATAACCCAT GGAAGGACT GATTTGAGGA TGAGGAAGAT  
 241 TTGAGGATGA GGAATTAGCA AATCGACTCG CTTTCTTCC TCCCTTCCCG TTCTTAGTCC AATGAAAACC TTTTTTTTAG  
 321 TAAGTGTTGG AACAAACGAA GTATTCGTA ATTGACATGA GACTCGGTAT CTCATTTTTC ATTCTTATTG GAAATTTCAA  
 401 TTGAAAAAT AAAATCCAGT TATAAATTAT AAATCGAATA TTTTTTTTTC TGTTTAGAAA TAGAGAAATT AATAAAAAAG  
 481 AAAATAAAAA ATAAGGGGTG AAGTGATAGA AAAAGAAGCT TGTTGATTT TTTTAGTCTA TCTATTTTAG TCTATCTATA  
 561 AGAGGAGATC ATATGAAAA TGTAACCGAT TCTTTCGTTT CCTTGGGTCA CTGGCCATCC GCCGGGGGT TCGGGTTTAA  
 641 TACCGATATT TTAGACCAA ATCCAWTAAA TCTAAGTGTA GtGCTTGGTG TATTGATCTT TTTTGAAAAG GGAGTGTGTG  
 721 CGAGTTGTTT ATTCCAGAAT AGGCTGGACT C

**PS4 (T7)**

1 GTCGACTATA ACCCCTAGCC TTCCAAGCTA ACGATGCGGG TTCGATTCCC GCTACCCGCT CAATATTATA CTCTATATTA  
 81 TTTAGAATGA AAATGAAATT AATGCATCAT TGAATTGAAT ATGCAATCC CGAAATCTC TCACACACAA TCCGATTCTT  
 161 TTTTCCGAA CAAGAGATAG GAAAGCAAAA ATGCGAAAAG AAAAAAAT CGGAATGAAA AGCGTCCATT GTCTAATGGA  
 241 TAGGACAGAG GTCTTCTAAA CCTTTGGTAT AGGTTCAAAT CCTATTGGAC GCAACTTATT TCCATATATT CTATTCTTTT  
 321 ATCAGTAAAG AAATTTGAA TGATTTAAAT CAGAGACGCT TTTATTTATT ATTTATTTAA TTTTCTTAAT ATTTAATATA  
 401 AAGGATTACC TTTTTATTAT TAAATATTCA TTTTTTTTAT TTATTATTAA ATTCATTAAT TAGACTAAGA ATTAAGAGTG  
 481 ATCAATTTCT TTATGCTGTG TCCTGAAGTA GAAAGCGTTC CATCTGTTCC TGAATAGTTT CTTTCAAAAAG GGTTTCTGCT  
 561 TCCTCGGTGA ATGTCTTGGT AGAAGATATG ATTTCTTGGT ACTGAGGTTT ATTCGTTTTT AAGTAATTAC GTAACCTCAGC  
 641 AAGAAATTC CTACCTGTC CAATTTCTAA TGAATCAGAT AACCATTTGT TCCAGTATAA ATAGTCATTA TCTGTCTTTC  
 721 CCCGTAGGAG GGGCTGCTGG GGATGTTTGA GCAATCACGT AATCGTTGAC CTCTGCCAAT GGATCTGAGT GCTTTTCAGA  
 801 TAGAAGAGAA TgGCAAgGC TCTATCTGCG AATGCCAGT CCATTTGATT G

**PS5 (T3)**

1 GTGTCCGAGT AGATACTGTT ACTTCTCTC GAACCATAGT AATATTACTT GATCAGATCA TTGAATCATT TATTTCTCTT  
 81 GAAATCTCTT CAATGTTTAT TTCTACACAG TTTATTCCTA TACACGCTT TTTTATAGGAG GTCTACAGCC GTTATGTGGC  
 161 ATAGGAGTTA CATCCCGTAC GAACTTAAT AGTATACCAC TTCTACGAAT AGCTCGTAAT GCTGCATCTC TTCCGAGACC  
 241 AGGACCCTTT ATCATGACTT CTGCTCGTTG CATACTTGA TCCACTACTG TACGAATAGC ATTTCTGCT GCGGTTGAG  
 321 CAGCAAATGG CGTCCCTCTT CTTGTACCC TGAATCCACA AGTACCGGCC GAGGACCAAG AAACCACCCG ACCCCGTACA  
 401 TCTGTAACGG TCACAATGGT ATTGTTAAAA CTTGCTTGAA CATGAATAAC TCCCTTTGGT ATTCTACGCG CACTCTTACG  
 481 TGAACCAATA CGTCCATTCC TACGTGAACC AATTCTTGGT ATAGGTTTTG CCATATTTTA TCATCTCATA AATATGAGTA  
 561 AGAGATATAT GGATATATCC ATTTTCATGTC AAAACATGAT TTTTTTTTAT TTGTACATCG GGTTCTTTAG AGAATCTCTT  
 641 TTCGAGAAAT TATCCTTGGC CTTTGTATG TCTCGGGTGG GACAAATTAC TATAATTCGT CCCCGTCTAC GGATCAGTCG  
 721 ACCTCGAGGG GGGGCCCGG ACCCAATTG CCCTATAGTG AGTCGTATTA CAATTCAGTG GCGGCGGTTT TACAACGTCG  
 801 TGACTGGGAA AACCTGGCGT TACCCAATT AATCGCCTTG CAGCACATCC CCCTTTCGCC AGGTGGGGTA ATAC

**PS5 (T7)**

1 GTCGACTGAT CCGTAGACGG GGACGAATTA TAGTAATTTG TCCCAACCCG AGACATAAAC AAAGACAAGG ATAATTTCTC  
 81 GAAAAGAGAT TCTCTAAAGA ACCCGATGTA CAAATAAAAA AAAATCATGT TTTGACATGA AATGGATATA TCCATATATC  
 161 TCTTACTCAT ATTTATGAGA TGATAAAATA TGGCAAAACC TATACCAAGA ATTGGTTCAC GTAGGAATGG ACGTATTGGT  
 241 TCACGTAAGA GTGCGCGTAG AATACCAAAG GGAGTTATTC ATGTTCAAGC AAGTTTTAAC AATACCATTG TGACCGTTAC  
 321 AGATGTACGG GGTCGGGTGG TTTCTTGGTC CTCGGCCGGT ACTTGTGGAT TCAGGGGTAC AAGAAGAGGG ACGCCATTTG  
 401 CTGCTCAAAC CGCAGCAGGA AATGCTATTC GTACAGTAGT GGATCAAGGT ATGCAACGAG CAGAAGTCAT GATAAAGGGT  
 481 CCTGGTCTCG GAAGAGATGC AGCATTACGA GCTATTCGTA GAAGTGGTAT ACTATTAAGT TTCGTACGGG ATGTAACTCC  
 561 TATGCCACAT AACGGCTGTA GACCTCCTAA AAAAAGACGT GTATAGGAAT AAAGTGTGTA GAAATAAACA TTGAAGAGAT  
 641 TTCAAGAGAA ATAAATGATT CAATGATCTG ATCAAGTAAT ATTACTATGG TTCGAGAGAA AGTAACAGTA TCTACTCGGA  
 721 CACTGCAGCC CGGGGGGATC CACTAGTTCT AGAGCGGCCG CCACGCGGTG GAGCTCCAGC TTTTGTTCCT TTTAGTGAGG  
 801 GTAATTCGAG CTGGGCGTAT CATGGTCATA GCTGTT

**PS5 Complete Sequence**

1 GTCGACTGAT CCGTAGACGG GGACGAATTA TAGTAATTTG TCCCAACCCG AGACATAAAC AAAGGCAAG GATAATTTCT  
 81 CGAAAAGAGA TTCTCTAAAG AACCCGATGT ACAAATAAAA AAAAATCATG TTTTGACATG AAATGGATAT ATCCATATAT  
 161 CTCTTACTCA TATTTATGAG ATGATAAAAT ATGGCAAAAC CTATACCAAG AATTGGTTCA CGTAGGAATG GACGTATTGG  
 241 TTCACGTAAG AGTGCAGGTA GAATACCAAA GGGAGTTATT CATGTTCAAG CAAGTTTTAA CAATACCATT GTGACCGTTA  
 321 CAGATGTACG GGGTCGGGTG GTTTCTTGGT CCTCGGCCGG TACTTGTGGA TTCAGGGGTA CAAGAAGAGG GACGCCATTT  
 401 GCTGCTCAA CCGCAGCAGG AAATGCTATT CGTACAGTAG TGGATCAAGG TATGCAACGA GCAGAAGTCA TGATAAAGGG  
 481 TCCTGGTCTC GGAAGAGATG CAGCATTACG AGCTATTCGT AGAAGTGGTA TACTATTAAG TTTTCGTACGG GATGTAACTC  
 561 CTATGCCACA TAACGGCTGT AGACCTCCTA AAAAAGACG TGTATAGGAA TAAACTGTGT AGAAATAAAC ATTGAAGAGA  
 641 TTCAAGAGA AATAAATGAT TCAATGATCT GATCAAGTAA TATTACTATG GTTCGAGAGA AAGTAACAGT ATCTACTCGG  
 721 ACAC

