

MARIE-CHRISTINE GAGNON

**SÉQUENÇAGE ET CARACTÉRISATION DU GÉNOME CHLOROPLASTIQUE ET
DU GÉNOME MITOCHONDRIAL DE L'ALGUE VERTE *Monomastix* OKE-1**

**Mémoire
présenté
à la Faculté des études supérieures
de l'Université Laval
pour l'obtention
du grade de maître ès sciences (M.Sc.)**

**Département de biochimie
FACULTÉ DE SCIENCES ET GÉNIE
UNIVERSITÉ LAVAL**

Avril 2001



**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**395 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**395, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-65404-4

Canada

RÉSUMÉ COURT

La classe d'algues vertes Prasinophyceae est connue comme celle qui contient l'ancêtre de toutes les plantes vertes. La caractérisation des génomes des organelles du prasinophyte *Monomastix* OKE-1, qui a fait l'objet de cette étude, avait comme but d'obtenir de l'information sur les génomes d'une algue se rapprochant de cet ancêtre ainsi que de clarifier la position phylogénétique encore mal connue de cette algue. Suite au séquençage et à l'assemblage de chacun des génomes de *Monomastix*, des analyses comparatives ont permis de mettre en évidence que l'ADN chloroplastique a perdu des caractéristiques conservées importantes de l'architecture des génomes chloroplastiques ancestraux mais que, à l'opposé, l'ADN mitochondrial est très conservé. Les analyses phylogénétiques réalisées à l'aide de la séquence protéique de chacun des génomes donnent des résultats différents des analyses préliminaires; *Monomastix* ne serait pas aussi basal qu'on le croyait.

RÉSUMÉ LONG

Les plantes vertes forment deux lignées évolutives principales qui ont divergé très tôt durant l'évolution : la lignée Streptophyta incluant les algues vertes charophytes ainsi que les plantes terrestres et la lignée Chlorophyta regroupant le reste des algues vertes. Parmi les quatre grandes classes d'algues vertes de la lignée Chlorophyta, la classe Prasinophyceae est celle qui regroupe les algues vertes les plus ancestrales. L'étude du génome des organelles des prasinophytes est utile pour obtenir de l'information sur l'ancêtre commun des streptophytes et chlorophytes ainsi que pour clarifier la position phylogénétique encore mal connue de plusieurs de ces algues. Dans le cadre de ce projet, le génome chloroplastique ainsi que le génome mitochondrial du prasinophyte *Monomastix* OKE-1 ont été séquencés et caractérisés. Afin de séquencer chacun de ces génomes, une banque de clones plasmidiques a été construite avec les fragments entre 1 et 2 kb de l'ADN coupé au hasard des organelles. Cet ADN a auparavant été isolé par ultracentrifugation en présence d'un gradient de chlorure de césium. Avant d'être séquencés, les clones étant les plus susceptibles de contenir une insertion chloroplastique ou mitochondriale ont été sélectionnés par hybridation avec une sonde construite à l'aide de l'ADN ayant servi à faire la banque de clones. L'assemblage de chacun des génomes a débuté avec l'accumulation des séquences des clones choisis. Des oligonucléotides et des produits PCR ont permis de refermer les génomes. Le génome chloroplastique de *Monomastix* OKE-1 a perdu des caractéristiques conservées importantes de l'architecture des génomes chloroplastiques ancestraux : l'élément d'ADN inversé répété, plusieurs gènes ainsi que l'arrangement en opéron de la plupart des gènes. Il a aussi acquis plusieurs introns et un grand nombre de courtes séquences d'ADN répété, éléments qui n'avaient pas été observés chez les autres prasinophytes étudiés. Les séquences répétées pourraient d'ailleurs avoir un lien avec la réorganisation de cet ADN. Les analyses comparatives de ce génome chloroplastique démontrent qu'il a évolué plus rapidement que les autres génomes chloroplastiques de prasinophytes étudiés. Contrairement au génome chloroplastique, le génome mitochondrial de *Monomastix* a conservé tous les traits d'un génome mitochondrial ancestral d'algue verte. La position phylogénétique de *Monomastix* basée sur l'ADN des organelles confirme que ce prasinophyte n'est pas aussi basal que le démontrait une analyse phylogénétique basée sur le gène 23S.

REMERCIEMENTS

Je tiens à remercier ma directrice de recherche, Monique Turmel, ainsi que mon co-directeur de recherche, Claude Lemieux, pour leur soutien tout le long de ce travail.

Je voudrais aussi remercier les membres de mon comité aviseur, messieurs Jean Bousquet et Michel Guertin pour leurs commentaires constructifs.

Enfin, merci à Christian Otis pour ses précieux conseils et son aide technique ainsi qu'à tous les membres du laboratoire soit Yannick Beauséjour, Jean-Charles de Cambiaire, Megumi Ehara, Patrick Lucas, Nick Martel et Jean-François Pombert pour leur aide et leur encouragement. Un remerciement aussi à Robin Drolet qui a aidé à préparer la banque de clones de *Monomastix* OKE-1.

TABLE DES MATIÈRES

RÉSUMÉ COURT	i
RÉSUMÉ LONG	ii
REMERCIEMENTS	iii
TABLE DES MATIÈRES	iv
LISTE DES FIGURES	ix
LISTE DES TABLEAUX	xi
LISTE DES ABRÉVIATIONS	xiii
 CHAPITRE 1	
INTRODUCTION	1
1.1 Les plantes vertes et leur origine	1
1.1.1 Les algues vertes	1
1.1.1.1 La classification des algues vertes	4
1.1.1.2 Les prasinophytes	5
1.1.2 Des modèles pour étudier l'évolution des plantes vertes	9
1.2 Le chloroplaste	10
1.2.1 L'origine des chloroplastes	10
1.2.2.1 Caractéristiques générales	13
1.2.2 Le génome chloroplastique des plantes vertes	14
1.2.2.2 Les régions répétées	14
1.2.2.3 Les gènes chloroplastiques	19
Les gènes impliqués dans la photosynthèse	20
Les gènes impliqués dans l'expression génétique	26
Les gènes de division	27
Autres gènes	30

ORFs conservés	30
1.2.2.4 L'organisation des gènes chloroplastiques	30
1.2.2.5 Les introns chloroplastiques	33
1.2.3 Les analyses phylogénétiques à l'aide de séquences de gènes chloroplastiques concaténés	34
1.3 La mitochondrie	35
1.3.1 L'origine des mitochondries	35
1.3.2 Le génome mitochondrial des plantes vertes	37
1.3.2.1 Caractéristiques générales.....	37
1.3.2.2 Les régions répétées.....	41
1.3.2.3 Les gènes mitochondriaux	42
Les gènes impliqués dans la chaîne respiratoire	42
Les gènes impliqués dans l'expression génétique	46
Autres gènes.....	46
ORFs conservés	46
1.3.2.4 L'organisation des gènes mitochondriaux	46
1.3.2.5 Les introns mitochondriaux	47
1.3.3 Les analyses phylogénétiques à l'aide de séquences de gènes mitochondriaux concaténés.....	47
1.4 Problématique de l'étude	49

CHAPITRE 2

MATÉRIEL ET MÉTHODES.....	51
2.1 Croissance cellulaire de l'algue <i>Monomastix</i> OKE-1	51
2.2 Isolement d'une fraction d'ADN enrichie en ADN chloroplastique et mitochondrial	51
2.2.1 Lyse des cellules	51
2.2.2 Extraction de l'ADN cellulaire total	52
2.2.3 Séparation de l'ADN cellulaire total sur gradient de chlorure de césium et extraction de la fraction d'ADN enrichie en ADN chloroplastique et mitochondrial	52
2.3 Digestion <i>EcoRI</i> de l'ADN des organites	53

2.4 Construction des banques de clones	56
2.4.1 Préparation des fragments d'ADN chloroplastique et mitochondrial.....	56
2.4.1.1 Nébulisation de l'ADN	56
2.4.1.2 Réparation des fragments d'ADN.....	56
2.4.1.3 Récolte des fragments d'ADN entre 1 et 2 kb	59
2.4.1.4 Phosphorylation des fragments d'ADN	62
2.4.2 Préparation du vecteur pFBS KS+	62
2.4.2.1 Digestion du vecteur	65
2.4.2.2 Déphosphorylation du vecteur	65
2.4.2.3 Test de ligature.....	65
2.4.3 Ligature de l'ADN nébulisé, réparé et phosphorylé et du vecteur déphosphorylé	66
2.4.4 Transformations	67
2.4.5 Repiquage des clones positifs et transferts sur filtres de nylon	68
2.5 Hybridation des filtres à une sonde d'ADN fabriquée à l'aide de la fraction originale d'ADN utilisée pour faire la banque.....	69
2.5.1 Préparation des filtres pour l'hybridation.....	69
2.5.2 Construction de la sonde.....	69
2.5.3 Hybridation	70
2.6 Purification des clones pour le séquençage	70
2.7 Séquençage	71
2.7.1 Méthodes et troussees utilisées pour le séquençage	71
2.8 Analyses des séquences et assemblage des génomes	72
2.8.1 Méthode utilisée pour les amplifications PCR	74
2.9 Caractérisation des génomes.....	75
2.9.1 Localisation des gènes et autres cadres de lecture ouverts ainsi que des introns	75
2.9.2 Construction des cartes de gènes	76
2.9.3 Analyses phylogénétiques.....	76
2.9.4 Caractérisation des séquences répétées chloroplastiques	79
2.9.5 Analyses de la composition et de l'utilisation des bases de la séquence de chaque génom.....	79

CHAPITRE 3

RÉSULTATS.....	80
3.1 Le génome chloroplastique.....	80
3.1.1 Structure et caractéristiques générales.....	80
3.1.2 Contenu en gènes.....	80
3.1.3 Les ARNs de transfert et la composition des codons.....	85
3.1.4 Organisation des gènes.....	90
3.1.5 Les introns.....	100
3.1.6 Les séquences répétées.....	109
3.1.7 Les analyses phylogénétiques.....	113
3.2 Le génome mitochondrial.....	119
3.2.1 Structure et caractéristiques générales.....	119
3.2.2 Contenu en gènes.....	119
3.2.3 Les ARNs de transfert et la composition des codons.....	124
3.2.4 Organisation des gènes.....	129
3.2.5 Les introns.....	134
3.2.6 Les régions intergéniques.....	140
3.2.7 Les analyses phylogénétiques.....	141

CHAPITRE 4

DISCUSSION.....	145
4.1 Le génome chloroplastique de <i>Monomastix</i> OKE-1 a perdu plusieurs caractéristiques ancestrales.....	145
4.2 Le génome mitochondrial de <i>Monomastix</i> OKE-1 a conservé une structure ancestrale.....	151
4.3 Le pourcentage en bases A + T des génomes d'organelles de <i>Monomastix</i> OKE-1 apporte certaines implications.....	153
4.4 Des événements de transfert latéraux et horizontaux à l'origine des introns de <i>Monomastix</i>	154
4.4.1 L'intron chloroplastique du gène <i>psbA</i>	154
4.4.2 L'intron chloroplastique du gène <i>trnK(uuu)</i>	157
4.4.3 Les introns chloroplastiques et mitochondriaux du gène de l'ARNr 23S.....	160

4.4.4 Les introns mitochondriaux du gène <i>cox1</i>	161
4.5 La position phylogénétique de <i>Monomastix</i> OKE-1 est surprenante	161
4.6 Conclusion et perspectives.....	164
CHAPITRE 5	
RÉFÉRENCES	165
ANNEXE 1 Séquence complète du génome chloroplastique de <i>Monomastix</i> OKE-1 en format GenBank	175
ANNEXE 2 Séquence complète du génome mitochondrial de <i>Monomastix</i> OKE-1 en format GenBank	223

LISTE DES FIGURES

Figure 1. Schéma évolutif présentant les deux lignées de plantes vertes ainsi que les classes les formant	2
Figure 2. Phylogénie des Viridiplantae basée sur la séquence de la petite sous-unité de l'ARN ribosomique nucléaire et présentant les quatre groupes de prasinophytes	7
Figure 3. Schéma présentant l'origine des plastides simples et complexes par endosymbiose primaire ou secondaire	11
Figure 4. Carte circulaire des gènes de l'ADNcp du prasinophyte <i>Mesostigma viride</i>	17
Figure 5. Structure secondaire standard d'un ARNt	28
Figure 6. Profil d'ADN obtenu suite à l'ultracentrifugation de l'ADN cellulaire total de <i>Monomastix</i> OKE-1 sur un gradient de CsCl en présence de bisbenzimidazole	54
Figure 7. Digestion <i>EcoRI</i> de l'ADN préparé par ultracentrifugation sur gradient de CsCl	57
Figure 8. Visualisation de la qualité de l'ADN avant et après nébulisation	60
Figure 9. Séquence du vecteur pFBS KS ⁺	63
Figure 10. Carte circulaire des gènes du génome chloroplastique de <i>Monomastix</i> OKE-1	81
Figure 11. Structures secondaires des ARNs codés par les gènes chloroplastiques <i>rnpB</i> et <i>rrf</i>	86
Figure 12. Structures secondaires des ARNt codés par le génome chloroplastique	88
Figure 13. Localisation des regroupements de gènes chloroplastiques conservés chez <i>Monomastix</i>	93
Figure 14. Regroupements de gènes chloroplastiques trouvés chez <i>Monomastix</i> OKE-1 ainsi que six autres algues vertes et dérivés des opérons <i>str-L2-spc-alpha-S9</i> de <i>E. coli</i>	96
Figure 15. Vestiges du regroupement de gènes ancestral <i>atpA</i> chez <i>Monomastix</i> OKE-1 et six autres algues vertes	98
Figure 16. Emplacement de l'intron dans la structure secondaire de l'ARNt codé par le gène <i>trnK(uuu)</i>	102
Figure 17. Structures secondaires des introns dans les gènes <i>psbA</i> et <i>trnK(uuu)</i>	104
Figure 18. Structures secondaires des quatre introns dans le gène <i>rrl</i>	106
Figure 19. Localisation des séquences répétées sur la carte chloroplastique linéaire des gènes de <i>Monomastix</i>	111

Figure 20. Liens hypothétiques entre les différentes familles de séquences répétées	114
Figure 21. Position phylogénétique de <i>Monomastix</i> déduite à l'aide de la séquence en acides aminés de protéines chloroplastiques concaténées	116
Figure 22. Carte circulaire des gènes du génome mitochondrial de <i>Monomastix</i> OKE-1	120
Figure 23. Structures secondaires des ARNs codés par les gènes mitochondriaux <i>rnpB</i> et <i>rrn5</i>	125
Figure 24. Structures secondaires des ARNs de transfert du génome mitochondrial.....	127
Figure 25. Regroupements de gènes similaires aux opérons <i>str-L2-spc-alpha</i> de <i>E. coli</i> dans l'ADNmt de <i>Monomastix</i> OKE-1 et d'autres ADNmt	132
Figure 26. Structures secondaires des quatre introns dans le gène <i>rnl</i>	136
Figure 27. Structures secondaires des quatre introns dans le gène <i>cox1</i>	138
Figure 28. Position phylogénétique de <i>Monomastix</i> déduite à l'aide de la séquence en acides aminées de protéines mitochondriales concaténées.....	142
Figure 29. Distribution des événements de perte de gènes chloroplastiques dans la lignée Chlorophyta	146

LISTE DES TABLEAUX

Tableau I. Caractéristiques générales des génomes chloroplastiques des plantes vertes dont les séquences entières sont disponibles	15
Tableau II. Distribution des gènes chloroplastiques répertoriés chez neuf plantes vertes dont les séquences entières de l'ADNcp sont disponibles.....	21
Tableau III. Distribution de regroupements de gènes chloroplastiques de l'ADNcp de <i>Mesostigma viride</i> chez cinq Chlorophytes et deux Streptophytes	32
Tableau IV. Caractéristiques générales des génomes mitochondriaux des algues vertes et d'un protiste dont la séquence entière est disponible.....	38
Tableau V. Distribution des gènes mitochondriaux répertoriés chez 11 plantes vertes dont les séquences entières de l'ADNmt sont disponibles.....	43
Tableau VI. Résumé des caractéristiques du génome chloroplastique de <i>Monomastix</i> OKE-1	83
Tableau VII. Gènes et ORFs uniques présents dans l'ADNcp de <i>Monomastix</i> OKE-1	84
Tableau VIII. Utilisation de codons chez 64 gènes codant pour les protéines conservées ainsi que pour les cinq ORFs introniques de l'ADNcp de <i>Monomastix</i> OKE-1	91
Tableau IX. Distribution des regroupements de gènes chloroplastiques conservés chez les algues vertes qui sont présents chez <i>Monomastix</i> OKE-1	92
Tableau X. Caractéristiques des introns présents dans le génome chloroplastique de <i>Monomastix</i>	101
Tableau XI. Classification et caractérisation des séquences répétées trouvées dans le génome chloroplastique de <i>Monomastix</i>	110
Tableau XII. Résumé des caractéristiques du génome mitochondrial de <i>Monomastix</i> OKE-1	122
Tableau XIII. Gènes codés par l'ADNmt de <i>Monomastix</i> OKE-1	123
Tableau XIV. Utilisation de codons chez 35 gènes conservés ainsi que pour les sept ORFs introniques de l'ADNmt <i>Monomastix</i> OKE-1	130
Tableau XV. Regroupement de gènes présents dans l'ADNmt de <i>Monomastix</i> OKE-1 ainsi que chez deux algues vertes ayant un génome mitochondrial de type ancestral et un protiste faisant partie des Jakobids	131

Tableau XVI. Caractéristiques des introns présents dans le génome mitochondrial de <i>Monomastix</i> OKE-1	135
Tableau XVII. Répertoire des introns localisés dans les mêmes gènes et insérés aux mêmes sites que les introns de l'ADNcp de <i>Monomastix</i> mais chez d'autres génomes.....	155
Tableau XVIII. Répertoire des introns localisés dans les mêmes gènes et insérés aux mêmes sites que les introns de l'ADNmt de <i>Monomastix</i> mais chez d'autres génomes	158

LISTE DES ABRÉVIATIONS

A + T : adénine + thymine

ADN: acide désoxyribonucléique

ADNcp: ADN chloroplastique

ADNmt: ADN mitochondrial

ARN: acide ribonucléique

ARNr : ARN ribosomique

ARNt : ARN de transfert

BE: acide borique-EDTA

BLAST: « basic local alignment search tool » ou outil de recherche pour les alignements de bases locaux

BSA: albumine de sérum de bœuf

Ci : Curie

dITP: désoxyinosine triphosphate

dNTP: désoxynucléotide triphosphate

DTT: dithiothréitol

EDTA: acide éthylènediaminetétraacétique

G + C: guanine + cytosine

HEPES: N-[2-hydroxyéthyl]pipérazine-N'-[acide 2-éthanesulfonic]

IPTG: isopropyl-β-D-thiogalactopyranoside

IR : région inversée répétée

JTT: Jones, Taylor, Thornton

kb: kilo base

LB: Luria Bertani

LSC : « large single copy » ou grande copie unique d'ADN

NZY⁺: « caseine hydrolysate enzymatic »

OPA: « one phor all »

ORF: « open reading frame » ou cadre de lecture ouvert

pb: paire de bases

PCR: « polymerase chain reaction » ou réaction en chaîne de la polymérase

PEG: polyéthylène glycol

PSI: « pound per square inch » ou poids par pouce carré

RELL: « resampling of estimated likelihood » ou rééchantillonnage de valeurs estimées de Likelihood

SDS: « sodium dodecyl sulfate »

SSC : « small single copy » ou petite copie unique d'ADN

20X SSC: « 3M sodium chloride, 0,3M sodium citrate » ou chlorure de sodium, citrate de sodium

TAE: tris base-acide acétique-EDTA

TE: tris-HCl-EDTA

U: unité d'enzyme

UTP: uridine triphosphate

INTRODUCTION

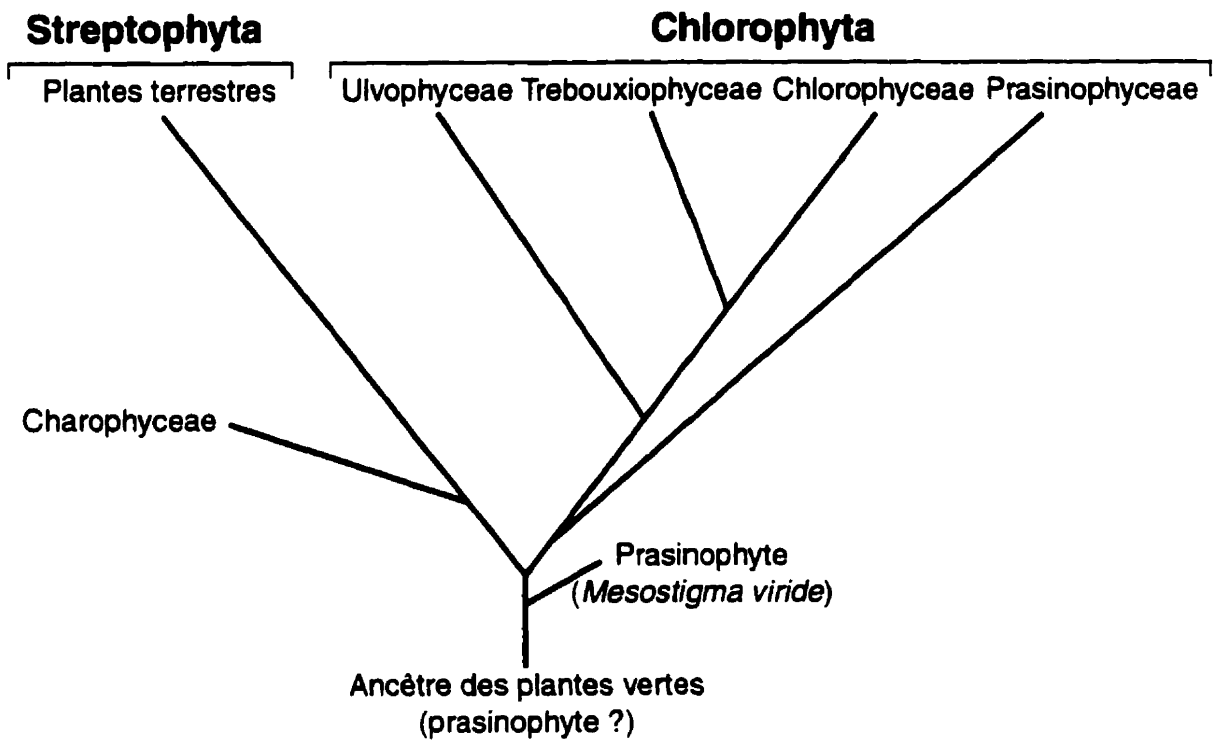
1.1 Les plantes vertes et leur origine

Les plantes vertes constituent le règne Viridiplantae. Elles se divisent en deux grandes lignées, Chlorophyta et Streptophyta (figure 1), qui diffèrent au niveau de l'architecture de leurs cellules flagellées. La lignée Streptophyta comprend les plantes terrestres (à ne pas confondre avec les plantes vertes) et la classe d'algues vertes leur étant le plus rapprochée, la classe Charophyceae. Les quatre autres classes d'algues vertes constituent la lignée Chlorophyta (Chapman *et al.*, 1998). L'idée qu'il y ait un seul ancêtre à l'origine de tout le règne Viridiplantae est maintenant acceptée. Cette origine monophylétique est supportée par les caractéristiques morphologiques des plantes vertes (Chapman *et al.*, 1998) ainsi que par des analyses phylogénétiques basées sur des comparaisons de la séquence de l'ADN ribosomique 18S de différentes lignées d'eucaryotes (Bhattacharya et Medlin, 1995) et d'autres, plus récentes, basées sur la séquence de protéines chloroplastiques concaténées de plantes vertes et autres algues (Lemieux *et al.*, 2000; Turmel *et al.*, 1999a).

1.1.1 Les algues vertes

Les algues forment un groupe d'organismes eucaryotes photosynthétiques unicellulaires ou multicellulaires ayant un éventail impressionnant de morphologies cellulaires, cycles de vie et habitats différents. Les principales lignées d'algues sont les lignées Chlorophyta (utilisé ici pour toutes les algues vertes), Rhodophyta (algues rouges), Glaucocystophyta, Euglenophyta, Chlorarachniophyta, Heterokonta, Haptophyta, Cryptophyta et les dinoflagellés (Bhattacharya et Medlin, 1998). Certaines caractéristiques permettent de distinguer clairement les algues vertes des autres lignées d'algues. Les algues vertes de la lignée Chlorophyta (n'incluant pas la classe Charophyceae) sont des organismes unicellulaires qui se caractérisent par la présence de chloroplastes avec une membrane double remplie de thylakoïdes empilés, de chlorophylles a et b, de caroténoïde et d'amidon comme principale réserve de polysaccharides (Friedl, 1997). Les xanthophylles lutéine, zéaxanthin, violaxanthin, anthéroxanthin et néoxanthin, des pigments accessoires, sont aussi présents. Ces algues ont des flagelles de structures similaires

Figure 1. Schéma évolutif présentant les deux lignées de plantes vertes ainsi que les classes les formant. La position du prasinophyte *Mesostigma viride*, situé à la base des plantes vertes et ne se groupant pas avec les autres membres de la classe Prasinophyceae, est indiquée.



mais de longueurs différentes et leur ADN circulaire est concentré dans un petit nucléole distribué dans tout le chloroplaste. Ces caractéristiques sont aussi présentes chez les plantes plus évoluées (Van Den Hoek *et al.*, 1995).

1.1.1.1 La classification des algues vertes

Comme les algues représentent un groupe d'organismes très diversifiés, il n'est pas facile d'en faire la classification. Entre 1935 et 1980, 14 auteurs ont proposé 16 classifications différentes basées essentiellement sur la morphologie et la biochimie des algues (Chapman *et al.*, 1998; Bold et Wynne, 1985). En 1984, une classification des algues vertes toujours utilisée comme modèle a été publiée (Mattox et Stewart, 1984). Les cinq classes proposées comprennent les classes Micromonadophyceae (incluant les genres *Micromonas*, *Pedinomonas*, *Pyramimonas*, *Mesostigma*, *Scourfielda*, etc.), Charophyceae (Chlorokybales, Klebsormidiales, Zygnematales, Coleochaetales et Charales), Ulvophyceae (incluant les genres *Ulothrix*, *Cladophora*, *Ctenocladus* et les genres du Trentepohliaceae), Pleurastrorhynchales (Tetraselmiales et Pleurastrales) et Chlorophyceae (Chlamydomonadales, Volvocales, Chlorococcales, Sphaeropleales, Chlorosarcinales, Chaetophorales et Oedogoniales) (Chapman *et al.*, 1998). Plusieurs analyses phylogénétiques ont été et sont encore de nos jours construites afin de tester ce standard (Chapman *et al.*, 1998).

Au milieu des années 80, la biologie moléculaire et en particulier le séquençage de gènes, a permis aux phycologistes d'obtenir relativement facilement un grand nombre de caractères utiles pour les analyses phylogénétiques (Chapman *et al.*, 1998). Le premier gène étudié a été celui de l'ARN ribosomique nucléaire 5S mais ce gène évolue rapidement et son utilisation s'est avérée peu appropriée pour l'étude des algues (Chapman *et al.*, 1998). Les études ont donc été concentrées sur le gène nucléaire de la petite sous-unité (18S) de l'ARN ribosomique ainsi que sur des gènes chloroplastiques (principalement celui de la grande sous-unité de l'enzyme Rubisco). Des analyses phylogénétiques, basées autant sur des caractères morphologiques que moléculaires, ont permis de classer les plantes vertes en deux grandes lignées, Chlorophyta et Streptophyta. Ces analyses présentent les plantes terrestres et les algues vertes de la classe Charophyceae dans la lignée Streptophyta et les quatre autres classes

d'algues vertes de Mattox et Stewart (1984) dans la lignée monophylétique Chlorophyta. Selon Friedl (Friedl, 1995), les quatre classes d'algues de la lignée Chlorophyta sont les classes Chlorophyceae, Trebouxiophyceae, Ulvophyceae et Prasinophyceae (figure 1). Plusieurs études supportent cette classification. Cependant, les liens entre certains organismes de chacune des classes (Chapman *et al.*, 1998), l'appartenance de certaines algues à une classe ainsi que leur positionnement phylogénétique à l'intérieur de cette classe ne sont pas résolus parfaitement.

1.1.1.2 Les prasinophytes

La classe Prasinophyceae (Micromonadophyceae selon Mattox et Stewart, 1984) représente une lignée hétérogène d'algues unicellulaires dont la plupart se situent à la base des chlorophytes (figure 1) (Friedl, 1997). Durant les 30 dernières années, toutes les algues vertes flagellées ne pouvant être classées dans un autre groupe étaient placées dans le groupe des prasinophytes. Environ 16 genres et 180 espèces, retrouvés en eau douce ou salée, font partie de ce groupe (Van Den Hoek *et al.*, 1995). La plupart de ces algues sont mobiles, possèdent entre un et huit flagelles et leur corps, ainsi que leurs flagelles, sont souvent recouverts d'écailles organiques non-minérales (Chapman *et al.*, 1998). Il n'y a pas de caractère morphologique unique qui unit ce groupe incluant les algues vertes les plus ancestrales (Friedl, 1997). Cette classe semble être paraphylétique, c'est-à-dire qu'elle serait constituée d'une série de lignées dérivées de l'ancêtre commun des plantes vertes et ayant un lien imprécis avec le reste des plantes vertes.

Au début du siècle, des hypothèses prônaient que la présence de flagelles était une caractéristique ancestrale et que les algues vertes en possédant auraient donné naissance à des niveaux organisationnels plus complexes d'algues vertes (Friedl, 1997). La présence d'écailles chez les prasinophytes appuie le fait que ces algues vertes soient à la base des streptophytes et des chlorophytes puisque des écailles typiques aux prasinophytes ont aussi été retrouvées chez des charophytes et des ulvophytes mais nul part en dehors des plantes vertes (Melkonian et Surek, 1995). Plusieurs analyses de la petite sous-unité de l'ARN ribosomique nucléaire démontrent que les prasinophytes forment un groupe paraphylétique à la base de la radiation

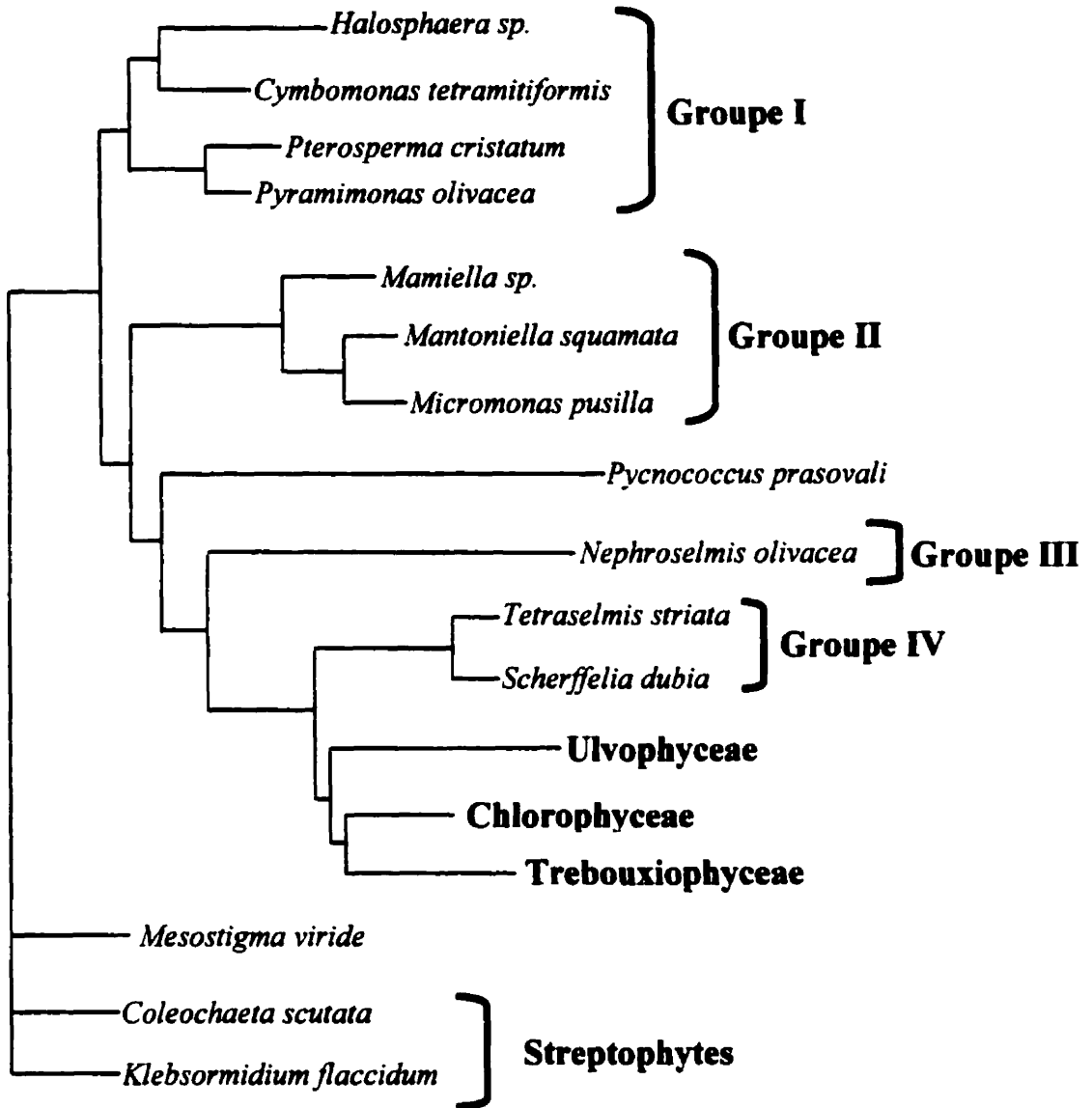
des chlorophytes, ulvophytes et trebouxiophytes (Bhattacharya et Medlin, 1998; Friedl, 1997; Melkonian et Surek, 1995; Melkonian *et al.*, 1995; Steinkötter *et al.*, 1994).

Certaines lignées de prasinophytes s'embranchent près du point de divergence des lignées Chlorophyta et Streptophyta. Récemment, des analyses phylogénétiques de séquences en acides aminés concaténées de gènes chloroplastiques et mitochondriaux ainsi que la caractérisation des génomes chloroplastiques et mitochondriaux du prasinophyte *Mesostigma viride* ont démontré que cette algue verte ne fait pas partie des chlorophytes et qu'elle aurait émergé avant la divergence des lignées Chlorophyta et Streptophyta (Lemieux *et al.*, 2000; résultats non-publiés du laboratoire). Des analyses similaires ont permis de placer le prasinophyte *Nephroselmis olivacea* à la base de la lignée Chlorophyta (Turmel *et al.*, 1999a).

En 1990, Melkonian classifiait les prasinophytes en quatre ordres, les Mamiellales, les Pseudoscourfieldiales, les Chlorodendrales et les Pyramimonadales (Melkonian, 1990). Les Mamiellales sont les prasinophytes exhibant la structure la plus simple. Ils présentent plus de caractères primitifs que les autres classes de chlorophytes. Ces prasinophytes sont phylogénétiquement reliés aux Pyramimonadales. Les Chlorodendrales sont des prasinophytes avancés ayant des liens avec la classe Chlorophyceae. Les Pseudoscourfieldiales et les Chlorodendrales ont des affinités clairement démontrées mais les Pseudoscourfieldiales ont aussi des caractéristiques communes avec les Pyramimonadales et les Mamiellales.

Des analyses phylogénétiques de la petite sous-unité de l'ARN ribosomique nucléaire révèlent qu'effectivement on peut diviser les prasinophytes en quatre groupes indépendants à la base des chlorophytes (Nakayama *et al.*, 1998) (figure 2). Le groupe I est représenté par les genres *Halosphaera*, *Cymbomonas*, *Pterosperma* et *Pyramimonas*, les genres *Mamiella*, *Mantoniella* et *Micromonas* constituent le groupe II, le groupe III comprend les genres *Pseudoscourfieldia* et *Nephroselmis* et le groupe IV les genres *Tetraselmis* et *Scherffelia*. Chaque sous-groupe représente un des quatre ordres de prasinophytes proposés auparavant par Melkonian (Melkonian, 1990) : les Pyramimonadales (groupe I), les Mamiellales (groupe II), les Pseudoscourfieldiales (groupe III) et les Chlorodendrales (groupe IV).

Figure 2. Phylogénie des Viridiplantae basée sur la séquence de la petite sous-unité de l'ARN ribosomique nucléaire et présentant les quatre groupes de prasinophytes. Cette figure est une version simplifiée de la figure 2 de Nakayama *et al.*, 1998. La position relative de *Mesostigma viride* est indiquée.



1.1.2 Des modèles pour étudier l'évolution des plantes vertes

Des caractères morphologiques et biochimiques ont d'abord été utilisés pour étudier l'évolution des plantes vertes. Un des problèmes rencontrés avec ces méthodes était la difficulté à assigner une valeur phylogénétique à certains caractères entre des groupes distants dont la morphologie varie (Bhattacharya et Medlin, 1998). Avec l'avènement de la biologie moléculaire et du séquençage, les chercheurs ont commencé à utiliser des séquences de gènes nucléaires ou chloroplastiques comme outils phylogénétiques. Même si ces séquences de gènes se sont avérées utiles pour étudier les liens entre les eucaryotes photosynthétiques, il ne faut pas oublier qu'elles révèlent l'histoire évolutive du gène en question et qu'elles ne représentent pas toujours celle des lignées étudiées (Douglas, 1998). Depuis les dernières années, des génomes d'organelles sont utilisés comme modèles pour étudier l'évolution des eucaryotes photosynthétiques. La taille relativement petite de ces génomes et leur abondance dans l'ADN cellulaire rendent leur séquençage en entier envisageable dans un laps de temps plutôt court (Clegg *et al.*, 1994). Les génomes d'organelles dont le séquençage est complété s'accumulent ce qui permet de faire des analyses comparatives (Clegg *et al.*, 1994) ainsi que des analyses phylogénétiques avec des séquences de gènes concaténés plutôt qu'avec la séquence d'un seul gène. Ceci augmente le nombre de sites informatifs disponibles et apporte des résultats plus concluants sur l'ensemble de la séquence de l'ADN utilisé. La caractérisation des génomes et les comparaisons entre ceux-ci permettent d'étudier l'architecture générale des génomes en plus d'apporter de nouvelles connaissances sur les forces qui dictent l'évolution du génome d'un organisme par rapport à un autre et sur l'origine endosymbiotique des organelles (Clegg *et al.*, 1994).

Le génome chloroplastique semble le modèle idéal pour étudier l'évolution des plantes vertes (Clegg *et al.*, 1994). La vitesse d'évolution des génomes chloroplastiques des plantes vertes est relativement lente ce qui implique que le contenu et l'organisation des gènes de ces génomes sont conservés. Ces deux caractéristiques permettent d'utiliser plusieurs espèces, dont certaines éloignées les unes des autres du point de vue évolutif, pour faire des analyses et en particulier des analyses phylogénétiques. Les génomes mitochondriaux sont moins intéressants pour l'étude évolutive des algues vertes. Ils contiennent moins de gènes et

l'organisation ainsi que le contenu en gènes diffèrent beaucoup plus entre les différentes espèces que chez les génomes chloroplastiques. Le nombre de séquences protéiques pouvant être utilisées pour faire des phylogénies est ainsi réduit. La vitesse d'évolution de l'ADNmt des algues vertes est beaucoup plus grande que celle de l'ADNmt des plantes terrestres et de l'ADNcp des algues vertes (Lang *et al.*, 1999). Les analyses entre des organismes de lignées parfois distantes sont moins intéressantes dans ce cas. Dans les arbres phylogénétiques, une vitesse d'évolution rapide se traduit en de longues branches. Le phénomène de l'attraction des longues branches peut être observé dans ces cas et fausser les résultats (Graur et Li, 2000).

1.2 Le chloroplaste

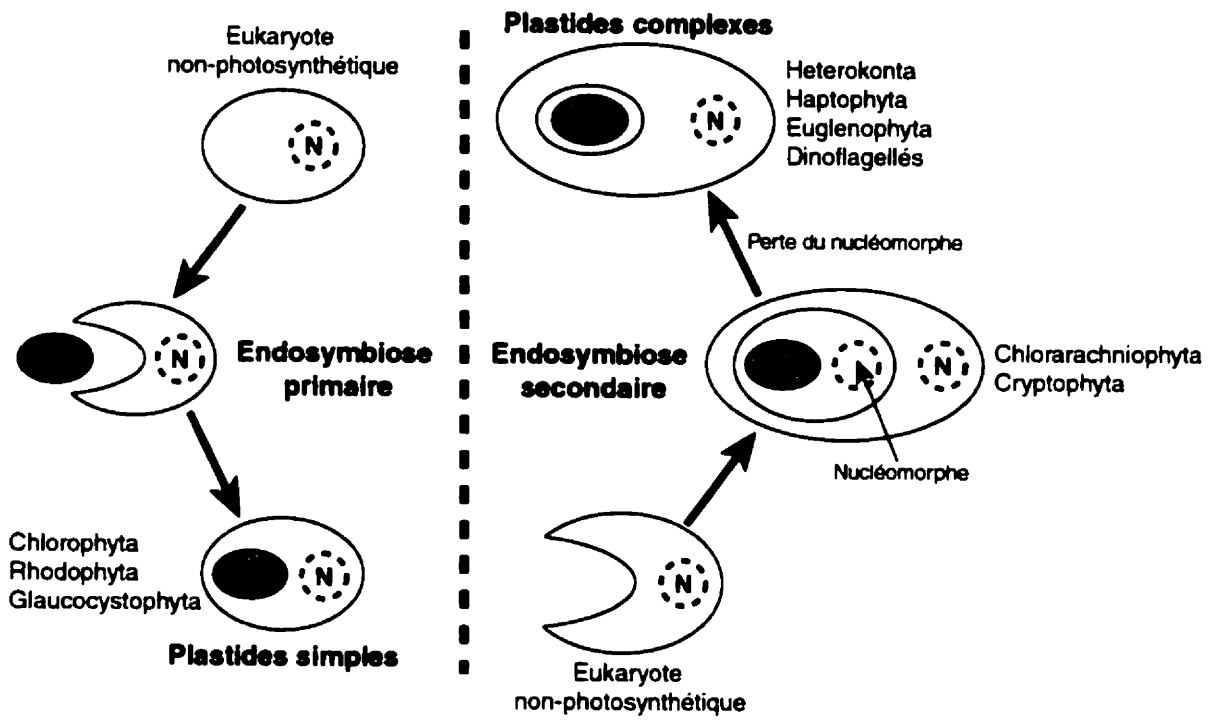
Les végétaux, les algues, certains protistes et procaryotes utilisent l'énergie solaire comme source d'énergie pour synthétiser des molécules organiques (glucides, lipides et protéines) et pour produire de l'oxygène à partir de dioxyde de carbone et d'eau. Les chloroplastes sont responsables de ce processus appelé photosynthèse (Campbell et Mathieu, 1995a). Ils possèdent leur propre génome ainsi que toute la machinerie nécessaire à la photosynthèse en plus de participer à la biosynthèse d'acides aminés, de nucléotides, de lipides et d'amidon.

1.2.1 L'origine des chloroplastes

Les plastides (chloroplastes) prennent origine d'un événement d'endosymbiose impliquant l'engouffrement d'une cyanobactérie photosynthétique par un eucaryote non-photosynthétique (figure 3). Suite à cet événement, la taille et le contenu en gènes du génome de l'endosymbiont ont été largement réduits. C'est d'ailleurs cette réduction et la dépendance consécutive du chloroplaste envers la cellule-hôte qui distingue cet organelle endosymbiotique d'un endosymbiont obligatoire (Delwiche, 1999).

Deux types d'endosymbiose ont donné naissance à deux types de chloroplastes : les chloroplastes simples des lignées Chlorophyta, Rhodophyta et Glaucocystophyta et les chloroplastes complexes des autres algues (Bhattacharya et Medlin, 1998). Les chloroplastes simples, entourés par une double membrane, proviennent d'un événement d'endosymbiose

Figure 3. Schéma présentant l'origine des plastides simples et complexes par endosymbiose primaire ou secondaire. Les lignées d'algues contenant le plastide approprié sont indiquées près de ce dernier. N, Noyau; CB, Cyanobactérie; Cp, Chloroplaste. Figure tirée de Bhattacharya et Medlin, 1998.



primaire impliquant une cyanobactérie et un eucaryote non-photosynthétique. Une endosymbiose secondaire, impliquant l'endosymbiose d'une algue rouge ou verte par un eucaryote non-photosynthétique, a donné naissance à des chloroplastes entourés par trois membranes ou plus (chloroplastes complexes) (Palmer et Delwiche, 1998) (figure 3). Les lignées Chlorarachniophyta et Euglenophyta ont pour origine une algue verte tandis que les lignées Cryptophyta, Heterokontophyta et Haptophyta proviennent d'algues rouges. Les dinoflagellés ont une évolution complexe mais la plupart font partie de la lignée des algues rouges. Une endosymbiose tertiaire impliquant un eucaryote non-photosynthétique et un haptophyte aurait même donné naissance à certains dinoflagellés (Delwiche, 1999).

L'hypothèse de l'origine endosymbiotique des plastides a été formulée en 1970 (Margulis, 1970) mais elle avait déjà été observée au début du siècle (Mereschkowsky *et al.*, 1905). Cette hypothèse est maintenant bien établie (Delwiche, 1999). D'après des évidences provenant de génomes chloroplastiques (Martin *et al.*, 1998; Palmer et Delwiche, 1998) et mitochondriaux (Palmer et Delwiche, 1998; Burger *et al.*, 1999), tous les plastides semblent provenir d'un même événement d'endosymbiose primaire. Récemment, des chercheurs ont apporté la première évidence fortement supportée provenant de gènes nucléaires d'une origine commune des algues rouges et des algues vertes. Ces résultats apportent aussi une faible évidence que les glaucocystophytes proviennent du même événement d'endosymbiose (Moreira *et al.*, 2000). L'origine unique de tous les plastides semblent maintenant certaine mais le nombre d'endosymbioses secondaires est encore controversé (Palmer, 2000). Plusieurs spéculations ont aussi été faites sur la période durant laquelle s'est produite l'endosymbiose primaire. Selon Cavalier-Smith, ce serait il y a 600 millions d'années mais des fossiles d'algues rouges datent de 1,2 billions d'années et la cyanobactérie la plus ancestrale a 2,7 billions d'années (Palmer, 2000).

1.2.2 Le génome chloroplastique des plantes vertes

1.2.2.1 Caractéristiques générales

Tout comme le génome des cyanobactéries, le génome chloroplastique est circulaire et

l'organisation de ses gènes est dense et compacte (Palmer et Delwiche, 1998; Sugiura, 1992). La majorité des ADNcp ont une taille variant entre 100 et 200 kb, ce qui est 20 à 30 fois plus petit que la taille des génomes des cyanobactéries (Palmer et Delwiche, 1998). Les génomes chloroplastiques sont aussi caractérisés par un pourcentage en bases A + T élevé. Les principaux traits des génomes chloroplastiques des plantes vertes sont résumés dans le tableau I. Ce tableau présente les caractéristiques de six algues vertes et trois plantes terrestres dont la séquence entière du génome chloroplastique est disponible. Le contenu en gènes, l'ordre des gènes et la séquence nucléotidique de l'ADNcp des plantes terrestres sont très conservés. Très peu de cas de perte de gènes et de réarrangement génomique ont été observés (Palmer, 1991). Les ADNcp d'algues vertes qui ont été étudiés ont une structure similaire à celle des plantes terrestres mais sont beaucoup plus variables du point de vue de la taille, de l'organisation des gènes et de la complexité du génome (Palmer, 1991). Chez les algues vertes, le génome chloroplastique du prasinophytes à la base des chlorophytes (*Nephroselmis olivacea* (Turmel *et al.*, 1999a)) et de celui à la base des plantes vertes (*Mesostigma viride* (Lemieux *et al.*, 2000)) sont les plus conservés et de bons modèles pour l'étude des algues vertes. Notre laboratoire a aussi séquencé et analysé le génome de deux autres prasinophytes, *Pycnococcus prasoali* et *Pedinomonas minor* (résultats non-publiés de notre laboratoire). Ces génomes sont petits et ont perdu plusieurs gènes (tableaux I et II) trouvés chez *N. olivacea* et *M. viride*. Le génome de *P. prasoali*, en particulier, est très réarrangé. Celui de *Pedinomonas minor* semble plus apparenté aux génomes des trebouxiophytes *Chlorella vulgaris* (Wakasugi *et al.*, 1997) et *Neochloris pseudoalveolaris* (résultats non-publiés de notre laboratoire) qu'aux génomes des prasinophytes disponibles.

1.2.2.2 Les régions répétées

Les génomes chloroplastiques contiennent une proportion beaucoup plus grande de séquences répétées que les génomes des cyanobactéries (Palmer et Delwiche, 1998). La majorité de l'ADN répété des chloroplastes consiste en un élément d'ADN inversé répété (IR) qui définit l'architecture de la plupart des génomes chloroplastiques de plantes vertes (Palmer et Delwiche, 1998). Ces génomes sont divisés en quatre segments : deux segments d'ADN inversé répété (IR_A et IR_B), un long segment d'ADN unique (LSC) et un court segment

Tableau I. Caractéristiques générales des génomes chloroplastiques des plantes vertes dont les séquences entières sont disponibles.

Traits	Algues vertes ^a						Plantes terrestres ^a			
	<i>M.vi</i>	<i>N.ol</i>	<i>P.pr</i>	<i>P.mi</i>	<i>N.ps</i>	<i>C.vu</i>	<i>M.po</i>	<i>P.th</i>	<i>N.ta</i>	
Longueur (pb)^b										
Totale	118 360	200 799	80 211	98 350	145 947	150 613	121 024	119 707	155 939	
IR	12 112	92 272	0	N/D ^c	13 570	0	10 058	495	25 340	
LSC	83 267	92 126	80 211	N/D ^c	115 979	150 613	81 095	65 696	86 689	
SSC	22 981	16 401	0	N/D ^c	16 398	0	19 813	53 021	18 575	
% répétitions	10,2	45,9	0	N/D ^c	9,2	0	16,6	0,8	32,5	
Nombre de gènes	137	127	98	104	106	111	120	110	113	
Introns										
groupe I	0	0	1	0	0	3	1	1	1	
groupe II	0	0	0	0	0	0	18	15	20	
Densité en gènes (nombre de gènes/kb) ^b	1,16	0,63	1,22	1,06	0,73	0,74	0,99	0,90	0,78	
% A + T	69,8	57,9	N/D ^c	N/D ^c	68,4	68,4	71,2	61,5	62,1	

^a *M.vi*, *Mesostigma viride*; *N.ol*, *Nephroselmis olivacea*; *P.pr*, *Pycnococcus prasovalii*; *P.mi*, *Pedinomonas minor*; *N.ps*, *Neochloris pseudoalveolaris*; *C.vu*, *Chlorella vulgaris*; *M.po*, *Marchantia polymorpha*; *P.th*, *Pinus thunbergii*; *N.ta*, *Nicotiana tobaccum*.

^b pb, paires de bases; kb, kilo bases.

^c N/D signifie que la donnée n'était pas disponible.

d'ADN unique (SSC) (Palmer et Delwiche, 1998; Sugiura, 1992). La figure 4, présentant la carte des gènes de l'ADNcp de *Mesostigma viride*, montre cette organisation. Les résultats recueillis jusqu'à maintenant permettent de dire que la taille de la région d'ADN inversé répété varie entre 5 et 76 kb. Chez les plantes terrestres, cette différence est responsable de la variation de la taille de l'ADNcp en entier (Sugiura, 1998). L'opéron des ARNs ribosomiques est présent dans chaque segment (IR_A et IR_B) d'ADN inversé répété. Le contenu en gènes, l'espacement entre les deux segments répétés (IR_A et IR_B) ainsi que la localisation et l'orientation de l'opéron des ARNs ribosomiques par rapport à la petite région d'ADN unique sont variables (Palmer et Delwiche, 1998). Les propriétés de recombinaison entre les deux segments d'ADN inversé répété (IR_A et IR_B) peuvent être responsables des variations et des insertions/délétions entre les régions répétées des différents génomes chloroplastiques.

Aucune région d'ADN inversé répété telle que celle trouvée chez les chloroplastes n'a été trouvée chez les génomes de cyanobactéries analysés mais l'opéron des ARNs ribosomiques a été repéré en deux copies inversées chez certains génomes de cyanobactérie (Sugiura, 1998). Palmer a suggéré, en 1985, que la présence de la région d'ADN inversé répété stabilisait l'ADNcp en diminuant le nombre de recombinaisons interchromosomales (Palmer, 1985). Il a cependant été observé que certains génomes chloroplastiques très réarrangés de plantes terrestres contiennent une région inversée répétée. Il est donc clair que la perte de cette région n'est pas la seule cause de déstabilisation de l'ADNcp (Cosner *et al.*, 1997).

D'autres types de séquences répétées ont été répertoriés dans l'ADNcp de plantes vertes. Des régions de répétitions dispersées dans tout le génome et composées d'éléments répétés plus courts arrangés de façon variable entre les différentes régions ont été trouvées chez *Chlamydomonas reinhardtii* et *Chlamydomonas gelatinosa*, des algues vertes de la classe Chlorophyceae. De courts éléments répétés se repliant en forme d'épingles à cheveu ont aussi été trouvés chez *Chlamydomonas moewusii* et *Chlamydomonas pitschmanii* (Boudreau et Turmel, 1996). Les génomes chloroplastiques des chlamydomonadales sont les seuls, chez les algues vertes étudiées, chez qui on a répertorié des séquences répétées ne faisant pas partie de la région d'ADN inversé répété. Chez les plantes terrestres, à l'exception de la région d'ADN inversée répétée, très peu d'ADNcp contiennent des séquences répétées, fait probablement

Figure 4. Génome chloroplastique du prasinophyte *Mesostigma viride*. Les segments d'ADN inversé répété (IR_A et IR_B), la longue séquence d'ADN unique (LSC) et la courte séquence d'ADN unique (SSC) sont indiqués. Les gènes à l'extérieur de la carte sont transcrits dans le sens horaire et ceux situés à l'intérieur du cercle, dans le sens anti-horaire. Figure tirée de Lemieux *et al.*, 2000.

lié à une contrainte de taille dans la molécule d'ADNcp (Palmer, 1985). Dans les régions intergéniques du génome chloroplastique du pissenlit, des répétitions directes ainsi que des régions répétées se repliant en épingles à cheveu et associées à des insertions/délétions dans le génome ont été trouvées (Mes *et al.*, 2000). De courtes régions répétées dispersées ont aussi été identifiées dans le génome chloroplastique très réarrangé de *Trachelium caeruleum*, dans celui du géranium, du blé, du riz, du trèfle, de la fève et du sapin de Douglas. Chez ces espèces, les séquences répétées ont été associées avec des points d'inversion dans l'organisation des gènes du génome, laissant supposer que les séquences répétées facilitaient les inversions et ainsi le réarrangement génomique par recombinaison homologue (Cosner *et al.*, 1997). L'origine de ces séquences répétées ainsi que le pourquoi de leur présence n'est pas connu (Cosner *et al.*, 1997; Palmer, 1991).

1.2.2.3 Les gènes chloroplastiques

Dans la transition entre l'endosymbiont eubactérien et l'organelle, la majorité de l'information génétique de l'endosymbiont (environ 90 % des gènes) a été perdue (Palmer et Delwiche, 1998). Trois mécanismes sont responsables de ces pertes (Delwiche, 1999; Palmer et Delwiche, 1998). Premièrement, les gènes qui n'apportaient plus d'avantage sélectif dans le contexte de la symbiose ont été perdus complètement, sans être remplacés. Dans le deuxième cas, des gènes ayant toujours une fonction essentielle au chloroplaste ont été perdus mais leur fonction a été remplacée par des gènes du noyau. Le troisième mécanisme est le transfert de gènes chloroplastiques au génome nucléaire. Les chloroplastes contiennent plus de 1000 protéines dont environ 90 %, chez les chlorophytes, sont codés par des gènes nucléaires, résultat des pertes massives de gènes (Palmer et Delwiche, 1998). Différents degrés de perte, réarrangements et transfert de gènes sont observés entre les différentes lignées d'algues et de plantes. Les génomes chloroplastiques des plantes vertes contiennent autour de 100 gènes (Palmer et Delwiche, 1998). La majorité de ces gènes sont impliqués dans la photosynthèse et l'expression génétique. Les principaux gènes retrouvés dans le génome chloroplastique des plantes vertes ainsi que leurs fonctions sont énumérés dans les paragraphes suivants. Le tableau II présente la distribution des gènes codant pour des protéines et des ARNs répertoriés dans le génome chloroplastique des plantes vertes dont la séquence complète de l'ADNcp est

disponible.

Les gènes impliqués dans la photosynthèse

Des gènes de l'ADNcp codent pour les sous-unités de quatre systèmes importants intégrés dans la membrane des thylakoïdes (Campbell et Mathieu, 1995a) et impliqués dans la photosynthèse : le photosystème I, le photosystème II, le cytochrome *b6/f* et l'ATP synthétase (Campbell et Mathieu, 1995a). Les gènes codant pour des sous-unités des protéines Rubisco et protochlorophyllide réductase sont aussi présents dans le génome des chloroplastes.

Parmi les gènes photosynthétiques, ceux des photosystèmes sont les plus nombreux. Les photosystèmes sont les unités photoréceptrices; ce sont eux qui captent l'énergie lumineuse (Campbell et Mathieu, 1995a). Leur différenciation s'est faite au cours de l'évolution. Les gènes *psa* et *psb* codent respectivement pour les composantes des photosystèmes I et II. Chez les plantes vertes, l'ADNcp contient six gènes codant pour des sous-unités du photosystème I et 14 gènes codant pour des composantes du photosystème II. Les gènes *pet* codent pour les six sous-unités qui composent le cytochrome *b6/f*. Ce système sert à faire la navette entre les deux photosystèmes (Campbell et Mathieu, 1995a). L'ATP synthétase produit l'ATP nécessaire à la deuxième phase de la photosynthèse (cycle de Calvin) à l'aide des protons produits par les réactions photochimiques (première phase de la photosynthèse) (Campbell et Mathieu, 1995a). Deux parties, le CF₀ et le CF₁ composent l'ATP synthétase. Le CF₀ comprend lui-même cinq différentes sous-unités et le CF₁ en comprend quatre. Les six gènes *atp* du génome chloroplastique des plantes vertes codent pour des sous-unités du CF₀ et du CF₁ de l'ATP synthétase (Sugiura, 1992).

La ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) est la protéine la plus abondante dans le stroma (liquide dense à l'intérieur des chloroplastes dans lequel baignent les thylakoïdes (Campbell et Mathieu, 1995a)) des chloroplastes (Sugiura, 1992). Cette protéine est impliquée dans la photorespiration (deuxième phase de la photosynthèse) (Campbell et Mathieu, 1995a). Elle est composée de huit grandes sous-unités identiques (LS) de 55 kDa et de huit petites sous-unités identiques (SS) de 12 kDa. Chez les plantes vertes, les sous-unités

Suite Tableau II

Gènes	Algues vertes						Plantes terrestres		
	<i>M.vi</i>	<i>N.ol</i>	<i>P.pr</i>	<i>P.mi</i>	<i>N.ps</i>	<i>C.vu</i>	<i>M.po</i>	<i>P.th</i>	<i>N.ta</i>
<i>P(ggg)</i>			•				•	•	
<i>P(ugg)</i>	•	•	•	•	•	•	•	•	•
<i>Q(uug)</i>	•	•	•	•	•	•	•	•	•
<i>R(acg)</i>	•	•	•	•	•	•	•	•	•
<i>R(ccg)</i>		•	•	•	•	•	•		
<i>R(ucu)</i>	•	•	•	•	•	•	•	•	•
<i>S(cga)</i>	•	•	•						
<i>S(gcu)</i>	•	•	•	•	•	•	•	•	•
<i>S(gga)</i>	•	•	•	•	•	•	•	•	•
<i>S(uga)</i>	•	•	•	•	•	•	•	•	•
<i>T(ggu)</i>	•	•		•	•	•	•	•	•
<i>T(ugu)</i>	•	•	•	•	•	•	•	•	•
<i>V(gac)</i>	•						•	•	•
<i>V(uac)</i>	•	•	•	•	•	•	•	•	•
<i>W(cca)</i>	•	•	•	•	•	•	•	•	•
<i>Y(gua)</i>	•	•	•	•	•	•	•	•	•
ARN polymérase									
<i>rpoA</i>	•	•	•	•	•	•	•	•	•
<i>rpoB</i>	•	•		•	•	•	•	•	•
<i>rpoC1</i>	•	•	•	•	•	•	•	•	•
<i>rpoC2</i>	•	•	•	•	•	•	•	•	•
Facteurs impliqués dans la traduction									
<i>infA</i>	•	•	•		•	•	•	•	•
<i>tufA</i>	•	•	•	•	•	•			
<i>clpP</i>	•	•	•	•	•	•	•	•	•
Protéines de division chloroplastique									
<i>minD</i>	•	•		•	•	•			
<i>ftsH</i>	•	•	•	•	•	•	•	•	•
<i>ftsI</i>	•	•							
<i>ftsW</i>	•	•							
Autres protéines									
<i>accD</i>		•		•	•	•	•	•	•
<i>ccsA</i>	•	•		•	•	•	•	•	•
<i>cemA</i>	•	•	•	•	•	•	•	•	•
<i>cysA</i>	•	•		•		•	•		

Suite Tableau II

Gènes	Algues vertes						Plantes terrestres		
	<i>M.vi</i>	<i>N.ol</i>	<i>P.pr</i>	<i>P.mi</i>	<i>N.ps</i>	<i>C.vu</i>	<i>M.po</i>	<i>P.th</i>	<i>N.ta</i>
<i>cysT</i>	•	•	•	•		•	•		
<i>odpB</i>	•								
Autres petits ARNs									
<i>rne</i>		•							
<i>rnpB</i>		•	•						
ORFs conservés									
<i>ycf1</i>	•	•	•	•	•	•	•	•	•
<i>ycf3</i>	•	•	•	•	•	•	•	•	•
<i>ycf4</i>	•	•	•	•	•	•	•	•	•
<i>ycf9</i>	•	•	•	•	•	•	•	•	•
<i>ycf12</i>	•	•	•	•	•	•	•	•	
<i>ycf20</i>	•			•		•			
<i>ycf61</i>	•								
<i>ycf62</i>	•	•			•				
<i>ycf65</i>	•								
<i>ycf66</i>	•						•		
<i>ycf81</i>	•	•							

^a *M.vi*, *Mesostigma viride*; *N.ol*, *Nephroselmis olivacea*; *P.pr*, *Pycnococcus prasoali*; *P.mi*, *Pedinomonas minor*; *N.ps*, *Neochloris pseudoalveolaris*; *C.vu*, *Chlorella vulgaris*; *M.po*, *Marchantia polymorpha*; *P.th*, *Pinus thunbergii*; *N.ta*, *Nicotiana tobaccum*.

^b Les cercles noirs indiquent la présence du gène chez cet organisme.

^c Les ARNt sont indiqués par le code à une lettre des acides aminés suivi de l'anticodon entre parenthèses.

LS sont codées dans l'ADN chloroplastique par le gène *rbcL* mais le gène *rbcS*, codant pour les petites sous-unités, se retrouvent dans l'ADN nucléaire. Une deuxième protéine photosynthétique importante est la protéine protochlorophyllide réductase qui est responsable de la biosynthèse de la chlorophylle. Les gènes chloroplastiques codant pour ses sous-unités sont les gènes *chlB*, *I*, *L* et *N* (Wakasugi *et al.*, 1997). Le gène *chlI* est retrouvé chez les algues vertes mais il n'a pas été répertorié chez les plantes terrestres.

Onze gènes chloroplastiques (*ndh*) codent pour des protéines similaires aux composantes de la chaîne respiratoire NADH déshydrogénase des mitochondries (Sugiura, 1992). Ils ont été trouvés dans l'ADNcp de plusieurs plantes terrestres (Sugiura, 1992) mais aussi chez les algues vertes *Nephroselmis olivacea* (Turmel *et al.*, 1999a) et *Mesostigma viride* (Lemieux *et al.*, 2000). Il a été montré récemment que les protéines Ndh des chloroplastes du tabac font partie d'un complexe fonctionnel indispensable à la croissance de la plante dans des conditions optimales (Burrows *et al.*, 1998; Shikanai *et al.*, 1998). Ce complexe semble participer à la chaîne cyclique d'électrons autour du photosystème I lors des deux phases de la photosynthèse (Turmel *et al.*, 1999a).

Les gènes impliqués dans l'expression génétique

La traduction des gènes chloroplastiques fait appel à des ribosomes 70S composés de deux sous-unités, une de 30S et une de 50S (Sugiura, 1992). Les gènes *rps* codent pour les protéines de la petite sous-unité (30S) du ribosome tandis que les gènes *rpl* codent pour les protéines de la grande sous-unité (50S). Il est à noter que l'ADNcp code seulement pour le tiers des 60 protéines ribosomales (Sugiura, 1992). Il y a trois principaux types d'ARNs ribosomiques dans le génome chloroplastique. Le 23S ou grande sous-unité de l'ARNr (LSU) et le 5S sont associés à la sous-unité 50S du ribosome. L'ARNr associé à la sous-unité 30S du ribosome (Sugiura, 1992) est le 16S ou petite sous-unité de l'ARNr (SSU). Le gène *rrl* code pour l'ARN ribosomique 23S, le gène *rrs* code pour l'ARN ribosomique 16S et le gène *rrf* code pour l'ARN ribosomique 5S. Un ARNr 4,5S, aussi associé à la sous-unité 50S du ribosome, a été trouvé dans l'ADNcp des plantes terrestres évoluées.

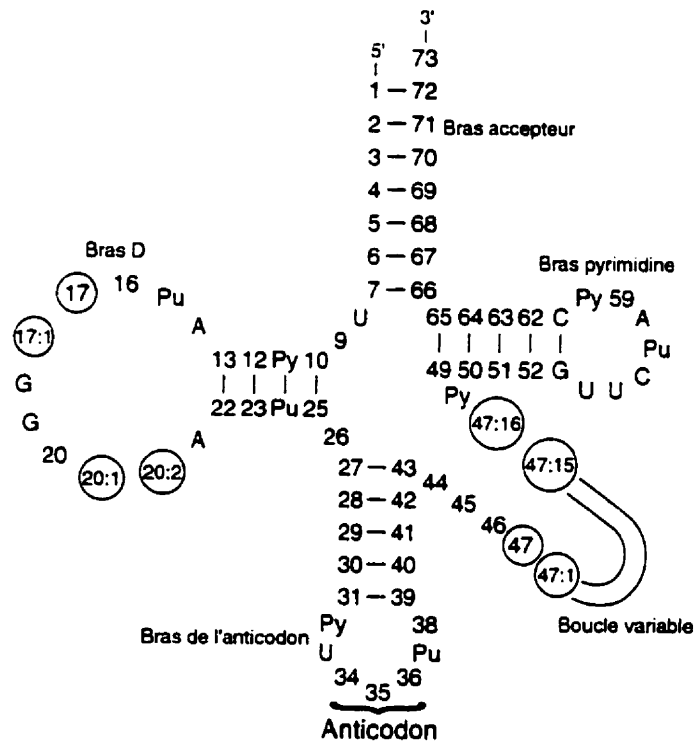
Il y a entre 20 et 40 gènes *trn* codant pour les 30-35 ARNs de transfert présents dans le génome chloroplastique. Normalement, ces ARNs de transfert sont suffisants pour lire les 61 codons du code génétique universel, en assumant des cas de dégénérescence entre la première base des anticodons commençant par un U et la troisième base des codons associés (Sugiura, 1992). Tous les ARNs chloroplastiques se replient en forme de trèfle à quatre feuilles, selon la conformation standard (figure 5).

Les gènes chloroplastiques *rpoA*, *B*, *C1* et *C2*, codent pour des sous-unités similaires aux sous-unités α (*rpoA*), β (*rpoB*) et β' (*rpoC*) de l'ARN polymérase de *E. coli*. L'homologue du gène *rpoC* dans l'ADNcp est divisé en deux sous-unité, soit *rpoC1* et *rpoC2*. La présence de ces gènes indique que les chloroplastes synthétisent plusieurs si ce n'est toutes les sous-unités de l'ARN polymérase. Le gène *tufA*, qui code pour le facteur d'élongation EF-Tu ainsi que le gène *infA*, qui code pour le facteur d'initiation IF-I, sont tous deux impliqués dans la traduction et sont aussi présents dans le génome chloroplastique des plantes vertes. Il est cependant à noter que le gène *tufA* n'a pas été répertorié chez les plantes terrestres. Le gène *clpP*, présent dans tous les ADNcp des plantes vertes, code pour la sous-unité protéolytique d'une protéase dépendante de l'ATP impliquée dans la maturation des protéines cytoplasmiques. Cette protéine dégrade les polypeptides incomplets et les protéines non-assemblées dans les chloroplastes (Sugiura, 1992).

Les gènes de division

Les gènes *minD*, *minE*, *ftsH* (ou *ycf2*), *ftsI* et *ftsW* sont des gènes homologues aux gènes bactériens de *E. coli* codant pour des enzymes impliquées dans la division du chloroplaste. La trouvaille du gène *ftsI* chez *Nephroselmis olivacea* et *Mesostigma viride* a été surprenante car le produit de ce gène catalyse la croissance d'une couche de peptidoglycan et aucune enveloppe chloroplastique d'algue verte ne semble posséder cette couche. La présence de ces gènes bactériens suggère que les mécanismes de division cellulaire ont été conservés au cours de l'évolution des chloroplastes (Turmel *et al.*, 1999a).

Figure 5. Structure secondaire standard d'un ARNt. Les chiffres indiquent les positions (du 5' au 3') qui sont toujours présentes et qui peuvent être comblées par n'importe laquelle des quatre bases. Lorsqu'une position est toujours occupée par une base en particulier, celle-ci est indiquée. Pour les bases semi-conservées, Py indique une pyrimidine (C ou U) et Pu une purine (G ou A). Les positions qui ne sont pas présentes dans tous les ARNt sont encerclées. Dans la boucle variable, toutes les positions 47 sont facultatives.



Autres gènes

On retrouve aussi les gènes *accD*, *ccsA*, *cemA*, *cysA*, *cysT*, *odpB*, *rne* et *rnpB* dans les génomes chloroplastiques des plantes vertes. Le gène *accD* code pour une acétyl-CoA carboxylase carboxytransférase, *ccsA* code pour une protéine impliquée dans la cytogénèse du cytochrome c, *cemA* (ou *ycf5*) code pour une protéine de la membrane chloroplastique alors que *cysA* et *cysT* codent pour des protéines probablement impliquées dans le transport des sulfates. Le gène *odpB* a été trouvé, parmi les plantes vertes, seulement chez *Mesostigma viride* mais il a aussi été identifié dans l'ADNcp de *Porphyra purpurea*, une algue rouge (Lemieux *et al.*, 2000). Le gène *rne*, codant pour la protéine RNaseE qui participe probablement à l'épissage de précurseurs d'ARNr et d'ARNm, a été identifié chez des algues rouges. Chez les plantes vertes, il n'a été répertorié que chez *Nephroselmis olivacea* (Turmel *et al.*, 1999a). Le gène *rnpB* code pour un ARN faisant partie du complexe RNaseP. Jusqu'à maintenant, il n'a pas été trouvé chez d'autres plantes vertes que les prasinophytes.

ORFs conservés

Les gènes *ycf* sont des gènes dont la fonction n'est pas encore connue mais qui sont conservés chez plusieurs organismes. Ceux qui sont retrouvés le plus souvent chez les plantes vertes sont *ycf1-3-4-9-12-20*. Les ORFs *ycf3* et *ycf4* ont peut-être un rôle à jouer dans l'assemblage du photosystème I. D'autres ORFs moins communs ont aussi été trouvés; *ycf61-62-65* sont similaires à des ORFs trouvés chez *Porphyra purpurea* et *ycf66* est similaire à l'orf135 de *Marchantia polymorpha*. *ycf81*, identifié seulement chez *N. olivacea* et *M. viride*, montre des similarités avec la protéine hypothétique slr1972 de *Synechocystis sp.* PCC6803.

1.2.2.4 L'organisation des gènes chloroplastiques

L'organisation fondamentale des gènes a été largement conservée entre les génomes chloroplastiques et le génome de la cyanobactérie ancestrale (Palmer et Delwiche, 1998). La plupart des gènes de l'ADNcp sont organisés en opérons (groupes de gènes co-transcrits et sur le même brin) ou regroupements de gènes (groupes de gènes pas nécessairement co-transcrits et sur le même brin mais qui sont spécifiques à une espèce et souvent conservés) dont l'ordre

et le contenu des gènes ressemblent de près aux opérons des cyanobactéries ou autres eubactéries. Dans la lignée Chlorophyta, une désintégration significative de la structure des regroupements de gènes de l'ADNcp a été observée seulement chez le genre *Chlamydomonas* (Palmer et Delwiche, 1998) ainsi que chez le prasinophyte *Pycnococcus prasoali* (résultats non-publiés). Chez les plantes terrestres, l'organisation des gènes est très conservée (Palmer, 1991); le contenu des regroupements de gènes diffère très peu entre les différentes espèces.

L'algue verte à la base des plantes vertes, *Mesostigma viride*, est celle dont le génome chloroplastique a conservé le plus de regroupements de gènes retrouvés chez les organismes des lignées Streptophyta et/ou Chlorophyta. Le tableau III présente les regroupements de gènes chloroplastiques présents chez *Mesostigma viride* et répertoriés chez au moins une autre des plantes vertes du tableau. Parmi les prasinophytes étudiés jusqu'à maintenant, *Nephroselmis olivacea* est celui dont le génome chloroplastique se rapproche le plus de l'ancêtre de la lignée Chlorophyta. C'est aussi celui qui contient le plus de regroupements en commun avec *Mesostigma*. La plupart des regroupements absents chez *Nephroselmis* sont des regroupements typiques aux plantes terrestres (tableau III). Peu de regroupements de gènes présents chez les prasinophytes *N. olivacea* et *M. viride* sont retrouvés chez le prasinophyte *Pycnococcus prasoali*. Il en contient aussi moins que chez les autres algues vertes étudiées (tableau III). Ce génome, qui ne contient pas de région d'ADN inversée répétée et qui a perdu un bon nombre de gènes, est très réarrangé comparativement aux génomes des autres prasinophytes et même des autres algues vertes séquencées. L'organisation des regroupements de gènes chez le prasinophyte *Pedinomonas minor* suit celle des trebouxiophytes *N. pseudoalveolaris* et *C. vulgaris*. La classification de cette algue verte est encore ambiguë. Chez les plantes terrestres, on peut voir que les regroupements sont très conservés.

Le regroupement 10b (*rps9-rpoA-rps11-rpl36-infA-rps8-rpl5-14-16-rps3-19-rpl2-23*) du tableau III représente l'opéron des protéines ribosomiques des Chlorophytes. Cet opéron très conservé correspond aux opérons L2-*spc*-alpha-S9 de *E. coli* (Lindhal *et al.*, 1990). Chez les eubactéries, l'opéron *str* est joint aux autres opérons dans l'ordre suivant: *str*-L2-*spc*-alpha-S9. Chez l'ancêtre commun des archéobactéries, l'opéron *str* ou S12/S10 a été séparé des autres opérons. C'est cet arrangement qui a été conservé au cours de l'évolution du génome

Tableau III. Distribution de regroupements de gènes chloroplastiques de l'ADNcp de *Mesostigma viride* chez cinq Chlorophytes et deux Streptophytes.

Regroupe- ments ^a	Chlorophytes ^b					Streptophytes ^b	
	<i>N.ol</i>	<i>P.pr</i>	<i>P.mi</i>	<i>N.ps</i>	<i>C.vu</i>	<i>M.po</i>	<i>P.th</i>
1						•	•
2			•			•	•
3	•	•	•	•	•	•	•
4	•	•	•	•	•	•	•
5	•	•	•	•	•	•	•
6						•	•
7	•		•	•	•	•	•
8	•						
9	•						
10a			• ^c			• ^d	
10b	•	• ^d	•	•	•	• ^d	• ^d
10c	•	•	•	•	•	•	•
10d	•	•	•	•	•	•	•
11a	•						
11b	•		•	•			
11c	•		•	•	•		
11d	•	•	•	•	•	•	
12	•					•	•
13	•						
14a	•				•		
14b	•	•	•		•		
15a	•					•	•
15b	•		•		•	•	•

^a 1, *rbcL-atpB-E*; 2, *rpoB-C1-C2-rps2-atpI-H-F-A*; 3, *psaA-B*; 4, *psbD-C*; 5, *psbE-F-L-J*; 6, *ycf9-G(gcc)*; 7, *psaJ-P(ugg)-W(cca)*; 8, *ftsW-N(guu)*; 9, *ftsI-psbA*; 10a, *clpP-psbB-T-N-H-petB-D-rps9-rpoA-rps11-rpl36-infA-rps8-rpl5-14-16-rps3-rpl22-rps19-rpl2-23*; 10b, *rps9-rpoA-rps11-rpl36-infA-rps8-rpl5-14-16-rps3-19-rpl2-23*; 10c, *psbB-T-N-H*; 10d, *petB-D*; 11a, *rps12-7-tufA-rpl19-ycf4-cemA-petA-L-G*; 11b, *rps12-7-tufA-rpl19-ycf4-cemA*; 11c, *rps12-7-tufA-rpl19-ycf4*; 11d, *petL-G*; 12, *psaC-ndhD*; 13, *ftsW-N(guu)-ccsA-chlL-N*; 14a, *rpl32-cysT-ycf1*; 14b, *cysT-ycf1*; 15a, *rrs-I(gau)-A(ugc)-rrl-rrf-R(acg)*; 15b, *rrs-I(gau)-A(ugc)-rrl-rrf*. Seuls les regroupements communs aux plantes vertes et à *Mesostigma viride* sont présentés.

Tableau tiré de Lemieux *et al.*, 2000.

^b *N.ol*, *Nephroselmis olivacea*; *P.pr*, *Pycnococcus prasovali*; *P.mi*, *Pedinomonas minor*; *N.ps*, *Neochloris pseudoalveolaris*; *C.vu*, *Chlorella vulgaris*; *M.po*, *Marchantia polymorpha*; *P.th*, *Pinus thunbergii*.

^c Le gène *rpl22* est absent des algues vertes et le gène *infA* n'est pas présent chez *Pedinomonas minor*.

^d Le gène *rps9* est absent chez les plantes terrestres ainsi que chez *Pycnococcus prasovali*.

chloroplastique avec de petites modifications propres aux différentes lignées d'algues et de plantes (Stoebe et Kowallik, 1999). L'opéron *str* a donné naissance, chez les algues vertes, aux opérons 11a, b et c du tableau III.

Le regroupement *atpA* est un autre exemple de la conservation des regroupements de gènes chez les génomes chloroplastiques (Stoebe et Kowallik, 1999). Ce regroupement de gènes est formé de plusieurs groupes de gènes non-reliés du point de vue fonctionnel et transcrits sur les deux brins. Un élément conservé intéressant de ce regroupement est la fusion de l'opéron *atpA* (*atpA-atpF-atpH-atpI* chez les algues vertes) avec les gènes *rps2*, *rpoB*, *rpoC1* et *rpoC2* (opéron 2 du tableau III). Cet élément n'est pas retrouvé chez les cyanobactéries, indication de son apparition suite à l'origine du chloroplaste (Stoebe et Kowallik, 1999).

Le regroupement *psbB* (regroupements 10c et 10d du tableau III) est un autre regroupement unique aux chloroplastes. L'arrangement des gènes *psbB-psbT-(psbN)-psbH-petB-petD* où le gène *psbN* est transcrit sur le brin inverse, est très conservé chez les génomes chloroplastiques. Chez les plantes terrestres, ces gènes sont retrouvés dans cet ordre mais chez les algues vertes, le regroupement *petB-petD* (regroupement 10d du tableau III) est souvent séparé (Löffelhardt et al., 1997). Les opérons *psaA-psaB*, *psbD-psbC* et *psbE-psbF-psbL-psbJ* (regroupements 3, 4 et 5 respectivement du tableau III) sont trois autres exemples d'opérons conservés présents chez tous les chloroplastes (à l'exception de l'ADNcp très réarrangé de *Chlamydomonas reinhardtii* (Löffelhardt et al., 1997)) mais pas chez les cyanobactéries. L'opéron des ARNs ribosomiques (opérons 15a et 15b du tableau III) est lui aussi très conservé chez les plantes vertes. Dans la plupart des ADNs chloroplastiques, il est dupliqué et chaque copie fait partie d'un segment d'ADN inversé répété.

1.2.2.5 Les introns chloroplastiques

Les génomes chloroplastiques des plantes vertes ont acquis des introns au cours de l'évolution; ils en contiennent entre 0 et 155 (Palmer et Delwiche, 1998), qui sont divisés en trois groupes (I, II et III). Un intron du groupe I, celui trouvé dans le gène de l'ARNt-Leu(UAA), pourrait être le plus vieil intron connu. C'est le seul qui est présent dans les

cyanobactéries de même que dans la plupart des ADNcp des chlorophytes et des plantes terrestres. Quelques algues vertes ont toutefois perdu cet intron. La présence d'un grand nombre d'introns dans le génome chloroplastique serait due à la capture d'éléments génétiques mobiles capables d'auto-épissage et à l'intégration de ces introns dans le génome (Palmer et Delwiche, 1998). Le contenu en introns varie beaucoup chez les algues mais est conservé chez les plantes terrestres (≈ 20) (Wakasugi *et al.*, 1997). Les génomes chloroplastiques des prasinophytes et des trebouxiophytes séquencés jusqu'à maintenant contiennent très peu d'introns (tableau I). Les algues vertes du genre *Chlamydomonas* faisant partie de la classe Chlorophyceae contiennent un grand nombre d'introns; autour de 21 du groupe I et deux du groupe II pour la plupart mais le contenu en introns des *Chlamydomonas* est variable. Contrairement aux plantes terrestres, l'ADNcp des algues vertes contient plus d'introns du groupe I que du groupe II (Palmer et Delwiche, 1998).

1.2.3 Les analyses phylogénétiques à l'aide de séquences de gènes chloroplastiques concaténés

Récemment, des analyses phylogénétiques faites avec la séquence peptidique concaténée de plusieurs gènes codant pour des protéines ont permis de placer le prasinophyte *Mesostigma viride* à la base des plantes vertes avec un bon niveau de confiance (Lemieux *et al.*, 2000). D'autres analyses faites auparavant avaient positionnées le prasinophyte *Nephroselmis olivacea* à la base de la lignée Chlorophyta (Turmel *et al.*, 1999a). Dans la phylogénie impliquant *M. viride*, la séquence du prasinophyte *Pedinomonas minor* a été incluse, ce qui a permis de le placer, avec un niveau de confiance élevé, avec le trebouxiophyte *Chlorella vulgaris*. Cependant, le manque de séquences de génomes chloroplastiques d'algues vertes et en particulier d'ulvophytes autour de cet embranchement ne permet pas de conclure que *P. minor* se groupe avec les trebouxiophytes. Aucune analyse phylogénétique publiée n'inclut la séquence peptidique concaténée du prasinophyte *Pycnococcus prasovali* ainsi que celle du trebouxiophyte *Neochloris pseudoalveolaris*.

Autant dans l'analyse phylogénétique de *M. viride* que dans celle de *N. olivacea*, la division entre les deux grandes lignées de plantes vertes Streptophyta et Chlorophyta est fortement

supportée. Dans l'analyse de *N. olivacea*, la séquence peptidique concaténée de trois algues rouges et d'un glaucocystophyte est incluse, montrant que les algues vertes et les algues rouges peuvent être considérées comme des lignées sœurs mais que l'origine plus ancestrale des glaucocystophytes est moins certaine (Turmel *et al.*, 1999a).

1.3 La mitochondrie

On trouve des mitochondries dans presque toutes les cellules eucaryotes. Tout comme les chloroplastes, ces organelles semi-autonomes convertissent l'énergie provenant du milieu où évolue la cellule en des formes que cette dernière pourra utiliser. Les mitochondries sont responsables de la respiration cellulaire, c'est-à-dire qu'à l'aide d'oxygène, elles produisent de l'ATP en extrayant l'énergie des glucides, lipides et d'autres substances (Campbell et Mathieu, 1995b). Ces organelles d'origine endosymbiotique possèdent, eux-aussi, une petite quantité d'ADN codant pour une partie des protéines et ARNs nécessaires à leur fonctionnement.

1.3.1 L'origine des mitochondries

La théorie de l'endosymbiose (Margulis, 1970) est le modèle expliquant l'origine des mitochondries qui est le plus accepté, surtout depuis que des analyses phylogénétiques ont clairement démontré que les mitochondries dérivent d'une lignée d'eubactéries, spécifiquement les α -protéobactéries (Gray et Doolittle, 1982). Selon cette théorie, l'acquisition de la mitochondrie aurait précédé celle du chloroplaste (Lang *et al.*, 1999). D'après les arbres phylogénétiques construits avec un gène de l'ARNr ou des gènes codant pour des protéines (Lang *et al.*, 1999; Gray *et al.*, 1999), un seul événement d'endosymbiose aurait donné naissance aux mitochondries. Les analyses de la séquence du génome de *Rickettsia prowazekii* (Gray, 1998; Andersson *et al.*, 1998; Müller et Martin, 1999), la première séquence à être complétée chez une α -protéobactérie, ont apporté la confirmation de l'origine α -protéobactérienne unique des mitochondries. Les résultats de ces analyses montrent que *Rickettsia prowazekii* est la bactérie qui, parmi les bactéries dont le génome a été étudié aussi en détail, est le plus près des mitochondries. Ces résultats ne veulent cependant

pas dire que les mitochondries dérivent directement des Rickettsiales, mais plutôt que les génomes des *Rickettsia* et des mitochondries descendent d'un ancêtre α -protéobactérien commun. La comparaison des deux génomes indique d'ailleurs qu'ils ont évolué séparément de façon réductive.

L'origine α -protéobactérienne des mitochondries est bien établie, mais celle de la cellule-hôte est moins claire. Suite à l'hypothèse endosymbiotique de Margulis en 1970, on s'est intéressé aux eucaryotes ne possédant pas de mitochondrie. Si les mitochondries, comme les chloroplastes, prenaient origine d'un événement d'endosymbiose, il était possible de trouver des eucaryotes sans mitochondrie ayant divergé avant cet événement. En 1983, Cavalier-Smith proposait la classe Archezoa, contenant des eucaryotes chez qui on n'avait pas trouvé de mitochondrie et pouvant avoir divergé avant l'événement d'endosymbiose ayant donné naissance aux mitochondries (Cavalier-Smith, 1983). Suite à cette proposition, des analyses phylogénétiques ont apporté des évidences de l'origine pré-mitochondriale des Archezoa (Brown et Doolittle, 1995; Vossbrinck *et al.*, 1987; Sogin, 1989; Viscogliosi *et al.*, 1993; Yamamoto *et al.*, 1997). L'idée que la cellule hôte ayant engouffré une α -protéobactérie pour donner naissance aux mitochondries soit un eucaryote primitif possédant un noyau et capable de phagocytose a par la suite été proposée (Gray *et al.*, 1999). Des observations récentes vont à l'encontre de cette hypothèse. Des vestiges de mitochondries (gènes codant pour des protéines mitochondriales, hydrogénosome, etc.) ont été trouvés chez les Archezoa. De plus, de nouvelles analyses phylogénétiques supportent le fait que les Archezoa aient divergé suite à l'apparition des mitochondries (Gray *et al.*, 1999; Roger, 1999; Keeling, 1998). Si, en réalité, il n'existe pas d'eucaryote ne possédant pas ou n'ayant jamais possédé de mitochondries, cela implique que l'origine des mitochondries soit très près (si elle ne coïncide pas) de l'origine des eucaryotes. Une chimère procaryotique pourrait être l'ancêtre des mitochondries et des eucaryotes primitifs (Karlin *et al.*, 1999). L'hypothèse de l'hydrogène (Martin et Müller, 1998), selon laquelle il y a eu une association entre une archéobactérie anaérobie strictement dépendante de l'hydrogène et autotrophique et une eubactérie générant de l'hydrogène (Lang *et al.*, 1999), supporte cette idée sans toutefois éliminer la traditionnelle hypothèse endosymbiotique. Elle présente la possibilité que le noyau des eucaryotes et la mitochondrie dérivent du même événement d'endosymbiose (Lang *et al.*, 1999).

1.3.2 Le génome mitochondrial des plantes vertes

1.3.2.1 Caractéristiques générales

La taille, la séquence, le contenu en gènes et l'organisation des génomes mitochondriaux sont extrêmement variables entre les diverses espèces. L'ADNmt des plantes et des algues est celui qui a été le moins étudié. L'analyse de la séquence complète du génome mitochondrial de quelques plantes vertes a permis d'en ressortir les principales caractéristiques qui sont résumées, avec celles du protiste *Reclinomonas americana*, dans le tableau IV.

Les génomes mitochondriaux des plantes terrestres ont une très grande taille mais leur contenu en gènes est très peu différent de celui des génomes beaucoup plus petits des protistes (incluant les algues) (Lang *et al.*, 1999, Leblanc *et al.*, 1997, Gray *et al.*, 1998). Environ 10-20 % de la séquence, chez les plantes terrestres, est constituée de gènes ayant des fonctions connues. Les séquences intergéniques, regroupant des introns, des ORFs introniques, des ORFs de fonction inconnue, des pseudogènes et des bouts d'ADN d'origine nucléaire ou chloroplastique, sont responsables de l'expansion de la taille de ces génomes (Lang *et al.*, 1999). Même si l'ADNmt des plantes terrestres est sujet à un réarrangement génomique accéléré, il a conservé des caractéristiques ancestrales (des vestiges d'ARNr bactériens, un code génétique standard et des regroupements de gènes typiques aux bactéries) qui prouvent son origine bactérienne (Lang *et al.*, 1999). Contrairement aux autres eucaryotes, les séquences des gènes mitochondriaux des plantes terrestres ont un taux de mutation très bas (Leblanc *et al.*, 1997).

Dans la lignée Chlorophyta des algues vertes, la situation est différente; l'ADNmt est compacte et riche en gènes. Leur contenu total en bases A + T est élevé (> 65 %) et est normalement plus élevé dans les régions non-codantes que dans les régions codantes. La plupart de ces génomes sont circulaires, comme les génomes bactériens mais certains génomes mitochondriaux (dont celui de *Chlamydomonas reinhardtii*) prennent aussi une conformation linéaire. Il a été observé que chez les génomes de *Chlamydomonas eugametos*, *Pedinomonas minor* et *Pycnococcus prasovali*, tous les gènes sont transcrits sur le même brin, suggérant une

Tableau IV. Caractéristiques générales des génomes mitochondriaux des algues vertes et d'un protiste dont la séquence entière est disponible.

Traits	Algues vertes ^a													
	Jakobid ^a					Génomes ancestraux						Génomes dérivés-réduits		
	<i>R.am</i>	<i>M.vi</i>	<i>N.ol</i>	<i>P.wi</i>	<i>P.pr</i>	<i>S.ob</i>	<i>P.mi</i>	<i>C.re</i>	<i>C.eu</i>	<i>C.el</i>				
Longueur (pb) ^b	69 034	42 424	45 223	55 328	24 321	42 919	25 137	15 758	22 897	22 704				
Forme ^c	C	C	C	C	C	C	C	L	C	C				
Nombre de gènes	94	65	65	61	35	42	21	13	12	12				
Introns														
groupe I	0	4	4	5	0	2	0	0	9	6				
groupe II	1	3	0	0	0	2	1	0	0	0				
% codant ^c	91,3	N/D ^d	78,4	70,6	N/D ^d	60,6	60,9	83,1	84,6	89,1				
% A + T														
total	73,9	N/D ^d	67,2	74,2	N/D ^d	63,7	77,8	54,8	65,4	62,2				
codant ^c	N/D ^d	N/D ^d	65,7	69,9	N/D ^e	60,6	76,3	54,9	65,7	63,1				
non-codant	N/D ^d	N/D ^d	72,7	84,6	N/D ^d	69,3	80,2	54,5	63,6	54,2				

^a *R.am*, *Reclinomonas americana*; *M.vi*, *Mesostigma viride*; *N.ol*, *Nephroselmis olivaceus*; *P.wi*, *Prototheca wickerhamii*; *P.pr*, *Pycnococcus prasovali*; *S.ob*, *Scenedesmus obliquus*; *P.mi*, *Pedinomonas minor*; *C.re*, *Chlamydomonas reinhardtii*; *C.eu*, *Chlamydomonas eugametos*; *C.el*, *Chlorogonium elongatum*.

^b pb, paires de bases.

^c C, circulaire; L, linéaire.

^d N/D signifie que la donnée n'était pas disponible.

^e Représente le pourcentage du génome qui code pour les gènes de protéines, d'ARNt, d'ARNr, d'autres ARNs ainsi que pour les introns et les ORFs introniques.

seule unité de transcription. Sur ce plan, les génomes mitochondriaux de *Mesostigma viride* et *Prototheca wickerhamii* semblent être divisé en deux, où dans une première moitié du génome, tous les gènes sont transcrits sur un brin et dans la deuxième moitié, tous les gènes sont transcrits sur l'autre brin. Cependant, dans le génome de *M. viride*, chaque bloc de gènes est fragmenté par quelques gènes d'ARNt situés sur le brin complémentaire.

Deux types de génomes mitochondriaux ont été décrits parmi les lignées d'algues vertes étudiées jusqu'à maintenant, les génomes dits dérivés-réduits et les génomes dits ancestraux (tableau IV). Les génomes mitochondriaux des chlamydomonads (*Chlamydomonas reinhardtii*, *Chlamydomonas eugametos* et *Chlorogonium elongatum*) et de *Pedinomonas minor* ont une organisation de type dérivé-réduit. Ces génomes ont une taille réduite (16-25 kb), un nombre de gènes limité et une vitesse accélérée d'évolution de la séquence. Les gènes des protéines ribosomiques et de la sous-unité 5S de l'ARNr sont absents de la plupart de ces génomes qui ne contiennent que quelques gènes codant pour des protéines de la chaîne respiratoire et des ARNt. Les gènes des ARNr sont divisés et dispersés dans tous ces génomes. Une variation dans le code génétique standard a été observée dans le génome de *P. minor*. Les génomes dérivés-réduits ont aussi tendance à avoir une utilisation extrêmement biaisée des codons dans les gènes de protéines menant, dans quelques cas, à l'élimination de certains codons (Gray *et al.*, 1999).

Les génomes mitochondriaux de *Prototheca wickerhamii*, *Nephroselmis olivacea* et *Mesostigma viride* représentent le type ancestral. Leur taille est plus grande (45-55 kb) et ils ont retenus plusieurs caractéristiques de l'ancêtre procaryotique, dont la présence du gène de la sous-unité 5S de l'ARNr, un plus grand nombre de gènes que les génomes de type dérivé-réduit (incluant les gènes codant pour les protéines ribosomiques), un éventail complet (ou presque) d'ARNt, peu d'introns et des regroupements de gènes similaires à ceux des eubactéries. Le contenu en gènes de l'ADNmt des ces algues vertes est similaire à celui des plantes terrestres mais l'organisation de l'information génétique de ces génomes est très compacte. Les ARNr de ces génomes, contrairement à ceux des génomes dérivés-réduits, sont entiers. Le code génétique universel est utilisé chez ces ADNmt (Gray *et al.*, 1999). Le génome mitochondrial de *Pycnococcus prasovali* a aussi été classé ancestral puisqu'il contient

quelques gènes de protéines ribosomiques et que ses gènes d'ARNr ne sont pas fragmentés. Toutefois, ce génome est très réarrangé et sa taille, de même que son contenu en gènes, sont beaucoup plus petits que ceux des autres génomes mitochondriaux ancestraux (tableau IV).

L'organisation du génome mitochondrial de l'algue verte *Scenedesmus obliquus* (classe Chlorophyceae) suggère que ce génome soit à un stade intermédiaire entre les deux types de génomes mitochondriaux décrits. Il a une taille (42 919 kb) se rapprochant du type ancestral mais il ne contient pas de gènes de protéines ribosomiques et d'ARNr 5S. Ces gènes d'ARNr sont fragmentés et dispersés dans tout le génome, des caractéristiques typiques aux génomes dérivés-réduits (Nedelcu *et al.*, 2000). Tout comme *Pedinomonas minor*, il n'utilise pas le code génétique standard; il reconnaît le codon de terminaison UAG comme une leucine. L'ADNmt de *S. obliquus* est le seul à reconnaître le codon UCA comme un codon de terminaison. La caractéristique qui permet le plus d'assigner à *S. obliquus* le titre intermédiaire évolutif est la fragmentation de ses gènes d'ARNr, qui suggère une évolution graduelle entre les gènes d'ARNr entiers des génomes de type ancestral et ceux grandement fragmentés des génomes de type dérivé-réduit (Nedelcu *et al.*, 2000).

Le plus grand répertoire de gènes mitochondriaux a été identifié chez le protiste classé avec les Jakobids, *Reclinomonas americana* (Lang *et al.*, 1997). Le génome mitochondrial de ce protiste s'assemble en un cercle unique d'une taille de 69 034 pb, est riche en bases A + T et contient une forte densité de gènes distribués sur les deux brins. Un seul intron, un intron du groupe II dans le gène *trnW*, a été trouvé dans ce génome. Il contient aussi un nombre d'ARNt suffisant pour lire tous les codons du code génétique standard utilisés par *R. americana* sauf les codons de la thréonine. L'ADNmt de *R. americana* est celui dont le contenu en gènes et l'organisation du génome sont les plus similaires aux génomes bactériens. Ces caractéristiques indiquent que le génome de *Reclinomonas americana* est celui qui ressemble le plus, parmi les ADNmt étudiés jusqu'à maintenant, à l'ancêtre de tous les génomes mitochondriaux (Lang *et al.*, 1997).

1.3.2.2 Les régions répétées

Les génomes mitochondriaux avec un faible pourcentage en régions codantes contiennent souvent un grand nombre de séquences répétées de taille et de contenu variables (Paquin *et al.*, 2000). La plupart de ces séquences sont situées dans les régions intergéniques mais il y en a aussi qui ont été localisées dans des introns, des régions variables de gènes codant pour des ARNr et, exceptionnellement, dans des gènes codant pour des protéines. Ces séquences répétées prennent principalement la forme de répétitions en tandem (de type macro ou minisatellite), de répétitions dispersées de longueurs variables (formant parfois des régions d'ADN inversé répété) ou de courts éléments se repliant en une structure secondaire conservée (Paquin *et al.*, 2000). Plusieurs exemples de répétitions ont été répertoriés chez les plantes vertes. L'ADNmt de l'algue verte *Pedinomonas minor* contient une région de séquences répétées de 9 kb. Cette région, qui occupe plus du tiers du génome, est composée de 13 classes de répétitions dispersées de 6 à 389 nucléotides qui forment une superstructure élaborée (Turmel *et al.*, 1999a). Le chlorophyte *Prototheca wickerhamii* contient aussi des familles de séquences répétées constituées de 9 éléments de 30 à 200 nucléotides mais dispersés dans tout le génome dans ce cas (Wolff *et al.*, 1994). Chez *Chlamydomonas reinhardtii*, 11 structures répétées formant des palindromes dont la séquence primaire est conservée ont été identifiées. La distribution de ces séquences dans le génome suggère qu'elles aient un rôle dans l'expression des gènes (Boer et Gray, 1991). Dans l'ADNmt des plantes terrestres, deux types de structures répétées ont été répertoriés : des séquences répétées mobiles et ayant une séquence primaire très conservée formant des palindromes (PRS) ainsi que des structures répétées moins conservées, se repliant en structure en épingle à cheveux et qui pourraient être impliquées dans l'épissage d'ARN (Paquin *et al.*, 2000). Chez des levures de l'espèce *Saccharomyces*, les éléments répétés sont des séquences riches en G + C dispersées dans tout le génome. Ces éléments semblent être mobiles car ils ne sont pas présents chez tous les membres de l'espèce. En se basant sur la similarité de leur séquence primaire, ces séquences répétées peuvent être classées en familles dont plusieurs peuvent se replier en une variété de structures en épingle à cheveux.

