**SEQUENCIAMENTO E ANÁLISE DO GENOMA
CLOROPLASTIDIAL DE CANA-DE-AÇÚCAR**
(Saccharum officinarum L.)

TERCILIO CALSA JUNIOR

Engenheiro Agrônomo

Orientador: Profa. Dra. **HELAINÉ CARRER**

Dissertação apresentada à Escola Superior de
Agricultura “Luiz de Queiroz”, Universidade de São
Paulo, para obtenção do título de Mestre em Ciências,
Área de Concentração: Fisiologia e Bioquímica de
Plantas.

PIRACICABA

Estado de São Paulo – Brasil

Outubro - 2001

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Calsa Junior, Tercílio
Sequenciamento e análise do genoma cloroplastidial de cana-de-açúcar
(*Saccharum officinarum* L.) / Tercílio Calsa Junior. - - Piracicaba, 2001.
76 p. : il.

Dissertação (mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2001.
Bibliografia.

1. Biotecnologia vegetal 2. Cana-de-açúcar 3. DNA do cloroplasto 4. Genoma 5.
Melhoramento genético vegetal 6. Sequência de DNA I. Título

CDD 633.61

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

Do fogo do céu faz-se a vida na Terra.

*Acima de tudo está Deus,
depois dele a terra está,
porque nela tudo se fez,
nela tudo se fará.
Porque dela, sim, emana a paz,
A terra eu sempre irei amar.*

*À Maria Clara,
meu tudo, sempre,
e à Giovana, luz,
à elas meu amor e vida,
sempre.*

AGRADECIMENTOS

Agradeço a Deus, pela dádiva de viver, pensar e fazer.

Agradeço à minha esposa Clara, pela inspiração, incentivo, participação integral e apoio incondicional em todos os momentos, na mais sublime e perfeita tradução do amor.

À nossa filha Giovana, pelo sentido da vida renovado e multiplicado infinitamente na plenitude da família.

Agradeço aos meus pais, Tercilio e Ilma, pela vida e por todos os caminhos que eles indicaram e, muitas vezes, abriram para que eu pudesse continuar. À eles minha total gratidão, e a esperança de seguir este exemplo para com meus filhos. Aos meus avós, Umberto, Adelaide, Elia e Palmira, por me ensinarem os valores da terra e da vida. Aos meus sogros, Sérgio e Maria Antonia, e aos meus irmãos, Tania, Tarcisio, Isabel, Marcos, Cíntia, Carlos, Daniel e Danilo pelo incentivo e apoio.

À Profa. Helaine Carrer, agradeço por estes anos de contínua orientação, exemplo, apoio e condições para meu desenvolvimento pessoal e profissional na Pesquisa Científica.

Agradeço a todos os integrantes do Laboratório de Biotecnologia Agrícola, do Departamento de Ciências Biológicas da ESALQ-USP. Aos Professores Otto Jesu Crocomo, Murilo de Melo e Luiz Antonio Gallo, pela orientação e amizade em inúmeros momentos. Aos funcionários Maria Solizete G. Silva, Enio Tiago de Oliveira, Antonio F. C. do Amaral, Romeu A. Rocha, agradeço pela amizade e pelo auxílio nas várias fases deste trabalho.

À Valentina de Fátima de Martin Orelli, agradeço pela participação na rotina da pesquisa, disponibilização dos materiais, sugestões, críticas e pela amizade..

À Dirce Maria Carraro, pela co-orientação e participação no desenvolvimento deste trabalho. Ao João Paulo Kitajima e Alexandre C. Barbosa (Instituto de Computação-UNICAMP), pela participação com as ferramentas de bioinformática essenciais à realização deste trabalho.

Agradeço a todos os integrantes do Laboratório de Biologia Molecular do Laboratório de Biotecnologia Agrícola, Depto. de Ciências Biológicas – ESALQ-USP, pelo convívio em equipe, pelo auxílio em várias situações, pelas sugestões, críticas e opiniões em prol do desenvolvimento deste trabalho. Aos amigos Patrícia, Simone, Adriano, Henrique, Renata, Nelson, e Maria Lúcia pelo convívio alegre. À amiga Leandra M. Scarpari pelo convívio e auxílio.

Em especial à minha esposa Clara, companheira também no trabalho.

Ao grande amigo Irving Joseph Berger, pela experiência partilhada na pesquisa com sugestões sempre interessantes.

Ao irmão Matheus R. Benatti, agradeço pela participação vital neste trabalho, pelo aprendizado recíproco, pelo convívio, amizade e exemplo.

Aos amigos do Projeto Genoma-FAPESP/Grupo EQ, Adriano, Reimar, Raquel, Letícia, pela ajuda em vários momentos deste trabalho.

Aos amigos do laboratório do Prof. Ralph Bock, do Institut für Biologie–III, Universidade de Freiburg, Alemanha, em especial Martin Hager, pelo treinamento necessário à realização deste trabalho.

Agradeço, enfim, a todos que, direta ou indiretamente, colaboraram para a realização deste trabalho.

SUMÁRIO

Página

LISTA DE ABREVIATURAS E SÍMBOLOS.....	viii
RESUMO	xi
SUMMARY	xiii
1 INTRODUÇÃO	01
2 REVISÃO DE LITERATURA	04
2.1 Cultivo e evolução da cana-de-açúcar (<i>Saccharum officinarum</i>)	04
2.2 Origem e características dos plastídios	05
2.3 Interações Intracelulares	06
2.4 O DNA cloroplastidial	08
2.5 Regulação gênica nos cloroplastos	14
2.6 Transformação genética de cloroplastos	16
3 MATERIAL E MÉTODOS	23
3.1 Material Vegetal	23
3.2 Isolamento de cloroplastos	24
3.3 Isolamento do DNA dos cloroplastos	28
3.4 Bibliotecas de fragmentos de ptDNA	29
3.5 Sequenciamento e análise dos clones	31
3.5.1 Isolamento do DNA plasmidial dos clones	31
3.5.2 Reação de sequenciamento	33
3.5.3 Purificação dos produtos da reação de sequenciamento	34
3.5.4 Sequenciamento	35

3.5.5 Montagem do genoma cloroplastidial	36
3.6 Finalização da montagem	38
3.6.1 Fechamento dos <i>gaps</i> de sequência	39
3.6.2 Fechamento dos <i>gaps</i> de qualidade	42
3.7 Análises comparativas e Anotação dos genes e <i>ycf's</i>	43
3.8 Análise comparativa de edição de mRNA por RT-PCR	43
4 RESULTADOS E DISCUSSÃO	46
4.1 Material Vegetal	46
4.2 Isolamento de cloroplastos e do ptDNA	47
4.3 Construção de bibliotecas de fragmentos de ptDNA	49
4.4 Sequenciamento e análise dos clones (montagem do plastoma)	51
4.5 Anotação dos genes e <i>ycf's</i> (mapa do plastoma).....	56
4.6 Análise comparativa de edição de mRNA	59
5 CONCLUSÕES	63
ANEXOS	64
REFERÊNCIAS BIBLIOGRÁFICAS	68

LISTA DE ABREVIATURAS E SÍMBOLOS

5'-UTR	região 5' não-traduzida do mRNA
A	adenina
<i>aadA</i>	gene codificador da aminoglicosídeo-3-adenililtransferase
AT%	porcentagem de bases nitrogenadas adenina e timina
ATP	adenosina-trifosfato
BADH	betaína-aldeído-desidrogenase
<i>Bt</i>	gene do <i>Bacillus thuringiensis</i> codificador de toxina inseticida
C	citossina
C ₄	fotossíntese com fixação inicial de CO ₂ em composto de 4 carbonos
cDNA	DNA complementar ao mRNA
CTAB	brometo de hexadeciltrimetilamônio
ddH ₂ O	água destilada, deionizada (Milli-Q) e esterilizada
DEPC	dietilpirocarbonato
DH5-α	cepa de células competentes de <i>Escherichia coli</i>
DNA	ácido desoxirribonucleico
DNAse I	desoxirribonuclease I
EDTA	ácido etilenodiaminotetracético
EST	região de sequência expressa (<i>expressed sequence tag</i>)
<i>fiber</i> -FISH	hibridização <i>in situ</i> por fluorescência da fibra de DNA
G	guanina

GTP	guanosina-tri-fosfato
IPTG	β -D-isopropil-tiogalactopiranosídeo
IR	repetição invertida
Kb	kilobases
Klenow	Fragmento maior da DNA-Polimerase I
LMP	baixo ponto de fusão (<i>low melting point</i>)
LSC	região longa de cópia única
mRNA	RNA mensageiro
MS	meio de cultura MS (Murashige & Skoog, 1962)
NADH	nicotinamida-adenosina-di-fosfato
ORF	sequência de leitura aberta (<i>open reading frame</i>)
PAT	fosfinotricina-acetiltransferase
pb	pares de bases
PFGE	eletroforese em gel de campo pulsado
PPT	fosfinotricina
PS I	fotossistema I
PS II	fotossistema II
ptDNA	DNA cloroplastidial
rDNA	sequência de DNA codificador de RNA ribossômico
RFLP	polimorfismo do comprimento de fragmentos de restrição
RNase	ribonuclease
rRNA	RNA ribossômico
Rubisco	ribulose-1,5-bisfosfato carboxilase/oxigenase
SSC	região curta de cópia única
T	timina
TAE	tampão de eletroforese Tris-Acetato-EDTA
tRNA	RNA transportador
U	uracila
V/cm	volts/cm
xg	múltiplos da força gravitacional terrestre (rcf)

X-Gal	5-bromo-4-cloro-3-indolil- β -galactopiranosideo
<i>ycf</i>	genes putativos cloroplastidiais conservados (<i>hypothetical chloroplast frames</i>)

SEQUENCIAMENTO E ANÁLISE DO GENOMA CLOROPLASTIDIAL DE CANA-DE-AÇÚCAR (*Saccharum officinarum* L.)

Autor: TERCILIO CALSA JUNIOR

Orientador: Profa. Dra. HELAINE CARRER

RESUMO

Os cloroplastos são organelas exclusivas de plantas e algas fotossintetizantes, e desempenham funções essenciais na fisiologia e metabolismo vegetal, principalmente a fotossíntese e outras rotas biossintéticas. Além do núcleo e das mitocôndrias, os plastídios possuem seu próprio genoma (plastoma), envolvido na coordenação destes três distintos mas interdependentes sistemas genéticos das células vegetais. O DNA cloroplastidial existe como moléculas circulares fechadas de aproximadamente 150 kb (± 30), geralmente apresentando sequências repetidas invertidas, separando duas regiões de cópia única. Através da construção de bibliotecas *shot-gun* a partir do DNA cloroplastidial e do sequenciamento automático de 6266 clones, o plastoma de cana-de-açúcar (híbrido SP 80-3280) foi completamente sequenciado e as regiões codificadoras anotadas por homologia a outros sistemas vegetais. Por análises comparativas entre o plastoma de cana-de-açúcar e os de outras espécies, observou-se que todos os grupos funcionais de genes cloroplastidiais de milho foram localizados no plastoma de cana-de-açúcar, como também *ycf's* deste plastoma e de outras espécies, embora estas sequências não tenham função conhecida. Verificou-se,

ainda, maior identidade entre os plastomas de cana-de-açúcar e milho do que entre cana-de-açúcar e arroz com base na organização estrutural e na regulação da expressão gênica (edição de mRNA). Os resultados sugerem uma evolução comum entre os plastomas das gramíneas C_4 , gerando novos elementos para a pesquisa deste tipo de fotossíntese e metabolismo dos plastídeos, assim como para a tecnologia de transformação de cloroplastos de cana-de-açúcar.

SEQUENCING AND ANALYSIS OF SUGARCANE (*Saccharum officinarum* L.) CHLOROPLAST GENOME

Author: TERCILIO CALSA JUNIOR

Advisor: Prof. Dr. HELAINE CARRER

SUMMARY

Chloroplasts are exclusive plant and algae organelles, which perform essential roles in plant physiology and metabolism, mainly photosynthesis and other biosynthetic pathways. Beyond nuclei and mitochondria, plastids also have their own genome (plastome) involved in the coordination of these three distinct but interdependent genetic systems of plant cells. Chloroplast DNA exists as closed circular molecules of about 150 kb (± 30), usually presenting inverted repeat sequences separating two single copy regions. Through the construction of *shot-gun* libraries from chloroplast DNA and the automatic sequencing of 6266 clones, the sugarcane (hybrid SP 80-3280) plastome has been completely sequenced, and the coding regions annotated by homology to other plant systems. By comparative analyses between sugarcane plastome and those from other species, it was observed that all maize functional gene groups were identified in sugarcane plastome. Also *ycf's* found in other species plastome were identified, although their functions are still unknown. It was observed higher identity between sugarcane and maize plastomes than between sugarcane and rice ones, even based on structural organization and gene expression regulation (mRNA editing). The results suggest a common evolution among C₄ grasses, yielding new elements for

research on this type of plastid photosynthesis and metabolism, as well as for sugarcane chloroplast transformation technology.

1 INTRODUÇÃO

Os plastídios são organelas celulares presentes desde algas até angiospermas, que se multiplicam por divisão e são geralmente herdados maternalmente nas espécies cultivadas. Eles possuem o seu próprio cromossomo e também as enzimas necessárias para a replicação dos genes e síntese proteica. Durante a diferenciação celular, os proplastídios das células meristemáticas também se diferenciam em cloroplastos (folhas), cromoplastos (flores e frutos), leucoplastos (tecidos de reserva, para o armazenamento de amido ou lipídios) e outros tipos.

De acordo com a hipótese endossimbionte, os plastídios descendem de determinadas cianobactérias denominadas fotoclorófitas, as quais foram adquiridas via fagocitose para dentro de uma célula eucariótica primitiva. Durante a diferenciação dos cloroplastos, a membrana envoltória interna sofre invaginações para formar os tilacóides, os quais são subsequentemente fechados formando um sistema membranar amplo para o aparato fotossintético. Os cloroplastos realizam totalmente a fotossíntese, e também estão envolvidos em várias rotas metabólicas, como a assimilação de nitrogênio e enxôfre, e a síntese de aminoácidos, lipídios e pigmentos. A ampla vida vegetal na Terra depende da captação de energia luminosa desempenhada pelos cloroplastos, assim como a dos animais e de outros organismos heterotróficos.

As células vegetais exigem a coordenação entre três sistemas genéticos distintos mas interdependentes: DNA nuclear, DNA cloroplastidial e DNA mitocondrial. Os genomas organelares são similares tanto ao DNA procariótico quanto ao DNA nuclear eucariótico. Nem toda a informação necessária para garantir o crescimento e a replicação dos plastídios e mitocôndrias é codificada pelos genomas organelares. Genes

referentes aos componentes do aparato fotossintético e do sistema de tradução cloroplastidial foram identificados no genoma nuclear, indicando a ocorrência de transferência de genes para o núcleo e de rearranjos drásticos nos plastomas ao longo do processo evolutivo.

O DNA dos cloroplastos pode constituir uma parte substancial do DNA total em uma folha madura (até cerca de 30%). O DNA cloroplastidial (plastoma) existe como uma molécula circular fechada de aproximadamente 150 (± 30) Kb na maioria das plantas. Ele geralmente apresenta uma sequência repetida invertida, separando duas regiões de cópia única. Embora sejam encontradas diferenças de tamanho, o conteúdo e a organização gênica dos plastomas de espécies muito divergentes são notadamente similares. Entretanto, alguns genes podem ser rearranjados nas suas posições relativas no plastoma, de acordo com a espécie.

A comparação estrutural dos plastomas fornece não apenas informações sobre o processo de evolução dos cloroplastos, mas também elementos relacionados aos mecanismos responsáveis pela divergência entre genomas cloroplastidiais de várias espécies vegetais, e aos processos regulatórios da expressão gênica, destacando-se, no caso dos cloroplastos, os mecanismos de interação entre os três sistemas genéticos da célula vegetal e os fatores ambientais, e a ocorrência de edição de mRNA, um processo de modificação pós-transcricional já descrito nos cloroplastos das espécies testadas até o momento.

A manipulação genética do DNA cloroplastidial é considerada uma nova tecnologia alternativa que suplanta muitas desvantagens associadas à manipulação genética nuclear. A herança materna dos genes cloroplastidiais (e de eventuais transgenes) na maioria das espécies cultivadas torna-se desejável em situações onde existam riscos potenciais de escape dos genes exógenos para culturas ou plantas daninhas filogeneticamente próximas à espécie transgênica. Além disso, a possibilidade de expressão de vários transgenes em conjuntos policistronicos e os maiores níveis de expressão verificados para os transgenes inseridos no plastoma em relação àqueles integrados ao genoma nuclear são vantagens deste processo.