**PS6 (T3)**

1 GAGATCTAAC TCCTTCTATT TTGATTCATA GTTGTAAGTT CCTACGACAT AATAGATCGG TTTTGGAGAA AGGAGAATAA  
 81 TCTTTATTCT TTGATTTTAA AGAGACATTA TCAATAACCA AACAAATAGA GAAAAGCCAG CTATCGGAAT CGAACCAGT  
 161 ACCATCGCAT TACAAATGCG ATGCTCTAAC CTCTGAGCTA AGCGGGCTCA CCAGAAATG TACATGCATA GAAATTCAGT  
 241 AAAGTCTGG TATCTTAGCT ATTAACATTT CCTTCTTAGC TAGTCACAAT TAATATGAAT ATATAAAGA ATAAGAATAT  
 321 AGATTCAAAA AAAATTATTG ATTATTAACG ATTATTAGAT AGAGCAATAT AGAATTCCGA TCTATTTTCA ATTATATGTT  
 401 TAAAAATATC CATATCTTAT TAGATAGGA

**PS6 (T7)**

1 GTCGACGGAT TTTCTCTTA CTATAAATTT CATTGTTGTC GGTATTGACA TGTAGAATGG GACTCTATCT TTATTCTCGT  
 81 CCGATTAATC AGTTCCTCAA AAGATCTATC GGACTATGGA GTGAATGATT TGATCAATGA ATATTGATT CTTTTCTTTC  
 161 TTCAACTTGG AATCGATTTA CAACAATTCT TTCATTTTTC ATATCAAAAA ATACAGATTC GAGCCATCAT TAATCATTG  
 241 ATATAATATT TCAGTACCTA TGTATATAGG TTTATCTTTC ATCTTTCTG GTTCTGGAGT TTTGATGGAA GGATTCGTTT  
 321 AACACGCAA CATAGTCAAC TCCATTTGTT TTTGTTAGAA CAGCTTCCAT TGAGTCTCTG CACCTATCCT TTTTTTATTA  
 401 TCGTTTTCTG AACCTTGGT TGTTCCTGGA AACAGGATT TGGCTCAGGA TTGCCATTT TTTATTCCAG GGTTCCTCGT  
 481 AATTTGAAAG TTATCACTTA GTAGGTTTCC ATATTAAGGC TCAATCCAAT TAAGTCCGTA GCGTCTACCA ATTTCCGCAT

```

561 ATCCCCCCTT ATTTTGT TTT GAGATTGAA TCTCTTTGTG ATGCCGCTCC CCTTTTCCTT TCATTTATGC TTATGCTGGA
641 ACCTTTGAAT TCATTAGATA CCGATTCTTA AAAAGTCCTT CCTCCTTTTT TTCTTGCCTA TCTTCTTTCA TCTCATCTCT
721 CTTGGGGACC ATATTTGTAT TTGGTCCAAT CAGAAATGAT CCTATCATTC TGTATCTGCA ATTCAATATA CATGAAATAT
801 ATACCACACA TATATGCTTC CTTTCT

```

**PS7 (T3)**

```

1 GCTGCACGAG CCCTTCCCGC GTGCCACAAA TGACCTACGA ATAGGAAGAA TCCTAGAACA AAATGAGAAG TAGCTAACCA
81 ACTTCTAGGA GAGACATAAT TGACTGCATT GATCTCGGTA GCTACGCCAC CCACGGAATT TAAAGAACCT AAAGGAGCGT
161 GAGTCATATA TTCCGCGGAA CGCCGTTCTT GCCAAGGTTG TATGTCTTTT TTCAGCCTAC TCAAGTCCAA ACCATTGGGA
241 CCCCTTAGAG GTTCTAACCA GGGAGCACGC AGATCCCAAA AACGCATAGT TTCTCCTCCA AAAATGACTT CTCCAGTCGG
321 GGAACGCATT AGATATTTAC CTAACCAGT AGGTCCTTGA GCGGATCCCA CGTTAGCCCC AAGACGCTGG TCTCTAACTA
401 GAAAAGTAAA TGCTTGAGCT TGAGACGCTT CGGGTCCAGT TGGCCCGTAA AACTCACTAG GATAAGCGGT ATTATTGAAC
481 CAGACAAAGC AACAAGCAAT GAAACCAAAG ACAGCTAAAG CACCTAAACT ATAAGACAAG TAAGCCTCTC CAGACCATAC
561 AAGTGCACGG CGAGCCCACG CAAAGGGTTT GGTAAAGATA TGCCAGATTC CACCAAATAT ACAAATGGAA CCTAACCAT
641 CATGTCTCC AATTATATCT TCCAAATCGT CCACACTAAC CATTCCATCC TTCTCCTCCA AGGGCGATTT TAGTAATACC
721 AATATAACAC TTGGGCTAaG GTCAGGTGGT AATTTCTTA CATCTCCCTT CCGGGAGCCC AGGTATCATA TACGCCGCCA
801 AATAAGAGCC TGAATCTAG AagAAAGCCC TATCCTAA