La fonction des séquences répétées n'est pas connue. Elles pourraient servir d'éléments de

contrôle impliqués dans la régulation de la transcription, de la traduction ou de la réplication et être conservées durant l'évolution. Elles pourraient aussi simplement être le résultat d'erreurs lors la réplication de l'ADN mitochondrial et dans les mécanismes de réparation de l'ADN ou de la transposition d'éléments mobiles. Dans tous les cas, l'accumulation d'éléments répétés semble favoriser les événements de recombinaison responsables de la réorganisation des génomes mitochondriaux (Paquin *et al.*, 2000).

1.3.2.3 Les gènes mitochondriaux

Plusieurs gènes du génome pré-mitochondrial ont été perdus lors de l'établissement et de la diversification primaire du génome mitochondrial. Puisque le génome pré-mitochondrial devait contenir environ 1000 gènes, une perte massive de gènes a dû avoir lieu pour donner naissance aux génomes mitochondriaux que l'on connaît (Lang *et al.*, 1999). Les mécanismes de perte de gènes de l'ADNmt semblent être les mêmes que ceux qui ont été décrits pour l'ADNcp (section 1.2.2.3, Les gènes chloroplastiques). Les gènes retrouvés dans le génome mitochondrial des plantes vertes sont surtout impliqués dans la respiration mais des gènes responsables de l'expression génétique sont aussi présents. Le tableau V présente tous les gènes mitochondriaux présents chez les plantes vertes dont la séquence entière est disponible.

Les gènes impliqués dans la respiration

La majorité des gènes mitochondriaux codent pour des composantes de la chaîne respiratoire et de la phosphorylation oxydative. Le complexe respiratoire I est formé des sous-unités de la NADH: ubiquinone oxydoreductase. Les gènes mitochondriaux *nad* codent pour ce complexe. Ces gènes semblent être les plus souvent retrouvés dans l'ADNmt. Les gènes *sdh* codent pour des sous-unités de la succinate: cytochrome c oxydoreductase (complexe II), le gène *cob* pour l'ubiquinone: cytochrome c oxydoreductase (complexe III), les gènes *cox* pour des sous-unités de la cytochrome c oxydase (complexe IV) et les gènes *atp* pour des sous-unités de l'ATP synthétase (complexe V). Les gènes *ccb*, retrouvés seulement chez les plantes terrestres, sont impliqués dans la biogénèse du cytochrome c.

Suite Tableau V

Gènes	Algues vertes ^a									Plantes terrestres ^a	
	<i>M.vi</i>	<i>N.ol</i>	<i>P.wi</i>	<i>P.pr</i>	<i>S.ob</i>	<i>P.mi</i>	<i>C.re</i>	<i>C.eu</i>	<i>C.el</i>	<i>M.po</i>	<i>A.th</i>
Protéines ribosomiques											
<i>rpl2</i>										•	•
<i>rpl5</i>	•	•	•							•	•
<i>rpl6</i>	•	•	•							•	
<i>rpl14</i>	•	•									
<i>rpl16</i>	•	•	•							•	•
<i>rps1</i>	•									•	
<i>rps2</i>	•	•	•							•	
<i>rps3</i>	•	•	•	•						•	•
<i>rps4</i>	•	•	•	•						•	•
<i>rps7</i>	•	•	•							•	•
<i>rps8</i>		•								•	
<i>rps10</i>	•	•	•							•	
<i>rps11</i>	•	•	•							•	
<i>rps12</i>	•	•	•	•						•	•
<i>rps13</i>	•	•	•							•	
<i>rps14</i>	•	•	•							•	
<i>rps19</i>	•	•	•							•	
ARNs ribosomiques											
<i>rrn5</i>	•	•	•							•	•
<i>rnl</i>	•	•	•	•	•	•	•	•	•	•	•
<i>rns</i>	•	•	•	•	•	•	•	•	•	•	•
ARNs de transfert^c											
<i>A(ugc)</i>	•	•	•		•					•	
<i>C(gca)</i>	•	•	•	•	•	•				•	•
<i>D(guc)</i>	•	•	•	•	•					•	•
<i>E(uuc)</i>	•	•	•	•	•	•				•	•
<i>F(gaa)</i>	•	•	•		•	•				•	•
<i>G(gcc)</i>	•		•							•	•
<i>G(ucc)</i>	•	•	•		•					•	
<i>H(gug)</i>	•	•	•	•	•	•				•	•
<i>I(cau)</i>		•	•							•	•
<i>I(gau)</i>	•	•	•	•	•						
<i>I(uau)</i>					•						
<i>K(uuu)</i>	•	•	•		•					•	•
<i>L(aag)</i>					•						

Suite Tableau V

Gènes	Algues vertes ^a									Plantes terrestres ^a	
	<i>M.vi</i>	<i>N.ol</i>	<i>P.wi</i>	<i>P.pr</i>	<i>S.ob</i>	<i>P.mi</i>	<i>C.re</i>	<i>C.eu</i>	<i>C.el</i>	<i>M.po</i>	<i>A.th</i>
<i>L(caa)</i>	•				•	•				•	
<i>L(cua)</i>					•						
<i>L(cag)</i>					•						
<i>L(uaa)</i>	•	•	•							•	
<i>L(uag)</i>	•	•	•	•						•	
<i>Me(cau)</i>	•	•	•	•	•		•	•	•	•	•
<i>Mf(cau)</i>	•	•	•	•	•		•	•	•	•	•
<i>N(guu)</i>	•	•	•	•	•					•	•
<i>P(ugg)</i>	•	•	•	•	•					•	•
<i>Q(uug)</i>	•	•	•	•	•	•	•	•	•	•	•
<i>R(acg)</i>	•	•	•		•					•	
<i>R(ccu)</i>					•						
<i>R(ucg)</i>	•	•								•	
<i>R(ucu)</i>	•	•	•		•					•	
<i>S(gcu)</i>	•	•	•	•	•					•	•
<i>S(gga)</i>					•						•
<i>S(uga)</i>	•	•	•	•						•	•
<i>T(ggu)</i>		•								•	
<i>T(ugu)</i>			•								
<i>V(uac)</i>	•	•	•	•	•					•	
<i>W(cca)</i>	•	•	•	•	•		•	•	•	•	•
<i>W(uca)</i>						•					
<i>Y(gua)</i>	•	•	•	•	•	•				•	•
Autre ARN											
<i>rnpB</i>		•									
ORFs conservés											
<i>ymf16</i>	•	•	•							•	•
<i>ymf39</i>	•	•	•							•	•

^a *M.vi*, *Mesostigma viride*; *N.ol*, *Nephroselmis olivacea*; *P.wi*, *Prototheca wickerhamii*; *P.pr*, *Pycnococcus prasovalis*; *S.ob*, *Scenedesmus obliquus*; *P.mi*, *Pedinomonas minor*; *C.re*, *Chlamydomonas reinhardtii*; *C.eu*, *Chlamydomonas eugametos*; *C.el*, *Chlorogonium elongatum*; *M.po*, *Marchantia polymorpha*; *A.th*, *Arabidopsis thaliana*.

^b Les cercles noirs indiquent la présence du gène chez cet organisme.

^c Les ARNt sont indiqués par le code à une lettre des acides aminés suivi de l'anticodon entre parenthèses.

Les gènes impliqués dans l'expression génétique

L'ADNmt des plantes terrestres ainsi que celui des algues vertes ayant un génome mitochondrial dit ancestral contiennent jusqu'à cinq gènes codant pour des protéines de la grande sous-unité du ribosome (gènes *rpl*) et jusqu'à 12 gènes codant pour des protéines de la petite-sous-unité du ribosome (gènes *rps*). Trois gènes codant pour des ARNs ribosomiques sont présents dans l'ADNmt des plantes vertes : le gène *rrn5* codant pour l'ARNr 5S, le gène *rns* codant pour l'ARNr 16S et le gène *rnl* codant pour l'ARNr 23S. Chez les plantes vertes, entre huit et 26 gènes *trn* codent pour les ARNs de transfert présents dans le génome. Ils sont souvent insuffisants pour lire tous les codons utilisés dans le génome. L'importation d'ARNt du cytosol est le mécanisme le plus souvent invoqué pour combler ce déficit (Gray *et al.*, 1998). L'édition partielle de la séquence de l'anticodon de l'ARNt est un autre mécanisme permettant de pallier à un manque d'ARNt (Gray *et al.*, 1998). Il est aussi à noter que certains ARNt mitochondriaux présentent des déviations face à la structure standard des ARNt (figure 5).

Autres gènes

Le gène *rnpB*, codant pour l'ARN faisant partie du complexe RNaseP, a été répertorié chez quelques algues vertes dont le génome mitochondrial est dit ancestral.

ORFs conservés

Les gènes *ymf16* et *ymf39* sont des gènes dont la fonction n'est pas encore connue mais qui sont conservés dans l'ADNmt de plusieurs organismes. Ces gènes ne sont pas présents chez les algues vertes ayant des génomes mitochondriaux de type dérivé-réduit.

1.3.2.4 L'organisation des gènes mitochondriaux

Baucoup de déviations, souvent spécifiques à une lignée, distinguent l'organisation des gènes entre les génomes mitochondriaux et celui des eubactéries. La structure des opérons

eubactériens a été complètement ou partiellement désintégrée dans la plupart des génomes mitochondriaux (Palmer et Delwiche, 1998). Cependant, chez les algues vertes ancestrales et au moins une plante terrestre (*Marchantia polymorpha*), l'organisation des gènes des protéines ribosomiques a été conservée; elle suit celle des opérons L2-*spc*-alpha-S9 de *E. coli* (Lang *et al.*, 1999). Comme les génomes mitochondriaux des plantes terrestres ainsi que ceux de type dérivé-réduit des algues vertes sont très réarrangés, il y a très peu de regroupements de gènes communs entre les plantes vertes. Seul les génomes ancestraux d'algues vertes ont des regroupements conservés entre eux.

1.3.2.5 Les introns mitochondriaux

Les génomes mitochondriaux des algues vertes contiennent peu d'intron et la majorité de ces introns sont du groupe I. Chez les algues vertes du tableau IV, il y a environ quatre fois plus d'introns du groupe I que du groupe II. À l'opposé, les plantes terrestres contiennent un plus grand nombre d'introns dont très peu font partie du groupe I (Gray *et al.*, 1998; Unseld *et al.*, 1999). Des évidences suggèrent que plusieurs de ces introns aient été acquis par transfert horizontal. Cependant, un événement de transfert vertical semble avoir donné naissance aux introns du gène *cox1* dans l'ADNmt de champignons, de *Prototheca wickerhamii* et de *Marchantia polymorpha* (Gray *et al.*, 1998). Des événements relativement récents de transfert horizontal entre des introns du groupe I insérés à des positions correspondant aux introns des positions 1931 et 2593 dans la séquence de *E. coli* du gène codant pour la grande sous-unité de l'ARNr du génome chloroplastique et du génome mitochondrial de deux organismes distants ont aussi été répertoriés (Turmel *et al.*, 1995).

1.3.3 Les analyses phylogénétiques à l'aide de séquences de gènes mitochondriaux concaténés

Dans les analyses phylogénétiques utilisant les séquences mitochondriales de l'ADNr, les deux types de génomes mitochondriaux d'algue verte ne sont pas reliés entre eux (Denovan-Wright *et al.*, 1996). Sensiblement la même chose est observée dans les arbres phylogénétiques construits avec des séquences concaténées de protéines; les algues vertes

ayant des génomes mitochondriaux ancestraux se lient avec les plantes terrestres mais les algues vertes avec des génomes dérivés-réduits forment un groupe séparé des autres plantes vertes (Turmel *et al.*, 1999b; Denovan-Wright *et al.*, 1996). Un taux d'évolution accéléré des séquences de gènes caractérise les génomes mitochondriaux dérivés-réduits d'algues vertes. Dans les arbres phylogénétiques, la longueur des branches est proportionnelle au taux d'évolution. Comme la vitesse d'évolution de la séquence protéique des génomes mitochondriaux dérivés-réduits est beaucoup plus grande que celle de la séquence des autres génomes mitochondriaux étudiés, la longueur de la branche de ces taxa est significativement plus longue que celle des autres taxa présents dans les arbres phylogénétiques. Les longues branches ont tendance à se regrouper, ce qui est appelé le phénomène d'attraction des longues branches (Graur et Li, 2000). C'est ce qui explique probablement que ce groupe ne forme pas une lignée monophylétique avec les autres plantes vertes dans les analyses phylogénétiques utilisant des séquences mitochondriales (Turmel *et al.*, 1999b). L'analyse phylogénétique du génome mitochondrial de *Scenedesmus obliquus* supporte son statut d'intermédiaire évolutif (Denovan-Wright *et al.*, 1996).

Autant dans les arbres construits avec plusieurs gènes de protéines mitochondriales que dans ceux construits avec le gène *rri* chloroplastique, le prasinophyte *Pedinomonas minor* est lié avec des algues vertes de la classe Chlorophyceae plutôt qu'avec les prasinophytes *Nephroselmis olivacea* et *Tetraselmis subcordiiformis* ou le trebouxiophyte *Prototheca wickerhamii* (Turmel *et al.*, 1999b). Les résultats obtenus avec la séquence protéique concaténée de plusieurs gènes chloroplastiques sont différents, *P. minor* se groupe avec les trebouxiophytes (Lemieux *et al.*, 2000). A savoir si *P. minor* est un ancêtre flagellé de la classe Chlorophyceae, s'il fait partie de la classe Trebouxiophyceae ou s'il est un prasinophyte, comme les analyses morphologiques le laissent croire, reste des questions à répondre.

Il est à noter qu'aucune phylogénie faite à l'aide d'une séquence protéique concaténée de plusieurs gènes mitochondriaux et incluant les séquences de *Mesostigma viride* et *Pycnococcus prasoali* ne sont publiées mais que celle de *Nephroselmis olivacea* (Turmel *et al.*, 1999b) présente la séquence de l'ADNmt de cette algue comme étant la plus ancestrale

parmi celles des algues vertes de la lignée Chlorophyta étudiées.

Une analyse phylogénétique construite à l'aide de la séquence protéique concaténée des gènes mitochondriaux *cob*, *cox1*, *cox2* et *cox3* de plusieurs lignées a permis de déduire, avec un niveau de confiance élevé, que les algues rouges et les algues vertes étaient des lignés sœurs et que le protiste *Reclinomonas americana* faisait partie d'une lignée à la base des plantes vertes et des algues rouges (Burger *et al.*, 1999).

1.4 Problématique de l'étude

Le récent séquençage des génomes chloroplastiques et mitochondriaux des algues vertes *Mesostigma viride* et *Nephroselmis olivacea* ont apporté beaucoup d'informations sur l'évolution et l'architecture des génomes ancestraux des organelles de l'ancêtre des plantes vertes (Lemieux *et al.*, 2000; Turmel *et al.*, 1999a). Le présent projet vise à apporter de nouvelles connaissances sur la classe d'algues vertes Prasinophyceae (la classe la plus ancestrale de la lignée Chlorophyta) et par conséquent sur l'ancêtre de cette lignée.

Le prasinophyte qui a fait l'objet de cette étude est *Monomastix* OKE-1. Il a été nommé OKE-1 car l'espèce de cette souche n'est pas précisée et qu'elle provient de la collection privée de Charles O'Kelly. L'algue verte *Monomastix* a d'abord été considérée car elle présente des caractéristiques morphologiques ancestrales aux autres plantes vertes. Cette algue est uniflagellée et possède des écailles de structure beaucoup plus simple que chez les autres prasinophytes (communication directe de Charley O'Kelly, 1994). La disponibilité de ce prasinophyte et la possibilité d'isoler l'ADN de ses organelles avec un bon rendement ont été deux facteurs déterminants dans le choix de cette algue. Comme les analyses faites avec le gène codant pour l'ARNr 23S chloroplastique de *Monomastix* plaçaient cette algue dans un groupe plus basal que *Nephroselmis olivacea* (groupe II des prasinophytes de la figure 2), *Monomastix* semblait être un bon choix pour l'étude des caractéristiques de l'ADN des organelles des algues vertes se rapprochant de l'ancêtre de la lignée Chlorophyta (résultats non-publiés du laboratoire). Le séquençage de ces génomes allait permettre d'étudier leur architecture et leur organisation ainsi que d'identifier les forces qui les ont fait évoluer par

rapport aux autres génomes d'algues vertes disponibles.

La classification des algues vertes, et en particulier celle des prasinophytes, est encore ambiguë. L'ajout de la séquence entière du génome chloroplastique et du génome mitochondrial de *Monomastix* aux quelques génomes d'organelles d'algues vertes déjà séquencés visait aussi à préciser la position phylogénétique de *Monomastix* ainsi que celle des algues vertes basales et à apporter de nouvelles connaissances sur leur évolution. Globalement, ce travail allait permettre de vérifier si la structure des génomes des organelles de *Monomastix* se rapproche de celle de l'ancêtre de la lignée Chlorophyta.

MATÉRIEL ET MÉTHODES

2.1 Croissance cellulaire de l'algue *Monomastix* OKE-1

La culture originale de l'algue verte *Monomastix* OKE-1 provient d'un don du Dr Charles O'Kelly. Cette souche a été prélevée en Nouvelle Zélande et est non axénique. A partir d'un inoculum de 1 ml de la culture originale, une pré-culture de 25 ml dans du milieu Volvox (0,94 g de tampon HEPES, 1 ml d'une solution de $\text{Ca}(\text{NO}_3)_2 \cdot 7 \text{H}_2\text{O}$ 0,50M, 1 ml d'une solution de $\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$ 0,16 M, 1 ml d'une solution de Na_2 Glycérophosphate $\cdot 5 \text{H}_2\text{O}$ 0,20 M, 1 ml d'une solution de KCl 0,67 M, 1 ml de biotine 1×10^{-6} M, 1 ml de vitamine B_{12} 1×10^{-7} M et 6 ml d'une solution PIV de 1 litre contenant 750 mg de Na_2 EDTA, 97 mg de $\text{FeCl}_3 \cdot 6 \text{H}_2\text{O}$, 41 mg de $\text{MnCl}_2 \cdot 4 \text{H}_2\text{O}$, 5 mg de ZnCl_2 , 2 mg de $\text{CoCl}_2 \cdot 6 \text{H}_2\text{O}$ et 4 mg de $\text{Na}_2\text{MoO}_4 \cdot 2 \text{H}_2\text{O}$ et 989 ml d'eau bidistillée pour un volume total de 1 litre) a été préparée. La croissance de la pré-culture a duré une semaine. Un inoculum de 5 ml par culture de cette pré-culture a permis de faire croître 16 cultures de 500 ml chacune dans du milieu Volvox. Des flacons plats en verre transparent ont été utilisés durant la période de croissance de 15 jours selon un cycle 12 h jour/12 h nuit.

2.2 Isolement d'une fraction d'ADN enrichie en ADN chloroplastique et mitochondrial

2.2.1 Lyse des cellules

Chaque culture d'algue a été centrifugée à 5000 rpm/4°C pendant 10 min dans une centrifugeuse RC5C Sorvall Instruments (Du Pont) avec un rotor GS-3. Le culot de cellules total a été conservé dans la glace et resuspendu dans un volume minimal (entre 0,5 et 1 ml) de tampon A (Tris-HCl 10 mM pH 8, EDTA 10 mM pH 8, NaCl 10 mM). La suspension cellulaire totale ainsi obtenue a été déposée dans un mortier préalablement refroidi à -70°C , de l'azote liquide y a été ajouté et les cellules ont été broyées jusqu'à l'obtention d'une fine poudre. Cette poudre a été ajoutée par petites quantités et en brassant continuellement mais

doucement à 25 ml d'une solution conservée à 50°C de tampon A contenant du SDS 1,25% et 2,5 g de protéinase K (100 mg/ml). Lorsque le mélange est devenu homogène et visqueux, il a été incubé à 50°C pendant 2 h. Le lysat a été conservé à 4°C.

2.2.2 Extraction de l'ADN cellulaire total

Le lysat de cellules a été placé au moins 2 h à la température de la pièce avant d'être transféré, avec un volume de phénol, dans un tube de 100 ml et d'être mélangé par rotation pendant 15 min. La préparation a ensuite été versée dans une bouteille GSA, centrifugée 15 min à 6000 rpm/température de la pièce dans une centrifugeuse RC5C Sorvall Instruments (Du Pont) avec un rotor GSA et la phase supérieure a été transférée dans un nouveau tube de 100 ml. Un volume de phénol/chloroforme : alcool isoamyl 24 :1 a été ajouté au tube avant de mélanger la solution 30 min par rotation à vitesse moyenne. Ce mélange a été centrifugé 15 min à 6000 rpm et la phase supérieure a été transférée dans un autre tube de 100 ml. De nouveau, un volume de phénol/chloroforme : alcool isoamyl 24 :1 a été ajouté et la préparation a été mélangée par rotation 15 min. Celle-ci a ensuite été centrifugée 15 min à 6000 rpm et la phase supérieure a été transférée dans une nouvelle bouteille GSA. Une fois l'ADN cellulaire total exempt de protéines et de débris cellulaires, des volumes de PEG 50% et de NaCl 5 M (concentrations finales de 10% et 1 M respectivement) ont été ajoutés afin de faire précipiter l'ADN. La solution a doucement été mélangée avant d'être placée à 4°C pendant 24 h. Suite à l'incubation, cette solution a été centrifugée 10 min à 7000 rpm/4°C dans la centrifugeuse Sorvall toujours avec le rotor GSA et le culot a de nouveau été centrifugé afin d'éliminer l'excédent de liquide. Le culot d'ADN a été resuspendu dans 6,75 ml d'eau stérile pendant 3 h avant l'ajout de 0,75 ml de TE 10X (Tris-HCl 100mM, EDTA 10 mM).

2.2.3 Séparation de l'ADN cellulaire total sur gradient de chlorure de césium et extraction de la fraction d'ADN enrichie en ADN chloroplastique et mitochondrial

L'ADN des organites a été séparé de celui du noyau par centrifugation isopycnique en présence de chlorure de césium et de fluorochrome bisbenzimidazole trihydrochlorure. Le CsCl crée un gradient de densité ce qui permet de séparer des molécules avec des densités

différentes. Afin de séparer les molécules d'ADN entre elles, le fluorochrome bisbenzimidazole trihydrochlorure, qui augmente la distance de séparation des bandes, a été utilisé. Ce composé se fixe aux régions riches en A + T, rend les bandes d'ADN fluorescentes et permet de les visualiser aux ultraviolets à 302 nm.

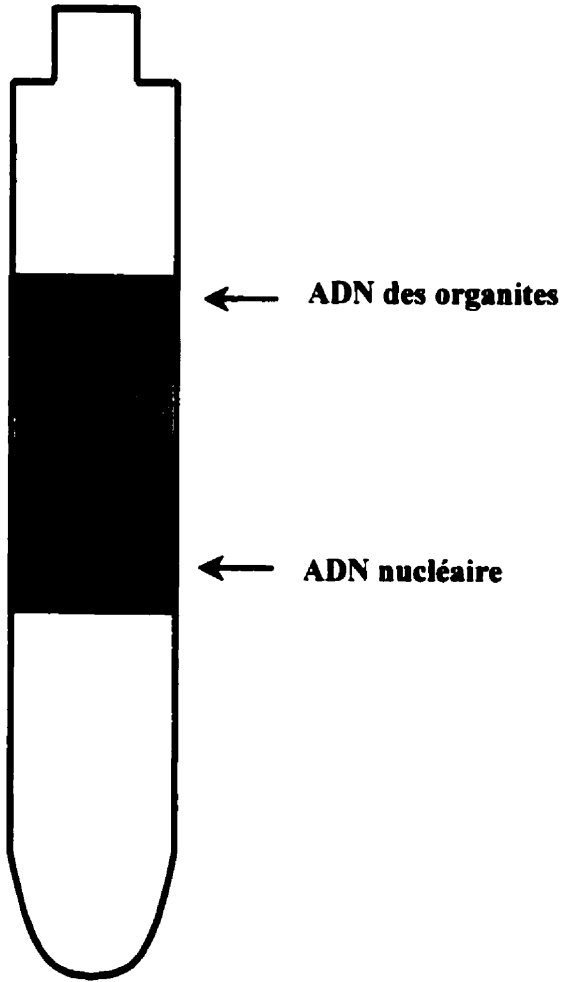
Dans un tube de type Falcon de 15 ml, 2,5 ml d'ADN, 1,2 ml de TE 1X (Tris-HCl 10 mM, EDTA 1 mM) ainsi que 4,565g de CsCl ont été placés afin d'obtenir un volume final de 4,9 ml et une densité de CsCl de 1,67g/ml. 0,01 mg de fluorochrome bisbenzimidazole trihydrochlorure ont été ajoutés au tube qui a par la suite été mélangé jusqu'à ce que le CsCl soit parfaitement dissout. Cette préparation a été centrifugée pendant 20 h à 50 000 rpm/20°C dans une ultracentrifugeuse L8-70M (Beckman) avec un rotor NVT 65,2. Les bandes séparées ont été visualisées dans une chambre noire avec un éclairage ultraviolet à 302 nm (figure 6). Une seule bande contenant autant l'ADN chloroplastique que l'ADN mitochondrial a été observée. Un trou a été percé à l'aide d'une aiguille #16 juste sous la bande d'ADN des organites et cet ADN a été recueilli avec une seringue de 3cc. Une nouvelle centrifugation de l'ADN recueilli a été faite dans un tube « Optiseal » selon les mêmes conditions que pour la première centrifugation isopycnique afin d'éliminer tout l'ADN nucléaire visible sous éclairage ultraviolet.

La solution ADN des organites-CsCl a ensuite été transférée dans un tube Pollyallomer de 12 ml et le volume a été ajusté à 1 ml avec une solution de CsCl d'une densité de 1,67 g/ml. La densité de la solution d'ADN-CsCl ainsi obtenue a été modifiée en ajoutant 1 ml de TE 1X (Tris-HCl 10 mM, EDTA 1 mM) afin de faire précipiter l'ADN. Le tube a été rempli avec une solution de NaCl 2,5 M avant d'être centrifugé pendant 20 h à 35 000 rpm/20°C dans une ultracentrifugeuse L8-70M (Beckman) avec un rotor « swinging bucket » TH-641. Le culot obtenu et débarrassé du surnageant a été resuspendu dans 100 µl d'eau stérile pendant 30 min. L'ADN a été conservé à -20°C.

2.3 Digestion *EcoRI* de l'ADN des organites

La digestion a eu lieu dans un volume de 10 µl incluant 1 µl d'ADN, 1 µl de tampon OPA

Figure 6. Profil d'ADN obtenu suite à l'ultracentrifugation de l'ADN cellulaire total de *Monomastix* OKE-1 sur un gradient de CsCl en présence de bisbenzimidazole. La bande du bas représente l'ADN nucléaire et celle du haut, l'ADN des organites. Comme l'ADN mitochondrial et l'ADN chloroplastique ne sont pas bien séparés, une seule bande est observée pour l'ADN des organites. La région en gris représente une région moins claire entre l'ADN des organites et l'ADN nucléaire. Des polysaccharides sont peut-être responsables de cette région trouble.



10X (Tris-acétate 100 mM pH 7,5, magnésium-acétate 100 mM, potassium-acétate 500 mM) (Amersham Pharmacia Biotech Inc.), 1 µl d'enzyme *EcoRI* (Amersham Pharmacia Biotech Inc.) et 7 µl d'eau stérilisée. Le mélange réactionnel a été incubé pendant 2 h à 37°C Une aliquote de 25 µl de « Stop-mix » 5X (EDTA 100mM, SDS 0,5%, glycérol 25%, bleu de bromophénol 0,1%) a été ajouté et le tout a été visualisé sur gel d'agarose 0,8% (figure 7). Le tampon de migration utilisé est du tampon TAE 1X (242 g de tris base 1M, 57,1 ml d'acide acétique glacial, EDTA 50 mM et complété à 1 litre avec de l'eau distillée pour avoir un tampon 50X) contenant 250 µg de bromure d'éthidium. Cette étape a permis d'évaluer la taille du génome des organites et de doser l'ADN.

2.4 Construction des banques de clones

2.4.1 Préparation des fragments d'ADN chloroplastique et mitochondrial

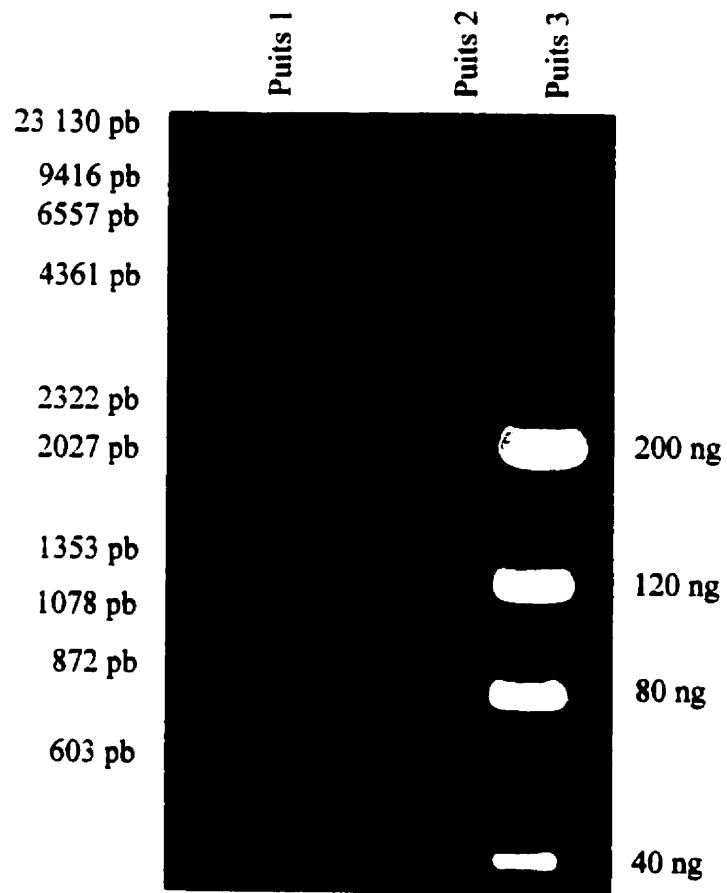
2.4.1.1 Nébulisation de l'ADN

Une aliquote de 1 µg d'ADN a été coupée de façon aléatoire dans un nébulisateur (ou « Up Mist Medication Nebulizer ») (Hospitak Inc.) contenant 12,5 µl d'ADN (80 ng/µl), 287,5 µl de TE 1X (Tris-HCl 1 mM, EDTA 0,1 mM) et 300 µl de glycérol 50% stérile. La nébulisation a duré 110 secondes sous jet d'azote à 5 PSI. Afin de récupérer l'ADN, le nébulisateur a été centrifugé 15 sec à 300g/température de la pièce dans la centrifugeuse EPPENDORF 5810R. L'ADN recueilli a été transféré dans deux tubes de 1,5 ml avant de le faire précipiter à -20°C pendant 12 h en présence de 0,5 volumes de NH₄(OAc) 7,5 M, 2 volumes d'éthanol 95% et 20 µg de glycogène. L'ADN précipité a été centrifugé à 12 000 rpm/4°C dans une centrifugeuse RC5C Sorvall Instruments (Du Pont) avec un rotor SA-600 pendant 50 min, lavé avec 500 µl d'éthanol 80% et centrifugé de nouveau 10 min à 12 000 rpm/4°C afin d'éliminer tous les sels.

2.4.1.2 Réparation des fragments d'ADN

Pour des fins de clonage, les extrémités protubérantes 3' et 5' des fragments d'ADN nébulisés

Figure 7. Digestion *EcoRI* de l'ADN préparé par ultracentrifugation sur gradient de CsCl. Puits 1 : standard de poids moléculaire λ ADN/*HindIII*, ϕ X174 RF ADN/*HaeIII*, Puits 2 : ADN des génomes des organites de *Monomastix* digéré avec *EcoRI*, Puits 3 : marqueur de poids moléculaire.



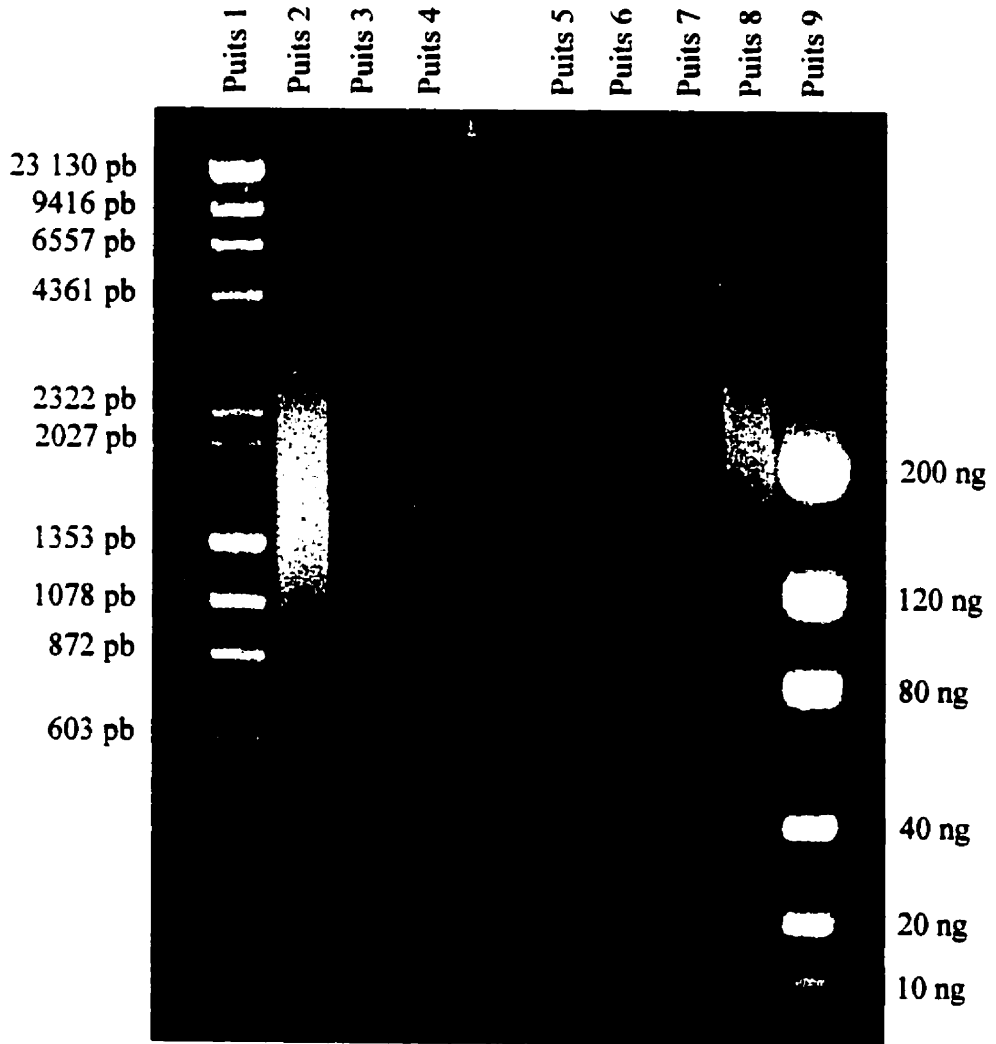
ont été converties en bouts francs. Pour ce faire, l'ADN a d'abord été resuspendu dans 18,5 µl d'eau bidistillée « molecular biology GRADE » (5 prime → 3 prime). Ensuite, cet ADN a été réparé en présence d'un mélange de dNTP 2,5 mM (dATP 10 mM, dCTP 10 mM, dGTP 10 mM, dTTP 10 mM), de 2,5 µl de tampon Klenow 10X (Tris-HCl 0,5 mM pH 7,6, MgCl₂ 0,1 mM, DTT 0,01 mM et 0,39 ml d'eau bidistillée stérile « molecular biology GRADE » (5 prime → 3 prime) pour un volume final de 1 ml), de 0,5 µl du fragment Klenow de l'ADN polymerase I 10U/µl (Fermentas) et de 1 µl d'ADN polymérase T7 5 U/µl (USB). Ce mélange a été incubé à 12°C pendant 30 min et la réaction a été arrêtée en ajoutant 6,5 µl de « Stop-mix » 5X (EDTA 100 mM, SDS 0,5%, glycérol 25% et bleue de bromophénol 0,1%) et en chauffant 5 min à 65°C avant de déposer sur la glace. Une aliquote de 1,5 µl d'ADN réparé a été conservée pour visualiser sur gel d'agarose 0,8% la qualité de l'ADN nébulisé avant et après électroélution (figure 8). Le tampon de migration utilisé pour ce gel est du tampon TAE 1X (242 g de Tris base 1M, 57,1 ml d'acide acétique glacial, EDTA 50 mM et complété à 1 litre avec de l'eau distillée pour avoir un tampon 50X) contenant 250 µg de bromure d'éthidium.

2.4.1.3 Récolte des fragments d'ADN entre 1 et 2 kb

Afin de recueillir les fragments d'ADN nébulisés et réparés, un gel d'agarose « low melting » (BRL) 0,8% contenant 12,5 µg de bromure d'éthidium (10 mg/ml) a été soumis à l'électrophorèse dans du tampon TAE 1X (242 g de Tris base 1M, 57,1 ml d'acide acétique glacial, EDTA 50 mM et complété à 1 litre avec de l'eau distillée pour avoir un tampon 50X) contenant 250 µg de bromure d'éthidium. Une fois la migration terminée, le gel a été déposé sur un transilluminateur et la portion d'ADN contenant les fragments entre 1 et 2 kb a été découpée à l'aide d'une lame stérile.

Les fragments d'ADN contenus dans le bloc d'agarose ont été récupérés par électroélution. L'ADN a été électroélué pendant 7 h à 120 V dans un électroéluteur (CBS Scientific Co.) en présence de tampon TAE 1X pour électroélution (Tris-HOAc 25 mM pH 7,5, EDTA 5 mM pour avoir un tampon 5X). La solution d'ADN concentré a ensuite été récupérée et l'ADN a

Figure 8. Visualisation de la qualité de l'ADN avant et après nébulisation. Puits 1 : standard de poids moléculaire λ ADN/*Hind*III, ϕ X174 RF ADN/*Hae*III, Puits 2 : fraction 1/20 de la préparation d'ADN réparé et nébulisé, Puits 3 : fraction 1/60 de la préparation d'ADN réparé et nébulisé, Puits 4 : fraction 1/20 de la préparation de fragments d'ADN entre 1 et 2 kb électroéluée, Puits 5 : standard d'ADN λ nébulisé 6,25 ng/ μ l, Puits 6 : standard d'ADN λ nébulisé 12,5 ng/ μ l, Puits 7 : standard d'ADN λ nébulisé 25 ng/ μ l, Puits 8 : standard d'ADN λ nébulisé 50 ng/ μ l, Puits 9 : Marqueur de poids moléculaire.



été précipité à -20°C pendant 12 h dans du NaCl 200 mM et deux volumes d'éthanol 95% froid avant d'être centrifugé à 12000 rpm/ 4°C pendant 30 min dans la centrifugeuse Sorvall RC5C avec un rotor SA-600, lavé avec de l'éthanol 80% froid et centrifugé de nouveau 10 min. Le culot, exempt de sels et légèrement séché, a été resuspendu dans 18 μl d'eau bidistillée stérile « molecular biology GRADE » (5 prime \rightarrow 3 prime) et 2 μl de TE 10X (1 : 0,1). Une fraction (1/20) d'ADN nébulisé, réparé et électroélué a été déposé sur gel d'agarose (figure 8).

2.4.1.4 Phosphorylation des fragments d'ADN

Les fragments d'ADN nébulisés et réparés ont été phosphorylés afin de faciliter le clonage. La phosphorylation des extrémités des fragments d'ADN diminue les chances que deux fragments se lient ensemble et améliore le pourcentage de ligature avec un vecteur linéarisé dont les deux extrémités sont déphosphorylées. Le mélange de réaction de la phosphorylation contenait 50 ng d'ADN, 5 μM d'ATP (Pharmacia), 0,5 μl de tampon kinase 10X (Tris-HCl 500 mM pH 7,6, MgCl_2 100 mM, DTT 50 mM, spermidine 1 mM, EDTA 1 mM), 0,25 μl de T4 polynucléotide kinase 10 U/ μl (Fermentas) et 1,25 μl d'eau bidistillée stérile « molecular biology GRADE » (5 prime \rightarrow 3 prime) pour un volume final de 5 μl . La réaction s'est déroulée pendant 30 min à 37°C et a été arrêtée par chauffage à 65°C pendant 15 min.

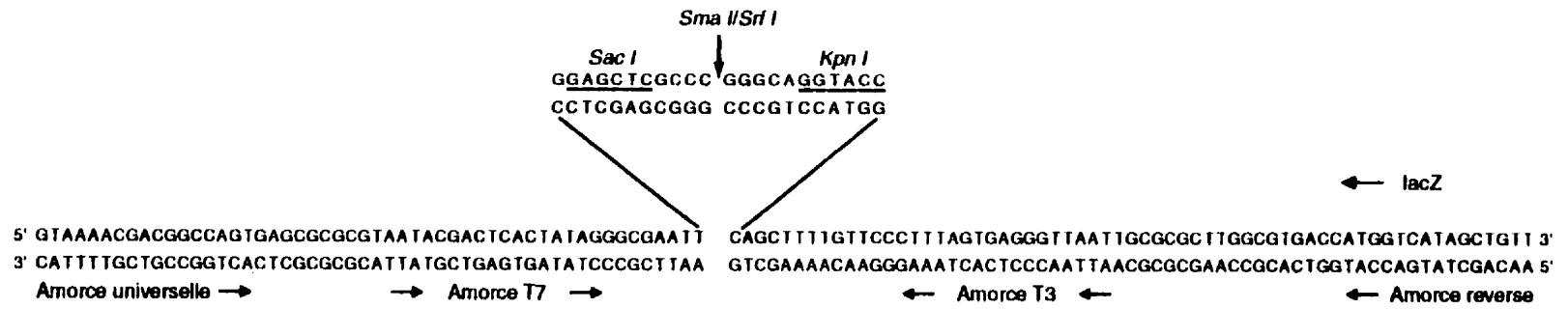
2.4.2 Préparation du vecteur pFBS KS⁺

Le vecteur utilisé pour le clonage est le plasmide pFBS KS⁺ (figure 9). Ce vecteur est une version modifiée au niveau du site de clonage multiple du vecteur pBluescript II KS⁺. Le site de clonage multiple du vecteur pFBS KS⁺ est beaucoup plus court que celui du vecteur pBluescript KS⁺. Il contient seulement les sites de coupure des enzymes *KpnI*, *SacI* (situés aux extrémités du site de clonage multiple), *SmaI* et *SrfI*.

2.4.2.1 Digestion du vecteur

Afin de faciliter le clonage, le plasmide (20 μg) a été digéré pendant 2 h à 30°C avec 10 μl

Figure 9. Séquence du vecteur pFBS KS⁻. La séquence des amorces universelles, l'emplacement du gène *lacZ* ainsi que les sites de reconnaissance des enzymes de restriction *SacI*, *SmaI/SrfI* et *KpnI* sont indiqués.



d'enzyme de restriction *SmaI* 100 U/ μ l (Fermentas) dans une solution de 40 μ l de tampon Y⁺/Tango (Tris-acétate 33 mM pH 7,9, acétate de magnésium 10 mM, acétate de potassium 66 mM, BSA 0,1 mg/ml) (Fermentas) et 302 μ l d'eau bidistillée stérile « molecular biology GRADE » (5 prime \rightarrow 3 prime) pour un volume total de 400 μ l. La réaction a été arrêtée avec 8 μ l d'EDTA 500 mM. Le vecteur linéarisé a été purifié par deux extractions avec un volume égal de phénol/chloroforme et une extraction avec un volume égal de chloroforme. Il a ensuite été précipité avec du NaCl 5M (concentration finale de 200 mM) et deux volumes d'éthanol 95% à -20°C pendant 12 h avant d'être centrifugé 30 min à 12 000 rpm/ 4°C dans une centrifugeuse RC5C Sorvall Instruments (Du Pont) avec un rotor SA-600, lavé avec 200 μ l d'éthanol 80%, séché et resuspendu dans 50 μ l d'eau. Une aliquote de 10 μ l d'ADN a été conservée dans du tampon TE 1X (Tris-HCl 1 mM, EDTA 0,1 mM).

2.4.2.2 Déphosphorylation du vecteur

Toujours pour des fins de clonage, une aliquote de 15 μ g de vecteur pFBS KS+ linéarisé a été déphosphorylée pendant une heure à 37°C avec 0,5 μ l de phosphatase alcaline 20 U/ μ l (Boehringer) dans une solution contenant 5 μ l de tampon de déphosphorylation 10X (Tris-HCl 100 mM pH 7,5, MgCl_2 100 mM) et 4,5 μ l d'eau. Après l'incubation, 1 μ l d'EDTA 0,5 M a été ajouté et le mélange a été chauffé à 65°C pendant 10 min afin d'arrêter la réaction. Le vecteur a été purifié par l'ajout de 150 μ l d'une solution de TE 1 X (Tris-HCl :EDTA (10 :1)) et de NaCl 200 mM ainsi que deux extractions au phénol/chloroforme : alcool isoamylique (24 : 1) et une extraction au chloroforme : alcool isoamylique (24 : 1). L'ADN a été précipité avec deux volumes d'éthanol 95% à -20°C pendant 12 h et centrifugé à 12000 rpm/ 4°C dans une centrifugeuse RC5C Sorvall Instruments (Du Pont) avec un rotor SA-600 pendant 30 min. Le culot a ensuite été lavé avec 100 μ l d'éthanol 80% et resuspendu dans 100 μ l de TE 1X (Tris-HCl : EDTA (10 : 0,1)).

2.4.2.3 Test de ligature

Un test de ligature a été fait pour vérifier l'efficacité de la digestion et de la

déphosphorylation. Pour ce faire, cinq mélanges de ligature d'un volume final de 10 μ l contenant 25 ng de plasmide, 1 μ l de tampon de ligature 10X (Tris-HCl 200 mM pH 8,0, Tris-HCl 200 mM pH 7,5, MgCl₂ 100 mM, DTT 100 mM, BSA 250 μ g/ml acétylé), 1 μ l de PEG 4000 50% (Fermentas), 0,5 μ l d'ATP 10 mM, 5,5 μ l d'eau bidistillée stérile « molecular biology GRADE » (5 prime \rightarrow 3 prime) et 1 μ l d'ADN ligase T4 5U/ μ l (Fermentas) ont été préparés. Le premier mélange contenait du vecteur pFBS KS+ coupé avec *Sma*I mais non-déphosphorylé, le deuxième, du vecteur pFBS KS+ coupé avec *Sma*I et déphosphorylé, le troisième, du vecteur pFBS KS+ coupé avec *Sma*I, déphosphorylé et sans ligase, le quatrième, du vecteur pFBS KS+ non-coupé et sans ligase et le quatrième, du vecteur pUC18 (contrôle). Ces mélanges ont été incubés à 14°C toute une nuit et chauffés à 65°C pendant 10 min afin d'arrêter la réaction. Ensuite, une aliquote de 200 μ l de cellules compétentes *Escherichia coli* XL 10-Gold (Stratagene) a été incubé sur la glace pendant 30 min en présence de 1 à 3 μ l d'un des mélanges de ligature dans des tubes Falcon 2059. Les cellules ont ensuite été placées à 42°C pendant 30 sec et ensuite, 2 min sur la glace (choc thermique). 0,8 ml de milieu SOC (bacto tryptone 2%, extrait de bacto levure 0,5%, NaCl 10 mM, KCl 2,5 mM, MgCl₂ 10 mM et glucose 20 mM) préchauffé à 42°C a été ajouté et le tout a été agité à 250 rpm à 37°C pendant 45 min. 100 μ l de cellules ont été étalées sur des boîtes de pétris remplies avec ~25 ml de gélose LB (10 g de NaCl, 10 g de bacto-tryptone, 5 g d'extrait de levure et 15 g de bacto-agar pour un volume final de 1 litre) contenant une solution stérile d'IPTG 0,2 mM, une solution de Xgal 80 μ g/ml diluée dans du diméthylformamide et une solution stérile d'ampicilline 75 μ g/ml. Les résultats de ces transformations ont permis de voir que la digestion et la déphosphorylation avait bien fonctionné et que le vecteur pouvait être utilisé.

2.4.3 Ligature de l'ADN nébulisé, réparé et phosphorylé et du vecteur déphosphorylé

Comme trois banques de clones ont dû être faites afin d'accumuler suffisamment de clones pour séquencer le génome chloroplastique et le génome mitochondrial en entier, cette étape ainsi que l'étape de transformation, celle de repiquage des clones positifs et celle d'hybridation ont dû être répétées trois fois. Pour la première banque, qui était une petite banque préliminaire construite pour vérifier la composition des clones, le mélange de ligature

contenait 8 ng d'ADN phosphorylé, 8 ng de vecteur pFBS linéarisé et déphosphorylé, 0,3 µl de tampon ligase 10X (Tris-HCl 200 mM pH 8,0, Tris-HCl 200 mM pH 7,5, MgCl₂ 100 mM, DTT 100 mM, BSA acétylé 250 µg/ml (Sigma)), 0,3 µl d'ATP 5 mM (Pharmacia), 0,3 µl de PEG 4000 50% (Fermentas), 0,3 µl d'ADN ligase T4 5U/µl (Fermentas) et 0,68 µl d'eau bidistillée « molecular biology GRADE » (5 prime → 3 prime) pour un volume total de 3µl. Pour la deuxième banque, le mélange de ligature contenait 16 ng d'ADN, 16ng de vecteur, 0,6 µl de tampon ligase 10X, 0,3 µl d'ATP 10 mM (Pharmacia), 0,6 µl de PEG 4000 50% (Fermentas), 0,6 µl d'ADN ligase T4 5U/µl (NEB) et 2,1 µl d'eau bidistillée « molecular biology GRADE » (5 prime → 3 prime) pour un volume total de 6 µl. Pour la troisième banque, trois mélanges de ligature ont été faits selon les mêmes conditions que pour les deux autres banques sauf pour la provenance de l'ADN ligase T4. Dans le premier et le troisième mélange, l'ADN ligase T4 qui a été utilisée est celle de Fermentas tandis que dans le deuxième, c'est celle de NEB qui a été utilisée. Le rapport molaire insert : vecteur était de 2 : 1 pour toutes les réactions de ligature. Ces mélanges ont été incubés à 14°C toute une nuit et chauffés à 65°C pendant 10 min.

2.4.4 Transformations

Afin d'optimiser la transformation, plusieurs types de cellules compétentes ont été utilisés. Pour la première banque, 50 µl de cellules ultracompetentes *Epicurian coli* XL 10-Gold (Stratagene) ont été transformées tandis que pour la deuxième banque, 75 µl de cellules compétentes *E. coli* XL 10-Gold résistantes à la kanamycine (Stratagene) ont été utilisées. Les deux types de cellules ont été transformés avec les mélanges de ligature de la troisième banque. Pour le premier mélange de ligature de cette banque, les deux types de cellules ont été utilisés, les cellules non résistantes à la kanamycine ont été transformées par le deuxième mélange de ligature et les cellules résistantes à la kanamycine ont été utilisées pour le troisième mélange de ligature. Pour toutes les transformations, 2 µl de β-mercaptoéthanol ont été ajoutés aux cellules pour chaque 50 µl de cellules et ces dernières ont été incubées sur la glace pendant 10 min avec des agitations circulaires à toutes les 2 min. Le β-mercaptoéthanol est fourni avec les cellules compétentes de Stratagene et semble augmenter de façon

significative l'efficacité de la transformation. Une aliquote de 5 ng de mélange de ligature a été ajoutée aux cellules et celles-ci ont été incubées sur la glace pendant 30 min. Les cellules ont par la suite été incubées à 42°C pendant 30 sec et placées sur la glace 2 min (choc thermique). Du milieu NZY⁺ (10g de NZ amine, 5 g d'extrait de bacto levure, 5 g de NaCl, 965 ml d'eau bidistillée et ajustement du pH à 7,5 avant d'ajouter 12,5 ml d'une solution stérile de MgCl₂ 1M, 12,5 ml d'une solution stérile de MgSO₄ 1M et 10 ml d'une solution stérile de glucose 2 M pour avoir un volume final de 1 litre) préchauffé à 42°C a été ajouté aux cellules (450 µl/50µl de cellules) et le tout a été incubé à 37°C pendant 45 min avec agitation à 250 rpm. Un nouvel ajout de milieu NZY⁺ a été fait afin de pouvoir étaler 100 µl de cellules sur des plaques de gélose LB-IPTG-Xgal-Ampicilline (voir section 2.4.2.3, Test de ligature).

2.4.5 Repiquage des clones positifs et transferts sur filtres de nylon

Environ 400, 2000 et 1500 clones positifs (colonies blanches) des banques 1, 2 et 3 respectivement ont été recueillis avec des cure-dents. Chaque clone a été déposé dans un puits d'une microplaques de 96 puits contenant 100 µl d'une préparation de milieu LB (200 ml de milieu LB et 10 mg de carbenicilline (50 mg/ml)). Au total, 39 microplaques ont été utilisées. Les plaques ont été incubées toute une nuit à 37°C. Des filtres de nylon ont été préparés et 10 µl de chaque culture de clone y ont été transférés à l'aide d'une pipette à huit canaux. Une fois séchés, les filtres ont été traités 3 min dans une solution de SDS 10%, 5 min dans une solution dénaturante (NaCl 1,5M, NaOH 0,5N), deux fois 3 min dans une solution neutralisante (NaCl 1,5M, Tris-HCl 0,5M pH 7,5) et 5 min dans une solution de lavage (2X SSC). L'ADN a été fixé aux filtres secs par exposition aux rayons ultraviolets à 254 nm pendant 2 min 30 sec. Une aliquote de 30 µl de glycérol 50% stérile a été ajoutée à chaque puits des microplaques qui ont ensuite été agitées (Brinkmann : Titermix) pendant 3 min. Les clones ont été conservés à –70°C.

2.5 Hybridation des filtres à une sonde d'ADN fabriquée à l'aide de la fraction originale d'ADN utilisée pour faire la banque

Cette étape est essentielle car l'ADN des organites utilisé n'est pas parfaitement pur; il peut contenir des contaminants de nature bactérienne ou nucléaire. En hybridant les clones avec l'ADN original ayant servi à fabriquer la banque, on peut localiser ceux ayant incorporé un ou plusieurs fragments d'origine mitochondriale ou chloroplastique. Normalement, ces clones vont montrer un signal d'hybridation fort ou moyen en raison du plus grand nombre de copies de fragments chloroplastiques et mitochondriaux dans l'ADN utilisé.

2.5.1 Préparation des filtres pour l'hybridation

Les filtres de nylon ont été incubés à 65°C pendant 4 h dans 300 ml de solution de préhybridation contenant 0,3 g de SDS, 0,6 g de BSA, 0,6 g de polyvinylpyrrolidone, 0,6 g de Ficoll 4000 (Pharmacia), 90 ml de 20X SSC (1402,4 g de NaCl et 705,6 g de citrate de sodium · 2 H₂O pour un volume final de 8 litres, ajusté à un pH de 7,0) et 0,6 ml d'EDTA 500 mM. Une aliquote de 1 ml d'une solution d'ADN de sperme de saumon 10 mg/ml a été ajoutée pour chaque 100 ml de solution de préhybridation (pour avoir une concentration finale de 100 µg/ml) afin de bloquer les sites non-spécifiques d'hybridation. La composition du milieu utilisée pour l'hybridation est la même que celle utilisée pour la préhybridation.

2.5.2 Construction de la sonde

La trousse NEBlot (BioLabs) a été utilisée pour préparer la sonde. Une aliquote de 25 ng d'ADN (160 ng/µl) a été ajoutée à 32 µl d'eau et dénaturée en la plaçant dans de l'eau bouillante pendant 5 min et ensuite 5 min dans la glace. Après l'addition de 5 µl de tampon 10X incluant les « Random Octadeoxyribonucleotides », 6 µl du mélange de dNTP (2 µl de dCTP, 2 µl de dTTP et 2 µl de dGTP), 5 µl de dATPα [³²P] 10 mCi/µl et 1 µl du fragment Klenow de l'ADN polymérase I (5 U), le mélange a été incubé durant 1 h à 37°C et la réaction a été arrêtée en ajoutant 5 µl d'EDTA 0,2 M, pH 8,0. Deux sondes ont été fabriquées pour

l'hybridation des filtres de nylon de la première banque et quatre pour chacune des deux autres banques. Les dNTPs non-appariés ont été enlevés du mélange réactionnel avec la trousse QIAquick Nucleotide Removal (Quiagen).

2.5.3 Hybridation

Les sondes ont été ajoutées au tampon d'hybridation et l'incubation a été poursuivie à 65°C durant environ une journée. Les filtres ont par la suite été lavés deux fois pendant 20 min avec 200-300 ml d'une solution de 2X SSC, 0,1% SDS à température de la pièce et deux ou trois fois pendant 30 min avec 200-300 ml d'une solution de 0,1X SSC, 0,1% SDS à 50°C. Suite aux lavages, les filtres ont été exposés à des films à rayons X et placés plusieurs jours à -70°C avant d'être développés et que les clones positifs soient localisés.

2.6 Purification des clones pour le séquençage

Suite à l'analyse des films à rayons X, des clones dont le signal d'hybridation était moyen ou fort ont été choisis et placés en culture pendant une nuit dans 3 ml d'une préparation contenant 200 ml de milieu LB + 20% d'extrait de levure et 10 mg de carbenicilline (50 mg/ml). Les cultures ont été transférées dans des tubes Eppendorf de 3 ml et centrifugées à 10 000 rpm/température de la pièce pendant 2 min dans une centrifugeuse Eppendorf 5417C afin de faire précipiter les cellules. Les plasmides des clones ont été extraits et purifiés à l'aide de la trousse QIAprep Miniprep (Quiagen). Le protocole qui a été suivi est celui utilisant un QIAvac 65 Manifolds. Les plasmides purifiés ont été dosés afin de connaître la concentration de leur ADN et de les ajuster à 50 ng/μl. Une aliquote de 100 ng des clones a ensuite été digérée à 37°C pendant 1 h dans un thermocycleur Gene Amp PCR system 9600 (Perkin Elmer) avec 0,5 μl d'enzyme *KpnI* 10 U/ml (Biolabs) et 0,5 μl d'enzyme *SacI* 11,3 U/ml (Amersham Pharmacia Biotech Inc) en présence de 1 μl de tampon OPA 10X (Tris-acétate 100 mM pH 7,5, magnésium-acétate 100 mM, potassium-acétate 500 mM (Amersham Pharmacia Biotech Inc)) pour un volume final de 10 μl ajusté avec de l'eau stérile. Cette étape a permis de libérer l'insertion du vecteur pFBS KS⁺. Les digestions de plasmides ont été déposées sur gel

d'agarose 0,8% afin de déterminer la taille des fragments d'ADN clonés et de choisir les clones qui seront séquencés (ceux dont la taille de l'insertion varie entre 1 et 2 kb).

2.7 Séquençage

2.7.1 Méthodes et trousse utilisées pour le séquençage

Les séquences ont été obtenues à l'aide de la méthode « cycle sequencing » et avec un séquenceur automatique 373 DNA sequencer stretch avec Option XL (ABI) du laboratoire d'analyse et de synthèse d'acides nucléiques du pavillon Charles Eugène Marchand de l'Université Laval. Cet appareil, qui permet de séquencer environ 800 bases par séquence, nécessite un gel « Long Ranger Singel Pack » (ABI) pour le type de séquenceur 373-48 cm. Le tampon requis pour ce gel est le tampon 10X BE (Tris base 890 mM, acide borique 890 mM, Na₂EDTA· 2 H₂O 20 mM) (ABI).

Les réactions de séquence ont été faites avec la méthode utilisant des fluorochromes distincts attachés à des terminateurs didéoxy. La trousse qui a d'abord été utilisée pour faire les séquences est la trousse « ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction » (ABI) avec l'ADN polymérase AmpliTaq FS. Chaque réaction de séquence contenait 500 ng d'ADN, 3,2 µM d'amorce et 8 µl du mélange Big Dye (A-Dye terminateur marqué au dichloro[R6G], C-Dye terminateur marqué au dichloro[ROX], G-Dye terminateur marqué au dichloro[R110], T-Dye terminateur marqué au dichloro[TAMRA], dATP, dCTP, dITP, dUTP, ADN polymérase AmpliTaq FS avec pyrophosphatase thermostable, MgCl₂ et tampon Tris-HCl pH 9,0) dilué 1/3 avec du tampon pour séquençage 2,5X (Tris-HCl 200mM pH 9,0 et MgCl₂ 5 mM). Les réactions ont eu lieu dans un thermocycleur Gene Amp PCR system (Perkin Elmer) selon la méthode suivante: un cycle de dénaturation d'une min à 96°C, 25 cycles d'appariement et d'élongation (10 sec à 96°C, 5 sec à 50°C, 4 min à 60°C) et un cycle à 4°C d'au moins 15 min. Les fragments d'ADN ainsi obtenus ont été précipités selon la procédure de précipitation à l'éthanol dans les plaques MicroAmp (ABI). Avant le dépôt sur gel, 3 µl d'une solution contenant 200 µl de formamide déionisé et 40 µl d'une solution

d'EDTA 10% et de Bleu Dextran 200 (40 µg) ont été ajoutés à chaque échantillon. Les fragments d'ADN ont été dénaturés en les plaçant 3 min à 95°C et ensuite sur la glace.

Comme l'ADN à séquencer contenait des structures riches en G + C formant souvent des structures en épingle à cheveu difficiles à dénaturer, l'ADN polymérase n'arrivait pas à amplifier toutes les régions avec cette première trousse. Lorsque l'enzyme rencontrait une de ces structures, elle se décrochait ou devenait très imprécise. Des séquences beaucoup plus courtes que 800 bases et tombant suite à ces structures étaient ainsi obtenues. Pour contrer ce problème, la trousse de ABI contenant des dGTP en remplacement des dITP a été essayée. Le protocole des réactions de séquence pour cette trousse est le même que pour la première trousse. Dans la plupart des cas, avec la trousse dGTP, l'ADN polymérase réussissait à passer la structure, donnant une séquence d'une longueur autour de 800 bases mais, dans la région de la structure, la séquence n'était pas précise et les bases étaient compressées. Pour réussir à lire parfaitement et avec certitude ces séquences problématiques, on a dû se servir d'oligonucléotides choisis à environ 200 bases de chaque côté de la structure afin de séquencer celle-ci sur les brins. La trousse « DYEnamic ET Terminator cycle » (Amersham pharmacia biotech) a aussi été essayée pour séquencer les structures. Le protocole utilisé pour la préparation des réactions de séquence est celui fourni avec la trousse Amersham et est différent de celui des autres trousse. Le cycle d'élongation de la réaction PCR est aussi différent : 95°C pendant 20 sec, 50°C pendant 15 sec et 60°C pendant 1 min. Cette trousse a été moins efficace que la trousse dGTP (ABI) pour passer les structures mais, pour les quelques séquences qui ont fonctionné, aucune compression de bases n'a été observée.

2.8 Analyses des séquences et assemblage des génomes

La stratégie utilisée pour séquencer le génome chloroplastique et le génome mitochondrial de *Monomastix* OKE-1 est le séquençage aléatoire. Parmi les clones donnant un signal d'hybridation fort ou moyen, les clones à séquencer ont été choisis au hasard. Les plasmides choisis et purifiés ont été séquencés à l'aide de l'amorce T3 en premier lieu. L'amorce T7, qui est située de l'autre côté du site de clonage (figure 9), a servi à prolonger les séquences utiles pour l'assemblage du génome chloroplastique ou mitochondrial.

Les séquences obtenues de cette façon ont d'abord été filtrées et légèrement éditées à l'aide du logiciel Factura 2.0.1. Elles ont ensuite été comparées avec la banque de données nucléotidiques ou peptidiques de GenBank . L'outil BLAST 2.1 du serveur NCBI a permis de faire cette comparaison et de trouver les séquences similaires à des séquences connues. C'est ainsi que les séquences contenant potentiellement des gènes ou introns chloroplastiques ou mitochondriaux ont été repérées. Les séquences n'ayant aucun lien avec les séquences de GenBank ont été classées comme étant inconnues. Ces séquences inconnues pouvaient être des contaminants d'origine bactérienne, des séquences chloroplastiques ou mitochondriales intergéniques ou des séquences trop courtes (dues aux structures riches en G + C) pour donner un résultat de comparaison.

Une fois les séquences placées dans leur classe respective (chloroplastique, mitochondriale ou inconnue), elles ont été assemblées et éditées à l'aide du logiciel Autoassembler 2.1.1 (ABI). Les séquences inconnues se sont ajoutées presque entièrement aux séquences chloroplastiques et mitochondriales au fur et à mesure que ces dernières s'accumulaient. Celles qui se sont assemblées entre elles ont formé des groupes se chevauchant nommés îlots. L'accumulation de séquences à partir des amorces T3 et T7 a permis de construire des îlots de plusieurs paires de bases.

Lorsque le séquençage à l'aide de ces deux amorces est devenu redondant (environ 500 séquences et une dizaine d'îlots), des oligonucléotides de 22 nucléotides ayant une séquence complémentaire aux extrémités des îlots ont été synthétisés. Tous les oligonucléotides ont été fabriqués chez Gibco BRL Life Technologies. Ces amorces ont servi à prolonger les îlots et à relier ceux dont la contiguïté était évidente entre eux. Comme mentionné plus haut, des oligonucléotides ont aussi servi à séquencer les structures problématiques. Afin de retrouver l'ordre de la dizaine d'îlots restants et de refermer chaque génome, des produits PCR ont été synthétisés à partir des amorces complémentaires aux extrémités des îlots. Les combinaisons d'amorces permettant de synthétiser un produit PCR étaient celles de deux îlots contigus. Comme quelques amorces étaient complémentaires à des régions répétées dans les génomes, il a été difficile de retrouver l'ordre des îlots et on a dû procéder par élimination pour trouver les

bonnes combinaisons. Les produits PCR entre 100 et 4000 bases ont été séquencés à partir des oligonucléotides ayant servi à leur amplification mais aussi à l'aide d'oligonucléotides complémentaires à des séquences à l'intérieur des produits PCR pour ceux qui étaient plus long que 1500 bases. Deux produits PCR mitochondriaux étaient beaucoup plus gros (8000 et 10 000 bases). Une sonde marquée au ^{32}P a été préparée à partir de ces deux produits PCR afin de cribler la banque complète. Le protocole de la première hybridation a été utilisé pour construire la sonde et pour l'hybridation (section 2.5). Cela a permis de trouver les clones contenant les fragments identiques aux produits PCR et de séquencer plus rapidement ces régions.

2.8.1 Méthode utilisée pour les amplifications PCR

Afin d'améliorer l'efficacité de l'appariement et de l'élongation des amorces, la méthode « Hot start » a été utilisée pour les amplifications PCR. Cette méthode consiste à chauffer tous les réactifs avant d'ajouter l'ADN polymérase et l'ADN à amplifier. Chaque mélange réactionnel a donc été divisé en deux phases; une phase inférieure et une phase supérieure. Pour un volume total des deux phases de 50 μl , la phase inférieure contenait 7 μl d'eau bidistillée, 6 μl de tampon 3,3 XL fourni avec la trousse GeneAmp XL PCR (ABI), 1 μl de chaque dNTP 2,5 mM (dATP, dCTP, dGTP, dTTP) (ABI), 0,5 μl de chaque amorce 50 mM et 2,2 μl de MgOAc 25 mM fourni avec la trousse GeneAmp XL PCR (ABI). Une bille de cire a été ajoutée sur cette phase avant qu'elle soit chauffée 5 min à 80°C. La phase supérieure contenant 0,5 μl d'une fraction 1/100 de l'ADN original qui a été utilisé pour faire la banque, 1 μl d'ADN polymérase rTth fourni avec la trousse GeneAmp XL PCR (ABI), 9 μl de tampon 3,3 XL (ABI) ainsi que 19,5 μl d'eau bidistillée, a été ajoutée sur la bille de cire fondue. L'appareil ayant servi aux amplifications PCR est le thermocycleur Gene Amp PCR system (Perkin Elmer) et la méthode utilisée est la suivante : un cycle de dénaturation d'une min à 94°C, 16 cycles d'appariement et d'élongation de 15 sec à 94°C et de 12 min à 65°C, 14 cycles d'appariement et d'élongation avec ajout de 15 sec à chaque cycle, un cycle de 10 min à 72°C et un cycle à 4°C pendant au moins 15 min. Afin d'amplifier les régions répétées, plusieurs modifications du cycle d'élongation ont été essayées : température de dénaturation à 92°C, élongation de 2 ou 6 min, retrait de l'ajout de 15 sec et modification du nombre de

cycles d'élongation. Pour les régions difficiles à amplifier, l'ajout de 0,2 M de bétaine 1 M à la phase supérieure a aussi été essayé mais a été peu efficace.

2.9 Caractérisation des génomes

2.9.1 Localisation des gènes et autres cadres de lecture ouverts ainsi que des introns

La comparaison des séquences avec l'outil BLAST 2.1 du serveur NCBI a permis de trouver grossièrement l'emplacement des gènes et introns de chaque génome. Tous les gènes, à l'exception des gènes codant pour les ARNs de transfert, ont pu être localisés avec plus de certitude en comparant la séquence nucléotidique de *Monomastix* avec la séquence nucléotidique ainsi que les séquences peptidiques de *Nephroselmis olivacea* et de *Mesostigma viride* et ce à l'aide des logiciels Bestfit et Framealign du groupe de logiciels GCG version 10.1 pour UNIX (Wisconsin Package Group) sur le terminal MacX du Pavillon Marchand. La séquence peptidique des protéines a été trouvée avec le logiciel Translate de GCG. Les ARNs de transfert ont été localisés à l'aide du logiciel tRNAscan du terminal MacX et leur structure secondaire a été construite afin de s'assurer que l'emplacement trouvé par le logiciel est bon et que leur forme est standard. L'outil ORF finder du serveur NCBI ainsi que le logiciel Frames de GCG ont permis de trouver l'emplacement des cadres de lecture ouverts de plus de 60 codons (65 codons pour le génome mitochondrial) ne codant pas pour une protéine connue et n'ayant pas d'homologue chez d'autres organismes. Seuls les ORFs de plus de 60 ou 65 codons ont été considérés car il est peu probable de trouver des protéines avec une séquence protéique de moins de 60 codons. La limite minimale du nombre de codons pour la recherche des ORFs a été placée à 65 pour le génome mitochondrial car c'est un standard utilisé dans le laboratoire mais elle a été abaissée à 60 pour le génome chloroplastique puisqu'il y a plusieurs ORFs entre 60 et 65 codons dans cet ADN, ce qui n'est pas le cas pour l'ADNmt.

Chaque séquence peptidique a été alignée avec les autres séquences peptidiques chloroplastiques ou mitochondriales disponibles de chaque gène à l'aide du logiciel ClustalW 1.81 (Thompson *et al.*, 1994). Les séquences des génomes proviennent de GenBank à l'exception des séquences chloroplastiques et mitochondriales de *Pycnococcus prasovali* et de

Pedinomonas minor, de la séquence chloroplastique de *Neochloris pseudoalveolaris* et de la séquence mitochondriale de *Mesostigma viride* qui sont des résultats non publiés de notre laboratoire. Les alignements multiples ont permis de corriger la longueur de quelques séquences peptidiques de *Monomastix*. Ils ont de plus servi à faire les analyses phylogénétiques.

Le gène *rnpB*, les introns et les gènes codant pour les ARNs ribosomiques autant chloroplastiques que mitochondriaux ont été délimités en construisant leur structure secondaire. Les gabarits qui ont été utilisés sont les structures secondaires de ces gènes chez *Nephroselmis olivacea*, *Mesostigma viride* et d'autres algues vertes. Les fichiers informatiques des structures secondaires des ARNs de transfert, des introns ainsi que du gène *rnpB* autant chloroplastiques que mitochondriaux ont été réalisés à l'aide du logiciel Adobe Illustrator 8.0.

2.9.2 Construction des cartes de gènes

Une carte des gènes de chacun des génomes a été construite à l'aide du logiciel Adobe Illustrator 8.0. Ces cartes circulaires se basent sur la position angulaire de chaque gène, intron et cadre de lecture ouvert présent. La position 0, qui a été déterminée préalablement et qui est la première base du gène *trnE*(uuc) pour le génome chloroplastique et la base suivant la dernière base du gène *trnV*(uac) pour le génome mitochondrial, a été placée au point le plus élevé du cercle et le positionnement des autres gènes suit le sens des aiguilles d'une montre. Les gènes transcrits dans le sens horaire ont été placés à l'extérieur du cercle et ceux transcrits dans le sens anti-horaire, à l'intérieur du cercle.

2.9.3 Analyses phylogénétiques

Afin de faire l'analyse phylogénétique autant du génome chloroplastique que du génome mitochondrial, les alignements multiples de plusieurs protéines ont été concaténés. Deux analyses, impliquant un nombre différent de taxa et de gènes, ont été faites pour chaque génome. La première analyse du génome chloroplastique se base sur 11 taxa (*Cyanophora paradoxa*, *Mesostigma viride*, *Nephroselmis olivacea*, *Pycnococcus prasoali*, *Monomastix*

OKE-1, *Pedinomonas minor*, *Chlorella vulgaris*, *Neochloris pseudoalveolaris*, *Marchantia polymorpha*, *Pinus thunbergi*, *Nicotiana tabacum*) et sur la séquence protéique de 47 gènes (*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *clpP*, *petA*, *petB*, *petD*, *petG*, *psaA*, *psaB*, *psaI*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbN*, *psbT*, *rbcL*, *rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl36*, *rpoA*, *rpoC1*, *rpoC2*, *rps2*, *rps3*, *rps4*, *rps7*, *rps8*, *rps11*, *rps12*, *rps14*, *rps18*, *rps19*, *ycf3*, *ycf9*) de chacun des taxa. La deuxième analyse chloroplastique se base sur les mêmes taxa que la première à l'exception de *Pycnococcus prasoali* et sur la séquence protéique de *ccsA*, *psaJ*, *psbM* et *rpoB* en plus des 47 gènes de la première analyse. Le glaucocystophyte *Cyanophora paradoxa* ainsi que le prasinophyte *Mesostigma viride* ont servi à enracer les deux arbres chloroplastiques. *Cyanophora* a été choisi puisqu'il y a des évidence phylogénétiques qui indiquent que le chloroplaste des glaucocystophytes a évolué avant celui des algues rouges et des algues vertes (Martin *et al.*, 1998). Le choix de *Mesostigma* comme racine s'appuie sur le fait que le chloroplaste de ce prasinophyte a une position phylogénétique à la base des plantes vertes (Lemieux *et al.*, 2000). Les deux analyses du génome mitochondrial ont été faites à l'aide de neuf taxa (*Cyanidioschyzon merolae*, *Porphyra purpura*, *Chondrus crispus*, *Mesostigma viride*, *Nephroselmis olivacea*, *Monomastix OKE-1*, *Prototheca wickerhamii*, *Marchantia polymorpha*, *Arabidopsis thaliana*) et 19 protéines (*atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *rpl16*, *rps3*, *rps12*, *ymf16*, *ymf39*) ainsi que 11 taxa (*Porphyra purpura*, *Mesostigma viride*, *Nephroselmis olivacea*, *Pycnococcus prasoali*, *Monomastix OKE-1*, *Pedinomonas minor*, *Prototheca wickerhamii*, *Scenedesmus obliquus*, *Chlamydomonas reinhardtii*, *Marchantia polymorpha*, *Arabidopsis thaliana*) et sept protéines (*cob*, *cox1*, *nad1*, *nad2*, *nad4*, *nad5*, *nad6*) pour la première et la deuxième analyse respectivement. Puisque le génome mitochondrial des algues rouges aurait évolué avant celui des plantes vertes, les algues rouges (*Cyanidioschyzon merolae*, *Porphyra purpura* et *Chondrus crispus*) présentes dans chacune des analyses ont permis d'enracer l'arbre. La séquence concaténée des protéines chloroplastiques de la première analyse totalise 8544 sites (4761 caractères constants et 2206 informatifs) et la deuxième, 9523 sites (5274 caractères constants et 2452 informatifs). Celle de la première analyse mitochondriale totalise 4135 sites (1560 caractères constants et 1577 informatifs) et la deuxième, 2221 sites (607 caractères constants et 958 informatifs).

Toutes les analyses phylogénétiques ont été faites à l'aide de la méthode Maximum Likelihood. Une seule méthode d'analyse a été utilisée ici afin de faire des études préliminaires. Les autres méthodes d'analyses phylogénétiques (parcimonie, distance) devraient cependant être utilisées pour confirmer les résultats. La méthode Maximum Likelihood a été choisie puisqu'elle est fiable, moins controversée au niveau de l'analyse des résultats et que plusieurs modèles de substitution sont disponibles avec cette méthode. Les analyses phylogénétiques avec peu de taxa se prêtent bien à cette méthode puisqu'elle permet d'utiliser un maximum de 12 taxa. Comme pour les analyses de *Monomastix* moins de 12 taxa étaient utilisés, la méthode Maximum Likelihood pouvait être choisie. Le programme PROTML 2.3b3 (Adachi et Hasegawa, 1996) avec un modèle de substitution JTT-F a permis de trouver toutes les topologies possibles pour chaque analyse. Les topologies avec les valeurs de « RELL bootstrap » les plus élevées ont été choisies afin d'être analysées de nouveau avec le logiciel AAML (CODEML avec seqtype=2) du groupe de logiciels PAML 3.0b. La méthode des « bootstrap » est une technique qui permet d'estimer le niveau de confiance d'une hypothèse phylogénétique par rééchantillonnage de la distribution des données initiales (Graur et Li, 2000). Ainsi, pour les analyses chloroplastiques, 27 topologies ont été considérées pour la première analyse et 12 pour la deuxième. 11 topologies ont été choisies pour la première analyse mitochondriale et 24 pour la deuxième. Plusieurs modèles de substitution (cpREV45, mtREV24 et JTT) ont été essayés pour ces analyses. Ceux avec les valeurs de Likelihood les plus élevées ont été retenus. Le modèle cpREV45 avec une distribution gamma des vitesses d'évolution entre les sites a été utilisé pour les analyses chloroplastiques. Pour les analyses mitochondriales, le modèle JTT avec distribution gamma a été choisi. Les modèles avec distribution gamma ont été préférés puisque ce paramètre permet d'introduire des variations dans la vitesse de substitution entre chaque site de la séquence et que les valeurs de Likelihood calculées avec ceux-ci sont plus élevées (Baxevanis et Ouellette, 1998). Lorsqu'on exclut ce paramètre, on suppose que la vitesse de substitution à chaque site de la séquence protéique utilisée est la même, ce qui semble être moins réaliste. Les valeurs de « bootstrap » des analyses finales ont été estimées par rééchantillonnage des valeurs de Likelihood estimées. Les quatre arbres retenus ont été édités avec les logiciels TreeView 1.6.2 et Adobe Illustrator 8.0.

2.9.4 Caractérisation des séquences répétées chloroplastiques

Les séquences riches en G + C présentes dans les régions intergéniques et ayant créés des problèmes lors du séquençage ont été repérées à l'œil nu lors de l'assemblage du génome. À l'aide du logiciel Findpatterns de GCG, la séquence chloroplastique entière du génome a été criblée avec ces séquences afin de trouver celles qui sont répétées et de les localiser. Les séquences similaires (entre zéro et dix paires de bases non appariées pour les séquences plus longues) et répétées plus d'une dizaine de fois ont été regroupées en familles.

2.9.5 Analyses de la composition et de l'utilisation des bases de la séquence de chaque génome

Pour chacun des génomes, un tableau de l'utilisation des codons chez les gènes codant pour des protéines ainsi que chez les cadres de lecture ouverts trouvés à l'intérieur des introns a été construit à l'aide du logiciel Codonfrequency 1.0 de GCG. Le logiciel Assemble de GCG a permis d'extraire la portion dite codante de chaque génome, soit celle codant pour des gènes de protéines, des ARNs de transfert, des ARNs ribosomiques, des introns et leur cadre de lecture ouverts ainsi que pour le gène *rnpB*. Le pourcentage en base A + T et G + C a été calculé pour la séquence complète, la région codante et la région non codante de chaque génome à l'aide du logiciel Composition de GCG.

RÉSULTATS

3.1 Le génome chloroplastique

3.1.1 Structure et caractéristiques générales

Le génome chloroplastique de *Monomastix* OKE-1 est circulaire et d'une taille de 114 528 pb. La carte circulaire des gènes de ce génome est présentée en figure 10. La région d'ADN inversée répétée est absente de cette séquence chloroplastique. Par conséquent, la structure quadripartite typique aux génomes chloroplastiques des plantes vertes est brisée. La séquence complète de l'ADNcp de *Monomastix* sous format GenBank se retrouve en annexe 1 et sera éventuellement déposée dans cette banque de données.

Le tableau VI résume les caractéristiques du génome chloroplastique de *Monomastix*. Ces 94 gènes contenus dans 55,5% de la séquence font de lui un génome peu dense. Trois de ces gènes contiennent un ou plusieurs introns. Le pourcentage en bases A + T de ce génome est de 61,0%. Le pourcentage en bases A + T de la région codante (61,7%) ainsi que celui des régions intergéniques (60,2%) sont similaires à celui du génome en entier.

L'ADNcp de *Monomastix* contient un grand nombre de séquences répétées. Ces séquences riches en G + C sont dispersées dans tout le génome mais elles se regroupent surtout dans les régions intergéniques. Ces séquences ont été classées en six familles distinctes et des liens ont été trouvés entre les différentes familles (voir section 3.1.6).

3.1.2 Contenu en gènes

Les tableaux VI et VII présentent le contenu en gènes de l'ADNcp de *Monomastix*. En plus des 94 gènes codant pour des protéines ou ARNs de fonction connue, ce génome contient 13 ORFs uniques de plus de 60 codons qui n'ont ni fonction connue, ni homologue chez d'autres organismes. Tous les gènes et ORFs sont présents en une seule copie.

Figure 10. Génome chloroplastique de *Monomastix* OKE-1. Les gènes à l'extérieur de la carte sont transcrits dans le sens horaire; ceux situés à l'intérieur de la carte sont transcrits dans le sens antihoraire. Les ARNt sont indiqués par le code à une lettre des acides aminés suivi de l'anticodon entre parenthèses. Les ORFs uniques à l'ADNcp de *Monomastix* ne sont pas indiqués à l'exception de ceux situés à l'intérieur des introns et codant possiblement pour une endonucléase. Les exons des gènes contenant des introns sont indiqués par ex.

Tableau VI. Résumé des caractéristiques du génome chloroplastique de *Monomastix* OKE-1.

Traits	Génome chloroplastique
Longueur (pb) ^a	114 528
Contenu en gènes	
total	94
protéines	
photosynthétiques	32
ribosomiques	18
transcription	6
/traduction	6
division	1
autres	7
ARNr	3
ARNt	26
Autres ARNs	1
ORFs introniques (I/II) ^b	5/0
ORFs uniques ^c	13
Introns	
groupe I	5
groupe II	1
% codant ^d	55,5
densité en gènes (nombre de gènes/kb ^a)	0,82
% A + T	
total	61,0
codant ^d	61,7
non-codant	60,2
Séquences répétées	
total	493
régions intergéniques	471
dans un gène	22

^a pb, paire de bases; kb, kilo bases.

^b Les ORFs introniques I sont ceux appartenant à un intron du groupe I; les ORFs introniques II sont ceux appartenant à un intron du groupe II.

^c Les ORFs de plus de 60 codons, de fonction inconnue et n'ayant pas d'homologue chez d'autres organismes sont notés. Ils ne sont pas calculés dans le contenu en gènes et n'incluent pas les ORFs introniques.

^d Représente le pourcentage du génome qui code pour les gènes de protéines, d'ARNt, d'ARNr, d'autres ARNs ainsi que pour les introns et les ORFs introniques.

Tableau VII. Gènes et ORFs uniques présents dans l'ADNcp de *Monomastix* OKE-1.

Produits des gènes	Gènes
Photosystème I	<i>psaA, B, I, J, M</i>
Photosystème II	<i>psbA, B, C, D, E, F, H, I, J, K, L, M, N, T</i>
Cytochrome <i>b6/f</i>	<i>petA, B, D, G, N</i>
ATP synthétase	<i>atpA, B, E, F, H, I</i>
Rubisco	<i>rbcL</i>
Protochlorophyllide réductase	<i>chlI</i>
Protéines de la grande sous-unité ribosomale	<i>rpl2, 5, 14, 16, 20, 23, 32, 36</i>
Protéines de la petite sous-unité ribosomale	<i>rps2, 3, 4, 7, 8, 11, 12, 14, 18, 19</i>
ARN polymérase	<i>rpoA, B, C1, C2</i>
Facteurs impliqués dans la traduction	<i>infA, tufA, clpP</i>
Protéines de division chloroplastique	<i>ftsH</i>
Autres protéines	<i>ccsA</i>
Protéines de fonction inconnue	<i>ycf1, 3, 9, 12, 20</i>
ORFs uniques ^a	<i>orf60, 62, 65, 66, 67, 71, 73, 75, 78, 82, 84, 122, 125</i>
ARNs ribosomiques	<i>rrf, rrl, rrs</i>
ARNs de transfert ^b	<i>A(ugc), C(gca), D(guc), E(uuc), F(gaa), G(gcc), G(ucc), H(gug), I(cau), I(gau), K(uuu), L(uaa), L(uag), Me(cau), Mf(cau), N(guu), P(ugg), Q(uug), R(acg), R(ucu), S(gcu), S(uga), T(ugu), V(uac), W(cca), Y(gua)</i>
Autre petit ARN	<i>rnpB</i>

^a Seuls les ORFs uniques à l'ADNcp de *Monomastix* de plus de 60 codons sont notés.

^b Les ARNt sont indiqués par le code à une lettre des acides aminés suivi de l'anticodon entre parenthèses.

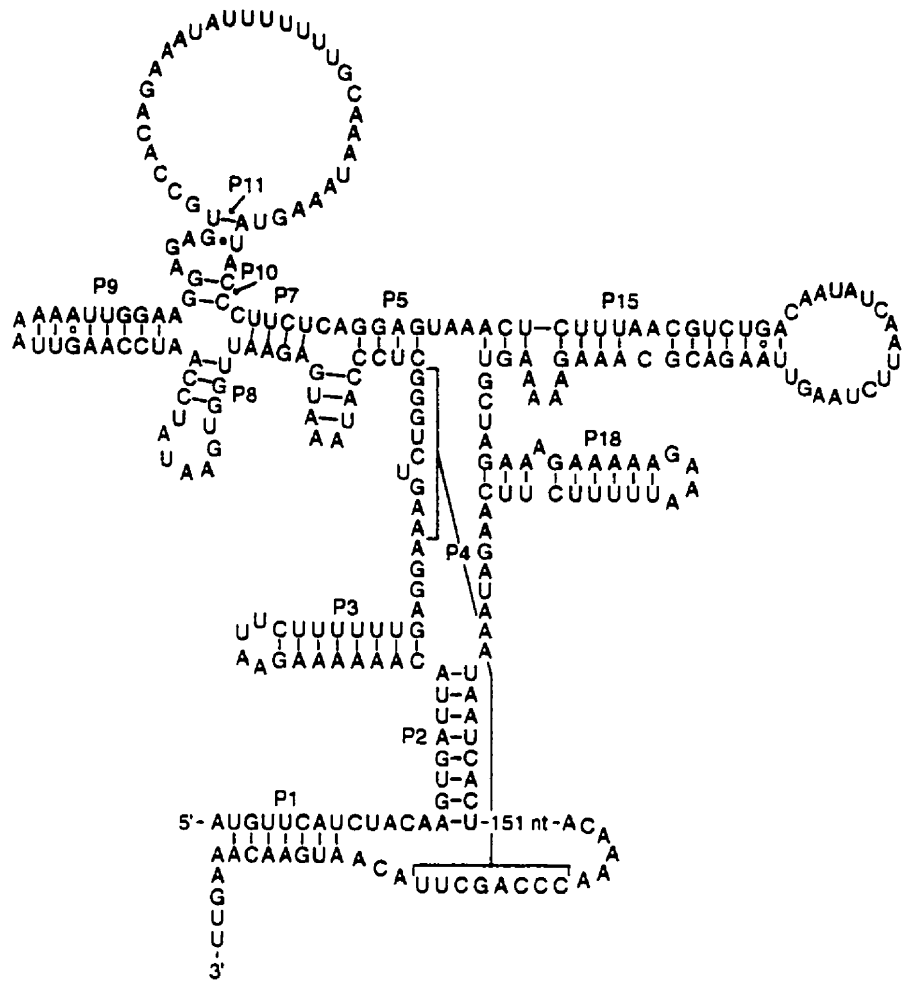
Le génome chloroplastique de *Monomastix* comprend autant de gènes codant pour des protéines photosynthétiques (32 gènes) que de gènes codant pour des protéines impliquées dans l'expression génétique du chloroplaste combinés aux gènes *ycf* de fonction inconnue (32 gènes). Parmi les gènes photosynthétiques, cinq codent pour des protéines du photosystème I, 14 codent pour des protéines du photosystème II, cinq codent pour le cytochrome *b6/f*, six pour l'ATP synthétase, un pour la sous-unité magnésium-chélatase de la protochlorophyllide réductase et un pour la grande sous-unité de la protéine Rubisco (tableau VII). Les gènes impliqués dans l'expression génétique sont les gènes codant pour des protéines ribosomiques (huit codant pour la grande sous-unité et dix pour la petite sous-unité), des sous-unité de l'ARN polymérase (quatre), des facteurs impliqués dans la traduction (deux), une protéine de division, la sous-unité protéolytique d'une protéase dépendante de l'ATP (*clpP*) ainsi que d'une protéine impliquée dans la biogénèse du cytochrome c (*ccsA*). Ce génome contient aussi cinq gènes codant pour des protéines de fonction inconnue. Les gènes codant pour les trois sous-unités de l'ARN ribosomique chloroplastique ont été localisés. La structure secondaire de la sous-unité 5S de l'ARN ribosomique (codée par le gène *rrf*) est présentée en figure 11b. Les structures secondaires des extrémités 3' et 5' des sous-unités 16S et 23S de l'ARN ribosomique (gènes *rrs* et *rrl* respectivement) ont aussi été construites afin de s'assurer de la localisation de ces gènes. Puisque ces structures ne sont pas complètes, elles ne sont pas présentées. Le gène *rnpB*, codant pour l'ARN composant l'enzyme RNaseP, a aussi été localisé. La structure secondaire de cet ARN est présentée en figure 11a. Les gènes codant pour 26 ARNt ont été trouvés dans l'ADNcp de *Monomastix*. Les ARNt sont décrits dans la section suivante.

3.1.3 Les ARNs de transfert et la composition des codons

Les gènes codant pour les ARN de transfert de l'ADNcp de *Monomastix* (tableau VI et VII) sont dispersés dans tout le génome, seuls ou en groupes de deux à cinq. La structure secondaire de chacun des ARNt chloroplastiques est présentée en figure 12. Ils assument tous, avec très peu de déviations, la structure standard en trèfle à quatre feuilles des ARNt conventionnels (figure 5). La modification la plus fréquente est au niveau du bras D, où il y a un appariement de paire de bases de moins que dans la structure standard. Elle est présente

Figure 11. Structure secondaire des ARNs codés par les gènes chloroplastiques *rnpB* et *rrf*. Les appariements entre nucléotides de type Watson-Crick (A:U, G:C) sont indiqués par (—). Les appariements G:U sont indiqués par un cercle plein et les appariements purine : purine (G:A) sont indiqués par un cercle vide. a) Structure secondaire de l'ARN codé par le gène *rnpB* et composant l'enzyme RNaseP. Chaque région d'appariement est indiquée par la lettre P suivi d'un chiffre. b) Structure secondaire de l'ARNr 5S codé par le gène *rrf*. Chaque région d'appariement est indiquée par un chiffre romain et chaque région variable est identifiée par une lettre.

a)



b)

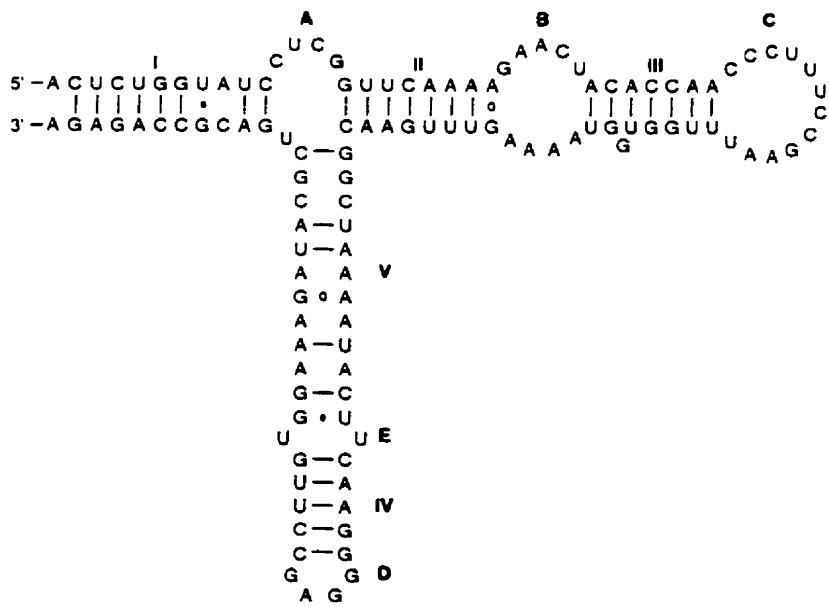


Figure 12. Structure secondaire des ARNt codés par le génome chloroplastique. Les appariements entre nucléotides de type Watson-Crick (A:U, G:C) sont indiqués par (—). Les appariements G:U sont indiqués par un cercle plein et les appariements purine : purine (G:A) sont indiqués par un cercle vide. Pour chaque ARNt, le gène le codant est indiqué. Le code à une lettre des acides aminés est utilisé. L'anticodon est en lettres minuscules. Le gène *trnMe(cau)* code pour l'ARNt^{met} élongateur et le gène *trnMf(cau)* code pour l'ARNt^{met} initiateur.

chez les ARNt^{Cys}(GCA), ARNt^{His}(GUG), ARNt^{Leu}(UAA), ARNt^{Leu}(UAG), ARNt^{Gln}(UUG) et ARNt^{Ser}(UGA). Ce genre de modification est aussi présent dans le bras accepteur des ARNt^{Met}(CAU) initiateur et ARNt^{Arg}(ACG) ainsi que dans le bras pyrimidine de l'ARNt^{Met}(CAU) élongateur. Un autre trait non typique aux ARNt est la présence d'un U et d'un A aux positions 53 et 63 respectivement de l'ARNt^{Glu}(UUC) ainsi que la présence d'un A et d'un U aux positions 11 et 24 respectivement de l'ARNt^{Met}(CAU) initiateur.

Dans le tableau VIII, se retrouve la fréquence d'utilisation des codons en relation avec les ARNt présents chez les 64 protéines et les cinq ORFs introniques de l'ADNcp de *Monomastix*. Ce tableau montre que les 26 ARNt chloroplastiques sont suffisants pour décoder les 61 codons du code génétique universel. Sept cas de dégénérescence (« superwobble ») à la première position de l'anticodon sont observés, soit pour les acides aminés leucine, valine, sérine, proline, thréonine, alanine et glycine. Tous les codons sont utilisés en quantité appréciable à l'exception de 24 codons qui sont présents entre 0 et 12 % des cas. Tous ces codons ont un C ou un G en troisième position sauf les codons Leu-CUA et Ile-AUA qui sont utilisés dans 8 et 5 % des cas respectivement. Le ratio des codons se terminant par un A ou un U versus ceux se terminant par un G ou un C est de 5,4: 1. Parmi les codons se terminant par une purine, le ratio des codons se terminant par un A versus ceux se terminant par un G est de 6,9: 1 et parmi ceux se terminant par une pyrimidine, le ratio des codons se terminant par un U versus ceux se terminant par un C est de 5,0: 1. Ces résultats montrent qu'il y a un biais d'utilisation des bases A et U en troisième position des codons par rapport aux bases G et C. Parmi les trois codons de terminaison, seul le codon UGA n'est pas utilisé.

3.1.4 Organisation des gènes

Les gènes du génome chloroplastique de *Monomastix* sont dispersés sur les deux brins en deux grands blocs entrecoupés par quelques gènes (figure 10). Plusieurs unités de transcription semblent donc nécessaires pour transcrire ce génome. Le tableau IX ainsi que la figure 13 présentent les regroupements de gènes conservés chez six algues vertes. Seulement huit des 14 regroupements de gènes conservés chez la plupart de ces algues vertes sont présents chez

Tableau VIII. Utilisation de codons chez 64 gènes codant pour les protéines conservées ainsi que pour les cinq ORFs introniques de l'ADNcp de *Monomastix* OKE-1.

AA ^a	Codon	Anticodon ^b	% ^c	AA ^a	Codon	Anticodon ^b	% ^c	AA ^a	Codon	Anticodon ^b	% ^c	AA ^a	Codon	Anticodon ^b	% ^c
F	UUU	gaa	65	S	UCU	uga	26	Y	UAU	gua	63	C	UGU	gca	88
F	UUC	gaa	35	S	UCC	uga	6	Y	UAC	gua	37	C	UGC	gca	12
L	UUA	uaa	53	S	UCA	uga	35	*	UAA		90	*	UGA		0
L	UUG	uaa	6	S	UCG	uga	6	*	UAG		10	W	UGG	cca	100
L	CUU	uag	28	P	CCU	ugg	34	H	CAU	gug	44	R	CGU	acg ^d	48
L	CUC	uag	3	P	CCC	ugg	9	H	CAC	gug	56	R	CGC	acg ^d	9
L	CUA	uag	8	P	CCA	ugg	51	Q	CAA	uug	92	R	CGA	acg ^d	16
L	CUG	uag	2	P	CCG	ugg	6	Q	CAG	uug	8	R	CGG	acg ^d	0
I	AUU	gau	81	T	ACU	ugu	38	N	AAU	guu	64	S	AGU	gcu	22
I	AUC	gau	14	T	ACC	ugu	6	N	AAC	guu	36	S	AGC	gcu	6
I	AUA	cau	5	T	ACA	ugu	48	K	AAA	uuu	90	R	AGA	ucu	22
M	AUG	cau ^c	100	T	ACG	ugu	7	K	AAG	uuu	10	R	AGG	ucu	5
V	GUU	uac	50	A	GCU	ugc	49	D	GAU	guc	71	G	GGU	gcc+ucc	52
V	GUC	uac	4	A	GCC	ugc	7	D	GAC	guc	29	G	GGC	gcc+ucc	12
V	GUA	uac	37	A	GCA	ugc	37	E	GAA	uuc	85	G	GGA	ucc	25
V	GUG	uac	8	A	GCG	ugc	6	E	GAG	uuc	15	G	GGG	ucc	11

^a Acides aminés (code à une lettre). Les astérisques indiquent les codons de terminaisons.