Um pressuposto básico para a transformação genética de cloroplastos é o conhecimento da sequência completa de nucleotídeos do plastoma a ser transformado, pois este processo ocorre através de recombinação homóloga, permitindo a inserção dirigida do transgene. Assim, pode-se escolher os sítios potenciais para a inserção de transgenes no plastoma. A identificação da sequência de nucleotídeos do genoma desta organela permite a localização dos genes e das regiões intergênicas. Estas informações sobre a localização dos genes e das regiões intergênicas têm auxiliado como ferramenta fundamental na manipulação genética do plastoma.

O sequenciamento do DNA cloroplastidial de cana-de-açúcar (*Saccharum officinarum* L.) envolve a análise da estrutura e organização deste plastoma, de modo a fornecer elementos para a análise de aspectos evolutivos e aplicações biotecnológicas (como a transformação de cloroplastos), as quais também podem ser úteis para estudos de regulação da expressão gênica neste sistema. Isto se torna particularmente interessante, desde que a cana-de-açúcar é a espécie vegetal que apresenta o maior número de ESTs (*expressed sequence tags*) já identificados (Projeto SUCEST/FAPESP).

Os objetivos deste trabalho foram:

- a) Construção de bibliotecas genômicas cloroplastidiais (plastômicas) de cana-de-açúcar (*Saccharum officinarum*).
- b) Sequenciamento das extremidades dos fragmentos clonados de DNA cloroplastidial.
- c) Montagem da sequência do plastoma de cana-de-açúcar.
- d) Anotação dos genes e *ycf*s (*hypothetical chloroplast frames*), identificando genes homólogos pela comparação com outros sistemas vegetais já conhecidos.

2 REVISÃO DE LITERATURA

2.1 CULTIVO E EVOLUÇÃO DA CANA-DE-AÇÚCAR (*SACCHARUM OFFICINARUM*)

A família *Poaceae* é uma das mais importantes famílias de espécies vegetais, ocupando aproximadamente 1/3 das áreas terrestres do planeta, e inclui as culturas de cereais de elevada importância na alimentação humana. A tribo *Andropogoneae* é uma das mais definidas taxonomicamente, dividida em três subtribos: *Saccharinae*/*Germainiinae*, *Sorghinae* e *Ischaeminae*/*Dimeriinae*/*Coicinae* (Clayton & Renvoize, 1986). A subtribo *Saccharinae* é tipicamente alta e higrófila, o que se considera ser a condição ecologicamente primitiva. Dentro desta subtribo, encontra-se o gênero *Saccharum*, o qual inclui os genótipos agronomicamente importantes de *Saccharum officinarum*, que têm sido cultivados durante milhares de anos devido ao seu alto teor de sacarose. A cultura da cana-de-açúcar e a agroindústria relacionada apresenta extrema importância para a economia do Estado de São Paulo e para o Brasil, através da produção de principalmente etanol e açúcar, e de outros produtos. De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE-<http://www1.ibge.gov.br/ibge/estatistica/indicadores/agropecuaria/lspa>), a área cultivada no Brasil em 2001 é de 4.928.736 ha, com uma safra esperada de 339.585.596 toneladas. No Estado de São Paulo, a produção dos 2.488.686 ha cultivados em 2001 está estimada

em mais de 189.540.537 toneladas (55,8% da produção nacional, em 50,5% da área nacional cultivada).

As análises taxonômicas e evolutivas de *Saccharum* são dificultadas devido à ampla ocorrência de poliploidia e da reprodução vegetativa opcional. Análises filogenéticas fundamentadas nos genomas nucleares e organelares (cloroplastos e mitocôndrias) têm auxiliado na elucidação dos eventos evolutivos para a formação dos genótipos atualmente cultivados com considerável importância (Sobral et.al., 1994). A taxonomia de *Saccharum* é ainda complexa (Al-Janabi et.al.,1994). Obviamente, a atividade humana nos últimos 10.000 anos desempenhou uma função crucial na dispersão, seleção e propagação vegetativa de determinados genótipos de cana-de-açúcar, incluindo híbridos interespecíficos e intergenéricos. Estudos utilizando ferramentas moleculares no complexo *Saccharum* demonstraram que os gêneros *Saccharum* e *Sorghum* possuem alta proximidade genética, apesar de o gênero *Sorghum* ter se dispersado da África, e o gênero *Saccharum* ter se estabelecido no sudeste asiático. Springer et.al. (1989) analisou a organização genômica do rDNA nuclear em *Sorghum* e em espécies próximas através de RFLP's (polimorfismo do comprimento de fragmentos de restrição), e verificou que sorgo, cana-de-açúcar e milho possuem tamanhos de monômeros de rDNA e mapas de restrição muito similares, e que sorgo e cana-de-açúcar estão mais proximamente relacionados um ao outro que com milho.

2.2 ORIGEM E CARACTERÍSTICAS DOS PLASTÍDIOS

Os plastídios compreendem um grupo de organelas específicas de células vegetais, que se multiplicam por divisão e apresentam herança materna na maioria das espécies cultivadas (Daniell et.al., 1998; Scott et.al.,1999). Durante as fases de diferenciação celular nos diversos tecidos vegetais, os proplastídios da célula ovo diferenciam-se nos vários tipos de plastídeos existentes, geralmente em concordância com as funções metabólicas e fisiológicas do órgão e tecido em que se localizam. Seguem, portanto, as rotas de diferenciação em cloroplastos (folhas), cromoplastos

(flores, frutos e órgãos pigmentados), etioplastos (tecidos estiolados), leucoplastos (tecidos de reserva, para o armazenamento de amido ou lipídios) e outros tipos (Pyke, 1999).

De acordo com a teoria endossimbionte, os plastídios descendem de determinadas cianobactérias fotossintetizantes, as quais foram eventualmente integradas ao interior de células eucarióticas primitivas, que provavelmente já haviam adquirido de maneira semelhante outras bactérias aeróbicas (ancestrais das mitocôndrias). O resultado deste processo forneceu vantagens adaptativas ao eucarioto primitivo fotossintetizante, iniciando o processo evolutivo no qual se desenvolveram interações entre os compartimentos intracelulares cada vez mais específicas, capazes de permitir o estabelecimento e a diversidade dos organismos vegetais. A ordem cronológica e o número de eventos endossimbiontes cientificamente propostos ainda são questionados (Palmer, 1992).

Dentre os plastídios, os cloroplastos são melhor caracterizados e estudados sob aspectos celulares, metabólicos e genéticos. O cloroplasto é uma organela celular vegetal que contém a maquinaria completa para a fotossíntese, além de desempenhar funções relacionadas ao metabolismo de nitrogênio, enxôfre e fósforo, e a vários mecanismos de interação da célula vegetal com fatores ambientais, especialmente a luz (Taiz & Zeiger, 1998). Desta forma, os cloroplastos atuam em vários mecanismos de regulação metabólica, fisiológica e genética como mediadores entre as variações da luminosidade ambiental e as respectivas respostas do organismo vegetal a elas submetido.

2.3 INTERAÇÕES INTRACELULARES

Vários componentes cloroplastidiais, principalmente proteínas, originam-se em outros compartimentos celulares, e são direcionados para as organelas através de mecanismos relativamente conhecidos. Até o momento, não foram detectados processos de exportação de proteínas pelos cloroplastos, entretanto, o complexo enzimático de

importação de proteínas responsável pela entrada de polipeptídeos na organela tem sido descrito em detalhes. De acordo com Chen & Schnell (1999), o modelo atual que descreve o processo de importação de proteínas para os cloroplastos foi estabelecido através de ensaios bioquímicos com cloroplastos isolados de ervilha (*Pisum sativum*).

As proteínas cloroplastidiais codificadas por genes nucleares são sintetizadas inicialmente no citoplasma e acrescidas de uma cauda oligopeptídica, denominada peptídeo de trânsito, que permite o seu direcionamento e adesão à superfície cloroplastidial. Em um processo estimulado por GTP, a proteína precursora associa-se com os componentes do complexo de importação de proteínas (*translocon*) localizados na membrana externa da organela. A hidrólise de ATP no citoplasma e/ou no espaço intermembranar do cloroplasto desencadeia a interação da proteína precursora também com os demais componentes do *translocon* localizados na membrana interna da organela. Considera-se que esta fase é auxiliada por chaperonas que existem no espaço intermembranar. A hidrólise de ATP do estroma resulta na translocação completa da proteína precursora para o interior do cloroplasto, onde o peptídeo de trânsito é removido da cadeia proteica, em parte pela ação de fatores do estroma Chen & Schnell (1999).

Com base nas sequências dos genes codificadores dos componentes do *translocon* cloroplastidial de importação de proteínas, determinadas em ervilha (*Pisum sativum*), Jackson-Constan & Keegstra (2001) identificaram os mesmos genes em *Arabidopsis thaliana*, cuja sequência genômica completa foi recentemente publicada (*Arabidopsis* Genome Initiative, 2000). Foram identificados homólogos múltiplos para mais da metade dos genes correspondentes aos componentes do complexo de importação de proteínas. Com exceção de um dos componentes, observou-se que mais de uma isoforma putativa para cada subunidade era expressa. Assim, estes autores sugerem que múltiplos tipos de complexos de importação proteica existam dentro da mesma célula, cada um com uma afinidade específica para diferentes precursores cloroplastidiais.

A célula vegetal depende da coordenação entre os três sistemas genéticos nela existentes: o genoma nuclear, o genoma cloroplastidial (plastoma), e o genoma mitocondrial, em interações específicas influenciadas por fatores internos e externos ao

longo do desenvolvimento celular. A própria biogênese do cloroplasto a partir dos proplastídios não-diferenciados é induzida pela luz (Taiz & Zeiger, 1998).

2.4 O DNA CLOROPLASTIDIAL

Os cloroplastos possuem seu próprio cromossomo, constituído de DNA circular associado a proteínas, em estruturas denominadas nucleóides, geralmente próximos à membrana interna da organela. Entretanto, nem toda a informação necessária à replicação e à totalidade das funções cloroplastidiais é codificada pelo plastoma, a exemplo do que ocorre com as outras organelas semi-autônomas, as mitocôndrias (Keegstra & Cline, 1999; Mackenzie & McIntosh, 1999). Os genomas organelares apresentam características comuns a DNAs procarióticos, além de outras comuns a DNAs eucarióticos. A organização nucleoproteica do DNA nuclear não é encontrada, mas vários genes organelares possuem introns (Palmer, 1992).

Embora os cloroplastos mantenham o plastoma, remanescente do genoma ancestral, a maioria dos genes codificadores de proteínas cloroplastidiais foram evolutivamente transferidos para o núcleo (Bauer et.al., 2001). Isto impõe à célula vegetal dois desafios principais durante a biogênese dos cloroplastos: primeiramente, os fatores luminosos e de desenvolvimento devem ser interpretados para expressar de maneira coordenada a informação genética contida em dois compartimentos celulares distintos; também, os cloroplastos em desenvolvimento devem ser eficientes na importação de proteínas codificadas pelo núcleo e sintetizadas no citossol. Outras proteínas, incluindo aquelas codificadas pelo plastoma, devem ser adequadamente distribuídas e estruturadas nos compartimentos dos tilacóides para a montagem dos sistemas fotossintéticos ativos.

Por outro lado, a coordenação entre o desenvolvimento dos cloroplastos e a expressão de genes nucleares relacionados a este processo envolve uma rota de transdução de sinais da organela para o núcleo. Mochizuki et.al. (2001), estudando mutantes de *Arabidopsis thaliana* para o gene *gus5* (*genomes uncoupled* - codificador da

subunidade H da enzima Mg-quelatase), obtiveram evidências genéticas sugerindo que os precursores da biossíntese de clorofila podem funcionar como moléculas sinalizadoras neste mecanismo. Estes autores verificaram que determinadas variações na rota biossintética dos tetrapirróis geram um sinal celular, a partir dos cloroplastos, capaz de causar a repressão transcricional de genes nucleares que codificam para componentes plastidiais.

O DNA cloroplastidial pode, dependendo da espécie, constituir até 30% do DNA total em uma folha madura. O DNA cloroplastidial apresenta-se tipicamente na forma circular fechada, constituído por 120-220 Kb na maioria das plantas, observando-se um conteúdo gênico extremamente conservado mesmo entre espécies filogeneticamente distantes (Palmer & Stein, 1986). De acordo com a espécie, pode ocorrer alterações estruturais no plastoma, envolvendo o reposicionamento relativo de alguns genes. O plastoma é considerado um sistema genético extremamente poliplóide, pois em uma única célula vegetal podem existir até 10000 cópias do DNA cloroplastidial (Palmer & Stein, 1986). Estruturalmente, o plastoma da maioria das espécies é constituído de um par de regiões repetidas invertidas (convencionalmente denominadas IR_a e IR_b), as quais são separadas por uma região de cópia única menor (SSC) e uma região de cópia única maior (LSC). De acordo com Kato et.al. (2000), o conteúdo AT% varia de aproximadamente 61,3% nas gramíneas até 62,9% nas dicotiledôneas, sendo que este valor é menor nas repetições invertidas (IR) do que nas regiões de cópia única (SSC e LSC).

A análise do DNA cloroplastidial através de microscopia eletrônica e de eletroforese em gel de campo-pulsado (PFGE) foi inicialmente utilizada para a investigação da sua conformação em tabaco (*Nicotiana tabacum*) e espinafre (*Spinacia oleraceae*) (Kolodner & Tewari, 1975; Deng et.al., 1989). A análise em microscopia eletrônica revelou a existência de apenas monômeros circulares e uma reduzida porcentagem de moléculas dímeras. Da mesma forma, a análise por PFGE demonstrou uma menor porcentagem de formas multiméricas do DNA cloroplastidial, nunca acima de tetrâmeros. Mais recentemente, Lilly et.al. (2001) mostraram uma maior plasticidade estrutural do genoma cloroplastidial de vegetais superiores [tabaco, ervilha (*Pisum*

sativum) e *Arabidopsis thaliana*] através da análise citogenômica por *fiber-FISH* (hibridização *in situ* por fluorescência da fibra de DNA), sugerindo fortemente uma considerável dinâmica *in vivo* entre as formas do DNA cloroplastidial. Do total de moléculas visualizadas, apenas 45% eram circulares e, destas, 60% eram monômeros, sendo observados dímeros e até hexâmeros. Moléculas circulares e lineares de até 1000 Kb foram detectadas, podendo ser resultantes de recombinação e/ou replicação de DNA. A funcionalidade destes multímeros ainda é desconhecida assim como se os mesmos são posteriormente clivados em subunidades menores. As moléculas multiméricas e atípicas podem ter origem em eventos de recombinação ou em clivagens aleatórias seguidas da fusão dos intermediários da replicação. A presença de regiões repetidas e invertidas na maioria dos genomas cloroplastidiais favorece a recombinação intramolecular (Palmer & Thompson, 1982). Através da abordagem citogenômica foram reveladas estruturas e níveis de plasticidade anteriormente ignorados em um genoma até então considerado extremamente estável. Esta informação pode ser útil na análise detalhada de plantas transgênicas que possuam o DNA cloroplastidial modificado (plantas transplastômicas).

As sequências completas de nucleotídeos já foram determinadas para o plastoma de algumas espécies (Quadro 1), como *Nicotiana tabacum* (Shinozaki et.al.,1986), *Oryza sativa* (Hiratsuka et.al., 1989), *Zea mays* (Maier et.al., 1995), *Triticum aestivum* (Ikeo & Ogihara, 2000), *Arabidopsis thaliana* (Sato et.al.,1999), *Spinacia oleraceae* (Schmitz-Linneweber et.al.,2001), *Lotus japonica* (Kato et.al.,2000), *Pinus thunbergii* (Wakasugi et.al.,1994), do musgo *Marchantia polymorpha* (Ohyama et.al.,1986), de algumas algas: *Chlorella vulgaris* (Wakasugi et.al.,1997), *Euglena gracilis* (Hallick et.al.,1993) e de outras espécies vegetais inferiores, assim como para um representante dos possíveis ancestrais dos plastídios, a cianobactéria *Synechocystis* sp. PCC6803 (Kaneko et.al.,1996).

Os registros fósseis indicam que as monocotiledôneas e dicotiledôneas divergiram há 100-140 milhões de anos, e se estima que as fanerógamas dividiram um ancestral comum com os vegetais inferiores há 350-400 milhões de anos. Entretanto, na análise comparativa das sequências de leitura aberta (ORFs-*open reading frames*) contidas nos plastomas publicados de tabaco e de *Marchantia polymorpha*, as mesmas

severamente truncadas, além de mudanças estruturais também evidentes: as regiões invertidas (IRs) são expandidas para a inclusão de outros genes, e eventos de recombinação intramolecular favoreceram a diferenciação (Hiratsuka et.al., 1989).