```

**PS7 (T7)**

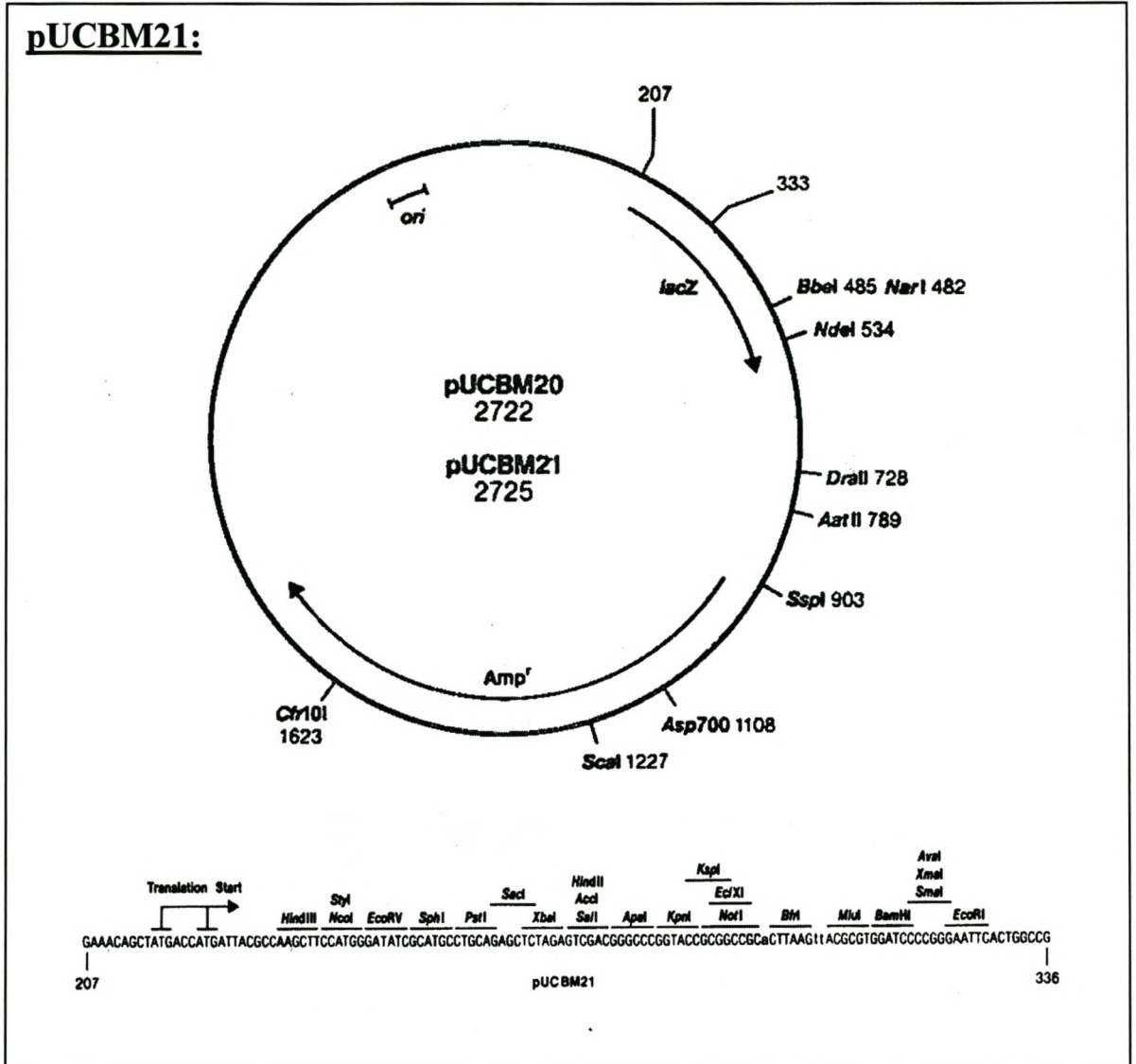
```

1 GTCGACCTAT CCTTCTAAT TCACATCTTT GTTAAAGAA TTTCTTTTGT AATTCCTTAC ATAATGATTC AGAAAATACC
81 GGATCCCCAC CTACACAAGC AAATTGTTGA TAAACTCCA AAATGGCACT TTCTTTTGAC CCAATTTTTT TTTTCTCCTT
161 ATCATTGAGG AAAGACAAAA AAATCTCAGG ATAGCAAACA TTCTCTAGAA TTTCTCTTAG ATTCTGAACC ATAGCTGATG
241 ATAGAACTAG AATAGATATT TTTTGTTC TACTCACACG AGCCCATATC CTTGCTTTTC TATCAATCTC TAATCTAAT
321 CTTCTCCCC AATCTGAGAC TATGGTGCCG GTATAGACCG AAATTCGTT ATGGTCCAAT TCTGACCGGT AATAAATGCC
401 GGGGCTTTGC AATATTTGAT TAATCACAAT TCTGTATAGT CCATTTACTA TAGAAGTCC CAGGGAATTC ATTAGAGGAA
481 TGTTTCCAAT AAAAATTGTT TGTTCTTGCA TATCTCTGCT GGTTTTCCAA ATTAATCCCG CAGATACATA TAATCAGAA
561 GAATATGTGA GTGATTCATA CACGGCATCT CTTTCTTTA TCAATGGTTC CACCAATTGA TATGTTTCCA CAAATAATTG
641 AAATTC AATT TCTTGATCTG TATCTTCAAT TTTTGAAAC TTATAAGGTT CTTCTGTCAA GCCCTGATCA ATAAACCTAC
721 AAAATCCTTC AAATGGGATC TGATTAAACC CAGGTATTGT AGACATTCCC TCATTTCCAC CCGGAACATT TTTGGTTCCC
801 ATTATCGAAA AATCCCATTA TGTATGGGAT GGGTCGTCTT CATCGAATCA TATGGATCGA TCGATCGAGA TCTAGCATGA
881 TG

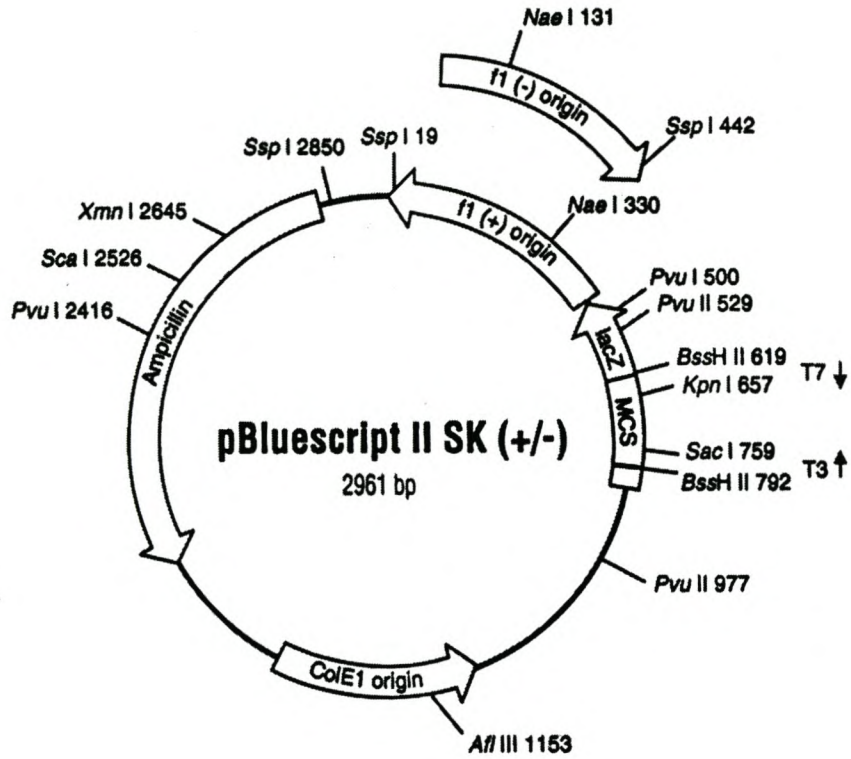
```