^b Les anticodons sont en lettres minuscules. Des cas de dégénérescence en première position de l'anticodon sont observés pour les ARNt avec un U en première position de l'anticodon. Cela leur permet de décoder tous les membres des familles de quatre codons (*i.e.* leucine, CUN; valine, GUN; sérine UCN; proline, CCN; thréonine, ACN; alanine, GCN; glycine, GGN).

^c Pourcentage de chaque acide amine spécifié par un codon donné dans l'ADN chloroplastique de *Monomastix*.

^d Le A à la première position de l'anticodon est converti post-transcriptionnellement en un I (inosine), résultant en un ARNt^{Arg} (ICG) capable de décoder les quatre codons arginine CGN (Pfitzinger *et al.*, 1990).

^e Un ARNt^{Mct} élongateur (e) ainsi qu'un ARNt^{Mct} initiateur (f, formyl) séparés sont présents.

Tableau IX. Distribution des regroupements de gènes chloroplastiques conservés chez les algues vertes qui sont présents chez *Monomastix* OKE-1.

Regroupement ^a	Prasinophytes ^b			Trebouxiophytes ^b		
	<i>N.ol</i>	<i>M.OK</i>	<i>P.pr</i>	<i>N.ps</i>	<i>P.mi</i>	<i>C.vu</i>
1	• ^c	•	•	•	•	
2	•			•	•	
3	•	•	•	•	•	•
4	•	•	•	•	•	•
5	•		•	•	•	•
6	•	•	•	•	•	•
7	•		•	•	•	•
8	•	•	•	•	•	•
9a	•			•	• ^d	•
9b	•		•	•	• ^d	•
9c	•	•	•	•	• ^d	•
10	•			•	•	•
11a	•			•	•	
11b	•			•	•	•
11c	•	•	•	•	•	•
12	•		•	•	•	•
13a	•					•
13b	•		•		•	•
14	•	•		• ^c	•	•

^a Les regroupements sont les suivants: 1. *atpB-E*; 2. *rpoB-C1-C2-rps2*; 3. *atpI-H-F-A*; 4. *psaA-B*; 5. *psbD-psbC*; 6. *psbJ-L-F-E*; 7. *psbB-T-N-H*; 8. *petB-D*; 9a. *rps9-rpoA-rps11-rpl36-infA-rps8-rpl5-14-16-rps3-19-rpl2-23*; 9b. *rpoA-rps11-rpl36-infA-rps8-rpl5-14-16-rps3-19-rpl2-23*; 9c. *infA-rps8-rpl5-14-16-rps3-19-rpl2-23*; 10. *psaJ-P(ugg)-W(cca)*; 11a. *rps12-7-tufA-rpl19-ycf4-cemA*; 11b. *rps12-7-tufA-rpl19-ycf4*; 11c. *rps12-7*; 12. *petL-G*; 13a. *rpl32-cysT-ycf1*; 13b. *cysT-ycf1*; 14. *rrs-I(gau)-A(ugc)-rrl-rrf*.

^b *N.ol*, *Nephroselmis olivacea*; *M.OK*, *Monomastix OKE-1*; *P.pr*, *Pycnococcus prasovali*; *N.ps*, *Neochloris pseudoalveolaris*; *P.mi*, *Pedinomonas minor*; *C.vu*, *Chlorella vulgaris*.

^c La présence d'un regroupement de gènes est indiquée par •.

^d Le gène *infA* n'est pas présent chez *Pedinomonas minor*.

^e L'ordre des gènes *I(gau)* et *A(ugc)* est inversée chez *Neochloris pseudoalveolaris*.

Figure 13. Localisation des regroupements de gènes chloroplastiques conservés chez *Monomastix*. La carte de la figure 10 a été utilisée pour construire cette figure. Les regroupements représentés en rouge sont ceux du tableau IX.

Monomastix. Parmi ces regroupements, certains sont des fragments d'opérons ancestraux conservés. Le regroupement 9c (*infA-rps8-rpl5-14-16-rps3-19-rpl2-23*) est un vestige des opérons *str-L2-spc-alpha-S9* qu'on retrouve chez *E. coli* (Stoebe et Kowallik, 1999). La figure 14 présente l'opéron des protéines ribosomiques chez les algues vertes du tableau IX en comparaison avec les opérons *str-L2-spc-alpha-S9* de *E. coli* (figure 14a) ainsi que le regroupement de gènes dérivé de l'opéron *str* d'une algue verte à la base des plantes vertes, *Mesostigma viride* (figure 14b). L'opéron des protéines ribosomiques de *Monomastix* est celui qui présente le moins de similitudes avec celui de *Nephroselmis olivacea* et des autres algues vertes; il est scindé en trois et plusieurs gènes de l'opéron sont dispersés ailleurs dans le génome. Le regroupement 11c du tableau IX (*rps12-rps7*) est le vestige de l'opéron *str* retrouvé chez *Monomastix*. Le regroupement 11b, dérivé de l'opéron *str* n'est pas présent chez *Monomastix*, ni chez *Pycnococcus prasoali*. *Monomastix* est la seule algue du tableau IX qui n'a pas conservé le regroupement 12 (*petL-petG*).

Les regroupements 1, 2 et 3 du tableau IX sont des regroupements ancestraux conservés faisant partie du regroupement de gènes *atpA* (Stoebe et Kowallik, 1999). La figure 15 présente le regroupement *atpA* de quatre prasinophytes et trois trebouxiophytes. L'ordre des regroupements 1, 2 et 3 varie selon les organismes mais les regroupements 2 et 3 sont souvent adjacents. Le regroupement 2 de *Monomastix* a été brisé et est devenu le suivant : *rpoC1-rpoC2-rpoB*. L'ordre des gènes du regroupement *atpA* de *Monomastix* et *Pycnococcus* est différent de celui des autres algues vertes de la figure 15. Chez *Monomastix*, tous les gènes *atp* sont regroupés et le gène *rps2* ne se retrouve pas près des gènes *rpo* ou *atp*.

L'opéron très conservé des ARNs ribosomiques (opéron 14 du tableau IX) est présent chez *Monomastix*. Il est brisé chez *Pycnococcus*, en une seule copie chez *Monomastix* et en deux copies chez les autres algues. Dans le tableau IX, seuls *Pycnococcus* et *Monomastix* ne possèdent ni la structure d'ADN inversée répétée, ni un duplicata de l'opéron des ARNs ribosomiques.

Figure 14. Regroupement de gènes chloroplastiques trouvé chez *Monomastix* OKE-1 ainsi que six autres algues vertes et dérivé des opérons *str-L2-spc-alpha-S9* de *E. coli*. a) Opéron des protéines ribosomiques. Les lignes pleines rejoignent les gènes adjacents; les lignes pointillées indiquent la présence de gènes additionnels non montrés. Significations des abréviations: *E. coli*, *Escherichia coli*; *M. viride*, *Mesostigma viride*; *N. olivacea*, *Nephroselmis olivacea*; *M. OKE-1*, *Monomastix* OKE-1; *P. prasovali*, *Pycnococcus prasovali*; *P. minor*, *Pedinomonas minor*; *N. pseudoalveolaris*, *Neochloris pseudoalveolaris*; *C. vulgaris*, *Chlorella vulgaris* b) Regroupement présent dans l'ADNcp de *Mesostigma viride* et dérivé du regroupement *str* de *E. coli*. Figure inspirée de Turmel *et al.*, 1999.

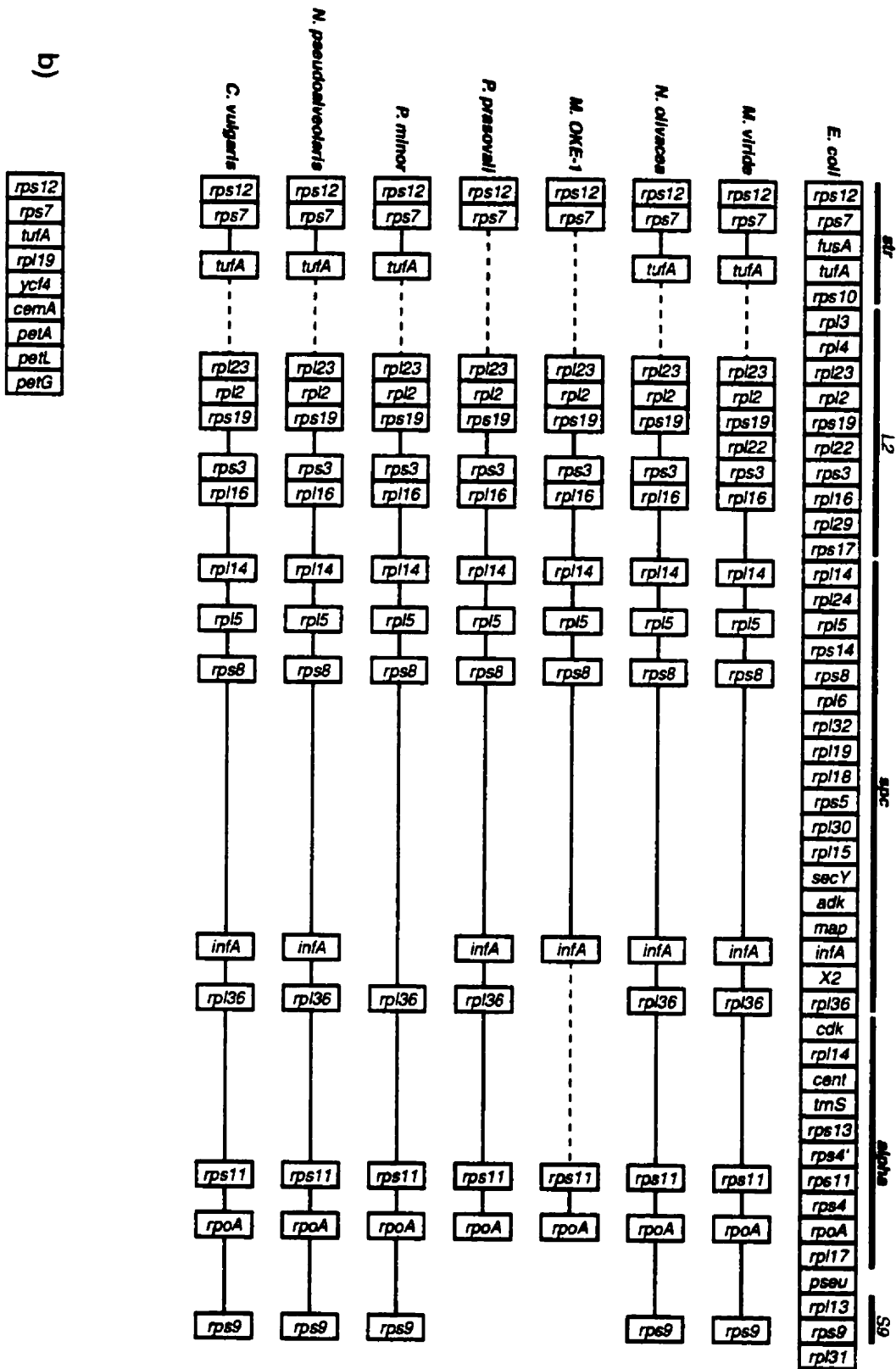
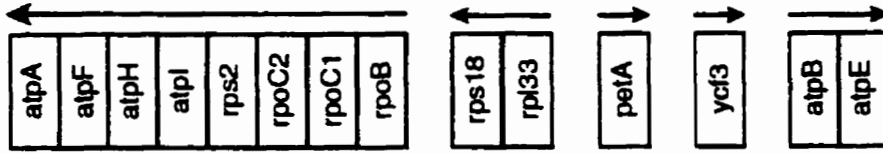
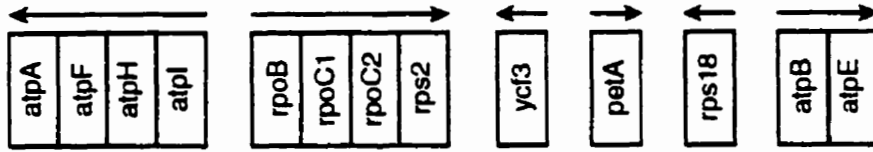


Figure 15. Vestiges du regroupement ancestral de gènes *atpA* chez *Monomastix* OKE-1 et six autres algues vertes. Les flèches indiquent la direction de transcription des cadres de lecture. Les blocs de gènes séparés sont ceux qui ne sont pas adjacents dans le génome. Significations des abréviations: *M. viride*, *Mesostigma viride*; *N. olivacea*, *Nephroselmis olivacea*; *M. OKE-1*, *Monomastix OKE-1*; *P. prasoali*, *Pycnococcus prasoali*; *P. minor*, *Pedinomonas minor*; *N. pseudoalveolaris*, *Neochloris pseudoalveolaris*; *C. vulgaris*, *Chlorella vulgaris*. Figure inspirée de Stoebe et Kowallik, 1999.

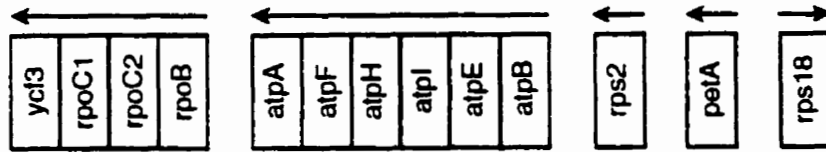
M. viride



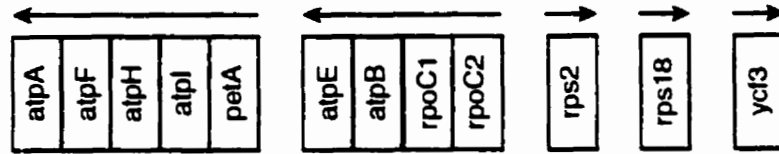
N. olivacea



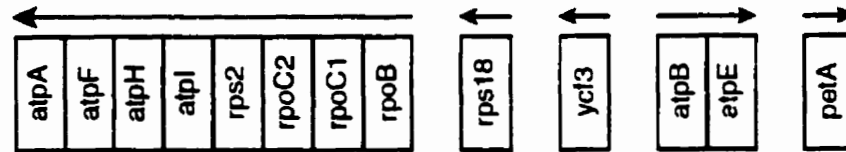
M. OKE-1



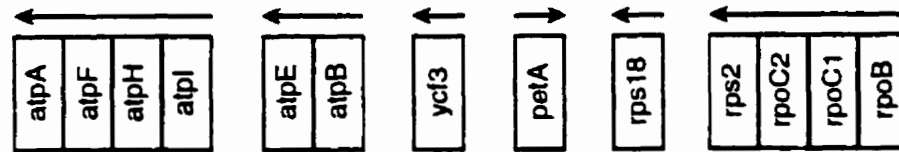
P. prasovali



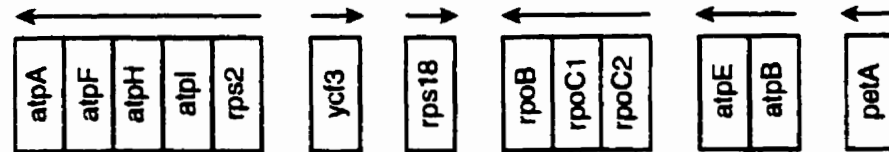
P. minor



N. pseudoalveolaris



C. vulgaris



3.1.5 Les introns

Le génome chloroplastique de *Monomastix* possède 6 introns : cinq du groupe I et un du groupe II (tableau VI). Ces introns sont inclus dans trois gènes différents, soit les gènes *rrl*, *psbA* et *trnK*(uuu). Le tableau X présente les caractéristiques générales de ces introns. L'emplacement exact de l'intron du gène *trnK* dans la structure secondaire de l'ARNt est montré en figure 16. La structure secondaire de cet intron (figure 17b) est typique aux introns du groupe IIB (Michel et Westhof, 1990) et il ne contient pas de cadre de lecture ouvert.

La structure secondaire des introns du groupe I est présentée en figure 17a (*psbA*·1) et figure 18 (*rrl*·1, *rrl*·2, *rrl*·3 et *rrl*·4). L'intron du gène *psbA* de *Monomastix* (figure 17a) adopte la même structure secondaire typique aux introns du groupe I que le troisième intron du gène *psbA* de *Chlamydomonas reinhardtii* (Holloway *et al.*, 1999) (*Cr. psbA*·3). Il est situé entre les acides aminés 174 et 175 et contient les domaines P9.1, P9.2 et P12 qui permettent de le classer dans le sous-groupe IA2 (Michel et Westhof, 1990) (tableau X), tout comme celui de *C. reinhardtii*. Il y a environ 63% d'identité de nucléotides entre les régions d'appariements conservées de *psbA*·1 de *Monomastix* et de *Cr. psbA*·3. La structure secondaire mais aussi primaire de l'intron *psbA* de *Monomastix* présente aussi des similarités avec l'intron du gène *sun Y* du bactériophage *T4* (Shub *et al.*, 1988). L'intron *sun Y* est classé dans le sous-groupe IA2 et possède un domaine P12 identique aux deux autres introns. La seule différence dans la structure secondaire de ces trois introns est l'absence de domaine P2 dans l'intron du bactériophage. Il y a environ 54% d'identité de nucléotides entre les régions d'appariements conservées de *psbA*·1 de *Monomastix* et de l'intron *sun Y*. L'ORF de l'intron *psbA* de *Monomastix* et celui de *C. reinhardtii* sont insérés dans la même boucle mais celui de *Monomastix* est plus long et ne contient pas le motif GIY-YIG (Holloway *et al.*, 1999). L'ORF de *Monomastix* présente un peu plus de similarités avec l'ORF du bactériophage mais il n'est pas inséré dans la même boucle (Shub *et al.*, 1988).

La structure secondaire des quatre introns du gène *rrl* (figure 18) est typique aux introns du groupe I. Le sous-groupe (tableau X) de chacun des introns du groupe I correspond à un des sous-groupes de Michel et Westhof, 1990. Le site d'insertion des introns est indiqué selon leur

Tableau X. Caractéristiques des introns présents dans le génome chloroplastique de *Monomastix*.

Introns	Groupe ^a	Emplacement	ORFs	
			Longueur (aa ^d)	Emplacement ^e
<i>psbA</i> -1	IA2	codon 175 (site 3) ^b	220	L6
<i>rri</i> -1	IB4	1931-1932 ^c	139	L8
<i>rri</i> -2	IA3	1951-1952 ^c	160	L8
<i>rri</i> -3	IB4	2500-2501 ^c	247	après P9
<i>rri</i> -4	IA3	2593-2594 ^c	171	L6
<i>trnK</i> (uuu)-1	IIB	bras D de l'hélice	— ^f	— ^f

^a Classé selon Michel et Westhof, 1990.

^b Correspond au troisième intron de *Chlamydomonas reinhardtii*.

^c Correspond aux positions (en nucléotides) du gène de l'ARNr 23S chez *Escherichia coli*.

^d aa, acides aminés.

^e La position des ORFs (ou cadres de lecture ouverts) est indiquée selon leur emplacement dans la structure secondaire de l'intron. P indique une région d'appariement; L suivi d'un chiffre désigne la boucle reliée à la région d'appariement numérotée.

^f Les traits horizontaux indiquent l'absence d'ORF.

Figure 16. Emplacement de l'intron dans la structure secondaire de l'ARNt codé par le gène *trnK(uuu)*. L'intron se situe entre les nucléotides 23 et 24. Les appariements entre nucléotides de type Watson-Crick (A:U, G:C) sont indiqués par (—). Les appariements G:U sont indiqués par un cercle plein.

Figure 17. Structures secondaires des introns dans les gènes *psbA* et *trnK(uuu)*. Les bases en lettres minuscules font partie des exons. Les appariements entre nucléotides de type Watson-Crick (A:U, G:C) sont indiqués par (—). Les appariements G:U sont indiqués par un cercle plein et les appariements purine : purine (G:A) sont indiqués par un cercle vide. a) Structure secondaire de l'intron du groupe I retrouvé dans le gène *psbA* de *Monomastix*. Les flèches indiquent les sites d'épissages 5' et 3'. Les astérisques montrent les codons d'initiation et de terminaison de l'ORF. P signifie une région d'appariement et L une boucle. Les boîtes noires indiquent les nucléotides identiques entre l'intron *psbA* de *Monomastix*, l'intron *psbA-3* de *Chlamydomonas reinhardtii* et l'intron du gène *sun Y* du bactériophage *T4*. Les boîtes blanches indiquent les nucléotides identiques seulement entre l'intron *psbA* de *Monomastix* et l'intron *psbA-3* de *C. reinhardtii* tandis que les boîtes grises représentent les nucléotides identiques seulement entre l'intron *psbA* de *Monomastix* et l'intron du gène *sun Y* du bactériophage *T4*. b) Structure secondaire de l'intron du groupe II de *trnK(uuu)*. Chaque domaine est indiqué par un chiffre romain. La jonction entre l'intron et l'exon est indiquée par des flèches.

Figure 18. Structures secondaires des quatre introns dans le gène *rrl*. Les bases en lettres minuscules font partie des exons. Les flèches montrent les sites d'épissages 5' et 3'. Les appariements entre nucléotides de type Watson-Crick (A:U, G:C) sont indiqués par (—). Les appariements G:U sont indiqués par un cercle plein et les appariements purine : purine (G:A) sont indiqués par un cercle vide. Les astérisques montrent les codons d'initiation et de terminaison de l'ORF. P signifie une région d'appariement et L une boucle. a) Intron *rrl*-1. Les boîtes noires indiquent les nucléotides identiques entre l'intron *rrl*-1 de *Monomastix*, l'intron *rnl*-2 de *Monomastix* et l'intron *rnl*-1 de *Mesostigma viride*. Les boîtes blanches indiquent les nucléotides identiques seulement entre les introns de *Monomastix* et les boîtes grises représentent les nucléotides identiques seulement entre l'intron *rrl*-1 de *Monomastix* et l'intron *rnl*-1 de *Mesostigma viride*. b) Intron *rrl*-2. c) Intron *rrl*-3. Les boîtes noires indiquent les nucléotides identiques entre l'intron *rrl*-3 de *Monomastix*, l'intron *rnl*-3 de *Monomastix* et l'intron *rrl*-1 de *Chlamydomonas humicola*. Les boîtes blanches indiquent les nucléotides identiques seulement entre les introns de *Monomastix* et les boîtes grises représentent les nucléotides identiques seulement entre l'intron *rrl*-3 de *Monomastix* et l'intron *rrl*-1 de *C. humicola*. d) Intron *rrl*-4. Les boîtes noires indiquent les nucléotides identiques entre l'intron *rrl*-4 de *Monomastix*, l'intron *rnl*-4 de *Monomastix* et l'intron *rnl*-2 de *Mesostigma viride*. Les boîtes blanches indiquent les nucléotides identiques seulement entre les introns de *Monomastix* et les boîtes grises représentent les nucléotides identiques seulement entre l'intron *rrl*-4 de *Monomastix* et l'intron *rnl*-2 de *Mesostigma viride*.

emplacement par rapport aux introns du gène *rri* trouvés chez *E. coli* (tableau XI). Les introns aux positions 1931 et 2593 présentent des similarités avec les introns trouvés aux mêmes positions dans le gène mitochondrial *rnl* de *Monomastix* OKE-1, de *N. olivacea* et de *M. viride* (figure 18). La structure secondaire de l'intron *rri*-1 de *Monomastix* (position 1931) est similaire aux introns *rnl*-2 de *Monomastix*, *rnl*-1 de *M. viride* et *rnl*-1 de *N. olivacea*. La seule différence dans la structure secondaire de ces introns est l'absence de domaine P9.0 chez l'intron *rri*-1 de *Monomastix* et l'intron *rnl*-1 de *M. viride*. Il y a environ 60% d'identité de nucléotides entre les régions d'appariements conservées de l'intron *rri*-1 de *Monomastix* et de l'intron *rnl*-2 de *Monomastix* et 66% d'identité entre l'intron *rri*-1 de *Monomastix* et les introns *rnl*-1 de *M. viride* et de *N. olivacea*. Il y a aussi un peu plus de similarités entre la séquence en acides aminés de l'ORF de l'intron *rri*-1 de *Monomastix* et celle de l'ORF de l'intron *rnl*-1 de *M. viride* qu'entre celle des deux introns de *Monomastix*. L'intron chloroplastique de *Monomastix* à la position 2593 (*rri*-4) a une structure secondaire similaire aux introns trouvés à la même position dans l'ADN mitochondrial de *Monomastix* OKE-1 (*rnl*-4), *M. viride* (*rnl*-2) et *N. olivacea* (*rnl*-2) (figure 18d). Il y a environ 67% d'identité de nucléotides entre les régions d'appariements conservées de l'intron *rri*-4 de *Monomastix* et l'intron *rnl*-4 de *Monomastix*, 62% d'identité entre l'intron *rri*-4 de *Monomastix* et l'intron *rnl*-2 de *M. viride* ainsi que 57% d'identité entre l'intron *rri*-4 de *Monomastix* et l'intron *rnl*-2 de *N. olivacea*. Les séquences en acides aminés des ORFs des deux introns de *Monomastix* sont aussi légèrement plus similaires que celle de l'ORF de l'intron *rri*-4 de *Monomastix* l'est avec celle de l'ORF de l'intron de *M. viride* (*rnl*-2).

Le troisième intron du gène *rri* chloroplastique de *Monomastix*, soit celui inséré à la même position que l'intron à la position 2500 du gène de l'ARNr 23S de *E. coli*, est similaire au troisième intron du gène *rnl* mitochondrial de *Monomastix* ainsi qu'à l'intron du gène *rri* chloroplastique de *Chlamydomonas humicola*, qui sont tous deux insérés à la position 2500. Il y a environ 57% d'identité de nucléotides entre les régions d'appariements de l'intron *rri*-3 de *Monomastix* et l'intron du gène *rri* de *C. humicola* (figure 18c). Ce pourcentage est d'environ 71% entre les deux introns de *Monomastix* (figure 18c).

Les orf139, 160 et 171, appartenant respectivement aux introns *rrl-1*, *rrl-2* et *rrl-4*, codent pour des ADN endonucléases potentielles de type LAGLIDADG. Ces protéines ont été nommées Mso1931c (orf139), Mso1951c (orf160) et I-MsoI (orf171). La seule de ces protéines dont la fonction d'ADN endonucléase LAGLIDADG a été démontrée est I-MsoI (Lucas *et al.*, 2001).

3.1.6 Les séquences répétées

Le génome chloroplastique de *Monomastix* est criblé de séquences répétées. Les séquences répertoriées ont une longueur variant entre 19 et 58 nucléotides et elles sont répétées entre 8 et 116 fois (tableau XI). Comme plusieurs séquences différentes sont répétées, elles ont été classées en six familles (A, A1, B, B1, C et D) et six familles doubles (AB, AC, AD, A1D, A1B, BD), c'est-à-dire contenant la séquence propre à deux familles différentes. Entre eux, les membres d'une même famille peuvent avoir entre zéro et une dizaine de bases différentes. Pour être classée dans une famille, chaque séquence devait avoir le patron typique aux séquences de cette famille. Le tableau XI présente cette classification. Ces séquences ont été trouvées sur les deux brins de l'ADN. Les familles A1 et B1 ont été nommées ainsi car elles contiennent respectivement la séquence des familles A et B mais elles sont prolongées par un bout de séquence n'appartenant à aucune autre famille. Les familles A, B et AB forment des structures en épingle à cheveu.

Les séquences répétées sont dispersées dans tout le génome. La figure 19 présente la distribution de ces séquences dans le génome. Elles sont surtout retrouvées dans les régions intergéniques mais les gènes *rpoC1*, *rpoC2*, *rpoB*, *ftsH*, *psbA* (dans la boucle L8 de l'intron, figure 17a) et *rrl* (dans la boucle 7.1 du quatrième intron, figure 18d) en contiennent aussi. Tous les ORFs uniques et non introniques de plus de 60 codons sont situés dans des régions riches en séquences répétées. Dans les régions intergéniques, les séquences répétées sont regroupées et selon des patrons différents d'une région à l'autre. Les séquences de la famille C ainsi que celles de la famille D sont presque toujours répétées en direct. Sur la figure 19, chaque trait horizontal indique une région de séquences répétées et chaque famille est

Tableau XI. Classification et caractérisation des séquences répétées trouvées dans le génome chloroplastique de *Monomastix*.

Familles	Séquences ^{b,c}	Longueur (nu ^d)	Nombre de répétitions
A	TGAGCGCTTTGACCTTTSAGGTCAKAHATHAAA*GCTSA	38	61
A1	AGCACACCAGACCTCCCGGTCCTTTSAGGTCAKAHATHAAA*GCTSA	45	20
B	CTY*CCCCCTTCGGGGGR*AG	19	72
B1	CTTCCCCCTTCGGGGGAAGGGGG	23	25
C	GCTSAAKCACACCAGGTCAAAGM	23	116
D	ACCTCTGGTCCTTTGGTCCTTGAG	24	43
Combinaisons de familles^a			
AB	CTCAAGGGGGTGAGCGCTTTGACCTTTSAGGTCAKAHATHAAA*GCTSACCCCCTTGAG	58	18
AC	AGGTCATATATTAAGCTCAATCACACCAGACCTTTSAGGTCAKAHATHAAA*GCTSA	57	9
AD	AGRCCTTTGGCCCTACGGTCCTTTSAGGTCAKAHATHAAA*GCTSA	45	26
A1D	TTTGAGACCTTTGGTCCTCCCGGTCCTTTSAGGTCAKAHATHAAA*GCTSA	50	8
A1B	AGCACACCAGCCTCCCGGTCCTTTCAGGTCTAAATAAAGCTCACCCCCTTGAG	55	1
BD	AGAGCTGGGGGGCCCCTTCAGRCCTTYGGYCCTC	34	70

^a Les familles de séquences répétées contenant la séquence de deux familles sont nommées par les lettres des deux familles correspondantes.

^b Les lettres en caractère gras et avec un astérisques sont des bases qui ne sont pas présentes dans toutes les séquences d'une famille.

^c D'après le code IUB : R = A ou G; Y = C ou T; K = G ou T; M = A ou C; S = G ou C.

^d nu, nucléotides

Figure 19. Localisation des séquences répétées sur la carte chloroplastique linéaire des gènes de *Monomastix*. Chaque famille ou famille double est représentée par la couleur suivante : A = ■, A1 = ■, B = ■, B1 = □, C = ■, D = ■, AB = ■, AC = ■, AD = ■, A1D = ■, A1B = ■. BD = ■. Les rectangles noirs représentent les gènes, les exons ainsi que les ORFs introniques. Les gènes orientés vers l'extérieur de la figure sont transcrits dans le sens horaire; ceux orientés vers l'intérieur sont transcrits dans le sens anti-horaire. Les ARNt sont indiqués par le code à une lettre des acides aminés suivi de l'anticodon entre parenthèses. Les exons des gènes contenant des introns sont indiqués par ex. L'extrémité supérieure de la section gauche correspond à la position 0 de la figure 1 et l'extrémité inférieure de la section droite correspond à la position 114 528.

représentée par un rectangle de couleur. Dans les régions contenant plusieurs séquences répétées, les rectangles de couleur appartenant à la famille de chaque séquence sont empilés.

La figure 20 présente chaque famille de séquences répétées en relation les unes avec les autres. Les couleurs et les flèches permettent de suivre les portions de séquences identiques entre les familles ou semblant venir d'une même séquence. Aucune des familles répertoriées semble ne pas avoir de lien avec au moins une autre famille. Les familles A, B et D sont celles qui ont le plus d'homologie avec les autres familles et qui ainsi ont le plus de chances d'être les mères de toutes les séquences répétées.

3.1.7 Les analyses phylogénétiques

La position phylogénétique de *Monomastix* déduite à l'aide de la séquence en acides aminées concaténée de protéines chloroplastiques est montrée en figure 21. La première analyse phylogénétique (figure 21a) présente *Monomastix* dans un groupe de prasinophytes formé par lui-même et *Pycnococcus prasovali*. Selon cette analyse, l'emplacement phylogénétique de *Monomastix* n'est pas plus basal que celui de *Nephroselmis olivacea*. La valeur de « bootstrap » de cet embranchement, ainsi que celle regroupant tous les chlorophytes à l'exception de *N. olivacea* et celle regroupant *C. vulgaris* et *Pedinomonas minor*, sont légèrement plus basses que celle des autres embranchements. L'embranchement des streptophytes ainsi que l'emplacement de *M. viride* à la base des streptophytes et chlorophytes sont supportés par des valeurs de « bootstrap » élevées. Il est aussi à remarquer, dans cet arbre, que *P. minor* est groupé avec les autres trebouxiophytes avec un niveau de confiance élevé.

La longueur de la branche de *Monomastix* est similaire à celle de *N. olivacea*, qui est légèrement plus longue que celle des autres plantes vertes à l'exception de l'embranchement menant à *P. prasovali*. La branche de ce dernier est d'ailleurs beaucoup plus longue que celle des autres plantes vertes utilisées dans cette première analyse. La deuxième analyse (figure 21b), dans laquelle *P. prasovali* a été exclus, a permis d'obtenir des valeurs de « bootstrap » plus élevées pour l'embranchement regroupant tous les chlorophytes à l'exception de *N. olivacea* ainsi que pour le groupe formé par *C. vulgaris* et *P. minor*. La branche de

Figure 20. Liens hypothétiques entre les différentes familles de séquences répétées. Chaque cercle de couleur entourant la lettre propre à une famille représente la ou les couleur(s) de la famille respective. Les couleurs utilisées sont les mêmes que celles de la figure 19. Les portions de séquence en couleur appartiennent à la famille représentée par cette couleur. Les flèches indiquent les séquences identiques ou similaires entre deux familles.

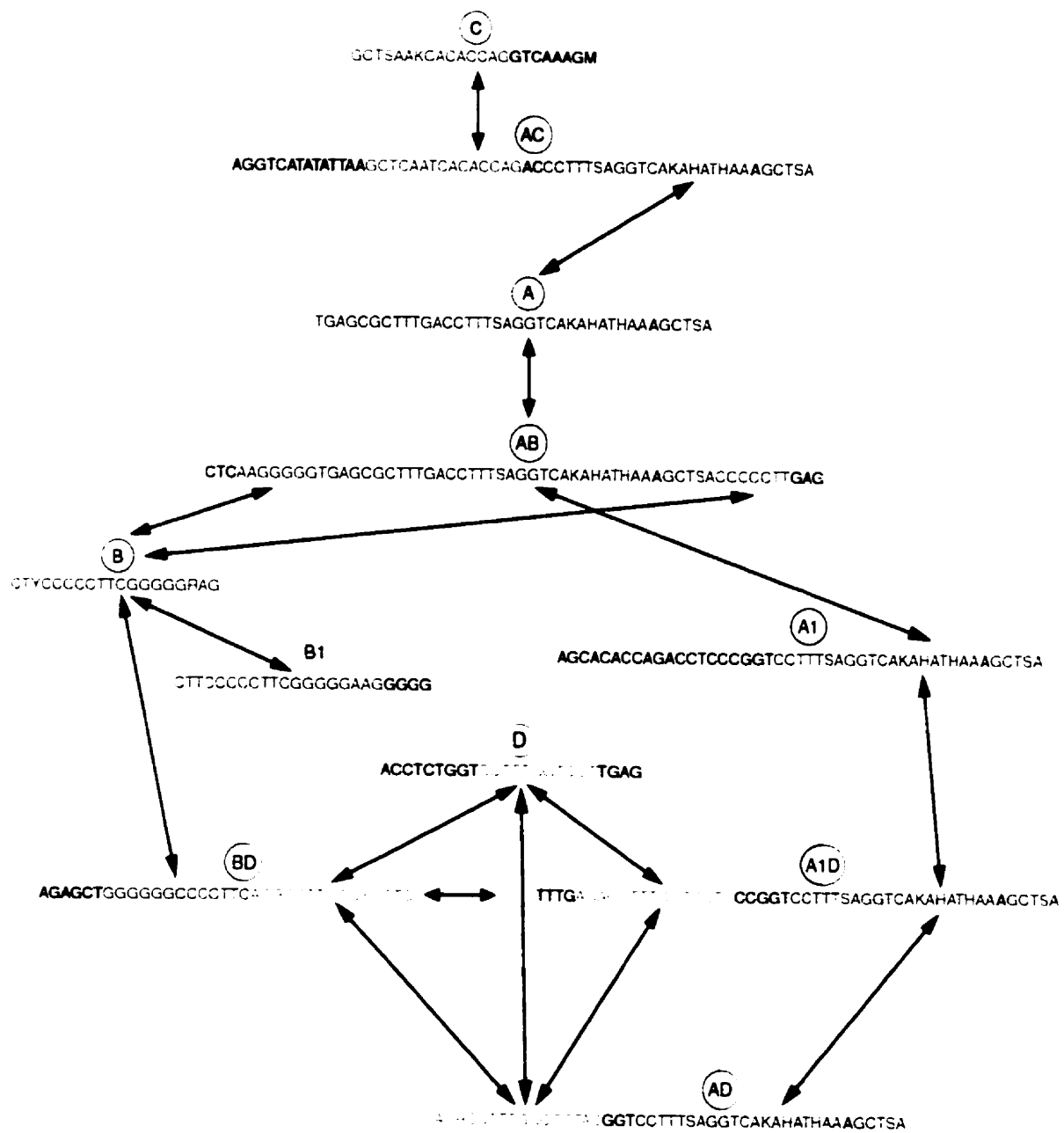
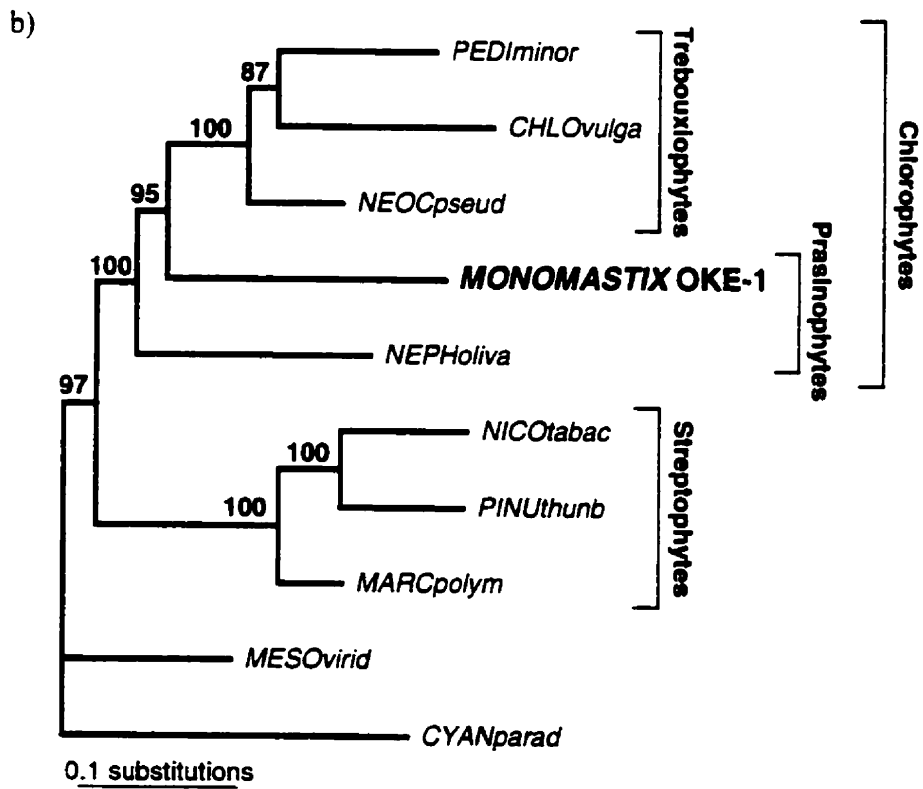
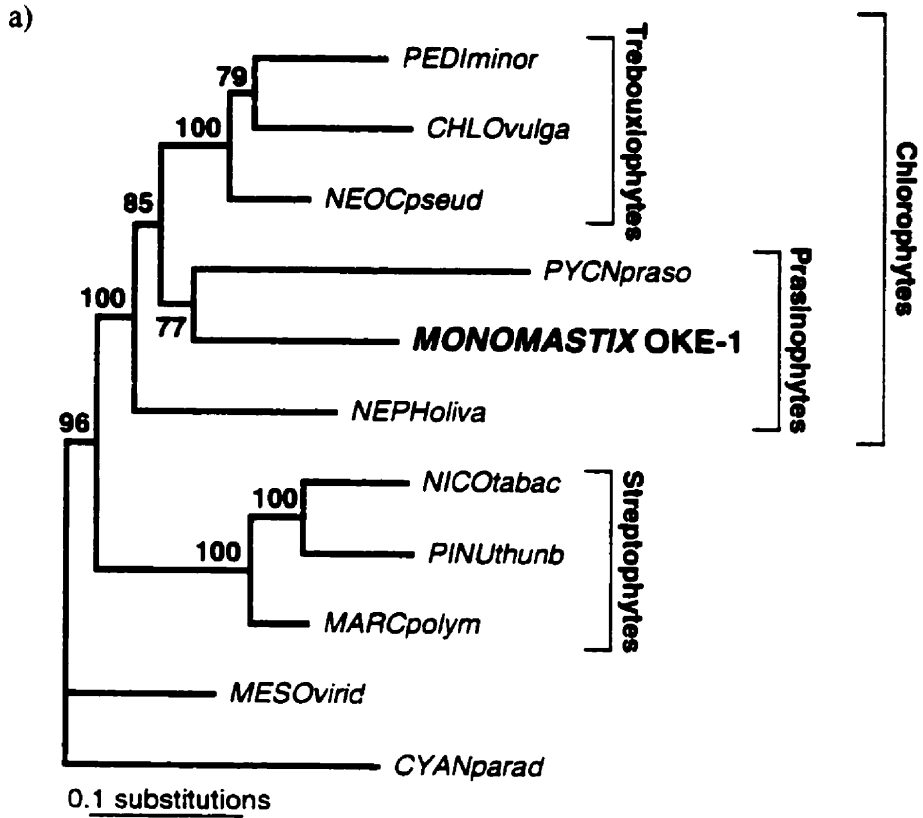


Figure 21. Position phylogénétique de *Monomastix* déduite à l'aide de la séquence en acides aminés de protéines chloroplastiques concaténées. Les valeurs de « bootstrap » sont indiquées à gauche de chaque embranchement. La figure montre, pour chaque analyse, le meilleur arbre Maximum Likelihood fait avec le logiciel AAML et le modèle cpREV45 avec distribution gamma. *Cyanophora paradoxa* ainsi que *Mesostigma viride* sont utilisés pour enraciner l'arbre. L'échelle sous chaque figure indique le nombre de substitutions par nucléotide pour une longueur de branche donnée. Signification des abréviations: *CYANparad*, *Cyanophora paradoxa*; *MESOvirid*, *Mesostigma viride*; *NEPHoliva*, *Nephroselmis olivacea*; *PYCNpraso*, *Pycnococcus prasoali*; *NEOCpseud*, *Neochloris pseudoalveolaris*; *CHLOvulga*, *Chlorella vulgaris*; *PEDIminor*, *Pedinomonas minor*; *MARCpolym*, *Marchantia polymorpha*; *PINUthunb*, *Pinus thunbergii*; *NICOtobac*, *Nicotiana tobacum*. a) Position phylogénétique de *Monomastix* basée sur la séquence en acides aminés de 47 protéines totalisant 8544 sites. 11 taxa sont utilisés pour l'analyse, incluant *Pycnococcus prasoali*. b) Position phylogénétique de *Monomastix* basée sur la séquence en acides aminés de 51 protéines totalisant 9523 sites. Les mêmes taxa que pour la première analyse sont utilisés à l'exception de *Pycnococcus prasoali*.



Monomastix est un peu plus longue que celle de *N. olivacea* dans cette analyse mais son emplacement phylogénétique est toujours le même.

3.2 Le génome mitochondrial

3.2.1 Structure et caractéristiques générales

La figure 22 présente la carte circulaire des gènes du génome mitochondrial de *Monomastix* OKE-1. La séquence complète de ce génome sous format GenBank se retrouve en annexe 2 et sera éventuellement déposée dans cette banque de données.

L'ADN mitochondrial de cette algue verte s'assemble en un cercle de 60 424 pb avec un contenu en bases A + T de 59,9%. Le tableau XII résume les caractéristiques de ce génome. Environ 70% de cette séquence code pour 65 gènes ayant une fonction connue. Deux de ces gènes, *rnl* et *cox1*, sont entrecoupés chacun par quatre introns du groupe I. Le pourcentage en bases A + T de la région codante (62,1%) est sensiblement le même que celui de la séquence en entier (59,9%) mais celui de la région non-codante est un peu plus bas (54,7%). Des cadres de lecture ouverts de plus de 65 acides aminés mais n'ayant ni fonction connue ni homologue chez d'autres organismes (16 en tout) ont été répertoriés. Les régions intergéniques de ce génome ont une longueur variant entre 1 et 3367 nucléotides. Tous les ORFs uniques, à l'exception de l'orf69b et de l'orf76, sont situés dans les régions intergéniques de plus de 500 nucléotides.

3.2.2 Contenu en gènes

Le contenu détaillé des 65 gènes et 16 ORFs uniques de l'ADNmt est répertorié dans le tableau XIII. Tous les gènes et ORFs uniques sont présents en une seule copie. Parmi les 16 ORFs uniques à *Monomastix*, seulement cinq ont plus de 100 acides aminés (orf101-103-106-110-127) et un seul a plus de 200 acides aminés (orf202).

La moitié des gènes codant pour des protéines dans l'ADNmt de *Monomastix* sont des gènes impliqués dans la chaîne respiratoire (phosphorylation oxydative et transport d'électrons). Parmi ces gènes, dix codent pour des sous-unités de la NADH déshydrogénase, un code pour l'apocytochrome b (sous-unité du cytochrome c : ubiquinol oxydoréductase), trois codent pour

Figure 22. Génome mitochondrial de *Monomastix* OKE-1. Les gènes à l'extérieur de la carte sont transcrits dans le sens horaire; ceux situés à l'intérieur de la carte sont transcrits dans le sens antihoraire. Les ARNt sont indiqués par le code à une lettre des acides aminés suivi de l'anticodon entre parenthèses. Tous les ORFs uniques à l'ADNmt de *Monomastix* de plus de 65 acides aminés sont indiqués. Pour les gènes contenant des introns, les portions noires indiquent les exons, les portions blanches les introns et les grises, les ORFs introniques.

Tableau XII. Résumé des caractéristiques du génome mitochondrial de *Monomastix* OKE-1.

Traits	Génome mitochondrial
Longueur (pb) ^a	60 424
Contenu en gènes	
total	65
protéines	
phosphorylation	
oxydative/transport	18
d'électrons	
ribosomiques	15
ORFs conservés non	
identifiés	2
ARNr	3
ARNt	26
Autres ARNs	1
ORFs introniques	7
ORFs uniques ^b	16
Introns	
groupe I	8
groupe II	0
% codant ^c	69,5
% A + T	
total	59,9
codant ^c	62,1
non-codant	54,7

^a pb, paire de bases.

^b Les ORFs de plus de 65 acides aminés uniques à l'ADNmt de *Monomastix* et n'ayant pas de fonction connue sont notés. Ils ne sont pas calculés dans le contenu en gènes et n'incluent par les ORFs introniques.

^c Représente le pourcentage du génome qui code pour les gènes de protéines, d'ARNt, d'ARNr, d'autres ARNs ainsi que pour les introns et les ORFs introniques.

Tableau XIII. Gènes codés par l'ADNmt de *Monomastix* OKE-1.

Produits des gènes	Gènes
NADH :ubiquinone oxydoreductase	<i>nad1, 2, 3, 4, 4L, 5, 6, 7, 9, 10</i>
Ubiquinone :cytochrome c oxydoreductase	<i>cob</i>
Cytochrome c oxydase	<i>cox1, 2, 3</i>
ATP synthétase	<i>atp1, 6, 8, 9</i>
Protéines de la grande sous-unité ribosomale	<i>rpl5, 6, 14, 16</i>
Protéines de la petite sous-unité ribosomale	<i>rps2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 19</i>
ORFs conservés de fonction inconnue	<i>ymf16, 39</i>
ORFs uniques ^a	<i>orf67, 69a, 69b, 71, 72, 73, 76, 85, 92, 95, 101, 103, 106, 110, 127, 202</i>
ARNs ribosomiques	<i>rrn5, rnl, rns</i>
ARNs de transfert ^b	<i>A(ugc), C(gca), D(guc), E(uuc), F(gaa), G(ucc), H(gug), I(cau), I(gau), K(uuu), L(gag), L(uaa), L(uag), Me(cau), Mf(cau), N(guu), P(ugg), Q(uug), R(acg), R(ucu), S(gcu), S(uga), T(ggu), V(uac), W(cca), Y(gua)</i>
Autre petit ARN	<i>rnpB</i>

^a Les ORFs uniques de plus de 65 codons et de fonction inconnue sont notés.

^b Les ARNt sont indiqués par le code à une lettre des acides aminés suivi de l'anticodon entre parenthèses.

des sous-unités du cytochrome c oxydase et quatre codent pour des sous-unités de l'ATP synthétase. Les autres protéines codées par ce génome mitochondrial sont les protéines ribosomiques (quatre protéines de la grande sous-unité et 11 de la petite) et deux protéines de fonction inconnue. Ce génome contient aussi trois gènes codant pour des ARNs ribosomiques, 26 gènes d'ARNt ainsi que le gène codant pour l'ARN de l'enzyme RNaseP. La structure secondaire de la petite sous-unité de l'ARN ribosomique ainsi que celle de l'ARN composant l'enzyme RNaseP sont présentées en figure 23. Les structures secondaires des extrémités 3' et 5' des gènes *rns* et *rnl* ont aussi été construites afin de s'assurer de la localisation de ces gènes. Puisque ces structures ne sont pas complètes, elles ne sont pas présentées.

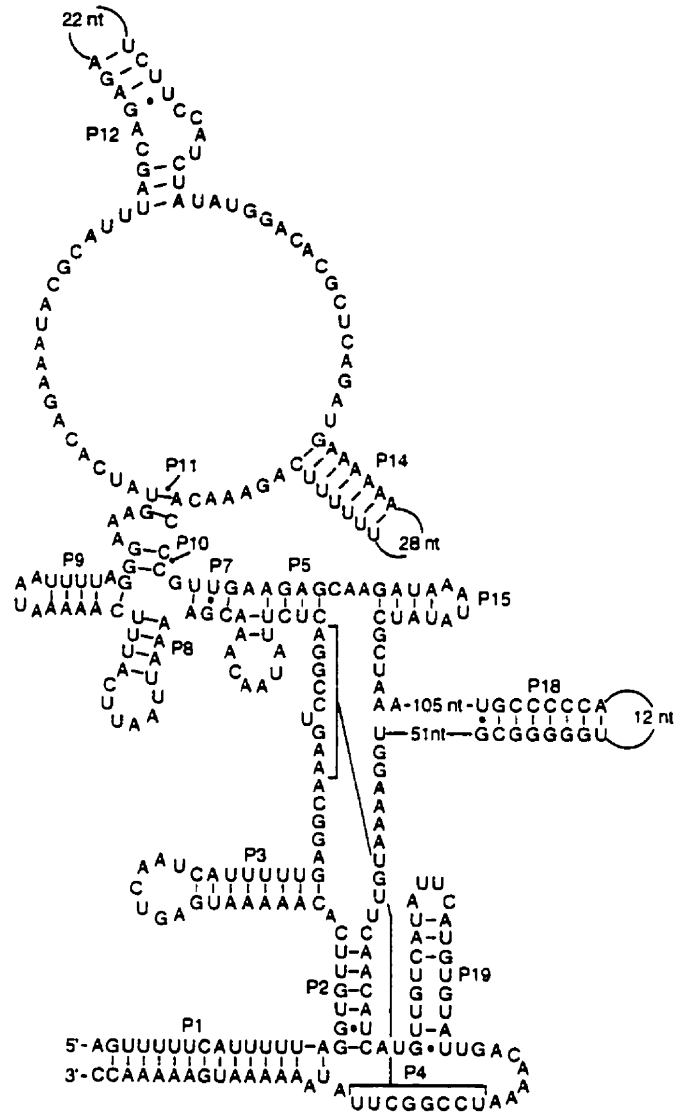
3.2.3 Les ARNs de transfert et la composition des codons

Les 26 ARNt de l'ADNmt sont dispersés dans tout le génome, seuls ou en groupe de deux ou trois. La structure secondaire de chacun des ARNt (figure 24) est conforme à la structure standard en forme de trèfle à quatre feuille d'un ARNt (figure 5). Quelques petites déviations à cette structure ont cependant pu être observées. D'abord, les ARNt^{Ala}(UGC), ARNt^{Met}(CAU) initiateur, ARNt^{Ser}(GCU), ARNt^{Ser}(UGA), ARNt^{Thr}(GGU) et ARNt^{Val}(UAC) ont un appariement de bases de moins dans le bras accepteur. Le même type de modification est présent dans le bras de l'anticodon de l'ARNt^{Cys}(GCA), dans le bras D de l'ARNt^{Ser}(UGA) et des trois ARNt^{Leu} ainsi que dans le bras pyrimidine de l'ARNt^{Lys}(UUU). La présence d'une purine en position 11 et d'une pyrimidine en position 24 alors que la purine devrait être en position 24 et la pyrimidine en position 11, a été observée chez les ARNt^{Gly}(UCC) et ARNt^{Met}(CAU) initiateur. Les autres traits non typiques aux ARNt sont les suivants : un G versus une pyrimidine en position 48 de l'ARNt^{His}(GUG), un G versus un A en position 21 de l'ARNt^{Ile}(CAU), un U versus un A en position 14 de l'ARNt^{Gln}(UUG), un C versus un A en position 14 de l'ARNt^{Trp}(UUG), un A versus une pyrimidine en position 32 de l'ARNt^{Thr}(GGU) ainsi qu'un A versus une pyrimidine en position 32 de l'ARNt^{Met}(GGU) élongateur. Cette dernière déviation est trouvée fréquemment dans l'ARNt élongateur mitochondrial des protistes (Burger *et al.*, 1999).

Le code génétique universel est utilisé dans le génome mitochondrial de *Monomastix*. Le

Figure 23. Structure secondaire des ARNs codés par les gènes mitochondriaux *rnpB* et *rrn5*. Les appariements entre nucléotides de type Watson-Crick (A:U, G:C) sont indiqués par (—) et les appariements G:U sont indiqués par un cercle plein. a) Structure secondaire de l'ARN codé par le gène *rnpB* et composant l'enzyme RNaseP. Chaque région d'appariement est indiquée par la lettre P suivi d'un chiffre. b) Structure secondaire de l'ARNr 5S codé par le gène *rrn5*. Chaque région d'appariement est indiquée par un chiffre romain et chaque région variable est identifiée par une lettre.

a)



b)

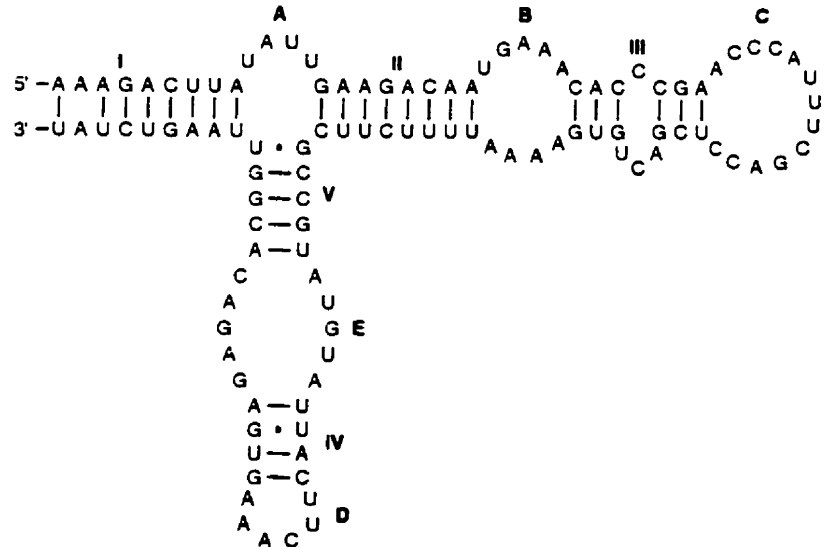


Figure 24. Structure secondaire des ARNs de transfert du génome mitochondrial. Les appariements entre nucléotides de type Watson-Crick (A:U, G:C) sont indiqués par (—). Les appariements G:U sont indiqués par un cercle plein et les appariements purine : purine (G:A) sont indiqués par un cercle vide. Pour chaque ARNt, le gène le codant est indiqué. Le code à une lettre des acides aminés est utilisé. L'anticodon est en lettres minuscules. Le gène *trnMe(cau)* code pour l'ARNt^{met} élongateur et le gène *trnMf(cau)* code pour l'ARNt^{met} initiateur.

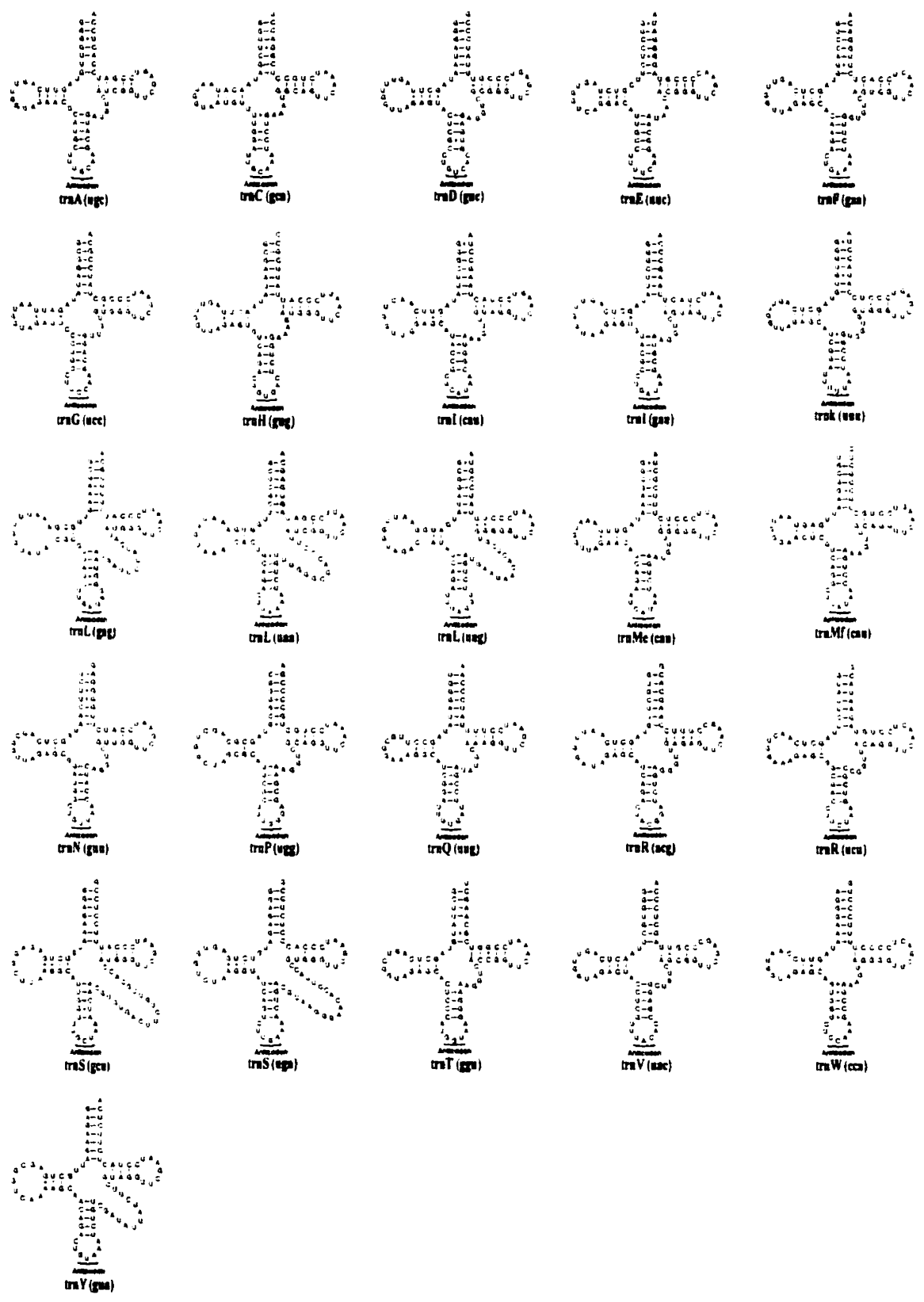


tableau XIV présente l'utilisation des codons de l'ADNmt en relation avec les ARNt présents. Tous les codons sont utilisés, incluant les trois codons de terminaison UAA, UAG et UGA. Les 26 ARNt de l'ADNmt ne peuvent décoder que 59 des 61 codons du code génétique puisque l'ARNt^{Thr}(GGU) n'est pas suffisant pour décoder les quatre codons de l'acide aminé thréonine. L'anticodon GGU ne peut que se lier avec les codons ACU et ACC car le G à la première position de l'anticodon ne peut être dégénéré. L'ARNt^{Thr}(UGU) est manquant. Six cas de dégénérescence (« superwobble ») du U à la première position de l'anticodon sont observés, soit pour les acides aminés valine, sérine, proline, alanine, arginine et glycine. Un léger biais d'utilisation des bases A et U versus les bases G et C en troisième position des codons est observé. Le ratio d'utilisation des bases A + U en troisième position des codons versus les bases G + C est de 3,0: 1, celui de la purine A versus la purine G est de 3,8: 1 et celui de la pyrimidine U versus la pyrimidine C est de 2,6: 1. On peut donc observer que les codons les moins utilisés sont ceux se terminant par un G ou un C.

3.2.4 Organisation des gènes

Les gènes de ce génome sont distribués non uniformément sur les deux brins (figure 22). Sur une première moitié du génome, tous les gènes se situent sur le même brin, suggérant la présence d'une seule unité de transcription. Sur la seconde moitié du génome, la distribution des gènes sur les deux brins prédit la présence d'au moins 11 unités de transcription.

Plusieurs regroupements de gènes retrouvés dans l'ADNmt de *Monomastix* sont aussi présents dans l'ADNmt ancestral de *Reclinomonas americana*, *Nephroselmis olivacea* ou *Prototheca wickerhamii*. Le tableau XV présente ces regroupements, à l'exception du regroupement des protéines ribosomiques, illustré en figure 25. Le génome ayant le plus de regroupements similaires à *Monomastix* est celui de *N. olivacea*. Seul le regroupement des ARNs ribosomiques diverge significativement entre ces deux algues vertes. Seulement deux regroupements présents chez *R. americana* sont retrouvés chez *Monomastix*. Les regroupements de *P. wickerhamii* sont seulement des vestiges de ceux de *N. olivacea* et *Monomastix*.

Tableau XIV. Utilisation de codons chez 35 gènes conservés ainsi que pour les sept ORFs introniques de l'ADNmt de *Monomastix* OKE-1.

AA ^a	Codon	Anticodon ^b	% ^c	AA ^a	Codon	Anticodon ^b	% ^c	AA ^a	Codon	Anticodon ^b	% ^c	AA ^a	Codon	Anticodon ^b	% ^c
F	UUU	gaa	76	S	UCU	uga	28	Y	UAU	gua	75	C	UGU	gca	71
F	UUC	gaa	24	S	UCC	uga	7	Y	UAC	gua	25	C	UGC	gca	29
L	UUA	uaa	36	S	UCA	uga	28	*	UAA		49	*	UGA		39
L	UUG	uaa	7	S	UCG	uga	9	*	UAG		12	W	UGG	cca	100
L	CUU	gag	33	P	CCU	ugg	44	H	CAU	gug	70	R	CGU	acg ^d	22
L	CUC	gag	13	P	CCC	ugg	9	H	CAC	gug	30	R	CGC	acg ^d	11
L	CUA	uag	4	P	CCA	ugg	37	Q	CAA	uug	83	R	CGA	acg ^d	26
L	CUG	uag	7	P	CCG	ugg	10	Q	CAG	uug	17	R	CGG	acg ^d	3
I	AUU	gau	56	T	ACU	ggu	34	N	AAU	guu	67	S	AGU	gcu	19
I	AUC	gau	27	T	ACC	ggu	18	N	AAC	guu	33	S	AGC	gcu	9
I	AUA	cau	16	T	ACA	—	42	K	AAA	uuu	85	R	AGA	ucu	30
M	AUG	cau ^c	100	T	ACG	—	6	K	AAG	uuu	15	R	AGG	ucu	8
V	GUU	uac	34	A	GCU	ugc	42	D	GAU	guc	73	G	GGU	ucc	27
V	GUC	uac	15	A	GCC	ugc	10	D	GAC	guc	27	G	GGC	ucc	7
V	GUA	uac	25	A	GCA	ugc	36	E	GAA	uuc	84	G	GGA	ucc	52
V	GUG	uac	26	A	GCG	ugc	12	E	GAG	uuc	16	G	GGG	ucc	14

^a Acides aminés (code à une lettre). Les astérisques indiquent les codons de terminaisons.

^b Les anticodons sont en lettres minuscules. Les traits horizontaux indiquent les gènes d'ARNt qui ne sont pas présents dans l'ADN mitochondrial de *Monomastix*. Des cas de dégénérescence en première position de l'anticodon sont observés pour les ARNt avec un U en première position de l'anticodon. Cela leur permet de décoder tous les membres des familles de quatre codons. (*i.e.*, valine, GUN; sérine UCN; proline, CCN; alanine, GCN; arginine, CGN; glycine, GGN).

^c Pourcentage de chaque acide amine spécifié par un codon donné dans l'ADN mitochondrial de *Monomastix*.

^d Le A à la première position de l'anticodon est converti post-transcriptionnellement en un I (inosine), résultant en un ARNt^{Arg} (ICG) capable de décoder les quatre codons arginine CGN (Pfitzinger *et al.*, 1990).

^e Un ARNt^{Mct} élongateur (e) ainsi qu'un ARNt^{Mct} initiateur (f, formyl) séparés sont présents.

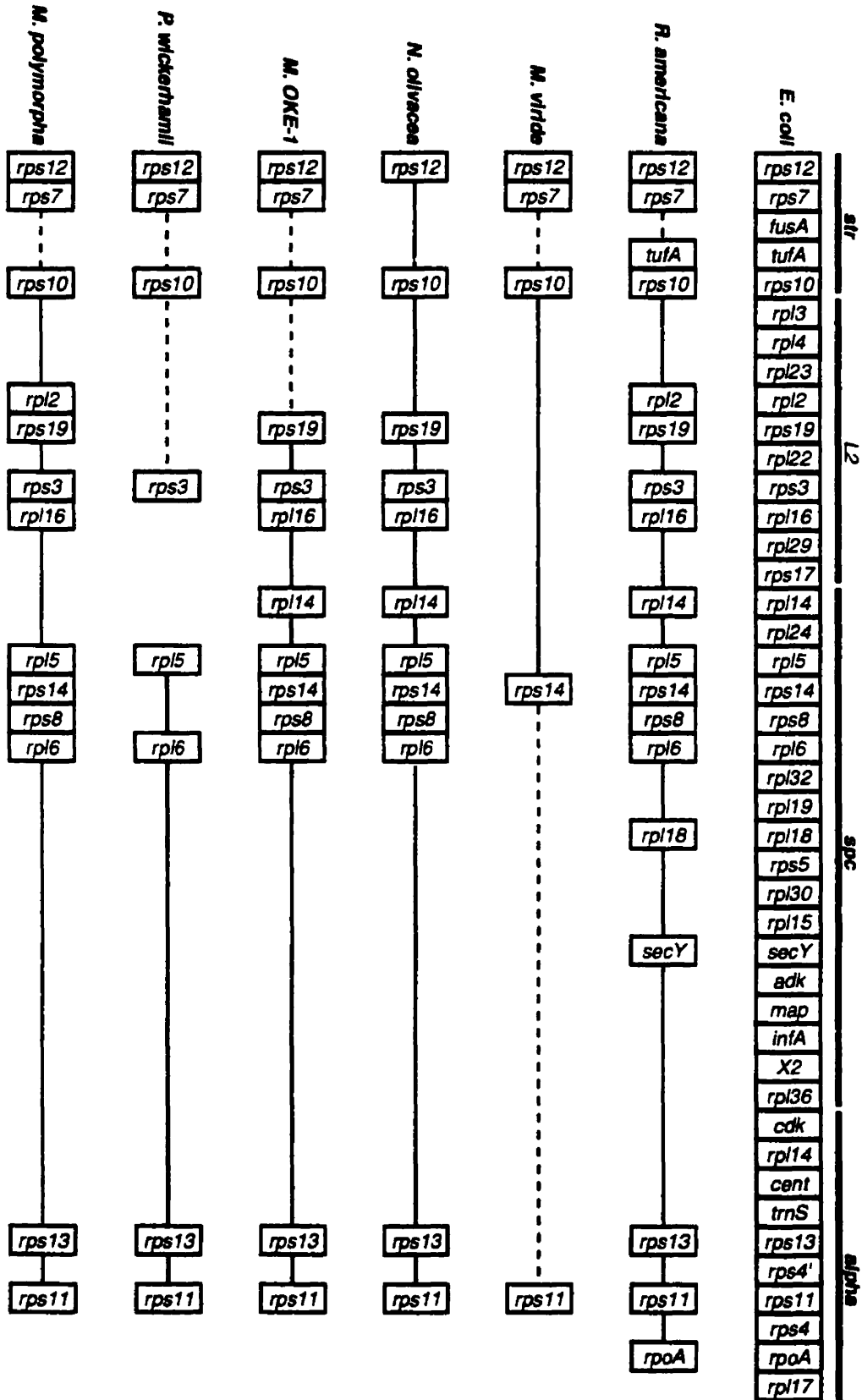
Tableau XV. Regroupement de gènes présents dans l'ADNmt de *Monomastix* OKE-1 ainsi que chez deux algues vertes ayant un génome mitochondrial de type ancestral et un protiste faisant partie des Jakobids.

Protiste ^a	Algues vertes avec un génome mitochondrial ancestral ^a		
<i>R.am</i>	<i>N.ol</i>	<i>M.OK</i>	<i>P.wi</i>
—	<i>cox3-cox2-N(guu)</i> ^b	<i>cox3-cox2-N(guu)</i>	<i>cox3-cox2</i>
<i>nad2-nad4-nad5</i>	<i>nad2-nad4-nad5- F(gaa)-Q(uug)- rps2</i>	<i>nad2-nad4-nad5- F(gaa)-rps2- Q(uug)</i>	<i>nad4-nad5</i>
—	<i>ymf39-atp8-nad4L- P(ugg)-R(acg)- rnpB</i>	<i>ymf39-atp8- nad4L-P(ugg)- R(ucu)-rnpB</i>	<i>ymf39-atp8</i>
<i>nad10-nad9</i>	<i>nad10-nad9</i>	<i>nad10-nad9</i>	—
<i>rrn5-rns-rnl</i>	<i>rrn5-S(uga)- L(uag)-I(gau)- Me(cau)-K(uuu)- Mf(cau)-rns-rnl</i>	<i>rns-rrn5</i>	<i>rns-E(uuc)- H(gug)-rrn5</i>

^a *R.am*, *Reclinomonas americana*; *N.ol*, *Nephroselmis olivacea*; *M.OK*, *Monomastix* OKE-1; *P.wi*, *Prototheca wickerhamii*.

^b Les ARNt sont notés selon le code à une lettre des acides aminés suivi de l'anticodon entre parenthèses.

Figure 25. Regroupements de gènes similaires aux opérons *str-L2-spc-alpha* de *E. coli* dans l'ADNmt de *Monomastix* OKE-1 et d'autres ADNmt. Les lignes pleines rejoignent les gènes adjacents; les lignes pointillées indiquent la présence de gènes additionnels non montrés. Significations des abréviations: *E. coli*, *Escherichia coli*; *R. americana*, *Reclinomonas americana*; *M. viride*, *Mesostigma viride*; *N. olivacea*, *Nephroselmis olivacea*; *M. OKE-1*, *Monomastix OKE-1*; *P. wickerhamii*, *Prototheca wickerhamii*; *M. polymorpha*, *Marchantia polymorpha*. Figure inspirée de Turmel *et al.*, 1999.



Les gènes codant pour les protéines ribosomiques de l'ADNmt de *Monomastix* sont distribués dans un ordre suivant celui des gènes des opérons *str-L2-spc-alpha* (figure 25). Cette organisation est très similaire à celle retrouvée chez *R. americana*, *N. olivacea* et *M. polymorpha*. Elle varie un peu chez *P. wickerhamii* mais elle n'est pas conservée chez *M. viride*.

3.2.5 Les introns

Le génome mitochondrial de *Monomastix* possède huit introns (tableau XII et XVI) inclus dans deux gènes différents, soient les gènes *rnl* et *cox1*. La structure secondaire des introns du gène *rnl* est présentée en figure 26 et celle des introns du gène *cox1* en figure 27. Elle a permis de confirmer que ces introns faisaient partie du groupe I et de leur attribuer un sous-groupe (tableau XVI) selon Michel et Westhof, 1990. Seul *rnl-1* ne contient pas de cadre de lecture ouvert.

Comme déjà mentionné dans la section 3.1.5, les introns mitochondriaux du gène *rnl* aux positions 1931 (*rnl-2*) et 2593 (*rnl-4*) présentent des similarités avec des introns situés aux mêmes positions et trouvés entre autres dans le gène chloroplastique *rnl* de *Monomastix* OKE-1 et dans le gène mitochondrial *rnl* de *N. olivacea* et de *M. viride* (figure 26). Il y a environ 60% d'identité de nucléotides entre les régions d'appariements conservées des introns *rnl-1* et *rnl-2* de *Monomastix* et 59% d'identité entre l'intron *rnl-2* de *Monomastix* et les introns *rnl-1* de *M. viride* et de *N. olivacea*. Il y a aussi environ autant de similarités entre la séquence en acides aminés de l'ORF de l'intron *rnl-2* de *Monomastix* et celle de l'ORF de l'intron *rnl-1* de *M. viride* qu'entre celle des deux introns de *Monomastix*. Environ 67% d'identité de nucléotides entre les régions d'appariements conservées des introns *rnl-4* et *rnl-4* de *Monomastix*, 65% d'identité entre l'intron *rnl-4* de *Monomastix* et l'intron *rnl-2* de *M. viride* et 63% d'identité entre l'intron *rnl-4* de *Monomastix* et l'intron *rnl-2* de *N. olivacea* ont été calculés. Environ autant de similarité est observée entre la séquence en acides aminés de l'ORF de l'intron *rnl-4* de *Monomastix* et celle de l'ORF de l'intron de *M. viride* (*rnl-2*) qu'entre les deux introns de *Monomastix*.

Tableau XVI. Caractéristiques des introns présents dans le génome mitochondrial de *Monomastix* OKE-1.

Introns	Groupe ^a	Emplacement	ORFs	
			Longueur (aa ^d)	Emplacement ^c
<i>rnl-1</i>	IB4	1922-1923 ^b	— ^f	— ^f
<i>rnl-2</i>	IB4	1931-1932 ^b	202	L8
<i>rnl-3</i>	IB4	2500-2501 ^b	271	après P9
<i>rnl-4</i>	IA3	2593-2594 ^b	174	L6
<i>cox1-1</i>	IB2	132-133 ^c	353	L8
<i>cox1-2</i>	IB2	208-209 ^c	273	L8
<i>cox1-3</i>	IB2	292-293 ^c	304	L8
<i>cox1-4</i>	IB3	372-373 ^c	317	L1

^a Classé selon Michel et Westhof, 1990.

^b Correspond aux positions (en nucléotides) du gène *rnl* chez *Escherischia coli*.

^c Correspond aux positions (en acides aminés) du gène *cox1* chez *Monomastix* OKE-1.

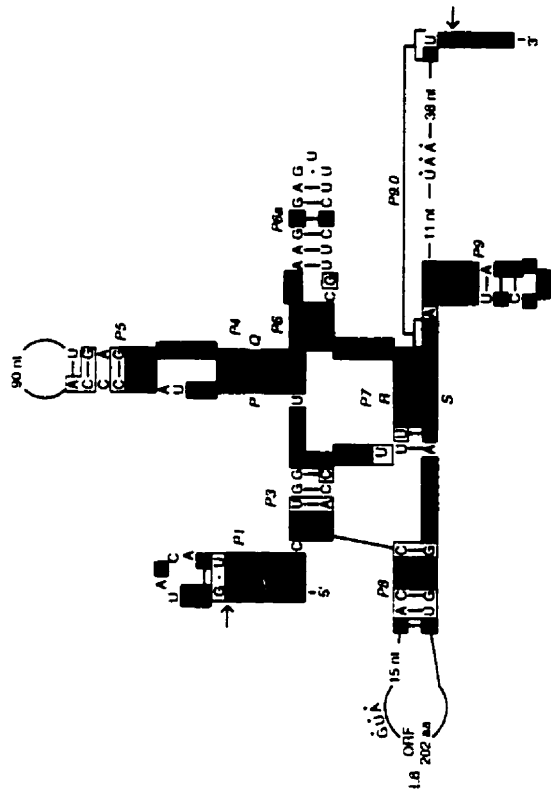
^d aa signifie acides aminés.

^e La position des ORFs (ou cadres de lecture ouverts) est indiquée selon leur emplacement dans la structure secondaire de l'intron. P indique une région d'appariement; L suivi d'un chiffre désigne la boucle reliée à la région d'appariement numérotée.

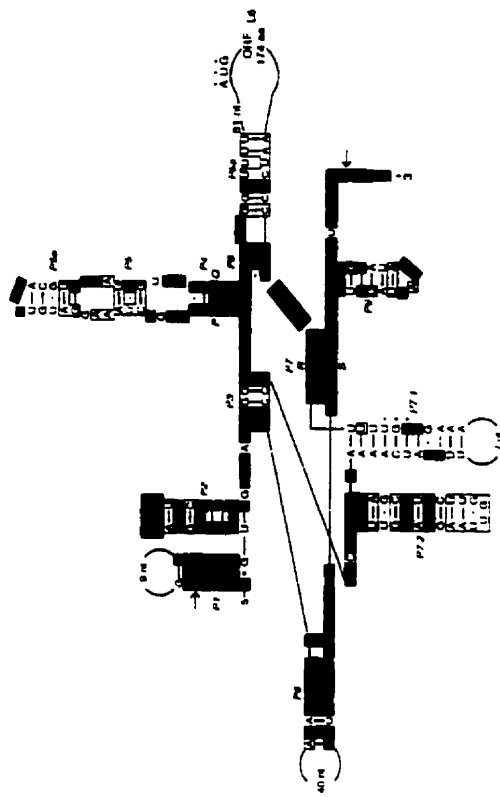
^f Les traits horizontaux indiquent l'absence d'ORF.

Figure 26. Structures secondaires des quatre introns dans le gène *rnl*. Les bases en lettres minuscules font partie des exons. Les flèches montrent les sites d'épissages 5' et 3'. Les appariements entre nucléotides de type Watson-Crick (A:U, G:C) sont indiqués par (—). Les appariements G:U sont indiqués par un cercle plein et les appariements purine : purine (G:A) sont indiqués par un cercle vide. Les astérisques montrent les codons d'initiation et de terminaison de l'ORF. a) Intron *rnl*-1. b) Intron *rnl*-2. Les boîtes noires indiquent les nucléotides identiques entre l'intron *rnl*-2 de *Monomastix*, l'intron *rrl*-1 de *Monomastix* et l'intron *rnl*-1 de *Mesostigma viride*. Les boîtes blanches indiquent les nucléotides identiques seulement entre les introns de *Monomastix* et les boîtes grises représentent les nucléotides identiques seulement entre l'intron *rnl*-2 de *Monomastix* et l'intron *rnl*-1 de *Mesostigma viride*. c) Intron *rnl*-3. Les boîtes noires indiquent les nucléotides identiques entre l'intron *rnl*-3 de *Monomastix*, l'intron *rrl*-3 de *Monomastix* et l'intron *rrl*-1 de *Chlamydomonas humicola*. Les boîtes blanches indiquent les nucléotides identiques seulement entre les introns de *Monomastix* et les boîtes grises représentent les nucléotides identiques seulement entre l'intron *rnl*-3 de *Monomastix* et l'intron *rrl*-1 de *Chlamydomonas humicola*. d) Intron *rnl*-4. Les boîtes noires indiquent les nucléotides identiques entre l'intron *rnl*-4 de *Monomastix*, l'intron *rrl*-4 de *Monomastix* et l'intron *rnl*-2 de *Mesostigma viride*. Les boîtes blanches indiquent les nucléotides identiques seulement entre les introns de *Monomastix* et les boîtes grises représentent les nucléotides identiques seulement entre l'intron *rnl*-4 de *Monomastix* et l'intron *rnl*-2 de *Mesostigma viride*.

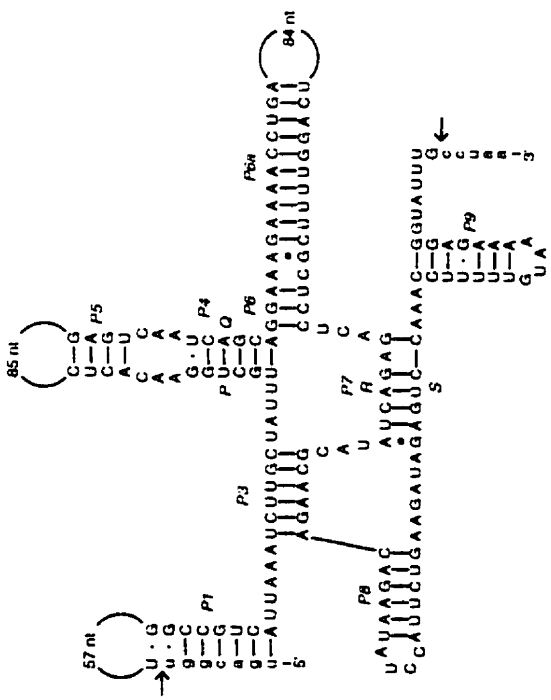
b)



d)



a)



c)

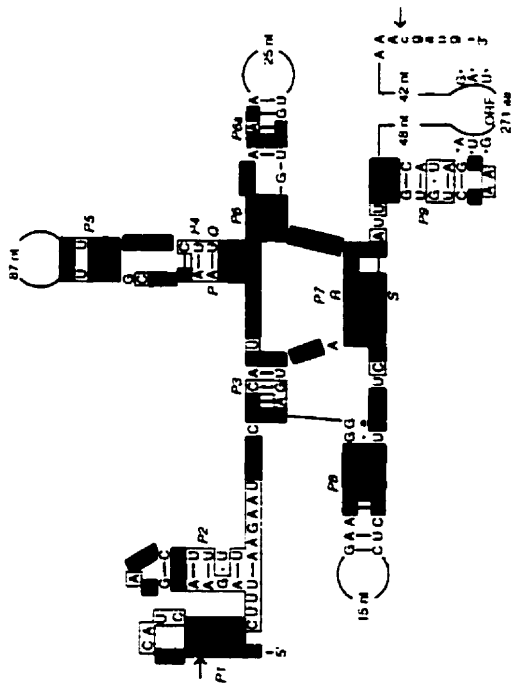


Figure 27. Structures secondaires des quatre introns dans le gène *cox1*. Les bases en lettres minuscules font partie des exons. Les flèches montrent les sites d'épissages 5' et 3'. Les appariements entre nucléotides de type Watson-Crick (A:U, G:C) sont indiqués par (—). Les appariements G:U sont indiqués par un cercle plein et les appariements purine : purine (G:A) sont indiqués par un cercle vide. Les astérisques montrent les codons d'initiation et de terminaison de l'ORF. a) Intron *cox1*·1. b) Intron *cox1*·2. c) Intron *cox1*·3. Les boîtes blanches indiquent les nucléotides identiques entre l'intron *cox1*·3 de *Monomastix* et l'intron *cox1*·1 de *M. viride*. d) Intron *cox1*·4. Les boîtes blanches indiquent les nucléotides identiques entre l'intron *cox1*·4 de *Monomastix* et l'intron *cox1*·2 de *M. viride*.

Le troisième intron du gène *rnl* mitochondrial de *Monomastix*, soit celui inséré à la position 2500 de *E. coli*, est similaire au troisième intron du gène *rnl* chloroplastique de *Monomastix* ainsi qu'à l'intron du gène *rnl* chloroplastique de *Chlamydomonas humicola*, qui sont tous deux insérés à la position 2500. Il y a environ 55% d'identité de nucléotides entre les régions d'appariements de l'intron *rnl-3* de *Monomastix* et l'intron du gène *rnl* de *C. humicola* (figure 26c). Ce pourcentage est d'environ 75% entre les deux introns de *Monomastix* (figure 26c).

Les introns *cox1-3* et *cox1-4* sont insérés au même site que les introns *cox1-1* et *cox1-2* respectivement de *Mesostigma viride*. Un pourcentage d'identité de nucléotides entre les régions d'appariements conservées de 60% a été calculé entre l'intron *cox1-3* de *Monomastix* et l'intron *cox1-1* de *M. viride*. La structure secondaire de ces introns est similaire (figure 27) mais l'intron *cox1-3* de *Monomastix* contient les domaines P2 et P9.1 qui sont absents chez l'intron *cox1-1* de *M. viride*. L'intron *cox1-4* de *Monomastix* est moins similaire à l'intron *cox1-2* de *Mesostigma viride*; l'intron de *Mesostigma* contient les domaines P3.1 et P3.2 qui sont absents de l'intron de *Monomastix* et il y a seulement 46% d'identité de nucléotides entre les régions d'appariements conservées de ces deux introns (figure 27).

L'orf202 de l'intron *rnl-2* a été identifié comme une ADN endonucléase potentielle de type LAGLIDADG (Lucas *et al.*, 2001). Cependant, il n'a pas été montré que la protéine Mso1931m (orf202) est bien une endonucléase de type LAGLIDADG.

3.2.6 Les régions intergéniques

Les régions intergéniques de l'ADNmt de *Monomastix* ont une longueur variant entre 1 et 3367 nucléotides. Leur longueur moyenne varie entre 20 et 300 nucléotides. Ces régions totalisent 19 204 nucléotides, soit environ 30% de la taille totale du génome et elles sont plus riches en G + C (45,3%) que les régions codantes (37,9%). La plupart des régions intergéniques de plus de 60 nucléotides contiennent des éléments riches en G + C. Ces éléments, qui ne semblent pas être répétés, peuvent prendre plusieurs formes; ils peuvent se replier sur eux-mêmes et former des structures en épingles à cheveu, ils peuvent être répétés en tandem ou être simplement de courtes séquences de G ou de C. Les éléments les plus

fréquemment retrouvés sont les structures en épingles à cheveu. Des éléments riches en G + C sont présents dans la plupart des ORFs non-introniques et uniques à *Monomastix*, soient les orf67, 69a, 69b, 71, 72, 73, 76, 85, 92, 95, 103, 106 et 127. Seuls les orf101, 110 et 202 semblent exempts de ces éléments.

3.2.7 Les analyses phylogénétiques

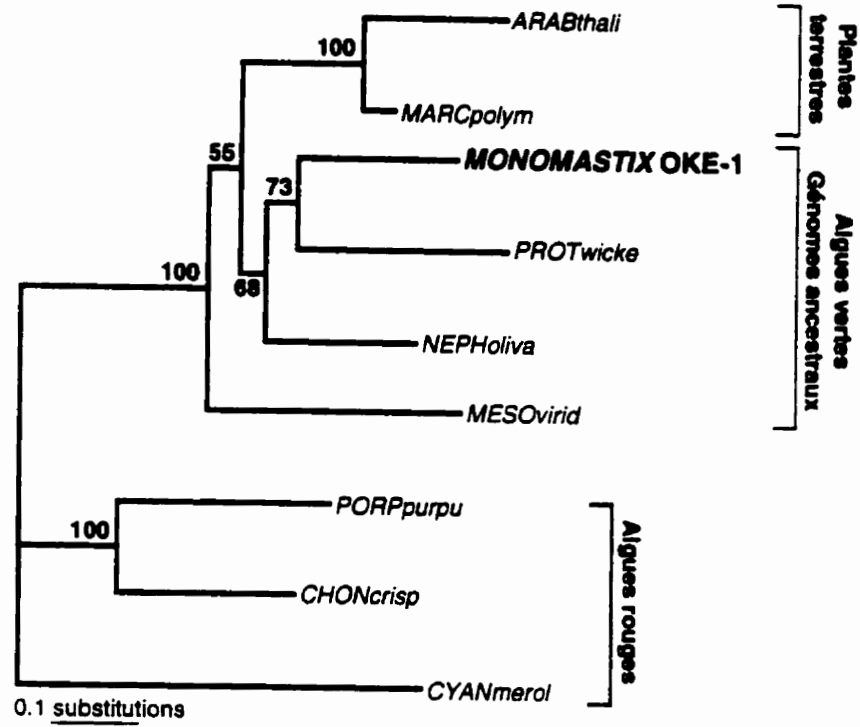
La figure 28 présente la position phylogénétique de *Monomastix* déduite à l'aide de séquences en acides aminées concaténées de protéines mitochondriales. La figure 28a montre l'emplacement de *Monomastix* parmi un groupe d'algues vertes de la lignée Chlorophyta dont le génome mitochondrial est dit ancestral. Les trois classes représentées ici (algues rouges, streptophytes et chlorophytes) sont très bien définies. Le groupe des algues rouges ainsi que celui des streptophytes sont supportés par des « bootstrap » élevés de 100. Les algues rouges sont représentées comme un groupe sœur aux plantes vertes.

Cette analyse place *M. viride* à la base des plantes vertes mais avec une valeur de « bootstrap » un peu faible. *Monomastix* est ici groupé avec le trebouxiophyte *P. wickerhamii*. Ce regroupement, ainsi que celui plaçant *N. olivacea* à la base de ce dernier, sont supportés un peu faiblement par des valeurs de « bootstrap » de 73 et 68 respectivement. La branche de *Monomastix* est d'une longueur similaire à celle des autres taxons représentés dans cette analyse à l'exception de *M. polymorpha* qui a une branche beaucoup plus courte. L'emplacement de *Monomastix* prévu par cette analyse est similaire à celui des analyses phylogénétiques chloroplastiques; il se retrouve avec les chlorophytes mais est moins basal que *Nephroselmis olivacea*.

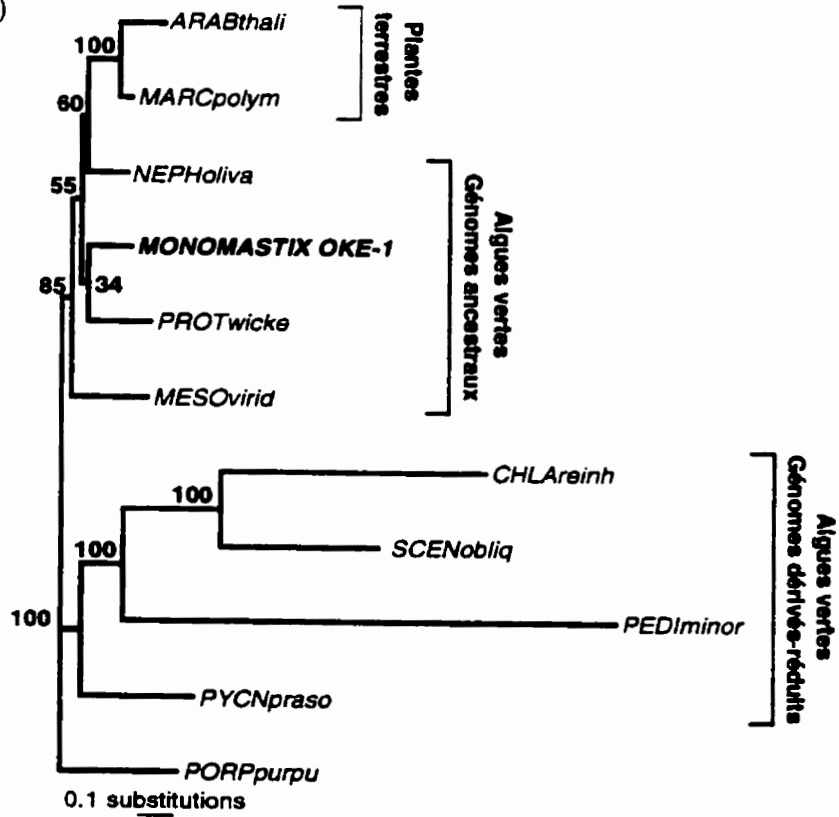
La figure 28b présente *Monomastix* parmi un échantillon de génome mitochondriaux d'algues vertes plus grand. Celui-ci inclus des algues ayant un génome mitochondrial dit ancestral (les prasinophytes *Mesostigma viride* et *Nephroselmis olivacea* et le trebouxiophyte *Prototheca wickerhamii*), intermédiaire (le chlorophyte *Scenedesmus obliquus*) ou dérivé-réduit (les prasinophytes *Pycnococcus prasoali* et *Pedinomonas minor* et le chlorophyte *Chlamydomonas reinhardtii*). Dans cette analyse, *M. viride* est toujours placé à la base des

Figure 28. Position phylogénétique de *Monomastix* déduite à l'aide de la séquence en acides aminés de protéines mitochondriales concaténées. Les valeurs de « bootstrap » sont indiquées à gauche de chaque embranchement. La figure montre, pour chaque analyse, le meilleur arbre Maximum Likelihood fait avec le logiciel AAML et le modèle JTT avec distribution gamma. L'échelle sous chaque figure indique le nombre de substitutions par nucléotide pour une longueur de branche donnée. Signification des abréviations du genre et de l'espèce de chaque taxon : *CYANmerol*, *Cyanidioschyzon merolae*; *PORPpurpu*, *Porphyra purpura*; *CHONcrisp*, *Chondrus crispus*; *MESOVirid*, *Mesostigma viride*; *NEPHoliva*, *Nephroselmis olivacea*; *PYCNpraso*, *Pycnococcus prasoali*; *PEDIminor*, *Pedinomonas minor*; *PROTwicke*, *Prototheca wickerhamii*; *SCENobliq*, *Scenedesmus obliquus*; *CHLAreinh*, *Chlamydomonas reinhardtii*; *MARCpolym*, *Marchantia polymorpha*; *ARABthala*, *Arabidopsis thaliana*. a) Position phylogénétique de *Monomastix* basée sur la séquence en acides aminés de 19 protéines totalisant 4135 sites. Neuf taxa ont été utilisés pour l'analyse. *Cyanidioschyzon merolae* a été utilisé pour enraciner l'arbre. b) Position phylogénétique de *Monomastix* basée sur la séquence en acides aminés de sept protéines totalisant 2221 sites. 11 taxa ont été utilisés pour l'analyse. *Porphyra purpura* a été utilisé pour enraciner l'arbre.

a)



b)



plantes vertes avec une faible valeur de « bootstrap ». Cependant, *N. olivacea* se retrouve anormalement à la base des plantes terrestres. *Monomastix* est toujours groupé avec *P. wickerhamii* parmi les génomes de type ancestral mais ce lien n'est supporté que par une faible valeur de « bootstrap ».

Une division entre les génomes mitochondriaux ancestraux d'algues vertes et les génomes dérivés-réduits ou intermédiaires est remarquée. Les génomes dérivés-réduits sont représentés comme un groupe sœur au regroupement incluant les génomes ancestraux ainsi que les plantes terrestres. Le regroupement des algues au génome dérivé-réduit est très bien supporté par les valeurs de « bootstrap ». Le prasinophyte *P. prasoali* est à la base de ce regroupement et *P. minor*, dont la branche est beaucoup plus longue que les autres, est à la base des algues vertes de la classe chlorophyceae. La longueur des branches du regroupement des génomes mitochondriaux dérivés-réduits est beaucoup plus longue que celle des génomes mitochondriaux ancestraux ainsi que des plantes terrestres, à l'exception de celle de *P. prasoali*, qui n'est que légèrement plus longue.

DISCUSSION

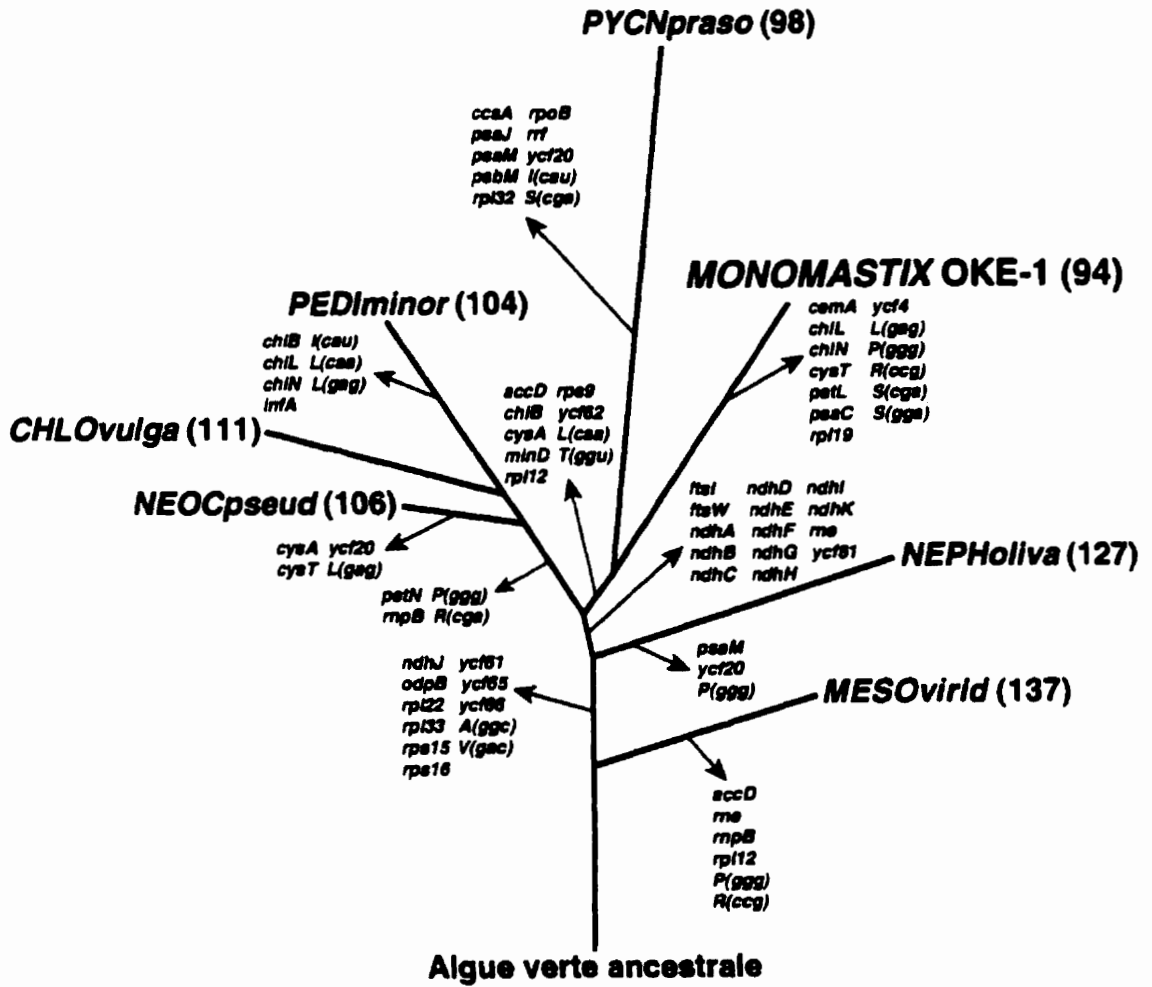
4.1 Le génome chloroplastique de *Monomastix* OKE-1 a perdu plusieurs caractéristiques ancestrales

L'étude du génome chloroplastique de *Monomastix* OKE-1 a permis de mettre en évidence la perte de plusieurs caractéristiques ancestrales retrouvées chez la plupart des prasinophytes étudiés jusqu'à maintenant.