No genoma cloroplastidial de *Pinus thunbergii*, observou-se a redução das IRs e a ausência de genes presentes nas outras espécies analisadas (tabaco, *Marchantia* e arroz): 11 genes funcionais das subunidades da enzima NADH-desidrogenase (4 genes completamente ausentes e 7 pseudogenes) (Wakasugi et.al.,1994). No plastoma de *Lotus japonica* ocorre uma inversão de 51 kb localizada na LSC, comum nas leguminosas, em relação às demais dicotiledôneas (Kato et.al.,2000).

No genoma cloroplastidial de *Chlorella vulgaris*, foi observada a presença de genes homólogos aos genes bacterianos envolvidos na divisão celular, estruturados da mesma forma que em *Escherichia coli*, o que sugere a similaridade entre o mecanismo de divisão de cloroplastos e a divisão bacteriana (Wakasugi et.al.,1997).

A análise de vários genomas cloroplastidiais estabeleceu a existência de uma ordem consenso dos genes cloroplastidiais entre as plantas terrestres vasculares idêntica àquela encontrada em tabaco (Palmer & Stein, 1986). As informações disponíveis para as espécies trigo (*Triticum aestivum*) e milho (*Zea mays*) demonstram que os seus genomas cloroplastidiais divergiram da ordem consenso dos genes cloroplastidiais através de uma série de inversões adjacentes na região longa de cópia única (LSC).

Na análise do genoma cloroplastidial de espinafre (*Spinacia oleraceae*), focalizou-se a análise em ORFs específicos de alguns plastomas e, portanto, evolutivamente não-conservados (Schmitz-Linneweber et.al.,2001). Apesar de demonstrar uma estrutura gênica extremamente similar àquela consenso entre os plastomas de dicotiledôneas já sequenciados, foram detectadas diferenças principalmente em relação ao tamanho dos *intros* e dos respectivos transcritos em alguns genes e *ycf's*. Já dentro o complexo poliplóide *Saccharum* (Andropogoneae) observaram-se reduzidos níveis de variação presentes nos genomas organelares (cloroplastos e mitocôndrias), sugerindo que o mesmo é relativamente muito recente ou as taxas evolutivas entre o complexo *Saccharum* e outras gramíneas são diferentes (Al-

Janabi et.al.,1994). Nenhuma variação foi observada nas variedades comerciais de cana-de-açúcar analisadas (híbridos de *Saccharum*), sugerindo a existência de uma monocultura citoplasmática mundial neste caso.

Considerando-se o plastoma de milho (140387 bp), o conteúdo gênico e a posição relativa de um total de 104 genes (70 genes codificadores de polipeptídeos, 30 genes de tRNAs e 4 genes de rRNAs) são idênticos aos de arroz, espécie filogeneticamente próxima. As diferenças são comuns nas regiões intergênicas. Além disso, os dois plastomas de gramíneas apresentam regiões localizadas de divergência. Uma delas pode refletir um evento de transferência gênica para o núcleo, e a degradação do respectivo gene cloroplastidial resultando em fragmentos do mesmo; a outra divergência parece ocorrer devido à perda completa de um gene plastidial e à sua substituição funcional por um homólogo nuclear (Maier et.al., 1995). O conteúdo gênico do plastoma de milho é representativo dentro da ordem *Poales*, dado o caráter conservado em relação ao genoma plastidial de arroz, e compreende as categorias descritas no Quadro 2.

CATEGORIA	GENES
RNAs ribossômicos	<i>rrn16, rrn23, rrn4.5, rrn5</i>
TRNAs	<i>trn's (30 espécies)</i>
Subunidade ribossômica pequena	<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12*, rps14, rps15, rps16*, rps18, rps19</i>
Subunidade ribossômica grande	<i>rpl2*, rpl14, rpl16*, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
RNA-polimerase plastidial	<i>rpoA, rpoB, rpoC1, rpoC2</i>
NADH-desidrogenase	<i>ndhA*, ndhB*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
PS I	<i>psaA, psaB, psaC, psaI, psaJ</i>
PS II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT</i>
Rubisco	<i>RbcL</i>
Citocromo b ₆ f	<i>petA, petB, petD, petG</i>
ATP-sintase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>
Fator de iniciação tarducional	<i>InfA</i>
Maturase	<i>MatK</i>
Protease	<i>ClpP</i>
Proteína membrana envoltória	<i>CemA</i>
ycf's	<i>3, 4, 5, 6, 7, 9</i>
ORFs	<i>148, 75, 69, 137, 85, 23, 133, 49, 63</i>

Quadro 2 - Categorias dos genes plastidiais em milho (*Zea mays*) (Maier et.al., 1995).

2.5 REGULAÇÃO GÊNICA NOS CLOROPLASTOS

A maioria dos genes cloroplastidiais estão organizados em operons e, assim, produzem mRNAs policistrônicos por co-transcrição, contrastando com o padrão de expressão gênica no núcleo das células vegetais. De acordo com Hager & Bock (2000), a origem procariótica dos cloroplastos é também evidenciada pelos mecanismos moleculares de praticamente todas as fases da expressão gênica plastidial, inclusive na estrutura e regulação da RNA-polimerase codificada por genes plastidiais, e na coordenação da atividade de outra RNA-polimerase codificada por genes nucleares que atua em plastídios não-diferenciados.

Os processos de regulação da expressão gênica plastidial envolvem mecanismos que atuam em níveis transcricionais, traducionais e pós-traducionais, geralmente ligados a outros fatores intra e extra-plastídicos. Mesmo em nível plastômico, podem ocorrer alterações regulatórias da expressão de determinados genes, como, por exemplo, através da metilação de resíduos de citosina em sequências promotoras. Ngerprasisiri et.al. (1988) relataram a presença de metilação no DNA de amiloplastos de tomate (*Lycopersicon esculentum*), embora ainda não tenham sido encontrados sinais de metilação de DNA em cloroplastos (Fojtová et.al., 2001). Isto sugere a existência de níveis discretos de metilação dos resíduos de citosina no DNA plastidial em estágios específicos do desenvolvimento vegetal.

Um processo de regulação pós-transcricional consideravelmente caracterizado em cloroplastos é a edição de mRNA. Este mecanismo foi inicialmente verificado em plastídeos para o gene *rpl2* de milho (*Zea mays*), e subsequentemente confirmado em todas as espécies de vegetais superiores testadas até o momento (Bock et.al., 1997). A edição ocorre pela substituição de uma base nitrogenada (C) por outra (U) no transcrito primário, e geralmente envolve o primeiro ou segundo nucleotídeo de um códon, levando à incorporação de um aminoácido diferente daquele codificado pelo DNA. Em casos mais raros, a edição pode criar códons de iniciação e terminação da tradução de um gene, e também pode ocorrer de forma silenciosa (terceiro nucleotídeo do códon) ou ainda em regiões não-traduzidas dos transcritos (5'-UTR) (Corneille

et.al.,2000). A edição de mRNA em plastídios é extremamente sítio-específica, e dependente de elementos *trans* e *cis*, assim como da espécie vegetal, para ocorrer.

Não foram ainda detectadas sequências de nucleotídeos comuns flanqueando os sítios de edição (citosinas) que possam ser utilizadas para a identificação putativa dos mesmos. Apesar disso, alguns autores têm demonstrado que são necessários apenas 16 nucleotídeos-5' e 6 nucleotídeos-3' flanqueando a citosina alvo para ocorrer a edição *in vivo* (Chaudhuri et.al., 1995; Chaudhuri & Maliga, 1996; Bock et.al., 1996), e que possíveis elementos *trans* responsáveis pela edição de uma determinada citosina podem ser inativados pela alteração de um único nucleotídeo da sua sequência (Reed et.al.,2001).

Corneille et.al. (2000) realizaram um estudo específico da ocorrência de edição de mRNA em plastídios de gramíneas, e escolheram os plastomas de arroz (*Oryza sativa*) e milho (*Zea mays*) devido às facilidades existentes para a análise comparativa entre estes dois plastomas: a disponibilidade da sequência do genoma plastidial de arroz (Hiratsuka et.al.,1989) e as informações sobre os sítios de edição de mRNA nos plastídios de milho (Maier et.al.,1995; Bock et.al.,1997). Com a análise, verificou-se a presença de 21 eventos de edição nos transcritos plastidiais de arroz, em menor número do que em milho (26), e a existência de 8 diferenças entre as duas espécies: 2 sítios específicos para arroz e 6 para milho. Na comparação entre a edição de mRNA em arroz e tabaco (*Nicotiana tabacum*), foram encontrados 9 transcritos editados apenas em tabaco, 4 editados apenas em arroz e 7 transcritos editados em ambas as espécies. Entretanto, em arroz e milho os mesmos 11 transcritos são editados (*atpA*, *ndhA*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *rpl2*, *rps8*, *rps14*, *rpoB* e *ycf3*) e, mesmo assim, existem 8 diferenças. Destas, 7 ocorrem devido à presença de uma timina ao invés de uma citosina no DNA, o que elimina a necessidade da edição do mRNA.

Considerando-se a edição de mRNA como mecanismo que restaura os aminoácidos evolutivamente conservados nas sequências gênicas, assume-se que este processo é essencial. Isto se aplica ao sítio de edição do gene *psbF*, desde que a perda da edição deste transcrito torna plantas de tabaco fotossinteticamente inativas (Bock et.al.,1994). Contudo, outros experimentos sugerem que a edição de mRNA pode ser

importante para alguns, mas não todos os sítios existentes, mesmo em genes essenciais. Um exemplo é a edição parcial do gene *rpoB*, cujos sítios estão localizados na região correspondente a um domínio não-essencial da subunidade β da RNA-polimerase plastidial (Corneille et.al., 2000). A edição tem se apresentado mais frequente em transcritos de genes plastidiais não-essenciais, como, por exemplo, os genes que codificam para as subunidades do complexo NADH-desidrogenase plastidial, os quais já se mostraram dispensáveis através da substituição de sua atividade por homólogos nucleares (Burrows et.al., 1998).

De acordo com Corneille et.al. (2000), a determinação dos sítios de edição são importantes também para o desenvolvimento de vetores de transformação de cloroplastos. Estes vetores baseiam-se em sequências de DNA cloroplastidial flanqueando o(s) gene(s) exógeno(s), e estas sequências flanquadoras podem derivar do plastoma da espécie a ser transformada, ou de uma espécie filogeneticamente próxima. Desde que genes plastidiais do vetor de transformação podem substituir os genes plastidiais da planta a ser transformada, apenas serão obtidas plantas funcionais se os sítios essenciais de edição introduzidos durante a transformação forem reconhecidos e editados na planta transformada (Chaudhuri et.al., 1995). Assim, diferenças específicas para cada espécie em relação à edição de mRNA podem limitar a gama de táxons passíveis de transformação por um dado vetor.

2.6 TRANSFORMAÇÃO GENÉTICA DE CLOROPLASTOS

A manipulação genética dos cloroplastos tem sido considerada uma tecnologia alternativa que contorna muitas questões polêmicas referentes à transformação genética nuclear. Uma questão ambiental rotineira é o escape de genes exógenos através da dispersão de pólen ou sementes de cultivos transgênicos para espécies filogeneticamente próximas (criando plantas daninhas mais resistentes, no caso de transgenes para a resistência a herbicidas), ou para outras culturas, resultando em “poluição” gênica (Daniell, 1999). De acordo com este autor, a herança maternal

característica dos genes cloroplastidiais (inclusive dos transgenes) apresenta-se como interessante em situações onde exista o risco potencial de fluxo gênico via pólen.

A utilização de culturas transgênicas nucleares que expressam as toxinas do *Bacillus thuringiensis* (*Bt*) em níveis sub-ótimos podem resultar no desenvolvimento de pragas mais resistentes. Entretanto, um estudo demonstrou que determinados insetos, com resistência 40000 vezes maior a outras toxinas *Bt*, foram totalmente controlados quando submetidos à alimentação com plantas expressando milhares de cópias de um novo gene *Bt* inserido no DNA cloroplastidial (Kota et.al., 1999). Apesar de o pólen de algumas espécies possuir plastídios metabolicamente ativos, o DNA plastidial é degradado durante a maturação do pólen e portanto não é transmitido à geração seguinte. De Cosa et.al. (2001), em um experimento semelhante, demonstraram que embora 47% da proteína total solúvel de cloroplastos transformados correspondesse ao produto do transgene *BtCry2Aa2*, esta proteína inseticida era ausente no pólen transgênico, o que evita a toxicidade deste pólen para insetos não-prejudiciais.

Embora atualmente seja possível transformar geneticamente várias espécies vegetais, os efeitos de posição e silenciamento gênico continuam limitando a expressão de genes exógenos no genoma nuclear. A demanda por novas plantas transgênicas relacionadas ao estudo dos componentes de diversas rotas metabólicas aumenta proporcionalmente à disponibilização de informações sobre novos genes, transcriptomas, proteomas e fatores de regulação gênica. A possibilidade da expressão de cadeias polistrônicas no genoma cloroplastidial (tipicamente procariótico) apresenta-se como oportunidade significativa para a expressão de rotas metabólicas exógenas em um único evento de transformação, não sujeito ao silenciamento gênico ou aos efeitos de posição. Isto já foi testado e verificado por De Cosa et.al. (2001), através da elevada expressão do operon de três genes (*BtCry2Aa2*) fusionado ao gene marcador seletivo *aadA*, que confere resistência ao antibiótico espectinomicina.

A transformação genética de cloroplastos também tem limitações. As principais desvantagens atribuídas a esta tecnologia têm sido o uso de genes de resistência a antibióticos como marcadores seletivos (igualmente à transformação nuclear), e o confinamento da atividade dos transgenes inseridos dentro da organela,

devido ao fato de ainda não se ter confirmado a existência de mecanismos de exportação de proteínas pelos cloroplastos.

A seleção dos transformantes transplastômicos utilizou inicialmente um gene *16SrRNA* mutante, resistente à espectinomicina (Svab et.al., 1990); subsequentemente, o gene *aadA* passou a ser utilizado como marcador seletivo (Svab & Maliga, 1993). Este gene codifica para a aminoglicosídeo-3-adenililtransferase, que inativa o antibiótico pela transferência de um grupo adenil proveniente de ATP para a espectinomicina ou estreptomicina. Contudo, estes antibióticos são também usados no controle de infecções bacterianas em humanos e animais, e especula-se que a probabilidade de transferência gênica de plantas para bactérias de solo ou do trato gastrointestinal aumente através da compatibilidade entre cloroplastos e bactérias em relação à sua maquinaria de síntese proteica, assim como pela presença de milhares de cópias do gene de resistência a antibióticos por célula, os quais, em sua maioria, são originários de bactérias.

Fisher et.al. (1996) propuseram uma metodologia para a eliminação de genes marcadores seletivos existentes em genomas cloroplastidiais transformados. A mesma demonstrou que a ocorrência de recombinação homóloga entre duas sequências repetidas (*direct repeats*) flanqueando o marcador seletivo possibilitou a remoção deste, em condições de crescimento não-seletivo.

Outra abordagem mais recente (Daniell et.al., 2001) relatou a primeira manipulação genética de cloroplastos sem a utilização de marcadores seletivos para resistência a antibióticos. O gene da betaina-aldeído-desidrogenase (BADH) de espinafre (*Spinacia oleracea*) foi utilizado como marcador seletivo. Esta enzima ocorre nos cloroplastos de algumas espécies vegetais, adaptadas a ambientes secos e salinos (Rathinasabapathi et.al., 1994), e o processo de seleção dos transformantes envolve a conversão de betaina-aldeído (composto citotóxico) em glicinobetaina (osmoregulador celular) pela enzima BADH cloroplastidial. A eficiência de transformação de cloroplastos foi 25% maior na seleção com betaina-aldeído em relação àquela conduzida em presença de espectinomicina. Além disso, a utilização de genes marcadores que naturalmente ocorrem nos vegetais deve reduzir a rejeição pública aos transgênicos. Ao

contrário do que se havia estabelecido no início da tecnologia de transformação de cloroplastos (Svab et.al., 1990), a seleção não-letal dos transformantes não é necessária, pois o acúmulo de betaina-desidrogenase é tóxico e letal para a célula vegetal, e, mesmo assim, permitiu a obtenção de transformantes estáveis. A seleção em betaina-aldeído é potencialmente útil para várias culturas, com exceção para aquelas que naturalmente expressam elevados níveis de BADH endógena como, por exemplo, os membros das famílias *Chenopodiaceae* e *Poaceae*, que naturalmente acumulam glicinobetaina e apresentam maior tolerância à seca (Daniell et.al., 2001).