## Addendum 4

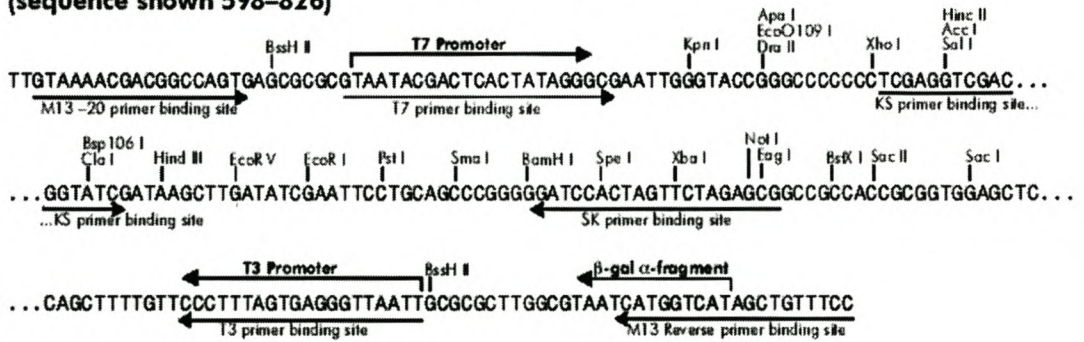
### Vector Maps

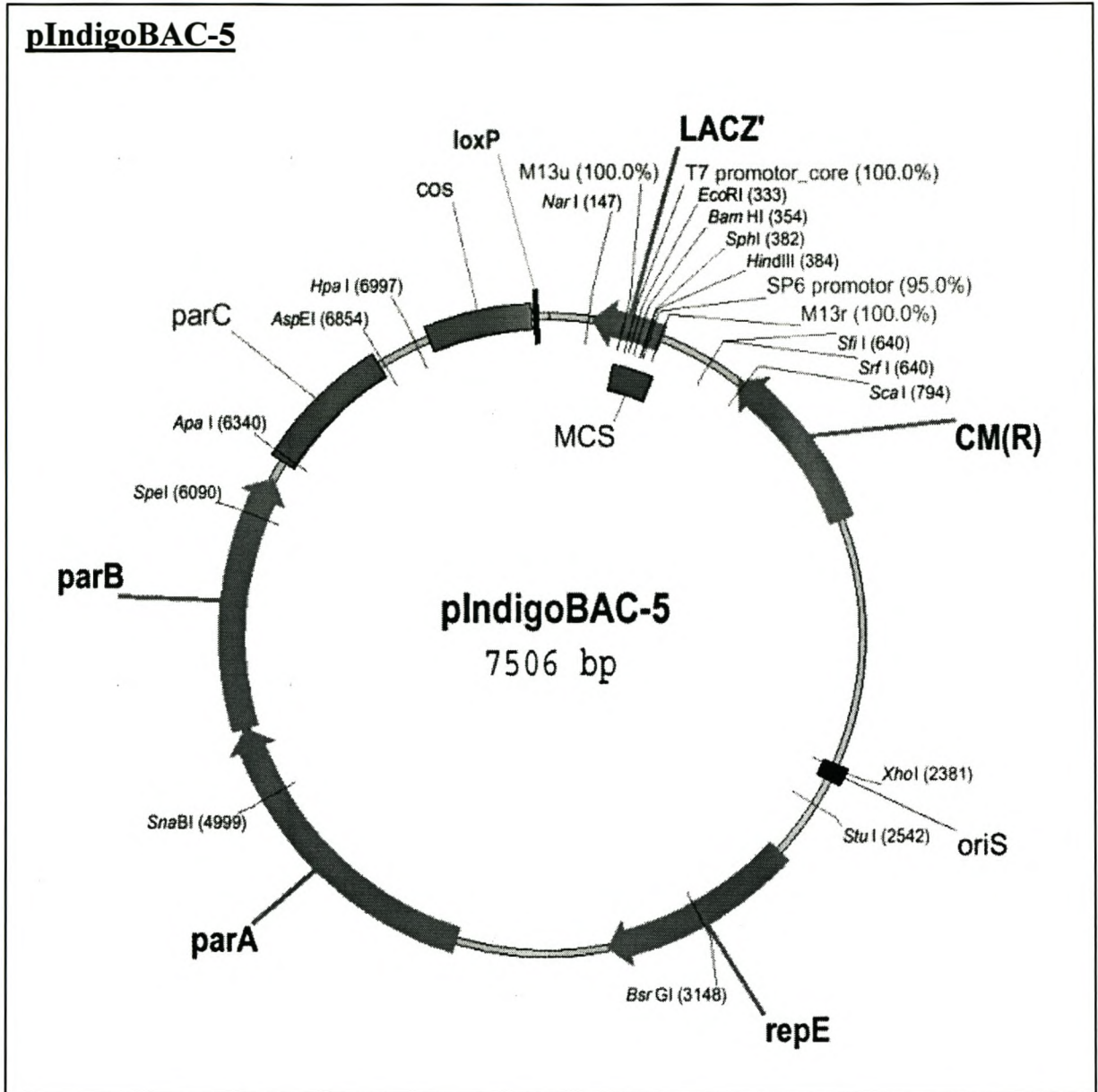


**pBluescript II SK (+)**



**pBluescript II SK (+/-) Multiple Cloning Site Region  
(sequence shown 598-826)**





## REFERENCES

- Abdallah, F., F. Salamini and D. Leister. 2000. A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. *Trends in Plant Science* 5: 141-142
- Ablett, E.M., G.G. Seaton, K.D. Scott, D.A. Shelton, L.S. Lee, G-H. Miao, M. Dolan, M. Hanefey, S.V. Tingey, R.J. Henry. 2001. The grape genome project. *Plant and Animal IX Genome Conference*, California
- Adams, M. D., S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, R.A. George, S.E. Lewis, S. Richards, M. Ashburner, S.N. Henderson, G.G. Sutton, J.R. Wortman, M.D. Yandell, Q. Zhang, L.X. Chen, R.C. Brandon, Y-H.C. Rogers, R.G. Blazej, M. Champe, B.D. Pfeiffer, K.H. Wan, C. Doyle, E.G. Baxter, G. Helt, C.R. Nelson, G.L. G. Miklos, J.F. Abril, A. Agbayani, H-J. An, C. Andrews-Pfannkoch, D. Baldwin, R.M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E.M. Beasley, K.Y. Beeson, P.V. Benos, B.P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M.R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K.C. Burtis, D.A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J.M. Cherry, S. Cawley, C. Dahlke, L.B. Davenport, P. Davies, B.de Pablos, A. Delcher, Z. Deng, A. D. Mays, I. Dew, S.M. Dietz, K. Dodson, L.E. Doup, M. Downes, S. Dugan-Rocha, B.C. Dunkov, P. Dunn, K.J. Durbin, C.C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A.E. Gabrielian, N.S. Garg, W.M. Gelbart, K. Glasser, A. Glodek, F. Gong, J.H. Gorrell, Z. Gu, P. Guan, M. Harris, N.L. Harris, D. Harvey, T.J. Heiman, J.R. Hernandez, J. Houck, D. Hostin, K.A. Houston, T.J. Howland, M-H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G.H. Karpen, Z. Ke, J.A. Kennison, K.A. Ketchum, B.E. Kimmel, C.D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A.A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T.C. McIntosh, M.P. McLeod, D. McPherson, G. Merkulov, Natalia V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S.M. Mount, M. Moy, B. Murphy, L. Murphy, D.M. Muzny, D.L. Nelson, D.R. Nelson, K.A. Nelson, K. Nixon, D.R. Nusskern, J.M. Pacleb, M. Palazzolo, G.S. Pittman, S. Pan, J. Pollard, V. Puri, M.G. Reese, K. Reinert, K. Remington, R.D. C. Saunders, F. Scheeler, H. Shen, B.C. Shue, I. Sidén-Kiamos, M. Simpson, M.P. Skupski, T. Smith, E. Spier, A.C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A.H. Wang, X. Wang, Z-Y. Wang, D.A. Wassarman, G.M. Weinstock, J. Weissenbach, S.M. Williams, T. Woodage, K.C. Worley, D. Wu, S. Yang, Q.A. Yao, J. Ye, R-F. Yeh, J.S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X.H. Zheng, F.N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H.O. Smith, R.A. Gibbs, E.W. Myers, G.M. Rubin, and J.C. Venter. 2000. The genome sequence of the *Drosophila melanogaster*. *Science* 287: 2185-2195
- Alleweldt, G., P. Spiegel-Roy and B. Reisch. 1991. Grapes (*Vitis*). *Acta Horticulturae. (ISHS)* 290: 291-330
- Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-3402.
- Angiosperm Phylogeny Group. 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 85: 531-553



- APG II. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141(4): 399-436
- Archibald, J.M. and P.J. Keeling. 2002. Recycled plastids: a 'green movement' in eukaryotic evolution. *Trends in Genetics* 18(11): 577-584
- Bedbrook, J.R. and L. Bogorad. 1976. Endonuclease recognition sites mapped on *Zea mays* chloroplast DNA. *Proceeds of the Natural Academy of Science USA* 73: 4309-4319
- Bedbrook, J.R., R. Kolodner and L. Bogorad. 1977. *Zea mays* chloroplast ribosomal RNA genes are part of a 22,000 base pair inverted repeat. *Cell* 11: 739-749
- Bhattacharya, D. and L. Medlin. 1995. The phylogeny of plastids: a review based on comparisons of small subunit ribosomal RNA coding regions. *Journal of Phycology* 31:489-498
- Bhattacharya, D. and L. Medlin. 1998. Algal phylogeny and the origin of land plants. *Plant Physiology* 116: 9-15
- Birky, C.W., Jr. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. *Proceeds of the Natural Academy of Science USA* 92: 11331 – 11338
- Birky, C.W., Jr. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanism, and models. *Annual Review in Genetics* 35: 125-148
- Blanchard, J.L. and J. Hicks. 1999. The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. *Journal of Eukaryotic Microbiology* 46: 367-375
- Bock, R. and R. Hagemann. 2000. Extranuclear inheritance: plastid genetics: manipulation of plastid genomes and biotechnological applications. *Progress in Botany* 61: 76-90
- Bogorad, L. 2000. Engineering chloroplasts: an alternative site for foreign genes, proteins, reactions and products. *TIBTECH* 18: 257-263
- Bookjans, G., B.M. Stummann and K.W. Henningsen. 1984. Preparation of chloroplast DNA from pea plastids isolated in a medium of high ionic strength. *Analytical Biochemistry* 141: 244-247
- Borsch, T., K.W. Hilu, D. Quandt, V. Wilde, C. Neinhuis and W. Barthlott. 2003. Noncoding plastid trnT-trnF sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16(4): 558-76.
- Boubals, D. 2000. Grapevine genetics and breeding facing the challenges of the 3<sup>rd</sup> millennium. *ISHS Acta Horticulturae* 528: VII International Symposium on Grapevine Genetics and Breeding.
- Bovenberg, W.A., A.J. Kool and H.J.J. Nijkamp. 1981. Isolation, characterization and restriction endonuclease mapping of the *Petunia hybrida* chloroplast DNA. *Nucleic Acids Research* 9(3): 503-517