D'abord, la structure d'ADN inversé répété, et par conséquent la structure quadripartite communément trouvée chez les génomes chloroplastiques (Palmer et Delwiche, 1998), n'est pas présente dans cet ADNcp. Parmi les génomes chloroplastiques d'algues vertes étudiés, celui du prasinophyte *Pycnococcus prasovali* ainsi que celui du trebouxiophyte *Chlorella vulgaris* ne possèdent pas non plus de structure d'ADN inversé répété.

Deuxièmement, la conservation du contenu en gènes ancestral du prasinophyte à la base des chlorophytes (*Nephroselmis olivacea*) n'a été suivie qu'en partie chez *Monomastix*. D'ailleurs, le génome chloroplastique de *Monomastix* est celui qui contient le moins de gènes parmi les génomes chloroplastiques d'algues vertes étudiés. La figure 29 présente un arbre montrant la position phylogénétique de *Monomastix*, de quatre autres prasinophytes (*Mesostigma viride*, *Nephroselmis olivacea*, *Pycnococcus prasovali* et *Pedinomonas minor*) et de deux trebouxiophytes (*Neochloris pseudoalveolaris* et *Chlorella vulgaris*) ainsi que les gènes perdus dans chaque embranchement. Seulement dix pertes de gènes séparent l'algue verte à la base de la lignée Streptophyta et de la lignée Chlorophyta (*M. viride*) et celle à la base de la lignée Chlorophyta (*N. olivacea*). Suite à l'embranchement de *N. olivacea*, les gènes de division *ftsI* et *ftsW*, le gène *rne* ainsi que tous les gènes *ndh* ont été perdus. Contrairement à l'embranchement menant aux trebouxiophytes *N. pseudoalveolaris*, *P. minor* (classé avec les trebouxiophytes pour les besoins de la cause) et *C. vulgaris*, beaucoup de gènes ont été perdus dans la branche menant aux prasinophytes *P. prasovali* et *Monomastix*. Parmi les gènes perdus par *Monomastix*, très peu sont des gènes impliqués dans la photosynthèse et plusieurs (*chlL*, *chlN*, *cysT*, *psaC*, *minD*, *ftsW* et les gènes *ndh*) font partie de la petite région unique (SSC) ou

Figure 29. Distribution des événements de perte de gènes chloroplastiques dans la lignée Chlorophyta. L'arbre de la figure 21a a été utilisé pour construire cette figure. La lignée Streptophyta a été exclue de l'arbre. Les nombres entre parenthèses à droite du nom des espèces indiquent le nombre de gènes présents dans l'ADNcp correspondant. Signification des abréviations du genre et de l'espèce de chaque taxon : *MESOvirid*, *Mesostigma viride*; *NEPHoliva*, *Nephroselmis olivacea*; *PYCNpraso*, *Pycnococcus prasovali*; *NEOCpseud*, *Neochloris pseudoalveolaris*; *CHLOvulga*, *Chlorella vulgaris*; *PEDIminor*, *Pedinomonas minor*. Figure inspirée de Turmel *et al.*, 1999a.



de la région d'ADN inversée répétée chez *N. olivacea* et *M. viride*. Cette observation suggère que la perte d'un segment d'ADN inversé répété ait pu entraîner, ou du moins avoir un lien, avec la perte de plusieurs gènes de la région SSC des génomes chloroplastiques ancestraux. Plusieurs gènes codant pour des ARNt et présents chez l'ADNcp des autres algues vertes sont aussi absents chez *Monomastix*. Quelques cas de dégénérescence des codons à la troisième base de l'anticodon semblent cependant suffisants pour pallier la perte de ces gènes. Le tableau IX confirme que tous les ARNt nécessaires pour décoder les 61 codons du code génétique sont présents. Il est aussi à noter que *Monomastix*, ainsi que *Pycnococcus* mais dans une proportion moins importante, ont perdu un nombre considérable de gènes chloroplastiques comparativement aux trebouxiophytes, qui ont évolué après les prasinophytes.

Le génome chloroplastique de *Monomastix* est aussi dépourvu de l'organisation ancestrale des gènes. Les résultats des analyses de l'ordre des gènes de l'ADNcp de *Monomastix* montrent que ce génome est radicalement réarrangé comparativement à celui des autres algues vertes étudiées; peu de regroupements de gènes chloroplastiques ancestraux ont été conservés chez *Monomastix* (tableau I, figure 14 et 15). On retrouve toutefois des vestiges des opérons conservés et brisés, indice qu'il y a eu du réarrangement dans ce génome. Ces vestiges incluent l'opéron des protéines ribosomiques (figure 14) et le regroupement de gènes *atpA* (figure 15). Chaque regroupement est reconnaissable mais réarrangé par rapport au génome des autres algues vertes étudiées. Un fait à remarquer est que l'opéron des ARNr (regroupement 14 du tableau IX) n'a pas été touché par ces réarrangements.

La perte de la région d'ADN inversé répété est souvent reliée avec le réarrangement génomique (Palmer, 1985). La présence de ces deux éléments chez *Monomastix* pourrait appuyer cette hypothèse. Toutefois, ce génome contient aussi un grand nombre de séquences répétées dispersées. Chez les plantes terrestres, des régions de séquences répétées ont été associées au réarrangement génomique (Cosner *et al.*, 1997). Donc, les séquences répétées de l'ADNcp de *Monomastix* pourraient aussi être responsables de la réorganisation de son génome. Comme une très grande prolifération de séquences répétées a eu lieu dans ce génome, aucun lien n'a pu être fait entre le bris de certains regroupements et la présence de séquences répétées (figure 19). Il est cependant à remarquer qu'il n'y a pratiquement pas de séquences

répétées dans l'opéron des ARNs ribosomiques. Une contrainte évolutive voulant que cet opéron reste intacte est peut-être responsable de cette observation. La perte de la région d'ADN inversé répété pourrait aussi être liée aux séquences répétées. Le fait que le génome chloroplastique de *Pycnococcus prasovali* ne contienne pas de région d'ADN inversé répété, ni de séquences répétées dispersées dans le génome et qu'il soit très réarrangé vient appuyer la première hypothèse mais ébranle la deuxième.

Le génome chloroplastique de *Monomastix* est le seul, parmi les ADNcp de prasinophytes et de trebouxiophytes étudiés, à contenir des séquences répétées dispersées dans tout le génome. Ces éléments sont surtout insérés dans les régions intergéniques, ce qui diminue la densité en gènes de ce génome par rapport aux autres génomes d'algues vertes (tableaux I et VI). Tous les cadres de lecture ouverts de plus de 60 codons ainsi que les gènes *rpoB*, *rpoC1*, *rpoC2* et *ftsH* contiennent des séquences répétées. L'insertion de ces éléments dans les cadres de lecture a peut-être altéré leur fonction, si présente, mais aucune analyse n'a été faite à ce sujet. Comme tous les ORFs de plus de 60 codons uniques à l'ADNcp de *Monomastix* et n'ayant pas de fonction connue sont situés dans des régions de séquences répétées, ils ne sont probablement pas codant et leur séquence doit constituer une région intergénique. Quant aux séquences répétées insérées dans les introns *psbA-1* et *rrl-4*, elles sont insérées dans des boucles (figures 17a et 18d) et n'altèrent en rien la structure secondaire. La fonction de ces introns n'est conséquemment probablement pas touchée. Une contrainte semble restreindre la prolifération de séquences répétées aux régions du génome sans fonction. Ni la provenance, ni la fonction de ces éléments ne sont connues mais des séquences répétées du même type que dans l'ADNcp de *Monomastix* ont été trouvées dans le génome chloroplastique de plantes terrestres (Cosner *et al.*, 1997). Celles-ci ont été associées au réarrangement génomique chez ces génomes mais leur provenance ainsi que le pourquoi de leur présence sont aussi ignorés.

Aucun point d'inversion dans l'organisation des gènes, impliquant que les régions répétées soient responsables du réarrangement génomique (Cosner *et al.*, 1997), n'a pu être décelé à l'emplacement des régions répétées dans le génome chloroplastique de *Monomastix*. L'association contraire entre les séquences répétées et le réarrangement génomique, c'est-à-dire que les séquences répétées ont pu être acquises lors du réarrangement génomique, doit

être considérée. Cela pourrait expliquer le fait qu'il existe plusieurs familles de répétitions et que quelques-unes semblent avoir donné naissance à d'autres familles, peut-être par recombinaison entre les différents membres d'une même famille. Ainsi, les familles A, B et D, dont la séquence est le plus souvent retrouvée chez les autres familles, pourraient être les mères des autres familles (figure 20). Les séquences répétées pourraient aussi être des éléments mobiles qui ont été acquis dans le génome chloroplastique de *Monomastix* pour une raison inconnue et qui ont ensuite été modifiés et proliférés au cours de l'évolution.

L'insertion de six introns, dont un du groupe II, dans le génome chloroplastique de *Monomastix* est un autre élément qui diffère des autres algues vertes du tableau I dont le génome chloroplastique contient peu ou pas d'intron du groupe I et pas d'intron du groupe II. D'après les connaissances actuelles, le génome de l'ancêtre des plantes vertes ne contenait pas d'introns (Lemieux *et al.*, 2000). Donc, l'acquisition de plusieurs introns dans ce génome a entraîné la perte d'une autre caractéristique ancestrale des génomes chloroplastiques.

Malgré ces observations qui démontrent que la structure de l'ADNcp de *Monomastix* a été très réorganisée et qu'elle a évolué plus rapidement que celle des prasinophytes à la base des algues vertes et des trebouxiophytes étudiés, la séquence du gène ancestral *rnpB* (figure 11a), qui n'a été répertoriée, dans l'ADNcp des plantes vertes, que chez les prasinophytes *Nephroselmis olivacea* et *Pycnococcus prasovali*, est présente chez *Monomastix*. La présence de ce gène ne change cependant pas le fait que *Monomastix* soit un prasinophyte dont le génome chloroplastique a beaucoup divergé par rapport à celui du prasinophyte à la base des chlorophytes (*N. olivacea*) et même des autres algues vertes étudiées. L'ADNcp de *P. prasovali* a, comme celui de *Monomastix*, beaucoup divergé mais de façon différente. Ce prasinophyte a un génome chloroplastique très réarrangé, sans région d'ADN inversé répété et a perdu presque autant de gènes que *Monomastix* mais cet ADN ne contient pas de séquences répétées, n'a pas suivi le même patron de perte de gènes (figure 29) et de réorganisation génomique (tableau IX) et ne contient qu'un seul intron du groupe I (tableau I). De plus, la séquence en acides aminées du génome chloroplastique de *Pycnococcus* a évolué beaucoup plus vite que celle de *Monomastix*, ce qui résulte en une branche beaucoup plus longue dans les analyses phylogénétiques construites à l'aide de la séquence protéique concaténée de

plusieurs protéines (figure 21a). Les forces qui ont fait évoluer chacun de ces génomes chloroplastiques semblent différentes.

4.2 Le génome mitochondrial de *Monomastix* OKE-1 a conservé une structure ancestrale

Contrairement au génome chloroplastique, le génome mitochondrial de *Monomastix* a conservé tous les traits d'un génome mitochondrial ancestral d'algue verte. La première caractéristique permettant de le classer parmi les génomes ancestraux est sa taille de 60,5 kb et l'organisation compacte de son information génétique (tableau XII). Sa taille est légèrement plus grande que celle des autres ADNmt ancestraux d'algues vertes mais elle se rapproche plus de ces génomes que de ceux de type dérivé-réduit (tableau IV).

Une autre caractéristique de l'ADNmt de *Monomastix* typique aux génomes mitochondriaux ancestraux est la présence du gène de la sous-unité 5S de l'ARNr, des gènes codant pour les protéines ribosomiques et d'un éventail presque complet de gènes d'ARNt. Le caractère ancestral du génome mitochondrial de *Monomastix* est aussi démontré par la trouvaille, dans l'ADNmt de *Monomastix*, du gène *rnpB* codant pour l'ARN composant l'enzyme RNaseP. Parmi les génomes mitochondriaux de plantes vertes étudiés, ce gène a seulement été répertorié chez *N. olivacea*. Les gènes codant pour les ARNs ribosomiques mitochondriaux de *Monomastix* sont entiers, autre trait typique aux génomes mitochondriaux ancestraux. À l'exception du gène *trnR(ucg)* absent chez *Monomastix* et du gène *trnL(gag)*, un gène qui n'avait pas encore été trouvé dans le génome mitochondrial d'une plante verte avant l'étude du génome mitochondrial de *Monomastix*, le contenu en gènes du génome mitochondrial de *Monomastix* est exactement le même que celui du génome mitochondrial ancestral de *N. olivacea*. Le répertoire des gènes de *Monomastix* contient beaucoup plus de gènes que celui des génomes de type dérivé-réduit (tableaux IV, V et XIII), caractéristique tout à fait typique aux génomes ancestraux.

Parmi les génomes mitochondriaux dont la séquence complète est disponible, le gène *trnL(gag)* n'a pas non plus été répertorié chez les algues rouges mais il a été trouvé chez le

protiste *Reclinomonas americana* (Lang *et al.*, 1997). Ce gène est aussi présent chez plusieurs génomes chloroplastiques d'algues vertes mais il n'a pas été trouvé dans l'ADNcp de *Monomastix*. Comme le gène *trnL(gag)* mitochondrial de *Monomastix* présente un peu plus de similarité de nucléotides avec celui de *Reclinomonas americana* (66 %) qu'avec le gène chloroplastique de *N. olivacea* (51 %), il est plus probable que l'ADNmt de *Monomastix* ait conservé ce gène depuis l'ancêtre des mitochondries que ce gène ait été acquis par transfert horizontal entre le génome chloroplastique d'une algue verte et le génome mitochondrial de *Monomastix*.

Tout comme les génomes mitochondriaux ancestraux des algues vertes, l'ADNmt de *Monomastix* utilise le code génétique universel. Les légères modifications dans la structure standard des ARNt ne semblent pas altérer leur fonction car ils semblent tous utilisés (tableau XIV). Les ARNt codés par ce génome ne sont cependant pas suffisants pour décoder les 64 codons qu'il utilise. Il y a un seul ARNt décodant la thréonine (ARNt^{Thr}(GGU)) et il n'est pas suffisant pour décoder les quatre codons de la thréonine. Un gène *trnT* est probablement importé du cytosol pour combler ce manque. Cette situation est commune puisque plusieurs autres algues, protistes et plantes terrestres ont aussi perdu un gène *trnT* (Gray *et al.*, 1998).

Des regroupements de gènes similaires à ceux des eubactéries de même que des regroupements conservés chez les algues vertes avec un génome mitochondrial ancestral sont présents dans le génome mitochondrial de *Monomastix*. Les regroupements de *Monomastix* du tableau XV sont presque identiques à ceux de *Nephroselmis*, autre indice qu'il y a beaucoup de conservation entre ces deux génomes. Par rapport à ces deux ADNmt, les regroupements du génome mitochondrial ancestral de *Prototheca* sont un peu plus réarrangés. Trois des regroupements du tableau XV conservés entre *Monomastix* et *Nephroselmis* sont aussi similaires à des regroupements du génome mitochondrial très ancestral de *Reclinomonas*, autre indication du caractère ancestral et peu réarrangé de l'ADNmt de *Monomastix*. Une autre caractéristique retrouvée chez les génomes mitochondriaux ancestraux d'algues vertes de même que chez *Monomastix* est la présence de l'opéron des protéines ribosomiques, similaire aux opérons *str-L2-spc-alpha-S9* des eubactéries (figure 25). La présence de cet ordre de gènes chez plusieurs ADNmt confirme le caractère eubactérien du génome ancestral proto-

mitochondrial. Encore ici, on peut voir que l'opéron de *Monomastix* est très peu réarrangé par rapport au génome mitochondrial le plus ancestral et à celui des autres plantes vertes présentées. Le fait que les génomes mitochondriaux ancestraux de *M. viride* et *P. wickerhamii* n'aient conservé qu'en infime partie l'arrangement eubactérien des gènes codant pour les protéines ribosomiques suggère qu'il y a eu du réarrangement génomique chez ces génomes, surtout chez *M. viride*.

La dernière caractéristique des génomes mitochondriaux de type ancestral des algues vertes présente dans l'ADNmt de *Monomastix* est le fait qu'il y ait peu d'introns (tableau XVI) et que tous soit du groupe I. Il est cependant à remarquer qu'un peu plus d'introns ont été insérés dans le génome mitochondrial de *Monomastix* que dans celui des autres algues vertes avec un génome mitochondrial ancestral (tableau IV).

4.3 Le pourcentage en bases A + T des génomes d'organelles de *Monomastix* OKE-1 apporte certaines implications

Autant le génome chloroplastique que le génome mitochondrial de *Monomastix* sont plus riches en bases A + T qu'en bases G + C. D'ailleurs, les pourcentages en bases A + T de chacun de ces deux génomes sont très similaires (59,9 % pour le génome mitochondrial et 61,0 % pour le génome chloroplastique). Comme la méthode de séparation de l'ADN cellulaire sur gradient de chlorure de césium en présence de fluorochrome bisbenzimidazole trihydrochlorure est basée sur la différence de densité entre les molécules d'ADN et que le pourcentage en bases A + T modifie la densité de l'ADN, la similarité entre le pourcentage en bases A + T de chaque génome est probablement la raison majeure qui a empêché la séparation de l'ADN des organelles sur gradient de chlorure de césium (figure 6).

Le génome mitochondrial de *Monomastix* contient un peu moins de bases A + T que celui des autres génomes d'algues vertes (tableau IV). Contrairement aux autres génomes ancestraux d'algues vertes du tableau IV, la région codante de l'ADNmt de *Monomastix* est plus élevée en bases A + T que la région non-codante (XII). Cette particularité est plus souvent trouvée chez les génomes mitochondriaux d'algues vertes de type dérivé-réduit que de type ancestral.

Chez *Monomastix*, elle peut être causée par la présence d'éléments riches en G + C insérés dans la plupart des régions intergéniques et augmentant le pourcentage en bases G + C des régions non-codantes. L'avènement de ces éléments dans les ORFs uniques au génome mitochondrial de *Monomastix* est d'ailleurs ce qui permet de croire que ces ORFs n'ont aucune fonction et font plutôt partie des régions non-codantes. Dans le génome chloroplastique, malgré le grand nombre de séquences répétées souvent riches en G + C, autant la séquence totale que la région codante et la région non-codante ont le même pourcentage en bases A + T (tableau VI).

Pour chacun des deux génomes d'organelles de *Monomastix*, le plus grand pourcentage en bases A + T versus les bases G + C résulte en une utilisation biaisée des bases A et T versus les bases G et C en troisième position des codons. Ce biais d'utilisation est cependant plus marqué dans l'ADNcp que dans l'ADNmt, même si le pourcentage en bases A + T n'est que légèrement plus élevé dans l'ADNcp.

4.4 Des évènements de transfert latéraux et horizontaux à l'origine des introns de *Monomastix*

4.4.1 L'intron chloroplastique du gène *psbA*

À l'exception de *Monomastix*, parmi tous les organismes photosynthétiques dont le gène *psbA* a été séquencé, *C. reinhardtii* est le seul à posséder un intron dans le gène *psbA* inséré au codon 175 (tableau XVII). Il a déjà été démontré que la structure secondaire de l'intron *Cr.psbA-3* et celle de l'intron du gène *sunY* du bactériophage *T4* étaient remarquablement similaires (Holloway *et al.*, 1999). Des similarités entre cet intron du groupe I chez ces deux organismes très distants et le fait que l'intron n'ait pas été trouvé chez d'autres *Chlamydomonas* suggéraient que l'intron de *C. reinhardtii* ait été acquis par un récent évènement de transfert horizontal du phage au chloroplaste ou même d'un troisième organisme, plutôt que d'un transfert vertical (Holloway *et al.*, 1999). Comme le pourcentage de similarité entre la structure secondaire de l'intron de *Monomastix* et celui de *C. reinhardtii* (63%) (figure 17b) ainsi que celui entre *C. reinhardtii* et le bactériophage (65%) (Holloway *et*

Tableau XVII. Répertoire des introns localisés dans les mêmes gènes et insérés aux mêmes positions que les introns de l'ADNcp de *Monomastix* OKE-1 mais chez d'autres génomes.

Introns de <i>Monomastix</i>			Introns des autres génomes	
Noms	Classe	Emplacement de l'ORF ^a	Source ^b	Emplacement de l'ORF ^a
Introns du groupe I				
<i>psbA</i> ·1	IA2	L6	<i>Chlamydomonas reinhardtii</i> (C) cp	L6
<i>rrl</i> ·1	IB4	L8	<i>Chlamydomonas frankii</i> (C) cp	L8
			<i>Chlamydomonas geitleri</i> (C) cp	L8
			<i>Chlamydomonas mexicana</i> (C) cp	L8
			<i>Chlamydomonas pallidostigmatica</i> (C) cp	L8
			<i>Chlorosarcina brevispinosa</i> (T) cp	L8
			<i>Chlorella vulgaris</i> (T) mt	L8
			<i>Monomastix</i> sp. M722 (P) mt	L8
			<i>Monomastix</i> sp. M722 (P) cp	L8
			<i>Monomastix</i> sp. OKE-1 (P) mt	L8
			<i>Nephroselmis olivacea</i> (P) mt	L8
			<i>Pedinomonas tuberculata</i> (P) cp	L8
			<i>Mesostigma viride</i> (P) mt	L8
			<i>Stimkania negevensis</i> nu	L8
			<i>Acanthamoeba castellanii</i> mt	L8
<i>rrl</i> ·2	IA3	L8	<i>Chlorosarcina brevispinosa</i> (T) cp	L8
			<i>Chlorella vulgaris</i> (T) mt	L8
			<i>Monomastix</i> sp. M722 (P) cp	L8
			<i>Acanthamoeba castellanii</i> mt	L8
<i>rrl</i> ·3	IB4	après P9	<i>Chlamydomonas humicola</i> (C) cp	L9.1
			<i>Monomastix</i> OKE-1 (P) mt	après P9
<i>rrl</i> ·4	IA3	L6	<i>Carteria luzensis</i> (C) cp	L6
			<i>Carteria olivieri</i> (C) cp	L6
			<i>Chlamydomonas agloeformis</i> (C) cp	L6
			<i>Chlamydomonas iyengarii</i> (C) cp	L6
			<i>Chlamydomonas reinhardtii</i> (C) cp	L6
			<i>Haematococcus lacustris</i> (C) cp	L6
			<i>Scenedesmus obliquus</i> (C) cp	L6
			<i>Chlorella vulgaris</i> (T) cp	L6

Suite du **Tableau XVII**

Introns de <i>Monomastix</i>			Introns des autres génomes	
Noms	Classe	Emplacement de l'ORF	Source	Emplacement de l'ORF
			<i>Pseudendoclonium akinetum</i> (U) cp	L6
			<i>Trichosarina mucosa</i> (U) cp	L6
			<i>Monomastix</i> sp. M722 (P) cp	L6
			<i>Monomastix</i> OKE-1 (P) mt	L6
			<i>Nephroselmis olivacea</i> (P) mt	L6
			<i>Scherffelia dubia</i> (P) cp	L6
			<i>Mesostigma viride</i> (P) mt	L6
			<i>Acanthamoeba castellanii</i> mt	L6
Intron du groupe II				
		<i>trnK(uuu)</i> -1 ^c	IIB	

^a La position des ORFs (ou cadres de lecture ouverts) est indiquée selon leur emplacement dans la structure secondaire de l'intron. P indique une région d'appariement; L suivi d'un chiffre désigne la boucle reliée à la région d'appariement numérotée.

^b mt, génome mitochondrial; cp, génome chloroplastique; nu, génome nucléaire. La classe à laquelle chaque algue verte appartient est indiquée entre parenthèses : C, Chlorophyceae; T, Trebouxiophyceae; U, Ulvophyceae; et P, Prasinophyceae.

^c L'absence de donnée dans une catégorie signifie que la catégorie considérée n'est pas applicable.

al., 1999) est environ le même mais que celui entre *Monomastix* et le bactériophage est plus faible (54%) (figure 17b), il est plus probable que l'intron de *Monomastix* ait été acquis par transfert horizontal entre son génome chloroplastique et celui de *C. reinhardti* que par transfert horizontal entre son ADNcp et le bactériophage. Le fait que la structure secondaire des deux introns chloroplastiques contiennent le domaine P2 et que leur ORF respectif soit inséré dans la même boucle alors que l'intron du bactériophage n'a pas de domaine P2 et que son ORF n'est pas inséré dans la même boucle que les deux autres introns appuie cette hypothèse. Ces résultats ne permettent cependant pas de dire si l'événement de transfert entre les deux génomes chloroplastiques est plus récent que celui entre *C. reinhardti* et le bactériophage *T4*. Il n'y a pas de similarité significative entre l'ORF de l'intron de *Monomastix* et celui de chacun des deux autres introns, ce qui peut être dû au fait qu'un second événement d'insertion non relié à l'acquisition de l'intron soit responsable de la présence de chacun des ORFs (Holloway *et al.*, 1999).

4.4.2 L'intron chloroplastique du gène *trnK(uuu)*

Il est commun de retrouver, chez les plantes terrestres, un intron du groupe II dans le gène *trnK(uuu)*. Cet intron, dont le cadre de lecture ouvert code pour la maturase K, est localisé dans la boucle de l'anticodon de la structure secondaire de l'ARNt (Neuhaus et Link, 1987). L'intron du groupe II trouvé dans le gène *trnK(uuu)* de l'ADNcp de *Monomastix* (figure 17b et tableau XVII) est différent de celui trouvé chez les plantes terrestres. Contrairement aux introns du gène *trnK* chloroplastique trouvés chez les plantes terrestres, l'intron de *Monomastix* n'est pas inséré dans la boucle de l'anticodon de l'ARNt, mais plutôt dans le bras D (figure 16). L'absence de cadre de lecture ouvert dans l'intron de *Monomastix* est un autre élément qui diffère des plantes terrestres. De plus, la structure secondaire de l'intron de *Monomastix* fait partie du sous-groupe IIB alors que l'intron des plantes terrestres est classé dans le sous-groupe IIA (Neuhaus et Link, 1987). On peut donc déduire que l'intron de *Monomastix* ne provient pas d'un événement de transfert horizontal entre le génome chloroplastique de cette algue verte et l'intron *trnK* d'une plante terrestre, puisque les deux introns sont significativement divergents. Rien de plus n'est connu sur la provenance de cet intron dans l'ADNcp de *Monomastix*.

Tableau XVIII. Répertoire des introns localisés dans les mêmes gènes et insérés aux mêmes positions que les introns de l'ADNmt de *Monomastix* OKE-1 mais chez d'autres génomes.

Introns de <i>Monomastix</i>			Introns des autres génomes	
Noms	Classe	Emplacement de l'ORF ^a	Source ^b	Emplacement de l'ORF ^a
<i>rnl-1</i> ^c	IB4		<i>Chlamydomonas eugametos</i> (C) cp	L6
			<i>Chlorococcum echinozigtum</i> (C) cp	L6
			<i>Chlamydomonas monadina</i> (C) cp	L6
			<i>Chlorogonium elongatum</i> (C) cp	L6
			<i>Chlamydomonas pallidostigmatica</i> (C) cp	L6
			<i>Chlamydomonas mutabilis</i> (C) cp	L6
			<i>Carteria luzensis</i> (C) cp	L6
			<i>Scenedesmus obliquus</i> (C) cp	L6
			<i>Ankistrodesmus stipitatus</i> (C) cp	L6
<i>rnl-2</i>	IB4	L8	<i>Chlamydomonas frankii</i> (C) cp	L8
			<i>Chlamydomonas geitleri</i> (C) cp	L8
			<i>Chlamydomonas mexicana</i> (C) cp	L8
			<i>Chlamydomonas pallidostigmatica</i> (C) cp	L8
			<i>Chlorosarcina brevispinosa</i> (T) cp	L8
			<i>Chlorella vulgaris</i> (T) mt	L8
			<i>Monomastix</i> sp. M722 (P) mt	L8
			<i>Monomastix</i> sp. M722 (P) cp	L8
			<i>Monomastix</i> sp. OKE-1 (P) cp	L8
			<i>Nephroselmis olivacea</i> (P) mt	L8
			<i>Pedinomonas tuberculata</i> (P) cp	L8
			<i>Mesostigma viride</i> (P) mt	L8
			<i>Stimkania negevensis</i> nu	L8
			<i>Acanthamoeba castellanii</i> mt	L8
<i>rnl-3</i>	IB4	après P9	<i>Chlamydomonas humicola</i> (C) cp	L9.1
			<i>Monomastix</i> OKE-1 (P) cp	après P9
<i>rnl-4</i>	IA3	L6	<i>Carteria luzensis</i> (C) cp	L6
			<i>Carteria olivieri</i> (C) cp	L6
			<i>Chlamydomonas agloeformis</i> (C) cp	L6
			<i>Chlamydomonas iyengaraii</i> (C) cp	L6
			<i>Chlamydomonas reinhardtii</i> (C) cp	L6
			<i>Haematococcus lacustris</i> (C) cp	L6
			<i>Scenedesmus obliquus</i> (C) cp	L6
<i>Chlorella vulgaris</i> (T) cp	L6			

Suite du Tableau XVIII

Noms	Classe	Emplacement de l'ORF ^a	Source ^b	Emplacement de l'ORF ^a
			<i>Pseudendoclonium akinetum</i> (U) cp	L6
			<i>Trichosarina mucosa</i> (U) cp	L6
			<i>Monomastix sp.</i> M722 (P) cp	L6
			<i>Monomastix</i> OKE-1 (P) cp	L6
			<i>Nephroselmis olivacea</i> (P) mt	L6
			<i>Scherffelia dubia</i> (P) cp	L6
			<i>Mesostigma viride</i> (P) mt	L6
			<i>Acanthamoeba castellanii</i> mt	L6
<i>cox1-1</i>	IB2	L8	<i>Marchantia polymorpha</i> mt	L8
<i>cox1-2</i>	IB2	L8	<i>Marchantia polymorpha</i> mt	L8
			<i>Allomyces macrogynus</i> mt	L8
			<i>Emericella nidulens</i> mt	L8
			<i>Podospora anserina</i> mt	L8
			<i>Neurospora crassa</i> mt	L8
<i>cox1-3</i>	IB2	L8	<i>Chlorogonium elongatum</i> (C) mt	L8
			<i>Mesostigma viride</i> (P) mt ^c	
			<i>Allomyces macrogynus</i> mt	L8
			<i>Podospora anserina</i> mt	L8
<i>cox1-4</i>	IB3	L1	<i>Marchantia polymorpha</i> mt	L1
			<i>Chlamydomonas eugametos</i> (C) mt	L9.1
			<i>Mesostigma viride</i> (P) mt	L1
			<i>Allomyces macrogynus</i> mt ^c	
			<i>Dictyostelium discoideum</i> mt	L1
			<i>Kluyveromyces lactis</i> mt	L1
			<i>Podospora anserina</i> mt	L1
			<i>Saccharomyces cerevisiae</i> mt	L9.1
			<i>Saccharomyces douglasii</i> mt	L9.1
			<i>Schizosaccharomyces pombe</i> mt ^c	

^a La position des ORFs (ou cadres de lecture ouverts) est indiquée selon leur emplacement dans la structure secondaire de l'intron. P indique une région d'appariement; L suivi d'un chiffre désigne la boucle reliée à la région d'appariement numérotée.

^b mt, génome mitochondrial; cp, génome chloroplastique; nu, génome nucléaire. La classe à laquelle chaque algue verte appartient est indiquée entre parenthèses : C, Chlorophyceae; T, Trebouxiophyceae; U, Ulvophyceae; et P, Prasinophyceae.

^c L'absence de donnée dans une catégorie signifie que la catégorie considérée n'est pas applicable.

4.4.3 Les introns chloroplastiques et mitochondriaux du gène de l'ARNr 23S

Les introns chloroplastiques ou mitochondriaux du gène de l'ARNr 23S de *Monomastix* n'ont pas été étudiés en profondeur mais la liste des introns publiés et insérés dans ce gène aux mêmes positions que ceux de *Monomastix* mais chez d'autres organismes est présentée dans les tableaux XVII et XVIII. Les introns des positions 1931 (*rri-1* et *rnl-2*), 2500 (*rri-3* et *rnl-3*) et 2593 (*rri-4* et *rnl-4*) sont insérés autant dans le gène chloroplastique que dans le gène mitochondrial de *Monomastix* (tableaux X et XVI). Un transfert latéral entre les compartiments mitochondriaux et chloroplastiques à l'intérieur d'une même cellule pourrait être à l'origine de la présence de ces introns (Turmel *et al.*, 1995). Une similarité modérée à forte (entre 60 et 75%) entre la structure secondaire des deux introns de chaque organelle de *Monomastix* à chaque position supporte cette idée. Pour les introns des positions 1931 et 2593, comme le pourcentage de similarité entre la structure secondaire des introns des génomes mitochondriaux ancestraux à la base des plantes vertes (*M. viride* et *N. olivacea*) et celle de chaque intron de *Monomastix* à la même position est très peu différent de celui entre les introns des deux compartiments de *Monomastix* (figures 18 et 26), il est impossible de dire, sur la base de cette observation, si l'événement de transfert entre les compartiments mitochondriaux et chloroplastiques est plus récent que l'événement de transfert entre un des deux compartiments de *Monomastix* et un autre organisme. On ne retrouve aussi que peu de similarité entre la séquence en acides aminés des ORFs des introns à chaque site, ce qui ne permet pas de conclure qu'ils ont été acquis suite à l'insertion des introns dans le gène. Ils pourraient avoir été transmis dans le même événement de transfert que les introns eux-mêmes. Pour les deux introns de *Monomastix* insérés à la position 2500, la situation est différente. Le pourcentage de similarité entre les deux introns de *Monomastix* est significativement plus élevé (~ 75%) que celui entre chacun de ces introns et *Chlamydomonas humicola* (~55%). Cette observation laisse supposer qu'il y ait eu un événement de transfert latéral récent entre les compartiments mitochondriaux et chloroplastiques de *Monomastix* puisqu'il y a moins de divergence entre les deux introns de *Monomastix* qu'entre ces introns et celui d'un autre organisme avec lequel il y a peut-être eu un événement de transfert.

Parmi les introns mitochondriaux, on remarque que celui inséré à la position 1922 (*rnl-1*)

n'avait été trouvé que chez des génomes chloroplastiques d'algues vertes de la classe Chlorophyceae avant l'étude du génome mitochondrial de *Monomastix* (tableau XVIII). Cette observation laissait supposer que cet intron était présent dans le génome chloroplastique de l'ancêtre de la classe Chlorophyceae et qu'il avait été transmis chez les membres de cette classe par transfert vertical. La présence de cet intron dans l'ADNmt de *Monomastix* laisse présumer qu'il y a eu un transfert d'information génétique via un échange intercellulaire entre *Monomastix* et un chlorophyte (Turmel *et al.*, 1995).

4.4.4 Les introns mitochondriaux du gène *cox1*

Le tableau XVIII présente les génomes ayant un intron *cox1* publié inséré à la même position que ceux de *Monomastix*. Les introns *cox1-3* et *cox1-4* de *Monomastix* sont aussi présents chez *M. viride*. La structure secondaire de l'intron *cox1-3* de *Monomastix* est similaire à son homologue chez *M. viride* mais celle de l'intron *cox1-4* diverge beaucoup plus de son homologue chez *M. viride* (figure 27c et 27d). Ce dernier résultat indique peut-être que ces deux introns ont beaucoup divergé depuis leur insertion et que l'événement de transfert n'est pas récent. Un autre fait à remarquer dans le tableau XVIII est que les introns insérés au même site que les introns *cox1-2*, *cox1-3* et *cox1-4* sont souvent présents chez des champignons, organismes éloignés des algues vertes.

4.5 La position phylogénétique de *Monomastix* OKE-1 est surprenante

Les analyses phylogénétiques basées sur la séquence en acides aminés de plusieurs protéines concaténées autant chloroplastiques que mitochondriales donnent des résultats différents de ce qui était attendu. Les analyses phylogénétiques préliminaires faites avec le gène de la grande sous-unité de l'ARNr chloroplastique émettaient l'hypothèse que *Monomastix* OKE-1 était placé à la base de la lignée Chlorophyta, dans un groupe encore plus basal que *Nephroselmis olivacea*, tout comme *Pycnococcus prasovali* (résultats du laboratoire non-publiés). Toutes les analyses faites avec la séquence concaténée de plusieurs protéines présentent plutôt *Monomastix*, avec un niveau de confiance élevé, dans un groupe ayant évolué après *N. olivacea*. Ces analyses phylogénétiques plus concluantes, puisqu'elles représentent une plus

grande portion de la séquence chloroplastique ou mitochondriale des taxa utilisés, apportent l'hypothèse que *Monomastix* soit un prasinophyte plus évolué que *N. olivacea* et donc qu'il soit moins près de l'ancêtre de la lignée Chlorophyta que ne l'est lui-même *N. olivacea*.

Dans les arbres phylogénétiques chloroplastiques de la figure 21, on peut observer que la branche de *Monomastix* est légèrement plus longue que celle des autres taxons. Comme la longueur des branches est proportionnelle au nombre de substitutions de nucléotides et donc à la vitesse d'évolution de la séquence en acides aminés des protéines utilisées (Graur et Li, 2000), cela indique que la séquence en acides aminés de l'ADNcp de *Monomastix* utilisée pour faire l'analyse a évolué un peu plus rapidement que celle des autres algues vertes de l'analyse.

Dans l'analyse incluant la séquence protéique concaténée du prasinophyte *P. prasovali* (figure 21a), *Monomastix* se lie avec ce dernier mais les valeurs de « bootstrap » sont un peu faibles. La séquence de *P. prasovali* semble avoir évolué beaucoup plus rapidement que celle des autres taxa puisque sa branche est beaucoup plus longue. Comme la branche de *Monomastix* est aussi plus longue que celle des autres taxa, le phénomène de l'attraction des longues branches (Graur et Li, 2000) peut être responsable du regroupement entre *Monomastix* et *Pycnococcus* ainsi que de la déstabilisation des autres regroupements. C'est pour éviter ce phénomène et vérifier les embranchements moins supportés de la figure 21a que l'arbre de la figure 21b a été construit. Les résultats de cette deuxième analyse indiquent qu'effectivement, l'absence de la séquence de *Pycnococcus* solidifie les embranchements moins bien supportés mais aussi qu'il manque des séquences de prasinophytes et d'ulvophytes pour confirmer la position phylogénétique des prasinophytes *Monomastix* OKE-1, *Pedinomonas minor* et *Pycnococcus prasovali*.

La position phylogénétique de *Monomastix*, d'après les arbres de la figure 28 construits avec la séquence en acides aminés de plusieurs protéines mitochondriales concaténées, appuie les hypothèses apportées par la caractérisation de ce génome mitochondrial. Il faut cependant noter que la séquence utilisée pour chacune des analyses mitochondriales comprend beaucoup moins de protéines, surtout pour l'analyse de la figure 28b, que celles des analyses

chloroplastiques. La séquence de la région codante entière de l'ADNmt est ainsi moins bien représentée. Chacun des deux arbres présente distinctement *Monomastix* parmi les algues vertes ayant un génome mitochondrial de type ancestral. On voit aussi clairement, sur ces arbres et en particulier dans l'arbre de la figure 28b, que la vitesse d'évolution de la séquence en acides aminés des protéines mitochondriales des plantes terrestres est lente, légèrement plus rapide pour les génomes mitochondriaux d'algues vertes de type ancestral et beaucoup plus rapide pour ceux d'algues vertes de type dérivé-réduit. Autre fait à remarquer dans la figure 28b est la présence, fortement supportée, du prasinophyte *Pycnococcus prasovali* parmi les algues vertes à génome mitochondrial de type dérivé-réduit, même si son génome mitochondrial contient plus de caractéristiques des génomes mitochondriaux de type ancestral et que la vitesse d'évolution de sa séquence en acides aminés est beaucoup plus lente que celle des autres génomes mitochondriaux d'algues vertes de type dérivé-réduit. En comparaison, dans les analyses chloroplastiques (figure 21a), *P. prasovali*, qui se groupe avec *Monomastix* et *P. minor* (faisant ici partie des algues vertes avec un ADNmt dérivé-réduit), se place avec les trebouxiphytes.

Plusieurs aberrations sont présentes dans l'arbre mitochondrial de la figure 28b. D'abord, *N. olivacea* s'embranchement anormalement à la base des plantes terrestres et des valeurs de « bootstrap » non-significatives ont été calculées pour les groupements entre les algues vertes avec des ADNmt ancestraux. Ensuite, les génomes dérivés-réduits sont étrangement positionnés comme un groupe sœur au groupe formé par les plantes vertes avec un génome mitochondrial dit ancestral. La vitesse d'évolution des génomes dérivés-réduits, beaucoup plus grande que celle des autres génomes mitochondriaux de plantes vertes, est probablement le facteur qui sépare le groupe des génomes dérivés des autres génomes de plantes vertes. Les branches très longues peuvent donner naissance au phénomène d'attraction des longues branches (Graur et Li, 2000) et déstabiliser tout l'arbre. C'est probablement ce qui s'est passé dans la figure 28b car, dans l'arbre de la figure 28a, le groupe formé par les algues vertes à génome ancestral est plus stable. Le nombre limité de séquences entières d'algues vertes ainsi que de données dont l'ordinateur pouvait traiter avec la méthode « Maximum Likelihood » sont deux autres facteurs qui ont contribué à l'obtention d'arbres non parfaitement résolus.

4.6 Conclusion et perspectives

Le séquençage et la caractérisation du génome chloroplastique ainsi que du génome mitochondrial de *Monomastix* OKE-1 ont permis de mettre en évidence que ce prasinophyte n'est pas aussi basal que certaines études phylogénétiques basées sur un seul gène semblaient indiquer. D'abord, l'étude de l'architecture de son génome chloroplastique a démontré qu'il avait perdu beaucoup de caractéristiques ancestrales et que des événements de recombinaison entre les séquences répétées pourraient être à la base du réarrangement génomique. La position phylogénétique de *Monomastix* basée sur son ADNcp confirme que ce génome chloroplastique n'est pas celui qui se rapproche le plus du génome de l'ancêtre de la lignée Chlorophyta. Il aurait plutôt évolué rapidement par rapport aux autres génomes chloroplastiques de prasinophytes. La caractérisation du génome mitochondrial de *Monomastix* a fait ressortir, à l'opposé, que ce génome avait conservé tous les traits d'un génome mitochondrial ancestral d'algue verte. Mais, l'analyse phylogénétique de *Monomastix* basée sur son ADNmt démontre aussi que le génome mitochondrial de *Monomastix* n'est pas celui qui se rapproche le plus du génome de l'ancêtre de la lignée Chlorophyta. Comme l'ADNcp de *Monomastix* a beaucoup divergé par rapport aux autres génomes chloroplastiques d'algues vertes mais que l'ADNmt est très conservé, l'étude du génome mitochondrial est plus concluante.

Chacune des analyses phylogénétiques de *Monomastix* a aussi permis de clarifier un peu la position des algues vertes dont la séquence entière est disponible. La caractérisation de ces deux génomes a apporté des éléments qui n'avaient pas été observés chez d'autres génomes d'organelles d'algues vertes (séquences répétées de l'ADNcp, gène *trnL(gag)* mitochondrial, introns, etc.). Le séquençage et l'analyse de nouveaux génomes chloroplastiques et mitochondriaux pourront peut-être, éventuellement, apporter des réponses sur l'origine et le pourquoi de la présence de ces éléments en plus de solidifier la position phylogénétique des algues vertes, en particulier celle des prasinophytes *Monomastix* OKE-1, *Pycnococcus prasovali* et *Pedinomonas minor*. Plusieurs projets sont d'ailleurs en cours dans notre laboratoire, dont celui du séquençage des génomes des deux organelles du premier ulvophyte à être séquençé.

RÉFÉRENCES

- Adachi, J., Hasegawa, M., MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood method. *Comput. Sci. Monogr.*, **28**, 1-150 (1996).
- Andersson, G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Pontén, T., Alsmark, U. C. M., Podowski, R. M., Näslund, A. K., Eriksson, A.-S., Winkler, H. H., Kurkland, C. G., The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria, *Nature*, **396**, 133-143 (1998).
- Baxevanis, A. D., Ouellette, B. F. F., *Bioinformatics, a Practical Guide to the Analysis of Genes and Proteins* (John Wiley and Sons Inc., New York, 1998).
- Bhattacharya, D., Medlin, L., The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions, *J. Phycol.*, **31**, 489-498 (1995).
- Bhattacharya, D., Medlin, L., Algal phylogeny and the origin of land plants, *Plant Physiol.*, **116**, 9-15 (1998).
- Boer, P. H., Gray, M. W., Short dispersed repeats localized in spacer regions of *Chlamydomonas reinhardtii* mitochondrial DNA, *Curr. Genet.*, **19**, 309-312 (1991).
- Bold, H. C., Wynne, M. J., *Introduction to the Algae: Structure and Reproduction, Second Edition* (Prentice-Hall, Englewood Cliffs, New Jersey, 1985).
- Boudreau, E., Turmel, M., Extensive gene rearrangements in the chloroplast of *Chlamydomonas* species featuring multiple dispersed repeats, *Mol. Biol. Evol.*, **13**, 233-243 (1996).
- Brown, J. R., Doolittle, W. F., Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications, *Proc. Natl. Acad. Sci. USA*, **92**, 2441-2445 (1995).

Burger, G., Saint-Louis, D., Gray, M. W., Lang, B. F., Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*. Cyanobacterial introns and shared ancestry of red and green algae, *Plant Cell.*, **11**, 1675-94 (1999).

Burrows, P. A., Sazanov, L. A., Svab, Z., Maliga, P., Nixon, P. J., Identification of a functional respiratory complex in chloroplasts through analysis of tobacco mutants containing disrupted plastid *ndh* genes, *EMBO J.*, **17**, 868-876 (1998).

Campbell, N. A., Mathieu, R., La photosynthèse, dans *Biologie*, éditeurs : Éditions du Renouveau Pédagogique Inc., pp. 199-217 (The Benjamin/Cummings Publishing Company Inc., Canada, 1995a).

Campbell, N. A., Mathieu, R., Conversion de l'énergie par les mitochondries et les chloroplastes, dans *Biologie*, éditeurs : Éditions du Renouveau Pédagogique Inc., pp. 135-136 (The Benjamin/Cummings Publishing Company Inc., Canada, 1995b).

Cavalier-Smith, T., A 6-kingdom classification and a unified phylogeny, dans *Endocytobiology II. Intracellular Space as Oligogenetic*, éditeurs: Schenk, H. E. A., Schwemmler, W. S., pp. 1027-1034 (De Gruyter, Berlin, 1983).

Chapman, R. L., Buchheim, M. A., Delwiche, C. F., Friedl, T., Huss, V. A., Karol, K. G., Lewis, L. A., Manhart, J., McCourt, R. M., Olsen, J. L., Waters, D. A., Molecular systematics of the green algae, dans *Molecular Systematics of Plants II DNA Sequencing*, éditeurs: Soltis, D. E., Soltis, P. S., Doyle, J. J., pp. 508-540 (Kluwer Academic Publishers, Norwell, 1998).

Clegg, M. T., Gaut, B. S., Learn Jr., G. H., Morton, B. R., Rates and patterns of chloroplast DNA evolution, *Proc. Natl. Acad. Sci. USA*, **91**, 6795-6801 (1994).

Cosner, M. E., Jansen, R. K., Palmer, J. D., Downie, S. R., The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families,

Curr. Genet., **31**, 419-429 (1997).

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

Denovan-Wright, E. M., Sankoff, D., Spencer, D. F., Lee, R. W., Evolution of fragmented mitochondrial ribosomal RNA genes in *Chlamydomonas*, *J. Mol. Evol.*, **42**, 382-391 (1996).

Douglas, S. E., Plastid evolution: origins, diversity, trends, *Curr. Opin. Genet. Dev.*, **8**, 655-61 (1998).

Friedl, T., Inferring taxonomic positions and testing genus level assignments in coccoid green lichen algae: a phylogenetic analysis of 18S ribosomal RNA sequences from *Dictyochloropsis reticula* and from members of the genus *Myrmecia* (Chlorophyta, Trebouxiophyceae *cl. nov.*), *J. Phycol.*, **31**, 632-639 (1995).

Friedl, T., The evolution of the Green Algae, *Pl. Syst. Evol. [Suppl.]*, **11**, 87-101 (1997).

Graur, D., Li, W.-H., Molecular phylogenetics, dans *Fundamentals of Molecular Evolution, Second Edition*, éditeur: Sinauer, A. D., pp. 165-247 (Sinauer Associates Inc., Massachussets, 2000).

Gray, M. W., Doolittle, W. F., Has the endosymbiont hypothesis been proven?, *Microbiol. Rev.*, **46**, 1-42 (1982).

Gray, M. W., *Rickettsia*, typhus and the mitochondrial connection, *Nature*, **396**, 109-110 (1998).

Gray, M. W., Lang, B. F., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Delage, E., Littlejohn, T. G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y., Burger, G., Genome structure and gene content in protist mitochondrial DNAs, *Nucleic Acids Res.*, **26**,

865-878 (1998).

Gray, M. W., Burger, G., Lang, B. F., Mitochondrial evolution, *Science*, **283**, 1476-1481 (1999).

Holloway, S. P., Deshpande, N. N., Herrin, D. L., The catalytic group-I introns of the *psbA* gene of *Chlamydomonas reinhardtii*: core structures, ORFs and evolutionary implications, *Curr. Genet.*, **36**, 69-78 (1999).

Karlin, S., Brocchieri, L., Mrazek, J., Campbell, A. M., Spormann, A. M., A chimeric prokaryotic ancestry of mitochondria and primitive eukaryotes, *Proc. Natl. Acad. Sci. USA*, **96**, 9190-9195 (1999).

Keeling, P. J., A kingdom's progress: Archezoa and the origin of eukaryotes, *BioEssays*, **20**, 87-95 (1998).

Lang, B. F., Burger, G., O'Kelly, C. J., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Gray, M. W., An ancestral mitochondrial DNA resembling a eubacterial genome in miniature, *Nature*, **387**, 493-497 (1997).

Lang, B. F., Gray, M. W., Burger, G., Mitochondrial genome evolution and the origin of eukaryotes, *Annu. Rev. Genet.*, **33**, 351-397 (1999).

Leblanc, C., Richard, O., Kloareg, B., Viehmann, S., Zetsche, K., Boyen, C., Origin and evolution of mitochondria: what have we learnt from red algae?, *Curr. Genet.*, **31**, 193-207 (1997).

Lemieux, C., Otis, C., Turmel, M., Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution, *Nature*, **403**, 649-652 (2000).

Lindhal, L., Sor, F., Archer, R. H., Normura, M., Zengel, J. M., Transcriptional organization

of the S10, *spc*, and α operons of *Escherichia coli*, *Biochim. Biophys. Acta.*, **1050**, 337-342 (1990).

Löffelhardt, W., Bohnert, H. J., Bryant, D. A., The complete sequence of the *Cyanophora paradoxa* cyanelle genome (Glaucocystophyceae), *Plant Syst. Evol. (Suppl.)*, **11**, 149-162 (1997).

Lucas, P., Otis, C., Mercier, J.-P., Turmel, M., Lemieux, C., Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases, *Nucleic Acids Res.*, **29**, 960-969 (2001).

Margulis, L., *Origin of Eukaryotic Cells* (Yale University Press, New Haven, 1970).

Martin, W., Müller, M., The hydrogen hypothesis for the first eukaryote, *Nature*, **392**, 37-41 (1998).

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M., Kowallik, K. V., Gene transfer to the nucleus and the evolution of chloroplasts, *Nature*, **393**, 162-165 (1998).

Mattox, K. R., Stewart, K. D., Classification of the green algae: a concept based on comparative cytology, dans *Systematics of the Green Algae*, éditeurs: Irvine, D. E. G., John, D. M., pp. 29-72 (Academic Press, London, 1984).

Melkonian, M., Phylum Chlorophyta. Class Prasinophyceae, dans *Handbook of Protoctista. The Structure, Cultivation, Habitats and Life Histories of the Eukaryotic Microorganisms and their Descendants Exclusive of Animals, Plants and Fungi*, éditeurs: Margulis, L., Corliss, J. O., Melkonian, M., Chapman, D. J., pp. 600-607 (Jones and Bartlett Publishers, Boston, 1990).

Melkonian, M., Surek, B., Phylogeny of the Chlorophyta: congruence between ultrastructural and molecular evidence, *Bull. Soc. Zool. Fr.*, **120**, 191-208 (1995).

Melkonian, M., Marin, B., Surek, B., Phylogeny and evolution of the algae, dans *Biodiversity and Evolution*, éditeurs: Arai, R., Kato, M., Doi, Y., pp. 153-176 (The National Science Museum Fondation, Tokyo, 1995).

Mereschkowsky, C., Über Natur und Ursprung der Chromatoporen im Pflanzenreiche, *Biologisches Centralblatt*, **25**, 593-604 (1905).

Mes, T. H. M., Kuperus, P., Kirschner, J., Stepanek, J., Oosterveld, P., Storchova, H., Nijs, J. C. M., Hairpins involving both inverted and direct repeats are associated with homoplasious indels in non-coding chloroplast DNA of *Taraxacum* (Lactuceae: Asteraceae), *Genome*, **43**, 634-641 (2000).

Michel, F., Westhof, E., Modelling of the three-dimensional architecture of group I catalytic introns based on comparative analysis, *J. Mol. Biol.*, **216**, 585-610 (1990).

Moreira, D., Le Guyader, H., Phillippe, H., The origin of red algae and the evolution of chloroplasts, *Nature*, **405**, 69-72 (2000).

Müller, M., Martin, W., The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes, *Bioessays*, **21**, 377-381 (1999).

Nakayama, T., Marin, B., Kranz, H. D., Surek, B., Huss, V. A. R., Inouye, I., Melkonian, M., The basal position of scaly green flagellates among the green algae (Chlorophyta) is revealed by analyses of nuclear-encoded SSU rRNA sequences, *Protist*, **149**, 367-380 (1998).

Nedelcu, A. M., Lee, R. W., Lemieux, C., Gray, M. W., Burger, G., The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome, *Genome Res.*, **10**, 819-31 (2000).

Neuhaus, H., Link, G., The chloroplast tRNA^{Lys}(UUU) gene from mustard (*Sinapsis alba*) contains a class II intron potentially coding for a maturase-related polypeptide, *Curr. Genet.*,

11, 251-257 (1987).

Palmer, J. D., Evolution of chloroplasts and mitochondrial DNA in plants and algae, dans *Monographs in Evolutionary Biology: Molecular Evolutionary Genetics*, éditeur: MacIntyre, R. J., pp. 131-240 (Plenum Press, New York, 1985).

Palmer, J. D., Plastid chromosomes: structure and evolution, dans *The Molecular Biology of Plastids. Cell Culture and Somatic Cell Genetics of Plants. Vol. 7A*, éditeurs: Bogorad, L., Vasil, I. K., pp. 5-53 (Academic Press, San Diego, 1991).

Palmer, J. D., Delwiche, C. F., The origin and evolution of plastids and their genomes, dans *Molecular Systematics of Plant II DNA Sequencing*, éditeurs: Soltis, D. E., Soltis, P. S., Doyle, J. J., pp. 375-409 (Kluwer Academic Publishers, Norwell, 1998).

Palmer, J. D., A single birth of all plastids?, *Nature*, **405**, 32-33 (2000).

Paquin, B., Laforest, M.-J., Lang, F., Double-hairpin elements in the mitochondrial DNA of allomyces : evidence for mobility, *Mol. Biol. Evol.*, **17**, 1760-1768 (2000).

Pfizinger, H., Weil, J. H., Pillay, D. T. N., Guillemaut, P., Codon recognition mechanisms in plant chloroplasts, *Plant Mol. Biol.*, **14**, 805-814 (1990).

Roger, A. J., Reconstructing early events in eukaryotic evolution, *Am. Nat.*, **154**, 146-163 (1999).

Shikanai, T., Endo, T., Hashimoto, T., Yamada, Y., Asada, K., Yokota, A., Directed disruption of the tobacco *ndhB* gene impairs cyclic electron flow around photosystem I, *Proc. Natl. Acad. Sci. USA*, **95**, 9705-9709 (1998).

Shub, D. A., Gott, J. M., Xu, M.-Q., Lang, B. F., Michel, F., Tomaschewski, J., Pedersen-Lane, J., Belfort, M., Structural conservation among three homologous introns of

bacteriophage *T4* and the group I introns of eukaryotes, *Proc. Natl. Sci. USA*, **85**, 1151-1155 (1988).

Sogin, M. L., Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs, *Am. Zool.*, **29**, 487-499 (1989).

Steinkötter J., Bhattacharya, D., Semmelroth, I., Bibeau, C., Melkonian, M., Prasinophytes form independent lineages within the Chlorophyta: evidence from ribosomal RNA sequences comparisons, *J. Phycol.*, **30**, 340-345 (1994).

Stoebe, B., Kowallik, K. V., Gene-cluster analysis in chloroplast genomics, *Trends Genet.*, **15**, 344-347 (1999).

Sugiura, M., The chloroplast genome, *Plant Mol. Biol.*, **19**, 149-157 (1992).

Sugiura, M., Hirose, T., Sugita, M., Evolution and mechanism of translation in chloroplasts, *Annu. Rev. Genet.*, **32**, 437-459 (1998).

Thompson, J. D., Higgins, D. G., Gibson, T. J., CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.*, **22**, 4673-4680 (1994).

Turmel, M., Côté, V., Otis C., Mercier, J.-P., Gray, M. W., Lonergan, K. M., Lemieux, C., Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion), *Mol. Biol. Evol.*, **12**, 533-545 (1995).

Turmel, M., Otis, C., Lemieux, C., The complete chloroplast DNA sequence of the green algae *Nephroselmis olivacea* : insights into the architecture of ancestral chloroplast genomes, *Proc. Natl. Acad. Sci. USA*, **96**, 10248-10253 (1999a).

Turmel, M., Lemieux, C., Burger, G., Lang, B. F., Otis, C., Plante, I., Gray, M. W., The

complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor* : two radically different evolutionary patterns within green algae, *Plant Cell.*, **9**, 1717-1730 (1999b).

Unsel, M., Marienfeld, J. R., Brandt, P., Brennicke, A., The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366 924 nucleotides, *Trends Plant Sci.*, **4**, 495-502 (1999).

Van Den Hoek, C., Mann, D. G., Jahns, H. M., *Algae : An Introduction to Phycology* (Cambridge University Press, 1995).

Viscogliosi, E., Philippe, H., Baroin, A., Perasso, R., Brugerolle, G., Phylogeny of trichomonads based on partial sequence of large subunit rRNA and on cladistic analysis of morphological data, *J. Eukaryot. Microbiol.*, **40**, 411-421 (1993).

Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A., Woese, C. R., Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes, *Nature*, **326**, 411-414 (1987).

Wakasugi, T., Nagai, T., Kapoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J., Nakashima, K., Tsudzuki, T., Suzuki, Y., Hamada, A., Ohta, T., Inamura, A., Yoshinaga, K., Sugiura, M., Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris* : the existence of gene possibly involved in chloroplast division, *Proc. Natl. Acad. Sci. USA*, **94**, 5967-5972 (1997).

Wolff, G., Plante, I., Lang, B. F., Kück, U., Burger, G., Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*, *J. Mol. Biol.*, **237**, 75-86 (1994).

Yamamoto, A., Hashimoto, T., Asaga, E., Hasegawa, M., Goto, N., Phylogenetic position of the mitochondrion-lacking protozoan *Trichomonas tenax*, based on amino acid sequences of

elongation factors 1alpha and 2, *J. Mol. Evol.*, **44**, 98-105 (1997).

ANNEXE 1

Séquence complète du génome chloroplastique de *Monomastix* OKE-1 en format GenBank

LOCUS MSPPCG 114528 bp DNA circular 30-AUG-2000
 DEFINITION Monomastix sp. chloroplast DNA, complete genome.
 ACCESSION MSPPCG
 VERSION
 KEYWORDS .
 SOURCE Monomastix sp.
 ORGANISM Chloroplast Monomastix sp.
 Unclassified.
 REFERENCE 1 (bases 1 to 114528)
 AUTHORS Gagnon,M.-C., Otis,C., Lemieux,C. and Turmel,M.
 TITLE Complete chloroplast DNA sequence from the Prasinophyte Monomastix
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 114528)
 AUTHORS Gagnon,M.-C., Otis,C., Lemieux,C. and Turmel,M.
 TITLE Direct Submission
 JOURNAL Submitted (30-AUG-2000) Biochimie et Microbiologie, Université
 Laval, Pavillon Charles-Eugene Marchand, Quebec, Quebec G1K 7P4,
 Canada

FEATURES Location/Qualifiers
 source 1..114528
 /organism="Monomastix sp."
 /organelle="plastid:chloroplast"
 /cell_line="Private collection of Charley O'Kelly"
 tRNA 1..73
 /gene="trnE(uuc)"
 /note="codons recognized: GAR"
 /product="tRNA-Glu"
 /anticodon=(pos:35..37,aa:Glu)
 gene 1..73
 /gene="trnE(uuc)"
 tRNA 491..562
 /gene="trnMe(cau)"
 /note="codon recognized: AUG"
 /product="tRNA-Met"
 /anticodon=(pos:523..525,aa:Met)
 gene 491..562
 /gene="trnMe(cau)"
 tRNA 659..731
 /gene="trnF(gaa)"
 /note="codons recognized: UUY"
 /product="tRNA-Phe"
 /anticodon=(pos:692..694,aa:Phe)
 gene 659..731
 /gene="trnF(gaa)"
 tRNA 735..819
 /gene="trnS(uga)"
 /note="codons recognized: UCN; 'super Wobble' codon
 recognition inferred, with U in the first
 position of the anticodon assumed to be unmodified"
 /product="tRNA-Ser"
 /anticodon=(pos:769..771,aa:Ser)
 gene 735..819
 /gene="trnS(uga)"
 gene 1522..2580
 /gene="psbD"
 CDS 1522..2580
 /gene="psbD"
 /codon_start=1
 /product="D2 reaction center protein of photosystem II"
 /translation="MTIAIGKSEEKRGWFDADDWLKRDRFVFGVSGILLPCAYFA"

```

LGGWFTGTTYVTSWYTHGLASSYLEGCNVLTAAVSTPANSMAHSLLEFLWGPEAGGDFT
RMCQLGGLWTFVALHGAFGLIGFMLRQFEIARAVQLRPYNAIAFSAPIAVFVSVFLIY
PLGQSGWFFAPSPFGVAGIFRFILEFFQGFHNWTLNPFHMMGVAGVLGAALLCAIHGATV
ENTLFEDGDGANTFRAFNPQSEETYSMTANRFWSQIFGVAFSNKRWLHFFMLFVFPV
TGLWMSAIGVVGLALNLRAYDFVVSQEIIRAAEDPEFETFYTKNILLNEGIRAWMAAQDQ
PHENLVFPEEVLPRGNAL"
repeat_region 2765..3118
                /rpt_family="A, A1, AC, BD, C"
                /rpt_unit=1 copy of A, A1, AC and BD, 4 copies of C
tRNA           3200..3288
                /gene="trnS(gcu)"
                /note="codons recognized: AGY"
                /product="tRNA-Ser"
                /anticodon=(pos:3234..3236,aa:Ser)
gene           3200..3288
                /gene="trnS(gcu)"
repeat_region 3613..4066
                /rpt_family="A, A1, A1D, AC, B, C"
                /rpt_unit=1 copy of A, A1, AC and B, 2 copies of A1D, 5
                copies of C
gene           complement(4171..4716)
                /gene="rps7"
CDS            complement(4171..4716)
                /gene="rps7"
                /codon_start=1
                /product="ribosomal protein S7"
                /translation="MSRRKTAKKRTITPDPPIYRSRLVSMVSRLLQDGKKSVASRLFY
                ESLDQITSGLGKKPASGAVNTNAQNPKEAPAVTQTATDALKVFQQAVLHATPVVEVKS
                RRVGGSNYQVPLEVNSERGTALALRLVQAARQRPGREMATKLAQEFLDASNKAGGAV
                RKREETHKMAEANKAFANYRF"
gene           complement(5049..5438)
                /gene="rps12"
CDS            complement(5049..5438)
                /gene="rps12"
                /codon_start=1
                /product="ribosomal protein S12"
                /translation="MFTIQQLVRSERTKREKKTAPALKTCPPRRGVCTRVYTTTPKK
                PNSALRKVARVRLTSGFEITAYIPGIGHNLQEHVVVLRGGRVDRDLPGVRYHIIRGTL
                DAAGVKDRKNSRSKYGATKASTTAAKK"
repeat_region 5531..5869
                /rpt_family="A1D, B, B1, BD, C, D"
                /rpt_unit=1 copy of A1D, B, B1 and D, 2 copies of BD and C
tRNA           6055..6127
                /gene="trnD(guc)"
                /note="codons recognized: GAY"
                /product="tRNA-Asp"
                /anticodon=(pos:6088..6090,aa:Asp)
gene           6055..6127
                /gene="trnD(guc)"
repeat_region 6195..6211
                /rpt_family="B"
                /rpt_unit=1 copy of B
repeat_region 6802..6906
                /rpt_family="A, B"
                /rpt_unit=1 copy of A and B
gene           6971..9226
                /gene="psaA"
CDS            6971..9226
                /gene="psaA"
                /codon_start=1
                /product="P700 apoprotein A1 of photosystem I"
                /translation="MTISPPNTTKVKIVVDRNPVETSFEKWARPGHFSRTLSKGPPTT
                TWIWNLHADAHDFDSHTDLEDISRKVFSAHFGQLGIIILWLSGMYFHGARFNYEAW
                LSDPTHIKPSAQVWVPIVQGEILNGDVGGGFQGIQITSGFFQLWRGSGITSELQLYST
                AIGGLILAGLMEFFAGWFHYHKAAPKLEWFQNVESMMNHLAGLLGLGSLAWAGHQIHV
                SLPVNALLDAGVDAKEIPLPHEFILNRALMAEIIYPSFARGLAPFFTLQWSEYSDFLTF
                KGGLNVPVTGGLWLTDTAHHHLAIAVLFIIAGHQYRTNWGIGHSMEKILEAHKGFPTGE
                GHKGLYEILTTSWHAQLGLNLAFFGSLSMIVSHHMYAMPYPYLATDYGTQLSLFTHH
                MWIGGFCVGTAAAHAAIFMVRDYPDTNNYNNVLDRIHRDAIISHLNWCIFLGFHS
                FGLYIHNDTMSALGRPQDMFSDTAIQLQPVFAQWIQNTHSLSISVTAPNALGSTSQS
                WGGDIVAVGGKIAMPISLGTSDFMVHHIHAFTIHVTVLILLKGVLFARSSRLIPDKA

```



```

NLGFRFPDGPGRGGTCQVSAWDHVFLGLFWMYNSLSVTIFHFSWKMQSDVWGS LGST
GTVSHITGGNFAQSANTINGWLRDFLWAQSSQVIQSYGSALSAYGLMFLGAHFVWAFS
LMFLFSGRGYWQELIESIVWAHNKLRVAPSIQPRALSITQGRAVGVVAHYLLGGIATTW
SFFLARIIVG"
repeat_region 9231..9249
                /rpt_family="B"
                /rpt_unit=1 copy of B
gene           9493..11700
                /gene="psaB"
CDS           9493..11700
                /gene="psaB"
                /codon_start=1
                /product="P700 apoprotein A2 of photosystem I"
                /translation="MATKFKFSQGLAQDPTTRRIWFGIATAHDFESHDMTEEKLYQ
KIFASHFGQLAIFLWTSGNLFHVAWQGNFEQWADPLHVRPIAHAIWDPHFQGPVE
AFTRGGASTPVNIAYSGVYQWYITIGMRTNVLDYNGSVFLLGAAAF LFAGWLHLQPQ
FQPSVSWFKNAESRLNHHLSGLFGVSSLAWTGHLVHVAI PESRGVHVHVNMLLTLPH
PQGLTPFFTGNWAA YAGSPDTATHLFGTSEGAGTALLTFVGGFHPQTQSLWLS DIAH
HLAIAVLFIIAGHQYRTNFGIGHSMKEILEAHVAPAGKLGSGHKGLFDTVNNLSHFQL
GLALASVGVITSLVAQHMYSLPAYAFLAQDFTTQAALYTHHQYIAGFMCGAFAHGA
FFIRDYDPELNKGNV LARMLLEHKEAII SHLSWASLFLGFHTLGLYHVHNDVMLAFGTPE
KQILIEPVFAQWIQA AHGKSLYGFVLLSSAGGAA YAAGEGIWLPGLD AINNAGANS
LFLTIGPGDFLVHHAIALGLHTTTLLILVKGALDARGSKLMPDKDFGYSFPCDGPGRG
GTCDISAWDAFYLA VFWMLNTIGWVTFFYHFKHITLWQGNPAQDESSTYLMGWLRDY
LWLNSSQLINGYNPFGMNSLSVYAWLFLFGHLVYATGFMFLISWRGYWQELIETLAWA
HERTPLANLVKWKDKPVALSIVQARLVGLVHFSAGYILTYA AFLLASTSGKFG"
repeat_region 11789..12045
                /rpt_family="A, B, B1, BD C"
                /rpt_unit=1 copy of A, B, B1 and C, 2 copies of BD
gene           complement(11961..12164)
                /gene="orf67"
CDS           complement(11961..12164)
                /gene="orf67"
                /codon_start=1
                /product="unknown"
                /translation="MLLISIFSLEKTSDFLIVCLNLIYNLKGPEGQRPQGPVKVFLR
GKGVSALTFQVINKAHPLETS GP"
repeat_region 12547..12589
                /rpt_family="BD"
                /rpt_unit=2 copies of BD
repeat_region 12821..12839
                /rpt_family="B"
                /rpt_unit=1 copy of B
gene           complement(12870..13256)
                /gene="rp120"
CDS           complement(12870..13256)
                /gene="rp120"
                /codon_start=1
                /product="ribosomal protein L20"
                /translation="MTRVRRGFVARKRRRVLK LKAGFRGSHSVLFRTAQQRTMRALS
LSYTDGRQFKRAMRRLWIRLNSALRNIGIQNTATNTNSKIPVVF RYSGFINNLKKA
SLLNRKVLSQLAILDPTSFQNFVKGL"
repeat_region 13368..13752
                /rpt_family="A, AB, AD, B, B1, BD, C, D"
                /rpt_unit=1 copy of A, AB, AD and C, 2 copies of B, B1, BD
                and D
gene           complement(13905..15134)
                /gene="tufA"
CDS           complement(13905..15134)
                /gene="tufA"
                /codon_start=1
                /product="translational elongation factor Tu"
                /translation="MARAKFERKKPHVNVGTIGHVDHGKTTLTAAITMTLAAIGGSVG
KKYDDIDSAPEEKARGITINTAHVEYETERHYAHVDCPGHADYVKNMITGAAQMDGA
ILVVSAAADGMPQTR E HILLAKQVGVNVLVFLNKEDQVDDPELLELDVMEVRETLS
YEFFGDEIPIVPGSALLALEAVVESPNITRGKNKWDKIYQLMDSVDTYIPTARETD
KPFMLAVEDVFSITGRGT VATGRVERGVVKGDSVEVVGLKTTKTTT V TGLEMFQKTL
DESVAGDNVGI LLRGIQKTDIERGMVLSAPGSITPHTKFESQVYIILTKEEGGRHTPFF
PGYRPQFYVRTT DVTGKIENFKADDGSESSMVMPGDRVKMTVELINPVAIENGMRFAI
REGGRTVGAGVVSQILA"
repeat_region 15308..15832

```

```

/rpt_family="A, A1, A1D, B1, C, D"
/rpt_unit=1 copy of A1 and A1D, 2 copies of A, B1 and D, 3
copies of C
tRNA complement(16000..16071)
/gene="trnR(ucu)"
/note="codons recognized: AGR"
/product="tRNA-Arg"
/anticodon=(pos:16036..16038,aa:Arg)
gene complement(16000..16071)
/gene="trnR(ucu)"
tRNA 16427..16507
/gene="trnL(uag)"
/note="codons recognized: CUN; 'super Wobble' codon
recognition inferred, with U in the first position of the
anticodon assumed to be unmodified"
/product="tRNA-Leu"
/anticodon=(pos:16460..16462,aa:Leu)
gene 16427..16507
/gene="trnL(uag)"
repeat_region 16641..16678
/rpt_family="A"
/rpt_unit=1 copy of A
tRNA 16805..16886
/gene="trnY(gua)"
/note="codons recognized: UAY"
/product="tRNA-Tyr"
/anticodon=(pos:16839..16841,aa:Tyr)
gene 16805..16886
/gene="trnY(gua)"
tRNA 16948..17021
/gene="trnP(ugg)"
/note="codons recognized: CCN; 'super Wobble' codon
recognition inferred, with U in the first position of the
anticodon assumed to be unmodified"
/product="tRNA-Pro"
/anticodon=(pos:16982..16984,aa:Pro)
gene 16948..17021
/gene="trnP(ugg)"
tRNA 17152..17222
/gene="trnC(gca)"
/note="codons recognized: UGY"
/product="tRNA-Cys"
/anticodon=(pos:17184..17186,aa:Cys)
gene 17152..17222
/gene="trnC(gca)"
tRNA complement(17330..17400)
/gene="trnG(ucc)"
/note="codons recognized: GGR"
/product="tRNA-Gly"
/anticodon=(pos:17366..17368,aa:Gly)
gene complement(17330..17400)
/gene="trnG(ucc)"
repeat_region 17407..17762
/rpt_family="AB, AC, AD, B, BD, C"
/rpt_unit=1 copy of AB, AC, AD and C, 2 copies of B and BD
tRNA complement(17877..17949)
/gene="trnW(cca)"
/note="codon recognized: UGG"
/product="tRNA-Trp"
/anticodon=(pos:17914..17916,aa:Trp)
gene complement(17877..17949)
/gene="trnW(cca)"
tRNA join(18195..18217,18758..18806)
/gene="trnK(uuu)"
/note="codons recognized: AAR"
/product="tRNA-Lys"
/anticodon=(pos:18767..18769,aa:Lys)
gene 18195..18806
/gene="trnK(uuu)"
exon 18195..18217
/gene="trnK(uuu)"

```

```

/number=1
intron 18218..18757
/gene="trnK(uuu)"
/note="group II intron"
/number=1
exon 18758..18806
/gene="trnK(uuu)"
/number=2
gene 19569..19655
/gene="petN"
CDS 19569..19655
/gene="petN"
/note="formally known as ycf5"
/codon_start=1
/product="subunit VIII of cytochrome b6/f complex"
/translation="MVYIFSWVFVILFFYSSIDLSKIARSLT"
repeat_region 19826..20140
/rpt_family="A, A1, AC, AD, BD, C"
/rpt_unit=1 copy of A, A1, AC, AD and BD, 5 copies of C
gene 20417..21844
/gene="rbcL"
CDS 20417..21844
/gene="rbcL"
/codon_start=1
/product="large subunit of Rubisco"
/translation="MSPQTETRTSAGFKAGVKDYRLTYTTPDYQVVDTDVLAAFRMT
PQGVFPPEEAGAAVAESSTGTWTVMWTDGLTSLDRYKGRCYDIEPVPGEDNQYIAYVA
YPIDLFEESVNTLFTSIVGNVFGFKALRALRLEDLRIQAYVKTFQGGPHGIVVERD
KLNKYGRGFLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVTSQPFMRWRDRF
LFVAEAIYKSAQETGEIKGHYLNVTAGTAEMLKRAQFAKDLGVPIIMHDYLTAGFTS
NTSLSHYCRDNGLLLIHHRAMHAVIDRQRNHGIFRVLAKALRMSGGDHLHSGTVVVGK
LEGERDITLGFVDLMDRAVVEKDRSRGIYFTQDWCGLAGTMPVASGGIHWHPALVE
IFGDDACLQFGGGTLGHFPWGNCPGAVANRVALEACTQARNEGRNLAREGGDIIRAACK
WSPELAAACEVWKEIKFEFQAMDTL"
repeat_region 22017..22155
/rpt_family="A, AD, B, BD"
/rpt_unit=1 copy of A, AD, B and BD
gene complement(22293..22802)
/gene="ycf3"
CDS complement(22293..22802)
/gene="ycf3"
/codon_start=1
/product="hypothetical chloroplast RF3"
/translation="MPRSQRNDNFIDKTFVTIADILLQVLPTTKREKEAFTYYRDGMS
AQAEGEYAEALQNYEALRLEVDAYDRSYILYNIIGLIHTSNGEHTKALEYFYQALESN
PSLPQALNNIAVIYHYRGEQAIEAGDTQASEILFEKAADYWREAIRLAPTNYIEAQNW
LARGWGEKS"
repeat_region 22882..23254
/rpt_family="A, AC, AD, B, B1, BD, C"
/rpt_unit=1 copy of A, AC, B, B1 and BD, 2 copies of AD, 6
copies of C
gene complement(23477..26977)
/gene="rpoC1"
CDS complement(23477..26977)
/gene="rpoC1"
/codon_start=1
/product="beta' subunit of RNA polymerase"
/translation="MKTFMQIGLASSEEQSWAERSLWLDGNEIKIGEVKKADTINYR
SFKPEKDGLFCERIFGPVKDWVCACGQSKLVGKQNPATANLSKNTGTEFFPRDPL
FSPVVRKAGLEGFPSSRVAGEGELISPFQFRQSENSLLNSGLPNLSVDKSNWAPQSP
TQEAQRPOVGGFLFIEGATKESTKSDENPNSSPQFSNREKFSQVGLGPHYICPNQVQV
LSRVRRYRMGYIELGCPVTHIWLNSRPNIFSVLLKIPTKYVKKITYRYGYSQAQKEF
ELSQEREQDLPLLSKMGKRPSTYTEKPHRLETKLQKELFVFGVGNPCVGGGDF
DNEWEFVHWFSLPRSSKPLPKGQTFPRDPLEDRRSPPIRGEAFGPHNAKKEKRAKLE
GGLWPFPGPFSFESGLILQNHKLGANTTDTELELNFSEKRTLRGGTKFTELSSSKLQK
SVKNVSWFSEKQALDSIAQRDQNI VPI NPSWLENTGARAFKEHFKTRNWKKEFEKLDQ
DLRRTPVFFSES IKETKTEFAEEGSYHVSGETESFSNENIWSLTPSFSMYENLLKRRK
RVRTLKLLNLLRTAGGAQSFDFGFI FTTL PVL PPELRPIVQLSAGQFASSDVNDLYRRL
ISRNRLKYYFTACGPHLVEFLVRSEQHLVQVAVDSLLDGSSAAEIGSLRRMSNSGS
GGGNKPLTKGNSLPYKSLSDRIGGKQGRFRQNLGKRVDSYSGRSVI VVGGPRLKLEHC
GLPYEMALELFOAFVIRHILELQLAKTIRGAKNLLKLNKPFARQILQNVVQSHPI LLN

```

```

RAPTLHRLGIQAFQPKLIYGRAIQLHPLVCSAFNADFDGDOMAVHVPLSPKARVEARL
LMLATTNWLSAATGQPSVLPQDMVLGFYYLTTLQPKTQIEFNLSLNERGGVNIPEGE
RSSQNFVFWKKSRAPAVNEIGDQNNERRQKQKVVDDKNSLLFHFIIEVENGPSQRNGL
DLHQPLWLQKTNLWPGEKRNWAGKQSIQVIANQNSGGTFFPRDPRGEPFEAFGPHN
GRKERVAAEEELRSSKTFEKQKAFLLNKAPGVWNGNANLKNQIFGENNANKIFNIEPVP
GENHQDKMNYSSVVGIEGGKGFNEPLYIRIFSSGFANQVYHSYKWTQDTKAYRRNKAI
RTTPGRVLVNEILQKLGDK"
repeat_region 23893..23952
                /rpt_family="B, BD"
                /rpt_unit=1 copy of B and BD
repeat_region 25850..25869
                /rpt_family="B"
                /rpt_unit=1 copy of B
repeat_region 27073..27670
                /rpt_family="A1, AB, B, B1, D"
                /rpt_unit=1 copy of A1, AB and B1, 2 copies of B and D
gene 27351..27719
     /gene="orf122"
CDS 27351..27719
     /gene="orf122"
     /codon_start=1
     /product="unknown"
     /translation="MWSFQVRYQEGGSPSFRILLDTVKRFCAVFSNLRGTGGLKDQR
TRGLKGVSFYELKQSAHPLEDRRSPPIRGEVFFPLRGKGVSFYELKGPFEFFPRDPP
PKGGGSRGKDRLVGIFGFQL"
gene complement(27776..31498)
     /gene="rpoC2"
CDS complement(27776..31498)
     /gene="rpoC2"
     /codon_start=1
     /product="beta'' subunit of RNA polymerase"
     /translation="MNFYNQVLDKGLKRIVSWFIENYGPTRTSLLLDDFKYIGFHVA
GKAGLSLGFDDLRIPAARKLLLTGAQKQVQCEKTYLMGRITAFERYQKLIDIWTTTS
EKLKDEVIQNFQKIQQGGPLASPLYMMAFSGARGNISQVRQLVGMRLMSDSSGGIIDF
PIRSNFREGLSVTEYVISCYGARKGLIDTALRTADSGYLTRRLVDVAHGIIIREIDCG
TSESIILMKISDKKNIIGRVLAESIKERTPTIWKNEI'SPNLALELKTLLKGSKASIQVR
SPLTCLSQSVCQLCYGNWLAQGRVALGDVAVGILAAQSIGEPGTQLTMRFTHTGGIFS
TEVEDKIYAPAGGTIWLQMGPPKTAIILSESQSQDKNHQSFFGSTHFLSASVGCKIRS
VHGQTAFVFFFAIQLKIEEEKGGTLQDINNQNNTISILQLPAQSILFITPGKTIARN
TLCAEISIKIQPAEKNEELTSDIPTLGLSIQKGGKATTQRGGSPLNNDGENSDGSGLA
QKKVFSEIQGEVVETDYFDLKKFTKPKLGGGNEVLNPLFEKGLKATSPKANPQPKGEG
YKPEIGTRTSSKARLVLAGEPFSFKRTKSAFAEEIQAGNKSQRPSKAKLSQLPFRSGDI
VFWRGKRLLSSKHFILAKDQKSPDNKLRALFLNTLSSESNSKDEIVNGKHFPMNEND
KGPNSLRIGKQGENSGTPLKGEAFGSWLQTSVNILKNKGTENSSGILTLLQKDFNQ
GGEQSSLSAFGEGKETKTKPWLNEESYSLTKTNSVHFVNPSFEIRKTNLQKKFHEN
LSFSGLVLEFSQKQEQRNGVLPTRPSRGTLRGLWPSLSNCEGQRPPPNWGEHRRPQG
ETLEGQRPPPNWGEGLVGRRTNRAKNLSLTSPLYSMLENTFSLSHLQKILQPYTFFV
GLPPAINVPTQIIRKRLNKLILRRVQTSLGGQFQGNVSRMSGNSPNTSEKQRTYI
KERDLLFQLVLPADSASSGDIVQGLPKIEQLFEARSANESSVQNTKLLANKTEKKS
INKLSKQEKFKNLQKPKQEQTSVFNENVKGLEEKTSNGPHNSIEGVTGEELRTLK
KTELRELQKRLIQEIQSVYNSQGVETADKHLEIIVRQMTSRVRIKGNTPFPYDDII
EAFGSNTLLSKEIQKSPNVLSDENSGFSYEWILLGITKVALSTTESYISAASFQETK
RVLMTAALQSRVDFFEGLKENVIVGRMIPAGTGYTS"
repeat_region 28887..29035
                /rpt_family="B, BD"
                /rpt_unit=1 copy of BD, 2 copies of B
repeat_region 31584..31998
                /rpt_family="A, A1, AD, A1B, BD, C, D"
                /rpt_unit=1 copy of A, A1, AD, A1B and BD, 2 copies of D,
                8 copies of C
gene complement(32170..38496)
     /gene="rpoB"
CDS complement(32170..38496)
     /gene="rpoB"
     /codon_start=1
     /product="beta subunit of RNA polymerase"
     /translation="MIDSFHKYPLFPDLENQRRSVRSFWEKGILDELQLFSTIIGYQ
QDKNTSKTNRATISSEILLEKSSSVQAFSGEGSMNQGF'PKLNGGSSVGNFEGPKDRRS
EGAPQLFPHSGEGGEPFDLSGRNKAHHLQAFGPPLETSGPSVHWFNQESICFSSSLI
PRKSMAYRQKSFKTQNNSENKLESFLEFSGAKFSLSKQRQSGSSSLINDLNRKKT
QNLQNFVPGTIPFSSKKNLDLYLVLHGSKFLIKSPQYSQIDAIRYQKTYSVGFPLPIE

```

```

KINQTONKFEDMTVLNQVRNTASLSDFRSENRSLESFAKSNIQLSIRKTQLKSVDNSS
SSSIFGVFDSNNKLNMTNHFVERSTRQKKQIFWYFGDIPLMTERGSFIINGAPRVLV
NQIIRCPSVYFKVKLDQKNRRSYTASFLSEYGSWLRLETDRLSNRLWARIDKSPRFFA
DILLMGLGYVLDSENLNSVKYGLSKNEFKLNAQTLSEFGKENKVVINKNIVKNSIFETA
STLNSTFPASPFTSLKKYLTAKYIQVYQENNNLPTRQHDEREVQRPPRNGGEGSVGRT
PFLDWIGKGLSKGLQSTAGEKQFAKLLNKVENKKISKEQNYLWSSVSARLAEPKISSI
QSPRPVLSGSIGKKFSSTKNFVFKNTVENDNTRTPRRTNAGSKSGITINTATALQALW
KCNPNGRWTSALGCYSFFYNKFLNPKRYSIGRVGRIRLNKRLRRNPECKIPTLTPEDV
LLAFHYLTELKIADAKENTFSLELDDIDHLKNRRVRLPGEILQNFRLALSRI PYSAI
RLQDNKTNIRNEVLQREKVVQPGVPLEGRVGRSPSPSGEARKVEQNQNSIRKNFDSV
SSRSSEDSVPPGIPAFQSTLREL FNTGQLSQYMDQTNPLAEITHKRLSSSLGPGGVGR
DQAGFAVREIHPSHFGRICPIETPEGQNAGLVGLSASARINDDGFLQSPALPIENQK
SRFFPNNNFEGDWLGFNSASARKYSKYILQNGPGKEIGLSSFQMNKANPRKNGGGSEK
LYSLEQSSLLKADDSFAVRQNISLFSSELEDEIYVCTGDTLNPCLLNTSKIVANTSLF
PPVFPVRHKQEFILTINSTDALQFIQAPTCLFTGICPIQLISIATSLIPFLEHDDANRA
LMGSNMQRQAVPMTPEKAFVGTGLELQAARDSSTLVLADESGQILFVDSQSILKLTE
KHGRTFSELRSPPGTGKQRPQWGLSLDLKQSAHPIEFSPERGGFSKEISSYELTDF
TPSNQSTCLKQRAVEI GEWVEKGSLLADGSGTRDGEVALGKNVLLAYMPWEGYNFED
AIVINERLVFEDIYTSVHVERFDVDRNTKIIQVRNPENSEFQLEDLTPENSILNIFRS
LVPGEYLTRDLDIVTNRKTSRNSNGENSTLKTSTKDSLEALEINLPPKLPKLRHLDA
NGI IKPGTWVQEGDILLGKLQIGPKDNLNNFDIKAQNTPFAQKIQTLPFPEYRLLAI
FDLQNRTPDKTLPQTSTNLEGGKLPWPTTFGEKLSKESTLPSQDKKEQIIIGEKSLSS
EEGERSPLSDLSAYKNTLPSQNFSGELTSGQDTLNLAAEKSSNSGLFAHQIKNTSFK
VGI GVKGRVLDYVVDGRISIFDGRQKANQNLPPRGSRET PPGLWVRSQREEREGRGGRT
PFLEKHKTLIGEGFIQSAATSVSIFLAHKKRIQLGDKMSGRHGNKGIVSLILPPQDMP
YLQDGKPVDDVVLNPLGVPSRMNVGVLETLFGLSAFYSHKVVYRIMPFERIAPSQSI E
KNTLQENASLSLRPETSRSLSVYGHLPRLKTKAGVPPKRAEISELGLQKGGYPRTLKS
IKKGSQLNWLFDPNPNPGKTHLFDGRTGEIFYQPALVGVSYMFKLIHLVDDDKIHARSTG
PYSLVTOQPLGGRSKKGQRLGEMEVWALEGFGTASILQEFLTVKSDIYSRNNFLFH
LMRRNWESLEKNENESFARDAIKARVAGEELRSLTKFSENQEPKTS PFGGGPKQRPE
GGELYFDLKGQRAHPLQVFGNTTEPRMNKQNGDIHLVLDLVPVGFDRFSNMEGPEQNLKV
LPETS NFSKVPESFRVLINELHALCLS VYSNLPS"
repeat_region 32398..32489
/rpt_family="AB, B"
/rpt_unit=1 copy of AB and B
repeat_region 33530..33567
/rpt_family="BD"
/rpt_unit=1 copy of BD
repeat_region 34752..34826
/rpt_family="AB, B"
/rpt_unit=1 copy of AB and B
repeat_region 35983..36049
/rpt_family="B, BD"
/rpt_unit=1 copy of B and BD
repeat_region 36817..36836
/rpt_family="B"
/rpt_unit=1 copy of B
repeat_region 38114..38205
/rpt_family="A, B, BD"
/rpt_unit=1 copy of A, B and BD
repeat_region 38563..39457
/rpt_family="A, AB, AC, AD, B1, BD, C, D"
/rpt_unit=2 copies of A, AB, AC, B1 and BD, 3 copies of D,
4 copies of AD, 14 copies of C
repeat_region 39849..39866
/rpt_family="B"
/rpt_unit=1 copy of B
gene complement(40039..40155)
/gene="rpl36"
CDS complement(40039..40155)
/gene="rpl36"
/codon_start=1
/product="ribosomal protein L36"
/translation="MKIRSSVRKICENCRVIRDRKVMICQNPKHKQKQGS"
tRNA complement(40722..40810)
/gene="trnI(cau)"
/note="codon recognized: AUA; C in the first position of
the anticodon assumed to be post-transcriptionally
modified to lysidine, which pairs with A rather than G"
/product="tRNA-Ile"
/anticodon=(pos:40774..40776,aa:Ile)

```

```

gene      complement(40722..40810)
          /gene="trnI(cau)"
repeat_region 40986..41246
          /rpt_family="A, AD, B, BD, C"
          /rpt_unit=1 copy of A, B, BD and C, 2 copies of AD
gene      complement(41433..42038)
          /gene="clpP"
CDS       complement(41433..42038)
          /gene="clpP"
          /codon_start=1
          /product="proteolytic subunit 2 of clp protease"
          /translation="MPIGVPKVAFRLPGEPTPQWIDLYNRLYRERVFLGSDLDDELA
          NQLVGIMIYLSAEDESKRIFIYINSPGGSVTCGLAVYDAMHHIQARVTTLCIGIAASM
          ASFVLGGERGHRJALPHSRVMIHQFQGGSQGASDVTAEAEVEMRIRRQIGRIYAER
          TGQSLATIAKSMRDRFMSAKEAKEFGLVDQVASSGSENRV"
repeat_region 42320..42459
          /rpt_family="AD, BD, C"
          /rpt_unit=1 copy of AD and BD, 4 copies of C
gene      43090..43467
          /gene="orf125"
CDS       43090..43467
          /gene="orf125"
          /codon_start=1
          /product="unknown"
          /translation="MNKELVCLSFYIDLKAPRSLVRLSFYIDLVCLLSALTFFQVIHTAH
          PLPPKGEELGGPFPRPSVLKGVSAITFQVIHTAHLPPKGEDLWSLRPLVLWSLRPLAL
          RSSKVTKLQGTGSHSSTPGRAH"
repeat_region 43112..43453
          /rpt_family="A, B1, BD, C, D"
          /rpt_unit=1 copy of BD, 2 copies of B1, 3 copies of A, C
          and D
gene      complement(43494..43607)
          /gene="petG"
CDS       complement(43494..43607)
          /gene="petG"
          /codon_start=1
          /product="subunit V of cytochrome b6/f complex"
          /translation="MVEPLLSGIVLGFIPVTIAGLFVTAYLQYRRGDQLNL"
gene      complement(43724..45247)
          /gene="atpA"
CDS       complement(43724..45247)
          /gene="atpA"
          /codon_start=1
          /product="CF1 alpha subunit of ATP synthase"
          /translation="MVKIRPDEISSIIRQQIESYNQEVKISNVGTVLNVGDGIARIYG
          LEKVMAGELLEFDQGTVGIALNLETDNVGVLMGDGLKIQEGDSVKATGKIAQIPVGE
          AFLGRVVDALARPIDGKGAIKADATRLIESPAPGIISRRSVYEPLCTGLVAIDAMIPI
          GRGQRELIIGDRQTGKTAVATDTILNQNGENVVVCVVAIGQKASSVAQIVTTLQERGA
          MKYTIIVAENADAPATLQYLAPYTGAAALAEYFMYTGRHTLVIIYDLSKQAQAYRQMSL
          LLKRPPGREAYPGDVFYLSRLLERAAKLSNALGEGSMTALFVIETQGGDVSAIYPTN
          VISITDQIIFLSADLFNSGIRPAINVGVISVSRVGSAAQIKAMKQVAGTLKLELAQFAE
          LEAFSQFASDLQATQNLARGVRLRELLKQSQSAPLRVEEQVVTIYTGIKGYLDKLA
          VDQVRDFLVGLREYIATNKTKLYETIRTTKAFGPDAAETLLKEAIAEYTIQIFLSSARK"
gene      complement(45449..45967)
          /gene="atpF"
CDS       complement(45449..45967)
          /gene="atpF"
          /codon_start=1
          /product="CF0 subunit I of ATP synthase"
          /translation="MAAESFGFNFDIETNIINLAVVIAVVLYLGGDVLTSLLRDRKQ
          KILSTIQSAEERYLEAQQKLEQAKQKVEIAKTKAIEIRQQGNLNAQAQTKLTFERTEE
          EIRRLEETSISGPQSTLRIIEGKAFQVQRQVVKLSIERAFTLLRKTAEASSQKRII
          DFNIALLGKMNPN"
gene      46048..46284
          /gene="orf78"
CDS       46048..46284
          /gene="orf78"
          /codon_start=1
          /product="unknown"
          /translation="MTKCYQNSVGALKNFLFEYFLGTPKKIGLLVNKHTAHRFEDFLK
          SLFFNKFKSLPPRERELRGSFRPSVLPSPGEGF"

```

```

repeat_region 46227..46383
               /rpt_family="A, B, BD"
               /rpt_unit=1 copy of A, 2 copies of B and BD
gene          complement(46841..47089)
               /gene="atpH"
CDS           complement(46841..47089)
               /gene="atpH"
               /codon_start=1
               /product="CF0 subunit III of ATP synthase"
               /translation="MNPLIGAASVVGAGLAIGLGAIGPGIGQGTAAGQAVEGIARQPE
AEGKIRGTLTLLSLAFMEALTYGLVVALALLFANPFAG"
repeat_region 47218..47557
               /rpt_family="A, AD, A1D, B, C"
               /rpt_unit=1 copy of AD, A1D and B, 2 copies of A, 5 copies
               of C
repeat_region 47930..47946
               /rpt_family="B"
               /rpt_unit=1 unit of B
gene          complement(47965..48684)
               /gene="atpI"
CDS           complement(47965..48684)
               /gene="atpI"
               /codon_start=1
               /product="CF0 subunit IV of ATP synthase"
               /translation="MLYDIAAVEIGQHWYWSVGGFQVHGQVLVNTWIVFAIILIVGFT
TTRQTLQVANLSGGQNFIEFFMEYMRDLASTQIHGADSKRWVPYLSLFLFIFVSNW
SGALVPWRIFEIPNGELGAPTNDINTTVALALLTSVAYFYAGLTKKGLNYFGRYVKPT
PILFPINVLEDFTKPLSLSFRLFGNILADELVVAVLITLVPLVIPILFLGLFTSAI
QALIFATLAGAYIGEALDHH"
gene          48734..48922
               /gene="orf62"
CDS           48734..48922
               /gene="orf62"
               /codon_start=1
               /product="unknown"
               /translation="MLRGKAQNLIFFPQTYAPPSGENFVPERLPQGPQGRGETFGPQG
GERFDLVCFSALTFEVRY"
repeat_region 48780..49350
               /rpt_family="A, A1, B, B1, BD, C, D"
               /rpt_unit=1 copy of A1, B, B1, BD and C, 3 copies of A and
               D
gene          complement(49473..49886)
               /gene="atpE"
CDS           complement(49473..49886)
               /gene="atpE"
               /codon_start=1
               /product="CF1 epsilon subunit of ATP synthase"
               /translation="MTLKVCIMAPDRVFNASAEIILPTNTGQMGVLTNHTPLITGL
DIGVMVVRPDKNTAPLAMLGGFALIQENTVTILVNEAELGSSINPTEAAAFTEAQ
SALEKAQNAGRKEKIEASLAFKRARARMQVLGQNR"
repeat_region 49993..50204
               /rpt_family="AB, B, BD, D"
               /rpt_unit=1 copy of AB, B and D, 2 copies of BD
gene          complement(50246..51694)
               /gene="atpB"
CDS           complement(50246..51694)
               /gene="atpB"
               /codon_start=1
               /product="CF1 beta subunit of ATP synthase"
               /translation="MNVQVKNEGRITQIIGPVMDVVFSSGNIPNIYNALVVEGKNEAG
DDLKITAQVQQLLGDNKVRAVALSATDGSMRGMKVIDTGAPLSVPVGEVTLGRIFNVL
GEPVDELGPVSTGKVKTLPIHRTAPAFVDLDTKLSIFETGIKVVLDLAPYRRGGKIGL
FGGAGVGKTVLIMELINNIKAHGGVSVFVGGVGERTRGNOLYEMKESKVINTEKMS
ESKVALVYQMNPEPPGARMRVGLTALTMAYFRDVNNQDVLFFIDNIFRFVQAGSEVS
ALLGRMPSAVGYQPTLSTEMGSLQERITSTRDGSITSIQAVYVPADDLTDPAPATFFA
HLDATTVLSRNLAAKGIYPAVDPLDSTSTMLQPWIVGDDHYKAAQSVKRTLQRYKELQ
DIIAILGLDELSEDDRLTVARARKVERFLSQFFVAEVFTGSPGRYVSLSEAIQGFKL
ILTGELDSLPEQAFYLVGNIEEAVERANKLAS"
repeat_region 51938..52206
               /rpt_family="A, AB, B1, BD, C, D"
               /rpt_unit=1 copy of A, AB, C and D, 2 copies of B1 and BD