Whitney et.al. (2001) testaram a possibilidade de expressar a enzima rubisco (ribulose-1,5-bisfosfato carboxilase/oxigenase) de algas não-verdes em cloroplastos de tabaco. As formas da enzima rubisco existentes em determinadas espécies de algas não-verdes apresentam maior afinidade de substrato pelo CO₂ do que pelo O₂. Portanto, a inserção de genes codificadores para estas enzimas mais eficientes no plastoma de um vegetal superior poderia incrementar sua eficiência de fixação de CO₂ e da fotossíntese. Através da transformação de cloroplastos de tabaco via biolística, foram inseridos de modo estável os operons *rbcL/rbcS* da algas *Galdieria sulphuraria* (rodófito) e *Phaeodactylum tricornutum* (diatomácea). As folhas das plantas transformadas continham aproximadamente a mesma quantidade de rubisco exógena e da forma endógena (mais de 30% das proteínas solúveis totais), entretanto, as subunidades da rubisco exógena (de algas) não eram submetidas à estruturação tridimensional e à montagem do complexo enzimático ativo. Posteriormente, verificou-se que as subunidades exógenas grandes (codificadas pelo gene *rbcL* exógeno) apresentavam a região N-terminal bloqueada, inapta para a oligomerização. Além disso, o acúmulo das subunidades inativas da rubisco exógena reduziu a síntese dos complexos endógenos ativos e, conseqüentemente, diminuiu a assimilação foliar de CO₂ e o crescimento das plantas transformadas. Estes resultados sugerem a existência de novos mecanismos de regulação pós-traducional em plastídeos, mais especificamente na região N-terminal das proteínas sintetizadas.

Outra abordagem tem sido a transformação de cloroplastos visando-se a inserção de genes de interesse agrônômico, principalmente de resistência a herbicidas.

Lutz et.al. (2001) obtiveram plantas transplastômicas de tabaco resistentes ao herbicida fosfinotricina (PPT) através da inserção do gene *bar*, cujo produto é a enzima fosfinotricina-acetiltransferase (PAT), inativadora do herbicida. Observou-se que a alteração dos códons do gene *bar* original para os padrões comuns de códons em transcritos cloroplastidiais (*codon usage*) permitiram o aumento da expressão do transgene nas organelas, ou seja, apenas a otimização da sequência do transgene para o padrão cloroplastidial de *codon usage* proporcionou maior acúmulo do produto do transgene. Detectou-se, também, que o gene contendo o *codon usage* alterado era pouco expresso em *Escherichia coli*. Além disso, os autores também sugerem a alteração do padrão de *codon usage* dos transgenes bacterianos de resistência a herbicidas ou antibióticos como estratégia para evitar a expressão destes transgenes após possíveis, embora raros, eventos de transferência horizontal de genes.

De acordo com Staub et.al. (2000), as plantas transplastômicas têm se tornado atrativas para a produção de proteínas de interesse farmacológico devido ao risco reduzido de contaminação com vírus de mamíferos, à ausência de disseminação do transgene via pólen em várias espécies, e à capacidade de produção em larga escala e com menor custo relativo. Estes autores obtiveram plantas de tabaco transplastômicas expressando a somatotropina humana, em forma solúvel e biologicamente ativa, constituindo mais de 7% do total de proteínas solúveis foliares (300 vezes a expressão obtida para o mesmo gene em plantas transgênicas nucleares).

A manipulação genética de cloroplastos tem sido dificultada também pela reduzida expressão de genes plastidiais em tecidos não-fotossintetizantes tais como nos frutos, tubérculos, caules e outros órgãos de armazenamento. Contudo, Ruf et.al. (2001) obtiveram recentemente plantas transplastômicas férteis de tomate (*Lycopersicon esculentum* Mill), e verificaram que a expressão do transgene (*aadA*) nos cromoplastos dos frutos era aproximadamente 50% daquela observada nos cloroplastos. De acordo com as elevadas taxas de acúmulo de proteína exógena em cloroplastos transgênicos (até mais de 40% das proteínas solúveis totais), os autores indicam este sistema como estratégia para a produção eficiente de vacinas, fármacos e anticorpos em frutos de tomate.

Ainda, a transformação de cloroplastos tem se mostrado útil para a investigação da função de genes putativos através da análise genética reversa. Devido à inserção dirigida característica, é possível inserir um gene marcador seletivo na região codificadora de uma determinada ORF cuja função é desconhecida, impedindo a expressão deste e permitindo a regeneração de mutantes deficientes para este gene putativo. Por exemplo, as ORFs *ycf1* e *ycf2* não apresentam genes homólogos em bactérias, indicando a sua origem a partir de eucariotos primitivos ou de transferência gênica lateral (Martin et.al.,1998).

Drescher et.al. (2000) verificaram que o nocaute dos genes putativos cloroplastidiais *ycf1* e *ycf2* não levava à formação de linhagens transplastômicas homoplásmicas, e mesmo após vários ciclos de regeneração sob pressão seletiva era mantida uma proporção entre as cópias do plastoma transformado e do plastoma selvagem. Com a eliminação da pressão de seleção para o transgene, as cópias do plastoma selvagem prevaleceram nas replicações subsequentes, indicando que os produtos destes ORFs são essenciais para a sobrevivência da célula vegetal.

Segundo Baena-González et.al. (2001), os mutantes obtidos para o gene putativo *ycf9* não apresentaram fenótipo diferenciado sob condições normais de crescimento, mas demonstraram taxas de crescimento severamente reduzidas em irradiâncias sub-ótimas, quando comparados aos não-mutantes. A inativação do gene *ycf9* resultou na diminuição da quantidade do complexo NAD(P)H-desidrogenase, e levou à aceleração do fluxo de elétrons para o fotossistema I (PSI), sem alteração da capacidade máxima de transferência de elétrons do fotossistema II (PSII). Apesar da menor quantidade de complexos NAD(P)H-desidrogenase nos tilacóides dos mutantes, não foram observadas diferenças na evolução de O₂ no PSII, mesmo em baixa intensidade luminosa. Os mutantes apresentaram menor intensidade no transporte de elétrons independente do PSII (transporte cíclico de elétrons dependente da ferredoxina e a redução da plastoquinona mediada pela NAD(P)H-desidrogenase). Assim, verificou-se que a inativação do *ycf9* gera anormalidades ligadas ao particionamento de elétrons entre os componentes da cadeia fotossintética transportadora de elétrons.

A cana-de-açúcar (*Saccharum officinarum* L.), além da significativa importância sócio-econômica para o sistema agroindustrial do Estado de São Paulo e do Brasil, tem adquirido considerável destaque no âmbito científico e tecnológico, com os resultados do sequenciamento de milhares de ESTs (*expressed sequence tags*) e da categorização de aproximadamente 40000 genes (Projeto SUCEST/FAPESP), cujas funções já estão em estudos visando-se a aplicação estratégica destas informações na obtenção de variedades e híbridos com novas características agronômicas. Além disso, viabilizou-se a utilização desta espécie como modelo para a investigação de diversos aspectos gênicos, bioquímicos e fisiológicos de plantas superiores, mais especificamente gramíneas de fotossíntese C₄. Neste aspecto, a determinação da sequência de nucleotídeos do plastoma de cana-de-açúcar favorece significativamente eventuais estudos envolvendo genes nucleares cujos produtos têm atividade nos plastídios (participando da estrutura e de diversas funções destas organelas), permitindo o melhor entendimento de fatores moduladoras da expressão gênica nuclear e cloroplastidial em termos de coordenação intracelular.

O sequenciamento do DNA cloroplastidial de cana-de-açúcar envolve a análise da estrutura e organização deste plastoma, de modo a fornecer ferramentas para aspectos evolutivos e aplicações biotecnológicas (como a transformação do DNA cloroplastidial), que podem também ser úteis para futuros estudos de regulação da expressão gênica.

3 MATERIAL E MÉTODOS

As atividades referentes à presente Dissertação foram desenvolvidas no Laboratório de Biotecnologia Agrícola, do Departamento de Ciências Biológicas da Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.

3.1 MATERIAL VEGETAL

A variedade comercial SP 80-3280 (gentilmente cedida pelo Centro de Tecnologia (CTC) - Copersucar, Piracicaba-SP) foi utilizada como material vegetal, e cultivada em casa-de-vegetação. Foram usados segmentos nodais de plantas adultas cultivadas em campo como material para a propagação vegetativa, através da imersão destes segmentos em água destilada a 50°C durante 30 minutos, e subsequente plantio em substrato comercial *Plantimax* (vasos contendo ~ 1000 cm³ / segmento nodal). Para o isolamento de cloroplastos, as plantas com idade entre 7 e 20 dias foram mantidas no escuro durante 2 dias anteriormente à coleta de tecido foliar, para reduzir a quantidade de amido armazenada em grânulos nos cloroplastos, que pode dificultar o processo de isolamento do DNA cloroplastidial. Também foram utilizadas folhas provenientes de plântulas de cana-de-açúcar (SP 80-3280) cultivadas *in vitro* [meio de cultura MS

(Murashige & Skoog, 1962) adicionado de 1,0 mg/L benzilaminopurina] em sistema de micropropagação, com 40 dias e mantidas em escuro por 2 dias.

Após a coleta das folhas, estas foram imediatamente mantidas em gelo ou congeladas em nitrogênio líquido, e subsequentemente utilizadas para o isolamento das organelas de acordo com os métodos testados.

3.2 ISOLAMENTO DE CLOROPLASTOS

Para a obtenção do DNA cloroplastidial, geralmente se procede ao isolamento de cloroplastos intactos através de centrifugação diferencial ou gradientes descontínuos, e posteriormente à extração do ptDNA. Estes métodos baseiam-se nas diferenças entre as densidades das organelas celulares vegetais. Poucos protocolos para essa finalidade utilizam-se de diferenças químicas entre as organelas, como, por exemplo, as composições das membranas nuclear e organelares.

Para o isolamento de cloroplastos de cana-de-açúcar, foram testados diferentes métodos existentes (tanto para o isolamento de cloroplastos quanto para o isolamento do ptDNA), os quais são apresentados a seguir. As soluções utilizadas em cada método testado são descritas no Quadro 3.

Métodos (referências)	Soluções	Composição
Centrifugação diferencial (Weiqi. et.al., 1996)	Tampão de homogenização (TH)	50 mM Tris pH 8,0; 25 mM EDTA; 10 mM 2-mercaptoetanol; 0,1 % albumina (p/v); 1,25 M NaCl/
	Tampão de Ressuspensão (TR)	20 mM Tris pH 8,0; 10 mM EDTA/
Gradiente de Percoll (Bock, 1998)	Tampão de Isolamento (TI) pH 6,8	1,65 M sorbitol; 5 mM Na ₂ PO ₄ ; 250 mM Hepes; 10 mM EDTA;
	Tampão de Isoascorbato (TS) pH 7,0	5 mM MgCl ₂ ; 5 mM 2-mercaptoetanol/
	Tampão de Trituração (TT) Percoll 40%	500 mM isoascorbato; 50 mM Hepes/
	Percoll 80%	100 ml TI; 5 ml TS; água destilada até 500 ml/ 8 ml Percoll; 4 ml TI; 2 g PEG 6000; 0,2 ml TS; água destilada até 20 ml/ 8 ml Percoll; 2 ml TI; 1 g PEG 6000; 0,1 ml TIS/
Gradiente de sacarose (Taylor et.al., 1993)	Tampão de Extração (TE)	0,35 M sorbitol; 50 mM Tris pH 8,0; 25 mM EDTA; 0,1% BSA (p/v); 0,1% 2-mercaptoetanol (v/v)/
	Tampão de lavagem (TL)	0,35 M sorbitol; 50 mM Tris pH 8,0; 25 mM EDTA/
DNase I (Triboush et.al., 1998)	Sacarose 52%	26 g sacarose; até 50 ml TE/
	Sacarose 30%	15 g sacarose; até 50 ml TE/
	STE pH 7,8	0,4 M sacarose; 50 mM Tris; 20 mM EDTA; 0,2 % BSA (p/v); 0,2 % 2-mercaptoetanol (v/v)/
	ST pH 7,8	0,4 mM sacarose; 50 mM Tris; 0,1 % BSA (p/v)/
CsCl-bisbenzimida (Taylor et.al., 1993)	TEN pH 7,2	100 mM NaCl; 100 mM Tris; 50 mM EDTA; 0,2 % 2-mercaptoetanol (v/v)/
	NETF pH 8,0	1,25 M NaCl; 50 mM Tris; 50 mM EDTA; 50 mM NaF/
	CTAB	2% CTAB (p/v); 100 mM Tris pH 8,0; 20 mM EDTA; 1,4 M NaCl; 2% 2-mercaptoetanol (v/v)/
	CsCl	1M CsCl/
Triton X-100 (Mariac et.al., 2000)	EtBr	0,2 mg/ml brometo de etídio/
	Bz	10 mg/ml bisbenzimida/
	Isopropanol saturado	60 ml Isopropanol; 40 ml água destilada/
	STEB pH 7,8	0,35 M sorbitol; 100 mM Tris; 5 mM EDTA; 0,5 % Na-bissulfito (p/v)/
Triton X-100 (Mariac et.al., 2000)	STEL pH 7,8	0,35 sorbitol; 100 mM Tris; 5 mM EDTA/
	NET pH 8,0	100 mM NaCl; 80 mM Tris; 30 mM EDTA/
	Triton X-100	20% Triton X-100 (v/v)/

Quadro 3 - Soluções utilizadas nos métodos de isolamento de cloroplastos/DNA de cloroplastos.

a) Centrifugação diferencial: este método apenas fundamenta-se nas diferenças de densidade existente entre as organelas vegetais (Weiqi et.al., 1996; Milligan, 1989; Taylor et.al.,1993; Bookjans et.al., 1984). Após a trituração de 50 g de folhas em 200 ml de TH a 4°C (ver composição das soluções no Quadro 3), em homogenizador (4 pulsos de 5 segundos cada), o homogenizado foi filtrado em 3 camadas de gaze e 1 camada de *Miracloth* (Calbiochem), e centrifugado a 200×g por 5 minutos a 4°C para a sedimentação dos núcleos. O sobrenadante foi coletado e centrifugado a 1700×g (15 minutos, 4°C) para a sedimentação dos cloroplastos, os quais foram ressuspensos em 10 ml de TR (4°C), novamente centrifugados a 1700×g (15 minutos, 4°C) e submetidos à extração do seu DNA (item 3.3).

b) Gradientes de *Percoll* ou sacarose: a fase inicial deste tipo de método (Taylor et.al.,1993; Bock, 1998; Phillips, 1985; Sakai et.al.,1998; Palmer, 1986; Schuler & Zielinski,1989) assemelha-se à centrifugação diferencial. Aproximadamente 50 g de folhas foram trituradas em 200 ml de TT ou TE (4°C), em homogenizador (4 pulsos de 5 segundos cada). Os homogenizados foram filtrados em 3 camadas de gaze e 1 camada de *Miracloth* (Calbiochem), e centrifugados a 50×g por 2 minutos a 4°C. Os respectivos sobrenadantes foram transferidos para tubos novos e centrifugados até 1700×g a 4°C. Os sedimentos contendo cloroplastos foram ressuspensos em 4 ml de TT ou TL (4°C), com o auxílio de um pincel macio de pintura (espessura de 0,5 cm). Após a ressuspensão, os cloroplastos foram aplicados sobre gradientes descontínuos (*step-gradient*) com soluções de substâncias que constituem matrizes fluidas [*Percoll* (Pharmacia), sacarose]. Foram utilizados gradientes de *Percoll* (40%, 80%) e de sacarose (30%, 52%), subsequentemente centrifugados a 2500×g por 10 minutos a 4°C. Após a centrifugação, observou-se a formação de sedimento no fundo dos tubos (núcleos e outros constituintes celulares), de material residual que não entrou nos gradientes (cloroplastos lisados) e das bandas correspondentes aos cloroplastos intactos (interfase), as quais foram coletadas com ponteiros de pipetadores cortadas na base (diâmetro de 3 mm). Esta banda foi transferida para tubos novos contendo 4 ml de TT ou TL (4°C), os quais foram invertidos lentamente 5 vezes. Em seguida, foram centrifugados a 1700×g

por 10 minutos a 4°C, e os sobrenadantes descartados. Os sedimentos foram então submetidos à extração do ptDNA.

c) DNase I: (Mourad & Polacco, 1989; Triboush et.al., 1998; Ma et.al., 1984). Aproximadamente 15 g de folhas, cortadas em pedaços de 1 cm² de área, foram maceradas manualmente em 10 ml de STE a 4°C, e o extrato resultante foi filtrado em 1 camada de *Miracloth* (Calbiochem). O filtrado foi diluído com 1 volume de ST e centrifugado a 1700xg por 10 minutos a 4°C. Após o descarte do sobrenadante, o sedimento foi ressuspensionado em 4 ml de TEN (4°C), e adicionou-se DNase I (a 10 µg/ml TEN), para incubação a 4°C por 1 hora. Nestas condições, a enzima DNase I (livre de RNase – Gibco) foi utilizada para degradar o DNA nuclear contaminante e, devido ao controle térmico para a baixa atividade enzimática (4°C), o DNA cloroplastidial não seria degradado por estar contido nas organelas intactas. A DNase I foi inativada através de 2 lavagens dos cloroplastos, constituídas da adição de 8 ml de NETF (NaF é inibidor de DNase I), centrifugação a 1700xg por 10 minutos a 4°C e descarte do sobrenadante. O sedimento final obtido foi utilizado para a extração do ptDNA.