- Bowe, L.M., G. Coat and C.W. dePamphilis. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceeds of the Natural Academy of Science USA* 97(8): 4092-4097
- Bowes, B.G. 1996. Ch 2: The plant cell. *A colour atlas of plant structure*. Mason Publishing, London UK, p26-28.
- Boynton, J.E., N.W. Gillham, E.H. Harris, J.P. Hosler, A.M. Johnson, A.R. Jones, B.L. Randolph-Anderson, D. Robertson, T.M. Klein, K.B. Shark and J.C. Sanford. 1988. Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science* 240: 1534-1538
- Bruce, B.D. 2000. Chloroplast transit peptides: structure, function and evolution. *Trends in Cell Biology* 10:440-447
- Burke, D.T., G.F. Carle and M.V. Olson. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236: 806-812
- Cavalier-Smith, T. 2000. Membrane heredity and early chloroplast evolution. *Trends in Plant Science* 5(4): 174-182
- Chapman, R.L. and D.A. Waters. 2002. Green algae and land plants – an answer at last? *Journal of Phycology* 38: 237-240
- Chaw, S.-M., C.L. Parkinson, Y. Cheng, T.M. Vincent and J.D. Palmer. 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proceeds of the Natural Academy of Science USA* 97(8): 4086-4091
- Chun, E.H.L., M.H. Vaughan and A. Rich. 1963. The isolation and characterization of DNA associated with chloroplast preparations. *Journal of Molecular Biology* 7: 130-141
- Cipriani, G., R. Testolin and R. Gardner. 1998. Restriction-site variation of PCR-amplified chloroplast DNA regions and its implication for the evolution and taxonomy of *Actinidia*. *Theoretical and Applied Genetics* 96: 389-396
- Cooper, G.M. 2000. Ch 10: Bioenergetics and metabolism. *The Cell: a molecular approach*. ASM Press, USA, p 402-409
- Corneille, S., K. Lutz, Z. Svab and P. Maliga. 2001. Efficient elimination of selectable marker genes from the plastid genome by the CRE-lox site-specific recombination system. *Plant Journal* 27: 171-178
- Cosner, M.E., R.K. Jansen, J.D. Palmer and S.R. Downie. 1997. The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Current Genetics* 31: 419-429
- Daniell, H. and A. Dhingra. 2002. Multigene engineering: dawn of an exciting new era in biotechnology. *Current Opinion in Biotechnology* 13: 136-141

- Daniell, H., M.S. Khan and L. Allison. 2002. Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. *Trends in Plant Science* 7(2): 84-91
- Daniell, H., S.B. Lee, T. Panchal and P.O. Wiebe. 2001. Expression of the native cholera toxin B subunit gene and assembly as functional oligomers in transgenic tobacco chloroplasts. *Journal of Molecular Biology* 311(5): 1001-1009
- Daniell, H. and B.A. McFadden. 1987. Uptake and expression of bacterial and cyanobacterial genes by isolated cucumber etioplasts. *Proceeds of the National Academy of Science USA* 84: 6349-6353
- Daniell, H., S. Varma, C.N. Stewart Jr. and C.S. Prakash. 1998. Chloroplast-transgenic plants: panacea – no! gene containment – yes! *Nature Biotechnology* 16(7): 602
- DeGray, G., K. Rajasekaran, F. Smith, J. Sanford and H. Daniell. 2001. Expression of an antimicrobial peptide via the chloroplast genome to control phytopathogenic bacteria and fungi. *Plant Physiology* 127: 852-862
- De Las Rivas, J., J.J. Lozano and A.R. Ortiz. 2002. Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. *Genome Research* 12: 567-583
- Delwiche, C.F. 1999. Tracing the thread of plastid diversity through the tapestry of life. *The American Naturalist* 154: S164-S177
- Delwiche, C.F., K.G. Karol, M.T. Cimino and K.J. Sytsma. 2002. Phylogeny of the genus *Coleochaete* (Charophyceae, Chlorophyta) and related taxa inferred by analysis of the chloroplast gene *rbcL*. *Journal of Phycology* 38: 394-403
- Douglas, S.E. and S.L. Penny. 1999. The plastid genome of the cryptophyte alga, *Gullardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. *Journal of Molecular Evolution* 48:236-244
- Douglas, S., S. Zauner, M. Fraunholz, M. Beaton, S. Penny, L.-T. Deng, X. Wu, M. Reith, T. Cavalier-Smith and U. Maier. 2001. The highly reduced genome of an enslaved algal nucleus. *Nature* 410: 1091-1096
- Edwards, D., J.G. Duckett and J.B. Richardson. 1995. Hepatic characters in the earliest land plants. *Nature* 374: 635-636
- Edwards, K., C. Johnstone and C. Thompson. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research* 19: 1349.
- Fast, N.M., J.C. Kissinger, D.S. Roos and P.J. Keeling. 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Molecular Biology and Evolution* 18(3): 418-426
- Ferris, C., R.P. Oliver, A.J. Davy and G.M. Hewitt. 1995. Using chloroplast DNA to trace postglacial migration routes of oaks into Britain. *Molecular Evolution* 4: 731-738
- Frey, J.E. 1998. Genetic flexibility of plant chloroplasts. *Nature* 398: 115-116

- Fujii, N., K. Ueda, Y. Watano, and T. Shimizu. 1997. Intraspecific sequence variation of chloroplast DNA in *Pedicularis chamissonis* Steven (Scrophulariaceae) and geographic structuring of the Japanese "alpine" plants. *Journal of Plant Research* 110: 195-207
- Garbary, D.J., K.S. Renzaglia and J.G. Duckett. 1993. The phylogeny of land plants: a cladistic analysis based on male gametogenesis. *Plant Systematics and Evolution* 188: 237-269
- Garcia, R.A., F. Lefort, M.T. de Andres, J. Ibaneaz, J. Borrego, N. Jouve, F. Cabello, and J.M. Martinez-Zapater. 2002. Chloroplast microsatellite polymorphisms in *Vitis* species. *Genome* 45(6): 1142-1149
- Garcia, R.A., L. Ruiz-Garcia, S. Agaoglu, R. Botta, F. Cabello and J. Cenis. 2002. The analysis of *Vitis* chloroplast genome polymorphisms around the Mediterranean Sea provide clues to understand grapevine domestication. *Plant and Animal Genome X Conference, California, P01\_47*.
- Gardner, M.J., N. Hall, E. Fung, O. White, M. Berriman, R.W. Hyman, J.M. Carlton, A. Pain, K.E. Nelson, S. Bowman, I.T. Paulsen, K. James, J.A. Eisen, K. Rutherford, S.L. Salzberg, A. Craig, S. Kyes, M.S. Chan, V. Nene, S.J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Perteu, J. Allen, J. Selengut, D. Haft, M.W. Mather, A.B. Vaidya, D.M. Martin, A.H. Fairlamb, M.J. Fraunholz, D.S. Roos, S.A. Ralph, G.I. McFadden, L.M. Cummings, G.M. Subramanian, C. Mungall, J.C. Venter, D.J. Carucci, S.L. Hoffman, C. Newbold, R.W. Davis, C.M. Fraser, and B. Barrell. 2002. Genome sequence of the human malarial parasite *Plasmodium falciparum*. *Nature* 419(6906):498-511
- Gewolb, J. 2002. Plant scientists see big potential in tiny plastids. *Science* 295: 258-259
- Gilson, P.R. and G.I. McFadden. 1996. Good things in small packages: the tiny genomes of chlorarachniophyte endosymbionts. *BioEssays* 19(2): 167-173
- Gounaris, I., C.A. Michalowski, J.J. Bohnert and C.A. Price. 1986. Restriction and gene maps of plastid DNA from *Capsicum annum*. *Current Genetics* 11: 7-16
- Graham, S.W. and R.G. Olmstead. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87(11): 1712-1730
- Gray, J.C. 2003. Chloroplast-to-nucleus signalling: a role for Mg-protoporphyrin. *Trends in Genetics* 19(10): 526-529
- Gray, M.W. 1999. Evolution of organellar genomes. *Current Opinion in Genetics and Development* 9: 678-687
- Grzebyk, D., O. Schofield, C. Vetriani and P.G. Falkowski. 2003. The Mesozoic radiation of eukaryotic algae: the portable plastid hypothesis. *Journal of Phycology* 39: 259-267
- Guda, C., S.B. Lee and H. Daniell. 2000. Stable expression of a biodegradable protein-based polymer in tobacco chloroplasts. *Plant Cell Reports* 19 (3): 257-262
- Gupta, M. and B. Hoo. 1991. Entire maize chloroplast genome is stably maintained in a yeast artificial chromosome. *Plant Molecular Biology* 17: 361-369