```

```

gene      complement(52627..52737)
          /gene="psaI"
CDS       complement(52627..52737)
          /gene="psaI"
          /codon_start=1
          /product="subunit VIII of photosystem I"
          /translation="MLSLYLPSILVPLVGLVFPAIAMASLFVYIEKDQIS"
gene      complement(52742..52858)
          /gene="psbI"
CDS       complement(52742..52858)
          /gene="psbI"
          /codon_start=1
          /product="I polypeptide of photosystem II"
          /translation="MLTLKIFVYTVVIFVSLFVFGFLSNDPGRNPKSKDFE"
repeat_region 53246..53301
          /rpt_family="AB"
          /rpt_unit=1 copy of AB
repeat_region 54270..54372
          /rpt_family="B, B1, D"
          /rpt_unit=1 copy of B, B1 and D
gene      complement(54449..55261)
          /gene="rps2"
CDS       complement(54449..55261)
          /gene="rps2"
          /codon_start=1
          /product="ribosomal protein S2"
          /translation="MPIEELLEAGVHFGHRVNKWNPKMAPYIFGERNGVHIIDIQT
          LGYLEESYLFLSKYTVKALRNSSSPSNPKTFGEGIGQKVLVFGTKKQVASIVQAEAE
          NSHANYVNRWLGGLLTNWSTIKLCIAKLTLRASGLFENTEISSVNNSTKTSLVTTK
          KEKALLKKQYELKKFFSGIENMSDIPDIVIIVGQDCEMNAVKECQKLASSWRSDTNS
          TKGGNALPRTITILDNTCDPSLADLFIPANDDSTKSVELILKKLGEAIRIPR"
gene      55418..55615
          /gene="orf65"
CDS       55418..55615
          /gene="orf65"
          /codon_start=1
          /product="unknown"
          /translation="MVLLSLGFHNLKKTITREELCCLESLGPKASRTKGPEVSRTK
          GPEVSRTKGLPPSGEGGELYL"
repeat_region 55501..55952
          /rpt_family="A, A1, AB, B, B1, BD, C, D"
          /rpt_unit=1 copy of A, AB and BD, 2 copies of A1, B and C,
          3 copies of B1, 4 copies of D
tRNA      complement(56278..56350)
          /gene="trnT(ugu)"
          /note="codons recognized: ACN; 'super Wobble' codon
          recognition inferred, with U in the first position of the
          anticodon assumed to be unmodified"
          /product="tRNA-Thr"
          /anticodon=(pos:56315..56317,aa:Thr)
gene      complement(56278..56350)
          /gene="trnT(ugu)"
tRNA      complement(56464..56535)
          /gene="trnQ(uug)"
          /note="codons recognized: CAR"
          /product="tRNA-Gln"
          /anticodon=(pos:56505..56507,aa:Gln)
gene      complement(56464..56535)
          /gene="trnQ(uug)"
tRNA      56769..56840
          /gene="trnN(guu)"
          /note="codons recognized: AAY"
          /product="tRNA-Asn"
          /anticodon=(pos:56801..56803,aa:Asn)
gene      56769..56840
          /gene="trnN(guu)"
repeat_region 56992..57455
          /rpt_family="A, A1, AD, BD, C"
          /rpt_unit=1 copy of A, A1 and AD, 3 copies of BD, 6 copies
          of C
gene      complement(57552..59630)

```


CDS /gene="ycf2"
 complement(57552..59630)
 /gene="ycf2"
 /note="shows limited similarity to bacterial and
 chloroplast ftsH gene products"
 /codon_start=1
 /product="hypothetical chloroplast RF2"
 /translation="MEVPISAKSRNLTLLFSLFVSNSSGSI FSSQLAQTTLTKESYS
 KKKSLLVKIKTLFSREDSLSGKRFHALNTSGRGLNPNFASSSHFLLSFLGLQGSQSK
 VPSFSLILFYVFFIGLEGSATRFLSHEQRLLLTKNLPGLVLGKNCCLKDRRNEVLP
 RDCEGQRPGGGEHIADLKGQSAHPLEGRVGRSPSGEDSRESSSLFFLTKLSNRLSTY
 SPIVIENKHQNSVNSLAFQNERTEVQYIQT VGTNFCSSKYLVKLSELNLDNKRHLRL
 THVFTQLFVFSVVQKIISDNNFGRNPLRQKDQKTNQARILWPKNKFSFWPFKSGNTN
 KLPNLPGIQELSPLIQVLIESLQNTKQLKTTSFAPKGYL FVGPPGTGKTL LAQEIAQI
 AKVPFICVSASEIQKQIEIGTRIGALRLRKLFEQARALTPCILFFDEIDAIARQQQPQ
 HDSKLTFFELVQMDSWGSGKGFKINSQAVLLGTTNYLSKLD SAFVRSGRFDRILAL
 SYPKKIRFDILKFYITKTSSVNAVLSNLLQEPKIVAGSKNKLIQQVKNFEGTEFF
 PRDPPPNWGEAFGRGEPFEAFGLFSPAVRAKGGGRVAGEGYSKVDIFLGSALGLNY
 FAYSTDGFSQAHLARMVNESLLFAISAQEKRQQT KIKKDSLHSFESLLHGLEKMNIHR
 ETTFKGAEIS"
 repeat_region 57860..57922
 /rpt_family="B, BD"
 /rpt_unit=1 copy of B and BD
 repeat_region 59073..59128
 /rpt_family="AB"
 /rpt_unit=1 copy of AB
 repeat_region 59939..60386
 /rpt_family="A1, AC, B, C"
 /rpt_unit=1 copy of A1 and AC, 2 copies of B, 5 copies of
 C
 gene complement(60553..61176)
 /gene="ycf20"
 CDS complement(60553..61176)
 /gene="ycf20"
 /codon_start=1
 /product="hypothetical chloroplast RF20"
 /translation="MFSQPTSFITILFLFFGFFIGNGGVVKSOLIQT FISFYEILIS
 ITKTLCIALNFQNSPLGIHILDNLPTLQIQEGNSSLNLT ERSNLSFP SRNVFLNFLN
 EFNSFTLFIITFLLECTNCFSSLEQLKQKLNLSVFP PSLISKELDPFFKVKFNKIE
 NFDENKNI PQSSLLTKKNSMFQLSSFLNVFKTGFLGFI FVDAFKVGS"
 gene complement(61245..62492)
 /gene="ycf1"
 CDS complement(61245..62492)
 /gene="ycf1"
 /codon_start=1
 /product="hypothetical chloroplast RF1"
 /translation="METQALSVFEGFYHGLVFSIPFSVPILICISRFLLEGFFPGLFA
 SLGTVLQGQVFFFIVNGTGASRSLIHVWYTFEPFLAFIGFIFAFKLASDFMSGTRTN
 RSNNTLAEKSLTPFLASFVLMFLNPPMGSTSSRILLTTG SFFENSSFLTIPYILVFG
 VTFGLGFVFLWPLIITLFVTKGGPTEKLLRFVNI PAESFGPGRAFQVPPEARILISFL
 LVGCLMSGGFQYSWRLFTQYPASFVDREGQSAWVREFP SFDSNIQHRDKNLPDRHLP
 VDKMNSRRVLSGRPPLSEQQKADASLKFQSF LNTLEEKIENQLLKYRSVSNSTSSFS
 GNKADITDFLLSLKSQGRSLGFFVSPDNSTRFVARPLENNTKGKPKFSYIQSDTLNQG
 KELGIRDILHDELQVYGALFAKS"
 repeat_region 62663..62810
 /rpt_family="A1, B"
 /rpt_unit=1 copy of A1, 2 copies of B
 repeat_region 63339..63461
 /rpt_family="AD, C"
 /rpt_unit=1 copy of AD, 2 copies of C
 gene complement(64096..64344)
 /gene="orf82"
 CDS complement(64096..64344)
 /gene="orf82"
 /codon_start=1
 /product="unknown"
 /translation="MTSKDQRAKGLKQPTRGLKGV SFIYDLKGQSAHPLEDRRSEGA
 PQLFPLRGKGVSCMYDLKGQSAQSHRAKALYHTKASVF"
 repeat_region 64145..64356
 /rpt_family="A, AB, AD, B1, BD, D"
 /rpt_unit=1 copy of A, AB, AD, B1, BD and D

```

repeat_region 64663..64842
                /rpt_family="A, AD, B, BD"
                /rpt_unit=1 copy of A, AD and B, 2 copies of BD
gene          65059..65304
                /gene="psbE"
CDS          65059..65304
                /gene="psbE"
                /codon_start=1
                /product="cytochrome b559 alpha subunit of photosystem II"
                /translation="MSGTTGERPFSDIITSIRYWVIHSITIPSLFIAGWLFVSTGLAY
                DVFGTPRPNEYFTETRQDIPLITDRFNALEQLDQLVP"
gene          65687..65812
                /gene="psbF"
CDS          65687..65812
                /gene="psbF"
                /codon_start=1
                /product="cytochrome b559 beta subunit of photosystem II"
                /translation="MSTNKPLTYPIFTVRWLSVHALAVPTVFFLGSITAMQFIQR"
repeat_region 66094..66132
                /rpt_family="A"
                /rpt_unit=1 copy of A
repeat_region 66407..66423
                /rpt_family="B"
                /rpt_unit=1 copy of B
gene          66505..66621
                /gene="psbL"
CDS          66505..66621
                /gene="psbL"
                /codon_start=1
                /product="L protein of photosystem II"
                /translation="MSNPNPKNQSVELNRTSLYWGLLLIFVLAVLFSSYIFN"
gene          66898..67026
                /gene="psbJ"
CDS          66898..67026
                /gene="psbJ"
                /codon_start=1
                /product="J protein of photosystem II"
                /translation="MSNTGTGRIPLWLVLVVGIAALGLLAILFYGSYVGLGSSL"
repeat_region 67068..67104
                /rpt_family="A"
                /rpt_unit=1 copy of A
gene          complement(67335..67634)
                /gene="rps14"
CDS          complement(67335..67634)
                /gene="rps14"
                /codon_start=1
                /product="ribosomal protein S14"
                /translation="MAKKGMIEREKKRQNLVSKYKNKRIEIKTLSKKSSPAQQFVLAA
                KMSNFPRNSAPTRLRNRFCVTRGRSKYSRDFGLSRHVLREMAVYQGLLPGVTRASW"
repeat_region 67793..67980
                /rpt_family="BD, C"
                /rpt_unit=2 copies of BD, 1 copy of C
gene          complement(68049..68675)
                /gene="rps4"
CDS          complement(68049..68675)
                /gene="rps4"
                /codon_start=1
                /product="ribosomal protein S4"
                /translation="MSRYRGPRLRITRRLGDLPLVNTAKSSGKTTLPQHGSSRPCK
                ASQYSIRLKEKQKLRFNVGVNESQLVRYVKKARNMKGSTGELVQLLEMRLDSIVFRS
                GLAPTLRAARQLVSHGHIAIETKTSQNKTE NSFLPVNI PSYQCRPGDNVRLINLPTGE
                ARSSTSLPAHIHFDNASKSYQVKQLIPRSEVSCSVNELLIIEYYSRKV"
gene          complement(68817..69032)
                /gene="orf71"
CDS          complement(68817..69032)
                /gene="orf71"
                /codon_start=1
                /product="unknown"
                /translation="MNKAHPFETSGPLVLEAFGPPVLSGHKQAQVTETSPRRGRPLAL
                KGEELRSSQGGGLWPFGLCVYNKVQA"
repeat_region 68890..69317

```

```

/rpt_family="A, A1, A1D, B, B1, BD, C, D"
/rpt_unit=1 copy of A1, A1D, B1 and BD, 2 copies of A, B
and D, 3 copies of C
gene complement(69531..70505)
CDS /gene="rpoA"
complement(69531..70505)
/gene="rpoA"
/codon_start=1
/product="alpha subunit of RNA polymerase"
/translation="MDTELLLEYRIEKLFPSPEDSLEAKKIQQVLGKNTFENSSSLHYGK
FYLSGLKKGQGITVANALRRVLLYDIMG LGITKVKFNSKNNENLSSQKLHEFSTISGI
RESILELLMNLRLLVWTTSP LDSNQKVS GDSSDPLLVHEELDISSPSLTSYLGQFS
LPEYFKSKDWLSEKNKNNQSSVFVLRGKHLISADQRTNQNPNFAQLINPEQYLATILP
IQKTATGLSKFPFEIEYKIGKGNLFENLDLQSNNEEFLLNSGPPFFVKKVNYTVEDI
KNSTNNFLWERGFSECVFFEIWTGDSLHPEKALSQAFDILLDFQKQD"
gene complement(70747..71124)
CDS /gene="rps11"
complement(70747..71124)
/gene="rps11"
/codon_start=1
/product="ribosomal protein S11"
/translation="MAKARKVLKVKVQTGIVHIRATFNNTLVTVDPNGNVISWSSAG
SCGFKGARKSTPFAAQEAATAAIRQSIEQGLRKVEVQVSGPGSGRETALRAIQAAVGV
IRLIRDITPIPHNGCRAPKRRRV"
repeat_region 71306..71796
/rpt_family="A, A1, AD, A1D, BD, C"
/rpt_unit=1 copy of A1, AD and A1D, 2 copies of A and BD,
4 copies of C
gene complement(71874..72827)
CDS /gene="petA"
complement(71874..72827)
/gene="petA"
/codon_start=1
/product="apocytochrome f of cytochrome b6/f complex"
/translation="MVQSSFFERVLSISVVCLSAFIVLGFSSHSSAYPVYAQQGYEN
PREATGRIVCANCHLAEKPVDI EVPQSVLPNTVFEAVVKIPYDQQVQVLGNGKGGPL
NVGAVLVLP EGFLAPADRIPEEMKTKVGKLYFSPYSESRSN ILVVGVPVPGKKYSEMV
FPILSPDPQTNPKAHYLNYP IYLGGNRGRGVYPDGKSNNTVYNSPAAGVIASITTG
ADGARTVTVKKADGSSVSDTIPAGPTLLVAEGQSVVADQPITNPNVGGFQGAETEIV
LQNPARIQGLIVFFTTVILAQIFLVLK KQFEKVQLSEMNF"
tRNA complement(72972..73044)
/gene="trnH(gug)"
/note="codons recognized: CAY"
/product="tRNA-His"
/anticodon=(pos:73010..73012,aa:His)
gene complement(72972..73044)
/gene="trnH(gug)"
tRNA complement(73187..73259)
/gene="trnV(uac)"
/note="codons recognized: GUN; 'super Wobble' codon
recognition inferred, with U in the first position of the
anticodon assumed to be unmodified"
/product="tRNA-Val"
/anticodon=(pos:73224..73226,aa:Val)
gene complement(73187..73259)
/gene="trnV(uac)"
repeat_region 73356..73664
/rpt_family="A, A1, AB, BD, C, D"
/rpt_unit=1 copy of A, A1, AB, BD and D, 3 copies of C
gene complement(73826..75154)
CDS /gene="ccsA"
complement(73826..75154)
/gene="ccsA"
/codon_start=1
/product="protein involved in cytochrome c biogenesis"
/translation="MLGILELLFKCFPSHSRPAVLRRIYLYFTLVFNMDSENIFLTL
FFSVVLSLWSWTSVKPNQPFLLVQESTGGNLSKNSKTLFQIKFEEFFIVENQFGILLS
SVFVFGANFLLSFLLMRWIDSGHPFNSNLYESLFLTWLFLFYFPTVGFQKDLK
TKNSSKLQEFNTNTEVLKRAEFGKSNEIHKTVLPMAIVPTGNPSLRISTSLKSVRQKES
LPPEVSSLAGIVLVAALFIYTFATWVLP AEMREIKPLVPALKSNWLLMHVSI MLLSY
SALLAGSLFSILYLVLFYFNKNRTEPTKLTTRKSSDQKRSVSEIPQLNGNISFPQLKQFS

```

KQRFLTDLSDLSYRSLAFAPFLLTLGIIISGAVWANEANGSYWSDPKETWALITWTF
 SIYHLRLRIQKSWEGEKAALVAAGFFVWVCYLVNLLGKGLHSYGMWNS"

repeat_region 75363..75716
 /rpt_family="A, A1D, B, C"
 /rpt_unit=1 copy of A, A1D and C, 2 copies of B

gene 76006..76098
 /gene="psaM"

CDS 76006..76098
 /gene="psaM"
 /codon_start=1
 /product="M polypeptide of photosystem I"
 /translation="MLSDSQVFTALFLALVTGIFAVRLGTQLYK"

repeat_region 76198..76627
 /rpt_family="A, A1, AD, A1D, BD, C"
 /rpt_unit=1 copy of A1D and BD, 2 copies of A, A1 and AD,
 5 copies of C

gene 76816..76947
 /gene="psaJ"

CDS 76816..76947
 /gene="psaJ"
 /codon_start=1
 /product="subunit IX of photosystem I"
 /translation="MKNFTTYLSTVPLSFLYFGFLAGLLIEINRFYPDALVFPFLG"

repeat_region 77240..77573
 /rpt_family="A, A1, BD, C, D"
 /rpt_unit=1 copy of A1, BD and D, 2 copies of C, 3 copies
 of A

gene complement(77676..77927)
 /gene="psbH"

CDS complement(77676..77927)
 /gene="psbH"
 /codon_start=1
 /product="10 kDa phosphoprotein of photosystem II"
 /translation="MATGTTAKTKISASSSTGKVTALGTALKPLNSEYKGVAPGWGTT
 VIMGVFMALFAVFLVILLEYNSSVVLEGVGVSWASLAS"

gene 77977..78177
 /gene="orf66"

CDS 77977..78177
 /gene="orf66"
 /codon_start=1
 /product="unknown"
 /translation="MKAGEPFQNFVLPKSGPESHANLPQLRGGRLRPSLSSRCEGQRPR
 VVPLEVRVGRTPFLRFGPFV"

repeat_region 77980..78062
 /rpt_family="B, BD"
 /rpt_unit=1 copy of B and BD

gene complement(78417..78524)
 /gene="psbM"

CDS complement(78417..78524)
 /gene="psbM"
 /codon_start=1
 /product="M protein of photosystem II"
 /translation="MEVNILGLIATALFIIPTSFLILYVKTSTKQD"

gene complement(78754..78942)
 /gene="ycf9"

CDS complement(78754..78942)
 /gene="ycf9"
 /codon_start=1
 /product="hypothetical chloroplast RF9"
 /translation="MLFLFQISLFAFIALSFAVVGVPVLLASPDGWSSSKQVVFSGT
 SLWFFLVFVVGVLNSFVV"

repeat_region 78994..79085
 /rpt_family="A, B"
 /rpt_unit=1 copy of A, 2 copies of B

repeat_region 79470..79725
 /rpt_family="A, A1, AD, B, C"
 /rpt_unit=1 copy of A1, AD, B and C, 2 copies of A

gene complement(79894..80037)
 /gene="psbK"

CDS complement(79894..80037)
 /gene="psbK"

```

/codon_start=1
/product="K protein of photosystem II"
/translation="MLNTSVFLLARLPEAYAPFDPLVDVLPPIPLFFLLAFVWQASV
SFR"
repeat_region 80301..80519
/rpt_family="A, B, B1, BD, C"
/rpt_unit=1 copy of A, B, B1 and C, 3 copies of BD
gene complement(80636..80737)
/gene="ycf12"
CDS complement(80636..80737)
/gene="ycf12"
/codon_start=1
/product="hypothetical chloroplast RF12"
/translation="MNLELVAQLVSLFLILGAGPIVVLLFARGGNL"
repeat_region 80971..81310
/rpt_family="A, AD, B, BD, C"
/rpt_unit=1 copy of AD and B, 2 copies of BD, 3 copies of
A, 4 copies of C
repeat_region 81815..82037
/rpt_family="A, C"
/rpt_unit=1 copy of A and C
tRNA complement(82378..82449)
/gene="trnG(gcc)"
/note="codons recognized: GGY"
/product="tRNA-Gly"
/anticodon=(pos:82415..82417,aa:Gly)
gene complement(82378..82449)
/gene="trnG(gcc)"
tRNA complement(82539..82612)
/gene="trnR(acg)"
/note="codons recognized: CGN; 'super Wobble' codon
recognition inferred, with A in the first position of the
anticodon assumed to be post-transcriptionally modified to
I (inosine)"
/product="tRNA-Arg"
/anticodon=(pos:82576..82578,aa:Arg)
gene complement(82539..82612)
/gene="trnR(acg)"
repeat_region 82670..82913
/rpt_family="B, D"
/rpt_unit=1 copy of B, 3 copies of D
repeat_region 83207..83358
/rpt_family="AB, B, D"
/rpt_unit=1 copy of AB, 2 copies of B and D
gene 83425..84951
/gene="psbB"
CDS 83425..84951
/gene="psbB"
/codon_start=1
/product="CP47 chlorophyll apoprotein of photosystem II"
/translation="MGLPWYRVHNVVLDNPGRLIAVHLMHTALVSGWAGSMALYELAV
FDPSDPVLNPMWRQGMFVLPFMTRLGVTGSWGGWSITGETIGNPGIWSYEGVAASHI I
LSGLLFLAAVWHHTFWDLELFRDPRTGDPALDLPKIFGIHLFLSGLLFCFGFAGFHVTVG
FFGPGIWSVDPYGITGSVQPVAPAWGPEGFDPYNPGGIASHHVAAGILGILAGLFHILT
VRPPQRLYKGLRMGNIETVLSSSIAAVFWAAFVVSGMTMWYGAATPIELFGPTRYQWD
QGFFQAEIDKRVQKSLSEGKNLSTAWASIEKLAFYDYIGNNPAKGGFLFRSGPMDNGD
GIAAGWLGHAVFTDKDGRELTTRRMPTFFETFPVILVDQDGIVRADVPFRAESKYSI
EQVGVSVTFYGGELDGITFNDPATVKKYARRAQLGEI FEFDRATLQSDGVFRSSPRGW
FTFGHLCFALLFFFGHIWHGARTIFRDVFAGIDPDLDEQVEFGAFQKLGDVTTTRQAV
"
repeat_region 85000..85099
/rpt_family="A, B, BD"
/rpt_unit=1 copy of A and BD, 2 copies of B
gene 85720..85815
/gene="psbT"
CDS 85720..85815
/gene="psbT"
/codon_start=1
/product="T protein of photosystem II"
/translation="MEALVYTFLLVGLTGIIFFSIFFREPPRIKD"
repeat_region 86010..86064

```

```

/rpt_family="A, C"
/rpt_unit=1 copy of A and C
gene      86163..86297
/ gene="psbN"
CDS       86163..86297
/ gene="psbN"
/ codon_start=1
/ product="N protein of photosystem II"
/ translation="METPAFIYAIFLWFLISVTGYSIYVGFPPSKELKDPFEDHED
"
repeat_region 96479..86623
/rpt_family="AB, B and BD"
/rpt_unit=1 copy of AB and BD, 2 copies B
gene      86691..87734
/ gene="chlI"
CDS       86691..87734
/ gene="chlI"
/ codon_start=1
/ product="magnesium chelatase subunit of
protechlorophyllide reductase"
/ translation="MSQNERPVFPFTAIVGQEMKLALILNVIDPKIGGVMVMGDRGT
GKSTTVRALADILPEIEIVANDPNSDPNDPEAMSDAVRDMIRQNKPFIEITKTKITMV
DPLGATEDRVCGTIDIEKALTEGVKAFEPGLLAKANRGILYDEVNLLDDHLVDVLL
DSAASGWNTVEREGVSIRHPARFILVSGNPEEGELRPQLLDRFGMHAQIGTKEPEL
RVKIVEQRTSFDQNAQAFQLEFQNSQKELRERI IQARALLPKIQLSYELRVNISKMCS
ELNVDGLRGDIVTNRASKALAAFEGRQNVTPKDIYTVAAALCLRHRLRKPDPLESIDTGT
KVQDAFTRVFGEV"
repeat_region 97851..88254
/rpt_family="A, AC, B, B1, BD, C, D"
/rpt_unit=1 copy of AC, B1 and C, 2 copies of B and BD, 3
copies of A and D
tRNA      88473..88553
/ gene="trnL(uaa)"
/ note="codons recognized: UUR"
/ product="tRNA-Leu"
/ anticodon=(pos:88506..88508,aa:Leu)
gene      88473..88553
/ gene="trnL(uaa)"
gene      89119..89766
/ gene="petB"
CDS       89119..89766
/ gene="petB"
/ codon_start=1
/ product="apocytochrome b6 of cytochrome b6/f complex"
/ translation="MSKVYNWFNERLEIQAIADDIGSKYVPPHVNIYFCFGGIVLTCF
LVQVASGFAMTFYYRPTVTEAFRSVEYIMTDVNFGLIRSVHRWSASMMVMMLHVF
RVYLTGGFKKPRELTVVTGVLASITVSGVGTGYSPLPVDQVGYWAVRIVTGPVDAIPG
VGSFVVELLRGGVGVGQATLTRFYSLHTFVLPPLAATFMLMHFLMIRKQGISGGL"
repeat_region 89882..90006
/rpt_family="B, B1, BD, C"
/rpt_unit=1 copy of B, B1, BD and C
gene      90074..90556
/ gene="petD"
CDS       90074..90556
/ gene="petD"
/ codon_start=1
/ product="subunit IV of cytochrome b6/f complex"
/ translation="MSVTRKPDLSDPVLRKAKLAKGMGHNYGEPAPNDLLYMFVVI
FGTVASCVGLGVLSPTQLGEPANPFATPIEILPEWYFFSTFNLLRTIPNKLGLVLSMA
AVPVGLLTVPFIEININKFQNPFRFPVATAVFLFGTFAAVWLIGIGAAMPIDKALTGLFF
"
repeat_region 90905..90995
/rpt_family="B"
/rpt_unit=2 copies of B
gene      91133..91432
/ gene="rpl23"
CDS       91133..91432
/ gene="rpl23"
/ codon_start=1
/ product="ribosomal protein L23"
/ translation="MSFWNFILFLSQNMILPIFTEKSTRILENNQYTFDVPDPMNSKQE

```

```

gene      IRQFVEKNFEVVKVLSINTHRPPRKARRVGFAGFRAQVKRVILTTLKAGDSIAFFE"
          complement (91614..91868)
          /gene="orf84"
CDS       complement (91614..91868)
          /gene="orf84"
          /codon_start=1
          /product="unknown"
          /translation="MKSFFLAIKDRRNGVLPTRPSRGTLRGSREKDQKTQGPKDRRPP
          PKVKDLPPNGGRAGGSPSGLWLSGSPRRGEAFGPQWLKL"
repeat_region 91643..91722
          /rpt_family="B, BD"
          /rpt_unit=1 copy of BD, 2 copies of B
gene      91911..92729
          /gene="rpl2"
CDS       91911..92729
          /gene="rpl2"
          /codon_start=1
          /product="ribosomal protein L2"
          /translation="MSIRSUKAYTPGTRNRSVDFNELTGALPQKKLTRFAHSSTGRN
          NRGIIITSRRGGGHRKRYRIDDFKRNKYDIPGIVKTVEYDPNRANISLICYQDGEKR
          YILHSAGLKIIGTEILSSQAPSTLGNLPLMNMPLGVEIHNVIEHPGKGGQLVRAAGT
          VAQIVAKEGPFVTLRLPSGEVRLFFKECWATVGQVGNIEHMNIRIGKAGRNRLGKRP
          HVRGSMNPVDHPHGGEGRCPIGHPRPLSPWGPALGKKTSAKKYSNAFRVI"
repeat_region 92794..93036
          /rpt_family="A, B, BD"
          /rpt_unit=1 copy of A, 3 copies of B and BD
gene      93098..93430
          /gene="rps19"
CDS       93098..93430
          /gene="rps19"
          /codon_start=1
          /product="ribosomal protein S19"
          /translation="MCFLVHNQRIMNLSRWIRIFAMPNTRPPFIANHLFKKIVELNNSK
          EKKVIQTSRSTIVPMIGHTIAVHNGKEHIPVFITDQMVGHKLGESPTRTFRGHI
          KTDKKGKR"
repeat_region 93566..94085
          /rpt_family="A, B, B1, BD, C, D"
          /rpt_unit=1 copy of B and B1, 2 copies of A and BD, 3
          copies of D, 6 copies of C
gene      complement (93852..94034)
          /gene="orf60"
CDS       complement (93852..94034)
          /gene="orf60"
          /codon_start=1
          /product="unknown"
          /translation="MGEEMGGPFRPSVLGGPLQAFGLKGTPLQLSPLQEERSFLPFWGR
          GGSNPPLLGWKFFTYF"
gene      94297..95004
          /gene="rps3"
CDS       94297..95004
          /gene="rps3"
          /codon_start=1
          /product="ribosomal protein S3"
          /translation="MGQKIHPGLRIGIFQKHKSEWFAPVGSNSNSQIPNYAAFIEQ
          DKGVRDFSSSELRDAGLSKILISRRTNRLFVELHVAQPKLVLGGENNKLQLTEKFQK
          NIQKLLSRHNNPSMSAPALSIQVVQIPQEEIATDATLLGQKIAEQLEKRVAFRRVMKQ
          SIQARAAGAEGIKIQIAGRLNGAEIARTEWAREGRVPLQTLKANIDYCYETAQTIYG
          ILGIKIWIFKGTNSRV"
repeat_region 95165..95181
          /rpt_family="B"
          /rpt_unit=1 copy of B
gene      95346..95762
          /gene="rpl16"
CDS       95346..95762
          /gene="rpl16"
          /codon_start=1
          /product="ribosomal protein L16"
          /translation="MLSPKRTKYRKHHRGRMKQASRGNKIAYGTALQALEPGWLTS
          RQIEAGRRTITRYAKRGGKLIWIRVFPDKPITMRPAETRMGSGKGSPEYVWAVIKPGRI
          LYEMNGVPLTLAKEAMRNASYKMPVKTTFLLTITEN"
repeat_region 95839..96115

```

```

/rpt_family="A, B, BD"
/rpt_unit=1 copy of A and B, 4 copies of BD
gene 95904..96125
CDS   /gene="orf73"
      95904..96125
      /gene="orf73"
      /codon_start=1
      /product="unknown"
      /translation="MRPSLSSRWRAKGLEGFPSRAKGLPANGGRAKGLEGFPSRVAVE
      GPHQLSRGPQALGGLNFISDLKSQNAPKP"
gene 96261..96626
CDS   /gene="rpl14"
      96261..96626
      /gene="rpl14"
      /codon_start=1
      /product="ribosomal protein L14"
      /translation="MIQPQTYLHVADNTGAQTIMCIRVLGSQREYASVGDVIGVVK
      ALPNMTLKKSDVVRVAVVVRTAHGVRRSNGMTRFDENAADVINQEGNPRGTRVFGPIA
      RELREKNFSKIVSLAPEVL"
gene 96637..97173
CDS   /gene="rpl5"
      96637..97173
      /gene="rpl5"
      /codon_start=1
      /product="ribosomal protein L5"
      /translation="MAQRLKTYIASIVQNPESEFGYSKKGTFEVPKIEKIVINSGIGE
      SSNIELASQEFLLSGQKTVVTKARKSIASFKIRENMPVGGICVTLRGRMYAFLDRLI
      QLALPRIRDFFQGLSQSFDGHGNYTLGLKEQLMFPEIDYDKIEKSRGMDISIVTSAKN
      DQDAFLLKALGMPFKAA"
gene 97324..97551
CDS   /gene="orf75"
      97324..97551
      /gene="orf75"
      /codon_start=1
      /product="unknown"
      /translation="MKFTLLLFLSKLYTLSSGGPKGQRPPPSGEERSSSHAILPPFAG
      RPSALKGKPPSSPFHFDLVCFSALWCASEL"
repeat_region 97396..97647
/rpt_family="A, AD, B, C, D"
/rpt_unit=1 copy of A, AD and C, 2 copies of B and D
gene 97845..98294
CDS   /gene="rps8"
      97845..98294
      /gene="rps8"
      /codon_start=1
      /product="ribosomal protein S8"
      /translation="MVNYAVSDMLTRIRNGIRVKKTKVSVISTNLTRSIAEILKREGF
      IENFATTQMTKPSVGNLATTGNKNSTLDIYLKWSQKDLGKAKQKNVLTNLKICISRPGV
      RIYANANNIPQVLGGLGITLSTSKGILTDEARNQKVGGEVLCMLW"
repeat_region 98348..98852
/rpt_family="A, AD, B, BD, C"
/rpt_unit=1 copy of AD and B, 3 copies A, BD and C
gene 99147..99353
CDS   /gene="infA"
      99147..99353
      /gene="infA"
      /codon_start=1
      /product="translational initiation factor 1"
      /translation="MEGIITESLPNAMFRVDLENGFNVLAHISGKIRKKNYIKILLGDR
      VLVQLTPYDLTKGRIIYRQKNSSS"
gene 99398..99586
CDS   /gene="rps18"
      99398..99586
      /gene="rps18"
      /codon_start=1
      /product="ribosomal protein S18"
      /translation="MKQTQIDYKDVAYLRRFITEQKGILSRFTGLKAKQQRELTRAV
      KQARILGLLGFPGSTKTR"
gene 99802..99972
CDS   /gene="rpl32"
      99802..99972

```



```

/gene="rpl32"
/codon_start=1
/product="ribosomal protein L32"
/translation="MAVPKKRVSRTRKLLRNNQWKLQIRPQALKALS LAKSVERGGLP
SPPENTSPNDES"
tRNA 100369..100442
/gene="trnMf(cau)"
/note="codon recognized: AUG"
/product="tRNA-Met"
/anticodon=(pos:100403..100405,aa:Met)
gene 100369..100442
/gene="trnMf(cau)"
gene complement(100622..102007)
/gene="psbC"
CDS complement(100622..102007)
/gene="psbC"
/codon_start=1
/product="CP43 chlorophyll apoprotein of photosystem II"
/translation="METLFNGTLVVGGRDQESTGFAWWAGNARLINLSGKLLGAHVAH
AGLIVFWAGAMNLFVVAHFVPEKPMYEQGLILLPHLATLGYGVGPGGEVLDTFYFVS
GVLHLISSAVLGFVGVYHSLIGPETLEESFPFFGYVWVKAKMTTILGIHLVLLGVGA
YLLVWKAMYFGGLYDTWAPGGGDVRIISNPTVTPGVI FGYLLKSPFGDGIWVSCDNL
EDVVGGHIWVGNLLIFGGIWHILTKPFAWARRALVWSGEAYLSYSLAAVSLMAAISAC
MVFNNTVYPSSEFFGPTGPEASQAQAFTELVVDQRLGANVASAQGPTGLGKYLMSRPT
GEIIFGGETMRFDWDRAPWVEPLRGPNGLDLNKLNKNDIQPWQERRSAEYMTAPLGS
NSVGVVATEINAVNFVNPRSWLATSHYVLFVFFVGHVHAGRARAAAAGFERGIDRD
TEPVLTMRPLD"
exon complement(102573..103109)
/gene="psbA"
/number=2
gene complement(102573..104718)
/gene="psbA"
CDS complement(join(102573..103109,104197..104718))
/gene="psbA"
/codon_start=1
/product="D1 reaction center protein of photosystem II"
/translation="MTATLRRENATIWSNFCEVFTSTENRLYIGWFGVLMIPTLLTAT
SVFIIAFIAAPPVDIDGIREPVSGSLLYGNIIISGAVIPTSNAI GLHFYPIWEAASVD
EWLYNGGPPYQLIVCHFFLGICSYMGREWELSFRLGMRPWI AVAYSAPVAAATAVFIIY
PIGQGSFSDGMPLGISGTENFMIVFQAEHNILMHPFHMLGVAGVFGGSLFSAMHGSLV
TSSLIRETTENESANAGYRFGQEEETYNIVA AHGYFGRLIFQYASFNNSRSLHFFLAI
WPVMGIWFTALGISTMAFNLNGFNENQSVVDSQGRVINTWADIVN RANLGMVEMHERN
AHNFPLDLAAVEAPSIVG"
intron complement(103110..104196)
/gene="psbA"
/note="group I intron"
/number=1
repeat_region 103308..103364
/rpt_family="AB"
/rpt_unit=1 copy of AB
gene complement(103448..104107)
/gene="orf219"
CDS complement(103448..104107)
/gene="orf219"
/codon_start=1
/product="putative site-specific DNA endonuclease"
/translation="MKAQSKDNPVLKTFSLSSLPVFPGIYSIKNTQTGFYIGQSKNI
IQRLRAHRSMLETHQNKMTEDCKIYGIHSFEVSI LLEGIEYESNDLRLSLEKEYI
KKFPIEKLYNTDRCVENNSFFGKIHTPELKKRFTEERTNI PNYELGRGICIPSFTSRK
GNISPGGSFLSLAEASRVTNMSRRDLRRLNDPTFPDWRELTPKEQEDLKKRQQQGH
A"
exon complement(104197..104718)
/gene="psbA"
/number=1
rRNA 105920..107421
/gene="rrs"
/product="small-subunit ribosomal RNA"
gene 105920..107421
/gene="rrs"
tRNA 107635..107708
/gene="trnI(gau)"

```

```

/note="codons recognized: AUU"
/product="tRNA-Ile"
/anticodon=(pos:107669..107671,aa:Ile)
gene 107635..107708
      /gene="trnI(gau)"
tRNA 107735..107807
      /gene="trnA(ugc)"
/note="codons recognized: GCN; 'super Wobble' codon
recognition inferred, with U in the first
position of the anticodon assumed to be unmodified"
/product="tRNA-Ala"
/anticodon=(pos:107768..107770,aa:Ala)
gene 107735..107807
      /gene="trnA(ugc)"
exon 107919..109779
      /gene="rrl"
      /number=1
rRNA join(107919..109779,110590..110609,111473..112020,
113148..113240,114058..114346)
      /gene="rrl"
      /product="large subunit ribosomal RNA"
gene 107919..114346
      /gene="rrl"
intron 109780..110589
      /gene="rrl"
      /note="group IB4 intron inserted at a position
corresponding to residues 1931-1932 in E. coli 23S rRNA"
      /number=1
gene 110055..110471
      /gene="orf138"
CDS 110055..110471
      /gene="orf138"
      /codon_start=1
      /product="putative site-specific DNA endonuclease"
      /translation="MSMHEQWITGFTDGEFCFHIGISKNNETKLGSQLLEFVITQHQ
RDEQLLNEIKNYFGVGVVRKSNRGRDILCYRVRSQKHLRDVILPFFFEKNILRTKKKFA
FQRFKQALVLIENKEHLTLEGLDKLRKLRDQVDDSD"
exon 110590..110609
      /gene="rrl"
      /number=2
intron 110610..111472
      /gene="rrl"
      /note="group IA3 intron inserted at a position
corresponding to residues 1951-1952 in E. coli 23S rRNA"
      /number=2
gene 110866..111351
      /gene="orf161"
CDS 110866..111351
      /gene="orf161"
      /codon_start=1
      /product="putative site-specific DNA endonuclease"
      /translation="MEPESKNTFDASWLVGFIDGEGCFCLSFSEERERLTLKIEVRPS
FSVSQNSKSMKILETMNTYFGCGGVRHRSKRENTHKYEVRLNLHTTVIIPFFFRKYP
LLTNKQRDFLLFDQACMIKENRHLSREGLKEILQLAAQMNTSGTRKNSIVNLLNILD
LDS"
exon 111473..112020
      /gene="rrl"
      /number=3
intron 112021..113147
      /gene="rrl"
      /note="group IB4 intron inserted at a position
corresponding to residues 2500-2501 in E. coli 23S rRNA"
      /number=3
gene 112368..113114
      /gene="orf248"
CDS 112368..113114
      /gene="orf248"
      /codon_start=1
      /product="putative site-specific DNA endonuclease"
      /translation="MEKNQEKLLAAGTLLSQYIQNPKPYTTRVLRGGKRTAYLEQFAN
QPLSFLQKSFITGHLGDCILYTDGITPFLKIDQQVSRKAYIDVIYSVFEEYVGSPP

```

SLRRDRMGNPHSYFRTLRLTEKLANAKQFYQLNAMGRRTKVPQNIHKWLNPIISLAV
WFMDDGGTEVSGYRLHTENFLSDVKILQQALGSVFNLANVRDRNRRTTGLYYLSIP
ANSKQRFQEIIVAPYMVPTMMYKLRLPDTE"

exon 113148..113240
/gene="rrl"
/number=4

intron 113241..114056
/gene="rrl"
/note="group IA3 intron inserted at a position
corresponding to residues 2593-2594 in E. coli 23S rRNA"
/number=4

gene 113351..113863
/gene="orf170"

CDS 113351..113863
/gene="orf170"
/codon_start=1
/product="site-specific DNA endonuclease I-MsoI"
/translation="MTTKNTLQPTAAAYIAGFLDGGDSIYAKLIPRPDYKDIKYQVSL
AISFIQRKDKFPYLQDIYDQLGKRGNLKDRGDGIADYTIIGSTHLSIILPDLVPLYR
IKKKQANRILHINLYPQAQKNPSKFLDLVKIVDDVQNLNKRADLSTNYDRLEEF
LKAGKIESSP"

repeat_region 113895..113931
/rpt_family="A"
/rpt_unit=1 copy of A

exon 114058..114346
/gene="rrl"
/number=5

rRNA 114407..114525
/gene="rrf"
/product="5S ribosomal RNA"

gene 114407..114525
/gene="rrf"

BASE COUNT 34898 a 22101 c 22525 g 35004 t
ORIGIN

```

1  gccccatcg tctagaggcc aaggacatct ccctttccag gaggtaacgg ggattcgaat
61  tccccctggg gtaaaagaaa ggggttcccc cttggtggtc cttgaccttc acactccttt
121 cccctaaacg actattggga tcttaaaatt ttgtatgttt taaaaattga atgcccggagt
181 agagaaccac tagtagcaaa gattcctgaa gatagataaa gggtaaagtg tattttcgtt
241 tctcaaaagca aagttttttc aaactctatt tattcttcaa agaataacgt cttatgtcaa
301 aaggtataga aaaatattcc cttcagaact ttttgtatc atttattatc ttctttaaaa
361 agtccccctg tgggactaag ttttgtatga aatggacttg gtccatgctt gagagttaaa
421 cctattttct tttattttct aataaaaacag ttttcataca taaataaaat tttatgtatta
481 taagacaaaa ggttcagtag cttagtggta gaggcgggga ctcataagcc caaggtcgca
541 ggttcaaatc ccgcctgagc caccagtgtt aatatcagtt tgttttcatt tttaaaaaat
601 gatcgtctat cttcgtccac ttgacgttta aaattttgtc gttgtatcat atttttatgc
661 tgagatagct cagtccgtag agcagaggat tgaaaatcct tgtgtcacca gttccgattct
721 ggttcttggc atatggaagg gtgtccgagt ggttgatggt cttggttttg aaaaccaacg
781 tggcgtaagt tacctgtggg tcgaatccca ccccttccga aaggaaagaa atgttataat
841 agtttgtaac gcagagaaaa aaaattatgg atcaaatcga gaagttttaa tacactttct
901 acttcgacct tttcgtcttt tggcaatgga atgaagactt ggccgaacaa gaagaagtat
961 atttctgaaa atttacagga aataagtgat attatttttc tcggtttttt ggtaaatag
1021 tgattttggt tctctctccg agaggaaaaca aaaaagggtt tacaagaagt gatgttctgt
1081 gaactgtggt ctctccaagc accaaaatct tcttcccttc gggcgggaga tttctgtcat
1141 tcgagtagaa agacttcttg tatttagtta gaccgaaaag gtcttaaatc taagatttac
1201 gaaccagaat ctcagccatc caaagatttt aaaaatgaga cttatgagtt tattgtaaca
1261 atattaaatt cttatgcttt ttttaggctt tggttacttt ttatccgcaa ctaaagagaa
1321 ataccgcttt tggtttttac tttagcaatat gctagtagta agttctttaa aattccgggt
1381 tgtttacttc ccgaaatttt tttggctata gtactttagt gcttatttgg tctaaggagc
1441 tctctgaaat ccattgaata ctttgtacgt taccaaaagtc gtaaatcttt agttcttcgg
1501 tcttttcatt taagaataaa tatgacaatt gcaatcggta aatcagaaga aaaacgtggt
1561 tggtttgatg acgcggatga ctggttaaaa cgtgaccgtt tcgttttcgt gggttggtct
1621 ggttattctat tacttctctt tgcgtatttc gcattagggtg gttggtttac tggactact
1681 tacgtaacct catggtatac tcacggttta gcaagttctt acttagaagg ttgtaacgtt
1741 ttaactgctg ctgtaagta cccagcaaac agtatggctc actcacttct attcctatgg
1801 ggtccagaag caggtgggtga ctttactcgt tgggttcaat taggcgggtc ttggactttc
1861 gttgctcttc acggtgcttt tggcttaatt ggcttcatgc tacgccagtt tgaaattgct
1921 cgtgctgtac aacttcgtcc atacaacgca attgctttct cagcaccaat tgctgtttc
1981 gtttctggtt tcttaattta cccattaggc caatcagggt ggttcttcgc accaagtttt
2041 ggtggtgctg gtattttccg tttcattctt ttcttccaag gtttccacaa ctggacactt
2101 aacccattcc acatgatggg tgtagctggt gttcttggtg cagctctttt atgtgccgatc
2161 cagggtgcaa ctgtagaaaa cactctattt gaagatggtg acggtgcaaa cactttccgt

```

2221 gctttcaacc caactcagtc agaagagaca tattcaatgg ttactgctaa ccgtttctgg
 2281 tcacaaattt tcgggtgtagc tttctctaac aaacggttggc ttcacttctt tatgctattc
 2341 gtccagtaa ctgggttatg gatgagtgct attggttag taggtttagc tctaaacctt
 2401 cgtgcttatg acttcgtatc acaagagatt cgtgctgctg aagaccaga attcgaaca
 2461 ttctacacga aaaacattct tctaaacgaa ggtatccgtg ctggatggc agctcaagc
 2521 caaccacacg agaaccttgt attccctgaa gaagtctctc ctcggtgtaa cgctctataa
 2581 tccagctctt ggaacaaag ttggttagaa tatcctcttt cttttagaat aaggatgctc
 2641 ctctcccgcg gccttagaat gatctaaggc cgcggggaca gtgcaaacgg cgcataaaaa
 2701 tatttactta ttttttggg ataagaacaa aggaacatgg tctgacctc ttggatgagg
 2761 accaaagaac cggaggtctc aaggggcccc cagcttcaa gggtttgagc ttaatatata
 2821 cgagctgaaa ggaccgggag gaccaaaggt ctcaaagggg tgagctttaa tttatgacct
 2881 gaaaggaccg ggaggtctgg tgtgcttcag ctttatttat gacctgaaag gaccttaggt
 2941 ctggtgtgat tgagcttaat atatgacctg ggtgcttga cgcctttgac ctctcaggtc
 3001 agatttcaag ctcaagcaca ccagacctcc ggtctcaaaa ggtcaaacgg ctcaatcaca
 3061 ccaggtcaaa gcgctgaagc acaccaggtc aaagcgtga agcacaccag gcaaaaacgc
 3121 tcaagccatt aaggaatgcc agatttcaat taattattta atctatttca aagaattatt
 3181 ctgatacact aaattatag gaagcgtgcc tgagaggctt ataggacaga atgtctaatt
 3241 ctgagagtg ttaaaaacgc tccggggggt cgaatcccc cgctccgaa aataacatat
 3301 aaaagagaaa aatgtcacta cacacaggtt gactgcccag gacatgcaga ttattggcat
 3361 staaactatt tgaatttaa aacctgttag ttgggatctg tttatttctc ttgtagattg
 3421 aaaagaaggt tataacaaa tattttttg ttttagaac tttccattt ttgtttctc
 3481 cgcgaaagtt tctttttgc tattgaaact atctatttct ctaacttagg gaatttagcc
 3541 actacctttt cttaagattc tttgtagact ttgacttctt gggagcctgg aggaccaag
 3601 gtccaagag gttgagcttt atatatgacc tgaagggacc gggaggacca aaggtcaag
 3661 gctccccccc ttcgggggga gacctcggg cctgtggtcc ttgaggtcgg aaggtgcccc
 3721 tgctctcggt tgctcgagcg ctttgacctg gtgtgattga gcgctttgac ctggtgtgct
 3781 tcagcgtctt gacctggtgt gatgagcgc tttgacctt gaggaccgga ggtctggtgt
 3841 gcttgagctt gaaatctgac ctgaaaggtc aaagcgtca agcacaccag gtcataat
 3901 aagctcaatc acaccagact acggctctt caggtcataa ataaagctga agcacaccag
 3961 actccccgct cttcaggtc ataaataaaa gctcaccct ttgagacctt tggctctccc
 4021 ggtcctttga ggtcagatat taagctcaag cacaccagac ctacggctct tctaggtcag
 4081 aaattaagct ttgacagacg aggttaagaa cccccagaac tttattctc gctctcttgg
 4141 gaagtaacaa aaaagattta agtttgctt ttgaaacga taatttgcaa aagcttctgt
 4201 agctcagctc attttatgag tttctctct tttcttact gctccacctg ccttatttga
 4261 agcatctaa aattcttgag caagttttgt agccatctca cgtccaggtc tttggcgtgc
 4321 ggcttgtaact aaccaacgta aagcaagtgc cgttccactg tccgaattta ctccaagagg
 4381 aacttgatag ttgatccac ccactcgtcg tgattttact tcaactacag gtgtgcgtg
 4441 aggtacagct tgcgaaata ctttaaggc atctgttctt gtttgagtaa ctgctggagc
 4501 tctctttggg tttgagcat ttgtatttac cgtcctgaa gcaggtttt taacctatcc
 4561 tgaggttaatt tgatctaaag atctataaaa taacagctgat gctacactt ttttccatc
 4621 ttgtaataaa cgactaacca tcatacttac taacgacta cgataaattg ggtcaggtgt
 4681 aattgttctt tttttgctg ttttctacg agacatacta attttttga ttaaaccca
 4741 aaaagaataa aaaactattc accaaagcgc aatggtttt ttggtgacac tttgtctcg
 4801 gttttaaaaa tattattcaa aggaaaattc gaagctttgg attcgaattt gctgcaata
 4861 acaatttttg ataggggact agtaccacg gacttgaat cctctgtca ggttctgga
 4921 gtgctttcgg agctttgtgt ttttaataag ttgatcgttt agatgcttt ttatagatc
 4981 gataactgat gatctaagcg aaaaaattt attttttct acgaatctc catcactga
 5041 acagaaactt attttttagc tgctgtggtt gaggctttag ttgctccata ttttgatct
 5101 gagtttttac gatctttaa cccagcggca tctaattgtc cacgaataat atgataacgg
 5161 acaccaggtt aatcacgaac acgcccacca cgtactaaaa ctactgagtg tcttgagg
 5221 ttatgtccaa taccaggaat ataagcgtg atttcaaac cagaagtaag tcttacacga
 5281 gcaacttttc ttaaaagcaga atttggtttt tttygcgtcg ttgtatacac tcttgcacat
 5341 accccacgct tttgaggaca agttttat gcagggtgct ttgtttttt ttgtctttt
 5401 gttctgtctg aacgaactaa ttgtgaata gtaggcatac ttggcaaaaga ataaatttt
 5461 aaagtaataa tccccaaag tagtaaggcc aaagcctatc taaacattaa aggaccacag
 5521 ggccaaaggg cccccccga aggggggagg accacaaggg caaaggcctg aagggcccc
 5581 tcagctttca aggggtgaac tttatataat aatgacctga aaggaccgga ggtctgaagg
 5641 gacccccccg ctctcaagg ggtgagcttt ttttatgac ctgaaaggac cgggaggacc
 5701 aaaggtctca aggaccaaag gcttcccc ttcgggggaa ggggggtgag cgtttggtc
 5761 ctgagacct stggtctctt ggtccttgag gcctttggc ctcgggtct caaaggtcaa
 5821 agcgtcaat tacaccaggt caaagcgtc aaacacaaa gggcaaatc cccgagcga
 5881 ccaggtggt ttatttgaa atattttgac ttttatgta tcaatagtt ttaaaattc
 5941 aggatactaa aattataaaa tgtttaaata taacacaaa tactgtgtt aacataaagt
 6001 atttgataaa aaaactatgg acaattcatt ttttaaac taaaaataaa gtacgggatc
 6061 atagttcaat ggttagagca ccgcccgtc acggcggag ttgctgggtc gaatccggt
 6121 gatctcgttt ttaacttga aaaaaattgt gagcttctt tttcaacgaa aatctaat
 6181 ttcacaataa agttctcccc cgaaggggga ggactgtct tttctctct tetaaacccc
 6241 gcaaatataa ctcaagatccc tattacgctt ttaaaattgg ataaacttat tcaatggaat
 6301 gttttttagc tcttctgtg tccctactc cttagtgat ttttaatta tttcttaga
 6361 taatagctat atcttgacaa tgaagttaaa aacttgttca gcacactata ctctcattt
 6421 ttgatttgg tccaatttgt gtttcaacaa gttctaggat cgattttct tttcattaaa

6481 aatTTTTgaa gaataaagaa aatTTTTctt attctTTTTgt tatcgatttt gtgatgttat
6541 gacttggtat ctcttcaaac aagTTTTgtt tttggaagaa actaagtctg gcgggacgag
6601 ttaatgattt gaaaaatcag ccatcttggt ccctcaggaa tgcaaatcaa aggagttatg
6661 gctagtaagc caattccatc tatatgggca atttgaggca caaagtgcctg gtgatgagaa
6721 gctatagaac tatctaaaag atcaattgct ttggtatcac ctattttagg accaaaggaa
6781 cgaagttctt cccacgcgac cctccccctt tgggggggag ggtcgcgtgg gaaggaccga
6841 agggtttgag cactttgacc tcgtctactc gagtgTTTTg aaatttcagg tcagaaataa
6901 aagctcgaac acagaagatc tcccaatcct ttctcaaat aatattcctc aatTTTTctag
6961 gagaataaaa atgacaatca gtcctccaaa cactactaag gtaaaaatcg ttgtagatcg
7021 taaccaccga gaaacttctt ttgaaaaatg ggccaaccg gccatttct ctcgtacgtt
7081 atcaaaaggt cctacaacta caacttggat ttggaattta catgctgatg cgcattgatt
7141 tgatagccac acaactgatc ttgaagatat ttcaagaaaa gtttttagtg cacactttgg
7201 tcaattaggt attatcctaa tttgggtaag tggatgtat tccatgggtg ctcgtttctc
7261 aaactatgaa ccatggttaa gtgatccaac acatattaaa ccaagtgcct aagttgtttg
7321 gcctattgta ggtcaagaaa ttctaaacgg tgacgttggg ggtggtttcc aaggatttca
7381 aattacttct ggtttctctc aactatggcg cggaaagtgg attacaagtg aacttcaact
7441 ttatagtact gctattgggt gattaattct tgcaaggact atgttctttg caggttgggt
7501 ccaccaccac aaagcagctc caaaacttga atggttccaa aacgtagaat caatgatgaa
7561 ccaccactta gctggtttac ttggtttagg aagtttagct tgggctggtc accagattca
7621 cgtttctctt ccagtaaacg cacttctaga tgctgggtgt gatgcaaaag aatctcctct
7681 tccgcacgaa tttattttta atcgtgcttt aatggcagaa atttatccaa gttttgcaaa
7741 aggtctagct ccattcttca ctctacagtg gagtgaatc agtgatttct taacatttaa
7801 aggcggttta aaccagtaa ctgggtggtt atggttaact gatacagcac accaccactt
7861 agctattgcy gttctattta ttatcgtcgg tcaccaatc cgcacaaact ggggtatcgg
7921 tcacagatg aaagaaatcc ttgaagcgca caaaggctca tttactggcg aaggctcaca
7981 aggtctttat gagattctaa caactctatg gcatgctcaa tttaggacta acttagcatt
8041 ctttggttca ctttcaatga ttgtaagta ccacatgat gcaatgctc cataccata
8101 cttagctacg gattatggtt cacaactgct actatctact caccatattg ggatttgggtg
8161 tttttgtgta actggtgctg ctgcacacgc tgctattttc atggttctgt attacgatcc
8221 tacaataaac tacaacaacy tattagatcg tgtaattcgt caccgtgatg caattatttc
8281 gcatttaaac tgggtatgta tttcttagg ctteccacgc tttggtttat atattcacia
8341 cgacacaatg agtgcattag gccgtcctca agatattgtc tcagatacag caattcaatt
8401 acaacctgta tttgctcagt ggatccaaaa taccactca ttatctagta tttctgtaac
8461 tgcaccaaat gctttaggca gtacaagtca gagctggggt ggtgatattg ttgccgtagg
8521 tggtaaaaatt gctatgatgc ctatttctat aggtacatcg gatttctatg ttcaccat
8581 ccacgctttt acaattcacy taacagtttt aatcctttta aaagggtgtt tatttgctcg
8641 tagttccagt ttaatccag ataaagcaaa cttaggtttc agattccctt gtgacgggtc
8701 aggcctgggt ggtacttctc aagtatctgc ttgggaccac gttttcttag ctctttctg
8761 gatgtacaac agcttatctg tgacaatttt ccactttagt tggaaaatgc aatcagactg
8821 ttggggttca ttagggtcaa caggaactgt gtctcacatt actggtggtt actttgctca
8881 aagcctaac acaatcaacy gttgggttac tgacttctta tgggtcfaat catcacaagt
8941 aatccaatca tacaggttca gcttatctgc atacggttta atgttcttag ggcacactt
9001 cgtatgggca ttcaagttta tgttcttatt cagtggcctg ggctactggc aagaacttat
9061 tgaatctatt gtttgggca acataaaatt acgtgtggct ccttcaatcc aaccacgtgc
9121 ttaagattt actcaaggtc gtgctgtggg tgttgcacac tttttgtag gtggtattgc
9181 aacaacatg tcaattctct tagctcgtat tattgctggt ggctaagaac ctcccccgca
9241 aggaagaagg cctttggccc ttgggtcttc aaaaaacaaa aagatcgaaa ggaaaacttc
9301 tcttcggtcg cgtttttttc ttaaaaaaaa gtcaggaaaa aacttaggaa ctaaggctcg
9361 aggtcccgcc cagatcattt gccacaatc tttctctgtc cttagattgt tccaataat
9421 ttcaaaaatt ctttttctt tttttgcat cattggtttt ttatcttggg ttcatctca
9481 ataggaggat taatggcaac taaattccca aaatttagtc agggtttagc tcaagacccg
9541 acaaccggcc gtatttgggt tggatttgca actgctgatg actttgaaag tcatgatggt
9601 atgactgaag agaagctata tcaaaaaatt tttgcttctc actttgggtca attagctatt
9661 attttctctt ggacttctgg taattttatc cacgtagcat ggcaaggtaa ctttgaacaa
9721 tggggagcag acccgcttca cgttcgtcca attgcacacg caatttggga cccacacttc
9781 ggtcaaccag ctgttgaagc attcacacgt ggtgggtgct caactcctgt aatatttct
9841 tattcaggtg tttatcaatg gtggtatacc attgggtatg gtacaaatgt ggatttatac
9901 aatggtctag tattctact tttaggagct gcagcattcc tttttgctgg ttggttacac
9961 ttacaaccac aatttcaacc aagtgtttct tggtttaaaa atgctgaatc acgtttaaat
10021 caccacttat ctggactttt tgggtgtgag tcatagcat ggactgggtca cttagttcac
10081 gtagegatcc ccgaatcacg tgggtttcat gtacgttggg acaacctttt aactacgcta
10141 cctcaccac aaggtttaac actttcttc actggaatg gggctgcata cgcaggtagt
10201 cctgatactg ctactcact ttttggcaca tcagaagggt caggtacagc attattaact
10261 tctggttggg gtttccacc acaaacgcaa agtttatggc taagtatat tgcaccac
10321 catttaggca ttgcagttt attcattatt gctggtcacc aataccgtac aaactttggg
10381 atcggccaca gtatgaaaga aattcttgaa gtcacgtag caccagctgg aaaaataggg
10441 tctggacaca aaggtctttt tgatactgta aacaactctt tacatttcca acttggttta
10501 gctttagcga gtgttgggtg aattacttca ttagtactc aacacatgta ctcgttacca
10561 gcttacgctt tcttagcgca agattttaca acacaagctg ctttatcac gcaccacca
10621 tatattgctt gttttatcat gtgtgggtct tttgcacag gtgctatttt ctccattctg
10681 gattacgacc cagaattaaa taaaggaaac gtgcttctc gaattgttag acacaaagaa

10741 gctattatctt ctcacttaag ctgggcaagt ttattcttag gttccacac tttaggtctt
10801 tacgttcaca atgacgtaat gcttgetttt ggtacaccag agaaacaaat tcttattgag
10861 ccagtatctt cacaaatgga tcaagctgct cacgggaaat cattatacgg ttttgatggt
10921 cttctatcaa gcgcagggtg tgacgcttac gcagcgggtg aaggcatttg gcttccagggt
10981 tggtrtagatg ctattaataa tgctgggtgca aactcacttt tccctactat tgggctggt
11041 gacttcttag ttcaccatgc gattgcaact ggacttcata caacaactct aatccttcta
11101 aaaggtgctt tagatgcacg tggatcaaaa ttaatgccag ataaaaaaga ttcgggtac
11161 agtttcccat gtgacgggtc aggtcgtggc ggtacttggt atatttcagc ttgggatgct
11221 ttctatcttg ctgtattctg gatgttaaat acaattgggt ggggtacggt ctactaccac
11281 ttcaaacaca ttacattatg gcaaggaaac cctgctcaat ttgatgagtc aagtacttac
11341 ttaatggggt gggtacgtga ctacttatgg ttaaactctt cacagttaat taacgggtac
11401 aaccatttgg gtaatgaact gttatctgtg tatgcttggg tattcttatt ttggcacctt
11461 gtatacgcac ctggtttcat gttcttaatt tcttggcgag gctattggca agaacttatt
11521 gaaacattag cttgggcaca cgaacgtact ccactagcaa acttagtaaa atggaaagat
11581 aaaccagtag cactaagtat tgttcaagca cgtttagttg gtttagtaca cttctcggca
11641 agttatcttt taacatacgc tgcattctta ctagcatcaa cttctggtaa attcggttaa
11701 acttttgggt ttgaaccaa gaaccgtggt tgcttagcgc tttgaactta tggggagcac
11761 accagcactc ttcagacctc aaggaccaca gggccgaagg cctgaagggg accccccagg
11821 gccgaaggcc tccccccgaa ggggggagga ccgaaggctc gaaggggacc cctgctcttc
11881 ccactttggg ggaaggggtg gacgccttgg acctggtggt cttcagcgtt ttgacctttt
11941 aggaccggag gccaaaggcc tcaaggacca gaggtctcaa gggggtgagc tttatttatg
12001 acctgaagg tcaaaagcgt caccaccttc ccccgagggt ggaagacctt tggctcttga
12061 ggctttggc cctccgggtc tttcaaatg taaatcaaat tcaagcacac aatgagaaat
12121 caggtcagag tttttctaa agaaaatatt gaaattcaaca gcataatgtaa ttcagtggtg
12181 tttacccttt atttctagt aaacagagta cttctctctg tttcacttta atgaaggata
12241 gaaaatttaa aaaaacaaaa atgaatgaca gaaaaagggt tttatgaaag cgttccccag
12301 tgttttcgga tctgttatac gagtgggttt aattaaatt tttctcgtaa tcaaaagaaat
12361 caaaattaaa ctgacttga gcaactcgaa ctccccataa ttgttaagag ctagaggccg atcaagtaat
12421 tgggttttcc cccaaagggt gggatagtga acatcgttcc cctctcata acttaggtct
12481 tttggtactc gtttgcctga gcttgattta tggcttaaaa aaaaggattg gaggattaaa
12541 gggccacagc gccaaaggcc tgaaggggccc acccagctct caagggggcc cccagctctt
12601 ctgtgcttga gcaactcgaa aagtcaaaat tcttccagc taagtataaa atcaagctca
12661 agcacacaag agctggagca cagtcaaac tcttaagcac acgaagcatt agccctttgg
12721 ggtgagaact ttgctgcttg ggggaaagaa aattcactgg tctttagagg taagaacttt
12781 aaaaaaatgt ataggaccgg aggtctcaag gaccaaaggt cttccccctt cgggggaggt
12841 tttactcttt tgetttaaaa aatcttgtct tatagacctt ttacaaaatt ttgaaatgat
12901 gttggatcga ggattgctaa ttgtgacaat actttacggt ttaaaagact tttcgtttt
12961 tttaaattat taatgaagcc actataacga aaaaactctg ggtttttaga atctgtattt
13021 gtggtggtat tttgaatgcc aatatttcta agagctgagt ttaaaccgag aatccataaa
13081 gtagcctctg cacgtttgaa ttgtcttcca tctgtatagc ttaatgataa agcgcgctt
13141 gttcgttgtt gggctgttctg aaataaaaaca gaatgagaac cacggaacc tttagctaat
13201 ttaagcactc tttttctctt tttacgtgct acaaaaccac gttttactct agtcataagt
13261 attattttat acgaacaaaa aaactcaatc aaactttact tttaaaattt tgatgtggt
13321 ttttgacgat ttttagttca aatgttaatt atctttctta cgcgaccctc cccccagggt
13381 ggggaggcct ttggccctga aggaccaaaag gaccgaagggt ctgaagggcc cccccagctc
13441 tttccccatt cggggggaagg ggggtgagcac tttgacctgg tgtgctcag cgtttgacc
13501 tttgaggta tatattaagc tcaatcacac cagacctccc cccgaagggg tggggccttt
13561 ggcctacggt cctttcaggt cgaataaaag ctccccctt tccccgaag ggggaagagc
13621 tggggggccc cttcagacct tgggtctca aggggggtgag cgttttgacc tttcaggtca
13681 taaataaagc tcaccacctt gagacctctg gtcctttggt ccttgagacc tccggtcctt
13741 tggctctttg tgatccttgt tttttgacc tggagtacga gaagcaccct aggtcaaaaag
13801 atcaaaagta gcaaaattgt tacagttttc cgtataaaaa gatttttctt tttccggtcc
13861 tttaaagatga agtctctaac aatttgttaa cttagcagga ttttttatgc aagaatttgt
13921 gatcaaacac cagcacctac agta-cgtccg ccttcacgaa tagcaaacg cataccattt
13981 tcaatagcaa ctgggttaat taattcaact gtcattttta cacggctcgc aggcattacc
14041 atagatgat cagaaccgtc atctgcttta aagttttcaa tttttctgt tacgtctggt
14101 gttcttcat agaatgtggt gcggttaacca ggggaagaatg gajtatgacg tccaccttct
14161 tcttttgtta gaatgtaaac ttgtgattca aatttagtat gagggtaat actgcccagg
14221 gctgaaagta ccattccacg ttcgatataca gttttttgaa taccacgaag taaaatcctt
14281 acgttatcac cagcaacact ttcatacaagt gttttttgga acattttcaag gccagttact
14341 gtatgtgttt ttgtgttttt taaacctact acttctacag aatcacctac tttacaacg
14401 ccacgttcta cagcactgtt tgctacagtc ccacgtccag taatagagaa aacatcctca
14461 acagccatta agaattggtt atctgtttct cgagcaggtg ttgggataa tgtatctact
14521 gaattccataa gttggttaaat tttatctacc catttatttt taccacaggt aatatttgg
14581 ctttcaacaa cagcttctaa tgctaagtgt gctgatctg gaacaatagg aatctcatca
14641 ccaggaaatt catatgaatt taaggcttca cgaacttcca tatsaacaag ttctagaagc
14701 tctgggtcgt caacttggtc tttttgtttt aagaatacaa ctaaatgtg tacaccaact
14761 tgttttctga atagaatgt ctcacgtggt tgtggcatg ggcacacag agctgatata
14821 actaaaatag cgccatctat ttgagcggca ccagtaatca tttttctac gtaatctgca
14881 tgtcctgggc agtcaacgtg tgcatagtga cgtttttctg tttcatattc cacgtgagct
14941 gtatttaattg tgataccacg tgctttttct tcaggtgctg agtcaatctc atcgtatttt

15001 ttacctacag aacctccgat agcagcaagt gtcattgtaa tagcagctgt taatgttgtt
15061 ttgccgtgat ctacgtgtcc aattgtacca acgtttacgt gagggttttt acgttcgaat
15121 tttgcgcgag ccattttgtt tttttaagta tgaaaaagag aacataaaat aaaacttttt
15181 ctcattgatga caactaagca tttttttact tctatttctt gttgtcttta aaactttatt
15241 catttggttac caccccccct gagcttcggt tctctgggtc ttaaggggac cacagctttt
15301 ttgtgctaga gcttgattta tgacctgaaa aggcaaagcg ctcgagcaca caagataaa
15361 gcgctcaagt acataagggt tttattttga gtttttgtt gttctcttaa acgtttgatt
15421 ttggaataaa atcttggaa gttctatatgg tgaatacaa ctgctctcga gctgctctgg
15481 gttgaattct acggcccaaa gcaaaggacc ataggaccac agggccaaag gctctcccc
15541 ttcgggggaa gggggtgagc gctttgacct ggtgtgctc agtgccttga cctggtgga
15601 ttgagcgctt tgacctttga ggtcagatat taagctcaat cacaccagac ctacggtcct
15661 ttcaggtcat atttaaagct gaagcacacc agacctccc gtcctttcag gtcataaata
15721 aaagctcacc cctttgagac cctttgctct cccggtcctt tcagggtata aataaagctc
15781 acccccttcc cccgaagggt gaagacctt ggtccttgag gctttggcc ctccggtcct
15841 caacggtcaa agggcttcac cctttgagag ctgaagggt caccctttct cagaaaaaag
15901 atattattat gtggcctgaa tgggatagtg aaagaagaat tttgtttcca agaagtctt
15961 tttcttgaat ttaagaaat ttaaaaggag ttaatttttt gagctcacta gtaaaagctc
16021 ctgatttaga aacttagaag gtttctgtcc tatccattaa acgataagcc cttaaaaaaa
16081 aagatttttt ggcctttgat tgggctttt ggcagttggg tcgaaaaagat cagtgaaac
16141 aacaattatc cgtaactgaa gatttgatcc tcagaatate tgaggaagt atttttttt
16201 ttattgaa gtaactgaa aagcattggt aaaaatttgt aggaacggat ttcagtttta
16261 ataaatatat tctttaagta taacacatgg tgttgcctt tttctcatt tgccttgact
16321 tcaataaacg gaacgataag atctcagtaa taaattcttt ggaatattta caagttttta
16381 tttcaactga ggatcaaat agaagtttat gctatattaa gaataagccc atgtggcgga
16441 atggtagcag cgcttgattt aggatcaagt accttccggt gtggagggtc aaactctctt
16501 atgggcaatc acaaaaataag ttaaatagc aaaaattgct tcaaaaaata gaactgaagc
16561 tcttccaaat attcaaaaat aggggaacaa agggccgcaca ttttgaggga tcatcttaa
16621 tagatatgat tttgtgtgct tgagcgctta tgacttttca ggtcataaat caagctcaag
16681 caaaaaagag ctggggaag gaacaaaatt cactctcccc agtcaaggaa taaatttttt
16741 gtttgggttc tgttaateca gcgaaagggt ggctagaagt aagaagggtt tttgctaaa
16801 atttgggaaa atgctcagat ggttaaaagag ggtagattgt aaatctatta gctttgctta
16861 cactgggttg aatccagttt tccccaaagt tcttatagta agatccatat aaattcttt
16921 aggttgtgct atataatagc aacatatcgg agtgtagcgc agtttggtag cgcatactgt
16981 ttgggacgat ggggtcgcag gttcgaatcc tgtcactccg ataggagtga caaagtttt
17041 ttaaaaggaa atcaaaactt gctctttttt tgtttttgtt gacgagaatc aaagtccagc
17101 gaagttagat tccagagaaa taaaaaaagg ctcaatggat aatgcttttt tgggtgtatg
17161 gccaaagtgt aaggcagag attgcaaatc ctttaacccc agttcgaatc tgggtaccgc
17221 ctttagcata acaatggcac cagatccggt aaaagcgcgt tccagataat taaaaaacat
17281 ctcaggaggg ccaaaaggacc aagaactatt ggagagaggt ttttaaaaga cgggatgacg
17341 ggactcgaac ccgtggcttc agcttggaa gctgaggtat tacccttata cgcataccgc
17401 aaacacctga tgtgattgag cactttgacc tttcaggtca taaaataaaag ctacacctt
17461 tgagaccttt ggtcctccc gttcctttgag gtcagatatt cagctcaatg acaccagagc
17521 tggggggccc cttcaggcct ttggccctac ggtcctttta ggccataaat caagctcaag
17581 cacaccagac ctcccggctc tttgaggtca taaatcaagc tcaaccacc gagacctta
17641 gtcctacggt cctttccttg agggccaaag gctcaagggt tttccccttg tgggccaag
17701 gcctcccccc ttcgggggga gggctctctg ggaagaactc cggctcctcc ccgaaggggg
17761 aggccttttg ccttttggtt cttaggtcct taaccccgaa caggatgcaa gttctctct
17821 tacaactttg atttgcacgg ggagaacttc gtgggtgctc tacaaaaatca agttttcag
17881 cctctatagga attgaaccga tgacatcagg ttttggaaac ctgcyttcta ccaactgaa
17941 taaaggcgct ttaattctata actaaattat actccttttc aaaaaatcaa atgaatgcat
18001 tattgtaata gtttggaaact ctaaaaagtt taaagtttga ctcaaaaagt agtataaac
18061 tcttctgggtt attatccaat tgtttttctc agaagttttc ccttctgctc aattcaagt
18121 gaaggttttt tttcttcttt attttaaaga tataaaaagt tttctctctc aagaattgct
18181 aaaaatcaa aaatggggtg ctaactcaat ggttagaggtg cgatctgtta taaaggttcg
18241 gattcctttt caaacctaac ttaaaattgg aatattcatt ttttgaatac aaattctttt
18301 caaagaaagc gaaagcaatt cgctggtgga gactcacta ctacagaaaa aaaagattat
18361 ctgctcagaa aacttgtgaa ataaaacgat catctataga atctaagag tctctatcct
18421 ttttagttca agacgtagt tcttagaaaa acgtaaaaca agaattctatt tttcaaaaac
18481 aattttagga acagataaaa atgaattaga gaaattttta taaaataaaa atttgtcttt
18541 tattcactaa aaggatgttt aaaaaagcca aagccgataa gtaatttatc tggatattgt
18601 aaatttaaac taagaattat ataaaacttt cttaaagtaa ggaattattat tcttaaggag
18661 ccgtgtgaaa tgtaaaattc aagcacggtt ttgaagtcaa aaaagcaag caatcaggac
18721 caaaagtctt taatgttttt gtttctttta gattaactac tccgctttta accgattagt
18781 tctgggttgc aatcccaggc atcccaaaat gaaatagaaa aaattacaac aaggacaaa
18841 gaaaaaagg gtttaccctt ttgggggaa gttttatatt tgttatcctt cagcttgggt
18901 taatgttttt aattgtttgt cttgacatgt ctaatcatcg gtaaataga catgtcaaga
18961 caactttttt tgttgttata taacaactcc aattttaaat atttagaatg aatgaaaaag
19021 gaactagcat tctgtgaagca caaaatcgtt cgatgaattt tcaagaacgt tttagttttc
19081 ttgtttttat tgggaaattt cttgtatcga aatttggact ttttttctg ttttaattt
19141 tacaccttct tcaagttat caagccaaaa ataattttat taagtttgc tctctgaaag
19201 aggaggggat caaaagcctt gttgtttctt ctacagaatc atttgttact aatgaaaatt

19261 ctctcgcaaga tcaaaaccac aatttatctc caggaaacaa aaataaggaa aaaacacaga
 19321 tttcgcgatgaa agaaggtaat aacagtttaa tattccactc ccctcaaat gaaaaagctc
 19381 aattgatagtg ttttgaaaa caaatatcat taaaagatat tgggtgcttt caaaaaatta
 19441 aatattattg tataatagct acaatagtta ctttatctt aaatccacaa caacaaaaat
 19501 attttggagg attggctggt gttttaaaac aagctgggtt ttttacgaaa tatggaaaat
 19561 cggaaaaaat ggtttatatt ttttctggg tttttggtt tctgttttc tatcttcaa
 19621 ttgatattaag caaaatagct cgttctttaa cgtaatttc ataaaactga agaagatata
 19681 ttgctgcttg acagaagtag tgttttgctt ctaaaatgaa tattaaagaa attacttatt
 19741 aaaataaaac ttggttttat cttaactttc ttgccttggt tcactgagat ctttgacctt
 19801 tgaggacagg gaagctcgtt atgtttcagc tttatttatg acctgaaagg accgtagggc
 19861 caaaggcctg aaggggcccc ccagctctgg tgtgattgag cttaatatct gacctcaag
 19921 gaccgggagg tctgggtgac ttcagcttta tttatgacct gaaaggaccg taggtctggt
 19981 gtgattgagc ttaatatg acctgggtg cttagcgcct ttgaccttc aggtcagatt
 20041 tcaagctcaa gcacaccaga cctcgggtcc tcaaaggta aagcgtcaa tcacaccagg
 20101 tcaaaagcgt gaagcacacc aggtcaaaag gctcaatcac aagacttca tttcttttca
 20161 tcgagtagac tagattatag agataaaaa cagtaaaagc tagaaattg aggaggttta
 20221 ttaggtctta atttattgt tgccatctaa tttttgattc aataaaaaat ctttttttaa
 20281 agaatcaacc ttgattttt agaataaaag caaaattgc ttgccgaagg attaaaattt
 20341 gctgtgcctt gcacactaag cctaaccgtg tcaaaaaaa ttcatttttc tactttacat
 20401 ataattggaga aataaaatgt caccacaaac agaaactaga acaaggcgtg gatttaaagc
 20461 aggtgtaaaa gattaccgtc ctacttacta cactcctgac taccagtag tagttactga
 20521 tgttcttgca gcattccgta tgactcctca accaggcgtt ccaccagaag aagctgggtc
 20581 tgctgtagct gctgaatcat caacaggtac ttggacaact gtatggactg acggtttaac
 20641 tagcttagac cgttacaag gccgttggta cgatatcgag ccagtacctg gcgaagataa
 20701 ccaatcacac cttacgggtg cttaaccaat cgacttattc gaagaaggtt gaagtaaaaa
 20761 cttatttagc tcaatcgtt gtaacgtatt cggtttcaaa gctctctgtg ctctacgtct
 20821 tgaagacctt cgtattccac aagcatactg aaaaacttcc caaggtcttc cacacggtat
 20881 cgttgtgtaa cgtgataaac taacaaaaa cggctcgtgt tctttaggtt gtacaatcaa
 20941 accaaaaacta ggtcttctag ctaaaaacta cggacgtgca gtatacgaat gttctcgtg
 21001 tggtttagac ttcacgaaag atgacgaaaa cgtaacatca caaccattca tgcgttggag
 21061 agaccgtttc ttattcgttg ctgaaagcaat ttacaaatca caagctgaaa ctggtgagat
 21121 caaaggctac tacttaaagc ttacagcagg tactgctgaa gaaatgtaa aacgtgctca
 21181 atttgcaaaa gatttaggtg tgccaatcat catgacgac tacttaactg actttctac
 21241 ttctaact tcaattatct actactgtcg tgataacggt cttctcttc acattcaccg
 21301 tgctatgcac gctgtaattg accgtcaacg taaccacggt attcacttcc gtgttcttgc
 21361 taaagctctt cgtatgtctg gtggtgacca cttcactca ggtactgtag taggtaaact
 21421 agaaggtgaa cgtgatatca ctttaggtt cttagactta atgctgtag ctgttgtga
 21481 aaaagacaga agccgtggta tttacttcac tcaagactgg tgtggtctag ctggtactat
 21541 gccagtagca tctggtggta ttcacgtatg gcacatgctt gctctgttg aaattttcgg
 21601 tgatgacgct tgtttacaat tctggtgggt tactttaggt cacccttggg gtaactgtcc
 21661 aggtgcccga gctaacctg tgctctaga agcttgtaca caagctcgt caagaaggtc
 21721 taacctgtct cgtgaaaggt gcgatattat tctgctgct tgcaaatgga gccctgaact
 21781 agctgctgct tgtgaaagtt ggaagaaat taaattcgaa ttccaagcta tggatactct
 21841 ataataacga gttttactt aataaaaaat ttatttttt tattttaagt tggggggggg
 21901 aacccccccc ccccgaagat aaaaacccga gcagaaatc ttaagatcca gaacaaaaact
 21961 acttgggtac aactttatgt ggtctgaaa ttactcaaag ttcactctca ggtgcttgag
 22021 cttaatatat gacctgaaag gaccctaggg ccaaggcct cccccctcg gggggaggac
 22081 caaaggctgc aaggcccccc cccaagttct ggtgtgctg agcttgatt ctgcataaaa
 22141 aaatcaagg gctcaagcac acaaagtcga agcaagcatc cgaatgatg cgtataaaaa
 22201 tttgttctg aaactaaaga taaatccaga ctattggtaa gaccattatt actaatttct
 22261 ttatttgtat tctgatcttc ttttactctg atttaagatt tttctcccc accgcgcga
 22321 agccagttt gagcttcaat ataattagtt ggtgctaaac gaatggctc cgcacaata
 22381 tcagccgctt tttcaataa aatttctgag gcttgcgtat cggcagctc aatagcttgt
 22441 tcaccacgat aatggtaaat gacagcaata ttatttagt cttgtggaag tgatggatta
 22501 ctttctaatag cttgatagta atattctaaa gcttttgtat gttctccgtt acttgtatgg
 22561 attaatacaa tattataaag aataaactt cggctcatat catcaactc caagcgcaaa
 22621 gcttcataat agttttgaag agcctcagca tattcacctt ctgcttgggc tgacattcca
 22681 tgcgataat atgtaaaagc ttcttttct ctttttgtt tgggtaaaac ttgaagtaga
 22741 atatccgcaa ttaactgtaa tgttttatca ataaaattgt catttcgtt tgaacgaggc
 22801 atagaattta attttttaa tagttgaaaa aagagagaag aacgaagtc ctttgaaac
 22861 ctcaaggggg tgggggtgag ctcttgacc ttggtgatt gagecctttg acctgggtg
 22921 cttcagcgtt ttgacctggt gtgattgagc gctttgacct ttgaggaccg gaggtctggt
 22981 gtgcttgagc ttgaaatctg acctgaaagg tcaagcgtc caagcacacc aggtcatata
 23041 ttaagctcaa tcacaccaga cctacgggtc tttcaggta taataaaaag ctacccccct
 23101 ttgctctccc ggtccttga ggtcatatat taagctcaat cacaccagac ctacggctct
 23161 ttcaggctcag aataaaagct ccccccttc ccccaaggg ggaagagctg gggggccct
 23221 tcagaccttc ggtcctcccc ccgaaggggg gagggtcggg tgggaagaac tccctcctc
 23281 ccacgaaggg ggaggcctt gccctaggg tcttgaact ctttgcgtag gtgcttcca
 23341 ccgcccttga agctatttta cctttttga gtaaagttct gctttctttt tctctcttt
 23401 tgatgatgaa caactgctag gtacgagaat caagataaga gttggcttaa atataattaa
 23461 gccactctt atcttgcctac tctttatctc caagttttg aagaatttca tttactaaaa

23521 cacgacctgg tgttgctcga attgctttat ttcgctggta ggcttttgta tcttgggtcc
 23581 atttataaga atgatatact tgatttgcga aaccagatga aaaaattcga atataaagtg
 23641 gttcattaaa attgcctttg ccacctcaaa ttcctactga actataattc attttatcct
 23701 ggtgattttc accaggctact ggctctatat tgaaaaattt gtttgatta ttctcaccaa
 23761 aaatctgggt ttttaagttt cgcgttccgt tccacactcc gggagctttg tttaaaagaa
 23821 aagctttttg tttttcaaaa gtttttgagg aacggagttc ttcctccgcg accctctctt
 23881 ttctcccgtt gtgagggcca aaggcctcga agggttcccc tcgagggctg cgggggaagg
 23941 gggtaacccc gcttttgttt tgatttgcta tttgtacgat ggattgtttt ccagcccaaa
 24001 gatttctttt ttctccaggc cataaatttg tttttgcag ccaaagaggt tgatgtaaat
 24061 cgagtccatt tctgttgat gggccgtttt caacaatttc tataaaatga aataacaaag
 24121 agttttttat atcaactttt tttgtttttt gttctctctc attgttttga tccaatttt
 24181 cattfactgc agcaggcctt gattttttcc aaggaacgaa gttctgtgaa ctctttccc
 24241 cctctggaat attaacgctt ccgcttcat tttagagaaa attgaattca atttggctct
 24301 ttggttgtaa agttgttaaa taataaaaac ctaaaacat atcttggctc ggtaaaacac
 24361 ttggttgacc agtagctgct gataaccaat ttgttgtagc caacataagt aatcgagctt
 24421 caactcttgc tttaggtgat aaaggcacat gaacagccat ttgatccccg tcaaagtctg
 24481 cattaaaagc tgaataccct aaacgggtga atgttgggtc tctatttaat aaaaatgggt
 24541 gttgaaaagc ttgaatccct aaacgggtga atgttgggtc tctatttaat aaaaatgggt
 24601 gagattgtac tacattttgt aatatttgtc ttgcaaaggg ctattttaat tttaaaagat
 24661 ttttagcgcc acgaatcgtt tttagcaagt gtaattctaa aatatggcga atcacaanaag
 24721 cttggaaaag ttccaatgct atttcaatag gtaaacccga ttcctgaagt tttagcgtg
 24781 gaccgacaaac aataacagaa cgaaccagaa aatctactcg ttttccaaac aaattttgac
 24841 gaaaacggcc ttgttttcca ccaattctat ccgataagga tttatagact ggaagattgg
 24901 acccttttgt taatggttta ttcccacctc cactcccaga atttgacata cgtcttaaac
 24961 ttccaatttc agaggcaga cttgagccat ctaacaagga atccacagca actgttcaaa
 25021 gatgttggct acttcgact aaaaattcaa ctaagttagg tccacaagct gtaaaaataat
 25081 attttaatcg attatttcta ctgattagac gacgatataa atcgtttaca tccgaaactg
 25141 caaatttctc tgcacttaat tgaacaattg gccttaattc tgggtgggaga actggaagag
 25201 ttgtaaaaaa aaaaccatca aaactttgtg cgcctccagc tgttcttaaa agatlaagaa
 25261 gttttaaagt acgtacgcgt ttttttcttc tttttaataa attttcatac attgaaaaac
 25321 ttggtgtaag agaccaaata ttttctgttg aaaaacttcc tgtaccggag acatgataac
 25381 ttcttctctc tgaacactcc gtctttgttt cttttatact ttactaaaa aagacaggtg
 25441 ttctgcgtaa atcatcttgt aatttttcaa attctttttt ccaatttctt gttttaaagt
 25501 gttctttaa agetcttctt ccagtatatt cttagccaaga aggatttatt ggtacaatat
 25561 ttggtctctt ttgcgcaatt gaatccaatg ctgttttttc tgaanaaccaa ctccatttt
 25621 ttacagattt ttgaagcttt gaggaagaaa gttctgtgaa ctctgttccc ccccttaaaag
 25681 tcttttttct tgatttaaaag ttttaattct aagatcagtt ggtatttgcg cctaaactat
 25741 gattttgcaa aatcaaacct gactcaaaaac ttgttagacc aaagggccaa aggccttccc
 25801 ctctgaggcc tttggccttc tctttttctg cgttgtgagg gccaaaaggcc tcccccgaa
 25861 tgggggggaga ccttcggttc tcaagggggt cgcgtgggaa ggtctgcccc ttccggagag
 25921 gttttgacga acgcggaaga ctaaaccata catgtaaaaa ttccatttca ttatcgaaaa
 25981 aatctctctc tggacctaga cagggatatt ctactccaaa aacaaaaagc tctttttgta
 26041 atttggtttc taatcgggtg ggaactttt ccgtatatgt tgaaggcttg tctctccca
 26101 ttttggatag caaaggggaga tcttgcctct ctcttctctg agaaagtctg aattctttt
 26161 gtgctggact ataccacggg taatatgtaa ttttttaac atattttgtt ggtattttta
 26221 gcaaaaactg aaaaatattt ggacgactat ttaaaatcca aatattgatt acaggacaac
 26281 ctaattcaat ataaccatt ctataacgct gtactcttga taaagtaact tccacttgac
 26341 aatttggaca aatataaggt ccaagctcta cactttgaaa tttttctcga tttgaaaatt
 26401 ggggtgaact atttggattt tctctcagatt tctttgtgga tcttttctca gctcttcaa
 26461 taataatacc tcccacttga ggcctttggg cctctgtccc agtagggctt tgaggagccc
 26521 aatctgattt atcaactgaa agattgggta gacctgaatt aaggagtga tttctgact
 26581 gcttaaaccc aaaaggtgaa ataagttctc ctccccgc gacctcag gggaaacctt
 26641 cgaggccttt tgcctcaca acgggagaaa agagagggtc gcgggggaa aactccgtc
 26701 ctgtattttt agataaattt gcagttgcgc gaggattttg gcttttccc actttaagtt
 26761 tggattgacc acatgcgcaa acccaattct tctctggacc aaaaattctt tcacaaaaca
 26821 aaccgtcttt tttagcttca aaactacgat aattaattgt atcagctttc ttaacttcc
 26881 caattttaat ttctattacca tetaaccata atgagcgttc agcccaactt tgaatttctt
 26941 ctgaactggc caatccaatt tgcataaaag tcttcatatt ttttgcatca aagattctag
 27001 gtgattaaca agttcacgcy gggagggccg aaggacagaa aagggagaa aaaagtccac
 27061 tggctttcga tctctccccg aaggcgagag gctcgggtgg cttagactt ttggactccc
 27121 ggtcccttca ggtctctatc aaagcccaaa cccacgagaa ctctgtttct tatcaactcc
 27181 tctctgattaa tggcttttgg tttaaagagg cttgaaacta aagacttctg gttctttcac
 27241 actagaaacg aaaataaggg atcttctctc agggaaagag aaagatggtt gaccgacttg
 27301 atttttaatt tataaggtca agtgcctcaag catacgagga ccaaaggccc atgtgttct
 27361 ttcaggctcag atatcaagaa gggggggggt cccctctct tagaaggatt ttgttagaca
 27421 cagtgaaagag attttgtgct tattttagca atctaaggac cggcggctc aaggacaaa
 27481 ggaccagagg tctcaagggg gtgagcttta tttatgaact gaaaggtcaa agcgtcacc
 27541 ccttgagga ccgaaggtct cccccattc gggggggaggt ctccccctt cggggggaag
 27601 ggggtgactt tatttatgaa ctgaaaggac cggagttctt cccccgcgac cctccccga
 27661 aggggggagag gtccccgggg aaggacagac tagttggcat ctttggtttt cagctctaac
 27721 caaaaaaaga gtttggaggg aggttctctc tgtccccctt tgaanaaac tatttttatg

27781 atgtataccc tgttccctgca ggaatcatac gacctacaat aacgttttct tttaaacctt
 27841 caaaaaaatc aacacgactt tgaagtgcag cagtcattaa tactcttttt gtttcttgaa
 27901 aaccttggcg tgaaatataa ctctcgggtg tagataaaag ggctactttt gtaattcccta
 27961 ataaaaatcca ttcatatgag aaccttgaat tctcatcact caatacattt ggactttttt
 28021 gaatttcttt tgataatagt gtgtttgaac caaaagcttc aataatatca tctggataaa
 28081 atgggggtgtt tcctttttct aaaattcgta cgcgtgaagt catttgacga acaataattt
 28141 ctaaatgttt atctgcaatt tctacacctt gggaattata tactgattga atttcttgaa
 28201 ttaaacggttt ttgtaattca cgaagttcag ttttttaac taaagtacga agtctctccc
 28261 ctgtgactcc ttctatcgag ttgtgagggc catttgaagt ttttctctc agcccttaa
 28321 cattcgaatt aaagactgat gtttgttctt gtgggttttt ttgttctaaa tttttaaatt
 28381 tttctttttg ctttgaaagt ttattgatgt cacttttttt ttcagttttg tttgctaaaa
 28441 gttttgtatt ttgaactgaa ctttcatttg cagagcgtgc ttcaaacaat tgttcaattt
 28501 ttggttaaacc ttgcacaata tctccagaac ttgcagaatc ggcaattgga agaactaatt
 28561 gaaataataa atcacgttct ttgatatac gtgtacgttg ttttctgctc gtattcggat
 28621 ttgatccaga cattcgtgat actccgtttt ttccttggaa ttgcccctct aaagaggttt
 28681 gaaacacgag aagaatgagt ttatttaaac gttttcgaat aatttgcgtt ggaacattaa
 28741 ttgctgggtg taaacctaca aaaaaagtat aaggttgaag aattttctct aaatgactta
 28801 aactaaaagt attctcaagc atgctataaa gtgggggaagt caatgaaaga tttggtttcg
 28861 ctctattttg ggtccttccc acgagaccct ccccccactt ggggggaggc ctttggccct
 28921 caaggggttt cccttgagcc ctttggccct ccccccactt ggggggaggc ctttggccct
 28981 cacaattcga ggaaagagag ggccaaaggc ctcgaagggt tcccctcgag gctcgcgtgg
 29041 gaagaactcc gttcctttgt tctttttggt gactaaattc taaaacaagg cctgaaaaag
 29101 aaagattttc atgaataaat tttttttgta aatttgtttt acgaaattca aaacttggat
 29161 ttacaataatg tactgaattt gtttttagtga gtgaagtgta agattctcca ttaagccaag
 29221 gtttagtttt cgtctctttt ccttccccga aggcagagag tgaactttgt tccccctt
 29281 gattaaaaatc tttttctgtt aaaagagtca gaatacctga ggagttttct gttccctttt
 29341 tatttttcaa aatattaacg gaagtttcta accaagagcc aaaggcctcc ccttccaggg
 29401 gagtacctga gttttcaccc tgttttccaa ttctcaaat atttggctct ttgtcatttt
 29461 cattcattgg aaaaattttg ccattgacaa tttcatcctt gctatttgat tcggaactga
 29521 gtgtattcaa aaaaagggct tttgttaatt tattatccgg agacttttgg tctctggcca
 29581 atataaagtg ttttgaagat aattttcgtt ttcctttcca aaaaacaata tctccacttc
 29641 gaaagggtaa ttgagaaggt ttggctttag aaggtctttg acttttattt cctgctttga
 29701 tttctttctg aaatgcgctt tttgtctggt taaacgaaaa aggttcacct gctaaaacaa
 29761 gacgagcttt tgaagttttt gttcctattt caggtttgta ccttctcctt tttggttggg
 29821 gattcgcttt tggactttgt gcctttaatc ctttctcgaa aagcgggtta agaactctgt
 29881 ttccccctcc aagcttttgg ttagtaaac ttttaagatc aaaaatattc gtttcaacaa
 29941 ctctcctctg aattttctgag aaaacttttt tttgagctaa accactacca tcagaatttt
 30001 caccatcatt ttcaaggggg ctaccccccc tctgagtagt tgcttttctt cctttttgta
 30061 ttgaaagacc taaagtttga atgtcactcg tcaattcttc atttttttca gctgggtgtt
 30121 taatttttga aattttcggca caaagcgtat ttttagcaat tgtttttcca ggtgtaataa
 30181 ataaaaatct ttgagctgga agttgaagaa tagaaattgt attgttattt tggttattga
 30241 tatcttggag ggtccccctt ttttctctt caatttttaa ttgaaatagc tcaaagaaaa
 30301 aaaaatgccg ttgtccatga acactacgaa ttttacaacc aactgaagca agagagaaat
 30361 ggggtggacc aaaaaatgat tgatggtttt tatcctggct ttgagactct gacaatattg
 30421 ctgttttggg ggggtcccc ctttttagcc aaaaattgttcc tccagccggg cgtgaaattt
 30481 tatcttctac ttctgttgg aaaattccac cgggatgaaa tgttccgatt gtttaattgtg
 30541 ttcggggttc accaatcgac tgagctgcca gaattccaac agcatctccg agtgcaacta
 30601 atcgacctg ggctaaatc cagccataac acaattgaca aactgattga cttaaacatg
 30661 ttaatggaga acgaaactga attgaagctt ttgaaacttt taaagttttt aattctaaag
 30721 ctaaatcttg agaaatttct tgatttttcc aaatagtcgg agttcgttct ttaatactct
 30781 cagcaagaac tctcccaatg atatttttct tatcagagat tttcattaaa atagattccg
 30841 aagtgccaca atcaatttca cgaatgataa tccatgtgc tacgtcaacg agtcggcgtg
 30901 ttaaatatccc agagtcagct gttcttaaag cagtatcaat taaacctttt cttgaccatt
 30961 aacatgaaat aacatattct gttacactta aaccttctct aaaaattgctt cgaattggaa
 31021 agtcaataat gccaccacct gaatcagaca ttaagccacg cattcccaca agttgtcgaa
 31081 cttgagaaat attacctcgg gctccagaaa aagccatcat ataaagtgga gatgctaagg
 31141 gccacccttg aatttttga aagttttgaa ttacttcatc ttttaatttt tcaacttgtg
 31201 ttgtccaaat atcaatcagt ttttgataac gttcaaaagc tgtaattctt cccattagat
 31261 aggttttttt acattcttga acttgttttt gagctccagt taaaagaagt ttttttgcag
 31321 cgggaattct taaatcatca aaacctaatg aaagacctgc ttttctgca taatgaaaac
 31381 ctatataatt aaagtcatcc aaaaggagac ttgtacgcgt ttgcccatag tggcccataa
 31441 accaggaaac aatacgtttc agtctttttt tatcaagtac ttgattataa aaattcattg
 31501 cttatatttg tataaaaaaa ttgaaatcgt aattttgaaat tccgttggag gctcccctga
 31561 actttgttct tgggggaaga actacctggg gtgattgagc gctttgacct ggtgtgcttc
 31621 agcgttttga cctgggttga ttgagcgttt tgacctttga ggaccggagg tctggtgtgc
 31681 ttgagcttga aatctgacct gaaaggtcaa agcgtcctaa cacaccaggt catatattaa
 31741 gctcaatcac accagacctc cggctctttc aggtcataaa taaagctcaa gcaccaccaga
 31801 gctggggggc ccttccaggc cttcggccct caaaggggtg agcgttttga cctggtgtgc
 31861 tctcagcctt tgacctttca ggtcatattt aaagctgaa caccaccagc ctcccgtctc
 31921 tttcaggtcc taaataaagc tcacccccct gagacctctg gctctttggt ccttggagcc
 31981 tctggctcca cggctccttc tttttcaaa aatcaactt aaacaacata gagtaagctc