d) CsCl-bisbenzimida : este protocolo independe do isolamento de organelas. O DNA total foi extraído de 5g de tecido foliar (ver item 3.3) e submetido a uma ultracentrifugação (108000xg por 22 horas a 20°C) em gradiente contínuo de CsCl. No preparo deste gradiente, adicionou-se bisbenzimida (Hoechst 33258, a 10 µg/ml Tris pH 8,0), a qual se liga às bases A e T do DNA, e fluoresce sob luz ultravioleta (Taylor et.al., 1993; Palmer, 1986). O DNA cloroplastidial apresenta maior quantidade de bases A e T do que o DNA nuclear e, quando ligado à bisbenzimida, esta origina uma diferença de densidade entre os dois tipos de DNA suficiente para a separação em gradientes de CsCl. A banda correspondente ao ptDNA foi coletada com agulha e seringa e purificada de acordo com Taylor et.al. (1993).

e) Triton X-100 : este protocolo (Mariac et.al., 2000) também gera inicialmente uma suspensão de organelas celulares em um tampão osmótico apropriado. O material foliar coletado (10 g) foi imediatamente congelado em nitrogênio líquido, triturado em almofariz até partículas menores que 1,0 mm, aliquoteado em tubos de microcentrífuga (aproximadamente 0,3 ml de material triturado por tubo de 1,5 ml) e

armazenado a -70°C por até 30 dias. Para cada isolamento, foram utilizadas 10 alíquotas. Cada uma foi ressuspensa em aproximadamente 1,2 ml de STEB (em gelo), e filtrada através de 1 camada de *Miracloth* (Calbiochem). O filtrado foi centrifugado a $4000\times g$ (10 minutos, 4°C) para a sedimentação dos núcleos e cloroplastos, e eliminação de mitocôndrias (descartadas no sobrenadante). O sedimento obtido (núcleos + cloroplastos) em cada tubo foi ressuspensa em 1 ml de STEL (em gelo) e centrifugado a $1000\times g$ (5 minutos, 4°C). Os sedimentos resultantes foram ressuspensados em 100 μl de STEL (em gelo) e reunidos em um único tubo, o qual foi novamente centrifugado a $1000\times g$ (5 minutos, 4°C). O sedimento foi ressuspensa em 200 μl de NET (em gelo), e foram adicionados 10 μl de 2-mercaptoetanol e 50 μl de Triton X-100 20% (v/v). A suspensão de cloroplastos e núcleos foi submetida à lise com o detergente não-iônico Triton X-100 sob condições [Triton X-100 4% (v/v), 0°C , 1-1,5 horas] que favorecem a lise específica das membranas dos cloroplastos, devido à maior afinidade de reação deste detergente por membranas de composição tipicamente procariótica. Após a lise, os núcleos e os cloroplastos não-lisados foram sedimentados a $1000\times g$ e o sobrenadante (contendo o ptDNA liberado na lise específica) foi utilizado para a extração do ptDNA.

3.3 ISOLAMENTO DO DNA DOS CLOROPLASTOS

A fração correspondente aos cloroplastos, obtida nos métodos descritos no item 3.2 (exceto o método com bisbenzimidazida), foi submetida ao isolamento do ptDNA. Para isto, utilizou-se o método aplicado à extração de DNA total de tecidos foliares descrito por Doyle & Doyle (1990), apenas com algumas modificações. Os sedimentos de cloroplastos (centrifugação diferencial, gradientes, DNase I) e o sobrenadante pós-lise do método com Triton X-100 foram ajustados para o volume de 500 μl com o tampão de extração de DNA [2 % CTAB (p/v); 100 mM Tris pH 8,0; 10 mM EDTA pH 8,0; 1,4 M NaCl; 1% 2-mercaptoetanol (v/v)]. Para a lise das organelas,

estas foram adicionadas de 500µl do tampão de extração e incubadas a 55°C durante 10 minutos, e subsequentemente incubadas em gelo por 1 minuto. Posteriormente, procedeu-se a duas extrações com clorofórmio:álcool isoamílico (24:1 v/v), e a fase aquosa final foi adicionada de 0,1 volumes de acetato de sódio 3 M (pH 5,2) e 1 volume de 2-propanol para a precipitação dos ácidos nucleicos (-70°C durante 1,0 hora). Após centrifugação a 12000xg (4°C, 30 minutos), o sedimento foi lavado em 0,5 ml de etanol 70%, seco à temperatura ambiente durante 20 minutos e ressuspendido em 20 µl de 10mM Tris pH 8,0. Alguns autores recomendam a purificação do DNA obtido em gradientes de CsCl (Aragão & Rech, 1998), mas esta não foi necessária.

3.4 BIBLIOTECAS DE FRAGMENTOS DE PTDNA

Os ptDNAs isolados foram digeridos com a endonuclease *EcoRI*, de acordo com o fabricante (*Gibco*), e o padrão de restrição foi analisado por eletroforese em gel de agarose 1,0% (TAE 1×, 10 V/cm), para a determinação da qualidade e pureza do DNA cloroplastidial, principalmente em relação à degradação e à contaminação com DNA nuclear. O padrão de restrição de uma preparação de DNA cloroplastidial pura é caracterizado por bandas nítidas isoladas, e a presença de contaminação com DNA nuclear é facilmente visualizada através de *smears*. Como controle, algumas bandas visualizadas foram coletadas, clonadas em vetor plasmidial [pGEM 3z/f (*Promega*) clivado com a endonuclease *EcoRI*] e tiveram suas extremidades sequenciadas.

O ptDNA classificado como de melhor qualidade (método com Triton X-100 + extração em CTAB) foi submetido à fragmentação mecânica por sonicação (em gelo, por 2 fases de 20 segundos a 75% / modelo CPX-600, *Cole Parmer Instr.*). Os fragmentos foram submetidos ao reparo das extremidades (*blunting* com T4-polimerase + *Klenow* + T4-fosfonucleotídeo-quinase), e fracionados por tamanho em gel de agarose 1,0% (LMP, TAE 1×, 10 V/cm). Os fragmentos de tamanho entre 1-2 Kb foram excisados do gel. Os mesmos foram purificados pela fusão da agarose a 65°C e adição de

0,1 volumes de 4M NaCl, seguida de duas extrações com 1 volume de fenol (equilibrado com Tris pH 8,0) e uma extração com 2 volumes de éter (saturado com ddH₂O); os fragmentos foram precipitados com 0,1 volumes de acetato de sódio 3 M (pH 5,2) e 2 volumes de etanol absoluto (-70°C, 30 minutos), e sedimentados a 14000xg (4°C, 20 minutos). O sedimento obtido foi lavado com 0,5 ml de 70% etanol (v/v), seco à temperatura ambiente por 20 minutos e ressuspensionado em 12 µl de ddH₂O. Posteriormente, os fragmentos foram ligados ao plasmídeo pUC18 desfosforilado, no sítio *Sma*I (*Sureclone kit*, *Pharmacia*), de acordo com o fabricante da enzima T4-DNA-Ligase (*Gibco*).

O produto da reação de ligação foi usado na transformação de células competentes de *E. coli* DH5- α preparadas em RbCl de acordo com a metodologia descrita por Doyle (1996), através de choque térmico, e o *screening* dos clones foi conduzido pela seleção em ampicilina (100 mg/L) e X-Gal/IPTG (operon *lacZ* do plasmídeo), durante incubação a 37°C por 16 horas.

As colônias brancas de *E. coli* obtidas contendo insertos nos plasmídeos (de acordo com a seleção em X-Gal/IPTG através do operon *lacZ* do plasmídeo) foram analisadas quanto ao tamanho dos insertos obtidos através de PCR de colônia, no qual pequenas porções das colônias brancas são individualmente coletadas através de *picking* manual com palitos de madeira novos e autoclavados, em câmara de fluxo laminar vertical esterilizada, e diretamente usadas como molde para a amplificação, estimando-se seu volume em 0,5 µl. A reação de PCR convencional (*polymerase chain reaction*) com os primers M13/pUC *Reverse* (5'-AGCGGATAACAATTCACACAGG-3') e *Forward* (5'-CCCAGTCACGACGTTGTAAAACG-3') (M13/pUC - *Gibco*) foi realizada nas seguintes condições: 94°C por 3 minutos; 30 ciclos de desnaturação a 94°C por 45 segundos, anelamento a 50°C por 30 segundos e extensão a 72°C por 1,5 minutos; extensão final a 72°C por 5 minutos; finalização a 4°C. Os produtos de PCR foram analisados por eletroforese em gel de agarose 1,0% (TAE 1 $\times$, 10 V/cm), para a estimativa amostrada do tamanho médio de insertos em cada biblioteca.

Os clones foram isolados através de *picking* manual aleatório das colônias brancas com palitos de madeira novos e autoclavados, em câmara de fluxo laminar

vertical esterilizada, e individualizados em microplacas de cultivo (96 *slots*, ou poços) contendo 150 µl de meio de cultura *CircleGrow* (Bio 101, Inc.; Cat. N. 3000-142) com glicerol a 8 % (v/v) e ampicilina (100 mg/L). Em seguida, as microplacas já inoculadas com os clones individualizados foram envoltas em filme plástico de PVC transparente e mantidas a 37°C por 16 horas, sendo posteriormente armazenadas a -75°C.

3.5 SEQUENCIAMENTO E ANÁLISE DOS CLONES

3.5.1 ISOLAMENTO DO DNA PLASMIDIAL DOS CLONES

Para o isolamento do DNA plasmidial dos clones, procedeu-se à minipreparação de plasmídeos pelo método de lise alcalina, em uso no Projeto SUCEST/FAPESP. Após o descongelamento dos clones mantidos a -70°C em microplacas (temperatura ambiente por 30 minutos), os mesmos foram inoculados em poços de microplacas de crescimento contendo 1 ml de *CircleGrow* seletivo (100 mg ampicilina / L), com o auxílio de replicador para 96 poços, respeitando-se as coordenadas de posição nas microplacas. Estas foram fechadas com adesivos plásticos transparentes, perfurados com agulha (diâmetro de 0,5 mm) para melhor oxigenação e consequente crescimento dos clones.

As microplacas de crescimento inoculadas foram mantidas a 37°C por 22 horas, sob agitação de 320 r.p.m.. Em seguida, as mesmas foram centrifugadas a 2250xg durante 4 minutos (20°C) para a sedimentação das células, e o sobrenadante (meio de cultura) foi descartado por inversão em recipiente contendo solução comercial de hipoclorito de sódio a 30%. Os procedimentos a seguir referem-se a cada poço das microplacas. Todas as centrifugações e agitações em vortex foram realizadas com as microplacas fechadas por adesivo plástico transparente.

As microplacas foram mantidas na posição invertida sobre papel absorvente por 5 minutos. Adicionou-se 240 µl de solução autoclavada GET [20 % glicose (p/v); 0,5 M EDTA pH 8,0; 1 M Tris pH 8,0] a 4°C, seguindo-se a agitação em vortex na rotação máxima (2 minutos). As microplacas foram novamente centrifugadas a 2250xg durante 4 minutos (20°C), e o sobrenadante descartado como citado acima.

Foram adicionados 80 µl da solução GET, seguida de agitação em vortex na rotação máxima (2 minutos). De acordo com as coordenadas de posição nas microplacas, foram transferidos 80 µl para a respectiva coordenada em microplacas de fundo “U” (96 poços, capacidade de 250 µl/poço), já contendo 2,0 µl de RNase A (10 mg/ml). Adicionou-se 80 µl da solução NS [4M NaOH; 10% SDS (p/v)], e procedeu-se à agitação em vortex por 30 segundos a 70% da rotação máxima. As microplacas foram incubadas à temperatura ambiente por 5 minutos e posteriormente centrifugadas até 150xg (20°C). Foram então adicionados 80 µl da solução autoclavada KOAc 3M [294,4 g/L acetato de K (3.H₂O); 15 ml/L ácido acético glacial], seguida de agitação em vortex por 30 segundos a 70% da rotação máxima, para a precipitação de proteínas e do DNA cromossômico. As microplacas foram novamente incubadas à temperatura ambiente por 5 minutos e posteriormente centrifugadas até 150xg (20°C).

Os adesivos foram removidos e as microplacas incubadas inicialmente a 90°C por 20 minutos, e depois no gelo durante 10 minutos. As mesmas foram centrifugadas a 2250xg por 4 minutos (20°C), e o sobrenadante (aproximadamente 230µl) foi coletado e transferido para a respectiva coordenada em microplacas-filtro específicas para adsorção de proteínas e eluição de ácidos nucleicos (MAGV N22, Millipore), montadas sobre microplacas de fundo “V”, alinhando-se os poços de mesma coordenada. Em seguida, estes conjuntos (filtro sobre fundo “V”) foram centrifugados a 2250xg por 4 minutos (20°C).

As placas-filtro foram removidas e descartadas, e foram adicionados ao filtrado 110 µl de 2-propanol. As microplacas foram agitadas em vortex por 1 minuto a 70% da rotação máxima, e centrifugadas a 2250xg por 45 minutos (20°C). Em seguida, o sobrenadante foi descartado por inversão das microplacas, adicionou-se 150 µl de

etanol 70% (v/v) a -20°C , e as mesmas foram centrifugadas a $2250\times g$ por 5 minutos (20°C). Novamente, o sobrenadante foi descartado por inversão, e o resíduo deste eliminado através da centrifugação rápida das microplacas invertidas sobre papel absorvente (até $150\times g$, a 20°C , e sob 20% da aceleração/desaceleração máxima).

Procedeu-se à secagem do DNA plasmidial à temperatura ambiente por 1 hora., e o mesmo foi ressuspensionado pela adição de $40\ \mu\text{l}$ de ddH_2O , agitação em vortex (70% da rotação máxima, 1 minuto), centrifugação rápida até $100\times g$ e incubação a 37°C durante 10 minutos. O DNA plasmidial foi armazenado a -20°C , e a sua quantidade e integridade foram verificadas através da eletroforese de uma alíquota (5%) em gel de agarose 0,8% (TAE 1x, 10 V/cm), para o subsequente uso como molde na reação de sequenciamento.

3.5.2 REAÇÃO DE SEQUENCIAMENTO

A reação de sequenciamento fundamenta-se nos protocolos fornecidos pelo fabricante do *kit* utilizado (*Perkin Elmer Biosystems*, *kit BigDyeTM Terminator Cycle Sequencing v2.0*, Cat.N. 4314419), com poucas modificações. Estas reações foram realizadas em microplacas para termocicladores (96 poços, $200\ \mu\text{l}$ /poço), seguindo-se as coordenadas de posição dos clones. Em cada poço foram adicionados: $2\ \mu\text{l}$ de ddH_2O , $2\ \mu\text{l}$ de tampão “*Save Money*” (200mM Tris-HCl pH 9,0; 5mM $\text{MgCl}_2\cdot 6\text{H}_2\text{O}$), $2\ \mu\text{l}$ de *primer* [M13/pUC *Reverse* (5'-AGCGGATAACAATTTTCACACAGG-3') ou *Forward* (5'-CCCAGTCACGACGTTGTAAAACG-3')], $2\ \mu\text{l}$ do *mix* de sequenciamento (*BigDye*), e $2\ \mu\text{l}$ do DNA plasmidial molde.

As microplacas de sequenciamento foram agitadas em vortex por 30 segundos a 50% da rotação máxima, e centrifugadas rapidamente até $100\times g$. Procedeu-se à reação em termocicladores GeneAmp 9700 (*Perkin Elmer*), nas seguintes condições: desnaturação inicial a 96°C por 2 minutos; 35 ciclos de desnaturação a 96°C

por 45 segundos, anelamento a 50°C por 30 segundos, e extensão a 60°C por 4 minutos; extensão final a 60°C por 2 minutos; finalização a 4°C. Os produtos das reações de sequenciamento foram imediatamente purificados, ou alternativamente armazenados a –20°C por, no máximo, 2 dias, antes da purificação.