- Hager, M. and R. Bock. 2000. Enslaved bacteria as new hope for plant biotechnologists. *Applied Microbiology and Biotechnology* 54: 302 – 310
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41:95-98.
- Hanson, L., R.L. Brown, A. Boyd, M.A.T. Johnson and M.D. Bennett. 2003. First nuclear DNA C-values for 28 angiosperm genera. *Annals of Botany* 91: 31-38
- Harper, J.T. and P.J. Keeling. 2003. Nucleus-encoded, plastid-targeted glyceraldehydes-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Molecular Biology and Evolution* 20: 1730 - 1735.
- Harris, E.H., J.E. Boynton and N.W. Gillham. 1994. Chloroplast ribosomes and protein synthesis. *Microbiological Reviews* 58(4): 700-754
- Heifetz, P.B. 2000. Genetic engineering of the chloroplast. *Biochimie* 82: 655-666
- Herrmann, R.G., H.J. Bohnert, K.V. Kowallik and J.M. Schmitt. 1975. Size, conformation and purity of chloroplast DNA of some higher plants. *Biochimie Biophysica Acta* 378 (2): 305-317
- Herrmann, R.G., H.K. Palta and K.V. Kowallik. 1980. Chloroplast DNA from three archegoniates. *Planta* 148: 319
- Hirai, A., T. Ishibashi, A. Morikami, N. Iwatsuki, K. Shinozaki and M. Sugiura. 1985. Rice chloroplast DNA: a physical map and location of genes for the large subunit of ribulose 1, 5-bisphosphate carboxylase and the 32-Kda photosystem II reaction centre protein. *Theory of Applied Genetics* 70: 117-122.
- Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C.R. Sun, B.Y. Meng, Y.Q. Li, A. Kanno, Y. Nishizawa, A. Hirai, K. Shinozaki and M. Sugiura. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular and General Genetics* 217: 185-194
- Howe, C.J. 1985. The endpoints of an inversion in wheat chloroplast DNA are associated with short repeated sequences containing homology to att-lambda. *Current Genetics* 10(2): 139-145
- Huang, J. C. and M. Sun. 2000. Genetic diversity and relationships of sweetpotato and its wild relatives in *Ipomoea* series *Batatas* (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theoretical and Applied Genetics* 100: 1050-1060
- Iamtham, S. and A. Day. 2000. Removal of antibiotic resistance genes from transgenic tobacco plastids. *Nature Biotechnology* 18: 1172–1176
- Iino, M. and H. Hashimoto. 2003. Intermediate features of cyanelle division of *Cyanophora paradoxa* (Glaucocystophyta) between cyanobacterial and plastid division. *Journal of Phycology* 39: 561-569

- Jakowitsch, J., C. Neumann-Spallart, Y. Ma, J. Steiner, H.E.A. Schenk, H.J. Bohnert and W. Löffelhardt. 1996. In vitro import of pre-ferredoxin-NADP<sup>+</sup>-oxidoreductase from *Cyanophora paradoxa* into cyanelles and into pea chloroplasts. *FEBS Letters* 381: 153-155
- Kaneko, T., Y. Nakamura, C.P. Wolk, T. Kuritz, S. Sasamoto, A. Watanabe, M. Iriguchi, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, M. Kohara, M. Matsumoto, A. Matsuno, A. Muraki, N. Nakazaki, S. Shimpo, M. Sugimoto, M. Takazawa, M. Yamada, M. Yasuda and S. Tabata. 2001. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Research* 8 (5): 205-213
- Kaneko, T. and S. Tabata. 1997. Complete genome structure of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. *Plant Cell Physiology* 38(11): 1171-1176
- Karol, K.G., R.M. McCourt, M.T. Cimino and C.F. Delwiche. 2001. The closest living relatives of land plants. *Science* 294: 2351-2353
- Katayama, H. and C. Uematsu. 2003. Comparative analysis of chloroplast DNA in *Pyrus* species: physical map and gene localisation. *Theoretical and Applied Genetics* 106: 303-310
- Kato, T., T. Kaneko, S. Sato, Y. Nakamura and S. Tabata. 2000. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Research* 7: 323-330.
- Kelchner, S.A. and J.F. Wendel. Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Current Genetics* 30: 259-262
- Kim, C.S., C.H. Lee, J.S. Shin, Y.S. Chung and N.I. Hyung. 1997. A simple and rapid method for the isolation of high quality genomic DNA from fruit trees and conifers using PVP. *Nucleic Acids Research* 25(5): 1085-1087
- Klug, W.S. and M.R. Cummings. 1997. *Concepts of Genetics (5<sup>th</sup> edition)*. Prentice Hall, New Jersey USA.
- Köhler, R.H., J. Cao, W.R. Zipfel, W.W. Webb and M.R. Hanson. 1997. Exchange of protein molecules through connections between higher plant plastids. *Science* 276: 2039-2042
- Kota, M., H. Daniell, S. Varma, S.F. Garczynski, F. Gould and W.J. Moar. 1999. Overexpression of the *Bacillus thuringiensis* (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. *Proceeds of the Natural Academy of Science USA* 96: 1840-1845
- Kugita, M., A. Kaneko, Y. Yamamoto, Y. Takeya, T. Matsumoto and K. Yoshinaga. 2003. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. *Nucleic Acids Research* 31(2): 716-721
- Kunnimalaiyaan, M. and B.L. Nielsen. 1997. Fine mapping of replication origins (*oriA* and *oriB*) in *Nicotiana tabacum* chloroplast DNA. *Nucleic Acids Research* 25(18): 3681-3686
- Kuzoff, R.K. and C.S. Gasser. 2000. Recent progress in reconstructing angiosperm phylogeny. *Trends in Plant Science* 5(8): 330-336
- Lang, M., K.E. Apt and P.G. Kroth. 1998. Protein transport into "complex" diatom plastids utilizes two different targeting signals. *Journal of Biological Chemistry* 273: 30973-30978

- Lee, Y-O., A. Kanno and T. Kameya. 1996. The physical map of the chloroplast DNA from *Asparagus officinalis* L. *Theoretical and Applied Genetics* 92: 10-14
- Leelavathi, S., N. Gupta, S. Maiti, A. Ghosh and V.S. Reddy. 2003. Overproduction of an alkali and thermostable xylanase in tobacco chloroplasts and efficient recovery of the enzyme. *Molecular Breeding* 11: 59-67
- Leister, D. 2003. Chloroplast research in the genomic age. *Trends in Genetics* 19(1): 47-56
- Lemieux, C., C. Otis and M. Turmel. 2000. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature* 403(6770): 649-652
- Lim, H., I. Gounaris, R.C. Haridison and C.D. Boyer. 1990. Restriction site and genetic map of *Curcubita pepo* chloroplast DNA. *Current Genetics* 18: 273-275
- Lodhi, M.A. and B.I. Reisch. 1995. Nuclear DNA content of *Vitis* species, cultivars, and other genera of the Vitaceae. *Theoretical and Applied Genetics* 90: 11-16
- Lodhi, M.A., G-N. Ye, N.F. Weeden and B.I. Reisch. 1994. A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter* 12(1): 6-13
- Ma, C. and M.A. Smith. 1985. The chloroplast genome of *Carthamus tinctorius*. *Molecular and Cellular Biochemistry* 68: 139-150
- Maier, R. M., K. Neckermann, G. L. Igloi and H. Kössel. 1995. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *Journal of Molecular Biology* 251: 614-628
- Maliga, P. 2002. Engineering the plastid genome of higher plants. *Current Opinion in Plant Biotechnology* 5: 164-172
- Maliga, P. 2003. Progress towards commercialisation of plastid transformation technology. *Trends in Biotechnology* 21(1): 20-28
- Manning, J.E., D.R. Wolstenholme, R.S. Ryan, J.A. Hunter and O.C. Richards. 1971. Circular chloroplast DNA from *Euglena gracilis*. *Proceeds of the Natural Academy of Science USA* 68: 1169-1173
- Mariac, C., P. Trouslot, C. Poteaux, G. Bezançon and J-F. Renno. 2000. Chloroplast DNA extraction from herbaceous and woody plants for direct restriction fragment length polymorphism analysis. *Biotechniques* 28: 110-113
- Marra, M., T. Kucaba, M. Sekhon, L. Hillier, R. Martienssen, A. Chinwalla, J. Crockett, J. Fedele, H. Grover, C. Gund, W.R. McCombie, K. McDonald, J. McPherson, N. Mudd, L. Parnell, J. Schein, R. Seim, P. Shelby, R. Waterston, and R. Wilson. 1999. A map for the sequence analysis of the *Arabidopsis thaliana* genome. *Nature Genetics* 22(3): 265-270
- Martin, W., T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa and D. Penny. 2002. *Proceeds of the Natural Academy of Science, USA* 99(19): 12246-12251