32041 aagcccttga gagctgaagg gaacgaagct cactggatgt tctcacttcg aaggaagata
32101 ctgaaacttcc ttttcttggc ttattactta cgaggtcata aatcctaaatg tatcaaat
32161 ataaaaactt taactcggta aatttgaata aacacttaaa cataaggcat gaagtctggt
32221 aattaaaact ctaaaacgatt caggaaacttt tgaaaaattg gatgtttcag gtaaaacttt
32281 tagattttgt tctggacctt ccataattga aaatctatcg aatcctggaa ccaaatcaac
32341 taaatgaatg ttttccatttt gtttattcat tcttggctca gtggattttc caaaaacctg
32401 aagggggtga gcaactgtgac ctttcaggtc gaaataaagc tcacccccct caggcctttg
32461 gccctttggg cctcccccca agggggagggt ctttggctct tggttttcag aaaactctgt
32521 caaggaacgg agttcttcgc cagcaaccct tgcttttate gcgtcgcggg cgaaggattc
32581 attctcatat ttttctaaag attcccaatt tctgcgctt aaatgaaata agaaattatt
32641 tcgggaataa atttcatctg attttacagt taaaaattct tgtaaaattg aagccgttcc
32701 gaagccttct aaagcccaaa cttccatttc tccctaatct tgaccacctt ttttgaacg
32761 tcttccctaa ggttgttggg taacaagtga ataaggacca gtcgatcagag catgaatttt
32821 tttctcaact aagtgaatca atttaaacat ataggaataa ccaacaagag ctgggtgata
32881 gaatatttct cctgttgcgc catcaataaa atgcgttttc cctgggttat tagggctaaa
32941 taaccaatfa agttgtgaa ctttttfaat actcttcaaa gtgcgggggt accccccctt
33001 ttgaagccca agttcggaaa tttctgctct tttagggggt accccccgtt ttgttttaa
33061 tttctgtaaa tgccccataa ctaaaacttcg tgatgtttct ggtcagagaca agaaactt
33121 tgcattttct tgaagcgtat tttttcaat actttgactc ggagcaatac gttcatcaaa
33181 aggcataaatt cgtatacactt tgtgtgagta aaaagcagat aaaccaataa aagtttcaag
33241 aacttgacca acattcattc gtgatggcac ccctaatggg tttaaaacca catctcagag
33301 tttgccatct tgtaaatcgc gcatatcttg tggagggaga ataaagggaga catcccttt
33361 atttccgtgt cgcactgaca ttttatcacc taattgaatt cttttttgtg gactaaaaa
33421 aattgatact gaagtttctg ctgattgaat aaaacttccc cctatgagcg ttttgtgtt
33481 tcttaggaac ggagttcttc ctccgcggcc ctctcttccc tgcgttctg agggcctaaag
33541 gccctggagg gtttccctcg agcctcgcgg gggaaagggtt tgatttgcct tttctgtcc
33601 atcaaaaatg ctaatacgc cgtcaacaac atagcttaaa actcgcactt ttacaccaat
33661 accaaactta aaagatgtat ttttaatttg atgggcaaaa agtctcgaat ttgaactttt
33721 ttcagccgct aacaaattca aagtattctg acccgaagta agttcagaac caaagtttg
33781 ggaaggaaga gtattttgt aagctgaaag gtcagaaggt ggactctgtt cctccctttc
33841 agaacttagg ctctctctc caataattg tctttttta tctgtgtac ttggaagagt
33901 acttcccttt tcagacaatt tttctccgaa tgtttaggc caaagcttcc cctctctag
33961 atttgaagtt tgagttgaa gagttttgtc ggggttctta ttttgaagat caaaaatagc
34021 taataaaaat ctatactcag gaaatggaag tgtttgaatt tcttgagcaa aaggagtatt
34081 ttgtgctttt atgtcaaaat ttttaaaat atcttttgg ccaatttga gttttcaag
34141 cagaatatca ccttcttgaa cccatgttcc cggtttgata attccattag catcccaatg
34201 acgaagttta ggtggaagtt taggtggaag attgatcag tctaaagctt ctaagaatc
34261 tttctgtgag gttttcaaa tcgaattctc atcccagag tttcttgaag tttatttgt
34321 aacaatgtct agatcacgtg ttaaatatc tccaggaact aaggagcgaa aatattcaa
34381 aataactgtt tcgggaagag tatcttctaa ctgaaacgaa ttttccgggt tctgtactt
34441 aataattttt gtattttgta cgtcaacatc aaaacgttca acgtgtacag aggtataaat
34501 atcttcaaaa accagacgtt cattaataac aatggcatct tcaaaaattat agcttccca
34561 tggcatataa gctaaacaaga cgtttttccc taatgcaact tctccatctc gagtctctga
34621 accatctgcg agaagactac ctttttcaac ccattcacca atttcaacag ctggtcgttg
34681 ttttaaacaa gtactttgat tccaaggagt gaaatccgtt aattcataag atgaaatctc
34741 tttgtgaaat cctccccctt caggggagaa ctcaatgggg tgagcacttt gatcttcaag
34801 gtctaaagaa agcttaccct attgaggcct ttggcccttc gttccagggg gggaaacgaag
34861 ttcagagaac gtccttccat gttttctgt ttttagctta attgattgag aatcaacaaa
34921 aagaattttt ccactttcat cagctaaaaa taaagtactt gaatcgcgtg ctgctgcaa
34981 tttcagttct gttcccaaaa aagcttttct tgggtgcata aatggaactg cctggcgttg
35041 catgtttgat cccattaaag ctcggtttgc atcgtcatgt tctaaaaaag gaattaaaga
35101 ggttgcgata gaaatfaat gaattgggca gatacctgta aaaagttttg ttgtgcttg
35161 aataaattga agcgcactct tggaaatfaat agttaaaaa tcttgtttgt gacgaacggg
35221 aaaaaacagga ggaataaaag aagtgtttgc tacaattttc gaagtattca gtaaacacgg
35281 atttaaagta tcacctgtac atacataaat ttcatcttct agttctgaag aaaaataaga
35341 aatattttgt cgaactgcaa aagaatcatc ggccttcagg agagaacttt gttcaaggga
35401 gtaaagtttt tcaactccac cccattcttt ctttggattt gctttattca tttgaaact
35461 tgaagagacca atctctttgc ccggtccatt ttgtaaaata tatttactgt atttacgaga
35521 agctgaattg aatcctaacc aatcaccctc aaaattattg ttattgaagg gaaatctact
35581 tttttgattt tcaattggaa gagctggact ttgtaagaaa ccatcatcat taattcgggc
35641 ataacttctt aaagaaccaa ctaaccacagc attttgacct tctgggtgtt caattggaca
35701 aatctgacca aagtggcttg ggtgaatttc acgtacagca aaaccagctt gatctctgc
35761 tacacctctg ggtcctaatt aacttaaacy gcgtttatgt gtaatttctg ctaatggatt
35821 ggtttgatcc atatattggg aaagtggccc tgtatfaat aattctctta aagtaacttg
35881 aaaggctgga attccagggg gacagaatc ttcactgcgg gagcttgaga ctgaatcgaa
35941 atttttacga atcgaaatttt gattttgttc aacctttttt gcctcccccg aagggggaga
36001 ccttcccacg cgacctcaaa ggggaacccc ttgaggcctt tggaccttct cctttgctg
36061 caaaactctg tttcttatgt ttgttttatt gctttgcaaa cgaattgctg aatagggaat
36121 tcgacttaaa gctaaacgaa attgattttg cagaatttca cccggtaaac gtagcgtct
36181 attttttaag tgatcaatgt cgtctaattc aagactaaaa gtattttctt tggcgtcagc
36241 tatttttaat tcagtcfaat aatgaaatgc caaagaaca tcttctggtg ttaaagctgg

36301 gattttacac tcaggatttc tacgtaaacg tttattttaa cgaattcgac caacacgccc
36361 aatagaatat cgttttgggt ttaaaaattt gttgtaaaaa aatgaataac aacctaaagc
36421 tgaagtccaa cgtccagggt tacatttttt ccaaagtgct tgtaatgctg ttgctgtatt
36481 gattgtaagt ccggattttg acccagcatt tgtctctcta ggcgttcgag tattgtcatt
36541 ctcgacggta tttttaaaga caaagttttt tgttgaagag aatttttttc caatagagcc
36601 agaaaggact gggcgaggag attgaatgga agaaatttta ggctctgcaa gccttgccga
36661 caccgaactc caaaggtaat tttgttcttt cgagattttt ttgttttcta ctttattaag
36721 taactttgca aattgctttt cgcttctgtg gctttgtaga cctttactta atcctttgce
36781 aatccaatca aggaacggag ttctccccac gctaccctcc ccccatgtgc ggggaggcct
36841 ttggacctct ctttcatcgt gttgtcagat gggaggtta ttattctctt ggtatacttg
36901 aatataatftt gctgttaaaf atttttttaa gctagttaag ggagatgcag gaaaagtaga
36961 atttaaagtt gaagctgtct caaaaatgga attcttaaca atgtttttat tttttacaac
37021 tttattttct tttccaaaag atagagtctg cgcgtttaat ttaaatctat tttttggcaa
37081 tccatacttt acagaaatfa agttctcaga atctaaaaag tatcctaate ccattataaa
37141 aatatacagct ggaatctctg gggattttat aattcttgcc cataaacgat tacttaaacg
37201 atcagtttct aaacgaagcc acgatccata ttcagataaa aaacttgctg tgaactctcg
37261 tctatttttt tgatcaagtt ttactttaaa gtaaacactt gggcaacgaa ttatttgatt
37321 cactaaaact cgtggtgctc cattataaat aaaactgcca cgttctgtca taagaggaat
37381 atcaccaaaa taaaaccaaa aaatttgttt tttttgacga gttgatcttt caacaaaatg
37441 atttgtcatt aaattcttat tgttcttatc aaagacacca aaaatagaac ttgaaagaaga
37501 attatccact gacttcaatt gatttttctg gattgacaat tgaatattac tttttgcaa
37561 actttctaat gatcgttttt cggaaacggaa atcagagagt gaggcagtat ttctcacttg
37621 atttaaaaca gtcattgtct caaaacttat ttgcgtttga tttattttct caatcggtaa
37681 gaaaaaacca acgctatag ttttttgata acgaattgca tcaatttgag aaattttggt
37741 tgatttcaatg agaatcttag aacatgaag aactaaataa tcaagatcgt tttttttgga
37801 agagaatgga atggtaccag gaacaaagt ttgcaaatt ttgagttttc gattttctaa
37861 atcgttttat aagcttgaac ttccagaacc ttgtctttgt tctttactta gggaaaactt
37921 tgctctgag aattcaagga aagactcaag tttattgttt tcaactattgt tttgtgtttt
37981 aaaacttttc tgacgataag ccatcgatt ttctggaatt aagctactgc taaaacaaat
38041 gctttcttga ttaaaaaacc aatggaccga aggaccggag gtctcaaggg gaggggccaaa
38101 ggcctgaagg tgggtgagct tatttcgacc tgaaggtca aagggctcgc ctccttcccc
38161 cgaatggggg aagagctggg gggcccttcc agacctcgg tcctttggtc cttcaaaagt
38221 cccaacggaa cctccattta gcttgggagt gaagccttga ttcatagaac tttctccaga
38281 aaaagcttgt actgatgaag atttttcaag tttttctgat gaaatagtgg cagttctgatt
38341 tgtttttgaa gtatttttgt cttgttgata accaataaatt gttagaaaaa gttgtaattc
38401 atcaagaatt cctttttccc aaaaagatcg aacacttctga cgttggtttt ctaataaate
38461 tggaaaaaag ggataattat gaaatgagtc aatcattatt tttattactg taaaaaaaac
38521 taaggaacaa aattcttccc cgtcgaactc cgcctcgaaa gctctttgac ctgggtgat
38581 tgagctcttt gacctgggtg gattgagctc ttgacctgg ttgtctcag cgccttgacc
38641 tgggttgatt gageccttg acccttgagg accggaggtc ttggtgtcct gagectgaaa
38701 tctgacctga aaggtcaagc cgtcgaagca caccaggtca tatattaagc tcaatcacac
38761 cagacctacg gtcctttcag gtcataaata aagctgaagc acaccagacc tcccggctct
38821 ttgaggtcag atattaaact caatcacacc agacctcaaa ggggtgagcg ctttgacctt
38881 tcaggtcata aataaaaagt cacccttttg agacctttgg tcttacggtc ctttcaggtc
38941 atattttaaag ctgaagcaca ccagagctgg ggggcccctt caggccttcg gccctcaag
39001 ggggtgagcg tttgacctgg ttgtctcag cgttttgacc ttggttgatt gageccttg
39061 acctttgagg tcatatatta agctcaatca caccagacct acggctcttt caggtcataa
39121 ataaagctga agcacaccag acctcccggg cctttgaggt cataaataaa agctcaccct
39181 tttgagacct ttggtctctc cggctctttc aggtcatata taaagctcac cctctcccc
39241 cgaaggggga agagctgggg ggccccctca gaccttcggg cctcaagggg gtgagcgtt
39301 tgacctttca ggtcatacat aaagctcacc ccttcccccc gaagggggaa gacctttggt
39361 ccttgagacc tctggtcctt ttggtccttga gacctcgggt cctttggttc ttgaggcctt
39421 tggccctctg gtcctttcag gtcagataga aaacttaagc acaccaggtc aaagcgtctc
39481 atcacagaag ggttgtgaaa gaaggattta tttttataa attaaattta gttttttgg
39541 tctcatttta aagccttctc ttgatttttag tctgtattgg agaataagct tcagtaggta
39601 atctttctaa aatctttgat tttcaatttt ttttttaata tgcggttggc aaaatagcca
39661 caaaaataaa ctttttttta ttatttgaga gaataaatt ttgtttcatt cttttttctt
39721 ttttaacaaa tcaaacaaat ctatatgaga gttgaaatgt aataagaatt tgactgattt
39781 tgtaggaagc aacgaagttc tcgggggctt tgatattctg cctgaaaagga ccttagggcc
39841 aaaggtctcc cccctgaagg ggggaggacc aaaggaccgg agggctaaa gctctaaag
39901 gggggtctg agttttaaag ttactcagc actgcagat atttagaata aacaaaaggg
39961 gatccccctt ttgtttttcg catttttggg tagctggagc cctacaaata ttgacgtgga
40021 gcccgattat gtccaaact atgagccttg tttttgttta ttgtttggat ttgacaaat
40081 aatcattact tttctatcac gacgaattac acgacaatt ttcaaaaatt tggaaactga
40141 gcttctaat ttcatattta atttgaatag atttttgttt ttaaaattct tgaactcttg
40201 aaatctgta cggatatata aatataatga ctttaataaa aggagaattt ttaaaaatca
40261 taggctagct gcctgtttta tcaaccagat tttcattctc cgggtatttg atttggggat
40321 ttatttatga cttaagaagt tgaaggactc agtctttttg taggactaaa gggagaaagt
40381 tctcaccctt ttcagtcgaa taagttcttc taaactttct tctttgatc ctaaaaattg
40441 taaattttct tgccctgta cattataaaa aaaagcaata taaggagcaa ataagtatgt
40501 tatacagatc cggaaaaatt gtgggccccat cctattttct tcccgaagaa aatgaagtcg

40561 tgctcctttt caggcctgac acagttggag ataaatctct aaacttacga aaagaaatta
 40621 aaatggggta actttaatat aacatatgtt ttttaaaaat acaaatttga aaacaaaata
 40681 tttttaaaaa atttataagt ttaagttttg aagtatttta atgcatatag caggattcga
 40741 acctgcgatg tcgataaatt atcgaagagg cttatgaggg ctatgcggtc tgccactccg
 40801 ccatatatac gggctaaaaa ttaaattgtt atttaacaaa aaagaattct ttgcagctta
 40861 agtctgaatt gttgcaaaga gaacccgaaa gaaatTTTTT ggctgcactg ctgttttgcc
 40921 tcgctcgttc taggattgag gtcttggttg ctcgaaacgt tcgacctttg cggactaaaa
 40981 gaccacaggg ccaaaaggct gaaggggccc cccagctctg gtgtgattga gctttattha
 41041 tgacctgaaa ggaccgtgg gccaaaagcc tcccccttc ggggggaggt ctgggtgat
 41101 tgagcttaat atatatgacc tgaagggtca aagcgtcaa tcacagcagg tcaaaagcgt
 41161 ccatgcattt aggtctgaag gggcaccagc ctttgtgtgc ttaagcttga tttacgacct
 41221 gaaaggaccg taggaccaaa ggtcgtaaa caagcttaag acacatagat tttcaatcct
 41281 ctgatgctrc gactcatctg cctttaggtt gaaaaatgtc cttcctaatt attttttagc
 41341 ttgtctgttt aaaagtttaa tataactaac ttttgcctct ctttccagtt tcccgagaaa
 41401 attgtgaatc atttaatgat tcataaaaag ctttaaacac ggttttccact tccgctagaa
 41461 gcaacttggc caactaaacc aaattcttta gctttttttg ctgacataaa tcgatcacgg
 41521 tccataactt tggcaatttt gcttaaaact tgctcagttc tttcagcata aattctacca
 41581 atttggcgtc taattctcat tacttcttct gcttctctg taacatcaga tgcttgcct
 41641 tggcttccac cttgaggttg gtgaatcatt acacgagaat gtggtaaaagc aattctatgt
 41701 ccacgttcgc ctccagcgag aacaaatgaa gccatagatg ctgcaattcc aatacataga
 41761 gtagttaacg gagcttggat atgatgcata gcatcataaa cagctaagcc acaagtaact
 41821 gaaccacctg gggaaattaa ataaataaaa attcgccttg attcatcttc agcacttaga
 41881 taaatcataa taccaacgag ttggtttgct aactcatcat ccaaatcaga acctaaaaat
 41941 aatacgcgtt ctctatacaa tctattatat aagtcaatcc attgtggagt tggttctccc
 42001 ggtaaaacgaa aagcaacttt tggaacacca attggcataa aattcctcgt atgtttattt
 42061 agttaaaaaa ataaaatTTTT tactgactac tatctcctaa ataagtgtaa acaaatTTTT
 42121 gattcggaaa tacctccctt tggggagaaa atttctgtat ctggggggac gaagttcgtc
 42181 ggcctttggc ccttaaccct ttataggctt tgctgtcagg tttggcacac gtgagtagga
 42241 ttcttggctc taggttggaa gactttagtt cgttggagtc aagcaatatt caaatcattg
 42301 aataattggc tatggcttct gtgcttgagc gctttgacct ggtgtgattg agcgttttga
 42361 cctttggctc ttgctggcct tggccctacg gtcctcaaag gtcagaaata aagctcaatc
 42421 acaccagagc tggggggggc cttcaggcct ttggccctct ggtcctttgg tctccggctc
 42481 gttcgaggtc atataataaa ttaacctcaa acccatgaaa tgctccttcg gtccttctca
 42541 ttaaaaactt ctcatcccaa aactttgttc attgtaagct gggttttgtt aaaagaaaga
 42601 tttcaggcca tagtttgttg tgttttaatg gaaagatttc tttttttgtt tattttttaa
 42661 acgagatcga tgaatgagc tcttttaaac attagatttt tccactgtat caaaaaatTT
 42721 ggaccttagt ctatacttct agtgattatt tatcttgaag aaaaatttct ttttcttct
 42781 agcaactttt ctttgcgtct taacttagaa ttatgatatt tcagacgtta aagagtttac
 42841 tcttgagaag ggtatacttt atttgcaaaa aaatatttct gtggcactct ccttccaatt
 42901 ttaacttggc atttgatatt caccaattct catttatggg agcccagact tctctcaaaa
 42961 aagaattctt ttttgaatc acttgtatag gaacatagaa ttattatata gaggttaaat
 43021 aaagaaacaa aattgtatat atatttatgg atcttattta tatcttaaaa cataagtaaa
 43081 agaacagtaa tgaacaaaga gctgggtgtc ttgagcttta tttatgacct caaagcactt
 43141 agaagctcgg tccgcttgag ctttattttat gacctggtgt gcttgagcgc tttgacctt
 43201 aggtcctac atacagctca ccccttccc ccgaaggggg aagagctggg gggcccttc
 43261 agacctcgg tctcaaggg ggtgagcgtc ttgaccttcc aggtcataca tacagctcac
 43321 ccccttcccc cgaaggggga agaccttgg tccctgagac ctctggtcct ttggctcttg
 43381 aggcctttgg cctctcggtc ctcaaaggtc aaaacgctca agcagaccgg gtcaacgcatt
 43441 tcaagcacac cagggcgtgc gcattgaaat ttcaatgcgc acgccccttg gctttacagg
 43501 ttaactgat caccacgacg atattgaaga tacgcccgtta caaataatcc agcaatagtt
 43561 actggaataa atcctaaaac aatacctgat aaaagtgggt caacctggg atttgattat
 43621 gattatgaat agactttatc ttattgtaac tcgaagtgc acaaaacaag gagaagctga
 43681 ccaataaaat attctttatt tggctcagctt ctcaagtttc ttattattha cgagcgttg
 43741 ataagaaaat ttgtgtgat tctgcaatag cttcttttaa cagagtttca gcatctggtc
 43801 caaaagcttt tgttgtacca atcgtttcgt aaagttttgt tttattagtt gcaatgtatt
 43861 cgcgaagacc tactaagaag tcacgaactt gatctacagc taatttatca agataccctt
 43921 taatacctgt ataaatagta acaacttgtt cttcaacgcg tagaggagct gattgtgatt
 43981 gttttaaata ttcacgaagt cttacacctc gagctaattg gttttgagtt gcttgatcta
 44041 ggtctgacgc gaattgggag aatgcctcaa gttctgcaaa ttgagcaagt tctagtttta
 44101 aagtaccagc tacttgtttc attgctttaa tttgagcagc agaacctaca cgtgatagc
 44161 aaataccaac gtttaattgtt ggtctaatac ctgagttaaa taagtacgct gataagaaaa
 44221 tttgacctc gtttaattgaa attacgtttg tcggaatgta cgctgatagc taccctctt
 44281 gcttttcaat tactggaagt gctgtcctac ttccttccac taaagcatta ctttaattttg
 44341 ctgcaacgctc aagaagacga gagtgtaaat agaaaaacat gcctgggtaa gcttcaagc
 44401 ctgggtggagc ttttagcaga agtgacattt gtcggtaagc ttgtgcttgt tttgataaat
 44461 cgtcatagat tactaaaagta tgacgaccag tatacatgaa atattcagct aatgctgcac
 44521 ctgtgtaagg agctaagtat tgtaatgttg ctggcgcac tcgcttttca gctacaataa
 44581 tagtatattt catagcacca cgttcttcta aagttgttac aatttgtgca actgaaagatg
 44641 ctttttctcc gatagcaacg taaacacaaa ctacgttttc accgttttgg tttaaaattg
 44701 tatctgttgc tactgcagtt ttacctgttt ggcggtcacc aatgataagt tcacgttggc
 44761 cacgaccaat tggaaatcata gcatcaatag ctacaagacc agtttgaat ggttcatata

44821 cagaacggcg tgaataata ccaggagctg gtgattcaat taaacgagtc gcgctagctt
44881 taattgcccc ttaccatca attggacgcy ctaaagcatc aaccacacga cctaagaaaag
44941 cttcacctac tggaaattga gcaattttac ctgttgcttt tacagaatca ccttcttgaa
45001 tttttaaacc atcacccatt aatactgcac caacgttatac agtttcaagg tttaatgcaa
45061 tacctacagt tccgtcttgg aattcttaata attctccggc cattactttt tcaagaccgt
45121 aaatcagcagc aataccgtca cctacgttaa gtacagtacc aacgttgcta attttactt
45181 cttgattata actttcgatt tgttggcgaa tgatacttga aatttcatca ggtctaactt
45241 ttaccataga ttatagaga tttgaggtgt cttcttaaag aacaccata attatagatt
45301 tattttaata tactgaaata cttagtaacta aaataaggaa aattcctttg gtctgagaag
45361 tctgctccta atctttgggt ctccttctt aaaggggagg agtctcgggg aggagactt
45421 tgatccgtgg gggtttaac tttgtctatt aaggattcat ttttctaat aaagcaatat
45481 taaaatcaat aatcagttt tgtgaaacty ccgcttctgt ttttgtcgt aaaagcgtaa
45541 aagctcttt aattgaaagt tttacaactt gttgacgaac ttgttgaaa gctttgcctt
45601 tttcaaatgc tagtggtagc tgtggtccgg aaatagaagt ttctcaaga cgcgcaattt
45661 cttctctgt tctttcaaac aaagttttg tagcttgagc tgcattaaga tttcttctgt
45721 gacgaatttc aatagctttt gttttcggaa tctcaacttt ttgtttcgt tgttctaact
45781 tttgttgagc tccagatat cgttcttcag cgtttgaaat tctcgaaga atttttctgt
45841 tctctctct taataaggat gttagaacat caccaccaag atacagaact acgcaataa
45901 ctactgctaa gttataata ttagtttcaa taatatcaa attgaaacca aacgattcag
45961 cagccataga taaatttaaa actgacacaa gaattctcct ggggaatccc ctcttatac
46021 agttgctttt tacaccaact tttttgaaat actaaatgtt atcaaaattc tgtaggggct
46081 ttgaaaagc tattgtaata tttttgggt acccccaaaa aaataggctt attggttaat
46141 aacatactg ctcatagggt cgaagacttc ttgttaaaaa gtttatttt taacaagaag
46201 tcaaaaagcc tcccccaag agagagagag ctgaggggct ccttcagacc ttcggtcctc
46261 cccctctcgg ggggaaggatt ttagcgcttt ggactttcag ttcataaata aagctaacct
46321 ttgtagagc tggggggccc cttcagacct ccggtccttt gatcctcccc acgaaagggg
46381 gaggcctttg gccctgtggt atttttagct cataaattat acagaaagt tttttattc
46441 ctatgattca agacaagttt atcaattttt ttaatgaaaa aagatttgat ctttttttg
46501 gcgggaaaga tttcccactc ccttcgggga taataaagt ctagaaatcc cagaataggc
46561 ccagagaaca aatctaactg ctcttcgctg atttctaagc ttatccatct acgaatggga
46621 gaaaagagaa aaacaagggc gtaaaaaatg tcaaaaaaaa agggagggg tattaatcta
46681 aagtttttaa gaaactttag attaataccc acttcccagg aattaagaac taaaaatcaa
46741 aatattacgt ttagaacgag tttcaaaaaga ttaaaaactc gaagaatgag atttcttatt
46801 ttgatgtct caaggcaaac ttagtctctt tttttaacta ttagccagca aatggttctg
46861 cgaaaagtag agctaagtct acaactagac cgtaaatagt taaagcttcc atgaaagcta
46921 atgaaagtaa taatgtacca cgaattttac cttcagcttc tggttgacgt gcaatacctt
46981 ctactgcttg accagcagca gtaccttgac caataccagg accaattgca cctagaccaa
47041 ttgataaac agcacctaca actgagagctg caccaattaa tgggttcag ttgggtctct
47101 attatgaaata ataataaagt tttggttaat tttgaccaa aagtaaattg agattaagaa
47161 agtttatctt attgttggta tgaagtttga aagcttgata tacgtatgtg ctaaaaacct
47221 ttgacctgggt gtgcttgagc gctttgacct ggtgtgcttc agccatttga cctgggtgga
47281 ttgagcgttt tgacctttt ggtcatatat taagctcaat cacaccagac ctacgtctct
47341 ttcaggtcat aaataaagct gaagcacacc agacctcaa ggggtgagcg ctttgacctt
47401 tcaggtcata aataaagct caccctttg agacctttg tctccccggt ccttttaggt
47461 cataaataaa agctcacctc tttgagacct ttggtctctc cggctctttc aggtcataca
47521 taaagctcta gcacaaaaga cttccccccg aaggggggga tcttcgactc tctggctctt
47581 tgaatgagaa aagtttataa tcttcaataa aaaagctgat tttaaaaata aattgaaata
47641 gtaatgttt gaatccaagt ttagttttgg aaaaaataaa gctttatttt ctgtatttat
47701 tttattattt agttatctgc ttaataacag tataccagaa atattatttt tcttatcgg
47761 atttcaatc ttaaaaattt aaagaagttt ttcgttctgc aagaactttt ttaactgtc
47821 atttctctag aattaaaaga atggttaaat tcaagaagat cagggtttag agattatctt
47881 gagagccgac tttctaaacc ccagctctgt taatcttag aacaaagttc tccccctcg
47941 ggggaggatt gttttttat ttttttaatg atgatcttca agagcttcac caatataagc
48001 accagctaaa gttgcaaaaa ttaatgcctg aattgcactt gtgaataaac ctaagaaaat
48061 taatggaata ggaatcacia gaggtactaa agtaataagt acagctacta ctaattcatc
48121 agctagaata ttaccgaata aacggaaact taaagataat ggttttgtaa atcttcaag
48181 aacgttaata gggaaataaaa ttggtgttgg ttttacataa cgacaaaaat aatttaaac
48241 ttttttggta aaaccagcat aaaaatagc tactgatgtt aatagtcaa gtgctactgt
48301 tgtatttaata tcatgttgg ggcacactaa ttcaccattt ggaatttcaa aattcggcca
48361 tggaaactaaa gcacctgacc agttggaac aaaaatgaat aaaaaaatg ttgaaaggta
48421 tggaaaccaa cgttttgat cagctccatg gatttgcgta gatgctaaat cacgcataa
48481 tccataaaa aattcaataa agttttgtcc accacttaa tttgcaactt gaaagttaa
48541 ttggcgagtt gttgtaaaac ctacaattaa aataattgca aatacaatcc aagtattaac
48601 taaaacttgt ccgtgaactt gaaatccacc aactgaccaa taccaatgtt ggccaatttc
48661 tacagcggcg atgtcatata acataggata aattcaataa aacgaaaaag ttacataact
48721 tttcttatgg gtgatgcttc gaggtaaaagc acaaaaattg attttcccc aaacatagc
48781 tccccctcgg ggggagaact ttgttctga aaggctccct caaggaccac agggccgaag
48841 gggggagacc ttcggtcctc aaggggtgga gcgcttgac ctggtgtgct tcagcgtttt
48901 gacctttgag gtcagatatt aagctcaatc acaccagacc tctcgtctct ttcaggtcat
48961 acataaagct caccctctc cccggaagg ggaagacctt tggctcttga gacctctggt
49021 cctttggctc ttgagacctc tggctcttgg gctcttgagg cctttggccc tcagaacttc

49081 gttcctcccc aaagactacg gttctcagaa actcaaaactt tcaactgtaa ggaccaaaagc
49141 tcccaccgga gagaatttcc tccgtttgaa ggggtatcgg ggaaggaag ggagactaca
49201 aagtgaggaa agaaccgaag gaacggagtc ctctgtggct tgagcttaat atatgcccctg
49261 aaagcaccga gaagaccacc gatctcaagg gtccccccac ctctgtgtg ctgagcttt
49321 tatttatgac ctgaaagatc aaagggcttg agcacacaag gagtgtgaga agaagaagtt
49381 cttaaatgat ctgacaaaata aagatatttt ttttaaaaaa ataggggctt gatttacaca
49441 gactatcaaaa aaactaaatt gttgctctac aattatcgggt tttgacctaa aacttgcata
49501 cgagcacgag cacgtttaa agctaaactt gcttcaattt tttctttct tccagcattt
49561 tgagcttttt caagagcgct ttgcttctt gtaaaagctg cttcagcttc agttggatta
49621 attgatgaac ctaattctgc ttcgcttact aaaatagtta cagtattttc ttgataaga
49681 gcaaaaaccac ccattaaagc cactgccaagt ggggctgtgt ttttatcagg acgaaccatc
49741 ataacaccga tatctaaacc agtaataaga ggagtatggt tagttaatac tcccatttga
49801 cctgtgtttg taggaagat aatttctct gctgaagcgt tccaaaaaac gcgatcaggg
49861 gccataatc acacttttaa agtcataatt ttaaatga aattttata cttaatagtg
49921 gaagcttaaa aagaaagcac ccaaaatcgg tatgtgaac gttttgacct ggtatgaagg
49981 accggagaa cagaggcccg gaggtctgaa ggggccccca gctctcaagg cggtagacc
50041 tcccccaaaa gaggggaggg tcttttaagg gctgcacgt tgctcaggaa tccctaaagg
50101 ccaaaggacc ggaggtctga agggcccccc cagctcttcc acctcagagg gaaggggggtg
50161 agctttattt cgacctttca ggtcgaataa aagctcacc ccttgaggcc tttggcccc
50221 acaattcttt ttttctttt acctattaag aagctaaatt gtttgccttt tctacggctt
50281 cttcgatgtt tctactaag tagaaagctt gtcccgttaa tgaatctaatt tccacttca
50341 gaattagttt aaaaccttga attgcttcag ataagcttac gtattttctt ggagaaccag
50401 tgaanaacttc agcaacgaag aatggctgtg ataagaacg tccactttt ctgacagtg
50461 ctactgttaa acgggtcatc tctgaaagt catcaagacc taaaatggca atgatattt
50521 gaagttcttt ataagcttga agagtacgtt ttacagattg tgcgtctttg tgctgatcat
50581 caccaacaa ccatggttgt aacattgttg atgttgaatc taatggatcc actgctggat
50641 aaataccttt tgcagctaaa tttcgagaaa gtactgttgt ggcatcaagg tgagcaaaag
50701 ttgttgcagg agctgggtca gttaaagtc cagcaggtac ataaacagct tgaatagagg
50761 taattgaacc atcttttgtt gactgtaaac gctcttgaag tgaacccttt tcaactta
50821 aagttggttg gtaacctacc gcagatggca tacgaccaag tagggctgat acctcagaac
50881 cagcttgtac gaaacggaaa atgttatcga tgaataataa tacgtcttgg ttatttcat
50941 ctctaaagta ttcagccatt gttaaagctg tttagaccac acgcatacga gcaccagtg
51001 ctctattcat ttgaccgtat actaaggcta ctttagattc actcatttta tctgtattga
51061 taacttttga tttttctatt tccatgtata agtcttccc ttcacgtgta cgttctcta
51121 caccaccaaa cactgaaaca ccaccgtgtg cttttgcaat gttgttaata agttccataa
51181 ttaaaaactgt ttttccaaca ccagcaccac caaataaacc aattttaccg ccacgacggt
51241 atggcgctaa taagtcaaca actttaatac ctgtttcaaa aatagaaagt ttagtattca
51301 aatcaacaaa agctggagct gttctgtgaa ttggaagcgt tttcactttt ccagtactta
51361 ctgggcctaa ttcacttact ggttcacctt atacgttgaa aatacgcctt aaagtactt
51421 caccaactgg aacacttaaa ggtgcaactg tgtcaataac tttcatacca cgcatactac
51481 catctgttag acttaagca accgcaactg ctttattgtc acctaaaagt ccttgaactt
51541 cagcagtaat ttttaaatca tcaactgctt ctttttccc ttaacaactt aaagcgttat
51601 agatgttttg aatatttctt gatgagaaaa ctacgtccat aactggtcca ataatttgtg
51661 taatgcccct ttcgttttta acttgtactg tcatatttaa actttatgag aatagaattg
51721 aaagcaattt cttaaaaga ctttccaag atttttttg tggaaagtaaa accaaaacat
51781 ggtttcttac tgtctaaaaa gcattcttga tatttggaaa agaacaaga ctcagtttga
51841 aattttttt ttatattaaa attgacaata tttcaaatat cataagcttt gtgatttcca
51901 ttgagcttgt ccaacggaca aaaggaccag aggaccggag ggccaaaggc cctcaaggcc
51961 aaagggaccag aggtctcaag gaccaaaagt cttccccctt cggggggaagg gttgtagctt
52021 tatgtatgac ctgaaaggtc aaagcgtc caaccttgag gaccgaaggt ctgaaagggc
52081 cccccagctc tcccccttc gggggaaggg ggtgagcttt tatttataaa ggaccgtagg
52141 tctggtgtga ttgagcttaa tataagacct gaaaggtcaa agctctcaat cacagcaggt
52201 caaagcactc atccccctga ggaacaatag catcagcact ttaagtcact gctttcaag
52261 ttttaataat taattcactt tctgtaatgc aaaactctta gtttttaac aatcगतaga
52321 ttgtgggaag aaaccacttt ttcattcttc aaatgtcatt tatcaaaact tattagtcaa
52381 gatttctgaa tatttttctt tgtatctgtt ttaacgaatc tttctttgaa caaataggat
52441 tgacacattt taaccttaag accaaggggc ggaggggaagg agagaaaaaac tttgaaata
52501 atataaaagc ttaaatagc tctcctttaa gaccttatag tcaaaaccgc ccttagaaa
52561 aagaacacag gttttcaata ttataatatt aagctctata attctttttt gataattcac
52621 ttgctactac gaaatttgggt ctttttcaat gtagacaaaa agacttgcca tagcaattgc
52681 cggaaaaaca agtccaacaa gaggaactaa aattgatgga aggtataaac tcagcatagt
52741 tttattcaaa atcttttagat ttagggtttc tgcttgggtc atttgataag aaacaaaaa
52801 cgaaaagtga tcaaaagaaa ataactactg tataaacgaa aatctttaat gttaacatag
52861 taaaagggt caaggttca cggggggccc cttttaaata agaaattttt ttattgaaag
52921 agaaaagcact caaacaggtt ggagggaaa gattaccaat actattggta agactttttt
52981 cactaggttt taaaattttt ttttttttcc aaacaggaac caaagacctg agggccaaag
53041 gaccaaaagat ctcggttagct tgagcttgat atctgacctg tcaaataggt tgggtctctt
53101 tccggcgctc gagctttaa cttagcgaag cacctcttcc ctcgaacctt gattaaataa
53161 gcttctctaa atttttgaat tctttgtcc aagtaggaaa agttataaga gtttttagtt
53221 ctatttttaa ctttccatgc caaacctgaa ggggggtgag accttgacct ttcaggtcga
53281 aataaagctc acccccttca ggcctttggc ccttcttctg tgacttttcc tagcactttt

53341 ggcaatgaaa ttaagactta gtgctatfff tcttatttac gcggttaaat aatgggggcc
53401 tgaatttcag accctcttat tcttaanaat ccaaatgagt tacgaccgaa gattttttgg
53461 tcttttagga tcgagaggac aagcttctca gaatttaatt ttttaaaagt tgaattttt
53521 ccctaaaaaa attatttcct tagaaaacca aaacaagaag tcgaaaagac gtacgggtga
53581 aatcccttac aaacaaaaa aaaattaaaa aattgttttt tagaccttt ttttgggtcc
53641 ccctcacaga gccgtttttt gattctcaag aagtggaaaa acrttctttt tctgtttgga
53701 cagtgaatta caaatataaa cttttaagaa tcttaaatat tcctcaagtt tcaatttaa
53761 ctttgattat ccgcagattt cttgttatcc tatggcttct ggaaaaacaa ggtttggaa
53821 atcttctttc ttcttattgt acactagatt cctttccaag ggtgaaaatt ctttttcaa
53881 aatctctgaa ttttcaagta aagtaaaaaa taaagtctta ttctagaaca gttatcacct
53941 gtttttattt aaattttaaa agcggaaaga agcaaatfff tagcaaaatc tttatttgat
54001 cagcaataaa agattcataa aagaacaaat accccagcac aaatttctca ctcgatacct
54061 gtttggagga acaaaagtca tcgaacttcg tccctccaa acaacaaaa cctttggaaa
54121 cccctattga tgaattgagc gccctgacct agtgtgcttg agcccttga ccattggagg
54181 ccggggaac caaatgaacg gagacctcgg tggcttgttt atgacctgaa aggaccggag
54241 ggccaaaggt ctccccaaaa aatgggggga gggccaaagg cctcgaaggg tccccctcga
54301 gggccaaagg cctccccgca aatgggggga ggggtggcgt ggaagaactc cgttccctcc
54361 ccgaaggagg agcctttggc cctttgggtc tccccgcaa taaagttaac tcatagagag
54421 tgcctccaat ggctcaccga ggtcattggt aacgtggaat acgaatagct tcacctaat
54481 tttttaagat taattctaca gactttgtag aatcatcatt ggcaagaaata aataaatcag
54541 ctaaagatgg atcacaattt gtatctaaaa ttgtgatagt tcttggaaaga gcaattcccc
54601 ctttagtcga atttgatca cttcgccaac ttgaagctaa tttttgacac cttttaactg
54661 cattcatttc acaatcttgg ccaacaataa taacaatate aggaatatca ctcatatttt
54721 caatccccga aaaaaatttt ttttaatttt catatgtgtt ttttagaaga gctttttctt
54781 ttttagtttg taccaggctc gtttttgttg agttattcac tgaagaaatc tctgtatttt
54841 taaaaagacc acttgcacgt aaggtttaa gtttagcaat gcaaaagctta atgttgacc
54901 aartcgttaa taaaccaccg agccaacgat gatttacgta atttgcatga gaattttcag
54961 cttctgcttg aaCaatagaa gctactgttt ttttagttcc aacaaataag actttttgtc
55021 caataccttc tccaaaagt ttgggattag aagggtgaat agaattttta agagctcctt
55081 taaccgataa tttgctaaa aataagtaag attctcttag ataccctaaa ttttggataa
55141 tatcgataat atgaacccca tttcgctccc caaaaatata tgggtccatt ttgggattcc
55201 atttatttac acgatgcccg aaatgcacgc sagcttctaa caattcttca attggaatca
55261 taaaatatgc tataagaatc ttttttcgaa taaaagcatt tgcatttttt gctctaaaa
55321 tatcttccct cgtaccctgt gtgcttaagt aatctgatct atacccttatt ctggtcagct
55381 ttaagcacac catagtcag tttctccctt tagaccatg gtcctgttgt ccttgggttt
55441 tcataacttg aagaaagaaa ctatcacgag agaagaactt tgttgtctgg aaagcttttt
55501 agggccaaag gcctcaagga ccaaaggacc agaggtctca aggaccaaag gaccagaggt
55561 ctcaaggacc aaaggctctc ccccttcggg ggaagggggt gacgttttat tatgacctga
55621 aagggtcaaa egctcaccct cttgaggtct tccctcttcg gggaaagggg gtgagcttta
55681 ttttagacct gaaaggaccg agaggtctgg tgtgcttcag ctttatttat gacctgaaa
55741 gaccgtaggt ctgggtgat tgagcttaat atctgacctc aaaggtcaaa gcgctcaatc
55801 acaccagctc aaagcgttga agcacacgag gtcaaaagcgc tcacccctt taggaccgaa
55861 ggtctccccc cattcggggg gaggtcttcc ccttcggggg gaagagctgg ggggcccttc
55921 agatcttcgg tctccccccc gaagggggga ggctctcggc cctgtgggtcc ttgaggattg
55981 ctctattcga aaaataaaaa ggtaattttt cttgggaaaa tccaattttg agttataaaa
56041 taattaaaat cttcagctta ggaccaaaaga accgaagtgt agagagattt cttttgact
56101 caaaattaaa atagaaatta tgtttctttt caaaaatata ataagttgtg ttctatctcc
56161 caaattagag aattagattt agttttgaga acaaaaaaac aaagttgact aaaccttgtt
56221 ccaaatcgta agatgtaaaa cttctggaaa ttcattttct tttgaaaaga aacacaaagc
56281 cccctgtcgg aatcgaaacc acgacctcag cattacaat acgatgctct agcccctgag
56341 ctaaggaggc aataaaaaaga gatttaataa tattatgtaa ttaatactac aaattcaaaa
56401 aaattttaag tacaaaactc tctttttaa gaatttaaat gacaaatcta ttttaagagt
56461 tgtctgggac ggaaggattc gaacctccga atgggtgggc caaaaaccac tgccttacca
56521 cttggccacg tcccatattg tcttcaagta taacaaaacc aaccggtaaa ctgagttcta
56581 aaaaaaagga actatacaaa agccaaagag aatttactcg gaaacctcaa aggagctttg
56641 gtccagagat cctccccata agcgagaaag ctagaactga agaataataa atgaaaaga
56701 ttttaataacg aaactgaaaa acaaagactt tgctatcact cgttttggta ttatcagagt
56761 actacttttc ctcaagctc cagtgttaga gcgttcagct gttaatcagag tggtcatagg
56821 ttcgaatcct atctggggag ataagaaaa cttttagactt cttttgttaa aaagatataa
56881 agtataacca tegttaaaaa caCaatttgc tctgaacat ctggaagtct taaaacatta
56941 aaccagtttt ctggatcttt tcaacaaacc ccaaaatctt taagaggacc acagggccaa
57001 aggcctgaag gggcccccaag tctcaaggg gcttcagct ctcgtgtgct tgaacttatt
57061 tatcacttga aaggaccgga gcggaggacc aaaggaccgg gggctcctgat ggcttgagct
57121 tgatataatga cctgaaagga ccgaagagcc aaaggtctga agggggcccc cagctctcaa
57181 aggtgtgagc ttttatttat gacctgaaag gaccgcaggg ccaaaaggcct gaagggggcc
57241 ccctcctctg gtgtgcttga ccttgaatc tgacctcaaa ggaccgggag gctcgggtgtg
57301 cttcagcttt atttatgacc tgaagggtca aagcgtgaa gcacaccagg tcaaaagcgt
57361 caatcagacc aggtcaaggg ctcaagcata ccaggtcaaa acgctcaatc agaccaggtc
57421 aagggctcaa gcataccagg tcaaaacgct caagccacca agggggaaga agattatctt
57481 cccaagaccg cgaagaaaag acaaggggtt atgggtctaa agttaagttt tataaaataa
57541 ctttcaatgt attaactgat ttcagccctt ttgaaagttg tttctctgtg aatattcatt

57601 ttttccaaac catgaagtaa accttcaaac gaatgtaaag aatctttctt tatttttggtt
57661 tgttgcggtt tttcttgagc tgaattgca aaaagtaaag attcattaac cattctcgct
57721 aaatgagctt gactaaacc atctgttgaa taagcaaagt agtttaaacc taaggcagac
57781 cctaaaaaaa tatctacttt ggaatccct tcccccgcca ccctcccccc acctttggcc
57841 ctcacagcag gaaaaaagag agggccaaag gcctcgaagg gttccccctc agggccaaag
57901 gcctcccccc aattgggggg agggctcgct gggagaagact ccgttccctc aaaatTTTT
57961 acttgttgaa taagcttatt ggattttgaa cctgcaacta ttttaggttc ttgtaataaa
58021 gaattagata aaactgcatt tacagaagat gtttttggta tgtaaaattt taaaatatca
58081 aaacgaattt ttttagaagg ataacttaaa gccaaaattc tatcaaaaacg tcctgagcgg
58141 acaaaaagctg agtcaagctt agaagatag ttcgttgctc ctaataatac agcttgtgaa
58201 tttattttaa aatctttccc ttttccgtaa ccccatgaat ccatttgaac taaaaactca
58261 gtaaaaaagt tactatcatg ctgagggtgt tgcctttgtg caattgcate aatttcatca
58321 aaaaaataaaa tgcattgagt taagctcgc gcttgttcaa ataatttttcg taaacgtaaa
58381 gctccaattc ttgtcccaat ttcaatttgt ttttgaattt cacttgcgga aacacaaaata
58441 aaaggaaact ttgctatttg tgcaattctc tgtgctaaaa gaggtttggc tgttccaggt
58501 ggtccaacaa ataaaaaac tttggagca aaagaagtgt ttttcagttg cttcgtattt
58561 tcaaaaactt caatttaaga ttgaataaga ggagaaggtt cttgaatacc aggtaaacata
58621 ggaagtttat ttgtatttcc tgatttgaac ggcaaaaaag aaaatttgtt ttttggccat
58681 aaaattctag cttgatttgt tttttgatct ttttgcgta acgggttatt tcgtccaaaa
58741 ttatttgcag aaataatttt ttgaacaaca gaaaaaaca aaagttgtgt aaaaacatga
58801 gttaatcgaa gcaaatgttt atttgaatca agatttaatt cactcaattt taactaatat
58861 ttgaaagaac aaaaatttgt cccaacagtt tgtatgtatt gcacctctgt tctttcatc
58921 tggaaaggcaa gagagtccac agaattttgg tgcctatttt cgattacaat tggtaatag
58981 gttgaaagac gattagataa tttgttaga aagaataagc ttgaagactc tcgagaatct
59041 tcccccggaag ggggagacct tccccgcga cctcaagggg ggtgagcact ttgaccttcc
59101 aggtcagcaa tatgtccacc ccttgaggc ctttggccct cacaatctcg cgtgggaaga
59161 acttctgtcc ttcggtcctt taacaatttt ttgccaaagca ctaaagggcc aggaagattt
59221 ttggtaagta aaaggcgttg ttcattgggt aaaaatcttg ttgctgaacc tcttaacca
59281 ataaaaaaa cataaaaaag attaagaac gaaaatgtg gaaactttga ttgggatccc
59341 tgtaaacctt aaaaactaag aagaaaatga gatgaacttg caaaattagg gttaaagac
59401 cctcttccag aagtattcaa ggcgtgaaat cttttccctg aaagactatc ttctctagag
59461 aaaagggttt ttatctttac taacaagaat ttttttttag aataagattc ttttgttaa
59521 gttttttgag caagtgtgta actaaaaatt gaaccggaag aattagaaac aaaaagagaa
59581 aataataatg ttaaaaattt tcttgatttt gcaactaatt gaacttccat tggatttaa
59641 gaaactccca aaagtgaaac tttagtttca aaatcttttg caaataaagt ttgcgatttg
59701 ttaattattt ggatttttag tcttttttga ttagtaacct cgcagttaag agtatcagta
59761 actaaactcag gcaacgtatg caattcgtcg aaagaattta attttttgtg gtcaacatga
59821 atgaaaaaag atgatgataa ggataaatca gttaatttgt ttaaattttt agtcacaaaa
59881 taatttgtgg tagaaggatc aaagggccaa aggaccagag gtctcaagga ccaaggtctt
59941 tcccccttcg ggggaagggg gtgagcttca tttatgacct gaaaggaccg ggaggtctgg
60001 gcttcttcag ctttatttga gacctgaaag gacctgaggt ctgggtgtgt ttgggatccc
60061 gacctttgag gtcaaaagcgc tgaagcacac caggtcaaaag cgtctacccc cttgaggacc
60121 gaaggtctcc cccscattcgg ggggaggtct cccccattt ttttggggag gcctttggcc
60181 ctcaaaaggg tgagctttat ttgacctgg tgtgcttgag cttaataact aatatatgac
60241 ctgaaaagggc agtaggtttg gtgtgcttga gctttctgta taacctgaaa aacctgaaag
60301 gaccggaggg ccaaaagacct caagcacacc aggtcaaaag gctcaaacac accaggtcaa
60361 agcgtcaag cataccaggt caaatccctt tagtacagaa agctctcaaa cacacaaggg
60421 ttacagggaa aaattttgag tttggaggaa tgatgacttc gttacttggg agaaagaagt
60481 ttaactgtct ttttcccttg tcttccctc tcttttgatt ttcgatggc tttttcttc
60541 aatatcaaaa agttaactcc ctactttaaa agcatctaca aaaataccaa ataaaaaac
60601 agttttaaat acattgagga atgaactcaa ttgaaacata gaattttttt ttgttttaag
60661 agaactctgc ggaatatttt tgttaaaatc aaagttttca attttaaaat tcttgaaaac
60721 tttaaaaaaa ggtaccaact cttttgaaat aagtgaagga gggaaaaaccg aagtatttaa
60781 ctttttttgt ttaagttgtt ccaatgaaaa aaaacaatta gtacattcaa ggaagaagaa
60841 agttataaaa agagttaaag aattaaattc atttaaaaaa ttttaaaaaa catttctgct
60901 tggaaaggat aaattcgaac gttcagttaa atttaagaa gaatctctt cttggatttg
60961 gagtgtttggc agatttaaat ccaagatgtg aatccctaaa ggagaatttt gaaagttag
61021 agctatacaa agcgttttgg taatagaaat caaaatttca taaaaactta taaaagtttg
61081 aatataaatca cttttgacaa cccccccatt cccaataaaa aaaccaaaaa agaaaagaaa
61141 taagattgta ataaaaagag taggttgaga gaacataaga aaagcaaat tttgtgaatg
61201 gcggattagc agttactgct cagccgacct ttcgatctta aaatctaaga ctttgcaaaa
61261 agtgctccat agacctgtaa ttcatcatgc aaaaatctc ggatacctaa tcttttctc
61321 tgattaaagag tgcagatttg aatatacga aatctcggtt ttccttttgt attattttct
61381 aaaggacgtg ctacgaaacg tgtcgaatta tcaggtgaaa caaaaaaac taaagaacgg
61441 ccttgagatt ttaactgagag aagaaaaatc gtaatgtccg cttattttcc taacgaaaaa
61501 gaagaagtat tagatactga acgatatttt aaaagtgtat tttcaatttt tctttcaaaa
61561 gtatttaaga aaaaatgattg aaattttaaa gaagcatcag ctttttgttg ttcactcagt
61621 ggtggacgct cacttaaaac agggcgtgaa ttcattttat ccacaggtaa atgtcagatc
61681 aagggtaaat ttttgcacg ctgttgaaata tttgaatcaa aacttggaaa ttcacgaacc
61741 catgcacttt gaccttctcg atccacaaaa gaagcaggat attgtgtgaa taatcgccaa
61801 ctatattgaa acccaccact catcaaaaaa ccaactaaaa gaaaacttat caaaattctt

61861 gcctcaggaa caggttgaag agctcgacct ggacccaaaag actcagctgg aatatttaca
61921 aaacgtaaaa gtttttcggg aggcctcctc tttgtaacaa ataaagtaat tattaaaggg
61981 cacaaaaata agacaaaacc taaaaatgtg acacccaaaa ctaaaatata agggattggt
62041 aaaaaagacg aattttcaaa aaaacttccg gtatgtaata agattcgact tgaagttgaa
62101 cccattgggt gatttaaaaa cattaaaaca aatgaagcta aaaaaggcgt taaactttt
62161 tctgctaaag tatttgagtt agatcgattt gttcttggtc cactcataaa gctcgaagct
62221 aattttaaag cgaagataaa accaataaaa gctaaagaaag gttcaaaaag ataccaaaaca
62281 tgaattaagc tacgacttgc accagttccg ttcacaataa aaaaagaaaa acaaaattgt
62341 cctaaaactg tcccaaaact agcaaataaa ccaggaaaaa aaccttcgag taaaaaacgt
62401 gaatacaaaa ttaaaatagg aacagaaaaa ggaatactaa aaactaaacc atgataaaaag
62461 ccttcaaaag cacttaaagc ttgtgttcc ataacattat tctcctcag aaaatttatt
62521 ttaaaacctg tagattaatg agcctcactt tttatctaga gcctatgatt tttggactt
62581 gaggttctat agaggagaa gcttcccaa ttagctcagt ggcttgagct tgaactctt
62641 gaaagaaccg gagggccaaa ggcctcccc ttcgggggag gaccggagtt tttcccacgc
62701 gaccctcccc ccgaaggggg gaggccttcg gccctcaagg accaaaggct ataaataaag
62761 ctgaagcaca ccagacctcc cggctccttc aggtcataaa taaaagctca cccctttgag
62821 acccttggtc cccctgctc tttcaggtcg aaataaagct caccctctc aggcctttg
62881 cctaaatgat actttcaaat ttgaaaatca aatataaaat aaactgaagc agctttctt
62941 ttaaaagttc tcaagtttat caattcttca aggttcaaaa tttttaaata taccttttt
63001 ttaaaaaaaa acggtgttct tccattgaag aaggagaaca aggttctaata taaactcgt
63061 cccccagca tcaaaagta taagtccaa aagggttaac ctaatgggtc ctaagagagc
63121 ctattgaaag taggcctaat agtagaatta ggctcgtacc cttatagtat aactctctaa
63181 aaaaaagcag tctaataata actgtgttgg caaaataaca aaatataaat cagatacaaa
63241 atggaattct ccccccaaaa aaaaattgag agctgaaaac aatgaagttc tttgaactc
63301 attgttgagg gggaccaacg ttcacagtc tttgagttg agcttgattt ctagactgaa
63361 aggagcgtgg gaccagagga cctgagatct cggtagcttg agcttgaaat ctgaccttc
63421 gatcgatcac accagggtcaa agcgtcctaa cacaccaggt cctcgcacaa gaattctct
63481 ctccccatg tgcagaaaaac atagtataat aattcataag cagaagatat gaagatcgt
63541 ctttttctta cgatcttatt tttattgat ctcatctcaa ataaaaaat gaataagaaa
63601 gggttcaagc tttcaactta tttaaatgtt tcatctttaa aaatgaaaag tttttaact
63661 gaagtttaag ctttgtacta tagataaata aacacttctt aattttaata ataaaaactt
63721 ttaagaacaa cgaactgtct gaattggaac caatgagtat ttaccggctc aaacatagtc
63781 acctctgaca aacgaactaa agagctcaaa gggtaaaac aaaaatccaa atctatagga
63841 caacacttaa tcttaagttc taaaactttt caagattact caaccagaat agaaatttt
63901 agagtcagct ctttcacaat gataaagctg caattttata ctaatgaatc ttgcacattt
63961 aagcactaaa gtattttaag gttctcgtta tactttaact agactaaagt agaatcttt
64021 atgtgattag caaactgaat aaaaagattca tattctttaa tttatgtatt atctatagga
64081 ctttgggtg gaggactaaa aaacggaagc tttcgtatga tagagtgcct tagcccggtg
64141 tgattgagcg ctttgacctt tgaggtcata cacaagctc acccctctc cccgaagggg
64201 gaagagttgg ggggccccct cagaccttcg gtcctcaagg gggtagcgc tttgacctt
64261 caggtcataa ataaagctca ccccttgag acctctgggt cttgggtct tagggcctt
64321 ggccctctgg tctttgagg tcatatatca agctctacga cagcagggtt ttacccttt
64381 aaactaaggt ctctaaagg tagactagaa cattgttctt ttgtaacct tagaccttc
64441 gccctgataa aaaagcgttt ttgctttta acttagtcaa taaacaaat ttattagatt
64501 tctttctcg gatcattaac tgagtaata cctttttta gctttcaaa ctctaaaaaa
64561 aagaacactt tttgctatga tttgaaagc ccaacaacc gttaatggga acgaaaaag
64621 aataaaaacc gaatctttaa tataaaaatt tcttgggtg cttgagcgtc ttgaccttc
64681 aggtcatata ttaagctcaa tcacaccaga gctggggggg ggggcttca ggcctttggc
64741 cctacgggtc tttgaggtca taacaaaagc tcacctcct ccccggaag ggaagacct
64801 ttggtcctg agagctgggg ccccttgagg cctttggccc tctgtctctc cgtacctttg
64861 aggtgagaga ttaacttaa acccctgtt tgattcttaa tctttttaa gcaagacat
64921 gctttgtcta acttataaga tattatgaat caaagaaagt atttttatc agtcggatt
64981 ttcacttttt aaaaatttaa agcttgtac taacaagctt tctgtgtat tcttattca
65041 cagaaacagg agagaaaaat gagcgggaact acgggagaa gtcctttctc agatattatt
65101 acttcgattc gctattgggt tattcactca atcacaattc cttcattatt tatcgcaggt
65161 tggctatttg taagtactgg tttagcttat gacgtgtttg gtacacctag accaaacgaa
65221 tattttactg aaacacgtca agatattcca ttaattactg accgtttcaa cgcattagaa
65281 caatttagat aatttagttc ataatttta aaatattacg taagtttga ctagtaagat
65341 agtaaatatt tgctattctt acaatctttc tttccaagat ttttaaaatc agaaactcta
65401 gattatttcg accgtctgga agagggtcag gggtgaaag gaggagaccg ccggtctgag
65461 ggtactaagc ttaaacctaa aaagttaaac gcaaggaag cctttggac aatgaaaaat
65521 gaattcaact tctcttacc aataagagaa agatattctt tttcgatttc gaacctcta
65581 taaaaacaaa agaaaagatt tggttctact gaaagagatg ttaagacat tctagtaaaa
65641 gcgaatcatt tctttagttg tagtggctat taatgtttta tttttatgt caactataa
65701 gcctctact tatccaattt cactgtctg ttggctatca gttcatgctt tagctgttcc
65761 tctcagtttt tcttaggtt caattactgc aatgcaattt attcaacggt agaaaagtc
65821 aagccttgaa cgatattcaa tctcacttaa aggcagctca aatcattct ttttcaaaag
65881 gaaaaataga aagtcgcatc aaaccgatat tcttataaaa atgtgaaaaa agaaatgatt
65941 tgactgaata aacatacaaa aaactgactt ttgtattttt atactgagaa tcaactctct
66001 attagaagaa aaaaatagct aatctttaa cgcaaaatat taaaataaag tttgagaatt
66061 attcaagctt tttagctaga cactattctg gattgagctc tttgacctt taaatcagat

66121 atattgagct caatccactt ggccctcccc ccccaaaaa aaccgagaac tgtcgggtcc
66181 atccagaggt ggcgcgccct tcttggggtc ttcattgcctt tggccctgtg gtcttttggg
66241 attggggatt tggctctctt taaaaaaaat ttaataaaca aagtatttag aagttatttt
66301 aattctgagt attttagtcc gtatagatgaa aaaaattcaa aaactcttct aagtttttta
66361 aatcacgaat attacattca aattaatgct ctataggacc aaaggtctcc cctctcgggg
66421 gaggaacaaa gttcttagtg gttttcaaac aggtttttga ataattatct tttctctat
66481 gggtacttca tactgataaa aaaaatgagc aacccaaatc caaataagca atcagtagaa
66541 ttaaaccgta caagtttata ctggggctta cttcttattt ttgtacttgc tgtattattt
66601 tcaagctata tttttaacta aaatTTTTTc aaatTTTaaC atgtaaaaa ataggacttt
66661 taggttcaaa accctTTTTg aaatccactt tatattttga aaataaataa agtaaaaaa
66721 gttttttcga aaggaaaaa acagtgtatt acaataaga atatttaata aaactaagaa
66781 agttgaatat ttaaaaaagc ttttgcata aaaaaaatat atttaacaat ttagaatttt
66841 actagttctg acattttaat cataaaatgt cagatttggg cacacaaata aagaattatg
66901 tctaatacag gaacaacagg ccgtattcca ttatggctag taggttttag agttggaatt
66961 gcagctttag gtctcttctg aattcttttc tatggttcat acgttggctt aggttcatct
67021 ctataaaaaa gtaaaagtggt ggggtacc ccccaacttt cctgtattga agcttctac
67081 ctgtgagatc atactctgag caaaacttag tttatattat tttcaacaat tgcagagcga
67141 gtctgtttat ggatctggtt ttagtgtgga tatcagaacc cagaattttt tagaaaaagt
67201 ttgagtcatt gataaattct tttcgggtgt ggatacttca tataatgaag atccaacacc
67261 ggaaaaaagt tacttttagt ttttaaaaaa cgttttttta tcaagtggaa aacttggatt
67321 tgttctttaa gaatttaacca acttggctttt gtaacaccag gtaataaacc ttgatatgcc
67381 attctcgcga gcacatgacg agataaacc aatcacgctg aataacctt tgcctacct
67441 gtaacaaaaa aacggtttct taagcgaggt ggagcgtat tgcgtggaaa attagacatt
67501 tttgcagcta aacaaatgg ttgtgctggt gaactttttt tactaagagt ttaatttca
67561 atacgcttgt ttttatattt tgaactaaa ttttgacgct tttttcacg tcaatcatt
67621 cctttttttg ccatttgact ctccaaaata aaacttaatg aaatggggac caaggaact
67681 gaggctctcg tgaacttgagc ttttttttga cctgcaaga accagaacag gtttccatt
67741 ttcagggaag atttttgttt tttagaactt aatttatgac ccaaaaggac cgcagggcca
67801 aaggtctgaa ggggccccca gctctggtgt gcttgagctt gatctctaga ttagattgat
67861 taagcattac aggtcaaatc ccttaatcac actaggtgga agcgtctccag tacacaaggt
67921 caaatcgagc aagtcaccga ggacctaaagg accaaaggct tcaaaaggcc ccacaacacc
67981 aaaggtaaag tgaatccat ttttaactca aaaagttaac gactatttga tacaatttt
68041 ttatgtactt aaactttacg tgaataatat tcaataatta gtaattcatt aacagaacaa
68101 ctaacctctg aacgaggaat taattgtttt acttgatatg attttgaagc attatcaaaa
68161 tgaatatgtg cagggaagtga agttgacgaa cgagcttccac ctggttgaag attaattaaa
68221 cgaacttat cactgggtaa actttggtaa cttggaatat taactggaag aatgatttt
68281 tctgtcttgt tttgacttgt ttttgtttca atagcaatat gcccatgget taccaattga
68341 cgtgcagctc ttaatgttgg tgcataaccg cttctaaaaa caactactatc aagacgatt
68401 tctaacaatt gaagaacaag ttcaccagtt gaacttttca tgttacgtgc tttttttaca
68461 tagcgaacta attgactttc attgacaccg taattaaaac gtaatttttg taacttttt
68521 agacgaatac tatactgtga agctttttga ggtcgagaac ctccatggtg acctggtaaa
68581 gtcgtttttc ctgatgattt ggcagtatta acaagacctg gtaaatctcc taaacgtctt
68641 gtaatacgtg aactgtggacc gcgatatcgt gacataatat aactcctgta aaaattaaag
68701 atcaaacgaa ctattttaaa caacaaaggt ctactggatt tatagtttta accacgaaa
68761 gaaagttttc gtaggaggaa aagcatcaaa aagcttattc tctcgatctt agtttctcaa
68821 gcttgaactt tgttgtaaac acacagaggc ccaaagggcc aaaggcctcc ccttggggag
68881 gaccggagtt cttccccctt gaggggcaaa ggccctcccc ttcgggggga ggtctcgggtg
68941 acttgagctt gtttatgacc tgaaaaggacc ggagggccga aggcctcaag gaccaaagga
69001 ccagaggtct caaagggtg agctttattc atgacctgaa aggtcaaaag cctcactccc
69061 ttgaggaccg aaggtctgaa ggggcccccc agctcttccc ccttcggggg aagggggtga
69121 gctttatgta tgacctgaaa ggaccgggag gaccaaaagt ctcaaaaggg tgagctttta
69181 tttatgacct gaaaggaccg ggaggtctgg tgtgcttcag ctttatttat gacctaaaag
69241 gtcaaaagcc tgaaagcacac caggtcaaaag cttcfaatca caccaggtca aaactctcaa
69301 tcaaccagg tcaaaagcct cgagcacaca agagaacaaa tttccgtaaa cttcgttccg
69361 gttggagcct ccattacctt ggtttctttt aactgaaat ttcgattata aaaatagaa
69421 aaccaagatt gttagactaa cctgttttgt taactgatta agacaactat ggtatagaaa
69481 agctcttctt ctatgactc aagtacacaa gggtcacgag gtcacaactt ttaactttt
69541 tgttgaaaaa agtctaataa aatatcaaaa gcttgagaaa gtgctttttc tggatgaagt
69601 gaacctctcg tccagatttc aaaaaacaca cattcagaaa aaccacgttc ccaataaaaa
69661 ttatttgttg aatttttaat atcttcaaca gtataattca cttttttaac tggaaaaaat
69721 ggccccgaat ttaaaagaaa ttcttcattg tttagattgta aatcaagggt ttcaaaaaga
69781 tttccccctt caattttata tcaattttca aattccggaa atttgctaag tccagttgcc
69841 gttttttgaa taggtaaaat tgtggctaag tattgttctg gatataaag tttagcaaa
69901 ttgggatttt gatttgcctt ttgatcagca cttatataat gttttccgag taatacaaaa
69961 accgaagatt gattattttt gttcttttct gaaagccaat ctttactttt gaaatattct
70021 ggtaaagaaa actgccccaa ataacttgtt aacgaggggt aggatgaaat gtctaattct
70081 tgcgtgaacca acagaggggt agaagaacta tccccggaga ccttttgatt caggtcaagg
70141 ggcttgtgt tccaaacaag taaacgaaa ttcattaaaa ttcccaaat tgaattctga
70201 atacctgaaa ttggtgaaaa ctctgtgagt ttttgcgaac ttaaaatttc gttattttt
70261 gaattaaatt tcaactttgt gatcccaagg ccatgatat cacacaataa aacacgtcgt
70321 aaagcgttag caacggtaat tccttgacct ttttttaact cagacaaata aaacttgcca

70381 taatgaagac tagaattctc aaaagtattt tttcctaataa cttgttgtat ttttttgc
70441 tcaagactat cctcaggaga aaaaggaagt tttcaatcc gatattctaa caattctgta
70501 tccatttttt tataaaacct aataagaact aaaaaattat tgttttaaac gcatgaaaga
70561 aattgtttta taacaagttt ttttatagat tctttttctt aagggaaaga ccacaatttt
70621 gtctgtctaa gaatgtcttt atagaaaatt tcaaaattct cattgaaaga cttctaagtt
70681 tttgtttgtt caaaacaaca aaaaaatgag aaaaaaccg aaatcaata gaaaagactg
70741 caaatattag acacgtcttt tcttaggtgc acgacaacca ttatgaggaa ttggagtaat
70801 atcacgaatt aaacggatac caacccagc agcttgaatt gctcttaag ctgtctcag
70861 accggaacct ggaccactta cttgcacttc aacttttcga aggccttgtt caatagattg
70921 tcgaatagct cctgtggcag cttcttgtgc agcaaaaggt gtactttttc gagcaccttt
70981 aaatccacaa gaccagctg aagaccatga atgacattt ccgtttggat ccgttactgt
71041 tactaatgta ttattaaaag ttgctcgaat atgaacaata ccggtttgaa ctttcttttt
71101 taaaacttta cgtgttttg ccataataaa atgattattt tcttaact ctgtaaagta
71161 ttaattttct tccattttca aggatcgtca aaccttcggt gaaaccaaga ggatttttag
71221 ccaatactcg aaaggcaaga tacaaggggc attcaaaatc taatctgtga tccgcaaggg
71281 ttttaagcgt ttgacctgt atgcttgagc tttatttata acctgaaagc accgtaggac
71341 cataggatta gaggaccaga gggccaaagg cctgaagggg cccccagct ctcggtggct
71401 ttatataatga cctgacaggg ccgggaggac caaagatctc gcttttaaat atgacctgaa
71461 aggaccgtag gtctgggtg attgagctta atatatgacc tgaaggacc gggaggacca
71521 aaggtctcaa aggggtgagc ttttatttat gacctgaaag gaccgggagg tctgggtgct
71581 ttcagcttta tttatgacct gaaaggtcaa agcgtgagc cacaccaggt caaagcgtc
71641 acccttttga acccttttga cctgaaaggg ccccccagct ctggtgtgct tcagctttat
71701 ttaattttat acctgaaagg tcaaagcgt gaagcacacc aggtcaaagc gctgaagcac
71761 accaggtcaa ggcgtcaat cacaccaggg ttaaccaaca aacttgcctt aggaattcta
71821 agaaaaagaa gttcaaaatg taaaaaat atgtacttgt gtacaaaaaa tcttagaaa
71881 tttcaattctg aaagtgaac tttttcgaat tgtttttct ttaactaa gaaaattgga
71941 gctaaaatta ctgttataa gaaaacaatt agacctgaa tacgagctgg gttttgtaat
72001 acaattttctg tttcagcttg gccaaaacca ccaacgtttg gattattagt aattggttga
72061 tcagcgacta cagattgacc ctccagcaact aaaagagttg ggcctgcag aattgtatca
72121 gatactgatg aaccgtcagc tttttcaca gttacggttc ttgcgccatc cgcacctgta
72181 gtgattgatg caattacacc agcagctggt gagttgtaaa cagtattgtt acctttctta
72241 ccatctggat atacttgacc acgaccacgg ttaccacct aataaattgg gtaatttaag
72301 tagtgagctt ttggattgt ttgtggatcc ggtgataaaa tagggaatac catttacta
72361 tattttttac ctggaacagg gccacaact aaaatgtttg aacggctttc actgtatggt
72421 gagaagtaaa gttttccaac ttttgttttc atttctctg gaattcgatc agcaggtgct
72481 aattggaaac cttctggaag aactaaaaca gcaccaactg ttaaaggacc ttttttacc
72541 ttacctaaaa cctgttgtac ttgtgtatcg taaggaattt ttacaactg tcaaatact
72601 gtatttggaa gaactgattg aggaacttca atatcaactg gttttctgc taaatggcag
72661 tttgcacata cgatacggcc tgtggcttca cgagggtttt cgtaaccttg ttgagcatat
72721 acaggatagc ctgaaactga atgagagaaa ctaaaaacaa taaaagcaga aagcaaaa
72781 actgaaattg ataataaact acgttcaaaa aaagaacttt gcaccataac taaaataaat
72841 tataataaaa cgaactttt tagggttata cgtttgactg ctaaaacaaa ttttctttt
72901 tataactaaa taataaagca aaacttaaaa aaaaacaaaa cttaaatagc ttggatctcc
72961 ttaaaaaaca gtggcgaacg acggggttcg aaccccgcaa tcatggagcc acaatccaat
73021 gccttaccgc ttggctacgc ccgcccgttt gttttttttt accactttgt aatcaaatg
73081 taccactagg tacagagagg tagctcatca gcatgtgag cttaaagttt aatcaaatg
73141 ttatgtcttt aatataagta gtatccaagg acaggtaaaa aatatctggg gctaatagga
73201 ctccgaaccta tgacatccac tttgtaagag tggcgtctca ccaactgagc tataaccaca
73261 tcaactgaa atagcatga aaaaaaaat gctttctttg cttttgatt cgttaaagtt
73321 ttatctgta atgaaaaccgt tttcaaggct tgatctcttt gacctgggtg gctcagcgc
73381 tttgacctgg tgtgattgag cgctttgacc tttgaggacc ggaggtctgg tgtgctgag
73441 cttgaaatct gacctgaaag gtcaaagcgc tcaagcacac caggtcatat attagctca
73501 atcacaccag acctacggtc ctttcaggtc ataaataaag ctcaagcaca ccagagctag
73561 ggggccctt caggccttcg gccctcaagg ggggtgagcgc tttgacctta caggtcataa
73621 ataaagctca ccccttgag acctctggtc ctttggctct tgaggccttt ggcctccct
73681 gtcctttggt tctcccggtc ctttctgggc aaagcgtca agccctaat aacaatctt
73741 taattttgac cttctctcga acgataacaa aggccagttt tttaccctaa caaacaagt
73801 tctttaaagc tcaccttttt gagggttatg aattccacca accataacta tctaagcctt
73861 ttcctaataa attcacacct aaataacaaa cccagacaac aaaaaacca aaagcagcaa
73921 ctaaagctgc tttttctccc tcccagatt tttgaaatag taaatgtaaa taaatagaaa
73981 aggtgaacca ggttaataaa gcccatgttt cttttggatc ccaactccag tatgaacccc
74041 aggcttcggt tgcccaact gcgcctgaaa taatacctaa agttaaactg ggaagcaaa
74101 aagctagact tcgataaact aaagaatcca aatctgttaa aaaacgctgt ttgaaat
74161 gcttttaattg tggaaaagaa atgtttccgt ttaattgagg gatctcgaa acagattttt
74221 gatctgagga cttttagttt aattttgttg gttctgtttt atttttatta aataaaaa
74281 gaaccaaata aagaattgaa aataaagaac ccgcaagtaa agccgaataa cttaatagca
74341 taatactaac atgcattaat aaccaatttg attttaaggc gggactaaa ggttaattt
74401 ctctcatttc tgcagggagc acctatgttg caaaagtata aataaataaa gctgcagaga
74461 ctaaaaacat accgctaaa gacgaaactt caggaggtaa actttctttt tgcctcaccg
74521 attttaatga ggttgatata cgtaatgaa gatttctgtg tggaaactata gccatcgaa
74581 gaactgtttt atgaatctca ttcgatttac caaactctgc tttttcaat acctctggt

74641 ttgtgaactc ttgtaatttg gaagagtttt ttgttttcga taaatctttt ccccaaatc
74701 ctactgtagg aaaatataga aataataaag accagggttaa aaaaagtaag gattcatata
74761 aattgctaaa aggtggatgg ccagagtcac tccaacgcat taataacaag aaacttaata
74821 aaaaatttag cccaataaca aaaactgaag ataataaaat tccaaaattt tgattttcaa
74881 caataaaaaa ctcttcaaat tttatctgaa ataaaagtct tgaaatttta ctgagatttc
74941 ctccagttga ttcttgaact aaaaaagggt ggtaggctt tacacttgc caactccaaa
75001 gactcaatac tacgctaaaa aataaagtta aaaaaattaa attctcgcata tccatattaa
75061 aaactagtgt aaaaatagagg taaattcttc tcaaaaccgc aggtctagag tgtgaaggaa
75121 aacatttgaa caaaagctca agtattccca gcatttctgg aacgaggtcc acaagggaaa
75181 gaccagagat tggaagtatt caacttagtt gttgatctaa aaaagcagcc tgttggctaa
75241 aacaatctca aagctttcat ccaactaaag atttatttta tccaatggca aaaaaccgga
75301 gagctaaaac atcctccgaa ataagacttc tacaatcca atagaaaccc aaggacccaa
75361 ggctccccc ttccgggggag gaccggagtt ctccccccg aacctgaaag ggggaaagg
75421 cgaagatctc agtaacttga gccttttgac ctgatccgct tgagaacttt gaccgttgag
75481 gtcataagct caagcggaca agaccctctc cctccggggg ggggacctcc ggtcctttgg
75541 tcccttcagg tgataaataa aagctcactt ctgaaagacc tccacctctc gtgggtagac
75601 ctttggtagt cgggtgcttt agcgttttga cctttcaggt catatattaa gctgaaagca
75661 ctgagacctc cggctctttg gtctctctgg tcttttgagg tcagaaatta aactcaagca
75721 gaccagacac cccaaaagag aacgaccttt aggccttagt tttcaaggt taaactcggg
75781 ttatcactcg acaaaagcct ccaaccacta gattttaaga cccaaattat taaaagctt
75841 gataaataaa caaacgtaa aagtgaagac ttcccaggag cagccactac gatctctcaa
75901 tgctattctg aaaaatagac agaacttagg agggggatac tctcgggtgat atgatcttta
75961 ttaattttaa ataataactt ttccggcttt ttcacaaaa caattatggt atctgatagc
76021 caagtgttta ctgctttatt tcttgccttt gttactggta tttttgctgt cagattagggt
76081 acacaactat ataaataaaa ttcaagtgcg aaacacaatc ataaagataa attatcttta
76141 ccttgaggac caaaggacca cagggccaaa ggcccgaaga ggcagctctc tttggttga
76201 gctttattta cgtctcaaaa agaccgggag gtctggtccg cttagcttta tctatgacct
76261 gaaaggaccg ggaggaccaa aggtctgaag gggcacccca gctctggtgt gattgacctt
76321 tattttctgac ctgaaaggac cgtaggacca aaggtctcaa aggggtgagc ttttattat
76381 gacctgaaa gtcaaaagcgc tcaccccttt gaggctcggg gtgattgagc ttaatatatg
76441 acctaaaagg accgggaggt ctggtgtgct tcagctttat ttctgacctg aaaggacctg
76501 aggtctggtg tgattgagct taatatatga cctcaaaagt caaagcgcctc aatcacacca
76561 ggtcaaacccg cttaagcaca ccaggctcaa ccgctcaatc acaccaggtc aaaccgctca
76621 agccacccggg gtttcaactc agtttcccac caaaaaattt ttatcacagt ttgataaact
76681 cttaaccctgt gatcctgtat taataaagac aggatcacgg ggtagaact ttttctctaa
76741 ttatctcaaa ttttttgttt attaaaagtct ttgaaaaact tttgaaatagt tttcaattct
76801 ttaatttatt tagttatgaa aaattttaca acatacttgt ctactgttcc tgattatctg
76861 tttttatatt ctggtttttt agcaggatta cttattgaaa ttaactcgtt ttatccagat
76921 gctctagttt ttccattttt aggcctaacac aagcattttt tgggggtata tagacaata
76981 aaagccttta taaggttttc atgcccatag cttttgattg gggggacttt aactgtcccc
77041 caaaacgcga attcagatcc tgtacttagg gattgtcata aaaaagtgt aacctttcaa
77101 agaaaggact aaaaacaatg tgccctgctt tattctatca ctataaaaact tcttccataa
77161 tttatggctg atctattaca agatcgacca taagtaagaa ccaaaaagaa aaacaatctc
77221 aaaaattggt cttggttagct gagcgctttg acctttcagg tcatatttaa agctcaagca
77281 cggcagagct ggggggccc ttttagacct tggctctcaa aggggtgagc gctttgacct
77341 ttcaggaccg gaggccaaa gcccacaagga ccaaaaggacc agaggtctta aggggtgag
77401 ctttatgtat gacctgaaag gtcaaaagcgc tcacccctt ggggtctggt gtgctgac
77461 tttatctctg acctgaaagg accgggaggt ctggtgtgct tgaactttat tctgacctg
77521 aaaggtcaaa acgctgaagc acaccaggtc aaagcgtca agcacaccag gtcataata
77581 aaagctcacc cctcggagac ctttggctct cctggctctt tgaggccagc gttaccctt
77641 aatcataga gacctagtgc tagtctttta aaaaattaa agctaaact tgcctagctt
77701 acacctacac cttcaagaac aactgaagag ttgtaaatct ctaataagat tacaaggaaa
77761 actgcaataa aagccataaa aacacctatg attacagttg taacctaac tggagctact
77821 ttaccgtact cagagtttaa tggttttaaa gctgttctta aagetgttac tttccagtc
77881 gaagaacttg ctgaaatttt tgttttgcct gtagtacctg ttgcatata gaatgaattg
77941 gttattgtct aatattgaaa ataaaaata tcaaaaatga aagctgggga gcttttcag
78001 aactctgctt ttccctcgaa ggagggtcct tcccacgcga acctcccca attgccccga
78061 ggccttcggc cctctctttc gtcgctgtgt gagggccaaa ggcctcaggt ggtccccctc
78121 gaggttcgag tgggaagaa tccgttctct cggttctgag gtcctttga agtatgacag
78181 ctttaaaagaa ttgtctttaa aaacttccag acttcaaaaa caaaagctct ccaacttgg
78241 tacacttctg ttttgcctaa aaaagactat ctttgtttaa cacatttttt ttcaagacaa
78301 ttttaaaaaat attttttact ttctttttga aataggaaaa cgaatatttt ttcaactaga
78361 aatagagtat tcaagaaact tacttcttta gcattatttt ctagtgaata aagctattag
78421 tctctgtttg tactttctgt tttacataaa agaataagta aaaaatgaag tggaaatag
78481 ataaaaagtg cagttgcaat taaccaaga atgtttactt ccatttttta tttgttttaa
78541 aagagggctc aagttttagt taactaaacg agtcgatctt tcagattact ttttaataa
78601 taaatactag aactttgtat ttaaaactga acaaagttca gtttaactat tttagataa
78661 ttttatttca taattttgt tcacttgaaa catttttatt taacaaaaa tttgttttg
78721 ctactagcaa aaatttctat aataaaaaact caattagaca caaatgagt ttaaaactcc
78781 tactacgaat actaggaaga accaaaagtga tgttctgaa aaaaacttt gtttagaaga
78841 agaccaacca tctggagatg ctaataaaac tggaaacca acaactaatg caaacgataa

78901 agcaatgaag gcaaaaagtg aaatttggaa taaaaaaagc ataacagaat gtttttaaat
78961 ttcccgcaaaa agaagratatt aaccttgggg acttgaacgc tttgacctat cagggtcatat
79021 atcaagctca agcccaacga gacctcccc cgaagggggg gagagctccc cccgaatggg
79081 gggagacctc cggtcctttg gtcctcagtt ctttcgggtcc ttttgcctc tggcccttta
79141 attccttgaat ctttagtaca cttctaaagt tttcaaaaag ccattgataa agactagaag
79201 aataaaaagt ctcaaaagt ctaaatcttc aaatttcagg tcatactaaa gcttagtggg
79261 gcctatttga tagtatgaat aaagtattta cgggtagccg aaaaaaagac tattgttctt
79321 ttaaaaaact tgaatttcaa acccaaaagt ttttaaaaag aacaagccat tttcgcacac
79381 agataaagct tccaagtgat ctcattctct caaccctac tcgaattgag taagattaaa
79441 ggagcaaaag accggaagtc tcggtgactt gatctttatt tacgacctga aaggacctta
79501 ggaccaaaag tctccaaggg gtgatcttta tttatgacct gacaggaccg ggaggtctgg
79561 tgtgattgag ctttagttat gacctgaaag gtcaaaagcga tcaatcacac cagggtcaag
79621 cgatcaagcc acggaggttt cgatggggga agctttattc acgacctgaa aagtcaaaat
79681 gctccaacccc ttgaaaacca taagtcttcc ccttggggg gaagggtttt gaaaaaaatt
79741 aaccaattct agaattctgg tttttgttca tttcgaagtt ttctggccac taaaagagca
79801 aaattcaaaa ccaaaagtct gaatgaaatt gatTTTTTt caataggcgc tcaacggaaa
79861 aaagcaaaa aaacctaaac ccgaagattt ttttatcta aaacttacag aagcttggca
79921 tacaaaagca agtaagaaga aaagaagtgg aataattgga agtacgtcaa caagtggatc
79981 aaatggtgag taagcttctg gtaaacgagc aagaaggaaa actgaagtgt tcagcatagc
80041 aggagttaaa ataaaaagaat agaataacct acatttgggt tttactctaa aaattgcccc
80101 gtcaaaaagg caaagtaaaa ttgcctatt ttagggagcc agaaatttca aaataaata
80161 gccaaaaatg tttaaaatga ttggttttaa aaggtctgaa agttaaaca agaaggtcta
80221 ttctttaagt aaagaaaagt aggtttggag aaaattggtt aaagactgct tgaagaaaga
80281 ccttgtgtgc aagagcttga gttatgacct ggtgtgcttg agccatttga cttttcaggt
80341 cataaataaa agctcacctc ttgagagct ggggggtccc cttcaggcct ccccccgaag
80401 aggaccacag ggccgaagcc ctgaagggga ccccccaggg ccgaaggcct ccccccgaag
80461 gggggaggac cgaaggtctg aaggggcccc ccagctcttc cccctcggg ggaagggggg
80521 aagcgtcttg atctttcaaa tcagatagct cagccccctg agacctccg tcctctggct
80581 ctttgcctct tcattttctc ggtttccatt aaaaactcaa gttaatgtct aacaattata
80641 gattgcccgc acgagcaaat aaaagtacaa ctacaattgg accggcgcct aaaaatagaa
80701 ataagatgac aagttgtgct acaagttcaa gattcatggt aataattgaa gagtgtttg
80761 ctaagaactt cgtgccccca acgggtgatt ggtttccacc caaagggaaa aggaaagcca
80821 ggctgagat ggagaaactt agtgttttaa taattgcttt tttttcaaa gttctgaaa
80881 ctctcagga ataacttcta taaaagtttt tagtttctat taaggccaa aaggccagaa
80941 tgaagtttca agaaaccaa gggccgaagg cctccccccg aaggggggag gaccgaagga
81001 ccaaaagggc tcaaggggct gatctttttt tatgaccgga acggtcaaaag tgcacaacc
81061 ctttgggacc aaaggtctga aggggtcccc agctatcgtt ggcttgagct gttatttga
81121 cctgaaagga ccggaagggc aaagccctga aggggcccc cagctctggt gtgctcagc
81181 tttatttctg acctgaaagg acagtaggtc tgggtgatt gagcttaata tatgacctca
81241 aaggtcaaaag cgctcaatca caccaggtca aatggctgaa gcacaccagg gcaaaaacct
81301 cagcactcca agaagcatca atggaacaaa gttataggtt gcaagttttt gattaataat
81361 ctgaaaagct caatcaaaaa cctttgtcca caactttttc caaataaatt taggggtctt
81421 gaaatatatc gaatatttta aaaactatta aaacagcaa ttttaaacag aatttcattt
81481 tatcgaatga ttatttgcga gtactttttt tcatttgata gttttataa ttaatatatt
81541 tttacaagcc agccactcct tatttttagt atcttaaca cactttgag taataaaca
81601 ttgattttctg tttattactt caaagtggtg aatcatgatt aaaaatgcat taagatcagt
81661 atttttaaatt taaaaacaat gaaccaaatt tttgaaaaac tttgtgaaa aagaaaagcg
81721 cagtgatttc aaaattattg caaacctctg agttattttt attataaccg aaggagtgat
81781 ttagcgatta agtcataacc aaaaaatact aggtatttga cctttcactt ttggaataa
81841 gctcttaaaag tttgcaatat tttgatcgat caatagctaa aggaagaaa ctcagagaga
81901 aacggttcaa aggagcaaa gccagcaaac tcggtcctcc cagtcggacc ttgtgtgctc
81961 gagagctttg atctttcaga aagctcaagc acactagacc tcaaaagggg gagcgtttg
82021 acctggtgct cttcagccat ttgactcagg tcataataat agagctccac ctttgagac
82081 ctgtggtcct tcggtcctac cgatccttta gtactaataa tcggaaaaaa gaaaaggttg
82141 aagaaaagct gcctaaaaaca gcaaaatttt ggtattttac tgataacttt ggattttaag
82201 aatagaaaa gttagcttct aggcggaagc aaattttatt tcaactcaagc aaaaaagcta
82261 aaacttgaat acaaaaaata aaagtgtttt ctaagaaaac actgaagtaa caactgaa
82321 aacaggctta attttaagt atatatagat ccagaaacat aacaaggcag ctttattagc
82381 gggtaacgag attcgaacgc gcgacctaat ccttggcaag gactcgtctt accactgagc
82441 tataccgctt ttaagctttc taatgattat acctagttt gcaaaaagaa cttttcattg
82501 gttcttaacc agagaattta tgatatctta tctgtatccg ggccaggagg gattcgaacc
82561 cccgactcca tcgtccgtga cgatgtgctc tagtccactg agctacaagc ccttgcactt
82621 atgatataat aaaaaaatag aaggggtata cacaaaaaat taaggaagta agggccaaag
82681 gcctcaagga ccacaagacc gaaggtcctt tggctccttg gacctctggt cctgtggtc
82741 ttgagggggc gaagtttttt ccgaacgatt tcgtctttgt cttaaaaaga tcaaaagtaa
82801 gccctactgg ttttggctct ggctcgctca gaaagaaaaa cgatctttaa attgagagct
82861 ctttattttt ataagataag cgacaaaagt ataaaggggg aagcaagccc cctcaaaaaa
82921 aaacgcaata ctaaacataa agaataacct tctttaagct ttgtccatgt gaatcaaaaa
82981 aagcaatcaa agttttaa gaacttcaact tgagttagt aggtataaa aaaaatcact
83041 tttataaca acaaaattta gagtgtttat aaaatgttat aaaaaactga aagtaagtca
83101 aatattcgtt tttataagga tgagaattat tctgaaaaaa agaatgaaa gttctctttt

83161 ataaactaaa tcagaaaaaa accaaaggac cataagggcc aaaggcctcc ccccttcggg
83221 gggaggacca aaggaccgga ggtcttcccc cgaaggggga agacctcaag ggggtgagcg
83281 ctttgacctt tcaggtcata cataaagctc acccccctga gacctctggt cctttggctc
83341 ttgaggccct tggcccttag aaatcttccc ctttttaggg ggctattatt ctctctttta
83401 ttgggcccga ggccaaaatt taatatgggt ttacctgggt atcgtgttca taactgtggt
83461 ttraaatgacc caggctggtt aattgcggtt cacttaatgc acacagctct tgatcaggt
83521 tgggctggtt caatggcttt atacgaatta gctgtttttg atcctagtga tccgttcta
83581 aacccaatgt ggagacaagg tatgtttgtt ctccattca tgactcgtt aggtgtaaca
83641 ggtcatggg gtggctggag tattactggt tttactggt gaaacaattg gtaatcctgg aatttggagc
83701 tatgaaggty tagcagcatc acacattatt cttcaggtc tattattctt agctgctgtt
83761 tggcactgga cattctggga cttagaatta ttccgtgatc caagaactgg tgatccagca
83821 cttgattttc caaaaatttt tggatttcac ttattcttat caggactctt ttgtttcggg
83881 tttggtgctt tccacgtaac aggattcttt ggtcctggta ttgggtatc ttactcgtac
83941 ggtattacty gaagtgttca accagttgca ccagcttggg gtccagaagg gttcgacct
84001 tacaaccag gtggtattgc aagtcaccac gtactgtctg gtattttagg tattttagct
84061 ggtttattcc acttaacagt acgtcccacca caaagacttt acaaaggctc tcgtatgggt
84121 aacattgaaa cagttttatc ttctcaatt gctgctgtat tctgggctgc tcttgyagt
84181 tctggaacaa tgtggtatgg ctacgtgca actccaattg agctttctgg gccaaacaaga
84241 taccaatggg accaagggtt cttccaagca gaaattgaca aacgtgtaca aaaatctctt
84301 tcagaaggaa agaattcttt aactgcatgg gcttcaattc ctgaaaaatt agctttctac
84361 gattacattg gaaacaaccc agcaaaagggt ggtctattcc gttcaggtcc tcttgataac
84421 ggtgacggta ttgctgctgg ttggttagga catgctgtat ttacagataa agacggccgt
84481 gaattaacaa cacgccgtat gccaacattc ttccgaaacat tcccagtaat tttagttagc
84541 caagatggta ttgctcgtgc tgactgtcca ttccgtctgt cagaatctaa atactcaatt
84601 gaacaagtag gcgtaagtgt tactttctac ggcggtgaac ttgatggcat tacatttaat
84661 gacccagcaa ctgttaagaa atatgcacgt cgtgcacaaac ttggtgaaat ttttgaattt
84721 gatagagcaa cattacaatc agatgggtgt ttccgtagta gccacgtgg ttggtttact
84781 ttcggtcacc tatgctctgc tctattattc ttctttggtc acatttggca cgggtctcgt
84841 actattttcc gtgatgtatt tgctgggtatc gacccagacc ttgatgaaca agttgaattc
84901 ttgacttccc agaaaactgg tgacgtaaca actcgtctgc aagctgttta attttctttt
84961 actgattggt tagggccaaa ggcctcaaga gggtagcgt ttatgacctg aaaggtcaaa
85021 gtgctcacc tcttcccccg aatgggggaa gagctggggt ccccttcaga ccttcgggtc
85081 tcccccttc ggggggagc cttcggctct gtygtcctt agggaaacag agtaccgacac
85141 taagatcaaa aaagtccaca atctgaaaac attacacaaa tcaattgttt agattttttt
85201 cgcttagtgt tcaaggatgc aattcttgcc gttcaattgt cccgcccccc ctaagactcc
85261 cctctttcag gggatataatt aaggggggag accgctctgg tctttgtgt tttttgggtg
85321 ggggtacccc ccttttgact aaaaattcaa aaatctaacc ccaaacttca gactctggt
85381 ttctctcaa tcaaaagaaa gtttttctt cagtttgttg aaactttcgg tctatcagta
85441 ctaacttaga agaggtcaga aaaaaaaga aaatagaat agaaactcga aaaagttaat
85501 ataaagtttt ctttaagaat tggattttat ccaaaaatag attttcacc tctttgttta
85561 gaacaattaa tgaaagattt aagaactaga agctatattt agagataaat caaaattctt
85621 tatcttaaat aagcaggcac tgaatcattc tgggctggtt tatcccgaaa gaagtgtgct
85681 ctagtctctg atgaattcaa aacaattcat ggagtaacta tggagcttt agtatacact
85741 ttccctcttg ttggaacatt aggtattatt ttctctcaa tttcttttag agaaccacct
85801 cgtattaaag attaatttta aaattttaaa ttaattcaaa gttaaaagat ctgaagacca
85861 aagctctccc tctcggggg aggaacggag ttctcccac gcgacctag aggcctttgg
85921 cctctctctt acatagcatt gtgagggcca aaggcctgaa gggggtgagc tttattcga
85981 cctgaaaagga ccgggaggtc tgggtgtgct cagctttaa taatgacctga aaggtcaaa
86041 cgctgaagca caccaggta aagtgtcac ccccttggg tctaactcgt tctttctct
86101 ttgctcattc gagtactctg gttaatgttc taaaactaaa cctgatttt ttgtaattaa
86161 ttatggaac accagctttt atttatgcca tttttttatg gtttctacta attagcgtaa
86221 ctggttattc aatttatggt gggtttggtc ctcttcaaaa agaactaaaa gatcctttt
86281 aagatcatga agactaaaat aaaaatacca atgtaataaa tttgtttata gaaaaaacg
86341 ccggagtttt ttgctccggc gaacatttga aaatatccaa aagttttcca tctctgaaa
86401 aaggaaaaaa aacaaataaa aaatttgcac acctcccac acaagcctca caacgagagg
86461 aaagagaggg ccaaaggcct cccccctcgg gggggagcgt cgcgtgagaa gaactccgtt
86521 cctccccca agggggagag ctgggggggc cttcagacct tccgtctca aggggtgag
86581 cgcttcgacc tttcaggtca taaataaagc tcaacccatt gagacctccg gttgtctgct
86641 ttttctgat ttgaaaaaa actttaaaaa taagaaaata tacgaaactt atgagtcaa
86701 acgaacgtcc agtatttcca tttactgcaa ttgttggaca agaagaaatg aaattagctt
86761 taattttaaa tgttattgat cccaaaattg gtggagttat ggttatggga gatcgtgga
86821 ctggaaaatc taacgacgta agagcactcg ctgatattct tccggaattt gaattctgt
86881 caaatgatcc atttaattcg gacccaaatg atcctgaagc aatgagcgt gctgtactgt
86941 acatgattcg tcagaataaa ccatttgaaa tcacaaaaac aaaaattacg atgggtgacc
87001 ttcatttagg agcaactgaa gatcgtgttt gtgggacaat tgatattgaa aaagcttaa
87061 cagaaggtgt aaaagctttt gaaccgggtt tattagcaaa agcgaactcg gatttttat
87121 acgtagatga agtaacttg cttgatgatc acttagtaga tgttttatta gactcagcag
87181 cttcaggatg gaatactgta gagcgtgaag gtgtttcaat tctcaccaca gcaccttta
87241 ttcttgttgg ctctggaac ccagaagaag gagaacttag accacaactt tttagacctt
87301 ttgctatgca cgcacaaatc ggaaccaatta aagaacttga actaagagta aaaaactgtg
87361 aacaaagAAC aagttttgat caaaatgctc aagctttcca attagaattc caaaattcac