Na reação de sequenciamento ocorre a extensão de todos os fragmentos amplificáveis a partir do DNA molde (neste caso, os insertos dos clones das bibliotecas plastômicas), produzindo fragmentos complementares a este, e com tamanho variando de 2 até por volta de 600 nucleotídeos, conforme a atividade da DNA-polimerase. No *mix* utilizado (*BigDye*), existem os reagentes necessários para a produção destes fragmentos: o tampão da reação, MgCl₂, dNTPs, *Taq*-DNA-polimerase e os dideoxiribonucleotídeos (ddNTPs) marcados com cromóforos fluorescentes. Estes cromóforos são específicos para cada base nitrogenada (A,T,C,G) e, quando um deles é integrado na fita complementar crescente pela polimerase, a extensão é interrompida pela falta de um OH livre para a ligação do próximo nucleotídeo na cadeia. Como os dNTPs e os ddNTPs encontram-se em concentrações otimizadas, garante-se a produção de todos os fragmentos amplificáveis a partir do DNA molde, e com ddNTPs marcados como último nucleotídeo de cada fragmento.

3.5.3 PURIFICAÇÃO DOS PRODUTOS DA REAÇÃO DE SEQUENCIAMENTO

Para a purificação dos produtos da reação de sequenciamento, adicionou-se a cada poço das microplacas 14µl da solução NG (15 ml NaOAc 3 M pH 5,2; 15 µg glicogênio; 70 ml ddH₂O) e 50µl de etanol absoluto a –20°C. Procedeu-se à agitação em vortex (70% da rotação máxima, 1 minuto), e à incubação no gelo durante 15 minutos. Em seguida, as microplacas foram centrifugadas a 2250xg por 45 minutos (4°C), e o sobrenadante descartado por inversão. O resíduo deste sobrenadante (principalmente sais e nucleotídeos fluorescentes não-incorporados) foi eliminado através da centrifugação rápida das microplacas invertidas sobre papel absorvente (até 100xg, a 4°C, e sob 15% da aceleração/desaceleração máxima).

Em seguida, adicionou-se 150 µl de etanol 70% (v/v) a -20°C, e as microplacas foram centrifugadas a 2250xg por 10 minutos (4°C). Novamente, o sobrenadante foi descartado por inversão, e o resíduo deste eliminado através da centrifugação rápida das microplacas invertidas sobre papel absorvente (até 100xg, a 4°C, e sob 15% da aceleração/desaceleração máxima). O sedimento foi submetido à secagem a 37°C durante 20 minutos, e ressuspensão em 3µl do tampão de sequenciamento [80% (v/v) formamida deionizada; 20% *loading-buffer* (kit)]. Os produtos da reação de sequenciamento foram armazenados a 4°C por, no máximo, 5 dias.

3.5.4 SEQUENCIAMENTO

Os produtos das reações de sequenciamento acima descritas foram sequenciados através da metodologia utilizada pelo Projeto SUCEST/FAPESP, em sequenciadores automáticos ABI-377 (*Perkin Elmer*), de acordo com as instruções do fabricante. Este equipamento realiza a eletroforese dos produtos da reação de sequenciamento em gel de uréia-poliacrilamida com resolução capaz de separar fragmentos de DNA com diferença de tamanho de apenas 1 base, os quais migram no gel até passarem por uma região onde se incide um feixe *laser* (40mW). A fluorescência emitida pelos ddNTPs quando atingidos pelo feixe *laser* (região de leitura) é captada por um sensor, eletronicamente interpretada e enviada a um programa de coleta de dados em funcionamento no microcomputador conectado ao sequenciador. Cada tipo de base nitrogenada possui um cromóforo que emite fluorescência em comprimento de onda diferente, permitindo ao sensor e ao programa de coleta dos dados detectar qual das 4 bases foi atingida pelo *laser*.

Os programas de coleta (*ABI-Prism Collection*, *Perkin Elmer*) e de análise (*Sequencing Analysis*, *Perkin Elmer*) dos sinais emitidos pelo sequenciador geram uma imagem virtual do gel, com base na sequência detectada de emissões de fluorescência (correspondente à sequência de bases nitrogenadas do DNA que foi

sequenciado). A partir desta imagem virtual, pode-se extrair os cromatogramas e as sequências de nucleotídeos referentes a cada amostra sequenciada. Os cromatogramas, ou eletroferogramas, são representações gráficas da intensidade da fluorescência emitida por cada nucleotídeo do DNA sequenciado, e indicam a confiabilidade dos dados obtidos durante o sequenciamento. As sequências de nucleotídeos são obtidas na forma textual correspondente à sequência de bases nitrogenadas do DNA sequenciado, e podem ser diretamente utilizadas para a análise em programas específicos.

3.5.5 MONTAGEM DO GENOMA CLOROPLASTIDIAL

Inicialmente, as sequências foram analisadas através do programa BLAST/NCBI (*Basic Local Alignment Search Tool - National Center of Biotechnology Information, USA*) (Altschul et.al., 1990) para a verificação de homologia com o DNA cloroplastidial de outras espécies (principalmente milho e arroz), e a análise de alinhamento e montagem do plastoma foi realizada através do conjunto de programas *Phred-Phrap-Consed* (Ewing et.al., 1998; Ewing & Green, 1998; Gordon et.al., 1998), comumente utilizados nos Projetos GENOMA/FAPESP.

O algoritmo *Phred* (*Base-calling software with quality information*) executa a leitura dos picos de fluorescência (detectados pelo sequenciador automático e registrados nos cromatogramas), associa a respectiva base nitrogenada a cada pico e registra os valores de qualidade (confiabilidade) para cada base. Após esta análise, este programa rescreve os dados em um formato próprio para a utilização no algoritmo *Phrap*, aumentando a precisão da montagem subsequentemente obtida.

Esta montagem foi, portanto, virtualmente construída pelo programa *Phrap* (*Phragment assembly program*), específico para o alinhamento de sequências de DNA provenientes de bibliotecas *shot-gun* (composta por fragmentos gerados aleatoriamente). Dentre as suas características, este algoritmo (i) permite a utilização das sequências inteiras e não apenas as regiões de alta qualidade, (ii) utiliza uma combinação dos parâmetros oriundos do próprio programa e também dos

cromatogramas para melhorar a precisão da montagem, e (iii) constrói os *contigs* pelo alinhamento das regiões sobrepostas das sequências. *Contigs* são agrupamentos de sequências que apresentam regiões sobrepostas e adjacentes formando uma sequência maior contínua. A Figura 1 mostra esquematicamente a construção de um *contig*.

Figura 1 - Esquematização da montagem de um *contig*.

Pelo fato de o programa *Phrap* não permitir a visualização gráfica da montagem obtida, utilizou-se para esta finalidade o algoritmo *Consed*, específico para a edição dos resultados criados pelo *Phrap*. Nesta visualização pode-se verificar os valores de qualidade para cada base, para o alinhamento das sequências e para a sequência consenso. Também é possível editar eventuais regiões, de acordo com variações na montagem.

Como o volume de sequências geradas tornou-se elevado, e para a otimização e racionalização dos processos de análise, comparação com bancos de dados e montagem do plastoma, foi criada pelo colaborador Prof. João Paulo Kitajima (Laboratório de Bioinformática - Instituto de Computação/UNICAMP) uma *homepage* específica (www.sucest.lad.dcc.unicamp.br/sc-chloroplast), para a sistematização, gerenciamento e análise do banco de dados produzido durante a fase de sequenciamento dos clones das bibliotecas plastômicas. Além de informações gerais sobre este Projeto de

Pesquisa (metodologias e resultados), esta *homepage* permitiu a realização de análises indispensáveis para seu desenvolvimento, tais como:

- a) automação do recebimento e montagem do genoma cloroplastidial.
- b) análises comparativas (BLAST/NCBI) da sequência de DNA cloroplastidial de cana-de-açúcar com plastomas de outras espécies (*Zea mays*, *Oryza sativa*, *Nicotiana tabacum*, *Pinus thunbergii*, *Marchantia polymorpha*), indicando as posições e os índices de homologia.
- c) identificação das regiões de menor qualidade, para a finalização da montagem do plastoma, permitindo a elaboração de *primers* (descritos no item 3.6) que flanqueiam estas regiões.
- d) anotação automática da maioria das ORFs encontradas.

3.6 FINALIZAÇÃO DA MONTAGEM

Durante o processo de montagem da sequência genômica cloroplastidial de cana-de-açúcar, foram progressivamente adicionadas sequências (sob a forma de cromatogramas) correspondentes às extremidades (5' e 3') dos insertos dos clones das bibliotecas plastômicas. Estas sequências eram automaticamente analisadas quanto à qualidade (*Phred*), agrupadas de acordo com as sobreposições para a formação dos *contigs* (*Phrap*), e a montagem geral envolvendo os *contigs* formados era visualizada e analisada (*Consed*). Com o aumento das sequências depositadas e integradas na montagem, ocorreu o aumento do tamanho dos *contigs* simultaneamente à diminuição do número de *contigs*, o que era esperado. Contudo, e como geralmente é observado em projetos de sequenciamento genômico, surgiram na montagem regiões não cobertas pelas sequências geradas a partir das bibliotecas, denominadas *gaps* de sequência, e

também regiões sequenciadas mas com qualidade (confiabilidade) inferior, denominadas *gaps* de qualidade.

Procedeu-se, então, às estratégias para o sequenciamento direcionado dos *gaps* de sequência, e para o aumento da confiabilidade da sequência dos *gaps* de qualidade, descritas a seguir. Esta fase constituiu o fechamento da sequência do plastoma de cana-de-açúcar, ou *finnishing*. Com o *finnishing* da sequência, espera-se a formação de um único *contig* maior contendo toda a sequência do genoma em estudo. A fase de fechamento da sequência genômica tem como objetivos principais:

- a) unir os *contigs* em um único, o qual representa a sequência total em estudo;
- b) indicar regiões de menor qualidade de sequenciamento;
- c) indicar regiões internas dos clones importantes para o fechamento;
- d) analisar discrepâncias entre sequências de alta qualidade (mutações ou erros na montagem);

3.6.1 FECHAMENTO DOS GAPS DE SEQUÊNCIA

Para o fechamento da sequência do plastoma de cana-de-açúcar, foram utilizados *primers* específicos, baseados nas sequências flanqueadoras dos *contigs* supostamente vizinhos, e na homologia com o plastoma de milho, para extensão dos fragmentos por PCR (Figura 2). Por serem fragmentos de tamanho entre 2 e 8 Kb e exigirem atividade *proof-reading* da DNA-polimerase, a sua amplificação foi realizada com o *kit Elongase (Gibco)*, através do protocolo recomendado pelo fabricante.

Para o sequenciamento completo destes fragmentos obtidos, os mesmos foram sonificados, clonados para a obtenção de mini-bibliotecas, e sequenciados conforme a metodologia descrita nos itens 3.4 e 3.5. Os *primers* utilizados para o fechamento e o tamanho dos produtos de PCR obtidos são apresentados no Quadro 4.

Figura 2 - Localização dos *gaps* de sequência e dos *primers* utilizados para seu fechamento.

Fragmento (pb)	Primers-Sequências (5'→3')	Posição no plastoma/milho
A (1760)	A - GGTATTTACACAGAGGAATGG	117784
	A' - CTTGGTCGACCCCTTACGTC	107008
B (8314)	B - CACTAAGGAGAAGATAGAATC	48631
	B' - GTTTCGTACTCCGGGGTGTAG	56945
C (6658)	C' - TGGATCTCGGTTGTTCAATCC	137503
	C - TCCCATTCCGCCCAAACCTT	78581
D (6346)	D - GCAAGGACTCCCAAGCACACG	1176
	D' - GTAAGTTACATATCGACGTCG	7522
E (8022)	E' - CTTAGCATTCGGGAAAGACA	22075
	E - GGTAGCCATTCGAGAAGGAGG	14053

Quadro 4 - Primers utilizados para amplificação de regiões de interesse do ptDNA.

Como observado na Figura 2, o sequenciamento das bibliotecas plastômicas *shot-gun* produziu uma montagem do plastoma composta por 4 *contigs* principais e diversos outros menores, homólogos às regiões repetidas invertidas. Os *gaps*

flanquados pelos *primers* B/B', D/D' e E/E' localizaram-se na região do plastoma de cana-de-açúcar homóloga à LSC do plastoma de milho e, portanto, eram *gaps* de cópia única.

O algoritmo utilizado pelo programa *Phrap* não diferenciou a origem das sequências correspondentes às IRa e IRb, devido ao fato de estas repetições invertidas constituírem uma parte proporcionalmente grande do tamanho total deste genoma (~33%), e não formou automaticamente um *contig* que compreendesse a região IRa. Embora já existissem no banco de dados de cloroplasto de cana-de-açúcar as sequências correspondentes às duas IR's, faltavam elementos fixados na montagem para que o algoritmo *Phrap* fosse capaz de completá-la na região de IRa.

Estes elementos foram identificados como as 4 regiões de intersecção entre as repetições invertidas e as regiões de cópia única, indicadas na Figura 2 por um retângulo tracejado. De acordo com a distribuição dos *contigs*, e se considerando a sua homologia posicional com o plastoma de milho, foi possível desenhar *primers* que amplificassem as intersecções de IRa com SSC (*primers* A/A') e de IRa com LSC (*primers* C/C'). Isto foi realizado supondo-se que os *primers* A' e C', desenhados a partir da sequência conhecida da IRb de cana-de-açúcar, também fossem passíveis de anelamento com a IRa, em orientação oposta, visto que as repetições do plastoma são reciprocamente invertidas.

Além do uso de *primers* para o fechamento dos *gaps* de sequência, foram identificados 19 clones cujas extremidades 5' e 3' localizavam-se nas extremidades de *contigs* adjacentes obtidos na montagem do plastoma, sugerindo que o sequenciamento total destes clones (*finishing*) favoreceria a junção destes *contigs* para o avanço da montagem do plastoma (Figura 3). Assim, estes clones principais na montagem tiveram seus insertos amplificados via PCR convencional (94°C por 2 minutos; 30 ciclos de desnaturação a 94°C por 45 segundos, anelamento a 50°C por 30 segundos e amplificação a 72°C por 1,5 minutos; 72°C por 10 minutos; 4°C) utilizando-se os *primers* M13/pUC *Reverse* e *Forward*, e os produtos desta amplificação foram submetidos à sonicação, reparo das extremidades, seleção de fragmentos (800 pb) e clonagem para a obtenção de mini-bibliotecas (itens 3.4 e 3.5). Destas, 96 subclones

foram sequenciados e utilizados para a montagem dos 19 clones principais, cujas seqüências completas obtidas nos programas *Phred-Phrap-Consed* foram incluídas na montagem geral do plastoma.

Figura 3 - Esquematização dos clones cujo *finishing* favoreceu a junção de *contigs*.

3.6.2 FECHAMENTO DOS GAPS DE QUALIDADE

Para o fechamento dos *gaps* de qualidade, ou seja, de regiões com tamanho reduzido (menores que 1 kb), procedeu-se à identificação dos mesmos através da análise no programa *Consed* da montagem gerada pelo programa *Phrap*. Assim, foi possível identificar quais seqüências depositadas apresentavam qualidade (confiabilidade) inferior àquela necessária para o progresso da montagem geral do plastoma de cana-de-açúcar. O programa *Consed* mostra o valor da qualidade das bases da seqüência correspondente à cada clone da biblioteca, permitindo desta forma a determinação de quais clones deveriam ser novamente sequenciados para melhorar a sua qualidade e consequentemente o progresso da montagem geral.

3.7 ANÁLISES COMPARATIVAS E ANOTAÇÃO DOS GENES E *ycf's*

Após a obtenção e determinação da montagem final do plastoma de cana-de-açúcar, procedeu-se às análises comparativas com os plastomas de algumas espécies vegetais (milho, arroz e tabaco) e à anotação das regiões codificadoras e *ycf's*. Estas análises foram realizadas com a utilização de *softwares* específicos disponibilizados na *homepage*. O programa *Glimmer* (Delcher et.al., 1999) foi utilizado para a localização das supostas regiões de leitura aberta presentes no plastoma obtido, e o programa *trnScan-SE* (Lowe & Eddy, 1997) para a detecção de genes de RNA transportador. O programa BLAST viabilizou (i) a identificação de cada sequência codificadora através da análise de homologia com genes descritos em outros plastomas, (ii) a determinação da similaridade dos genes e ORFs do plastoma de cana-de-açúcar com os respectivos *operons* do plastoma de outras espécies e (iii) a realização de análises comparativas para regiões intergênicas.

3.8 ANÁLISE COMPARATIVA DE EDIÇÃO DE MRNA POR RT-PCR

Para a análise da ocorrência de edição de mRNA no gene *rpoB* plastidial de cana-de-açúcar, foram desenhados *primers* com base em regiões específicas deste gene no plastoma de milho (*Zea mays*). Este *operon* é um dos genes plastidiais que apresentam maior frequência de edição de mRNA (4 sítios no plastoma de milho; Maier et.al., 1995). Os *primers* utilizados para a amplificação de uma região de 400 pb que inclui os sítios de edição do transcrito do gene *rpoB* de cana-de-açúcar, assim como para a amplificação da mesma região no respectivo cDNA foram: *rpoB21829* (5' ATACCGAATTGTGATCAATC 3') e *rpoB22229'* (5' CCTAATTCACATTTTGTG 3').