- Martin, W., B. Stoebe, V. Goremykin, S. Hapsmann, M. Hasegawa and K.V. Kowallik. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393: 162-165
- Matsuoka, Y., Y. Yamazaki, Y. Ogiwara and K. Tsunewaki. 2002. Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. *Molecular Biology and Evolution* 19(12): 2084-2091
- Maul, J.E., J.W. Lilly, L. Cui, C.W. dePamphilis, W. Miller, E.H. Harris and D.B. Stern. 2002. *The Plant Cell* 14: 2659-2679
- McFadden, G.I. 1999. Plastids and protein targeting. *Journal of Eukaryotic Microbiology* 46:339-346
- McFadden, G.I. 2001. Chloroplast origin and integration. *Plant Physiology* 125: 50-53
- McFadden, G.I. 2001. Primary and secondary endosymbiosis and the origin of plastids. *Journal of Phycology* 37: 951-959
- McFadden, G.I. and D.S. Roos. 1999. Apicomplexan plastids as drug targets. *Trends in Microbiology* 7(8): 328-333
- Mereschowsky, C. 1905. Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol Zentralbl* 25: 593-604
- Milligan, B.G. 1992. Ch 3: Plant DNA isolation. In: Hoelzel, A.R. (ed.). *Molecular Genetic Analysis of Populations: A practical approach*. Oxford University Press, Walton Street, Oxford, p 59-88
- Moreira, D., H. Le Guyader and H. Philippe. 2000. The origin of red algae and the evolution of chloroplasts. *Nature* 405:69-72
- Moreira, D. and H. Philippe. 2001. Sure facts and open questions about the origin and evolution of photosynthetic plastids. *Research in Microbiology* 152: 771-780
- Mourad, G.S. 1998. Chloroplast DNA Isolation. *Methods in Molecular Biology* 82: 71 –77
- Mozo, T., K. Dewar, P. Dunn, J.R. Ecker, S. Fischer, S. Kloska, H. Lehrach, M. Marra, R. Martienssen, S. Meier-Ewert, and T. Altmann. 1999. A complete BAC-based physical map of the *Arabidopsis thaliana* genome. *Nature Genetics* 22(3): 271-275
- Mpoloka, S.W. 2001. Genetic effect of prolonged UVB exposure in a Namaqualand daisy *Dimorphotheca sinuate*. *PhD thesis University of Cape Town, South Africa*
- Nishiyama, T. and M. Kato. 1999. Molecular phylogenetic analysis among bryophytes and tracheophytes based on combined data of plastid coded genes and the 18S rRNA gene. *Molecular Biology and Evolution* 16: 1027-1036
- Nozaki, H., M. Matsuzaki, M. Takahara, O. Misumi, H. Kuroiwa, M. Hasegawa, T. Shin-i, Y. Kohara, N. Ogasawara and T. Kuroiwa. 2002. The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *Journal of Molecular Evolution* 56: 485-497



- Ogihara, Y., K. Isono, T. Kojima, A. Endo, M. Hanaoka, T. Shiina, T. Terachi, S. Utsugi, M. Murata, N. Mori, S. Takumi, K. Ikeo, T. Gojobori, R. Murai, K. Murai, Y. Matsuoka, Y. Ohnishi, H. Tajiri and K. Tsunewaki. 2000. Chinese Spring wheat (*Triticum aestivum* L.) chloroplast genome: complete sequence and contig clones. *Plant Molecular Biology Reporter* 18: 243-253
- Ogihara, Y., T. Terachi and T. Sasakuma. 1988. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proceeds of the Natural Academy of Science USA* 85: 8573-8577
- Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi and H. Ozeki. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322: 572-574
- O'Kane, S.L., Jr. 1995. Clone bank and physical map of *Lopezia miniata* Lag. Ex DC. Ssp. *Miniata* (Onagraceae) chloroplast DNA. *Biochemical Systematics and Ecology* 23(5): 507-516
- Olson, M.V., J.E. Dutchik, M.Y. Graham, G.M. Brodeur, C. Helms, M. Frank, M. MacCollin, R. Scheinman and T. Frank. 1986. Random-clone strategy for genomic restriction mapping in yeast. *Proceeds of the Natural Academy of Science USA* 83: 7826-7830
- Palmer, J.D. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *The American Naturalist* 130: S6-S29
- Palmer, J.D. 1988. Ch 6: Isolation and Structural Analysis of Chloroplast DNA. In Weissbach, A. and H. Weisbach (eds.). *Methods for Plant Molecular Biology*. Academic Press, p.105
- Palmer, J.D. 2003. The symbiotic birth and spread of plastids: how many times and whodunit? *Journal of Phycology* 39: 4-11
- Palmer, J.D., R.A. Jorgensen and W.F. Thompson. 1985. Chloroplast DNA variation and evolution in pisum: patterns of change and phylogenetic analysis. *Genetics* 109: 195
- Palmer, J.D. and W.F. Thompson. 1981. Clone banks of the mung bean, pea and spinach chloroplast genomes. *Gene* 15(1): 21-26
- Peeters, N. and I. Small. 2001. Dual targeting to mitochondria and chloroplasts. *Biochimie Biophysica Acta* 1541: 54-63
- Pfannschmidt, T., K. Shutze, V. Fey, I. Sheraleti and R. Oelmüller. 2003. Chloroplast redox control of nuclear gene expression – a new class of plastid signals in interorganellar communication. *Antioxidant Redox Signal* 5(1): 95-101
- Porebski, S., L.G. Bailey and B.R. Baum. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* 15(1): 8-15

- Pryer, K.M., H. Schneider, A.R. Smith, R. Cranfill, P.G. Wolf, J.S. Hunt and S.D. Sipes. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* 409: 618-621
- Pryer, K.M., H. Schneider, E.A. Zimmer and J.A. Banks. 2002. Deciding among green plants for whole genome studies. *Trends in Plant Science* 7(12): 550-554
- Qiu, Y.-L., Y. Cho, J.C. Cox and J.D. Palmer. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* 394: 671-674
- Qiu, Y.-L. and J. Lee. 2000. Transition to a land flora: a molecular phylogenetic perspective. *Journal of Phycology* 36: 799 – 802
- Qiu, Y.-L. and J.D. Palmer. 1999. Phylogeny of early land plants: insights from genes and genomes. *Trends in Plant Science* 4 (1): 26-30
- Race, H.L., R.G. Herrmann and W. Martin. 1999. Why have organelles retained genomes? *Trends in Genetics* 15: 364-370
- Raven, J.A., A.M. Johnston, J.E. Kübler, R. Korb, S.G. McInroy, L.L. Handley, C.M. Scrimgeour, D.I. Walker, J. Beardall, M. N. Clayton, M. Vanderklift, S. Fredriksen and K.H. Dunton. 2002. Seaweeds in cold seas: evolution and carbon acquisition. *Annals of Botany* 90: 525-536
- Reece, R.J. 2004. *Analysis of Genes and Genomes*. John Wiley and Sons, West Sussex, England
- Reith, M. and J. Munholland. 1993. A high resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. *American Society of Plant Physiologists* 5: 465-475
- Rochaix, J.-D. 1997. Chloroplast reverse genetics: new insights into the function of plastid genes. *Trends in Plant Science* 2(11): 419-425
- Ruf, S., M. Hermann, I.J. Berger, H. Carrer and R. Bock. 2001. Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nature Biotechnology* 19(9): 870-875
- Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. *Molecular cloning: a laboratory manual, 2<sup>nd</sup> edition*. Cold Spring Harbour Laboratory Press, USA.
- Sager, R. and M.R. Ishida. 1963. Chloroplast DNA in *Chlamydomonas*. *Proceeds of the Natural Academy of Science USA* 50: 725-730
- Sato, N. 2001. Was the evolution of plastid genetic machinery discontinuous. *Trends in Plant Science* 6(4): 151-155
- Sato, S., Y. Nakamura, T. Kaneko, E. Asamizu and S. Tabata. 1999. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Research* 6(5): 283-290
- Schaffer, H.E. and R.R. Sederoff. 1981. Improved estimation of DNA fragment lengths from agarose gels. *Analytical Biochemistry* 115: 113-122