87421 aaaagaact ccgtagagaga attattcaag ctcgtagcatt attacaaaa attcaactaa
 87481 gttatgaatt acgagtaaat atctcaaaaa tgtgttcaga attaacgta gatggtttaa
 87541 gaggcgatat tgtaacaaac agagcgtcaa aagctttagc tgctttttaa ggacgtcaaa
 87601 atgtaacacc taaagacatc tatacagtg gacgtctttg tttacgccac cgtttaagaa
 87661 aagatccttt agaatcaatt gatacaggaa caaaagttca agacgcattt acaagagtat
 87721 ttgggtgaagt ataattgcct actggaacaa gagagcaaaa atgaaattca tacactgacc
 87781 aaaggaccgg aggtctcaag agggtagcg ttttgacctg atgtgattga gcgttttgac
 87841 ctgggtgtgat tgagcgcctt gacctttgag gtcatatatt aagctcaatc acaccagagc
 87901 tggggggccc cttcaggcctc tcggccctca aggaccaaag gaccgaaggt ctgaaggggc
 87961 cccccagctc tcccccttc gggggaaggg ggtgagcgtc ttgacctggt gtgcttcagc
 88021 gatttgacct ttcaggctcat aaataaagct gaagcacacc agacctcccc gtctttcag
 88081 gtcataaata aagctcaccc cctccccccg aagggggaag acctcccccc gaatgggggg
 88141 agacctttgga tccctcaaggg ggtgagcgtc ttgaccttc aggtcataaa taaagctcac
 88201 ccccttgaga cctctggctc tttggctcct gagacctctg gtcccttgggt cctcatcaaa
 88261 aaagggtcct gtgggtattgt ggtttccctc tgttgattc tgaatagggg caactgcaag
 88321 ttttcagctt gcgattctaa aaaaacaaat ttgacctcaa caatgaaatt attaacctaa
 88381 aagaaaaaac tttagcgcac atagatcaag gggaaagcac cctgtcata cttgcgtctc
 88441 caaaaaaat acgatactat taaaaatagat tggctcgtgt ggtgaaatgg ttgacctcgc
 88501 acacttaaaa tgcctgggac ttgtccgtgt cggttcagat ccgaccagcg gcataatgc
 88561 tttctttaag aaaatttgaa aaagcgacaa aaggggaaat tctctatcta cacatatgta
 88621 gatagagaaa taataaatag aagtaaaaga cttaaaaaa aaaataaccc ttggtatttt
 88681 cactgaactt ttggggggga aagaaattga ctcatagtca tccccctggg cgaaaaaatt
 88741 gtttcttcca tgctttagtt atcccccaag aggaagaagg tcgaggggaa agaaaaatcat
 88801 tgacctttgt tgcttccgaa aacctctca tttttggtca ctccacaatt ggaattggga
 88861 gattgtttac acaattttaa aatataagtc tgagtaaaag acatcgaaga taagttgctc
 88921 ttaaacatc ttcgacttag tgtaaaaatc aaaaatgaaa ataaacgaat aaaatccatt
 88981 gaccaactgc gaaaacgctc tttcaaaaata aaaaacttct cttttttct atctacaaa
 89041 gatgtagttt agtactaaaa gaccataatt tttttattcc cattgttaga acagtgttaa
 89101 aataaataat tcttttctat gagtaaaagt tataattggt ttaatgagcg tttagaat
 89161 caagtaactg ctgatgatc cgggtcctaa tacgttccgc cgcacgtaaa catctttac
 89221 tgttttgggt gtattgtatt aacatgtttc ttagtccaag tagcaagtg cttcgtctg
 89281 acattctact accgtccaac agtaacagaa gcgttccgtt ctggtgaaata cattatgacg
 89341 gatgtaact ttgggtgggt aattcgttca gtacatcgt ggagtgcttc aatgatggt
 89401 atgatgtaga tccctcagct tttccgtgta tacttaacgg gtgggttcaa aaaaccagct
 89461 gaattaact gggttaactg tgttctctt gcatcaatta ctgttcttt tgggtttaca
 89521 ggttattctt taccttggga ccaagtggga tactgggagg taaaaattgt aactgggtt
 89581 cctgatgcta tccctggcgt aggttcattt gtagtgaat tattacgtgg tggcgtaggt
 89641 ctgaggtcaag caacattaac tcgttctac agtttacaca cattctgact tccacttct
 89701 gccgcaacat tcatgcttat gcacttctta atgattcgt agcaaggtat ttcaggcca
 89761 ttataaaaa cttaaatca aagtaatttat aggacaaaa aagccgaaga accaagaac
 89821 agaaggacca acggggcgaa gcacacaagg gcaaaagcaa acgagcagtc gaagtcaag
 89881 cgctgaagca caccaggtca aagtgctcac ccccttcccc cgaatggggg aagagctggg
 89941 gggcccttc agacctcgg tcccttgggt cttcagggcc aaaggcctcc cccctcggg
 90001 gggagggtct cgggggaaga aatttcaaa atattttttg gggtacccc cctttctat
 90061 tcaatttaat aatattgtcag ttacaaaaaa accagattta tctgatccag tactaagagc
 90121 ctgatgtaga aaaggtatgg gtcaacaatta ctacggggaa ccagcttggc cgaattgatt
 90181 actttacatg tcccaagtg taatttttg tactgttgca agttgtgtg gtcttgggtg
 90241 gttatctcct acacaattag gtgaaccagc aaacccattt gcaactccaa tcgaaattct
 90301 tccagaatgg tatttctct caacattcaa tctcctcgt acaattccta ataaactatt
 90361 acaggtatta tcgatggcgg ctggtcctgt gggtttatta acggttccat ttttgaaaa
 90421 tatcaataaa tcccaaaacc cattccgtcg tccagttgca actgctggt tcttatttg
 90481 aacttttgca gccgtttggt tagggattgg tgctgctatg ccaattgaca aagcctaac
 90541 acttggtttc ttttaaaact aattaaaaaa tatctctccc cgagccttc ttaagaagg
 90601 cctggggaga gatattttag atctctttta gagtttagta cacagtatct atatttatat
 90661 atacggtaaa taccaagttt tttatacttt tcttagatgc gttcaagtaa tctattata
 90721 aacaagcttt atcaaacata agaaaaacta taaaaagaat attttgatag tgaagatgat
 90781 agcttatctg agtactttac aaaaatacaa attaaagttt ttcatttaa tctttgtac
 90841 ttgagcctt tgacctcgg tcctctggtc cttccccgc gacctcaag gggaaacctc
 90901 tgaggcctgt ggccctcccc ccaattgggg ggaggccttt ggccctcag gggaaacctc
 90961 cgaggccttt ggccctcccc ccaattgggg ggaggccttt ggcccttaa caaataattt
 91021 aatgaaaatt aaaaaataa aaagtgatct ttttaagac ttaaagaca aatttgtgca
 91081 aaactaaaata aaaagaatca aagcttagt tctctttcc agtccacctg catgtctctt
 91141 ttggaacttt attcttttct tgagtcagaa tatgatcctt cctatctta cagaaaaatc
 91201 aactcgaatt cttgaaaata accaatatac gtttgatgtt gacccaaata tgcataaca
 91261 agaaattcga caattttag aaaaaaattt tgaagtcaaa gttctttcaa ttaatacaca
 91321 ccgtccacca agaaaagctc gccgtgttgg acaatttgc ggttttagag cacaagtaaa
 91381 aagagtaatt ttaactctta aagccggtga tctatcgca tctttgaa atttgaaat
 91441 acaagaatg cttttaagtc atagcaattt ctctttaa gtcaagctt tttgattttc
 91501 atcgcgaatta tcaaaaagga tgtgaacaat taagcacaaa gctatgcatc aatattctc
 91561 tctgaagga agagaactta ggtctcgtt cgcttgacct tttgacctt tggttaaagt
 91621 ttaagccact gagggccaaa ggccctcccc cttcgggggg agggaccgga gagccaaagg

91681 cctgaagggg acccccagc tctccccca ttcgggggga ggtcttttac cttgggtggc
 91741 ggccttcgat cctttggccc ctgggtcttt tggctcctct cacgcgacct tcgaagggtt
 91801 cccctcgagg gtcgctgggg aagaactcgg ttccttcggt cctttatggc taaaaaaaag
 91861 cttttcattg getttaattc gcttagccta tagaattcta tctatataat atgagtatta
 91921 gatcatataa agcctataca ccaggcacia gaaatcgatc agtttcggat tttaatgaac
 91981 taacaggagc acttccacaa aaaaaattaa cacgttttgc acatagttcg acaggctcgt
 92041 ataatcgagg cattattaca agtcgctatc gtgggggtgg tcataaaaaga aaatatcgtg
 92101 acattgattt taaaagaaat aaatacagata tcccaggaat tgttaaacg gttgaatatg
 92161 atccaaacag aaatgcta atttcattaa tttgttatca agacggcgaa aaacgttata
 92221 ttcttcattc agcgggttta aaaaattgaa cagaaatcct tccaagtcct caagctccaa
 92281 gtacctagg aaactcttta ccattaatga atatgccttt aggcgttgaa attcataatg
 92341 tggaaattca tcttgaaaaa ggtggtcaac ttgtacgtgc agcaggtaca gttgcacaaa
 92401 ttgttgctaa agaaggctct ttgttaactc tacgtttacc atctggtgaa gtaagactct
 92461 tttttaaaga gtgttggggc actgttgggc aagttgaaa cattgaacat atgaaatattc
 92521 gtatcggaaa agcaggccga aaccgatggc ttggcaaacg acctcatggt cgtggttcag
 92581 taattgaacc agttgatcac ccacacgggg ggggtgaaag acgctgtcca attggtcatc
 92641 cacgtccatt atcaccttgg ggtaaaccag cactagggaa aaagactcgt tcagctaaaa
 92701 aatatagtaa tgcatttaga gtcattta atctttat ttcgaaggaaa agacaagctt
 92761 tctaaattgt gtgcttgagc tgaaggggcc gtcaaagctg gggagccctc tcagagccgt
 92821 ggtcctactc ggttctctcc ccgaagggga gagagaacta gaggccctc cagccctcaa
 92881 ggaccacagg gccgaaggcc tgaaggggac cccccagggc cgaaggcctc cccccgaagg
 92941 ggggaggacc gaaggtctga aggggcccc cagctcttcc cctttgggg gaaaggggtg
 93001 agcgtcttgg gctttcaggt catatataa gctccacccc ttgagacctc gtcctcttgg
 93061 tcttttgggt ctcaaagaat tctaaagact tattaatatg tgttctctgg ttcacaatca
 93121 aagaattatg aacctttcaa gatggagaat tttcgtatg ccaaatacac gtcctctctt
 93181 tattgcaaat cattatttta aaaaaattgt tgaattgaat aataaaaagt aaaaaaaagt
 93241 tattcaaat tgggtccagag gttcaacaat tgttccaatg atgattggtc atacaattgc
 93301 agttcataat ggaaaagaac atattctctgt ttttattact gaccaaaatg ttggtcataa
 93361 acttgggtgaa ttttcaccaa cacgaacatt tctgtggcac atttaaacag ataaaaagg
 93421 aaaacgttaa taaaacattt ttaatacttc gatcataggg attgagttt gaatgaataa
 93481 gatttttcag ctttaaaag aacttctttg tccaacttat gacttttaag tgaactcaat
 93541 gctatggcaa agtacttttg attacctcaa ggaccaaaag accagaggtc tcaaggacca
 93601 aaggaccaga ggtctcaagg accaaaggct tcccccttc gggggaagg ggtgagcact
 93661 ttgacctggg gtgcttgagc tttatttatg acctgggtg cttgagcgtc ttgacctttc
 93721 aggtcagatt tcaagctcaa gcacaccaga cctccggctc tcaaaggta aagggcccaa
 93781 tcaatcacc aggtcaaaag gatcaatcat ataccaggt caaaaacctc tatattattt
 93841 ttacataaaa actaaaaata tgtgaaaaat tccatctta agagcggggg gttacttccc
 93901 ccccttcccc aaaaggggag aaaacttctt tcttctgtta aaggggaaa cttgggggtc
 93961 ccttctcagg cgaaggcctg aaggggacct cccaggacct aaggtctgaa ggggcccccc
 94021 atctcttccc ccattcgggg gaaggggggtg agcgttttga cctttcaagt cagatattaa
 94081 gctcaaggag ccgagaactt tttagggtca aggtcttagg agggaaatat ttaaacgaa
 94141 acaaaaacta tcaatcattt tctaaaaaat aaaaaacata gatttaaagg caaaaatata
 94201 cgagctttaa aagcttgaat actcagaaac tattaacagt tttaaaaaag taaatgaatg
 94261 atcttttgtt actatctttt ttttgaggaa taaaaaatgg tcaaaaaaat tcacccttta
 94321 ggtttaaagaa tccgtatttt tcaaaaacat aaatctgaat ggtttgcacc agtgggttca
 94381 tcaaatcaaa accaatcaat tccataattat gctgctttta ttgagcagga taaaggtgta
 94441 agagattttt ttagtccaga acttcgctgat gctggacttt ctaaaaattct gattttctga
 94501 cgaacaaaat gtttatttgt tgaattacat gttgctcaac caaaaactgt ctttgagaa
 94561 aacggaaaaca acttaaaaaca gttaaaccgaa aaatttcaaa agaaccattca aaaactgta
 94621 agtcgccata ataactcttc tatgagcgc cctgctttat ctattcaagt ggttcaaatc
 94681 ccacaagaag aaattgcaac agatgccact ttattaggac aaaaaattgc ggaacagtta
 94741 gaaaaacgtg tagcattccg tctgtttatg aaacaaagta tcaacaagc cgcagctgct
 94801 ggtgcagaag gaatcaaaat tcaaaattgct ggtcgtttaa atggagcaga aattgctcga
 94861 acagaatggg cgcgtgaagg aagagtacct ctccaacat taaaagcaaa tattgatatt
 94921 tgttacgaga ctgctcaaac aatttatgga attttaggga ttaaaatttg gattttcaaa
 94981 ggaacaaata cgagtcgtgt ttaaaaagac taaagttaac attcttaaaa aatgtaatca
 95041 atattagaac cgcagatttt atactgcat ttatggctag agagaatctg gtgttcaata
 95101 gaacaccgaa ccaaagatta ggtatgttga aaagaaagca ctcttttggg gtaaaacttc
 95161 gttcctcccc cgaaggggga gagctggggg gggcccttca ggggtgctg ggaagacatc
 95221 tcttttcttt ttaccttggg tctaaggatt taaacttttt agcttattca taaagctcga
 95281 aactttaatc ctttgactct ataactttag tttagccaaa tcaacttttc acttgaataa
 95341 agagtattgt aagtccaaa cgaacaaaat atagaaaaca ccatcgtggt cgtatgaaag
 95401 gtaagcttc tctgtgaaat aaaatagcat atggaacttt tgcatataa gctcttgaac
 95461 caggttgggt aacatcccga caaattgaag ctggtcgtcg tacaattaca agatagcga
 95521 aaagaggtgg taaactttgg atctcgttt tcccagataa gccaaattct atgagactcg
 95581 ctgaaaccag aatgggttca ggcaaaaggt caccagaata ttgggtagca gtaatcaaac
 95641 cagggccgat cctctatgaa atgaacggcg tccattaac tttagcaaaa gaagctatgc
 95701 gaaatgcttc atataaaaat cctgtaaaaa caactttttt aacaacaatt gaaataaact
 95761 aatcaaaaaca tttgttaaaa caggaatggt ctaaaagcac ggcaaggggg gttagcccca
 95821 aaaaatctta caaaaccaa gggccgaagg cctgaagggg cctgccagc tctctcttc
 95881 gggggaggaa gagagttctt cccatgcgac cgtctcttcc ctcgcttgg agggccaaaag

95941 gcctcgaagg gttccctcag agggccaaaag gcctccccgc aaatgggggg agggccaaaag
 96001 gcctcgaagg gttccctcag agggttgcgt ggggaaggacc acaccagctc tcaaggggggc
 96061 cccaagctct tgggggtttg aactttatat ctgactttaa aagtcaaac gctccaaagc
 96121 cttaggaata aagttcagacc tcttttaggg ggcaaatgtc ttgtcttttg acgttttcgg
 96181 ggatttttga agaaggagca gaaattagag agtaaagctc aaacaatcga ccttaaaagcc
 96241 aagttttttt tagcaacctt atgattcaac cacaaactta tcttcagtga gccgataata
 96301 ctggggctca aactattatg tgtattcgtg ttctcggaa gcaaaagaa tatgccagtg
 96361 ttggggatgt tattattgga gttgttaaat ccgctctccc taatatgact ttaaaaaagt
 96421 cagatgttgt acgcgcagtt gttgtaagaa ctgcacacgg tgttcgcccgt tcaaacggaa
 96481 tgaccttaccg ctttgatgaa aatgctgctg ttgtaattaa tcaagaagga aatcctcgtg
 96541 gaactcgtgt ttttgaccg atcgcaagag aattacgtga aaaaaat ttt tccaaaattg
 96601 tgtctctagc ccttgaagtt ctttaattaa ataaatattg cacaaaaggt aaaaaatat
 96661 tacgcatcaa ttgttcaaaa tccagaatcg tttgggtatt ctaaaaagg aacttttgaa
 96721 gttccaaaaa ttgaaaaaatt tctgaaatg tctgaaatg tctgaaatg tctgaaatg
 96781 gaacttgctt cgcaagaatt tacacttcta agtggacaaa aaacggttgt taccaaaagca
 96841 agaaaatcaa ttgcaagttt taaaaattcgt gaaaacatgc ctgtaggat ttgtgtaaca
 96901 ctacggcgtg acagaatgta tgctttttta gatcgattaa ttcaacttgc tttaccacgt
 96961 attcagatct tccaagttt aagccaacaa agttttgatg gtcattgaaa ctatacctta
 97021 ggtttaaaaa aacaattaat gttcccagaa attgattacg acaaaaatcga aaaatctcga
 97081 gggatggata tttcaatcgt tacttcagcc aaaaatgatc aagatgcttt ttgtctcta
 97141 aaagcattag gaatgccatt taaagctgct taaaattaac tctttttatt tcttttaaat
 97201 taaaaattct tttactgtaa ttacaagtgt aaatattaaa ttgttcttt ttgttctttc
 97261 ggaggaaaaga agttctccct tttgcaactc atcggcttga atgtttttgt aaagagaata
 97321 gaaatgaagt ttactttgtt gctctttcta agcaaaactg atacccttct ctctggagga
 97381 ccaaaagggcc aaaggcctcc ccttcggggg gaggaaacgga gttcttccca cgcgatcctc
 97441 cccctatttg cggggaggcc tctggccctc aaggggaaaac ccagctcgtc ccccttccac
 97501 tttgaccttg tgtgctcag cgctttgacc tgggtgctt cagagctttg acctttcagg
 97561 ccataaataa aagctcacc ctttgagacc tttggctcta cggtcctttc aagtaataaa
 97621 taaagctcaa gcacaccagg tcaaaagcgg caaatgaaca aggtcctcaa aagtaaaaaac
 97681 tatccagagg aaacaagttt ataaaaactt aaagaaaaaa acgaagcaca ggggggaaat
 97741 taaaacgaaa gctccaacga tcaaaataatc taaaaaaaca aaaatataaa actgatatta
 97801 ctttaaaaaa ctgaaataaa atataattact ggttaacaat tcatatgggtg aattacgctg
 97861 taagcgatag gcttactcga atccgaaacg gaattcagat taaaaaaaca aaagtttctg
 97921 ttatttcaac aattttaaaca agaagtattg ctgaaatttt aaacagtgaa ggttttatgt
 97981 aaaaattcgc tacaactcaa atgacaaaaac cttctcgttg aaaaattagca actaccggaa
 98041 ataaaaactc tactttagat atttatctaa aatggtctca aaaagacggc ttaggaaaag
 98101 cacaaaaaaa tgtcttaaca aacctaaagt gtattagctg tccagggttt cgcatttatg
 98161 caaatgcaaa taatattccc caagtctctg gtggattagg catcacaatt ctactctctt
 98221 caaaaggcat ctttaactgat tatgaagcaa gaaatcaaaa agttgggtgga gaagtcttat
 98281 gtatgctttg gtaaaacctt ttaactagc caattgctga acaaccttac actcaagggg
 98341 tttgagcggg ttgacctggt gtgcttaagc gcttgaccct ggtgtgattg agcggacctt
 98401 tttaggtcaga tattaaagca aatcacacca gagctggggg gcccttccag gctctcggcc
 98461 ctcaaaaggg tgagcgtctt gacctttcag gtcataaata aaagctcacc cctttgagac
 98521 ctttggtcct acggtccttt caggacataa ataaagctca agccaccag agctgggggg
 98581 ccccttgagg cctttggccc tccggctcct tggctcctgta ttgagagaag aaaaataaagt
 98641 cctcatcaaa ccaaaaatag gttttcttaa atgtttttt tggattgggt cctctctttt
 98701 cagtttaatta ggaaggtcaa agaaacaaag ttgttccctt gtgatctttg tgtgcttgag
 98761 ctctttgacc tttcaggtca taaataaagc tcaatcacac cacagctggg gggcaccttc
 98821 agacctttg tctccccccc ttcgggggga ggcctttggc cctagggtcc tttcagggtc
 98881 ttcaaaagct tgaagagata gagggccaaa ggctgaaag ggaagctctg ggaagcatt
 98941 atgggatcaa ccttaaaaaa aaacgaaga taaaagatcg actttatcag gaggagaaaa
 99001 aactacttta tcaactaaata tttcaaaact cccctctttt acagttcaga acagttctac
 99061 agaacttctt cctcaaaaaa cagagataaa ctcaatgaag ttaaaatctc taaggaggaa
 99121 cactggaaag acaaaaattg atcgaaatgg aaggaattat tacggaatca ttacctaatg
 99181 ctattgttccg tgttgatctt gaaaatggat ttaacttctt agcacatatt tcaggaaaaa
 99241 tccgtaaaaa ttatatataa atttttatg gtgatagagt tttagttcaa ctacacctt
 99301 atgatttaac aaaaagggcga attatattat gacaaaaaaa tagttcatca taaacttatt
 99361 gtctttatga ataaagcaaa aataaattca acgaggaatg aaacaaacac aaattgatca
 99421 taaagatgta gcttatctac gtcgttttat taccgagcag ggaaaaaatt tatctcgtcg
 99481 ttttacgggt ttaaaagcca aacaacaacg agagcttaca aaagctgtaa aacaagcagc
 99541 aattttagga ttacttggtt ttggctcttc aacaaaaaac cgtaagtgg ggggtcccc
 99601 cccctctttt tagttcaaaa tagattttta aatgcaactc caaagtgcta aactgacgag
 99661 tgtattttca tcatccaaaa aagttcatag tttccccgaa tcatataggt taagacgggg
 99721 ggaaccccc cgtcttaacc aattgctaaa agagtggttt taccattct ttagatcatt
 99781 tattacaaat ttttaaaaat tatggcagtt ctaaaaaaac gagtgctacg tacacgcaag
 99841 cttttactga ataaccaatg gaaattcaaa attagaccgc aagctttaa agctttatca
 99901 tttagcaaat cagtagaag ggggggactt cctccccgc ctgaaaaaac ttttccaaac
 99961 gatgaatcat aagaacagga ataaaaaac agaaaaaac tcaaaatata tttgcccttc
 100021 ttttttttga taaccattc ttgaacaaga ttggctagga tccagttaag agagtacttt
 100081 caaaagctag aaaaaaactt tttttcgtg cctctattc ttttaaaat tcaagaagtc
 100141 ttttctattt aggtcaaaat tgcaaaatc taaagtggcg agctctttga tcttttatta

100201 cagctcgcca cccaattgct atcttttcaa tagcacctgc cctttgggt gtagaacgtt
100261 gttcctcagt agattgagca atatgatttt gaattcaaaa acttctgaga aagctatttt
100321 atataaaaca aaaatattac cacttttgtt atatactcta agtaacaacy cggggtagag
100381 cagtttggtg gctcgcgaag ctcataatct tgaggtcata ggttcaaatc ctgtccccgc
100441 cactactaaat ataaccggat cttagccgtat cgtgtcaaac agaatagctt ttattttttt
100501 ggtggggggt ttaaaaaatg cccccaccaa attgattcac ttttaaaaaa agaaactctt
100561 ttgtcagcaa atttgtcaac tcaatgagtc aacaatttgc caaaaagaag aaattgtaaa
100621 cttagtcgaag aggacgcata gttaaaactg gctctgtatc acgatcaata cctttttcga
100681 aacctgcaag cgctgcacgt gcacgaccag cgtgccataa gtggcctaca aagaagaaga
100741 agaaaagaac atagtgagaa gttgctaacc agcttctagg gtttacgaag tttacagcgt
100801 taatttctgt agctacgcca cctactgagt ttaatgaacc aagtggtgca tgagtcatgt
100861 actcagctga acgacgttct tgcccagggt gaatgtcatt tttcagttt tttaaactta
100921 aacctgttgg accacgaagt ggttctaccc atggagcacg gtttgaataa atacgcattg
100981 tttcaccacc aaaaatgatt tcaccagttg gtgaacgcat taagtattta cctagacctg
101041 ttggaccctt tgctgagca acgttagcac ctaaacgttg gtcacgaact aagaatgtga
101101 aagcttgcgc ttggcttgcc tctggacctg ttggacaaa gaattcactt gggtaaaccc
101161 taatttataa ccacaccatg caagcagaaa tagcagccat taaagatact cagcagtaag
101221 agtagcttaa gtatgcttca ccgctccata caagagcacg acgtgcccc gcaaatgggt
101281 tagtaagaat gtgccaata ccaccaaaaa ttaataaatt accaacccaa atgtggcccac
101341 caactacgct tccaagggtt tcacagctta caatccagcc gtcaccacca aatgggtgatt
101401 ttagaaggta accgaaaata acgctggag ttactgttgg gtttgaataa atacgcattg
101461 caccaccgctc tggagcccaa gtatcgtata aaccacaaa atacatcgtt tttcaacta
101521 ataagtaagc accaacacc agaagaacaa ggtgaatgcc tagaatagtt gtcatttttag
101581 ctttgcctt ccacacatag ccgaagaatg ggaagcttct ttttaattg ttcagggccga
101641 tttagagctg gtatagccca ccaaaaccaa gaacagctga tgagattaag tgaagaacac
101701 cagatacaaa gtatgggaaa gtatcaagaa cttcaccacc agggccaaca ccatagccta
101761 aagtagctaa gtgaggaaga agaattaaac cttgttcata cattggttt tctggaacaa
101821 agtgtgcaac ttcataaag ttcattgcac cagcccagaa aacaattaga cctgctgtg
101881 caactgtgag acstaaaagt tttctgata agttaattaa acgagcgttt ccagccccacc
101941 aagcgaatcc agttgattct tggctcggc cgcctactac aagagtaccg ttaaaaagag
102001 tttccataag aagagacttc tgagaaaatg aacttattac atgagacaag atgtgaaaa
102061 aaattcatcg aaggacctga ggtcaaaagt gcagagaaaa gaacaaaaa gttcaatttt
102121 ttagaacctt aaaaactacag tttttaaata actgtttcaa aagtatgatc ttttgaat
102181 agcagaaaagt taagattatt cgttttaaaa gatagaatca ctttaataa taataaagcg
102241 acatacctcc aaggaataat atagctctta atagaattat agcatataca caatggtaaa
102301 aagaactggg ctatcagagt tcataacaat acataggata tgactcaaaa agatttatgt
102361 tttgaaagc aaaaataaac ataaaaaagt ctcaaaaaa gttcagtat cttattotta
102421 atttaataa aaccctgtt tattattaat taagaataag atactgaagc tttttggat
102481 ctacacacga cataaaaagg aatgagtcag caattatggg ggggtattac cccccccatt
102541 agcaagctgc aagtcgcaac tttttcttaa tactagccaa cgattgaagg agcttcaact
102601 gcagcaaggt caagaggaa gttgtgagcg ttacgctcgt gcataactc cacaactagg
102661 ttagcacggt taacgatgc agcccaagtg ttgattcac gaccttgaga gtctacaaca
102721 gattggttga agttgaaacc gtttaaggtg aaagccatag tagagatacc tagagctgtg
102781 aaccagatc ccataactgg ccaaaatgct aggaagaagt gaagagaacy agagttgttg
102841 aaagaacgct attggaatc taaacgacca aagtaaccgt gagcagctac gatgtgtga
102901 gtttctctt cttgaccaa acggtaacca gcgttagctg attcgttctc agtagtttca
102961 cggataagtg aagatgttac taatgaacca tgcatagcag agaacagaga gccaccgaaa
103021 acgccagctc cacttagcat gtggaatggg tgcattagga tgttgtgctc agcttggat
103081 acgatctga agttgaaagt accagagatc attagtgtc agatgagttc gcaaaaactca
103141 cctcgttcag taaagaaact ttatcgtatt cttttataa gtttttctg aactgcctga
103201 cgtcaccgac aggatcagac catatcatgt tcatatggtt attctgaact ttggagattt
103261 ctaatgatat aaaaatcagtc gtcactcccc aaagcaaaa gaaatccctc aaggggggtga
103321 gcactttgac ctttcaggtc atacatacag ctcaaccctc tgaggcctt ggcctttta
103381 tttcttcaga actcctagaa cctcgcatt tcgagagcgc ttgctccta ctccctctg
103441 atggtcgtta cgcattgctt tgctgtgtc tttgttttt taaactctct tgtctctttg
103501 gagttaattc gcgccaatcc ggaatgtgc gatcatttaa acgtttctgt agatcacgctc
103561 gagacatatt tgtcacacga gatgcttcag caaggcttaa aaaagaacct ccaggggaaa
103621 ttttctctt cctagatgta aaagatggaa tacaattcc acgtcctagc tcgtaatttg
103681 gaataattgt gcgttcttca gtgaaacgct ttttcaattc tgggtgtgtg attttcccaa
103741 aaaaactatt attttcaacg catcgtctg tattataaag tttttcaata gaaattttt
103801 ttatgtattc tttttcaaga cttaaacgaa gatcattcga ctcgtactct atcttcaa
103861 taagaattga gacttcaaaa gaatgtatac cgtagatttt acaactctct gtcattttt
103921 tgttttgatg tgtttctcgt tctaacatag atcagatgtc acgaagctgt tgaatgat
103981 ttttagattg gccaatataa atgaaccctg tttgggtgtt ttttatagag taaataccag
104041 gaaaaacag aagagaggac aacgagaagg tttttagcac gggattgtct tttgattgtg
104101 ctttcaaat tggaaataa aaaacaaaag aggttccccg tttagacaag ttttaacctc
104161 gtataacttt cgtagtatag aggggcaatg ttattcactt aaaggcctac cgtcagagaa
104221 tgaaccttga ccgattgggt agatgatgaa tactgcagta gcagctgcaa ctggctgaga
104281 ctaagctaca gcgatccaag gacgcatacc taaacggaaa gaaagtctcc actcagggc
104341 gatgtaagaa caaattccta agaagaagt ccaaacgatt agttggtaag gaccaccgtt
104401 gtataaccac tcatctacag aagctgcttc ccaaatggg tagaagtgt aaccaatagc

104461 gtttgaagtt ggaattacag caccagaaat gatgttgta ccgtataata atgaaccaga
104521 aactgggtcca cgaataccat cgatgtccac tgggtggagca gcgatgaacg caatgatgaa
104581 tacactagta gctgttaata aagtaggaat cattaacaca ccgaaccaac cgatgtataa
104641 acgggtttcca gtgatgtta cgaattcaca gaagttagac caaattgtcg cgttttcgcg
104701 ctttaaagtt gctgtcatga tagttaaata tgtaataaaa gatatctctt cgaagagaaa
104761 agaacttaaaa attcctttca tttgaaqtta tttttattat aacagtagca aaacacaaaa
104821 attagtggtc aatagcccaa tttaaaacc ccactttggt gaaaaaatat cccgctgcct
104881 catcttaaag agatcaagaa aatcagttct ttttattttt gtattaaaca atacaactga
104941 aataaatggt cgattagtag ttaatgttta tacaagatcc agttaatat tatttgacaa
105001 caattaaaaa tttttataaa tttccatatac gtggagttaa gagaatgagt aaagaattat
105061 ttttagatt tagaacataa gtatattata aagagtaaac aatttatatt atactctcga
105121 tgttacctaa aaaggaacga attcattatt tttggaatct attctgattt tatttctatt
105181 tttagttttc taaaatctat ctaatgtttt ttaatgaata aaaaaccgga ggactaaaaa
105241 cctcctaaaa ctaagaaaga attcctttgtg tttgtgtttt ttaacagttc agtcttaag
105301 atttcataca agctttgtat tcccgaattg ggaatctttt taagtgaact aaaagtttga
105361 gctactaaat cttcgattta gtatttgata ttgatgtgaa ttggtccacc taggcaatat
105421 ttgtggatct ctaatgtttt aagtttcgta attatttata aggtagtctg cattagtaaa
105481 ttgatttate cttttttttt tttttttttt tcaagatttt ttaacaatgt atctctatga
105541 tgaacgaaag tctagtgtgt tctgagttga ttcattgaaag gaccgagaga aataaagatc
105601 tccgtggctt gggcttgatt tctgacttca aaggacctct ggtgtacttc gatttatgac
105661 tcgaaaaatc aaagcgcttt agcaccatgag gctctcaagc acatgaggct aattttaaac
105721 tggttcgttc tcttttttaa atgtatagtt ttaatggaaat ttttattttt ttcttttca
105781 aaaagagact ttttagagcg gacaatgttt tttttataca gggtagttga tataataact
105841 gcatgagaaa gtttttagaca aacaaatcaa cagaaaagcgt tgcctcctga aaacattgtt
105901 ttctaaaaat tttcttttga ttttttga ga ttttgatcct ggctcaggat gaacgctggg
105961 agtatgctta acacatgcaa actgtacgga agaaaagcaa ttttctttag tggcggacgg
106021 gtgagtaacy cgtaaagaatc tttccttttag agtggataaa cctttggaaa cgatggttaa
106081 tcccgcataat gaggcctctg gtcttgaaa agtgacgaat gtatcgtat aggatgagct
106141 tgcgtctgat tagctagtgt gtgaggtaat ggctcaccaa ggcgatgatc agtagctggt
106201 ttgagagaac gactagccac actgggactg agacacggcc cagactccta cgggagcag
106261 cagtgggaa tttttcgaa tgggcgcaag cctgacggag caatactgag tggaggatga
106321 aggctcgtgg gttgtaaaact ccttttaca gagaagaaaa cattgacggt atctcgtgaa
106381 taagcattgg ctaactacgt gccagcagct gcggaatac gtggaatgca agcgttatcc
106441 ggatttattt ggcgtaaagc gctgttaggt tgcttcgcaa gtttgcgttt aattgttctg
106501 gctcaaccga atcccagcag caaaaactat gaagcttgag tgtgtagag gaaaaaggaa
106561 ttccctgtgg agcggtgaaa tgcgtagaga tagggaagaa caccaatggc gaaaagcattt
106621 ttctggcctt gcaactgacac tgagagacga aagctagggg agcgaagcggg attagatacc
106681 ccgttagtcc tagccgtaaa cgtggaaac ttgatcttgg ggtccacaag agtctatggt
106741 cgtagctaac gcgtaagtt tcccgcctgg gaagtatgct cgaagagtg aaactcaag
106801 gaattgacgg gagcccgac aagcggtyga ttatgcggt taattcgtat caacacgaag
106861 aaccttacea gggcttgaca gtcacgaag aaggcagaaa cgtctttgtg ctggttact
106921 ccagaacgtg aacacaggtg gtcacggct gtctcagct cgtgtcgtaa ggttgatgtg
106981 taagtcacgc aacgagcgca acctcgtgc tttgttcaat ttagtcaag cagactgctg
107041 gtgagaaatc agaggaaggt gaggatgat tcaagtcagc atgcccccta cggccctggg
107101 tacacgcgta atacaatggc caatacaaa agaaagcaag gggcgacccc aagcaaacct
107161 caaaaaagtt gttcagttc agattgttag agattcgtt cctacatgaa ggtggaatcg
107221 ctagtaatcg ttggtcagcc atacaacgggt gaatcgttc cgggctttg tacacaccgc
107281 cctgcacacc atgggagctg actatgccc aagtcgttac tccaacggtt tacgagggg
107341 atgcctaagg cagagttagt gactggggtg aagtcgtaac aaggtagccg ttctggaag
107401 agtggtcgg acacctcctt aaaaaggaaa aatatactca aaagttcttt ttgactctt
107461 gagaaccgca aggttaaatc tagtatgttt atcactactag acaggggag catatagatt
107521 tttaaaaagc tttccc aaaag attttataag cttaaaaact aaagtcgtct aaaaactaa
107581 agcgtatcaa acgctctgtc taectgccat gaggttcttc gaaactcaat tggcgggcta
107641 ttagctcagc tggtaagagc gcacccttga taagggtag gcccctggtt cgaactcagg
107701 atagcccagg acttaggtct tataaaatgt tccggggggt atagctcagt tggtagagc
107761 ctgcctttgc aaggcagatg tcagcgggtc gactcctctt acctccagcg aatagaatga
107821 agataatcat cccagaaaaga ttaaagcaaa gactaaaatt ctccccttg tggatgaaaa
107881 ctgtcgtctt gtgtttcttt tttggctaat tttacagttt tcaaaaaat aaagcttat
107941 ggtggatacc taggcaacca aagacgatga agggcgtact taccgacgat acgcttcggg
108001 gagctggaat gacgctttga tcccgaagatt cccgaatagg gcaacctata taactcttt
108061 ttgaattcat agaaaagaga gagacaaccg agtgaattga aacatcttag tagctcaggg
108121 aagagaagc aaaaagcgtt cccggagtag cggcagcga aatgggaaca gcttaaggga
108181 tatagattct ttttgaatct atctcgggt gtgggacatg aaatggataa ttaaaaacac
108241 aaaactgaaa cagctgagtc ctgtaccaaa gaaggtgaga gtctgttag caattgtcgt
108301 ttttataact cccatgaatc ccgagtagca tggggcagct gaaatccctg gtgaatctgc
108361 gaggaccacc tsgtaaggct aaatactctt gggtagccga tagtgaaca gtacctgag
108421 ggaaaaggta aaagaacctc atcaggggag tgaatagac catgaaaccg taagcttaca
108481 agcaatggga gagtcctgag acttgaccgt gtgcctgttg aagaatgagc tggcgactta
108541 tagggagagc cttggttaag gaactcagtt cgcagccag agcgaagca tgtttgata
108601 gagcgttgt ctctcttat agaccgaaac ccgggtgatc taacctgac caggatgaag
108661 ctctcgtgat acggagtaga ggtcccaacc gaccgatgtt gaaaaatcga cggatgagtt

108721 gtggttaggg gtgaatacc aatcgaactc ggagctagct ggttctcccc gaaatgcatt
108781 taggtgcagc tgcttttcat gaactattct ggggtaaagc actgtatcgg tgcgggtgc
108841 gaaagcggta ccaaactcgt gcaactaag aatacagaat atgaatccga caagcaagtg
108901 agacagtggg cgataaggtt cattgtcagg agggaaacag cccagatcac sagctaaagg
108961 ccctaataaa gtaactaagt gtaaaggagg tgggagtgtc aagacagcca aaaggtttgc
109021 ctagaagcag ccattccttga acgagtgcgt aacagctcat tggtcgagcg ctcttgcgcc
109081 gaaaatgaac gggactaagt actttgccga agctgtgagg ggaggtttac ctctccggtg
109141 ggggagcggt ctgcttgcga gtgatggagg cgggtaactg tcctcttgac gaagcagaag
109201 tgagaatgct agcttgagta acgaaaacag cagtgtgat ctgctgcccc gaaaaccacg
109261 gggttctctc ggaaggttcg tccgcggagg gtgagtcagg acctagggtg tagccgaaag
109321 gcgagctga tggcaaatca ggcgaacatt cctgtactag tttgtattta gtcccggggg
109381 acagggaagg ggttcgcccc tctcttttgg aataaaaaac aaaagtatcc ttcaagaaaa
109441 gcccgctcga cttttgata caaattacct gtaccgtaaa ccgacacagg tgggtaagta
109501 gaataactc aggggcgca gataactctc tccaaggaac tccggcaaat gggccccgaa
109561 ctctcggaag aggggtggga acttatgggt ctccgcagtg acctgaccca agcagctgtt
109621 taccaaaaac acaggtctcc gcgaagtcgt aagacgatgt atgggggctg acgcttgcgc
109681 agtgccagaa ggttaaggaa gttggttcgg gaacttgttc ccaaagctgg cgaccgaagc
109741 cctggtgaac ggcggccgta actataacgg tcctaagggtg aacgcaagct gccttagtaa
109801 aacttaacta tatgctggaa actcctcaaa aaggtgcgtc tactcattta aaattcaatt
109861 tttgacaaaa aaagcacgtg aaacggtaag tccggttttgg ctttctttat tgaagaaaa
109921 ttttaagtga tgaaaatgag catcctaggg ggacaatcag caggaaacca aagatgcaaa
109981 tctgtagtag tctcagaga ctatacgtca agctcatgaa aagaacaaat atctaagaat
110041 tctctcgcga cattatgagt atgcatgaac aatggatcac cggctttact gacggtgaa
110101 ggtgctttca cattggaata agtaaaaaaca acgaaaccaa attaggttct caagttttgc
110161 tagaatttgt gattacacaa catcaagag atgaacaact tttaaatgaa attaaaaaat
110221 attttgattg tggcgtcgtt cgcaaaaagta atggccgtgg agatattctt tgcattcag
110281 tccgttcaca aaaacactta agagatgtca tttctcttt ttttgagaaa aatattctaa
110341 gaaacaaaaa aaaattcgcct tttcaagat ttaaacagc tcttgtttta attgaaaaa
110401 aggaacattt gactttggag ggtttagata aacttcgaaa attaagagat caagtgtatg
110461 attctgatta aagcgtttga atgagtacga gatgggagag agcataggtc tctacaaaac
110521 ttaagatat tgggtcttct tttgggaaga tagagtccaa caaagaacta atgatctttg
110581 atcaaaaacga gcgaaattcc ttgtcgggta aactgaactt gcccggttaa tttggcgaca
110641 aattatagca ctccagctca taacgggtgga atctaagag agtaattctc ctgataata
110701 cctggtggaag atttatttta taaaataaat ccccgtaacg actgatccga aaggtgaggg
110761 ctcgaaaaag aaatccaaca gaaaagatat aatccgaact gtggatttta tttcttggc
110821 caaaacgctg gcaactcagc tctctagaaa agaaaaatg aaatcatgga gccagaatct
110881 aagaatacat ttgattttgc tagttggctt gtaggtttta tcgatgggga aggttgtttt
110941 tctctttctt tttcggaaaag agaaagatta acactgaaaa ttgaggttcg accaaagttt
111001 tcagtttccc aaaattcaaa gtcgatgaaa atcttagaaa caatgaatc atattttggt
111061 tgtggtggtg tttagacatt aaaaagagag aatacgcata agtatgaagt aagaatacta
111121 aatcacctta ctactgttat tattccttt tttagaaaa atcccttct tacaacaaaa
111181 cgaagaagatt ttctgctgtt tgaccaagca tgtatgatta tcaagaaaa tctgcatctt
111241 tcaagaagaag gactgaaaga aattttacia ctccagctc aaatgaacac atcagggaca
111301 cgaagaagatt cgtatcgtgaa cttatataat attcttgatt tagactccta aaatttctga
111361 cgaagtgaaag gtatagtcg aacaatatgg aaacatatg cataactttg acactacaaa
111421 agcgaagctt ttttcaaaa agaaagtaaa tatcgttttt aagtcagaat tgaagttccg
111481 acctgcacga aaggcgtaac gatttgggta ctgtctcggg gagagactcg gtgaaataga
111541 catgtctgtg aagatcgagg ctacttcgac ttggacagaa agacctatg aagctttact
111601 gtaactttgg attgaattg ggttcatctt gcgcagctta ggtggaagc tatgaagcaa
111661 aggttccggc ttttctggag ccaccagtga gataccactc tggatgaact tgaattctca
111721 atgcacaggg taatcccttg cgtgacagtc tcaggcgggc agtttctctg gggcggagtc
111781 ctctcaaaag gtaacggggg tgcataaagg ttcctcagg ctggacggaa atcagtcaaa
111841 gagtgcattg gcagaagggg gcttgactgc gagacctaca agtcgagcag ggaagaaagt
111901 cggctatagt gatccgacgg tgcctcgtgg aaaggccgtc gctcaacgga caaaagttac
111961 tctagggata acaggtcaat ctctccaag agtccatctc gacgaggagg tttggcactt
112021 caacaaacat cggggcttta gaatcgaaaa gattcctaagt atgcatttct ctatattgca
112081 aggacctctg gaagcagctt tctcaccttt tcgaggagtt caaaagtatt taaactttc
112141 tccatagaca aaaagggaa atcagaaaaga gaacctgaca acttgcaggg aagaacgact
112201 gctgtactct tctggacact cggtcfaatca cctcaacga ctttatgaga aacaccact
112261 tccggtgatg acaaagtcta ctcttgggtg aaaaaatcaag gacaacctcg taatgaaaca
112321 aaacatgagt gtttttcttc agaagatgtt ggcatacaat agaataagtg gaaaagaatc
112381 aagaaaaact ttagctgctt ggaactttac tctcccagta tattcaaaat ccaaaagcct
112441 aactacaag agttctcaga ggcgggaagc gcacagctta tcttgagcag tttgctaac
112501 aacctttgtc ctttctcaca aaaagtttta ttactggaca tcttttagga gatgggtgta
112561 ttttgtatac tgatggaata accccgtttt taaaaattga tcaacaagta agccaaaaag
112621 ctctatagata tgtgatata agctgttttg aagaatatgt aggaagttca ccaagtttaa
112681 gaagagatag aatgggtaac cctcattctt actactttag gactttaagg actgaaaaat
112741 tagcaaatca tgcacaaaca ttctatcaac tcaacgcaat ggggagggct caaaaagttg
112801 tcccgcgaaa tatccataaa tggttaaatc ctatttcttt agcgttttgg tttatggagc
112861 atgggggaaac tgaagtatca gggatcgcct tacatactga gaatttcttt ttttccgatg
112921 taaagatttt acaacaagct ttaggttcgg tttttaaatt acaagcfaat gtcggaagag

112981 ataatcgaac aaccggaact ttgtattatc tttctattcc tgcgaattcg aaacaacggt
113041 ttcaggaaat cgttgccctt tacatgggtc caaccatgat gtataaactt cgattgcctg
113101 atactgaaag ttagagttta gcttttcatt tcaattgtat taattcgcga tgcgcggtcg
113161 tcgcatcctg gggcggtagt acgtcccaag ggtctggctg ttcgccagtg aaagcggcac
113221 gtgagctggg ttcagaacgt aaacatcatt gcgtttagtt agggagtaac ctttcttagt
113281 attggcttat atcggtgaac cctaaagagc ataagctcca tggcaacycc gaggggaagc
113341 ttttcagatt atgaccacga agaatacttt acaaccaaca gaagcagctt atattgcagg
113401 ttttctggat ggtgatgggt cgatatatgc aaaaacttatt cctcgtcctg attacaaga
113461 cattaaatat caagtgagtc tagcaatttc ttttatacaa agaaaagata aatttcata
113521 tttacaagat atctatgatc aattaggaaa acgaggcaat ttaagaaaag accggaggta
113581 tgycatcgca gattatacca ttatcgggtc aacacattta tcgattatat tgcagactt
113641 ggttccttat ttaaggatta aaaaaaaca agcaaatcgc attcttcata taattaatct
113701 ttatccgcaa gctcaaaaaa atccttccaa atttttagac ttagttaaaa tagtggacga
113761 tgttcaaaat ctcaaaaaa gggctgatga acttaaatct acaaattatg ataggttggt
113821 ggaggagttt ttaaaagcag gaaaaataga atcgtccccg tagagactgt gaaggaccca
113881 agatcttctg tgcttaaget tgatttttga cctgaaaggt caagaggctt aagcaaatag
113941 aaattctcct tccgatatct ctttgggta aaagagatat aatacgccaa ctccaacaag
114001 caaattggtg gatgaagaca tagtccatgc ccttttga aaatggggac aacatgcctg
114061 agacagttcg gtccatatec agtggaggcg ttagagcatt gagaggattt ctctcttcta
114121 cgagaggacc agaaggaacg caccctgggt gtaccagtta tggaaatttcg gtttcgtaga
114181 cgctgggtag ctatgtgcyg agtcaataag acctgaaagc atataagtt gaagtgcacc
114241 tcaagatgag tgctctcgcg atttatcgct aaggctcgcg caagactagc cgattgagga
114301 atcaagtgca tgctgagatt tccataaaga ccgaatgatt ttgattttga ctgtaaaaca
114361 ctacgagagt gtggcttctg ttaaaaaaga agccacactc tctcactc tggtatctc
114421 ggttcaaaaag aactaccca acctttccg aatttggtgg taaaagttg aacggctaaa
114481 atacttcaag gggagccttg tggaaagata cgctgacgcc agagaatt

ANNEXE 2

Séquence complète du génome mitochondrial de *Monomastix* OKE-1 en format GenBank

LOCUS MSPMTCG 60883 bp DNA circular 13-NOV-2000
 DEFINITION Monomastix OKE-1 mitochondrial DNA, complete sequence.
 ACCESSION MSPMTCG
 VERSION
 KEYWORDS
 SOURCE Monomastix OKE-1.
 ORGANISM Mitochondrion Monomastix OKE-1
 Unclassified.
 REFERENCE 1 (bases 1 to 60883)
 AUTHORS Gagnon,M.-C., Otis,C., Lemieux,C. and Turmel,M.
 TITLE Complete sequence of Monomastix OKE-1 mitochondrial DNA
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 60883)
 AUTHORS Gagnon,M.-C., Otis,C., Lemieux,C. and Turmel,M.
 TITLE Direct Submission
 JOURNAL Submitted (13-NOV-2000) Biochimie et microbiologie, Université
 Laval, Pavillon Charles-Eugene Marchand, Quebec, Quebec G1K 7P4,
 Canada

FEATURES Location/Qualifiers
 source 1..60883
 /organism="Monomastix OKE-1"
 /organelle="mitochondrion"
 /cell_line="Personal collection of Charley O'Kelly"
 exon 38..1309
 /gene="rnl"
 /number=1
 gene 38..6107
 /gene="rnl"
 rRNA join(38..1809,2191..2198,3061..3638,4842..4934,5812..6107)
 /gene="rnl"
 /product="large subunit ribosomal RNA"
 intron 1810..2190
 /gene="rnl"
 /number=1
 exon 2191..2198
 /gene="rnl"
 /number=2
 intron 2199..3060
 /gene="rnl"
 /number=2
 gene 2411..3019
 /gene="orf202a"
 CDS 2411..3019
 /gene="orf202a"
 /codon_start=1
 /product="putative site-specific DNA endonuclease"
 /translation="MKLTSDWICGFVEGEGCFCISISKIRSVQSKTPENLPGGQPTLL
 TEMTSKNQKSSFPEQSLEKKESVSKQVRLVFKVTQGVKNIQVLYALKKFFGVGHVKSQ
 QASGKVWEYTVSRFEHLHTKIIPFFEKNTLYTSKQDFYRFRKVALYMSRKEHLTFEG
 LAKIEKLTARMNTPECLEEVYEESFDEDKVRTFSKEKEEDTD"
 exon 3061..3638
 /gene="rnl"
 /number=3
 intron 3639..4841
 /gene="rnl"
 /number=3
 gene 3981..4796
 /gene="orf271"
 CDS 3981..4796
 /gene="orf271"
 /codon_start=1
 /product="putative site-specific DNA endonuclease"
 /translation="MVDRQHTLNQTALEYLFKFMQIPYIPETRDLRGSALISYLKQHL

TDLQPSV FVGNLLGDGTLCKQIGNNGSSNFKI DQAASEKGI EYVHFIY TILAPWVGT P
 PKLRYKNNKAHSIWFRTRFMKELDFYCKQFYSIDANGQRQKI P SLLYKWLDDVALAI
 WFMNDGGKTDYGYRIHTQCFTVFQIRQLQKILQHNFGLTTGIGRDRKSLIGT IPEVIK
 QERKLF SRLDDPTCKTYYYLEIHAPHREQFTSLVEPFILPGFRYKLSSTNKD"
 exon 4842..4934
 /gene="rnl"
 /number=4
 intron 4935..5811
 /gene="rnl"
 /number=4
 gene 5126..5650
 /gene="orf174"
 CDS 5126..5650
 /gene="orf174"
 /codon_start=1
 /product="putative site-specific DNA endonuclease"
 /translation="MNANTKSTLNPTDAA YLAGLLDGDGSI FVQIVPKPNYNCFFGLQ
 FTINFTQSTKRKHFM LAIEKLGSKASFRDRKDNCEIAIYGWQSVAVFLHEVHSYLK
 FKKPQADLVLKIVEKLPSTTTSYKNFLEV CALVDKLT EYNDSSKKRKYTSKYIVEYYKE
 RGIVIDEIAPVETF"
 exon 5812..6107
 /gene="rnl"
 /number=5
 gene 6332..7489
 /gene="cob"
 CDS 6332..7489
 /gene="cob"
 /codon_start=1
 /product="apocytochrome b"
 /translation="MKTRLSILKQLVTLVNDHLIEYPTPTNLSYFWGFGSMAGICLL
 LQILTG VFLAMHYTPHVDLAF LSV EHM RDVEGGWFLRYMHANGASMFFI VVYMHMER
 GLYYGSYASPREFVWIVGVILFLLMIITAFIGYVLPWGQMSFWGATVITSLASAIPIV
 GDELTHWLWGGFSDNATLNRFFSLHYLLPFLLVGISAVHIIALHQYGSNNPLGSLAS
 VDKISFYYPYFVKDLVGVAVAFALFFSFFVYFAPNLLGHPDNYIPANPMSTPAHIVPEW
 YFLPVYAILRSIPNKLGGVAAIGLVFVSL LALPFLNTSEIRSSSFRPIHKILFWFLGA
 DCVLLGWIGCQVPEAFVVTIGQIASIFFFFYFLVCVPLLGRIEKAFITYKA"
 gene 7834..8205
 /gene="rps12"
 CDS 7834..8205
 /gene="rps12"
 /codon_start=1
 /product="ribosomal protein S12"
 /translation="MPTFQQILSTRRQLV KRKSKAPALHKAPQKRGVCIRVYTKT PPK
 PNSALRKKVAVRLTNSIEVIAYIPGEGHNIQEHSVVMIKGGRIKDLPGVKYRILRGLV
 DLQGVKNRRKRSRSLYGSKKQI"
 gene 8212..8859
 /gene="rps7"
 CDS 8212..8859
 /gene="rps7"
 /codon_start=1
 /product="ribosomal protein S7"
 /translation="MEIKNQSKDGARKSVEPKFSLTPSGKGNSTTQKSNSSILSVEN
 QRIPKIFDPKNSHAQELRIQNK CINLLTKDGKKAKAFSLLKTAFTLVMKRRKTQSDGG
 KPSSIHSLTANMLFLQSIENVKPIFEVRKVRVAGNTLLVPALISFERQENKALRWIIE
 AAQERKRKNPHFLFECLAIEILEAVQHGGYAKQKRDELHKIAEANRAFAHYKWW"
 gene 9429..9749
 /gene="orf106"
 CDS 9429..9749
 /gene="orf106"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MKYEDTFLLMFFLSINKMQMTQVNVILTMSNHCMSPHHFKRKWD
 QGLISGPGRAPKICMSNFIVDLGEKQQQI ALSMHFGAFNAEHQLKCYFYFKYQSRAL
 LNTI"
 gene 9933..10265
 /gene="orf110"
 CDS 9933..10265
 /gene="orf110"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MPKTSWFLDRYWKQLQALIRSPRKAQNHTFVYQKTLDSYVTSTF

PLMPKRIYKILEDENLLIKFIRAYIDHKKPYLNKTMEKAVQEI KKHPPQEVEALRAYLQK
 HIDQNNKE"
 gene complement(10372..10650)
 /gene="orf92"
 CDS complement(10372..10650)
 /gene="orf92"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MRSAYNTKESLSENNFCPSDKNNPQIMTLCSSPENNFVSKCT
 PGCQARFWEVGRSYKVAVFLREGRFYFTQLSQSAGPSSSLGRIAPFHV"
 gene 11140..11355
 /gene="orf71"
 CDS 11140..11355
 /gene="orf71"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MNVLTPRCSVLFELKTHESLGAARGPFLRCGCVRHPILCAAAGR
 ALLSSRPLYAAGPLACSPSLGEAPHLE"
 gene 12227..13015
 /gene="nad6"
 CDS 12227..13015
 /gene="nad6"
 /codon_start=1
 /product="NADH dehydrogenase subunit 6"
 /translation="MIESQALQALFYLFSSIALISGVMVITAKNPVHVSVFLILVFCN
 AAGLLILLNLDFFAMI FLIVYVGAIAVLFVVMMLNIKLAEINESILRYLPVGGFIG
 LLFVFEILLVIDNDLIPVLVFPESNNHVFSGLGVLEVFFYKLGQVLSQSPGEYRKD
 LESLLSNSQAYGLPSTSI STYTEWALKTQSKTNI EALGQII YTYSSLV FILASLILLV
 AMIGAVLTMHKGVHVKRQEVFEQNRDRHSQT VHKIGKLSNAFK"
 gene complement(13951..14256)
 /gene="orf101"
 CDS complement(13951..14256)
 /gene="orf101"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MKISLFFFLYRNTSEIQTSPCRSSVPLELHPFRHTSKSITLEKR
 SRHTSTSGRAMVRRVKLQEKQQAARSILRKFFGRESSFLQIRGQQTACQRYGAES"
 gene 14277..14495
 /gene="orf72"
 CDS 14277..14495
 /gene="orf72"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MGYEHFRILMAAFILCQAFDPKRSRRHSTKSLAHRGVLTLLHGS
 TRSGSSFNRTL PYASREVAEDPLSCSLL"
 exon 14549..14944
 /gene="cox1"
 /number=1
 gene 14549..21532
 /gene="cox1"
 CDS join(14549..14944,16291..16518,17846..18097,19438..19677,
 21023..21532)
 /gene="cox1"
 /codon_start=1
 /product="cytochrome c oxidase subunit 1"
 /translation="MTNFANRWLFSTNHKDIGTLYLIFGAFSGVLGTCFSMLIRMELA
 SPGNQILAGNHQLYNVITAHAFMIFFMVMPALIGGFGNWFVPILIGAPDMAFPRLN
 NISFWLLPPSLLLLSSALVEVGVGTWTVYPPLSNITSHSGGSVDLAI FSLHLSGIS
 SILGAINFITTVLNM RPGMTMHRPLFVWAVFITAILLLSLPVLGGITMLLTDNRN
 LNTSFFDPAGGDPILFQHLFWFFGHPEVYILILPAFGIISQVISTFSRKPVFGYLG
 VYAMLSIGILGFIVWAHMYVVGLDVDTRAYFTAATMIAVPTGIKIFSWIATMWGGS
 IVLKT PMLFAVGFLFTVGGVTGVILSNSGLDIALHDTYYVVAHFHYVLSMGAVFGM
 FAGFYYWIGKITGLQYPETLGQIHFWTFFLGVNITFFPMHFLGLAGMPRRIPDY PDAF
 ADWNAIASYGSYLSI ISSLFFYVYKTLTGNEVCPNPPWAFDGNSTTLEWMLPS
 PPGFHTFEEI PAIKELDSTALKATEIAGKSAKK"
 intron 14945..16290
 /gene="cox1"
 /number=1
 gene 15122..16183
 /gene="orf353"
 CDS 15122..16183

```

/gene="orf353"
/codon_start=1
/product="putative site-specific DNA endonuclease"
/translation="MCITKKQSAWADGSTIFRRKEQWFTSSSHQRLNVKHPFTHHINL
SHWLVGFTDGDGCFMNETKDGWVQFTFQISASLYNLRVLYFIKSVLQVGSVSYAGER
MAQYRVRDRRYLKNSILPIFDGSSLSVSKAYDYEAQGLFIYEDPTLSYAEKSERLR
ALRAAALENKKKASPWGRGERPTKSWVLGFTAEAGSFYITKKEKERYCHGFGLTQKT
DEIILLFFREIFQIRAQVKWNSAGSTFWSLDTTNFRSIQNI EAFFKNTFKGVKSLFR
IWSRALKYRGQCETLKHTQDFLRRLRKS PFIGDEDRRS GGRSLCERPQDRRSAPMSVA
HGGLECCSAPHEHC SWGRE"
exon      16291..16518
/gene="cox1"
/number=2
intron    16519..17845
/gene="cox1"
/number=2
gene      16781..17602
/gene="orf273"
CDS       16781..17602
/gene="orf273"
/codon_start=1
/product="putative site-specific DNA endonuclease"
/translation="MSTSQFPFYLAGLIEGDGHFYS PARHRTNKGNTLYPSIQICFAL
KDLPLAQMVLAHLGHGSLCRTSGKQAYILTVIDREGCVLVHMINGRLRTSKCHQFSR
FLTCLNERGYGNFVVGPKDTS PVFSNAWLAGFIEAAGCFEIRCTLT PPRRISPSFVIE
QAYSHEEDRPLFTTISKSF LAPEPRI IQRKKGGEFWRI RTTSLAGNLCLLQYLKECSL
AGAKYMDFRDWCQVEMVREK KHLMDASFPIVLA IKKDFNDARTTFDWT HLKTF"
exon      17846..18097
/gene="cox1"
/number=3
intron    18098..19437
/gene="cox1"
/number=3
gene      18332..19246
/gene="orf304"
CDS       18332..19246
/gene="orf304"
/codon_start=1
/product="putative site-specific DNA endonuclease"
/translation="MKDCGSSETIRKTTFRFQNFNHLPEHIKTCSPFLEWFIGFCE
GDSCFSSSKPSSLEFIINQKEKRILQNI RTTLGFGRVTTTPYVKGKGNLYRFSVSKR
SDIDRIVHLFNGNIVLNKINDNFLLFLQNRNAQACKDGRPP IHFLGKHEFLFFENNAW
LSGFTDAEGCFYVGKSKDKRYS LGYRLRCRYILDQLAERWILEKIKNEFVKNGCISIR
KTKDKDGNVIRRVGQEMLR YCTHLKCLEFLMNYFKRYPL LSEKSIALHRFQKMMRY
IQNRKTVEWEGKVLKKVENLLNKLKEED"
exon      19438..19677
/gene="cox1"
/number=4
intron    19678..21022
/gene="cox1"
/number=4
gene      19688..20641
/gene="orf317"
CDS       19688..20641
/gene="orf317"
/codon_start=1
/product="putative site-specific DNA endonuclease"
/translation="MLCMSNSNIIVHKSSHGPQS FQDQLESFFLGLLEGDGSIQVNHW
RKASLQYRI I IKLNTEENYRMLTFIHRKLNIGNVKITAHGTRVSLVEDHRDRIPKIM
GII EKHLRLTKARIRYGFQYAFQNS IGYSEYLYLRTFRDMWEKAA YAADFHGRAPA
QHSDFSGAPHPLDFFKMLTPYKAQDILQRPHFVHWLCGI IEAEGCFSRKSGKSCSF
SLAQKYEEAFLLAVKQYFQLPNRVRVNSHMYMLET FNVTHVQNI LEFCRPGSLNALH
RKAYA EKRVDSFGLLGSKETQRQEF EKRFLLSPRLESEN"
exon      21023..21532
/gene="cox1"
/number=5
tRNA      21809..21882
/gene="trnMe (cau)"
/note="codon recognized: AUG"
/product="tRNA-Met"
/anticodon=(pos:21843..21845, aa:Met)

```

gene 21809..21882
 /gene="trnMe(cau)"
 tRNA 21961..22035
 /gene="trnD(guc)"
 /note="codons recognized: GAY"
 /product="tRNA-Asp"
 /anticodon=(pos:21996..21998,aa:Asp)
 gene 21961..22035
 /gene="trnD(guc)"
 gene complement(22202..22936)
 /gene="atp6"
 CDS complement(22202..22936)
 /gene="atp6"
 /codon_start=1
 /product="ATP synthase F0 subunit 6"
 /translation="MFSPLEQFTIVPLIPLGYFSFTNSAFFMLLTGGLVSLFYLVT
 NGGFLVPSRWQSFVEMIIYEFVLNLVVEEQIGSAGKKYFPFIFTTFTLLFSNLIGMIPY
 SFTTTSHFIVTFTLSISIFIGVTLIGFQKHGIFHSFFLPPGAPLALAPFLVIVVVS
 YSFRGISLGVRLFANLMSGHVLLKVLGFAWTMLCSGGVLLVASAIPVVVVYLLFFLE
 VGVAALQSYVFAILTCIYINDAIQLH"
 tRNA 23350..23423
 /gene="trnR(acg)"
 /note="codons recognized: CGN; 'super Wobble' codon
 recognition inferred, with A in the first position of the
 anticodon assumed to be post-transcriptionally modified to
 I (inosine)"
 /product="tRNA-Arg"
 /anticodon=(pos:23384..23386,aa:Arg)
 gene 23350..23423
 /gene="trnR(acg)"
 gene complement(23434..24171)
 /gene="rps4"
 CDS complement(23434..24171)
 /gene="rps4"
 /codon_start=1
 /product="ribosomal protein S4"
 /translation="MRHTLRASYKDIHMTQKIKFKRYRGLVESFGSPQYFPRKITHKQ
 GLLLQKLGSRKKLSYFGQNLQSKRLFSLLYGNLRRKHVFFLKKASQLSGSLSANF
 LSLLEERRLDIVVYRMFQFRTLESARQYILHGGILLNGLPVNLSLSLQIQPGDLIQIHNA
 QDFFSSKLSLLRNGATVPQKTEKQEMQSNAFKSSLLKKTALRPRRIKYSHFIEINFKLLV
 GIFLFSRQLYYPFQLKKEDFIPKNMK"
 tRNA complement(24235..24305)
 /gene="trnG(ucc)"
 /note="codons recognized: GGN; 'super Wobble' codon
 recognition inferred, with U in the first position of the
 anticodon assumed to be unmodified"
 /product="tRNA-Gly"
 /anticodon=(pos:24271..24273,aa:Gly)
 gene complement(24235..24305)
 /gene="trnG(ucc)"
 tRNA 24359..24439
 /gene="trnL(uaa)"
 /note="codons recognized: UUR"
 /product="tRNA-Leu"
 /anticodon=(pos:24392..24394,aa:Leu)
 gene 24359..24439
 /gene="trnL(uaa)"
 misc_RNA 24519..25047
 /gene="rnpB"
 /product="ribonuclease P RNA"
 gene 24519..25047
 /gene="rnpB"
 tRNA 25076..25149
 /gene="trnR(ucu)"
 /note="codons recognized: AGR"
 /product="tRNA-Arg"
 /anticodon=(pos:25110..25112,aa:Arg)
 gene 25076..25149
 /gene="trnR(ucu)"
 tRNA 25223..25296
 /gene="trnP(ugg)"

```

/note="codons recognized: CCN; 'super Wobble' codon
recognition inferred, with U in the first position of the
anticodon assumed to be unmodified"
/product="trnA-Pro"
/anticodon=(pos:25257..25259,aa:Pro)
gene 25223..25296
/gene="trnP(ugg)"
gene 25586..25795
/gene="orf69a"
CDS 25586..25795
/gene="orf69a"
/codon_start=1
/product="hypothetical protein"
/translation="MLGSRCKRSGTCSTSDAFGPSALRPYGPSIVGVRACSHPDTFSP
PIISFFVRKSVVSHTKSRRLFGQKM"
gene 25848..26162
/gene="nad4L"
CDS 25848..26162
/gene="nad4L"
/codon_start=1
/product="NADH dehydrogenase subunit 4L"
/translation="MFSSPDQVQFLAVSILLFLLGIWGIFLNRKNIIMLMSIELMLL
AVNLFVFLFSVYLDLVLGQMFALFILTVAAAESAIGLALLVVYRVRGSIAVEFINLI
KG"
gene 26402..26959
/gene="atp8"
CDS 26402..26959
/gene="atp8"
/codon_start=1
/product="ATP synthase FO subunit 8"
/translation="MPQLDQLTFFSQFFWLCFFYLGFYLVLVKYFLPKMSKILKVRNM
KMNSGNAGQTDFLMKEENEKVKRMRDEALSEAFKESKHFLNESYQRTSSVWSKILQD
INKKLQKMNKTYIHSVQNLSFAQTTLTLHNLKTVIAPSSYKASGSIASLSTSESVKTSK
DIFFNTVLLKSLTRKKAGLVKPTKK"
gene 27153..27770
/gene="ymf39"
CDS 27153..27770
/gene="ymf39"
/codon_start=1
/product="unidentified protein"
/translation="MIHMNPTFLQNLVSSLQKAFQPDNRLFVRLGILAFLLTSSKHIF
IYNEETLILICFIAFIFTCQKMLGESITQSLNERSQLIAQELQNFYNLKEQLVLELIE
EHQKQLSIHSFIKKIGQFSLTEVKHVSLLQKKALQSLLDQIEHKLNTLMSFSNVSNQ
LQEAISTSFVAVLEEFYKSKKNFQSVLVKQAVVNLERLANQKK"
rRNA 27976..29506
/gene="rns"
/product="small subunit ribosomal RNA"
gene 27976..29506
/gene="rns"
rRNA 29520..29629
/gene="rrn5"
/product="5S ribosomal RNA"
gene 29520..29629
/gene="rrn5"
gene 29931..30284
/gene="nad3"
CDS 29931..30284
/gene="nad3"
/codon_start=1
/product="NADH dehydrogenase subunit 3"
/translation="MFEYFPILIYLGISLILSLVILGSSFLFATRNPDMEKISAYECG
FDPFDDARGRFDIRFYLVAILFIIFDLEVTFLFPWAVSLSHIDFFGFWSMMIFLFILT
VGFFYEWKKGALDWE"
gene complement(30665..31663)
/gene="nad1"
CDS complement(30665..31663)
/gene="nad1"
/codon_start=1
/product="NADH dehydrogenase subunit 1"
/translation="MILALVCSIAQILGIIVPLLVAVAFLTLAERKVMASMQRRKGPV
VVGIFGVQPFADGLRLLVKEPVLPSNANSVIFLLAPVLTFTLSLISWSVIPFDSGVV

```



```

NNRINWQRLVDIGKVTAEQAMDWGFVGVMRLGSGVSWDLRKNMPYDAYDSDFEIPVG
TQGCYDRYLIRMEEMRQSLRIMLQALNTMPSGMIKTDDRKITPPSRGQMKESMESLI
HHFKLYTEGFSVSASETYTAVEAPKGEFGVYLVSNGTNKPYPYRCKIRAPGFVHLQGLDM
MSKHHMLADVVTIIGTQDIVFGEVDR"
trRNA 35078..35166
      /gene="trnS(gcu)"
      /note="codons recognized: AGY"
      /product="tRNA-Ser"
      /anticodon=(pos:35147..35149,aa:Ser)
gene 35078..35166
      /gene="trnS(gcu)"
trRNA 35197..35269
      /gene="trnT(ggu)"
      /note="codons recognized: ACY"
      /product="tRNA-Thr"
      /anticodon=(pos:35230..35232,aa:Thr)
gene 35197..35269
      /gene="trnT(ggu)"
trRNA complement(35436..35508)
      /gene="trnA(ugc)"
      /note="codons recognized: GCN; 'super Wobble' codon
      recognition inferred, with U in the first position of the
      anticodon assumed to be unmodified"
      /product="tRNA-Ala"
      /anticodon=(pos:35473..35475,aa:Ala)
gene complement(35436..35508)
      /gene="trnA(ugc)"
gene 35798..36205
      /gene="rps10"
CDS 35798..36205
      /gene="rps10"
      /codon_start=1
      /product="ribosomal protein S10"
      /translation="MTALSHKIQISLKAFEYCFVEKALHQLNMTISHASSHIESKTVS
      SSKPSAFSKYTLPLKVQRFVIRSPHIDKSRDQFEIRQYKVIILNHSADWSSDDLRLF
      LDNLKRNARFYGVQMQIKLCTKTFGPAKKKDRTY"
trRNA 36462..36543
      /gene="trnL(gag)"
      /note="codons recognized: CUY"
      /product="tRNA-Leu"
      /anticodon=(pos:36497..36499,aa:Leu)
gene 36462..36543
      /gene="trnL(gag)"
gene 36561..36947
      /gene="rps19"
CDS 36561..36947
      /gene="rps19"
      /codon_start=1
      /product="ribosomal protein S19"
      /translation="MTRSLWKGPFIDTFFIRSLRKVSKNLSAVGLEAKPQRVVTGSKT
      SVNASTLTKKSSEKITGHFQKMWSRRSVILPQFLDSQYEIYNGKQFIKLRVTEDMVGH
      KFGFALTRKKPKHPTLKKNLITKKK"
gene 36949..37896
      /gene="rps3"
CDS 36949..37896
      /gene="rps3"
      /codon_start=1
      /product="ribosomal protein S3"
      /translation="MGQKANPISLRLSFNRDFESCWYQDLNYSKILLEELTLRKYILS
      LFSSMNVYQGRVLFQIFPKKCVIYSFITNKKTQLNSPLKLAGGTRFRKTSLNSSAPR
      GTLRILPKNATKQDGIHLFMEYALQRFDSVNHLPASIASIAGFLYLISKSMQTSERKMV
      ALYGGSDKQKGRHAKSVPKLPLSLSSFPHVVASPLEAFTGSKVDVFPVKLHSRYQSAS
      FLSQYIAQSIKSMFRQIFRIIQKDI LRDSKILGMKIQCSGRINGAEIAKVESKKMG
      QTSLHTFSARVDFGTAEAYTMYGIIIGIKVWSSISEENL"
gene 38090..38506
      /gene="rp116"
CDS 38090..38506
      /gene="rp116"
      /codon_start=1
      /product="ribosomal protein L16"
      /translation="MLQPKRTKFRKFHKGKATGIKQNTSSLHFPTYGIKVVEAGRLSS

```

QVIEARRALTRKLRNRQIWIIRIFPDIAVSKKPAEVRMGKKGGAPEYVVARVQAGQI
 LFEIAGVSAELAKQAGLLASRKIALGTVFIKSEDAL"
 gene 38734..39114
 /gene="rpl14"
 CDS 38734..39114
 /gene="rpl14"
 /codon_start=1
 /product="ribosomal protein L14"
 /translation="MILVGTQLNVSDNSGAKIIECIKILGKKSSSFASIGDIIVVSVK
 ETTGKSKQTRKVEKGGIYKALVIETKKGIRKRDGSILRFHRNSALLLSEQGNPIGTRI
 LGNGTYELRKKNQTKVLSLALRVF"
 gene 39260..39865
 /gene="rpl5"
 CDS 39260..39865
 /gene="rpl5"
 /codon_start=1
 /product="ribosomal protein L5"
 /translation="MNHSRLEWYSEHIMKRDLLYKLCIPFAGNIYTLPKIQHLSLNV
 SSTTSGSQATPNTSSRQKMITLFSGLEILSGQKVKKTYAKKSIASFKLKRGQCIGGKV
 TLRNKEYMFLEKFFIVLPKIREFVCLQPNTHDSTHTLSSINFSGGSFLLYPELSNQ
 YELFESNKGFNATLIMNTNHCSSLHTVSGHSVGNLISFFVP"
 gene 39929..40228
 /gene="rps14"
 CDS 39929..40228
 /gene="rps14"
 /codon_start=1
 /product="ribosomal protein S14"
 /translation="MSNLVFRDKRRNLFKKYELQRVLYKSIHDLSEIEKKYRLESIL
 KINSLPRNSSTVRLKNRCILTGRGKAVLSTFKLSRISFRELARKGMLPGVSKASW"
 gene 40477..40857
 /gene="rps8"
 CDS 40477..40857
 /gene="rps8"
 /codon_start=1
 /product="ribosomal protein S8"
 /translation="MDNIANILSSIENAQRQKLCCLFPFSKRTQHILHILFIEGYIR
 GFKINQGFQVYLKYSIEDREEGSSLFKIKRISKPGKKIYIPAKKINKIKRGLGILILST
 SKGVLSDRDARYLQLGGEVVAAYV"
 gene 41398..41601
 /gene="orf67"
 CDS 41398..41601
 /gene="orf67"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MANQKAGLSTGGRSTWETRAAFPIGKRRNARVSES�KTQKERT
 ASPSSRGVLRPSRMMRKYHVRF"
 gene 41793..42506
 /gene="rpl6"
 CDS 41793..42506
 /gene="rpl6"
 /codon_start=1
 /product="ribosomal protein L6"
 /translation="MQKNTSVHFLNQSGCFHSHSIKIPSDIKLEIVEEGLPSSHHTHS
 VLKCTGPLGCVHVNLQKIDRYGLCFEIEKQSSQGVLESFLHVSIKSFDNFQKKS
 KKSTHTIAQQLAFSNVLTSMCQQMIEGVSKGFVLYLELHGVGFRASVHEKTLAGKTV
 KYLEFKLQSHDIFFEIPDNIQIFSVKPTLIGLYGIDREGITQLAAKMRHLKVPDSYK
 GKGIRYKDESVRTKIGKKK"
 gene 42509..42823
 /gene="rps13"
 CDS 42509..42823
 /gene="rps13"
 /codon_start=1
 /product="ribosomal protein S13"
 /translation="MVYILNKNLPNKKLRYALQEIFGIGPFVAQQISDQLGLSSHTL
 MEELTQSQIDTLVRIINNSITGSDLKRVVQQDIKRLSTIGSYRGIRHTQHLPVRGQR
 TH"
 gene 42860..43171
 /gene="orf103"
 CDS 42860..43171
 /gene="orf103"
 /codon_start=1

gene 46250..47044
 /gene="cox3"
 CDS 46250..47044
 /gene="cox3"
 /codon_start=1
 /product="cytochrome c oxidase subunit 3"
 /translation="MSIQKHPFHLVDPSPWPIMASLASFSSTTVGASMYMHGYTGKEL
 LLFGLTMVMYSMFVWRDVRVREGSFQGHHTSFVQLGLRSGMILFIVSEVMFFVAFFWA
 FFHSSLAPTVEIGAVMPPKGDVLPWEIPFLNTVILLSSGAAVTWAHHAIVAGYRQQ
 AIYGLIVTILLACVFTGLQGLEIYI EAPPTISDGIYGSTFYLATGFHGFHVFIGTVFLT
 ICLVRLINFHFTQSHHFGFEAAAWYWHFVDVWVLFLEVAIYWGGN"
 gene complement(47385..49022)
 /gene="nad2"
 CDS complement(47385..49022)
 /gene="nad2"
 /codon_start=1
 /product="NADH dehydrogenase subunit 2"
 /translation="MKTQDYLFHQVNSAEDAFATSFKNDLALFPEIFLIMCTLFL
 LMYGVHYSTVENKVENNFRPILTENVGYSLLTLGCTFILLFNNPLNQRIFFGSLI
 LDDFTLFFKGLLLSCLFVVMISDFDLKESLNAFEYHILILLSTCSMLFMISYDFL
 SLYLALFQSLCFVYLAACKRNSEFSTEAGLKYFLGAFSSGILLFGISLIYGFGTIT
 NFEELTLLCASVDINHPAGILLGMLFIAVGLLFLKLTAAFPHMADPVYEGAPTSVTAY
 FSVAPKIALLALFIRLFYAGPFGALGTGIENGSALEQILVLCIASMFIGAFSALS
 QQIKRLLFAYSSIGHIGYMFGLIAANREGLESVIVYLIYIMMTLNVFALLSLRAK
 STSLISENNENIKNESTQLRYISDLSLTKNPMFLAFTLTITLFSMAGIPPLAGFCG
 KFYVFSALSSFFILAFVGI FTSVISCFYIIRLIKIMYFEKESTTGYQPINSFEPID
 KTKSIVLGITFFFTLFFLYPAPLFLVAHKVALSLLF"
 gene complement(49120..49407)
 /gene="orf95"
 CDS complement(49120..49407)
 /gene="orf95"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MSNAHQDRSYKIVELNGKGLCCVLDHNLKGSYNTQKFCR
 SWSNIQSSKSRGSKISCPKHKFLGQGSKDHKAKGPKGRMHQKWNLFHF"
 gene complement(49521..50999)
 /gene="nad4"
 CDS complement(49521..50999)
 /gene="nad4"
 /codon_start=1
 /product="NADH dehydrogenase subunit 4"
 /translation="MPSLLMSILFLPLVGVFLFLFIPTRQTVIRITGLLFSLLNFIL
 SLFLWICFDNSSSRFQFVESLSLRTAFSGGFPNIDFVAGIDGSLFFVLLTTFVLPIC
 LLVGWSSVQNYVKEYFIAFLVLETCLMGVFCILDLLLFYVFFESVLI PMFIIIGVWGS
 RERKIRAAAFQFFLYTLIGSVLMLLALLFIYFQTGTTDVMILLTDFSEKRQILLWIAF
 FASFAVKVPMIPVHILWLEAHVEAPTAGSVILAGILLKLGTYGFLRFSIPMPMASIY
 FSPLIYTLVSLGIIYASLTLRQIDLKIIAYSSVAHMGYVTLGLFSLNTQGIEGSI
 LMLSHGFVASALFLCVGLYDRHKTRLVKYGGCVHVMPLFSFFLFTLGNISLPGT
 SSFIGELLILAGIFQNTFFIATLACSGMVLGAAVSLWL YNRVVFNGFKTHSIHFFADL
 NRREFFIFLPFGIAVLWMMGIYPEIFLDTMHTSVAQTLASFQR"
 gene complement(51111..51329)
 /gene="orf85"
 CDS complement(51111..51329)
 /gene="orf85"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MFHFFCIRSGTCSTSDAFGPSILDHRGASMLAPRYLTATSARG
 DQIFDPAHELCSWAPELPSFGPLIIFR"
 gene complement(51456..53483)
 /gene="nad5"
 CDS complement(51456..53483)
 /gene="nad5"
 /codon_start=1
 /product="NADH dehydrogenase subunit 5"
 /translation="MYLLIVSLPLFGFLSASLFGFRFLGYRGSALITTCVSLTACLSA
 VAFYEVGLCGSPCYIEFAPWIQSEMFDAWGFDFDSLTVVMLVVVTFVSSLVHLYSL
 YMSDPHLPRFMSYLSIFTFFMLMLVTADNFIQLFLGWEGVGLASYLLINFWFTRLQA
 NKSAIKAMIMNRIGDFGLALGILGIFSLFKSVNFATVFATAHQFADTTFSECTFEFDT
 LTLCLFLFIGSIGSAQLGLHTWLPDAMEGPTPVSAIHAATMVTAGVFLLRCSAL
 FEHAPLALIGVTFIGAMTSFFAATTGVVQNDLKRVIAYSTCSQLGYMVFACGISQYSV
 GVFHLMNHAFFKALLFLSAGSVIHALSDEQDMRKMGGRLKLLPFTYSMMVIGSLALIG

FPFLTGFYSKDVILEVTYAKYTIHGNFAYWLGSLSALFSTSYYSFRLLYLFTIAPTNAF
 KQSIKXVHDAPILMAIPLILLAIGSIFVGYVAKDMMIGFGTFFWGNALFQFPENALLI
 ESEYIQQFIKMIPLIFSLGAMIAYQINIAPHLAYSMTSSIGMALYTFLNKRWLF
 KVTYDFIGYVPLKMGYDITFKTIDKGVLELIGPSGVIQLLGSKHPDSFLSSVKNLQSG
 AIYHYAVVMLIGLTFFFIAIVGLWDLVLSFWIESRIFFLHLLTFTFLCTQKNI"
 tRNA complement (53721..53794)
 /gene="trnF(gaa)"
 /note="codons recognized: UUY"
 /product="tRNA-Phe"
 /anticodon=(pos:53758..53760,aa:Phe)
 gene complement (53721..53794)
 /gene="trnF(gaa)"
 gene 54100..54483
 /gene="orf127"
 CDS 54100..54483
 /gene="orf127"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MTKRPKLKKVELFGIFQRKSLYLKGFKIFSMNDNKITFYSNTE
 FTSYIFHDVSYEYIGHREWNLFHSRYLTATSIARGEQIFDPAPIRIPYGGRTSFQPPM
 SNAHGGLIFDPGGAHRVSAHQIFEP"
 gene 54503..55450
 /gene="rps2"
 CDS 54503..55450
 /gene="rps2"
 /codon_start=1
 /product="ribosomal protein S2"
 /translation="MNLILKELNNTKAHVGHKKNLETSPYLLGRRASHPLNKTSLIP
 QNKDQKSGVVISEKARQHSEKPKTTGHSMEIEKKLRITFRKKRALFNI FNLEMTLFCL
 SQVIDFLQNLITLKTSLKTSKTTLSSSPRRPHILIMAGRGKTPSGHFHLDIAKILSKW
 NTLGNPQAQPSKTSLQKPILSFVQEKWVGMLTNWKQVSESLKIYSKFKYKFEFDFLE
 KHKLRFPLYEKYNKRYLGLKEISTSLPDVIIITHPEDNDILIQEAFILKPIPIAFVNS
 DLPKNLLHSIQYPIPGNNASPYFIYFCLNMLLMALSTKK"
 tRNA complement (55459..55530)
 /gene="trnQ(uug)"
 /note="codons recognized: CAR"
 /product="tRNA-Gln"
 /anticodon=(pos:55496..55498,aa:Gln)
 gene complement (55459..55530)
 /gene="trnQ(uug)"
 tRNA complement (55591..55661)
 /gene="trnC(gca)"
 /note="codons recognized: UGY"
 /product="tRNA-Cys"
 /anticodon=(pos:55627..55629,aa:Cys)
 gene complement (55591..55661)
 /gene="trnC(gca)"
 tRNA complement (55846..55918)
 /gene="trnH(gug)"
 /note="codons recognized: CAY"
 /product="tRNA-His"
 /anticodon=(pos:55882..55884,aa:His)
 gene complement (55846..55918)
 /gene="trnH(gug)"
 tRNA 55976..56057
 /gene="trnY(gua)"
 /note="codons recognized: UUY"
 /product="tRNA-Tyr"
 /anticodon=(pos:56010..56012,aa:Tyr)
 gene 55976..56057
 /gene="trnY(gua)"
 gene complement (56108..56716)
 /gene="orf202"
 CDS complement (56108..56716)
 /gene="orf202"
 /codon_start=1
 /product="hypothetical protein"
 /translation="MSLSPPQFFRKHFLQKKLMQKVQLYPYTYIFQYNNISANDWKHIK
 NTIHLHHGSTMNIQIIPSKLRKSLHDSVQDNRTQSQSALFFDMKGRCCFLNCTSPAAL
 EVFYQIIHHTVPTKELKTQSVHAPLHVNSFFHVALDSCENNHSTLLNVFDIKKRLGL
 SETSVYAHFFQTVHLPVLNLTSLPNALTQRILWMKAQPASLA"

```

gene      56777..56998
          /gene="orf73"
CDS       56777..56998
          /gene="orf73"
          /codon_start=1
          /product="hypothetical protein"
          /translation="MKRKSLKTKSVNLI FVSSKDQKIQNLWPKKAMLFWARGLLFPSS
          ARGRRMTAFCPRAKLQEEKVVLSSVKETE"
gene      57393..57869
          /gene="nad10"
CDS       57393..57869
          /gene="nad10"
          /codon_start=1
          /product="NADH dehydrogenase subunit 10"
          /translation="MNSKADYVISKLDTLVNWARKGSLWPMTFGLACCAVEMMSAAS
          RYDLDRFGIIFRPSPRQSDVMIVAGTLTNKMAPALRKVYDQMPPEPRWVISMGCSCANGG
          GYHYYSVVRGCDRIVPVDIYVPGCPPTAEALLYGLLQLQKKIKRSQNVRVWYNR"
gene      58254..58883
          /gene="nad9"
CDS       58254..58883
          /gene="nad9"
          /codon_start=1
          /product="NADH dehydrogenase subunit 9"
          /translation="MRYLPPFKQQDNASRASTHVLIEKFSKSLTLLVPKWIQSFQFN
          DELVIVIYPEYVLPFFQFLKDHTNTQFKMLMDITAVDYPSREKRFEIVYNLLSPQVNS
          RIRVKTYGDEITPIESLVGIYSSSNWFEREVWDMFGIFFVNHDPDLRRIITDYFGQGHP
          LRKDFPLSGYVEVRYDDSEKRVITEPLEITQEFRSDFDFANPWEITEKKI"
gene      59221..60753
          /gene="atp1"
CDS       59221..60753
          /gene="atp1"
          /codon_start=1
          /product="ATP synthase F1 subunit alpha"
          /translation="MAVAAGELSKLLEKRISNYTQLNMDEIGRVLSVGDGIARVYGL
          KKIQSGEMVEFANGLKGMALNLENDNVGVVCFGNDSGIKEGDIVKRTRAIVDVPVGRG
          MLGRVVDALGAPIDGKALKDVTRRRVEVKAPGIIARKSVHEPMQTGIKAVDSLVPIG
          RGQRELIIGDRQTGKTAIAIDAILNQKGINSGSDESKLYCVYVAIGQRSTVAQLVK
          TLMNADALNYTIIVAATASDPAPLQFLAPYSGCAMAFFRDNGMHALIYYDDLKQSV
          AYRQMSLLRRPPGREAFPGDVLYLHSRLLERAAKMSDKAKAGSMTALPVIETQAGDV
          SAYPTNVISITDQGI FLETEL FYKGI RPAINVGLSVSRVGSAAQMRAMKQVCGSLKL
          ELAQYREVAFAQFGSDDLDASTQYLLNRGARL TEVLKQGQYAPVPIEKQVVVIFAATK
          GYLDKLNIAEILPFETALLKEIDPSIFTTIRDEKVLSPATSTKLSQFLDKFTANFIAT
          RQ"
tRNA      60811..60883
          /gene="trnV(uac)"
          /note="codons recognized: GUN; 'super Wobble' codon
          recognition inferred, with U in the first position of the
          anticodon assumed to be unmodified"
          /product="tRNA-Val"
          /anticodon=(pos:60844..60846,aa:Val)
gene      60811..60883
          /gene="trnV(uac)"
BASE COUNT 19045 a 11986 c 12444 g 17408 t
ORIGIN
1 tagatgaaaa ctgagcctta gagacacaca acaatataga tcatgtaaaa aggacatcga
61 agggatgcct tgatatggaa agaaaaggacg cataacgatg cgataagtca ttactgagaa
121 tcgaaaaaat ctacgacgaa tgaatctcct tatggggaaa cccaaaggct gtgatggtag
181 aaataccacc aagcaaacaa ctcagggaac tgaaacatct aagtacctga aggaaaagaa
241 atcaatagag agtccgcttag tagtggcgag cgaacatgga tcaggcagta gaaagaacac
301 aattttaaaa aagtttggaa tatcttgcaa tataagggtga tagccctgta taaagaagat
361 gttctttcga aatgagtaga gagtttttta gaattcttga agatagggggg cccaccctct
421 aagcctaagt atatttccat aatcgatagt gaacaagtac cgtgagggaa aggtgaaaag
481 aaccctaagt aqggagttaa aagaacactg aaattcgatg tttacaagca catgaagtac
541 acagaacacg aatcacgtgt gtgcagtatg ccatggtacc ttttgcataa tgggtcagcg
601 agttaaataa aaaagcctgc ttaagccgat aggtggagge aaagtaaaaag cgaacatgaa
661 gagtgtgatt gttttttat taagaccgca aactgagtga tctagctatg agcaggttga
721 agaaaaggta aaactttttg gaggaccgaa cccgtgtgtg ttgcaaaaacg ctgggatgac
781 ttgtagctag gggtgaaaagg ctaatcaaac tcagaaatag ctggttttcc gcgaaatcta
841 ttgaggtaga gcgtatgaaa gatgcattga tggggtagag aactcgatga agaaagggtg
901 cccaaaagcct aactgacttc aaggaaactc cgaatacgtc aatgtgaaca catacagaca
961 gacttttaggt gctaaggctc aaagtcgaga gggaaacagc ccagatcgtt cgttaaggtc

```

1021 ttgaagcgt aactgagtgg aaaaggaagt atataagctt tgacaaccag gaagttggct
1081 tgggaagcagc catcttttga agaaatcgta atagatcact ggtctatgct tgtgtgcacc
1141 gaagatgtat cgaggctgaa gttatccacc gaaacgacga aaagcacaca ctcagtgtgt
1201 gctttggtag cggaaacgctt catacatcat aaaagatatt ttgaaaaaaa tattggagag
1261 attggaagtg agaatgctga catgagtaac gtaatatcat gtgaaaaaca tgcacgcccg
1321 aaatccaagg gtttcagcgc tccgtaaatc gacgctgagt aaggcggccc ctaaaagaca
1381 aagccaacgc tttttctctg atgggttgtg gaaaagactc tttcccccg agaatacagt
1441 gacgaggagc cattggctcc cccgggaaat aacgtattca aagattttaa aaaatctaca
1501 ccgtacccta aaccgacact ggtggatagg ttgaaaaaac caaggcgttc agagaaccat
1561 cttgaaggaa ctcgcaaaaa taaccttcta actttggaag aaaaggtgag gtgagcaagg
1621 gttgactctg tccccctggt catcttggca cagaaatggg ggtaacggct gtttaataaa
1681 aacacaggac tctgcaaaagt ggaaacacca tgtatagggt ctgacacctg cccgggtgctg
1741 gaagcttaac aggagaggtg aaagcttcaa attgaaggcc cagttaaacgg cggcgtaac
1801 tctgacgggt agcctcaata gacgcaataa cataaaattc tattgagtac atgatacaat
1861 tcaaaagggc cgtcattaaa tcttgcattt tgctggaaca ctcgggaaag agactttccc
1921 aacattctct ctactgtgcg gaaaaaaaaa gacaattttt tcttgcgct gtaaaaaatga
1981 gccgaatgga gtcaatcagc aggaaagaaa acctgacagc cacgagcaat gctcgtggcc
2041 attagaacaa tgttccctgag tccccgagag catagctcgc ggggaccaa acttcgtttc
2101 tcaggttttc gctcctcaga gactatacgc aagacagaat atccattctg aagatagagt
2161 ccaaaccttt ttgtaaaaag agggatattg cctaagggtg ataacattgc cttccttggc
2221 tatttgcggg aatatccccc aatatgcaga cgtactcgtt gacacaggtg atgcttggg
2281 gcttggccca agacgctctc acccgctctt cagtaaaaat cgtcgcata tgaggcaat
2341 ccgcaggaaa aagagagttt ctcttgcctc tcagagactt tacgccaagc atcaataaaa
2401 aaatacgtat atgaaactga catcagattg gatctgtgga tttgtagaag gagaaggatg
2461 cttttgtatc tcaatatcaa agatccgaag tgttcaatca aaaaacactg aaaaactccc
2521 aggaggtcaa ccaacgcttt tgacagagat gacttccaaa aaccagaagt cttctttccc
2581 cgagcagagc ttggagaaga aggaaagcgt tcaaaaacag gttcgtctcg tctttaaagt
2641 cacacaggga gtgaaaaata tcaagtggtt atatgctctc aaaaaattct ttggagtgg
2701 tcatgtaaaa agccaacaag catcagggaa ggtttgggaa tatacagtga gtcgatttga
2761 acatcttcac acaaaaaat tccattttt tgaaaaaac acctgtata aacgtaaaaa
2821 atttgatttt taccgttttc gaaaagtggc tctatacatg tcacgaaaag aacatctcac
2881 ttttgagggt cttgcgaaga tagaaaaact cactgctcgc atgaacactc ctgaatgct
2941 tgaagaagta tacgaagaa cttttgatga agataaagtc cgaactttt cgaagaaaa
3001 agaagaagat acagattaaa gagtattcac tcatttcaat gtatgctttt aacaatattg
3061 agcgaatttc cttgtcgcac aagtagcgac ccgcatgaat ggtgtaacga ttgccccact
3121 gtctccaaga tggctctgagt gaaattgaat tctccgtgaa gatgctggagt acaaatgatt
3181 agacggtgaa accctgtgca cttttactct agctttgcat tgaaaaatag acaacttgt
3241 gtaggatgag tgggagaag atgtaaaact ttcggccttg aaataaccac cttttgtct
3301 gctgtttctc aggaatgaag ttgtgaaaca gtgagcttat cgtgtttca aaatggttc
3361 caacagtgca tgggtgggag tttgactggg gcggtcgcct cctaaacagt aacggaggcg
3421 tgtgaaggta agctcaaaact ggtcggacat cagttggaga gcgcaatagt ataagcttg
3481 ctgactgtaa gatggacaca tcaagcagag acgaaaagtcg attatagtga tccgggagtc
3541 ctttgggaa aggctctcgc ttaacggatc aaaggtacgc cagggataac aggctgatac
3601 cccccaaagag ctcatatcga cgggggtggt tggcacctca acatcggggc tttagaatgg
3661 aaacatttta agaatgcact tcacttata gcaagaactg ctggatagaa gctcttccac
3721 tttttctgaa aaggaacaga taccactgaa ttaaaaagct ttaaaagcga atcaatacaa
3781 tcgaagagag ctgacaactt gcaggaaaaa acaaaaagaa catcaaaaaa acttctgatg
3841 tatgtctctc acgactaaac gtgaaggacc taaagagaca cagagctgaa ctctaggtta
3901 tgtctatgct tattctgtgt cgaagatac agttttcatt ctgtgaagga atgaaaaaca
3961 ttactaaca ataacacatt atggtagatc gtcaacatac gttaaatcaa acgacattag
4021 aatatcttaa attcatgcaa ataccctata ttcctgagac tcgagatctt cgagggagtg
4081 cacttatttc ctatttga aaacaacacc tcacagatct tcaaaaggct gttttgtgtg
4141 gaaatctctc aggtgatggc acattatgca acaaaatagg aaataaccgga agttccaatt
4201 ttaaaattga tcaagctgcc tcagaaaaag gaatcgaata cgttcaattt atctatacaa
4261 ttctggcacc ttgggttggg acaccaccaa aattacgtta caagaataat aaagctcata
4321 gtatttgggt tcgaactttt cgtatgaaag aacttgattt ttattgtaaa caattctata
4381 gcattgacgc caatggctca cgacaaaaaa tcatcccttc tcttttat ataatggctg
4441 atgatgtagc tttagcgata tggtttatga atgatggagg caaaacagac tatgggtatc
4501 gaatacatac tcaatgcttc actgtcttcc aaattcgtca acttcaaaaa atcttgaac
4561 acaatttttg attaacgact ggaataggaa gagatcgaaa atcattgata ggtacaatac
4621 ctgaagtcat aaaacaagag agaaagcttt tctctcgttt ggatgatccc acatgtaaaa
4681 cttattacta tttagagatt catcgccac atcgtgaaca atttaccgagt ttagtggagc
4741 cttttattct tccaggtttt cgatacaaat tgcattcttc acaaaacaa gactaaaatc
4801 ttaaagtggg gaagttaaat gtcggttagt aaatacttaa acgatgtcgg ctcatcacat
4861 cctggggctg aagaaggtcc caagggttcg gctgtctcgc gataaaaagt gtcagtggc
4921 tgggtttaaa acgtagacaa caaaactcgc tgtctttgta gtaatacaaa gagaagaatt
4981 ggcttatatc ggtgaagcct aaaggatgtc aaacgtcga tggtaatacc gagggaggc
5041 ttttgaatca tactcaacaa aaatcaagac tctattcata tcttgtaaa catacgcaa
5101 tacagctcaa actattaata acacaatgaa tgcaaacaca aaatcaactt taaatccaac
5161 ggaagctgca tatttagcga gattgctcga tggagacggc tctatttttg tacagattgt
5221 accaaaaact aactataact gtttttttgg tttacaattt acgatcaatt ttactcaag