Através da comparação das sequências genômica e do respectivo cDNA deste gene sem *introns*, é possível detectar os eventos de edição do mRNA nos cloroplastos, devido à ocorrência de substituição de uma base nitrogenada (C) por outra (U) nos sítios de edição. Isto também acarreta, nos respectivos sítios do cDNA sintetizado, substituições de uma base (G) por outra (A).

O material foliar coletado de cana-de-açúcar (SP 80-3280, obtida conforme item 3.1) foi imediatamente congelado em nitrogênio líquido, triturado em almofariz até partículas menores que 1,0 mm, aliquotados em tubos de microcentrífuga (aproximadamente 0,1 ml de material triturado por tubo de 1,5 ml) para o isolamento de RNA total com o reagente Trizol (*Gibco-BRL*), pelo método recomendado pelo fabricante. O RNA isolado foi analisado em gel de agarose 1,2 % (TAE 1x, tratado com DEPC 0,01 %, 8 V/cm). Aproximadamente 5,0 µg do RNA total foram tratados com DNase I livre de RNase (*Gibco-BRL*), de acordo com as instruções do fabricante. O tratamento do RNA total com DNase I é indicado para eliminar possíveis contaminações com DNA genômico, o que pode comprometer a amplificação correta da sequência dos cDNAs.

Posteriormente, o RNA total foi utilizado para a síntese do cDNA total foliar de cana-de-açúcar, obtido através do kit *SuperScriptTM Preamplification System* (*Gibco-BRL*), utilizando os hexâmeros randômicos do *kit* como *primers*.

Como controle da presença de DNA plastômico contaminante, e também para futura análise de edição, foram também desenhados *primers* para a amplificação da região editada em transcritos dos genes *ndhA* e *ndhB*, que possuem *intron*, sendo que cada *primer* localiza-se em um *exon* diferente, permitindo a distinção por tamanho dos produtos amplificados a partir do cDNA (sem *intron*) e a partir do DNA contaminante (com *intron*). Os *primers* utilizados foram: *ndhA114320* (5' TTCTTACTGTTATTGTATTC 3'), *ndhA116430* (5' GGGTAGAGGTAGAAACTATC 3'), *ndhB89230* (5' AGACCTAGCAGCTAAAAGAG 3'), e *ndhB91030* (5' GTGGTGCTAACGATTTAATA 3').

O cDNA total foi usado como molde para a amplificação da região de interesse com os *primers* citados, via PCR convencional (94°C por 2 minutos; 30 ciclos

de desnaturação a 94°C por 45 segundos, anelamento a 43°C por 30 segundos e amplificação a 72°C por 1,5 minutos; 72°C por 10 minutos; 4°C). Como controles, procedeu-se à amplificação da mesma região usando-se como molde o ptDNA de cana-de-açúcar, o seu RNA total não-tratado com DNase I, e uma alíquota do cDNA total sintetizado a partir do RNA total não-tratado com DNase I.

Uma alíquota de cada produto de RT-PCR foi analisada em gel de agarose, e, após a verificação do tamanho esperado para cada produto de amplificação, estes foram diretamente utilizados para a reação de sequenciamento (ver item 3.5). O alinhamento das sequências foi analisado através dos programas *Vector NTI Suite v.6.0* (Informax Inc., USA) e *ClustalW* (<http://www2.ebi.ac.uk/clustalw/>).

4 RESULTADOS E DISCUSSÃO

4.1 MATERIAL VEGETAL

A metodologia de propagação utilizada apresentou-se adequada, e não foram observadas diferenças significativas na qualidade do ptDNA obtido a partir de material proveniente de casa-de-vegetação ou de cultivo *in vitro*. Observou-se que as folhas entre 5,0 e 15,0 cm de comprimento são mais favoráveis ao isolamento de cloroplastos e do seu DNA, o que pode ocorrer devido à menor quantidade de fibras e sílica no tecido, e também provavelmente à menor ação de DNAses endógenas em folhas neste estágio.

A manutenção das plantas no escuro (2 dias) favoreceu o rendimento do isolamento de cloroplastos em relação ao material utilizado sem este tratamento, possivelmente à menor ruptura de organelas devido à presença de grânulos de amido no seu interior. Para o método que utiliza Triton X-100 não se observou a necessidade de um período no escuro, porque o método não requer trituração do material vegetal em homogenizador.

4.2 ISOLAMENTO DE CLOROPLASTOS E DO PTDNA

Considerando-se os resultados dos vários testes realizados com diferentes métodos para o isolamento dos cloroplastos de cana-de-açúcar, observou-se que a eficiência dos procedimentos comumente utilizados na literatura (item 3.2) era inadequada, pois em todos os testes foram obtidas misturas do DNA cloroplastidial com o DNA nuclear. Lâminas de microscopia óptica preparadas contendo material foliar homogenizado proveniente de cana-de-açúcar e tabaco (*Nicotiana tabacum*) indicaram que os cloroplastos de cana-de-açúcar são aparentemente maiores que os cloroplastos de tabaco (Figura 4), o que justificaria o co-isolamento do DNA nuclear e cloroplastidial observado nos testes anteriores com os métodos fundamentados em características físicas (tamanho, densidade) das organelas. Outra possibilidade é a formação de agregados de cloroplastos (Pike, 1999), dada a existência de comunicação física por microtúbulos entre estas organelas (Gray et.al., 1999).

Figura 4 - Visualização (400x) de cloroplastos de células próximas à epiderme foliar de plantas adultas de cana-de-açúcar (A) e de tabaco (B), em microscópio com filtro de profundidade (barra=10µm).

O protocolo com CsCl-bisbenzimidazila apresentou resultados mostrando que a separação entre as bandas correspondentes ao DNA nuclear e ao DNA cloroplastidial é mínima, da ordem de alguns milímetros. Isto pode ocorrer quando o método é utilizado para espécies vegetais superiores (Taylor et.al.,1993), que apresentam o DNA nuclear metilado, o que reduz a capacidade da bisbenzimidazila em criar

diferenças significativas de densidade entre o DNA nuclear e o cloroplastidial na sedimentação durante a ultracentrifugação.

O método que utiliza Triton X-100 forneceu os melhores resultados em relação à integridade e pureza do DNA cloroplastidial. O mesmo não se baseia em características físicas, mas na composição lipídica diferencial entre as membranas do cloroplasto e a carioteca, principalmente em relação aos glicerolipídios.

A base deste método (Mariac et.al., 2000) é a composição lipídica diferencial entre as membranas do cloroplasto e a carioteca, principalmente em relação aos constituintes glicerolipídicos. A composição lipídica das membranas do cloroplasto também difere consideravelmente das membranas do retículo endoplasmático e das mitocôndrias (Taiz & Zeiger, 1998). Aproximadamente 87% dos constituintes das membranas cloroplastidiais são galactosildiacilgliceróis e sulfolipídios, ambos ausentes na carioteca e nas membranas do retículo endoplasmático e das mitocôndrias. Após a lise, os núcleos e os cloroplastos não-lisados foram sedimentados a 1000×g e o sobrenadante (contendo o ptDNA liberado na lise específica) foi utilizado para a extração do ptDNA.

Figura 5 - DNA cloroplastidial de cana-de-açúcar isolado com Triton X-100. M=1 kb DNA ladder (*Gibco*); 1: 800 ng de DNA plasmidial clivado com *EcoRI* (demonstrando eficiência da enzima); 2: DNA cloroplastidial de cana-de-açúcar clivado com *EcoRI* (com pouco *smear*).

A lise específica dos cloroplastos (0°C, 40 minutos) aparentemente liberou o DNA cloroplastidial para o tampão, do qual foi isolado e purificado posteriormente pelo método com CTAB (ver item 3.3). A análise de restrição em gel de agarose apresentou resultados favoráveis, onde várias bandas nítidas isoladas eram visíveis e pouco *smear* resultante da digestão do DNA nuclear foi detectado (Figura 5). O período de lise com Triton X-100 mais favorável foi de 40 minutos. O ptDNA isolado

através deste método foi prioritariamente utilizado para a construção de bibliotecas de DNA cloroplastidial.

4.3 CONSTRUÇÃO DE BIBLIOTECAS DE FRAGMENTOS DO PTDNA

O processo de fragmentação mecânica através de sonicação apresentou-se adequado para a construção de bibliotecas plastômicas, principalmente devido ao seu caráter aleatório e uniformizador do tamanho dos insertos de ptDNA a serem clonados nos vetores plasmidiais (bibliotecas *shot-gun*). Em relação ao controle inicial (clonagem de bandas produzidas por restrição enzimática com *EcoRI*, sequenciamento e análise), este permitiu a verificação de elevada homologia dos insertos com os plastomas de milho (*Zea mays*) arroz (*Oryza sativa*) e cevada (*Hordeum vulgare*), nas regiões de *operons* de rRNA, genes de tRNAs e do gene *rpoC*, localizando-se na região repetida invertida.

A seleção dos clones foi realizada via seleção em X-Gal/IPTG (operon *lacZ* do plasmídio). Visando-se ajustar a metodologia, foram primeiramente construídas 2 bibliotecas plastômicas *shot-gun* (PT1 e PT2) a partir do ptDNA (método com Triton X-100), utilizando-se o sistema de microplacas (96 *poços*) para o armazenamento e manipulação dos clones. As bibliotecas plastômicas PT1 e PT2 foram amostradas pelo sequenciamento de 384 clones, os quais apresentaram ainda considerável contaminação com DNA nuclear (30%) e tamanho médio de inserto de aproximadamente 600 pb, inferior ao tamanho desejado (dados não apresentados). Deste modo, foi necessária a obtenção de novas bibliotecas que apresentassem menores níveis de contaminação nuclear e insertos entre 1-2 Kb (Figura 6), os quais favorecem a determinação de suas sequências de nucleotídeos a partir das duas extremidades, utilizando-se os *primers Forward* e *Reverse*, e permitindo a sobreposição na região interna do inserto.

Figura 6 - (A) Seleção de fragmentos de ptDNA com tamanho entre 1,0 e 1,6 kb após a sonicação, através de excisão da região do gel (➡). (B) DNA plasmidial dos clones, utilizado para a reação de sequenciamento. 1, 2 e 3: plasmídios com insertos de 2,0 , 1,5 e 1,0 kb, respectivamente.

A nomenclatura utilizada para a identificação dos clones foi estabelecida com base nos parâmetros referentes ao Projeto SUCEST/FAPESP, de modo a viabilizar a funcionalidade das análises através da *homepage*. Os clones passaram, então, a ter a denominação PSEQXXYYNNN.z , onde: PS = *plastid sugarcane*, EQ = grupo EQ-SUCEST/FAPESP , XX = número da biblioteca plastômica (00, 01 ou 50) , YY = número da microplaca , NNN = coordenada da posição do clone na microplaca , e z = orientação da extremidade sequenciada do clone (g=5' e b=3').

As bibliotecas construídas consistiram de:

- PSEQ00 e PSEQ01 : fragmentos de ptDNA total sonicado;
- PSEQ50: fragmentos do produto de PCR correspondente a uma região amplificada do ptDNA de cana-de-açúcar a partir de *primers* desenhados com base em uma região específica do plastoma de milho (interna à IRa), a qual se apresentou homóloga em ambas as espécies.

- PSEQ51 a PSEQ55: fragmentos do produto de PCR correspondentes às regiões importantes para o fechamento da montagem, amplificadas via *primers* desenhados com base na sequência do plastoma de milho (intersecções entre as regiões IR, SSC e LSC).

- PSEQ56 a PSEQ64: mini-bibliotecas para o *finishing* dos clones principais para o fechamento da montagem do plastoma.

As verificações do tamanho médio dos insertos nas bibliotecas demonstraram que o mesmo era de aproximadamente 1,0 Kb, o que favorecia o sequenciamento (exceto para PSEQ56-64, cujo tamanho médio dos insertos selecionado e verificado foi de 800 pb).

4.4 SEQUENCIAMENTO E ANÁLISE DOS CLONES (MONTAGEM DO PLASTOMA)

As sequências dos clones obtidos foram analisadas através do programa BLAST/NCBI (*National Center of Biotechnology Information*) para a verificação de homologia com o DNA cloroplastidial de outras espécies, principalmente milho e arroz.

Com o sequenciamento das bibliotecas plastômicas PT1 e PT2, vários novos clones foram caracterizados, e a sua análise (*BLAST/NCBI*) demonstrou elevada homologia (acima de 97%) com várias regiões dos plastomas de, principalmente, milho (*Zea mays* e *Zea perennis*), arroz, cevada e trigo. Foi observada alta homologia dos clones obtidos tanto com regiões intergênicas quanto codificadoras de *operons* cloroplastidiais relacionados com subunidades dos fotossistemas (genes *psbA*, *psbE*, *psbG*, *psbF*, *psbK*, *psbI*, *pslA*, *psaB*), ATPase (*atpA*, *atpH*, *atpI*, *atpB*), RNA-polimerase plastidial (*rpoB*, *rpoC*), NADH-desidrogenase plastidial (*ndhB*, *ndhC*, *ndhI*), ribulose-1,5-bisfosfato-carboxilase/oxigenase (*rbcL*), proteínas ribossômicas (*rps2*, *rps16*), RNA ribossômico plastidial (*rRNA4.5S*, *rRNA16S*, *rRNA23S*), RNA transportador plastidial (*tRNL*, *tRNC*, *tRNK*, *tRNS* e *tRNI*), citocromo *b₆f* (*petE*) e os ORF's (genes putativos) denominados *orf31* e *orf48*.

Após a implementação e disponibilização dos serviços da *homepage* no Laboratório de Bioinformática (UNICAMP), a rotina de sequenciamento intensificou-se, aumentando exponencialmente a análise de sequências, que passou a ser realizada automaticamente através de programas específicos para este objetivo (*Phred-Phrap-Consed*), permitindo a otimização do processamento das sequências obtidas e a racionalização da montagem do plastoma de cana-de-açúcar.

Os resultados do sequenciamento e da montagem do plastoma foram atualizados diariamente, de acordo com o recebimento de novos lotes de sequências (microplacas). Foram sequenciadas 61 placas (23 da biblioteca PSEQ00, 12 da PSEQ01, 2 da PSEQ50, 13 de PSEQ51-55 e 11 de PSEQ56-66). Algumas placas foram sequenciadas mais de uma vez. Totalizou-se 6266 sequências (correspondentes aos

insertos de 6266 clones). Deste total de sequências, 4880 foram incorporadas na montagem de *contigs* e 1386 (28,4%) mantiveram-se como *singletons*, ou seja, não foram incluídos na montagem de *contigs*. A análise destes *singletons* indicou que são sequências cloroplastidiais de baixa qualidade ou são regiões de DNA nuclear, caracterizando a contaminação já esperada para esta espécie significativamente poliplóide nas bibliotecas, de acordo com o padrão de restrição das amostras de ptDNA obtido (ver item 4.2).

Foram obtidos 382 *contigs*, e o maior deles apresentou 142033 pb. Todos os demais *contigs* são menores que 4,0 Kb, e, apesar de apresentarem alinhamento com várias regiões do *contig* maior, não foram incluídos na montagem devido à não-concordância com os parâmetros estabelecidos pelo programa de montagem (*Phred-Phrap*). Isto indicou que o *contig* de 142033 pb compreendia o plastoma de cana-de-açúcar, o que também foi sugerido pelo padrão de homologia com o plastoma de milho. O alinhamento dos *contigs* e a análise comparativa com outros plastomas de interesse são apresentados na Figura 7.

Como observado nestes gráficos de homologia estrutural com os plastomas das 3 espécies vegetais consideradas (milho, arroz e tabaco), verificou-se os seguintes aspectos principais:

a) A homologia com tabaco ocorre envolvendo apenas trechos dos *contigs*, os quais apresentam identidade com várias regiões do plastoma desta dicotiledônea. Isto ocorre envolvendo regiões codificadoras conservadas em plastomas de diferentes grupos vegetais.

b) O plastoma de arroz mostrou-se pouco homólogo à maioria dos *contigs* de cana-de-açúcar, sendo observada homologia apenas nas regiões codificadoras. Isto pode ter ocorrido devido à descontinuidade de homologia entre as regiões intergênicas de cana-de-açúcar e de arroz.

c) A homologia estrutural entre os plastomas de milho e cana-de-açúcar apresentou-se significativamente elevada. Isto indica que tanto regiões codificadoras como intergênicas são muito similares nestas duas espécies.