- Schmitz-Linneweber, C., R.M. Maier, J-P. Alcaraz, A. Cottet, R.G. Herrmann and R. Mache. 2001. The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. *Plant Molecular Biology* 45: 307-315
- Schmitz-Linneweber, C., R. Regel, T.G. Du, H. Hupfer, R.G. Herrmann and R.M. Maier. 2002. The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation. *Molecular Biology and Evolution* 19: 1602-1612
- Schopfer, M. 1995. Ch 22: Development of Chloroplasts. *Plant Physiology*. Springer-Verlag, Hong Kong, p375-382
- Schuler, M.A. and R.E. Zielinski. 1989. *Methods in plant molecular biology*. Academic Press, San Diego
- Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B. Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada and M. Sugiura. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *The EMBO Journal* 5(9): 2043-2049
- Shizuya, H., B. Birren, U.J. Kim, V. Mancino, T. Slepak, Y. Tachiiri and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proceeds of the Natural Academy of Science USA* 89: 8794-8797
- Sidorov, V.A., D. Kasten, S-Z. Pang, P.T.J. Hajdukiewicz, J.M. Staub and N.S. Nehra. 1999. Stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. *The Plant Journal* 19(2): 209-216
- Simpson, C.L. and D.B. Stern. 2002. The treasure trove of algal chloroplast genomes. Surprises in architecture and gene content, and their functional implications. *Plant Physiology* 129: 957-966
- Smith, M.A. and C. Ma. 1985. Restriction endonuclease cleavage site map of Safflower (*Carthamus tinctorius L.*) chloroplast DNA. *Theoretical and Applied Genetics* 70: 32-41
- Soltis, E.D. and P.S. Soltis. 2000. Contributions of plant molecular systematics to studies of molecular evolution. *Plant Molecular Biology* 42: 45-75
- Soltis, P.S., D.E. Soltis and M.W. Chase. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402-404
- Staub, J.M., B. Garcia, J. Graves, P.T.J. Hajdukiewicz, P. Hunter, N. Nehra, V. Paradkar, M. Schlittler, J.A. Carroll, L. Spatola, D. Ward, G. Ye and D.A. Russell. 2000. *Nature Biotechnology* 18: 333-338
- Stiller, J.W. and B.D. Hall. 1997. The origin of red algae: Implications for plastid evolution. *Proceeds of the Natural Academy of Science USA* 94: 4520-4525

- Stiller, J.W., D. C. Reel and J.C. Johnson. 2003. A single origin of plastids revisited: convergent evolution in organellar genome content. *Journal of Phycology* 39: 95-105
- Stoebe, B.W. and Kowallik, K.V. 1999. Gene-cluster analysis in chloroplast genomics. *Trends in Genetics* 15: 344-347
- Stoebe, B.W., K.V. Martin and K.V. Kowallik. 1998. Distribution and nomenclature of protein-coding genes in 12 sequenced chloroplast genomes. *Plant Molecular Biology Reporter* 16: 243-255
- Sugiura, M. 1992. The chloroplast genome. *Plant Molecular Biology* 19: 149-168
- Sugiura, C., Y. Kobayashi, S. Aoki, C. Sugita and M. Sugita. 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. *Nucleic Acids Research* 31(18): 5324-5331
- Svab, Z., P. Hajdukiewicz and P. Maliga. 1990. Stable transformation of plastids in higher plants. *Proceeds of the Natural Academy of Science USA* 87: 8526-8530
- Svab, Z. and P. Maliga. 1993. High frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proceeds of the Natural Academy of Science USA* 90: 913-917
- The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815
- Tregoning, J.S. 2003. Expression of tetanus toxin fragment C in tobacco chloroplasts. *Nucleic Acids Research* 31: 1174-1179
- Triboush, S.O., N.G. Danilenko and O.G. Davydenko. 1998. A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower. *Plant Molecular Biology Reporter* 16: 183-189
- Turmel, M., C. Otis, C. Lemieux. 1999. The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: Insights into the architecture of ancestral chloroplast genomes. *Proceeds of the Natural Academy of Science USA* 96: 10248-10253
- Turmel, M., C. Otis and C. Lemieux. 2002. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proceeds of the Natural Academy of Science USA* 99: 11275-11280
- Turmel, M., C. Otis and C. Lemieux. 2002. The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Molecular Biology and Evolution* 19(1): 24-38
- Turner, S., K.M. Pryer, V.P.W. Miao and J.D. Palmer. 1999. Investigating deep phylogenetic relationships among Cyanobacteria and plastids by small subunit rRNA sequence analysis. *Journal of Eukaryotic Microbiology* 46: 327-338

- Van Ham, R.C.H.J., H. Hart, T.H.M. Mes and J.M. Sandbrink. 1994. Molecular evolution of noncoding regions of the chloroplast genome in the Crassulaceae and related species. *Current Genetics* 25: 558-566
- Venter, J.C., H.O. Smith and L. Hood. 1996. A new strategy for genome sequencing. *Nature* 381: 364-366
- Wakasugi, T., J. Tsudzuki, S. Ito, K. Nakashima, T. Tsudzuki and M. Sugiura. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proceeds of the Natural Academy of Science USA* 91: 9794-9798
- Watson, M.D. and D.J. Murphy. 1999. Ch 10: Genome organization, protein synthesis and processing in plants. In Lea, P.J. and R.C. Leegood (eds.). *Plant biochemistry and molecular biology*, 2<sup>nd</sup> edition. John Wiley & Sons, England, p237-243.
- Weber, A. and U-I. Flügge. 2002. Interaction of cytosolic and plastidic nitrogen metabolism in plants. *Journal of Experimental Biology* 53: 865-874
- Weber, A., J.C. Servaites, D.R. Geiger, H. Kofler, D. Hille, F. Gröner, U. Hebbeker, and U-I. Flügge. 2000. Identification, purification, and molecular cloning of a putative plastidic glucose translocator. *Plant Cell* 12: 787-802
- Whitney, S.M. and T.J. Andrews. 2001. Plastome-encoded bacterial ribulose-1,5,-biphosphate carboxylase/oxygenase (RubisCO) supports photosynthesis and growth of tobacco. *Proceeds of the Natural Academy of Science USA* 98:14738-14743
- Wolfe, K.H., W.H. Li and P.M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proceeds of the Natural Academy of Science USA* 84: 9054-9058
- Yang, D., A. Parco, S. Nandi, P. Subudhi, Y. Zhu, G. Wang, N. Huang. 1997. Construction of a bacterial artificial chromosome (BAC) library and identification of overlapping BAC clones with chromosome 4-specific RFLP markers in rice. *Theoretical and Applied Genetics* 95: 1147-1154
- Yeoh, H.H. and S.S. Joseph. 1995. Cassava chloroplast DNA: isolation, comparison and cloning. *Acta Physiologiae Plantarum* 17(4): 381-386
- Yoon, H.S., J.D. Hackett, G. Pinto and D. Bhattacharya. 2002. The single, ancient origin of chromist plastids. *Proceeds of the Natural Academy of Science USA* 99(24): 15507-15512
- Zhang, H-B. and R.A. Wing. 1997. Physical mapping of the rice genome with BACs. *Plant Molecular Biology* 35: 115-127
- Zhang, H-B. and C. Wu. 2001. BAC as tools for genome sequencing. *Plant Physiology* 39: 195-209
- Zhang, Z., B. Green and T. Cavalier-Smith. 1999. Single gene circles in dinoflagellate chloroplast genomes. *Nature* 400: 155-159

Deep Green – *The Green Plant Phylogeny Research Coordination Group*. [World Wide Web] Available: (<http://ucjeps.berkeley.edu/bryolab/GPphylo/>) [2004, January 27]

Entrez Genomes – *Complete chloroplast genome sequences*. [World Wide Web] Available: ([http://megasun.bch.umontreal.ca/ogmp/projects/other/cp\\_list.html](http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list.html)) [2004, January 27]

*International Grape Genome Project (IGGP)*. [World Wide Web] Available: (<http://www.vitaceae.org/>) [2004, January 27]

Kim, U-J. *Caltech Genome Research Laboratory: Manipulation and Analysis of BACs*. [World Wide Web] Available: (<http://wheat.pw.usda.gov/~lazo/methods/caltech/caltech.html>) [2004, January 24]

Lefort, F. *Characterization of Grapevine with Universal Chloroplast Microsatellite Markers (The Greek Vitis Database)*. [World Wide Web] Available: ([www.biology.uoc.gr/gvd/contents/general-info/06b.htm](http://www.biology.uoc.gr/gvd/contents/general-info/06b.htm)) [2004, January 27]

Olesen, K. *p-Draw 32 1.0 – AcaClone Software*. [World Wide Web] Available: (<http://www.acaclone.com>) [2004, January 27]

Wait, D. (2003) *South African Wine Industry Statistics (SAWIS)*. [World Wide Web] Available: ([http://www.sawis.co.za/SAWISPortal/uploads/Extent of the South African Wine Industry 2003.pdf](http://www.sawis.co.za/SAWISPortal/uploads/Extent%20of%20the%20South%20African%20Wine%20Industry%202003.pdf)) [2004, January 27]