5281 tacgaaacga aaacatttca tgcttgcgat tgaaaaatta ataggatcaa aagcatcctt
5341 ccgcgataga aaggacggca attgtgaaat agctatttat ggttggcagt ctggtgcttg
5401 gtttcttcat gaagttcatt catatttgaa attcaaaaaa cctcaagcag atcttgtttt
5461 aaaaattgtt gagaaactac cctctacaac aacttcttac aaaaactttt tagaagtatg
5521 tgcttttagt gataaactca ctgagtacaa tgactccaaa aaaagaaaat accaaagtaa
5581 gtattttgtt gaatactaca aagaaaagagg gatcgtcatt gatgaaatcg cccctgtaga
5641 gactttttga tgaatacat acttttatca aaagatagct ttcttgggta agaaagtat
5701 aacacgccaa cccctaaaag atgaatacac cgtgaattga aagaacgctt tcatggtgt
5761 tttagggtga agatatagtc catgctttta tgaaaataag agaataacat gcgtgagaca
5821 gtttgggtcc tatctgtcat ttgtgttga gaattgagaa aaaggggttg attgtacgag
5881 aggaccttaa ccccttaacc actggtgtat cgattgtatg attttaaaca tgcaacgctg
5941 agaagctaaag tttagcagcag ataactgctg aaagcatata agcgggaaac ttctttcaaa
6001 acaaatcttc atcatcgtag aagaaaacta cgtaaatagg caaaacgtat atcttttaa
6061 tctgaaaaga cagcgttttc gtactaaaat gaaaactttt tcaaagactt ctgaaaagtc
6121 gaagagccat agggccgaaa ggccgaaggg atcagaagtg gaacaagttc cacttctgat
6181 gcatcgggaa tggaaacacat tccattccct atcggggcga gagcccttct ctggaggcgc
6241 tcatgtgtg agcgcacc acccttctga atctttgaa aacttcttat gtgtgatgg
6301 ggaaaatact ccaatataaa atacgcact catgaaaaca cgtctttcta ttttaaaaca
6361 accacttgta acacttgtaa atgatcatct tatcgaatat ccaaccctta caaatttaag
6421 ttatttttgg ggatttggat caatggctgg aatttgttta ctctccaaa tcttaacagg
6481 agtgtttctc gcgatgcact atacaccaca tghtgacctt gctttctaa gcgtagaaca
6541 tattatgcga gatgtggaag gagggtggtt tcttctgtac atgcatgcaa acggtgcaa
6601 tatgttttctc attgttctt atatgcatat gtttctgtgt ttatactatg gaagctatgc
6661 aagcctctga gaattcgtat ggattgtagg ggtgatcctt ttctattga tgattatcac
6721 agcttttctc ggttatgtgt taccatgggg tcaaatgtct ttctggggtg caacagtaat
6781 tacaagttaa gccagcgcca ttccatagat agggagacgaa ttaactcatt ggctctgggg
6841 aggatctctc gtggacaatg caacattaaa tctgttctt agtcttctt acctgttacc
6901 tctctctctc gttggaatca gtgcagtga catcatcgtct ctccaccaat atggttcaaa
6961 caaccctta ggttcattag cgagtgtgga taagattctt tttatctct acctttatgt
7021 gaaaagattta gtaaggtggg tagcttttgc ttatttttct tcttttttg tgtatttgc
7081 gccgaactta ttaggctcatc cagacaacta tattctctgca aatctctatg ctactccagc
7141 tcatatcgtt ccagaatggt atttctctcc tgtatatgct attctgagaa gtatcccaaa
7201 caaacttgga ggattgtgt ctattggctt agtttttgtg ttttattgg cacttcttt
7261 tttaaatact tctgaaatta gaagctcaag ttctcgtctt attcataaaa tcttattctg
7321 gttctcttga gctgattgtg ttttacttgg atggatcggg tgtaacctg tagaagcacc
7381 ttttgttaacc ataggacaaa tagcttctat ctttttcttc ttctattttt tagtgtgtgt
7441 tccactgtta ggaagaatcg aaaaagcctt tataacatat aaagcttaat atattgtacc
7501 tttcttttctc gtaagatgaa ggttgaaaaa ccacgtgcat cgtggggctt acattgagcc
7561 cttctgtgaa ggcagatctg tttttcatga gttctgctca tagaaagcaa gacttcgatc
7621 ttctcttgca ttcatgtgtc agcatattgt catcgggggtg cgagcatgct cgaccctcga
7681 tgacaaaagg agcctagcat gcgatgcgcc cgatgcatca gaagtggaac ctgttccact
7741 ttttttgcaa aagaagtgga acaggttcca cttctgatgc ctctggccct ctctgcatag
7801 agcattctaa aaataaaaaca ctcaaaaata aaaatgcaa cttttcaaca aattctatca
7861 acgagacgcc aacttgtcaa acgaaaaatca aaagcaccag ctttacacaa agctcctcaa
7921 aaacgtggag tatgcattcg agtatatacc aaaactccta aaaaaccaa tttctgtttg
7981 cgaaaagttag ctaaaagtac cttaaccaat tctatagagg tgattgctta tattccaggt
8041 gaaggacaca atattcaaga acattccgta gttatgatta aaggtggaag aatcaaagat
8101 ttacctgggtg tgaatatatag gatccttcca ggtgttttag atctccaagg tgtgaaaaat
8161 cgtcgaaaaat ctctgtctgct atacggaagt aaaaaacaaa tataatacaa aatggaaaata
8221 aaaaacccaaa gtaaaagtgg tgctcgaaaa agtgtagaac ccaaatctc acctaccoca
8281 agtggaaaag gaaacaatag cactacgcag aaatccaatt cttcaatctc cagtgtagaa
8341 aatcaacgca taccaaaaat ctttgatccg aaaaactcac atgcacaaga attacgtatt
8401 caaaaataat gtataaaactt gttaacgaaa gatggaaaaa aagcaaaagc tttttcttg
8461 ttaaaaaacgg ctttcacttt agtaatgaa aaaaaaagaa cacaaaagta tggagaaaaa
8521 ccgtcatcaa tacactctct cactgccaat atgttatttc ttcaaagtat tgaaaatgtt
8581 aaacctatct ttgaaqtctg aaaggtgaga gtcgctggaa atacactggt agttcctgct
8641 ttaattctct ttgagagaca agaaaaataa gcacttctgt ggattattga agcagcgcag
8701 gaaaagaaaa gaaaaaaccc acattttctt ttgaaagat gcttagccat tgaaatctta
8761 gaagcagtac agcaccagg ctatgcaaaa caaaaaagag atgagcttca taaaatagct
8821 gaagcaaatc gagcttttgc tcattacaaa tgggtgtaga ttaggaaatca tttgaatctc
8881 caatgtgaat atcgttttgc cacacaagaa aaggacgcaa tgagcgagga atacagtgat
8941 caggggagcg aacactctcg ctctccctct cactatggct caaccacacc aaaaagtgtt
9001 tcaactcttc ggaaatacaa tgggtgtgata caacaaagt tttcattgat tcttactaa
9061 tcaatttata ttgttctcac atactcaagg tgacgctctt tcagtaaaaa gaggtattaa
9121 aactttttag ttaaagaaat aaaaaatatg caaatattta aaaaaacaa agctgttgcg
9181 aatacatctc tcacattgtg agagttgaa cygggtgac cagggagtgg gaaaggtcaa
9241 agatattctaa gctttttgtg tcagcttttt caagtatttc aatatagaat acaggttttc
9301 taacaataaa catcccattc gcgggcaaaa atttgccgac agcgtgcaat ctcccaatgg
9361 tatacaatta ttgttgaata cactgacctt tggttttcga tacctgatct tattctgtat
9421 tttagactat gaagtatgag gatacttttt tactatggt tttctatca atcaataaaa
9481 tgcagatgac ccaagtaaat gtcattctca ctatgtctaa ccattgtatg agtccccacc

9541 acctcaagag gaaatgggat caaggcttaa tatcgggccc ggaagagct cctaaaaat
 9601 gtatgagtaa ttttattggt gacctggag agaaacaaca gcaacaaata gccctgagca
 9661 tgcattttgg ggcgtttaac gctgagcatc aacttaagt ctattatttt aaataccaat
 9721 ctctgtccgcg attattgaa atcacatgaa agggaaagcg gctaaaacct ttgtcttaaa
 9781 tctgtttacg agattcttga cggctgtgaa aagtaataac aatcagatga gtttctcatt
 9841 acatacagaa gatctcaaag ccttgtttta cgacatgaaa tcttatcttg gaaatatgca
 9901 atctcttggg cgactaatcc gctgtgcctg aaatgccaaa aacatcttgg ttttagata
 9961 gatattggaa acagttgcaa gctctcatca gaagcccccg caaagctcaa aaccacactt
 10021 ttgtgtatca aaaaactctg gattcatatg taacatcaac tttccctta atgccaaaa
 10081 tatataaaat ccttgaagat gagaacttac ttatcaaat tatccgtgca tatatcgacc
 10141 acaaaaaacc ctacctgaac aaaacaatgg aaaaagccgt gcaagaaatt aagaaacatc
 10201 ctcaggagggt ggaggccctt cgtgcatatt tacaanaaca tattgatcaa aacaacaag
 10261 agtaaaaaa aagcaatgct atagcctttg tctatttaag tatctctctt tcccaagtg
 10321 caccgcaaac acattgcgac agctgcttgg gaaggagaga tacttatttc actaaacatg
 10381 gaaaggagct atccttccca gagaagaagg accggcactt tgcgatagct gggtaaaata
 10441 aaaccttctt tcccgaaaga aaacggccac tttatacga cgagctcccta cctcccagaa
 10501 acgagcctg caccaggtg tttacaaga cacaaagtta tttctgggg tttctgcaaa
 10561 ggtcattatc tgagggttat tttgtctctc tgagggacaa aagttatttt ctgagagtga
 10621 ttcttctgta ttataataag cgcttctcat tttctgggga ttagggggtgc taatccccag
 10681 aatcctaaga gatctctctg ttcagaccac atcaatcttt agggggctcg ccccccttt
 10741 tgcgaagcc gcgcaaagcg aacggctctc tctctactc ttcttcacac aactcactt
 10801 tctgtttcaa cacaaaaat agaggggcaa caccgagcaa aagacgctc aagtgtagt
 10861 ccaaaaaaga tgaagaact cgttcacgcc cagcgcagtg cagtcataag agacaataga
 10921 cccgagcaaa gcttgcgca gagagtgtgt acgcatcctt tcagctgctc tgcagctgaa
 10981 gggatgctgg ctgtcatacc atttttctc tgtactttg aatgattgct tgaaacggc
 11041 tgccgcttct cccacaatcc gaagacctc cctgtggca ccttacaatg tctgaacgtt
 11101 ttgacaaaaa cgttcagagc acctggtcat tttctgggaa tgaacgtctt aactcccagg
 11161 tgctcagttc tttttgagct caaaacacat gaatccttag gtgcgcccg cggaccttt
 11221 ctcaggtgcg gctgtgtgcg gcacccaatc ctatgcgctg ccgcgggccg cgcctcctt
 11281 tctcaccgpc cgtttacgc ggccggccct cttgcttgg ccccttccct aggagaggcc
 11341 ccccatctag agtagtgagc taggagctcg gatcgatttt cttgttttc tctaaaagtc
 11401 attctctgtt tttagagctt tcccccttcc gaggttcgcc tttgaagagc tttttctct
 11461 tcttaataata ggctttactt aaaagtcgga aattccgggg ttttagagcg aaaaacgctt
 11521 ctaaaaacccc ggaatacag ggagtaaaaa aatggaatc ggccgtttcg ggctttatc
 11581 gtttttttac tccaaaaacc atcaaaatc aagggattta tttgtacct caggcgtcac
 11641 tttgtcacct gcataatgta tgaactca aaaatcctat gattcactt ggaactcagg
 11701 cgatgtgaac caaatatc atatttttt tctctataa gagcaattgt ctctggatc
 11761 caaaaaacag aagcttactt taaaaaaaag cacatacatg aaaaagaaaa aaaggctcaa
 11821 tcttataaaa gttccccctc tctcacgta tttttacctc tgattgagge agagaatcct
 11881 ggccctcagc agcattgctc gtgggggcca gaactctggt cctggggggcg ctaccaggga
 11941 cttcgttctc cagctcaatc gtgtgacca tgtgagtaaa tgaagcattc acatgaatac
 12001 gatacaagct ctccacaaa caaggagaaa gagtctctgaa ataaaaataca aaaaataaaa
 12061 aaaagtgtaa tattaaagag actgacagtc tggaaaacte accttctctt ctccgtggac
 12121 tctcttatcc ttcaacgagc ctctgtaaga aacacagagc ctggtcatgt accgactcaa
 12181 gctttactct ttgcacagct ctacacatac tataaaaagc aaaactatga tcaaatccca
 12241 agcacttcaa gcgcttctt atcttttctc aagcatcgca ctgatttcag gtgtgaggt
 12301 tattacagca aaaaaaccag ttcattcagt tctctttctc attcttctct tctgtaatgc
 12361 tgccgggatta ctgattttat tgaatctcga tttttctgct atgatttttt taattgtcta
 12421 cgtaggagcc attgctgttt tatttttatt cgtagtcag atgttaaaaca ttaattagc
 12481 agagattaac gaaagtattt taagatatct ccccgttggt ggatttatcg gcttttctt
 12541 cgtctttgaa attcttttag tgattgacaa tgatctcacc cccgtctctg tgttctctga
 12601 atcctcaaac cacgtttttt catccggtct tcttgggtgt ctggaagtgt tcttttataa
 12661 acctcaaggt gctttagctc aatctctctg agaataataga aaggatcttg aaagtcttct
 12721 ctcgaaattct caagcatacg gccttctctc aacttcaata tcaacctaca cagaatgggc
 12781 actcaaaact cagagtaaaa ccaatattga agcttttagga caaatcattt atacctatta
 12841 ttcactgtat tttatccttg caagtttaat ctacttgggt gcgatgatcg gagctattgt
 12901 tttacaatg cataaagggt ttcagtgtgaa aagacaagaa gtctttgaa aaaaataccag
 12961 agatcactca caaactgtac ataagattgg aaaaacttctg aatgcattca aatacaagg
 13021 tgcgagactg gcagaagagg gtgagtggtt tttgcccctt caccctcatg aagatgcctt
 13081 gaaagtgggg aaaaacataca caaatcaaaa gagaagaaaa aagagacggc acgcccagcc
 13141 cgtttgctac gcaataccaa taggttaggc ctctctcatg tgaggagggg ggttaggggg
 13201 ggaggtgtaa aagggggagt tgaactccc cctcaaggcc ctctcaaggc tttgctctt
 13261 aggcctataa agagggccaa ggggctccca aagagtcctt tggctcagct cactggcctg
 13321 ggggctctct ggagagggcc ctgggctcca aacggaggtt gaaagcctct tcatatattc
 13381 aagggcgaag ccctttttca tagagaaacg ggcgaagccc ttaatgttta gggcaaggc
 13441 ctctgtaata gagggggcga gccccttctc tcaagctctg gtgcgagcct gaaggtctcg
 13501 accagagctc tcatagggac ttgagtcaag gcaaaagcctt gactcaagga ccagaacgga
 13561 tctctgggtgt gttcatatgc gcttttctgt cgggaatgga aaaacttcca tctccatga
 13621 ttgcacccct atgatcggag gatcgaagga ccgtagggcc gaaggcatca gaagtggaaac
 13681 ctgttccact tcttttgcac cgggacccta gcatgctagg gtccctttgt atcggggctg
 13741 gggcttgcct gcacccctat gaccgaaaa cttttcacga aacaggcagc aagcggcgtg

13801 ggttcttcat cttttgagag ctggtgagcc aggtgaacta ctcagtgtga cacagcttgt
13861 gctcagtaaa ggtgttccac ctgggtgtttt tcatttgagt ggaagcgaaa tgcgtattc
13921 cagcccttca tgttatgtgt tcttccaaaag tcagctttcc gctccatc tttgacaagc
13981 tgtccattgc tggcctctta tttgtaagaa agagctttct cttccaaaaga actttcggag
14041 tataagcggg gcagcttgtt gtttttctctg aagtttgact cttctacca tccatgctct
14101 cctggaagtc gatgtgtgac ggctcctttt tccagggtg atacttttac tegtatgccg
14161 aaaagggtgt agctctaagg gcacagagct tctacaaggc gacgtctgta tctcacttgt
14221 gtttctatat aagaagaaga agaggctgat cttcatcaaa ggtttttgtc aactccatgg
14281 gatacgaaca tttccgcatc cttatggcag ccttcatttt gtgccaggct tttgatccga
14341 agagatctcg aaggcattca accaaaagcc tggcacacag ggtgtattg acacttctgc
14401 atggaagcac tcgaagcggg tcttctttca atagaactct gccttatgct tcacgtgaag
14461 tcgcgaggga cccgctatct tgctctctcc tegttagctct gtttgcagac cggggagcac
14521 tttccgtgct gtattataaa aactgcgtat gacaaaattt gcaaatcgtt ggccttttc
14581 aaccaatcat aaagatattg gtactctgta tcttattttt ggtgcttttt cgggagttct
14641 tggaaactgt tttcaatgc ttattcgtat ggagtagct tcaccaggtt atcagatatt
14701 agctgggaac catcaattgt acaacgttat catcactgct catgctttct taatgatttt
14761 cttcatggtg atgcttctc ttattggagg atttgggaac tttgtcgtt caattcttat
14821 tgggtgctcg gatattggcat tttccgcttt aaataacatt agtttttggg tattaccacc
14881 atcactttta cttcttttaa gctcagcact tegttagggt ggagtaggaa cagggtggac
14941 cgtttgcatg ggataagcgg tcccgaagtg tcaaaaaaca agactgaagt ttttgaagc
15001 tgcaggaaga tctgcccgtg aaagcttctt tttgaatgct ttgtataaaa aaaactcac
15061 tcgatgcccgg aactttttg cgaagtaatg gactactcga tattcaaaaa aatcagtaaa
15121 aatgtgtatt acaaaaaaac aatccgctcg ggcagatggg tccacgattt tccgaaaagg
15181 agagcagtg tttacatcat cttcccatac gagacttaac gtgaagcacc cattcactca
15241 ccatatcaat ctttctcact ggctcgttgg tttaccgat ggagatggct gttttctata
15301 gaatgaaacc aaggatggcg tatggcagtt cacttttcaa atttctgcca gcctttataa
15361 ccttcagatc ttatatttta taaaatctgt gcttcaagtg ggttcgggat cttatgctgg
15421 agaacgcatg gctcaatata gactcagaga tccgcatat ctcaaaaaact ctattcttcc
15481 cacttttggc ggtcttctct tattctcggg gaaagcttat gattacgagg cctttaagca
15541 aggtcttttc atctatgagg atcccactt gtcttaccgt gaaaaatctg accgtcttcg
15601 tgccttctga gcagcagctt tagagaacaa aaagaagct tccctgctt ggggtcggg
15661 tgaacgaccg accaagctct gggttcttgg ctttactgaa gcggagggga gcttttcat
15721 tacaaaaaaa gagaaagagc gatattgcca tggatttggg ctcactcaaa aaaccgacga
15781 gattatctta ctgttttttc gagagatttt tcaaatagga gcacaggtga aatggaattc
15841 tgcgtggtct actttttgga gcttagatac cacgaacttt cgtaccatac aaaaatcga
15901 agcttttttt aaaaatacct tcaagggggg gaaatccctt gaggtttagaa tatggagtcg
15961 agctttgaaa tatagagggc agtgtgaaac attgaaacac actcaggatt tcttctgacg
16021 tcttctgaaa ctgtcttttc taggagatga ggtcgaaga ggtcgggggc gctcactttg
16081 tgagcgcctc caggatcgaa gatcagcccc catgagcgtt gctcatgggg ggctggaatg
16141 cagttctgct ccccacgagc attgctcgtg ggggagagag taggtgttgt gaatgagatg
16201 atggtatagt cegatctcta tggtagata gacgctgtaa tgagagtgct ctcatcaat
16261 gagaacatgt tgattacca cttttttgt tatccccctt tgagcaatat cacagccac
16321 tctggtggtt ctgttgattt agctattttt agtcttctc taagcgggat tagtctatt
16381 ttagggtgcta ttaactttat taccactgta ttaaatatga gaggctccag tatgactatg
16441 cacagacttc ctttatttgt atgggctgta tttattactg cgattttat actattgtca
16501 cttctctgtt tagccggtaa attgaatct gccgcatccc tataagccaa cccagggca
16561 cttctctgtc aaggcttgt agatgaaac ctggctatat gctggaaccg cctcctgata
16621 cggcaccatt gccctatcac gccaatcagc aggtcacgct ggtacgatcg aaaggtctgt
16681 ggagaggtga agattccat gtattcccca acatcgtttt gacttccagt ggaacctcag
16741 agactacacg ccaggatgta tgtataaca caaaattctt atgtcaactt ctcaatttcc
16801 tttttatctc gctggcttga ttgaaggtga tggccatttt tatagcccag cacgccaccg
16861 acaaaaacag gggaaacctc tctatccgtc cattcaataa tgccttggct tcaaaagattt
16921 gcctcttgca cagatggtct tagctcatct tgggcatgga agcctgtgtc gaacatcggg
16981 aaaacaagcc tatattttaa cgggtgacga tccgagaaggc tgcgttggct ttgttcatat
17041 gatcaatgga aaacttcgca catcaaaatg tcatcaattt tcacgtttct tgacatgctt
17101 caatgagcgc ggatattgaa actttgttgt agggcccaaa gacacttccc ctgtgtttct
17161 gaatgcgtgg ttggctgggt ttattgaaag tgccggatgt tttgaaatcc gatgactct
17221 cacaccacca cgtcgaattt cgcttctggt tgtcattgaa caagcgtact cacacgaaga
17281 agatcgtcct ttgttacta caatttccaa atctttctta gctcccgaac ctgcacatcat
17341 acagcgcaaa ggaaggggag aattctggcg cattcgcaca accagcctcg ctggcaatct
17401 gtgtctactg cagtatctga aagaatgttc gcttgcctgg gcaaaatata tggatttccg
17461 agattgggtg caggttgggt aaatgggtcg tgagaaaaaa cacctgatgg atgcaagttt
17521 tcccattgtt cttgccataa agaagactt taatgacgca cgaacgacat ttgactggac
17581 acatttgaaa actttttttt gaataacatt gtgttgacac acaaaaagaat atacatcat
17641 atgatatagt ccgacccact atgaaaatag tggcgtactt gccctaaatt cagctccgat
17701 ataggggtgc gagcaagcgc gcaccccgat aactacgac atcggtatga atttttgagg
17761 agaaagaaaa aggcattctca aaggcttttg gttttttgtg gttttttctg ctccatgctc
17821 acatgagagc aagtcaacag tacagggtat tacaatgctt ttaactgacc gtaattttaa
17881 tacaagtttt tttgatctcg ctggaggagg agaccctatt ctgttccaac atcttttttg
17941 gttcttttga caccctgaag tctatattct cattcttctc gcgtttgga tttatgtca
18001 agtgatttct actttttctc gtaaacagct ttttggatat cttggaatgg tatacggat

18061 gttaagtatt ggaatthttag gattttattgt atgggcttaa cagaataaga tgggccccgc
18121 accttgtgaa gagtgtctgac aaatthttgct atatgctgga accgcctgaa gttaaaaagtc
18181 cgtcataaca gaaaacaaag caaaaagctta gattttccgg aaaaatcttt taaattggcc
18241 aatcagcagag aaacacaatt gaaagacgag tgctttcccc aagcctaaag aggtactaaa
18301 attgggatat ttatacataa aaaaaatata tatgaaagat tgtggatcct cagagactat
18361 acgcaaaaaca actthttcgat ttcagaattht tttcaaccac ttgccagaac atataaaaaac
18421 atgttcgccc actthtttag aatggthttat cggatthttgt gaaggtgact ctgtthttag
18481 ttcttcaaaa ccgtccctct tagagthttat tatcaatcag aaagagaaac gaatthttca
18541 gaacattcga accacattag gatttgcccg tgtgacaacc accccttatg tgggaaagtg
18601 cggtaattct tactatcgat tttcagtgtc aaaaagatcg gatatagaca gaattgttca
18661 tttgtttaat gggaacattg tttcaataa gataaatgac aactthttct tttthttca
18721 gaaccgaaac gcacaagcct gcaaaagatgg aagacctcct atthctthttc ttggaaagca
18781 tgaatthctt tttthtgaaa ataacgcgtg gctthtcagga thtactgatg cgggggagtg
18841 cthttatgtg gaaaaatcca aagataaaa atattcactt ggttatctgc thcagatctg
18901 ttatatttta gatcaacttg ctgaaagatg gatattagaa aaaaataaaa atgaggtctg
18961 aaaaacggg tytatttcaa tccgtaagac aaaaagataaa gatggcaacg tcaattccg
19021 aagagtcgga caagaaatgc tgcgttatac atgcacacat taaaatgtc thgaatthct
19081 tatgaattat ttaaaaaat atcctthgtc ttcagaaaa tctatctgct tccactgttt
19141 tcaaaaaatg atgagatata tcaaaaatcg aaaaactgtg gaatgggaa gaaggtthtt
19201 aaaaaaagtg gaaaatthac tgaataaact caaagaagaa gactaaacac aaatthagcc
19261 tcgcttctgg gacggggcag cacttgccgg thtattthtt atthctthgaa caatgtgctt
19321 gtgaggcttg gacccacggg caatgaaaag ttgatgatag agtccatgta atatacga
19381 gatattactt ttgagcatga gttgggggtt ctcccccaat tctgtthttc atatatgcat
19441 cacatgtatg ttgttggtht agatgtatag actcagagct atthttacagc agctaccatg
19501 atthattgtg thctacag gattaaaat thtagctgga thgctactat gtggggagga
19561 agtattgtat taaagacacc aatgctthtt gctgtaggth tctthttctt thcactgta
19621 ggaggagtga ccggtgttat thtatcaaac tcaggcctag atatthcctt ccacgataaa
19681 acattaaatg thatgtatgt ctaactcaaa tatcattgtt cacaaaagt ctcatggacc
19741 tcaatcttht caggatcaac thgaatctth thctctcgga ctthctggaag gagatggctt
19801 thtcaagtg aatcattgga gaaagcgtc ctthcaatat cgaatthtca taaactgaa
19861 caacacagaa gaaaactata gaatgcttac ctthtatcat cgaaaatthaa acatthgaa
19921 tgtgaagatt accgcacatg gaacgagagt gtccctggth gaagatcaca gagaccgtat
19981 accgaaaat atgggcatta thgaaaagca tgggttgaga thgacaaaag ctthgaatctg
20041 atatggthtt thtcagratg ctthtcagaa thcagatcggg tactctgaa atctctactt
20101 gcgaacttht cgagatagt gggaaaaagc tgcatactgc gcagatthtt atggaagagc
20161 cccagctcag catthcagact thctctgtgc thctcatcca thgttagatt thttcaaaat
20221 gttgacacca tataaagcac aagatthttt acaacgacct catthtthtgc atthgtthtgc
20281 cggaaatcatt gaagcggag gatgthttth thttcgttaag thctgggaaat ctthgtcttht
20341 thcactcgcg caaaagtatg aagaagcctt thctctgtct gthcaagcagt atthtcagct
20401 thcaaataga gthcgaagag thaatthcga thtgtatath thgaaaactt thaatgthac
20461 acatgtthcag aatthcttag aatthttgtc acctggaagc thgaatgctc thcataagaa
20521 agcatatgca gaaaaacgtg thgattcatt tggattgttg ggaatcgaag ggaatcgaag
20581 gcaagagtht gaaaaacgct thctctctgt thctctctgt thtagagctcg agaactcta
20641 gagggggctt acagatatht gaaaacatag ggtgtctgtg thaaatthctg ctatathcaa
20701 gaaactccta aaggccaaca gthattgaagc acatctthctg atctgccaag cghatthgtg
20761 thcgttgcac acgggcatag aacaatggac aatthtgcag aaaggacaac agthtthgaaa
20821 gaggggctct ccgathcctt ctthtcgacc cacacagaa atathtgcct ggtthctctc
20881 thcagagacta cagcgaaac acaaaaagac aagcagacgt thggggtgat gacacagtht
20941 atcaccttht ctthcaccatg thttthttgt gatgacatag thcactctth catgaaagtg
21001 aaagcccgct thttatctga ggaactthta thgtggttht cactthctat thttcttht
21061 aatgggtgca gthattthgaa thgtthctgg atthttattc thgathcggaa aaatthcagg
21121 thtacagtht cctgaaacat thaggtcaaat thaatthttg actthttthct thaggatgaa
21181 thcatccttht thttcaatgc atthctthgg actthcgggt atthccaagac gthatthcaga
21241 thtactctgat gctthctcag atthggaatgc aatcgcgagt thagthctt atthtctat
21301 thttctctct thattctctt thtatgtgtg thataaaaaa thgacagga atgagthtgc
21361 thcgaataat cctthggctt thgatggaaa thcaaatgca thcaactact thgagthgat
21421 gctthcacta cctthcaggtt thcathctth thgaagatth cctthcaatta aagaatthaga
21481 thctactgtt thaaaagcaa cagaaatthg thgaaaatca gcthaaaaat aagacagagt
21541 agggcttata cctgagthc actctcactc thcagatctgc thctctgcaag ctthctthgca
21601 agagggaagc ctthgtthcca aaaagcthaag ctthttthggg gthcagatathg ataatcggg
21661 thcgggctthg cccgcacccc gthgagthct thcagctthg acgthctat ggtthgthac
21721 aatthcaata cghatcagag agctthgacgg ththccccaaa aatctthtt thcagattht
21781 aatthtcaaa aagaaaaaaa cthcgtatgg thgaaatagth thaatgthta gaacatcaga
21841 atcataatct gaatthtggg gthtcaatth cctthttctg thagaaaagc aaagthgagth
21901 thaatagaga gthttthttt thgtcagtht cgtthgtaca aatthcagtht thcatttht
21961 gggataatag ctthagthtgg thagactact cgcctthctc gthgagagtht cgggthctga
22021 thctccgttht thccgthcatt cgtthgtgth atactthctg thataatthg aaatthacc
22081 thtaacaaag atthttthatt actthgagth ctthgagaaca thaacatctg thttagatc
22141 gthgagtht cgcgaatgth cagggcccaa thagthgggc cctthtcaa thgattatac
22201 atthagthg thgaaatgct thgtthgat thatacaagth thaatagthg aacagthcag
22261 atthgagagc thgcaactca actthcaagga aaaaataaa atataccacc actactgtht

22321 ttgctgatgc gacaagtaag actccccag agcaaagcat ggtccatgca aaaccagcaa
 22381 gaacttttaa taaaacgtga ccgctcataa gattggcgaa cagacggaca cccagcgaga
 22441 ttccgcgaaa actataagat accacttcta ttacaacaag aaaaggcgcg agagctaaag
 22501 gagctcctgg tggcaagaaa aaactgaaaa aatgaattcc atgtttttgg aagccaataa
 22561 gagtccacac aataaaaaata gagattgata aagtaaaagt gacaataaaa tgacttggtg
 22621 ttgtaaaact atacggaatc attccaatta aattactaaa taataaaaaa gtaaattgtg
 22681 taaatatgaa aggaaagtat tttttcccag cacttcctat ttgttcttca actaaatcca
 22741 gtacaaatcc atagatcatt tctacaaatg attgccatct gcttggaaact aaaaaatccc
 22801 catttctggt cactaaaata aataaaaaaac tcacaagtcc agttgtgaga agcatgaaaa
 22861 atgcagaatt tgtaaaagaa aaatatccga gtgggataag aggtacaatg gtaaactggt
 22921 caagagtgga aaacatagag tattattttt cttttttaa acacgcaagt gtcccgctga
 22981 aagggacgct gattggcatc cccatgcttt cggccgggta gccgtgtaag ctctacagag
 23041 aaagctgctg cactgaagca ttacgaagca aagagctttt tgagattgag aagctcacag
 23101 tgatcgagtc aatcttcatc tattcatttg tatgcacaaa gcaagagttt ctgaatgtct
 23161 ttctctgact catcacgctc tgctgctttt tggccgtatt ggtgatagag tcaaagaaac
 23221 aaattttcta actaatagga agatcgaagt cctgcccccc caagagcaat gctcttgggg
 23281 ggacgcaacg aagttagctt tctcacgagc aatgctctgt gggagcagga ttaattttac
 23341 acacaaagtg tgctttagtc tcaattggat agagcatcag actacggatc tgagggttga
 23401 gaggttcaagc ctttccaagc acggttgatct gtattatttc atatttttgg ggataaagtc
 23461 ttctttcttt aactgaaaag gataataata ttgtcgtggg gagaataaaa agattctctac
 23521 taagagttta aaattttatt caaaaatgaga atattttatt tttctaggac cttgggcaat
 23581 ctttttcaga ctggacttga atgcattgga ttgcatttct tgtttttctg tcttttggg
 23641 tactgttgcg ccatcttcta aaagcgaagc cttgcttgaa aaaaaatcct gtgcattgtg
 23701 aatctgaata agatcaccag gttgaatttg cagacttgaa agattgacag gtaatccatt
 23761 aagtaagatt ccacctgaa gaatatattg acgggcagac tctaagtac tcaatgaaa
 23821 cattctataa acaacgatat ctaactctcg ttctaaaaga gagaggaagt tgcacttaa
 23881 ggatcctgaa agttgagagg ctttttttaa aaagaacacg aagtgttttc gccgaagatt
 23941 tccgtatagt agagaaaaata gtcgttttga ttgtaaat tgaacaaagt acgacagttt
 24001 ttttttcta cgacttccga gtttttggag gagtaaacct tgtttgtgag ttaattttacg
 24061 tggaaaatac tgaggagagc caaaaagattc aactaatcca cggtatcgtt taaattttat
 24121 tttttgtgtc atgtgtatgt ctttataatga agctcgaagc gtgtgacgca taaacctaaa
 24181 tggctgcctt taaaagcagt gcttttgaga gccgattttc gagcggaaaga aaagtgcgga
 24241 tagcgggatt cgaaccaca caacctgctt ggaagggcag gaacttacca ttaactata
 24301 tccgcaaaaa acactatgat atgaacgata ggtgcaaaagc ttctctatca gatgacatgc
 24361 tccgttgggtg aaatggtaaa cactgcagac ttaaaatctg ttggggcgac ccttatcggg
 24421 tcaattccga tagtgagcac tctttatagg ttcagttcac tgatcatagg ggtgcgggca
 24481 agcccgacc ccgatattct tgatcataat caaatcaag ttttctatt ttagggttct
 24541 acaaaaatga gtcaatcatt tttgaggcaa agtccggact cttataacaa acgaaaatta
 24601 attcatttca aaaataattt tagggaaagt atcacagaaa tacgcattta agcagagagt
 24661 gtagaacctc cacaacgctg tcttccatct atatggacac gctcagatga aaaaaataac
 24721 gcattttata aaagaacatt ttcttttttc agaaaacacc gttgaaagag gataaataa
 24781 tatcgctaaa gggaaagata tcgaaagacc gaagggccga aggcacgaga agtggaaacac
 24841 gtgtacgagt tcatgaacct ttggtctttg aatacgtcag acgctatcat cttcttgcct
 24901 ccagagcatt gcttgtgggg gcgctcacgc ccactagaac ttcgttctct agagaaatga
 24961 acaccgcaat tcatggaaa atgttcaaca tcatgttctc atattcatgt atattgacaa
 25021 aatccggctt ataaaaaatg aaaaaccctt cttccgcaca cagtgtctgc cgcatgtatt
 25081 cttagctcaa ctggaaagag cggcgacctt ctaagtcgcc ggttacaggt tcaactcctg
 25141 tagaatacga tagtgatgtc gtatgatgag ggaagcagat ccttgatccc cgcgaagctt
 25201 gtaaccatg ttctctccc ttccggggcg agcgcagctg gctagcgcac tccttggggg
 25261 agcgaagagt cgcaggttca aatcctgctg ccccgagagg tgactggtct tcgaccttcc
 25321 ctttgagtct ggtcttcaa gagtgaact cgcagggagc agatctttat aaaaaaatg
 25381 tgagtcttca tagagacaaa tggataaaaa gttctcatac ataggggttc gagcatgctc
 25441 cccggatacc tgacagccac gacattgct cgcgggaacc aaatcttca tccctggcgc
 25501 cacgagcaat gctcgtgggc gccagaactt cgttccctca gtcataggag tgcctggctg
 25561 ccagcactta caaaggaagc ctagcatgct aggtcccga tgcaaaagaa gtggaacctg
 25621 ttccacttct gatgcttctg gccctcagc ccttccggcc tacggtcctt cgtctgtagg
 25681 ggtgcgagca tgctcgcacc ccgatacttt tagtctctcg atcataagct ttttgtgag
 25741 aaagctgtct gtctcccata caaagagccg aaggtctctc ggccaaaaga tgtgaaagt
 25801 cacatgtgct tgttctctca cgttttccat aaaagaaaaa ataaactatg ttttcatcca
 25861 ctgatcaagt ccaatttttg gctgtttcaa ttctgttatt tctcttagga atttggggga
 25921 tttttttaa tagaaaaaat atcatcatta tgttaatgct aatagaatta atgtttttg
 25981 cagtcaacct taattttgtt ctttttcag tttattttaga tgacttagtc ggtcaaatg
 26041 tcgctttatt tatcttaact gttgcagctg cagaatcagc cattggatta gctcttttag
 26101 tgggtgatta ccgctgaaga ggaagtatag cagttgaatt tatcaattg attaaagatt
 26161 agggttcttc ctatccaata caggcttga cagtgctga gctgtctct cgtgtgctc
 26221 ttgtatttga aagtcattcca ttcatatgga gaaaaacac agtggaccca ggaacaaagt
 26281 tctgacggcc acgagcaatg ctctggtccg tcaggatttg tctccccg agctttgttc
 26341 gcaagaggac aggtcttctg tgtgggaact ctctcatat ctataataa aagaacgaag
 26401 catgctcaa ttatgatcaac tcaccttttt ttcacaattt tctgtgctt gttttttta
 26461 cctcgggttt tatttagttc ttgttaataa tttcttcca aaaatgagta agattctaaa
 26521 agttcgtaat atgaaatga actcaggaaa tgctggacag actgattttt tacttatgaa

26581 agaagaaaaa gaaaaagtaa aaaaaatgag agatgaagct ctcagcgaag ctttcaaaga
 26641 atcgaaacat tttttaaagt aaagctatca aaaaacatct tcttgggtct cgaaaattct
 26701 tcaagatata aacaaaaaaa aacttcaaaa aatgaataaa acctatattc attcagttca
 26761 aaatttaagt tttgccccaa cottaacact tcataacctc aaaacagtga tagctccatc
 26821 ttcgtataaa gcttcaggaa tcgcttcttt aagtaacttct gaaagtgtaa aaacttcgaa
 26881 agatattttt ttttaactcg tcttacttaa aagtttgaca cgtaaaaaag ctggacttgt
 26941 caaaccaact aagaaataaa aaaagttctc gccgcgattc gatgtcatag ggaatggaac
 27001 ctgttccact tctgatgcct tcggcccttc ggccctcggg tcataggggt cgcgagcaggc
 27061 tcgcaccccg atactttcag tccctcgcgc ctcttctccc gagctctgta ttcgatgagc
 27121 ttgtatcgtc gaacacgggc attttttaaa atatgataca catgaatcca accttctctc
 27181 aaaatttagt atcttcgctt caaaaagcct tccaacctga taatcgtctc tttgttcgat
 27241 tagggatttt agcattttta actttaagtt caaaacatat ttttatctat aatgaagaaa
 27301 ctttaattct catttgttcc atttgctttt tttcacctg tcaaaaaaat ctgggggat
 27361 caatcacaca atcttccaat gaaagaagtc agcttattgc tcaagaactt caaaatttt
 27421 ataactaaa agagcaactt gttcttgaac ttatcgaaga acaccaaaaa caactttcaa
 27481 ttcattcttt tatcaagaaa atgggtcaat tttctttaac agaggtgaaa catgtgagtc
 27541 ttcaaaagaa aaaagcattt ttaaagatca aacttcaaga agctatttca actagttttc
 27601 ctttaatgag ttttagcaat gtatcaaatc aacttcaaga agctatttca actagttttc
 27661 gaggagcagt gctggaggaa ttttacaagt caaaaaaaa ctttcaatct gtattagtca
 27721 aacaagctgt tgtgaattta gagcgtctcg caaatcaaaa taaaaataa ttcagcctag
 27781 ttcgagttt ctgggttctc cacatagtga gcgccctcag gcttgtgggt aggcctgag
 27841 gtgctttgtg tttaaaagt ttgattcctt tagagtgtac atcaaaaaata tgataagcgt
 27901 atgccaattg cagcattgag acaaaaaact gtattctttg gtttttttcc aatcaatata
 27961 ataacaatat agactgaaca tatgggtttg atcttggccc agaatgaacg cgagcaatag
 28021 gcttaacaca tgcaagttga atggcctaag aacagaagca attgtaagct aggcctatag
 28081 gaacgggtga gtaaacctg ggaatctacc ataaaagttc gaataaaaatt ccgaatactc
 28141 aaaggcaaaag actgctttat gatgagccta cggaggatta ggtagttgga agggtaaaaag
 28201 cctcccaagc caatgatctt tagctggctt gagaggacgt tcagccacac tagaactgag
 28261 acacggteta gagatttacc gatcgcagca gtgaggaatt tttagacaatg ggcgaaaagcc
 28321 tgatctagct ctattgctg ggtgatgaag gctcttttgg tctgtaaaacc ctttgatgaa
 28381 agaatgatgt tgacagtatt tcaaaaaaaa gtccctggct aacctgtgctc agcagcccg
 28441 gtaatacagg tagggcaagc gttattcaga atgactgggc gtaaaagata cgtagggcgt
 28501 atttgaagt gaaaatgaaa accgagagct tgacttttgg actgttttca aaacaaaaat
 28561 acttgagtat gaaagaagaa ggtagaattt ctcatggaag agtgaaatct gtagagatga
 28621 gaaggaatac caaaagcga ggcagcctc tagttcatca ctgacgctga ggtatgaaag
 28681 tatggggagc gagcgggatt agatacccc ttagtccata cagtcaacta tctcggttcc
 28741 ctctcgttgc aaagctgctg ggggacagct aacgcacaaa ccgaccgctc gaggagtagc
 28801 gccgcaaggt taaaactcaa aggaattgac gggggctcac acaagttggt gagtatgctg
 28861 tttaatgcca aacaacgctc aaaaccttac cagtccctga cattcattgc gaacacatag
 28921 gttatactat gttctattct ttgaacaggt gttgcatggc tgtcgtcagc tctgtctgctg
 28981 agatgtcggg ttaagttcta gaacgagcgc aacctcatt tgaaatctaa cacaggctgt
 29041 gttgagcctt tgggttttcc aaaaactgcc agtgatagac tggaggaagg cgaggatgac
 29101 gtcaggctcg cagcaccctt atgggctggg ctacacgctg actacaatga tagctacaat
 29161 ggggaagcaag agtgcgagct ggagcaaatc ctcaaaagtt atctcaatc agattgtttg
 29221 ctgcaactcg acagcatgaa ggtggaatca ctagttaatc cggatcagca tgcgctgctg
 29281 aattcgtacc tgagccttgt acacaccgctc tgtcagccta cggaaatcaa tttgactgga
 29341 cgtttgaaaa aaaacaatgg tctcatcttt tgccttgaga acaatgaagg tggttctctc
 29401 tacaagaat gtcctttgat ttttcttac tagagtacaa ttttaacta gaggcaagtc
 29461 ataacaaggt aaccgtaggg gaacctgctg ttggattgac tctttcttt acaacataca
 29521 aagacttata ttgaagacaa tgaaacacc gaacctatt cgacctgac tgtgaaactt
 29581 tctctgctcg tatgtattac ttcaaagtga gagacacggt taagtctatt ttgaaactta
 29641 tcaccaatca tgaaaaaagc agtgtgattg gaaaaacttt cctggtgggc gctcatagcc
 29701 caccagaact tctaccgccc caccagcaat gctcgtgggg gccagaactt cgttccctgg
 29761 ggcgctcaca gagtgcgctc ccaccagatc ttcgatcctg gcgccacga gcaatgctcg
 29821 tgagcgcagc aactcgttcc ctccacatta tttttgagct taatttctaa gcatcctac
 29881 gactgaatct gttatctaa gctcagctct aaacaagaaa caaaacaaat atgtttgaa
 29941 attttctctat tctcatttat ctcggtataa gtttaattct atctcttctg atacttgggt
 30001 catctttctc ttttgcaaca cgaatccag atatggaaaa gatctctgca tacgaaatgtg
 30061 gctttgatcc ttttgatgac gcacgagggc gatttgatat tgccttttat cttgtagcga
 30121 ttttattctat cattttctgat ttggaagtga cttttttatt tccatgggct gttcttttaa
 30181 gtcataatga tttctttggg ttttggctta tgatgatatt tttatttctc ttaactgttg
 30241 gattttttca tgagtggaag aaaggggctc tggactggga ataacaact taagggtgtt
 30301 ctacagactg agttttgaa tgaagctctc agtgtttgac ctccatttta tgtgatgaa
 30361 aaatcttgag staaagagcc accgggacaaa gaaggtgcca aagtgtctgt ccgaccccg
 30421 aagcatcaga agtggacat accccacttc tgatgctctg aacctccgat ctacgaattc
 30481 taactgatta atatgctttt agacctctgg ccttatgggg ccagagcttt tttctctggg
 30541 gcgctcacag agtgagcgc caccagatct tccgatcctg cccccagag cattgctcgt
 30601 gggggccaga acttctgtcc tctggctctt tccgggtttc caaacagagg agtgtattga
 30661 cgttttatgg taaccaatca aatgctctga gttattccagc tgctaagaca acccatgca
 30721 gcgacagcgg aaggaaacct tccatccta atctcattaa ttgatcatat ctgtatcgag
 30781 ggaatgccc acgaacccaa atgtaaagaa agagaaaaaa tacagttttg actgccagcc

30841 aaaagatcc tggatccac gagagaaata aaggtccttc aagttcaccg aatggagaaa
30901 gccaacacc taagaataaa atagaacaca tcgagctcat cataatcata ttggcatatt
30961 caccaagaaa aataaagga atcccattg aagaatattc tacattatac ccagctacaa
31021 gttctgcttc tgccctcagga agatcaaaaag gtgctcgatt tgtttcagca agacaagata
31081 caaaaaaaa aattaataat ggaaagagag gaacacaaaa ccacacatct tgttgagctt
31141 ttacaatgtc tgataaattt aaagatcccta cgcatagaag cactgtgatt aaaattaaac
31201 caatagagac ttcatatgac accatctgag cagctgatcg taagcttccct aaaaaagcgt
31261 atttcgagtt acttgaccaa cccgctgtaa taattccata tacacctaaa gatgagatag
31321 caaatagga aagaattcca acatttaaat cagcaagaac aacaccagaa tcaaaaggaa
31381 taacagacca agaaattaaa cttaatgtaa atgttagtac tggagctaat aaaaagataa
31441 cgctgtttgc gttactggga agtacgggtt ctttcacaag gaggtttaat ccaccgcga
31501 aaggttgaaa aacaccaaaag ataccaacta cattedggctc ttttcgctct tcgcatagaag
31561 ccatcacttt tctttctgca agtgttagaa aagctacagc tacaagtaaa ggaactataa
31621 ttccaagaa ttgtgcaata ctcaaaaaca gagcaagaat catatatttt cttttttta
31681 attgttgcca aagacataag accctttggt cttatgtctt tttgaccaa ggtctgagga
31741 tcggtcatag ggaatggaac atgttccatt ccctatgacc gaaagatcga agatctggtc
31801 cccgagagca atgctcgtgg tcgctcaggaa tcgggagcct gaggcttttt tgactctgga
31861 tttcactttt caggagacca aggtccgcag ggaaaagggg agtcgaaccc ctactctcgg
31921 ctttgagac cactgctcta ccattggaac tatttccctt tttgtatccc ctacggggtt
31981 cgaacccgtg ttattgcttt gaaaaaggcaa tgcctgacc actagacgaa gaggacttag
32041 tttcaactat gggcttaagt agatttgaac tactgacctt acgcttata cgcgtacact
32101 cttaccaact gagctataag ccccttttta tgtatatccg tgaacgcatt cctacttttc
32161 tccccgggt tttggttttt atgtttgtgt gtgtttctct tttctcgtc tctttggctc
32221 ttcggcccta cggtcatagg ggtgcgggca agcccgacc ccgatacaaa gagatccctag
32281 catgtgagge tcccagggca tcagaagtgg aacaggttcc acttctgatg ccttcggccc
32341 atcggccgct cggccactcg gccgatcggc ccactcggcc actcggccctt cggccctcgg
32401 gcccttcggt gctctctctc aatccttggc tcaaggtgct gattttgtgt gtttcttcag
32461 aaaaatgttca aatcacgaaag tgaataggat caatatttct ataaagacca gatgaaatcc
32521 aattagaaaa gcttgaacat aaaggtcagc aggggagcag acacttgtga tgagaatgag
32581 agatacataa atatatattc gttttctgga taaaaactgt gagcttattt tatctctttt
32641 gagaaggata taaaaaaaaa gaggaagctg acaaaaacat tgagtcaaga atataatttt
32701 aatatataaa gttactgtgt ttagaatcag tgcttcgagt tgaagattga gggtaaatct
32761 atgaaatttca aagtttgaaa aaaaagacca gaattcacgg acacatattt ttaatccgaa
32821 aaagaaagct tcgatagact tgacaatgag ggtgaacaga agaaatcgcc aggaagctct
32881 ttcaaaagag aagcatctcg gagcaaaaga acaaaaaaca tggtaaagca caagagggaa
32941 tacaagaaaa cccgaaaagt caaaagctac atatacatag gtcttaaaag ctctgaaat
33001 tctgtacat tgaagtgaa gccctgaaaa ggaagcaagg tctttcgac tctgagcatt
33061 tcttaagatg taataggtt tggttaaaaa ccatagatat tcagttgaaat ctgatatagag
33121 gatgagaaa gttacacagca aggaaaatac caatgaaaaa gctctcaaac gaagttcttt
33181 tgaatatatt tttagtttta tcatgtctct acgtgttttt ttgatccgga gagggtggga
33241 ttcgaaccca cggtaaggtg ttcctctacg tcagttttca agactgatac cttagaccac
33301 tcagacatct ctccaaacta tgagcggag caggacttga acctgcaact ttcaggttat
33361 gagcctgacg agttcccaat tactcaattc cgcaatgctg tacatacaat gactttaggt
33421 atggagggac tcgacccacc caccaccgga ttaaaaaatcc gatgtctctaa ccaactgagc
33481 tacacaccgg ataaccggaa gaacacgctc tttgactttt ttgtcacaca aaaatactga
33541 ttttccgaa tcctgectcc cagcagcatt gctcgtgggg gggcagtaact tccactctct
33601 tgccccgcga gcattgcttt ttgggacaag ttgggagatt cagggctgaa agcccaagtc
33661 ttcacgcttt tcgatcacgt gtcccggtgt gtatgtgata tgctgatctt tgacttatct
33721 gtcaacttca ccaaatataa tgtcttgtgt accaatgata gttacaacat ctgcaagcat
33781 gttgtgcttc gacatcata ccagtccttg taagtgcaca aaaccaggag ctctaactct
33841 acaacgatat ggtttgtttg taccattact cacaaggtat actccaaact caccctttgg
33901 tgcctcaaca gcggtatatg tttctgacgc agagacagaa aacctctctg tatagagttt
33961 aaaaatgatga atcagtgatt ccattgattc tttcatttgc ccgctgaa gaggagtaat
34021 cttgcgatca tetgttttga tcaattcctga tggcatggta ttgagagctt gtaacataat
34081 tcgaaagact tgtctcattt ctccatagc aataagatat cgatcataac aatcaccttg
34141 tgtaccactt ggaatttcaa aatcaaaaga gtcataggca tcatagggca tgtttttctc
34201 cagatcccaa gacaccccag atcctcgaag catcacacca gtaaatcccc aatccattgc
34261 ttgttcagct gtcacttttc caatatcgac aaggcgttgt tccatattc tattatgggt
34321 taacatttct tcgagttcat caaggcgtga ggcaaatgtg tgacaaaata aataaatgtc
34381 ttcacagagc ccaattggaa ggtcttgaga cactcctcca gggcgaatgt aagcagcgtg
34441 catacgcgca ccagacactc gttcatagaa ttccatgagt tttctcgtt cttcaaaggc
34501 ccaaaagcatt ggagttagcg ccccccagtc cattgcatga caagtcagag caagttaattg
34561 atttaaaaa cgtgtaattt cacaaaaaag gacacgaata tattgtgctc gaagtggaac
34621 tcttcgctgc aataattttt caacagcaag tgaatatcca tgttcttgag ccatcataga
34681 tacataactt aagcgatcaa aataaggtaa agcttgtaaa taggttttat attcgataag
34741 ttttctcagtt cctctatgca gtaaacgat atgtggatct gctcgttcaa ccacctctcc
34801 gttcatctca agaacaagtc gtaaacacacc gtgagcagca ggatgctgag gtccaaaatt
34861 taaggtaaag ttttttactt tttttgacac aaaggtggat tcaactgttt ttgaaagagc
34921 agctgagggg agtgaagaaa gaagatctga tttttgagat tccatcgcca taattaaatt
34981 ttttcttaa taagtaaaag ctaagactg tttctagggg aaacacggaa gagaaggccc
35041 tttccttccac aaagctttct tgagtctctt gtttttccgg aggaaatggg attcgaaccc

35101 atggtgcaac agaagtccaca cggtgattta gcaaaccaat gccttaagcc tctcagccat
35161 ttctccacag aagcaccaga gtgccggagg gtacacgctt atgtagctca gtcggtagag
35221 cactccccat gtaagggaaa ggtcaccggg tcaaatccgg tcgcaagcta cacaatgaca
35281 caagtgatct tacacaagag attgagagtc tgacagagtg atacatagga tcttctgtc
35341 aaagaagagc gaagttcttc gttcttcttt gacatctatt gaaatcaagg cagtgtgagg
35401 ggcacttctg ttggaagcgc cctcacaana gagtatggag tggatcggac tcgaaaccgat
35461 gacatcatgc ttgcaaaagca tgcgttctac caactgaact acaaccctgt ctgtgtaagc
35521 ctttgatata tctcggcttc ettctgtgtg atgaaatttg taatataaag agggaagaac
35581 cagaggaacy aagatcagcc cccctcaagc attgctcatg ggggccagga tcacagatct
35641 cttgctccga gcagagctca gggagaagga tgctggcttt gaattgtttt cattcttgtc
35701 aatcatctat gtcctatcgt ccaagatctc aagcttggta gggactcgtt ttttctctcc
35761 cgttttctgc tctttataa aatacaaaaa tagaaacatg acagctctct ctcataaaat
35821 tcaaaattct ttgaaagctt ttgaatattg ttttgtttaa aaaagcttgc atcaactaaa
35881 tacaatgatt tcgcatgcat ctccacatat agaaaagcaaa accgtgtctt cgtcgaagcc
35941 atctgcattt tctaagtaca cattgccttt aaaaagttcaa cgatttactg tgattcgttc
36001 tcccataata gataaaaaat tctcgatca atttgaatc cgacagtata aggtcatttt
36061 aaatcatctg ccgagactgt ctccggatga tcttcgattg tttttagaca atctgaaaaa
36121 tgetcgtatt tatggagtcc aaatgcaaat caaactctgt acaagacctc ttggacctgc
36181 aaaaaagaaa gatcgaactt actaatctga ctcatataa agaatgcaaa gatttttata
36241 tcccgagtga agagagctgg tcttggggag caaagctgcg gaggagagtt tctttaaagc
36301 ttcatagggg atggaaggtg ctccactctc gatgccctcg gcccttcggc tcttcggccc
36361 tacggtcata ggggtgcggg ctgcccgtta ccccgacatt ttcagtcctc cgtctgttta
36421 cactagggtt tttggagtat accggataga gaatgaatcc cgcgaaaagt gcgaaatttg
36481 gtatacgcgt tatcttgagg ggttaatgat tctatcatat gggttcaagt cccacttctc
36541 gcagaaaaaa aaagaaaaat atgacaagat ctcttggtaa aggtcctttt attgatactt
36601 tttttatccg tagcttacgc aaagtgtcaa aaaatctctc agcagttggt ctgaaagcca
36661 agcctcaacy ggtggttact ggtcaaaaa catcagtgaa tgcgtccact ctcaaaaaa
36721 aatcctcgga aaaaataact ggacacttcc aaaaaatgtg gtctcgtcgc tcagttatcc
36781 ttctcaattc cttagattct cagtatgaaa ttacaatgg aaaaatattt atcaactctc
36841 gagtgaacta agatattggt ggtcataaat ttggagaatt tgccttaaca cgtaaagaac
36901 cgaaacatcc tactttaaaa aaaaacttaa ttcaaaaaaa aaaaataat atgggtcaaaaa
36961 gctaactcct taagtcttcc ttaagtttt aatcgagact ttgaaatcag ttggtatcaa
37021 gatctcaatt attcaagat attactttaa gaactgacgc ttcgaaaaata tatcctttca
37081 ctttttctat caatgaatgt ctatcaagga caggttcttt tcaaaattt ccttaaaaaa
37141 tgtgtgattt attcttttat taccaataaa aaaaccctaa tgaatagctc ttgaaagctg
37201 aaagcaggag gtggaacaag attcaaaaac agttaaattt ctctgctcc acgaggaaca
37261 ctctgtatac tcccgaaaaa tgcaaaaaaa caggacggta tcttctattt tatggagtat
37321 gcattacaga ggtttgacag tgtgaatcat ctgcacctt cagcatccat tgcaggtttt
37381 ctgtatttga tttcaaaatc gatgcagaca tcagaaaaac gcattggtggc tctgtatggg
37441 tctggtgaca aacagaaggg aaagcatgca aagtcagttc caaaaacttcc ctttctgctc
37501 tcaagtttct cacatgtggg ggctcaccct ctcgaaagctt tcacaggttc taaggtagac
37561 tttgttccgg tgaattaca ttcgagatc cagagtgctt cttttttatc gcagtatac
37621 gcgcaatcct tcgaaaaatc gatgtcttcc cgacagattt tctgtattat tcaaaaagac
37681 attcttctg atagcaaaat acttggaaatg aaaattcaat gttcaggtcg tatcaatgga
37741 gcagagatag cttaaagtaga tcccaaaaaa atgggccaata ctctattaca tacccttctc
37801 gcacgtgttg attttggaa cgcagaagca tatactatgt atggaattat cgggattaaa
37861 gtatggttgt caagtatttc cgaagaaaac ctataaatga aagctttgat atgagggaga
37921 taccttctgt gctgcccccc ccccacgagc aatgctcgtg ggagggcagga ctttgatcaa
37981 agggcgggca acatgttccg cgccttaata ccatcgtttc tgccccccat gagcaatgct
38041 cgtggggggc aggactttgt tcataacaga aaataaaaaa gaaaaaaaca tgttacaacc
38101 aaaacgcaca aaatttcgaa aatttcataa aggaaaaagca acaggaataa acaaaaatc
38161 aagctcttta cattttggaa cctacggaaat caaagtatga gaggctgaa gactgtctc
38221 tcaggttatt gagcagctc gaagagcgtc gactcgaaaa ctcccgagaa atggccaaat
38281 ttggatacgc attttctctg atatcgcagt cagtaaaaaa cctgctgaag ttcgaatggg
38341 gaaaggaana ggagcacctg aatattgggt tgctcgtgtt caagctggtc aaataactct
38401 tgaatctcgt ggagtttccg ctgaattagc aaaacaagcc ggtcttttag caagtccgaa
38461 aatcgcactt ggcactgttt ttattaaaac ggaagacgca ttataaaata tggctctcaa
38521 gatggagctg tagagccctt ggtctctccg gtcactactg ctccaggtgt acaatacatt
38581 gtcgaaaaaa gattgtgtga tcttcggaga ttctactcac atgccaatat cgtattctct
38641 gtatgataca gaagatcagt tggaaactct ttttttgaca gagggtgact taaaaaatg
38701 tatgaattat tttagaaata caaaaatata tatatgatac tcgtaggaac tcaactcaat
38761 gtttcagata actcaggtgc taaaaatcatt gaatgtatta aaattttagg aaaaaaatcg
38821 tcatcttttg ctccaatcgg tgacattata gtggtttcag tcaaggaaac aactggaaaa
38881 agtaaacaga ctcgaaaagt tgaaaaaggg gggatttaca aagccttgt gatagaaaac
38941 aaaaaaggga tcaaacgaaa agatgggagt attcttctgt ttcatagaaa ctcagcactt
39001 cttctttcag aacaaggaat tctatcggga actcgaatc tcggaaatgg aacgtatgaa
39061 cttcgaaaaa agaatacaaac caaagtgtct tctttagctt tacgagtatt ttaaagactc
39121 ttgggatggg gggggaggca atgctctctc ctctgcttgt ccactttctg accagcatac
39181 ttcactttat tacgaatggt cgtctgccaa ttttgtgtgg cgaccatacc tttctttaa
39241 gtatatatat taaactctta tgaatcatag cagactcgaa tggattctg aacatataat
39301 gaaacgtgat cttctctata aattaatttg cccattcgtt ggaaatatct acacacttcc

39361 aaaaatacaa catctttcct taaacgtgag ttcgagcacc acttcaggtt ctcaagcaac
39421 tccaaataca agttctcgcc aaaaaatgat cacattattc agtggactgg aaatcttaag
39481 cggccaataa gtgaaaaaaa catacgcaaa aaaatcgata gcatctttta aactgcaaaa
39541 aggccaatgt ataggaggta aagtactttt acgaaataaa gagatgtatt tttttcttga
39601 gaagtttttc tttatagtgc ttcctaaaat tcgtgaattt gtatgtttac aaccaaaatc
39661 acatgattcg actcatacct tgcctcgat caacttttca gggggaagt tcttctgtga
39721 tcctgaactc tcaaatcaat atgaactctt cgaatcgaac aaaggattta acgcaactt
39781 gattatgaat acaaacact gttctcttca cacagtgtca gggcattctg ttggaaatat
39841 tcttatcagt ttttctgtac cataacaccg aaggcttaag tcatcaaaag ccttcgggga
39901 agggaaacaca caaatataaa aaatacacat gtctaactta gtttttcgcy ataaaaaacg
39961 tagaaatctt ttaagaatat atgaacttca acgtgtactc tataaatcta ttattctga
40021 tctcagatc gaaaaaaat atcgttaga atccattctt aaaaacaaca gtcttccacg
40081 aaatagtagt acagttcgtc tcaaaaatcg atgtatttta acgggcccgt gaaaagcggg
40141 attgagcact ttaaaactct ctcgaatttc attcctgaa cttgacgaa aagggatgct
40201 tcctggtgtc agtaaaagca tttggtaata cggcttaagg cgtttccact caaagaagag
40261 ttggatcaaa gatctgctgc tccagagcaa tgttctgtaag cggcaggatc gaaggaccgt
40321 agggccaag gcatacagaag tggaaacatgt tccacttctt ttgcatcggg agcctagcat
40381 gctaggctcc ctttgtatcg ggtgctgggc aagcccgcac ccctatgatc gaagctctac
40441 ttccttctga aattgaaaca aataaaaca caaagtatgg acaaatatgc aaaaaactta
40501 tcaagcattg aaaatgctca acgacgacaa aaactgtgct gctttcttcc attttcaaaa
40561 aagacacaac atattcttca cttctttttt attgaaggtt atattcgagg atttaaaatc
40621 aatcaaggct ttcaagtgtat tctgaagtat tcagaagatc gagaagaagg aagctcactt
40681 ttaaaaaata aaagaatctc aaagcctgga aaaaaatct atattccggc aaaaaagata
40741 aataaaatca aaagaggctt aggtattctt attctttcca cttcaaaaagg agtgtgtcgt
40801 gatcgagatg caagatcact tcaactcggg ggtgaagtcg tggctgctgt ctattaatta
40861 ttgagcaaaa aagtttagag atatgatccg gtgtgaacc ggccattcat caggcgatgc
40921 gtgcctatct gtacatgac acttgagtaac gaggtgtgt gaagctacgg agacaatgt
40981 ccaagaccgc cgagctaac ctcgggaccg acagcgaagg gagcagtgcc ccagagccat
41041 gggctaggca atgggaacaa ggataacgaa agtgaacacg catattcaga agctctgtga
41101 agacagctct cgactaaagc agcttaaaag cgggaaatga ggtaggggcc cgtcttagtt
41161 atacgggtac cagcttgata caacacctcc ggagtagagc gtccgtccta aagctcta
41221 taaagcattt ttctaagaac ttgtaagcc ccatggattc tcgctgtaaac gagaggacc
41281 ggagcaatcc gtcttatcgt ccagggggta aaggacgga acaaaaagct agtggccatg
41341 agtaatgatg gggatagtcg gttatccgat ggctcgaag agccctaact tgtttcaatg
41401 gcaaaccaaa aagcaggact gagtactgga gggcgtcca cttgggaaac caggcgacg
41461 ttccctatag gacgaaaaga gcgcaatgcc cgaqtaagtg agagcttaaa aactcaaaaa
41521 gaaaggacgg catctctctc ctcgaggggg gtgcttggtc gtccgagccg tatgatcga
41581 aagtatcagc tacggttcta aaggggggaa gtcccgagg accctacctt ccaaggcttt
41641 ctatagggaa aagctagggc atcgcattggc gtgtgcccc ggctctctcc accagattcc
41701 tagacctga tcagagagcc aaagggccta agagccaagg aaaggcttac acaagaagg
41761 gcccttgcc tttttataaa aaatacaatt caatgcaaaa gaacacatct gtacattttc
41821 tcaatcaatc tgggtgtttt cttctcatt ctatcaaaat accctcagat attaaacctg
41881 aaatcgtaga agaaggtttg ccttctagtc atcacactca ttcagctta aagtgcacg
41941 gaccattagg ctgtgttcat gtcaatctcc aaaaaatcga taaatattgt ctttgtttt
42001 tcgaaattat agaaaaacag tcttcccaag gcgtacttga atcttctctg cacgtgtcga
42061 tcaaaagttt ttctgacttc aatcagaaaa agagcggcaa aaaaagcaca catacaatag
42121 cacagcaatc tctggcatt tcaaatgtat tgacctccat gtgtcaaca atgattgaag
42181 gagtctctaa ggtttctgtg ttataacttag aactacatgg tgtgtggttt cgtgcaagt
42241 tacatgaaaa aacactggct ggaaaaacag tcaaatatct tgaatttaaa cttggacaaa
42301 gtcatgatat ttttttgaa atccctgaca atattcaaat tttttcagtc aagcctactc
42361 ttatcggctt gtaggaata gataaagaag gaatcacaca actcgtgca aaaatgcgtc
42421 atcttaaggt gcctgattca tcaaaaggaa aaggaattcg ctataaagat gaaagtgtta
42481 aaacaaaaat tggaaagaaa aaataaacat ggtctatatt ctcaataaaa atttaccgaa
42541 caaaaaaaaa cttcgttatg ctttcaaga aatctttggg attgggcccc ttgtagcaca
42601 acaaatctct gatcaattag gattaagttc acatacttta atggaagaat tgacacaatc
42661 acagattgat actctgttac gtattattaa taataattct atcacagggt cagacttta
42721 aagagtgtgt cagcaagata ttaaaagatt gagtactata ggatcataca gaggtattcg
42781 ccatacacaa catcttctctg ttcgagggca aagaactcat taagattaat agagctctg
42841 acaagaagg aaaaaggaa tgcctattca tgaataccag accatcgaga gaagccat
42901 attatgtggc ttcttgagaa caggctcaca tacacatttt ccaagtttcc gggagctttg
42961 tgtatactgt tcgtgtgact cactctttct ccaaggcca gaggacgaa gtcctggtag
43021 gcgctcacag agtgagcgc cccagcaacg aagttctgct cctcacgagc aatgctcgtg
43081 gggagcagga tcaaaaatct cttgctcga gctctgctt gctcggggac agaaggaaaa
43141 cagtacagaa aagaattgtc ttctttata agatatacag agagttgttt tgccctctct
43201 cttcaagcc cgtattcaac gtggatatgc ttgcaagaa gtctatcttc agaggagtgc
43261 gagcatgctc gcaccctgat atcgtctctt cggctctgtg gccctatggt caagtgcctg
43321 cattgacatc gaaataaata tataaaaaac atggctgcaa taccgtattt tccgaaaaa
43381 aaaaaaacact tcgaactct ctctcttacc caaaacaatc accatattgt tcatattcaa
43441 agtacactga ataatactat ttaacatta acacaacctt cagggaatc tctttttgg
43501 gcaagtgtc gatctgctgg ttttaaaaac tctagaaaat ccacaagtta tgcgtctcag
43561 gctgctgag aaaaacttgc tcaattttgt ctgacaaaa atatcaaaaa gattaatgct

43621 aaacttaaaag gcctagggta cggaaaagaa gcctctttaa aaggtttgca attgaaatgga
43681 attctcattc ataaaaattga agatgtcaat actattccac ataacgggtg tagacttccg
43741 aaaaaaaagaa gaatgtaagc ccacacacat ctgctgcctc gcgagttata ctccgagggc
43801 agcaggtaga aggacaagag atgtttttct cccaagcttt gctcggggga gaaaaaaac
43861 aaaacattcc tcctcgtatc tatcagcgag tctgacaaag actctagaaa gatagggcgg
43921 aggggcaaaag cccttccgcc cacaaatac agttaagagt ctatctctta catcattcga
43981 cagaagatca cacatcgcta cacaacaaga gagaactatg gcagaaggag ctaaaattaat
44041 tggagcaggt tgtgcaacta tcgcttagc aggtgctgga gcaggaattg gtattgtgtt
44101 cggatctttt attaactctg ttgcacgtaa cccatcttta acaaaaaacac tttttggata
44161 tgcaatttta gggttcgctt taactgaagc gatcgcttta ttcgcattaa tgatggcttt
44221 ccttattctt ttcgtattct aattgtgttc gaagggaat agtgtttttt ctattgccct
44281 gtgggactgt aggtccctt ttgaacggat tttccctgtg cagcgggctt atagttaac
44341 tggttggagc ataccgctca taacggtaaa gttgtaggtt caagtcctac taagcctagg
44401 aaattgaaag tgtcgaagat gcggatatga tggaaattgt agacatatca gatttaggtt
44461 ctgatgata acaactcgtg gggttcaaat ccttttatcc gtaagattga tttgatttca
44521 gggggctctt gtcctccctt aatccaacat ggtgtcagat tcttcaatag ctccagcgggtt
44581 agagcggatg gctgttaacc atcaggtcgt tggttcgaat ccatctggg gagatctttt
44641 atttctagaa gaaagcattg cttttcattg tgtgtacac atgaagaaca cagtatcga
44701 aagcctcagc tgctcgtctt cttctgtctc agtctttcat actatgaaag aggaacaag
44761 tctctgtggc gctcacaag tgagcggcca cagcatcaaa gatctcttgc ctctaaaagc
44821 attgctttta ggagcaggta taaggctttt tttttgtgga aactttcatg ttcttcaat
44881 gagctgctct ctgatctatg cacttttcta acatgatctc actgtatttt aagaaaaaac
44941 ctatgtactc tttttctctt tttgtgttt tctgtttttt tctgtttttt gtaaccct
45001 tgtgtagcac ttatggatgc acctgagcca tggcaaatga cattccagaa cccggcgtcc
45061 cctattgcac aaggaattca agatctccac catgatctct tttttttat ggtctcgtt
45121 gtaacctttg ttttatggat gctaggacgt accgtgtggc attttaacgt agtatacaat
45181 ccagtttcaa gtaaaattgt gcctggact gctagttgaa ttgctttgac ttgctttgac
45241 agtcttatct taattattat tgccttcca tctttgtctt tactgtattc aatggatgaa
45301 gtacaagatc ctgcgattac attaaaagtc attgagacc aatggattg gacttatgaa
45361 tattcagatt atacttcagc agatgatcaa tctatcgttt ctgatagcta tatggttcca
45421 actgatgact tagaacctgg acaattacga cttttagaag tagataatcg agtagttta
45481 cctatcaaca cacacgttcg tgtgattatt actgcagcag atgttcttca cagttgggca
45541 atcccttctc ttggtgtgaa aaccgatgga gttcctgggc gtttaacca agccaatctt
45601 tttatataaa gagaaggagt tttttacgga caatgtagt aaatttgctg aacaaccat
45661 ggatttatgc ctatagtat tgagggtgtt tctttagaag attatatctc ttggatctg
45721 agtaataaat gattcaaaaga catacggcgg acaatgagcg aggtcgacac tgacctgctt
45781 ccatatatcc ttgagtcacc gtatggattt gaaagtatgt cttttttgt atcgttgc
45841 tgcacatagt actccttctt gcaagtctct ctcgtgggaa gcacgggagc tcataatg
45901 accccttact gattgtttc cttgagctt tggcatctac aaattgtctt ggagcgtaac
45961 gcttcagcac taagtctat aggggtgcga gcatgctcgc accccgagaa aggtttgta
46021 tgttcatctt ctctatgcaa agctttattt ttagaatgct ttgatcagag atccctgtc
46081 ttgctggcat cattcaatgc tgagtttgat cgcaacacac aatcataaaa tgacttttag
46141 gaaggagact agatgtgaat cccgttacac tatttggttc ctctccccg agcctgctc
46201 ggggagagaa gaccattggt ccttttattg tctataaaaa aatgtattta tgagtattca
46261 aaaaaccctt tttcatttag ttgatccaag tccatggcca ataatggctt cacttgcac
46321 atttagcact actgtaggag caagcatgta tatgcatggt tatacaggag gcaaaagaact
46381 tcttttattt ggacttaca gttgtatgta tctatggtt gtttgggtg gagatgtgtg
46441 tctgtaaggt tcttttcaag ggcaccacac tctttttgtt caacttggct tctgtctg
46501 tatgattctc tttattgtct ctgaaagtgt gttttctgta gttttctttt gggcattctt
46561 tcatcaagt ttgagcggca ctgttgaaat tgggtcgggtg tggcctccaa aaggaattga
46621 tgtattaaat ccatgggaaa tctcttctt aaatactgtc atttacttt ctctggagc
46681 tgcgtgtaact tgggcacatc atgctattgt tgcagggtat cggcaacaag caatttatg
46741 ccttattgtg actattttat tagcatgtgt tttcacagga cttcaagggt tagaatatat
46801 cgaagcacct tttacaattt ctgatggaat ttatgggtct actttctatt tagcactgg
46861 atttcacgga tttcacgtat ttattgggtac tgttttttta accatttgt tagctgact
46921 tattaatttt cattttacac aaagtcacca ttttggtttt gaagcagcag cttgttattg
46981 gcattttgtg gatgtcgtat ggctttctt atttgttga atttattggt ggggagggaa
47041 ctagtcttct cattttaccac atatacagtg aagaacagac ttccattgtt ggagtctctt
47101 gagatattgt gtttccacac agtattaaga acgattttac gaggaggctg acgtctccaa
47161 aggggcgctg gtaagcggcc tctttgcaaa gggattttca aaatgggca taaaaaatgc
47221 gtcagtttat gtttaatatg ctgccaatg aaaaacctt ctatcttatg agcagcttac
47281 aacaccattg agaaatctc tctggaagag tctgaaatctg agagattcga ctcccaaaag
47341 gcatgctctc catttttca actcaaagt ccccttgtt tttttaaaa cagcagagag
47401 agtgcactct tgtgagccac taaaaagagt ggtgctgggt aaaaataagaa aaacagagta
47461 aaaaagaaag tgattccaag aacaatagat tttgttttat cgattggttc gaaagattg
47521 atcggttgat atccggttgc actttcttt tcaaagtaca taatttttat aagtcaata
47581 taataaaaac aactgatcac acttgtaaaa ataccaaca aggccaagat aaaaaagaa
47641 gaacttaagc cagagaaaaa cacataaaat tccccaaaa atcctgcaag agggagata
47701 cctgccatcg aaaaacatgt aatagtaag gtaaaagcta acataggat tgtttctgat
47761 aaaagactta aatcagaat gtatctgagt tgatcagatt cttttttat gatatttctg
47821 ttattttcac tgattaatga agtacttttt gctcgtaaa ataacagag agcaaaaaa

47881 ttcagagtca tcataatata gataataaga tacacgatga cagattcaag tccttctcta
47941 ttcgctgcta ttaagcctat aaacatatat ccaatatgtc caattgaact ataagcaaaa
48001 agacgtttta tttttgttg agataaagcc gaaaaagccc cgataaacat tgaagcaata
48061 ctacacagga caagaatctg ctctaaaggt gctgatccat tttcaatccc tgtgccaagg
48121 gtcctcaaaa gtcctccata atagaaaaga cgaataaata atgctaaaag agcaatctta
48181 gtcgctactg aaaaatattg agtgactgaa gttggtgctc ctccatacac gtcagggtgcc
48241 cacatatgga atggagctgc tgtgagtta aaaagtaatc caacggcaat aaaagcatt
48301 ccaagaagaa ttcagctgg atgattaata tcaactgatg cacataggag agttaattct
48361 tcaaaagtgg tgattccagt aaatccataa ataagagaaa taccgaaaag taaaatacca
48421 gatgagaaaag ctccgagtaa aaaatatttt aaaccagctt cagtcgaaaa ttcagaattt
48481 cgcttgacag cagctaatac ataaaagcaa agactttgaa attcagctgc taatagagt
48541 gataaaaaat catatgaaga tatcataaaa agcatyctac atggtgagag caagattaaa
48601 atatgatatt caaatgcgtt taaggattct ttttgagaa aatcaaaaag gatcataaca
48661 agcaaaaatc aactgagaag gagtaaaacct ttaaaaaaaa gagtaaaatc atctaaaatc
48721 agacttccaa aaaagatttc tctttggttt aatggattgt taaataaaaag gatgaaaagta
48781 catctcaaa taagaaggct taataacca acgttttctg ttaagatcgg acgaaaaaga
48841 ttgttttcaa ctttattctc gactgacta taatgaacac catacataag caaaaaataa
48901 gatcacataa tgaggaaaat ttcaggaaaat aaagcgagaa aatcattttt aagaacgaa
48961 gtagcaaaa catcttctgc agagtgcaca tgctgagaaa ataaaataatc ttgtgttttc
49021 atcttacatt ttttttaat acattctgaa gatcaaaagg atgcgagtat gctcgcacc
49081 cgaacaaaag ggagcctagc atgctaggct cccaatgcat cagaagtggg acaagtcca
49141 ctctcgatgc attcggcctt tcggcccttt ggccctatgg tectcgatc cttgccccaa
49201 aaagcaatgc tttttggggc aagagatctt tgatcctcgg cgagacttgg aagactgtat
49261 atctcactc caagatctgc agaatttttg tgtgtgtat gaacctttgt tcaagtgatc
49321 atccaggaca caacagagaa gaccttttcc attgagtct acgatcact tataactcg
49381 atctctgctc ccatgagcgt tgcctatggg gagcagaact ttcgatcctg ctccccagga
49441 gcattgctcg tgggaagcag gatcgaaaag tctggggggc ctccacacag gagcgcctat
49501 cagaacttcg atcttagaga ttacttttga aaacttgcca atggttgagc gacagaagta
49561 tgcatagtat ctaagaaaat ttcaggataa atccccatcc atagaacagc aatccaaaa
49621 ggaagaaaaa taaagaactc tctctattt aaatcagcga agtgatggat agaatgagt
49681 ttaaagtttc caaagactac tccgttataa agccacagtg aatatgcagc tccataaccc
49741 ataccgctac aagcgagtgat tgctataaat gtgtttgttt ggaaaaatcc agctaaaatt
49801 aaaagtccac cgataaaaact acttgctcca ggaagactaa tatttctcaa tgaataagat
49861 aagaagaaaga atgaaaacag aggcatacaca tgtacacatc ctccatagta tttcacaagt
49921 cgagttttat gtcgatcata gagaactcca acacataaga aaagagcggg agcgcaaaa
49981 ccatggctta acattaaaag aatacttctc tcaattctct gtgtatctaa gctaaaataag
50041 cctaaagtca catagcccat gtgcgctact gaggagtacg caatgatatt tttcagggtca
50101 atttgtcgta atgtggttaa tgaagcataa atgatcccta agacactgag ttgtatagatt
50161 aaggggtgaa aataaatgct tgcatttggg aacataggaa ttgaaaatcg taaaaatcca
50221 tatgtcccaa gttttaaataa aataccccgt aaaaataactg atcctgctgt aggcgcttca
50281 acgtgagctt ctggaagcca gatatgtaca ggaatcattg gaactttcac ggcaaaaagat
50341 gcaaaaaggc atgaaactaa aagaatttga cgtttttcac taaaatctgt agtgagaaga
50401 atcatgacat ctgttgctcc agtttgaaag tatataaata aaagagccag aagcatcaga
50461 actgacccga tcaaggtata aaggaagaat tgaaaagctg cgcggatttt tcttctcga
50521 gatccccaca ctctataat aataaacata ggaataagca cgctttcaaa aaagacgtaa
50581 aaaaagtata gatcaaggat acagaacact ccgagcatac aggtttcaag aactaaaaaa
50641 gcaataaaat actctttcac gtagttttga aactcagacc aacctactaa taaacaaata
50701 ggaactaaaa aagtagtcaa taatacaaaa aagagagaaa tcccatcaat tccaactgca
50761 aaatcgatat ttggaatcc agaacaaaag gcagttcgtg aagaaaggct ttcgacaaat
50821 tgaaatctgg aagatgaatt atcaaaactga atccataaaa aaagtatag ttaaaaattg
50881 aggggggaaa aaagaagacc tgtgatacga ataactttgt tttgacgagt tggataaaag
50941 agtaaaaaa agactccaac caaaggaagg aaaaagaatgc tcaataaaag tgaaggcatt
51001 gtatatttct tttttttaa tatctctgtt tgagggggac tcacatgtga gcctcaaaa
51061 caatttgaaa cgcgtcctaat gctcatgtgc tccacccccg gtgtagaagt ccagcgaaaag
51121 atgatcaaa gaccgaagga aggaagttct jggccccacg agcatagctc gtgggcccga
51181 ggatcaaaaga tctgttcccc gcgagctatg ctctgtggccg tcaaggtatcg ggggtcgcagc
51241 atgctcgcac cccatgatc gaggatcgaa gggccgaagg catcagaagt ggaacatggt
51301 ccacttttga tgcaaaagaa gtggaacatg tccatgtct tttgtatcgg gaatggaaagc
51361 agttccattc cctatgacca gaaaaggtaa tgaagaactt ggttctctc ccagaagaat
51421 tctaagacct tgagctttca ctattgactt gatttttata tatttttttg tgtacataaa
51481 aatgtgaaa tgagaagatg aagaagaaa attcggcttt caatccaaaa agaaaagaca
51541 tccccataac cgacgatggc aataaaaaaag gttaaaccaa ttaacatcac aaccgataa
51601 tggtagatgg caccactttg taattttatt actgaagata aaaagctgtc cgggtgtttc
51661 gatccaaaga gttgaatcac accagaaggt ccgataagtt ctaacacacc tttatcaatg
51721 gttttgaaa tgatgtcata ccccattttt aacacaggat atcctataaa atcagtatat
51781 actttatcaa aaagccatcg tttgtttaa aaagtataaa gagccatccc gatagaaactt
51841 gttttcatag aatagcgaag gtgtggcttc gcaatattga tttgatagc gatcatagcc
51901 ctaaaaaaac taagatcaa aggaatcatt ttaataaatt gagggatata tctgactca
51961 attaaaagt cattttctgg aaattgaaaa agcgcatttc ccaaaaatgg agttccaaaa
52021 ccgatcatca tatcttctgc tacatatcca acaaaaaatc tccaatagc gagaagaatg
52081 agtggaaatg ccattaaaat tggagcgtca tgcacgtgtt tgactattg ttaaacgca

52141 tttgtcgggtg caataaaagt gagatataag agtcgaaatg aataaatga agtaataag
52201 gctgaaagac ttccctaacca ataggcaaaa tttccgtgaa tggatatatt ggcataggctc
52261 acttcaagga tgacatcttt tgaatagaat cctgttaaga atggaaatcc gatcagagcgc
52321 agagatccaa tgacctcatc agaatatgta aaaggaagca attttgcaag tcctcccat
52381 tttctcaca cttgtctcacc tgataaggca tgaataactg aacctgctga taagaacaga
52441 agagctttta aaaaagcatg attcattaaa tgaaacacac caactgaata ttgtgaaatt
52501 ccacacgcaa acaccatata accaagctga ctacatggtg aatacgcgat gacacgcttt
52561 aagtcattct gaaccacacc agttgttgca gcaaaagaaag aagtcatggc tcctataaaa
52621 gtgacaccga taagggcaag aggggcatgt tcaaacagg cagaacaacg agctaataaa
52681 aagactcctg ctgttaccat ggtagcagcg tgaataagcg cagacaccgg agttggacct
52741 tccatagcat caggaagcca tgtatgaagt cctaattgag ctgactttcc aatggatccg
52801 ataaataaaa aaagacagag taaggttaag gtgtcaaat caaaggtaaa aatgaaaaat
52861 gttgtatctg caaactggtg tgtgttgca aacacagtg caaaattgac acttttaaat
52921 aaagaaaaaa tccctaaaaat tccgagagca agtccaaaat caccaattcg attcataatc
52981 atcgtcttga tagctgattt atttgcttgt aaacgggtaa accaaaaatt aatataaagg
53041 tacgaagcga gaccaacacc tccccacct aaaaaaagt ggataaaatt gtcggctgtt
53101 accagcataa gcatgaaaaa tgaaaaaatt gaaagataac tcataaagcg tggagatgc
53161 ggtactctctg acatataatga taaagagtag agatgaaacta agctacttac aaaggctacc
53221 accacaagca taactacagt taaactatca aagaaaaaac cccagcttgc atcaaacatt
53281 tctgatgtaa tccatggcgc aaattctata taacaaggty atccacatag accgacctca
53341 taaaaagcta ctgcagatag acatgcggtg agacttacac atcctgtggg aataagtga
53401 gatccgggtt atcctagaaa tctgcaaaa agactggcag atagaaaacc aaaaagagga
53461 agagagacaa taagaaggtg catataattt tttcttttg atgaactgag tcaaggggt
53521 gcaccatag gtgcgctgt ctgcacaata caacgcgatg accttctccc ccgagcagag
53581 cttgggggga gaagacctcc gatcctaat acaagagatg gttcccaat ctctgttct
53641 atagcttttg cttggggaaa tcaagcccc gatgcttga agctcttgg tcttatgta
53701 tagaaaagat gtcatacact tgtccaagag tggagtgaa ccaactgacac caagattttc
53761 agtcttctgc tctaaccaac tgagctacct ggacctaat atacgagcat acacacaaaa
53821 tatcatcttc agaaatggtc aacccgagac tccagcatca gaataaattc atattcttg
53881 ttaatttaaa cttaggggta tgatcaaaag atcggaggac taaaagtatc ggggtgtag
53941 catgctcgca cccctatgat cgaagatcg aaggatcgaa ggaccgtaga gccaaaggca
54001 tcgaagtggt aagatgttcc acttctttg catcggggag ctgacatgct aggtctctt
54061 tgtatcggga cgtcactga gtgagcttc agcttctca tgacgaaaag gccgaaattg
54121 aaaaaagtggt aactcttgg catcttctag agaaaaagcc tctatctggg gaaatttaaa
54181 atttctctct ttatgaatga caataaaatc acttttact caaatacaga gttacctca
54241 tacatatttc acgatgtgag ctacgaatat ataggtcata gggaaatgaa ctgttccat
54301 tcaagatacc tgacggccac gagcatagct cggggggaac agacttttga tctgcccc
54361 ataaggtatc cttagggggt cagaactctg tccagcccc ccatgagcaa cctcatggg
54421 gggtgatctc tccatctgg gggcgtcac agagttagcg cccaccgat ctccgacct
54481 taataataaa tacaatcaaa ctatgaatct catctcaaaa gaacttttaa atacaaaagc
54541 ccatgtgggt cataaaaaat gtaatctcga aacttcacce tatctctag gaagcagagc
54601 ttcacaccca ctgaataaaa catcttggat tcccaaaaat aaagatcaaa agtccgggtg
54661 ggtcatctct gaaaaagcac gtcagcacag tgaaaaacca aagaccactg gacactgat
54721 ggagatcgag aaaaaacttc gaatcacctt ccgcaaaaaa cgcgcactct tcaactttt
54781 taacctcgaa atgactttat tttgtcttcc tcaagtgat gatttctac aaaaacttat
54841 cacactcaaa accagcttta aaacatcgaa gacctctc tcttcaagcc caagactcc
54901 tcatatctta atcatggctg gaagaggaaa gactccatcc ggacattttc atttagatat
54961 cgcaaaagatt ctttcaaaat ggaatacact tggaaatcca caggctcaac catcaagcaa
55021 aacttcaatt caaaaaccaa ttttgagctt tgttcaagaa aaatgggttg gaggatgct
55081 tacaatttgg aaacaagtct ctgaatcatt gaaaatctat ccaaaaattt aatacaaat
55141 cgaagatttt ttgaaaaaac ataaattacg ctttctctc tatgaaaaat ataataagcg
55201 atacttagga ttaaaagaaa taagtacatc tttaccagat gtgatcatta tcaactatcc
55261 agaagacaat gacattttta ttcaggaagc ttttattctt aagattccaa ttattgctt
55321 tghtaaatag gatctgccta aaaaatctatt acactctata cagatccta ttcctgaaa
55381 caacgcactc ccatatttta tttattttg tttgaatag cttcttagg cactttctac
55441 gaagaaataa cactgatttt gggataaaaag gattcgaacc tttgcatatc ggaatcaaaa
55501 tccgatgctt tccgcttgg ctatatccca gaagcttgtg cttggcacac agcaaaaaat
55561 gaagatctac tgttcgcaa ctttttttaa aggtccgag cagatttgaa ctgccatttc
55621 aggatttgca atccaaaaca ttaccattat gttatcgaa cctgtttatg gctttcagg
55681 gactgaaaag actttcacat taaggagccg agcctcggct cggggggcgg gggaaaggag
55741 ctccacattg aagcctctta gtacaaaca aacctcgtgt gtaagacttc tgaagcttc
55801 aaacctcacc tatatgcttt tgaagataaa gctgaggggt gtagggcgaa ataatggaa
55861 ttgaacccat ctttccagag tccagctctg acactctaac caactaagt atattcgccg
55921 tgaatttaaa agaccgacc ctccgaccga aaggccaagg gcccttcgat gttctgagaa
55981 gatgtctgag cgggtcaaaag caacagactg taaatctgt gatattatct tctgaggttc
56041 gaactctact ctctcaacc acaggctcac gcagtgagcg tctccccctc tctgccccctg
56101 actggctcta tgcagtgaa gcaggttgag ctttcatcca aagaatgctg tgagtcaagg
56161 cattcggag actcagtgta ttaaacacag gaagatgcac tgtttgaaaa aaatgagcat
56221 agacggaagt ttctgaaaga ccaagtcttt ttttgatgtc aaagacattc aaaagagtg
56281 agtgattgtt ttttcacaa ctatccagtg ctacatgaaa aaaaagattg acatgaagag
56341 gagcatggac actttgtgtt ttcaactctt tgggtggaac tgtgtggtg atgatttgg

56401 aaaaaacctc taaagctgca ggtgacgtac aatttataaaa acaacagcga cctttcatat
56461 cgaaaaacaa agccgattga ctctgcgtgc gattgtcttg aacctgtcg tgaagagatt
56521 ttctgagttt tgaagggata atctgaatat tcatggttga accatggtga agatgaatgg
56581 tatttttaat gtgcttccaa tcattggctg aaatattggt atattgaaag atataggtat
56641 atgggtacaa ttgtaccttt tgcaataatt tttttgaaag aaagtgttta cgaaaaaact
56701 gaggtgacaa agacatatat ttttttttt tcttataaag agcaaaactc gcgttaaac
56761 tgagaggctt cttccaatga aaagaaaaag cttgaaaact aagctctgaa atctaatttt
56821 tgtttctctt aaggaccaga agatccaaaa tctctggccc aaaaaagcaa tgcctttttg
56881 ggccaagagt cttctcttcc cgagctctgc tcgggggaga agaattgacag cctctgttcc
56941 gcgagcaaaag ctccaagaag aaaaagtggg tctttcttca gtaaaagaga cggatgata
57001 ctctctgatt cctgtgtcct tcaggcaaga gaatcatggt cttccgctct tggccgtcat
57061 tactttgttc ctgagctccc gcgagcatag ctgcgggaa ccagatcttt aatcttctct
57121 aagtcctccc caagcaatgc tcgctgggtac cagatctgtg atcatagggg tccgagcatg
57191 ctgcaccctt tatacgaagg gagctctgca tgcgaggtct ccaatgtaga aaaaagtggg
57241 acacgttcca cttctgatgc cttcggccct tcggccttct tttccacgat gaaaacggac
57301 tctcaagatc gaagatagac ttccatgagt ttgtctatgg aaaacagcac atagccctg
57361 ttaactctac ttaataaata aaaaaaagaa atatgaactc aaaaagcagat tatgtgattt
57421 ctaagttaga tactcttcta aattgggcac gaaaaggctc tctttggccc atgacctttg
57481 gattagcatg ttgtgctggt gagatgatgc atagtgcagc aagctgatac gatttagatc
57541 gatttgggat tattttttag ccaagcccga gacaatctga tgtgatgat gttgcaggaa
57601 caataaccaa taaaaagctc taagtccttc gaaagataa tgatcaaatg cccgaacctc
57661 gttgggtgat ctcaatggga agttgtgcaa atggaggtgg atattaccat tattcatatt
57721 cagttgtctg tggatgtgat cgaatcgttc ctgttgatat ttatgtccca ggatgcccac
57781 ctaccgctga agcactcttc tacggattac ttcaattaca aaaaaaattt aaactgactc
57841 aaaaactcag tgtttgggat aatcgataat agccttctcg ccgcccacaa ggttacttg
57901 tgggctggcag aactctgttc cagctcccca tgagcaacgc tcatgggggc gctcactttg
57961 tgagcgtccc ccagctacttt gttccatcaa gtgaagaaag tcttttaata ttataaaaaat
58021 attctgtcta taggctcgaa gactggcccc tttgagagca gagttcgcag gaagcgaagc
58081 ccacgagttt tsaggattct tttccaaggg gtgcgagcat gctcgcacct cgatatcttc
58141 gatcctggcg cccacgagca tagctcgtgg gcgccagaa ttcgttctct cgatcattga
58201 aagaccgaac accaaagtat cggctcacttc ccaaacaaaa aaacaaaaaa aatatgagat
58261 atcttctctt cgtaaagcaa caggacaacg ctctctgcgc aagtactcat gtattgatag
58321 agaaattttc aaagtctctc actcttcttg ttccgaaatg gattcaaatg tttctcttcc
58381 aaaacgatga gcttgtgatc gtgatctatc ctgaatacgt gctcccattt tttcaatttt
58441 taaaagatca tacaataaca caattcaaaa tgctttatgga tatcacagca gttgatattc
58501 ctctctcgca aaaaagattc gaaattgtat ataatttatt aagtcctcaa tataactctc
58561 gaattcgagt aaaaacttat ggagatgaaa ttaccccatt tgaattctta gttggcatct
58621 atagttcttc aaattggtt gaacgcgaag tgtgggatat gtttggaaat tttttgtgga
58681 atcatctctg tttaaagcga atctcgactg attatgggtt tcaaggtcat cctctccgta
58741 aagactttcc tttaaagtga tatgtcgaag tacgctatga tgattctgaa aaacgagtga
58801 tcaactgaac tttagaattt actcaagaat ttcgttcttt tgatttcgca aaccttgggg
58861 aatcaccga aaaaaaata taatactctg tctgtagaga gagtctcatg acctagagac
58921 tctcaacctg atcggctgtg tggacctttg atcaaaagatc tgccttccaa gagcttgtgc
58981 ttgcggaggt tggctgctt ttcacatgct gtgcgtgtga aaaaaaaga cttttgtgct
59041 ctggttgcaa tccaacaaaa acttttgta ttaataagag aattgaaagt tattctatc
59101 aagagacaaa ctgctccgcg agactgacgg gagtacgaca cacggaatc aaaaagtcag
59161 aaaaccagag ttctggaatt gaccttgcga gtgtacgata aaactaaaaa gaatgaaact
59221 atggctgtag cagcaggaga attatcgaaa ttactagaaa aaagaatc aaattattat
59281 actcaactca acatggatga aatcggacgt gttttatctg taggtgacgg aattctcgt
59341 gtatacgggt taaaaaaat tcaatcagga gaaatgggtg aatttgcaa tggtttaaaa
59401 ggaatgctt taaacttaga aaacgataac gttggggtag tgtgttttgg gaacgacagt
59461 ggaatcaaa aaggtgacat tghtaaaact accagagcta ttgttgacgt tccagttgga
59521 agaggaatgc ttggtcgcgt tgttgatgct ttagggagcg caatcgatgg aaaaagtgct
59581 ttaaaagatg tgacacgacg ccgtgtttaa gtgaaagctc ccggaattat tgcctgtaa
59641 tccgtacacg aaccgatgca aacagggatc aaagctgtgg acagtttagt accaattggg
59701 cgcggacaa gtgaacttat tattggagac cgtcaaacag gaaaaaccgc gattgcaatc
59761 gatgctattt taaacaaaa aggaattaat agcggaaagt atgaaaagcaa aagttatc
59821 tgtgtgatg ttgctattgg acaaaaactg tcaaacagtg ctacagctgt aaaaacttta
59881 atgaatgctg acgctttaa ttaaccattt attgtagccg cgaactgcgtc tgaccagcgy
59941 cctttacaat tcttagcacc atattctgga tgtgcaatgg ctgaattctt ccgtgataat
60001 aatgtaattt ctttaattat ttatgatgac ttaagttaaac aatctgtggc atactgtcaa
60061 atgtctttac ttttaagaag acctcttggg cgtgaagctt tccctggaga cgttttctat
60121 cttcactctc gtttactcga acgtgcagcg aaaatgtctg ataaagcga agcaggatca
60181 atgactgac ttcctgtgat tgaaaacaaa gctggagacg tctcagccta tctctacc
60241 aatgtaattt caattacaga cggacagatt tcttagaaa cagaattatt ttcaaaagg
60301 atctgacctg cgatcaactg aggactctct gtgagtcgag taggatcagc tgcctaaatg
60361 aaagctatga aacaagtgtg cggatcttta aaacttgaat tagctcagta tccgagaagt
60421 gctgcttttg ctcaatttgg atctgattta gatgcttcaa cacaataact ttaaataga
60481 gaagctcgtt taacagaggt tctgaaacaaa gggcaatatg ctccagttcc aattgaaaa
60541 caagctgttg tgatttttgc agcgacaaaa ggatacttag acaaaactca cattgctgaa
60601 attttacctt ttgaaaccgc acttttaaaa gaaattgatc cgtcaatttt tactacaatt

```
60661 cgcgacgaaa aagtattatc tccagcgaca agtactaaat taagccaatt tttagataaa
60721 tttacagcga actttatagc aactcgtcag taatctgtgt agacacaacg agggggaaca
60781 ctgttcccc ttgtatacaa tctccactat gggtgcttaa ctcagttggt agagtacctc
60841 gtttacaccg aggctgtcag cggttcgagc ccgttagctc cca
```