Observou-se, ainda, que a montagem obtida para o plastoma de cana-de-açúcar (*contig* com 142033 pb) apresentava uma sobreposição de 296 nucleotídeos entre o início e o final da mesma (graficamente representado na Figura 8). O programa utilizado (*Phrap*) pode gerar estes artefatos de montagem, devido a parâmetros computacionais internos. A formação de sobreposição entre o início e o final da montagem indicou o fechamento da estrutura circular do plastoma, como esperado, e o tamanho total de 141737 pb.

Figura 7 - Análise de homologia (*crossmatch*) entre o plastoma de cana-de-açúcar e os plastomas de milho (A), arroz (B) e tabaco (C). As linhas verdes espessas representam os plastomas destas 3 espécies, e as linhas vermelhas representam os *contigs* constituintes da montagem do plastoma de cana-de-açúcar.

Através da análise de homologia da sequência total obtida contra ela própria (BLAST), foi possível identificar a localização das repetições invertidas (IR's). Deste modo, obteve-se que o plastoma de cana-de-açúcar é constituído por uma molécula de DNA circular de 141737 pb, que possui duas regiões repetidas invertidas

(IRa e IRb) cada uma com 22794 pb, separando uma região de cópia única longa de 83604 pb (LSC) e uma região de cópia única curta de 12545 pb (SSC).

Figura 8 - *Crossmatch* linear entre os plastomas de cana-de-açúcar e milho. Observar sobreposição a partir de aproximadamente 142000 pb.

De acordo com análises comparativas realizadas entre os plastomas de arroz e milho, existem regiões de extrema divergência entre estes dois sistemas genéticos, envolvendo deleções, substituições e mesmo migrações de genes cloroplastidiais para o núcleo. As regiões referentes aos genes cloroplastidiais *ycf2* e

accD (forma procariótica da acetil-CoA-carboxilase) descritas como divergentes entre os plastomas já sequenciados de espécies vegetais superiores (Maier et.al., 1995) mostraram-se praticamente idênticas entre cana-de-açúcar e milho. Outros dados obtidos sugerem que o plastoma de cana-de-açúcar assemelha-se consideravelmente ao de milho (Quadro 5).

Região / Parâmetro	ptDNA <i>Saccharum</i>	ptDNA <i>Zea</i>
LSC	83604	82355
SSC	12545	12536
<i>lra</i>	22794	22748
<i>lrb</i>	22794	22748
Total de Bases	141737	140387
%GC	38,48	38,46

Quadro 5 - Composição de nucleotídeos dos plastomas de cana-de-açúcar e milho.

A similaridade com milho também foi evidenciada pela amplificação via PCR convencional de um fragmento do plastoma de cana-de-açúcar correspondente a uma região exclusiva do plastoma de milho e ausente no de arroz. Esta região compreende aproximadamente 3700 pb entre os genes *trnI* e *trnL*, nas IR's. Este fragmento do ptDNA de cana-de-açúcar foi utilizado na construção da biblioteca PSEQ50, e apresentou identidade maior que 98% com a região específica do plastoma de milho, indicando maior homologia de organização estrutural com o plastoma desta espécie, assim como a divergência em relação ao plastoma de arroz.

De acordo com a estimativa inicial do tamanho total do plastoma de cana-de-açúcar de aproximadamente 140000 pb, e para uma cobertura de 8 vezes do mesmo, estabeleceu-se uma projeção do mínimo de 4000 clones para o seu fechamento, de acordo com as condições materiais de sequenciamento. Como citado, foram obtidos 4880 clones integrados em *contigs*.

A presença de regiões repetidas invertidas proporcionalmente grandes em relação ao tamanho total do genoma cloroplastidial ($\cong 30\%$) dificultou a montagem do

plastoma devido a alguns parâmetros dos programas disponíveis incompatíveis com esta característica, mas o sequenciamento contínuo das bibliotecas, o sequenciamento direcionado das regiões dos *gaps* (via *primers*) e o *finishing* dos clones-"chave" para a montagem auxiliaram na obtenção da estrutura final do plastoma de cana-de-açúcar.

4.5 ANOTAÇÃO DOS GENES E *ycf*'S (MAPA DO PLASTOMA)

Para a fase de anotação dos genes e *ycf*'s, foram utilizados programas e algoritmos específicos para a localização, identificação e análise de homologia de todos os trechos potencialmente codificadores. O programa *Glimmer* foi utilizado para a detecção das ORFs, e o *tRNScan* para a localização de genes de RNA transportador. O algoritmo BLAST foi utilizado para a localização dos genes de RNA ribossômico (comparativamente aos genes de rRNA de milho), e para a identificação de todas as ORFs codificadoras de proteínas através da homologia com genes já conhecidos em outros plastomas.

O Anexo A apresenta a identificação dos genes, *ycf*'s e ORFs encontrados no plastoma de cana-de-açúcar, e as respectivas homologias com os mesmos genes de outros plastomas. O Anexo B apresenta os genes de tRNA encontrados no plastoma de cana-de-açúcar. As sequências e a anotação destes genes, *ycf*'s e ORFs estão em processo de depósito no GenBank/NCBI.

A maior homologia das regiões codificadoras com os respectivos genes de milho ocorre devido à total disponibilidade da sequência do plastoma desta espécie no GenBank, sendo a única representante, até o momento, de um DNA cloroplastidial completo de gramínea com fotossíntese C₄. Apesar disto, observou-se que os genes *ndhF* (NADH-desidrogenase), *psaA* (PS I), *atpB* (ATP-sintase) e *matK* (maturase) do plastoma de cana-de-açúcar apresentam maior identidade com os respectivos homólogos de outra gramínea C₄, sorgo (*Sorghum bicolor*), do que com os de milho. Alguns genes de cana-de-açúcar, porém, mostraram maior identidade com homólogos de arroz (*ndhE*,

psaB, *psaC*, *psbA*, *psbD*, *psbE*, *psbF*, *psbL*), trigo (*psaI*, *ycf9*) e cevada (*psbC*). Dos genes e *ycf*'s identificados no plastoma de cana-de-açúcar, 80,2% mostraram-se mais homólogos aos genes correspondentes de milho do que de qualquer outro plastoma. As Figuras 9 e 10 mostram as categorias gênicas encontradas no plastoma de cana-de-açúcar e a identidade (%) das categorias gênicas encontradas no plastoma de cana-de-açúcar com homólogos de mono e dicotiledôneas, respectivamente.

Figura 9 - Categorias gênicas encontradas no plastoma de cana-de-açúcar.

Semelhantemente ao plastoma de milho, observou-se a presença de 40 genes de RNA transportador (30 espécies de tRNA) no plastoma de cana-de-açúcar (Anexo B), com número de cópias variável. Estes genes possuem anti-códons correspondentes aos 20 aminoácidos proteicos e são suficientes para a leitura dos 61 códons para aminoácidos, de acordo com Maréchal-Drouard et.al. (1991), analisando espécies cloroplastidiais de tRNAs em vários outros organismos vegetais superiores.

Identidade dos grupos gênicos cloroplastidiais de cana-de-açúcar com os homólogos de milho

Figura 10 - Identidade (%) dos genes do plastoma de cana-de-açúcar com os respectivos homólogos de milho.

De acordo com a anotação obtida, foi construído um mapa genético virtual do plastoma de cana-de-açúcar (programa *Vector NTI Suite v.6.0*), com a localização dos genes e *ycf's*, assim como a identificação das regiões LSC, SSC, IRa e IRb (Figura 11). Os genes representados na parte externa do mapa são transcritos no sentido horário.

Figura 11 - Mapa genético virtual do plastoma de cana-de-açúcar.

4.6 ANÁLISE COMPARATIVA DE EDIÇÃO DO MRNA

Os produtos amplificados com os primers *rpoB21829* e *rpoB22229'* a partir do cDNA total apresentaram-se com o tamanho esperado (~420 pb). O RNA total de cana-de-açúcar e os produtos amplificados a partir do cDNA são apresentados na Figura 12.

Figura 12 - RT-PCR para análise de edição em transcritos do gene *rpoB*. (A) RNA total de cana-de-açúcar isolado de plantas de campo (1-3), *in vitro* (4-6) e de estufa (7), este sendo usado para a síntese de cDNA. L=RNA ladder (*Gibco*). (B) Bandas correspondentes às regiões editadas dos genes *rpoB*, *ndhA* e *ndhB* de cana-de-açúcar, amplificadas a partir do ptDNA (1, 5, 9), RNA total não-tratado com DNase I (2, 6, 10), cDNA obtido de RNA não-tratado com DNase I (3, 7, 11), e cDNA obtido de RNA tratado com DNase I (4, 8, 12). Observar as bandas amplificadas a partir de DNA contaminante (amostras 6, 10 e 11), cujo tamanho inclui os *introns* de *ndhA* e *ndhB*; para a análise de edição em *rpoB*, utilizou-se a amostra 4, oriunda do RNA tratado com DNase I (4, 8, 12). M=1 kb DNA ladder (*Gibco*).

Foi detectada a presença de DNA contaminante no RNA total)e, portanto, no cDNA obtido a partir deste, indicando a necessidade do tratamento com a DNase I anteriormente à síntese do cDNA. Na Figura 12.B (amostras 4, 7, 11, 12) é possível identificar as bandas correspondentes às regiões amplificadas a partir do cDNA, com tamanhos esperados de 420 pb (*rpoB*), 1086 pb (*ndhA*) e 1095 pb (*ndhB*).

As sequências relativas à região do gene *rpoB* foram analisadas e confirmadas quanto à qualidade dos cromatogramas gerados (programa *Vector NTI Suite 6.0*), e procedeu-se ao seu alinhamento pelo programa *ClustalW*. Como controle, este alinhamento (Figura 13) incluiu as sequências do gene *rpoB* amplificadas (Figura 12.B) a partir do DNA plasmídico (*rpoBDNA*, 1), do RNA total (*rpoBRNA*, 2), do cDNA obtido a partir do RNA não-tratado com DNase I (*rpoBcDNA*, 3), e do cDNA obtido a partir do RNA tratado com DNase I (*rpoBcDNA*, 4).

			22084*	
rpoBDNA	...	AAAACTCCAAAATAGCTTTTTCTTTGACTCAATCCTCTTCTCTCCTTAGCATTCG	120	
rpoBRNA	...	AAAACTCCAAAATAGCTTTTTCTTTGACTCAATCCTCTTCTCTCCTTAGCATTCG	120	
rpoBcDNA-	...	AAAACTCCAAAATAGCTTTTTCTTTGACTCAATCCTCTTCTCTCCTTAGCATTCG	120	
rpoBcDNA+	...	AAAACTCCAAAATAGCTTTTTCTTTGACTCAATCCTCTTCTCTCCTTAGCATTCG	120	
	...	*****		
			22027*	
rpoBDNA	GGAAAGACAAGAAAATTTTCGGGGTAGGAAACATTATCTAAAATTTCTCTTAGATTCCAAC	180		
rpoBRNA	GGAAAGACAAGAAAATTTTCGGGGTAGGAAACATTATCTAAAATTTCTCTTAGATTCCAAC	180		
rpoBcDNA-	GGAAAGACAAGAAAATTTTCGGGGTAGGAAACATTATCTAAAATTTCTCTTAGATTCCAAC	180		
rpoBcDNA+	GGAAAGACAAGAAAATTTTCGGGGTAGGAAACATTATCTAAAATTTCTCTTAGATTCCAAC	180		
	*****	***		
			22012*	
rpoBDNA	CCATAGCTGATGATAGAAGTAAAATAGATATCTTTTGTTTTCTACTCACGCGAGCCCATA	240		
rpoBRNA	CCATAGCTGATGATAGAAGTAAAATAGATATCTTTTGTTTTCTACTCACGCGAGCCCATA	240		
rpoBcDNA-	CCATAGCTGATGATAGAAGTAAAATAGATATCTTTTGTTTTCTACTCACGCGAGCCCATA	240		
rpoBcDNA+	CCATAGCTGATAATAGAAGTAAAATAGATATCTTTTGTTTTCTACTCACGCGAGCCCATA	240		
	*****	*****		
			21935*	
rpoBDNA	TCCTTTCTTTTTTATCAATTGCTAATTCCGACCTTCTCCCCAATCTGATATTATAG...	300		
rpoBRNA	TCCTTTCTTTTTTATCAATTGCTAATTCCGACCTTCTCCCCAATCTGATATTATAG...	300		
rpoBcDNA-	TCCTTTCTTTTTTATCAATTGCTAATTCCGACCTTCTCCCCAATCTGATATTATAG...	300		
rpoBcDNA+	TCCTTTCTTTTTTATCAATTGCTAATTCCAACCTTCTCCCCAATCTGATATTATAG...	300		
	*****	*****		

Figura 13 - Alinhamento das seqüências do gene *rpoB* para análise de edição.

(*) posição do sítio de edição no plastoma de milho.

Observou-se a edição dos 4 sítios nas mesmas posições já descritas para este gene no plastoma de milho (*Zea mays*). Dos 4 sítios de edição já detectados para este gene (localizados em uma região desnecessária para a atividade da subunidade β da RNA-polimerase plastidial), 3 são editados também em arroz (*Oryza sativa*), cevada (*Hordeum vulgare*) e tabaco (*Nicotiana tabacum*) (Corneille et.al., 2000), e apenas em milho e cana-de-açúcar o sítio referente ao aminoácido 206 da proteína é editado, levando à substituição de uma prolina por uma leucina. Estes resultados indicam, além da significativa similaridade estrutural entre os plastomas das espécies gramíneas C₄, uma possível identidade entre os mecanismos de regulação da expressão gênica plastidial, especialmente em relação ao mecanismo de edição de mRNA.

Como continuidade deste trabalho, pode ser citado o detalhamento da análise de edição de mRNA em cana-de-açúcar, a investigação das funções dos *ycf's* encontrados, a realização de análises evolutivas e filogenéticas com base em regiões intra e intergênicas específicas entre os plastomas de diferentes espécies, e o

desenvolvimento de vetores de transformação para a inserção dirigida de genes exógenos no DNA cloroplastidial de cana-de-açúcar.

5 CONCLUSÕES

O plastoma de cana-de-açúcar é constituído por uma molécula de DNA circular de 141737 pb, que possui duas regiões repetidas invertidas (IRa e IRb) cada uma com 22794 pb, separando uma região de cópia única longa de 83604 pb (LSC) e uma região de cópia única curta de 12545 pb (SSC).

O plastoma de cana-de-açúcar sequenciado apresenta todos os grupos funcionais de genes cloroplastidiais de milho, além de *ycf's* e ORFs também encontradas no plastoma de outras espécies, embora ainda com função desconhecida.

Observou-se maior identidade entre os plastomas de cana-de-açúcar e milho do que entre cana-de-açúcar e arroz, inclusive com base na organização estrutural. Isto sugere uma evolução comum entre os plastomas das gramíneas C₄, gerando novos elementos para a pesquisa deste tipo de fotossíntese e metabolismo dos plastídeos, assim como para a tecnologia de transformação de cloroplastos de cana-de-açúcar.

ANEXOS

**ANEXO A - CONTEÚDO GÊNICO E HOMOLOGIA DO PLASTOMA DE CANA-DE-AÇÚCAR
COM OS RESPECTIVOS GENES DO PLASTOMA DE OUTRAS ESPÉCIES.**

* pseudogenes / genes fragmentados

** aa=aminoácidos da proteína

Grupo Funcional	Gene	> % Identidade (aa**)	Plastoma homólogo
RNAs ribossômicos	<i>rrn16</i>	99,0	<i>Zea</i>
	<i>rrn23</i>	98,0	<i>Zea</i>
	<i>rrn4.5</i>	100,0	<i>Zea</i>
	<i>rrn5S</i>	100,0	<i>Zea</i>
	<i>rps2</i>	99,0	<i>Zea</i>
Subunidade ribossômica menor	<i>rps3</i>	100,0	<i>Zea</i>
	<i>rps4</i>	100,0	<i>Zea</i>
	<i>rps7</i>	100,0	<i>Zea</i>
	<i>rps8</i>	100,0	<i>Zea</i>
	<i>rps11</i>	99,0	<i>Zea</i>
	<i>rps12</i>	100,0	<i>Zea</i>
	<i>rps14</i>	87,0	<i>Zea</i>
	<i>rps15</i>	82,0	<i>Zea</i>
	<i>rps16</i>	97,0	<i>Zea</i>
	<i>rps18</i>	100,0	<i>Zea</i>
	<i>rps19</i>	100,0	<i>Zea</i>
	<i>rpl2</i>	91,0	<i>Zea</i>
	<i>rpl14</i>	99,0	<i>Zea</i>
Subunidade ribossômica maior	<i>rpl16</i>	100,0	<i>Zea</i>
	<i>rpl20</i>	88,0	<i>Zea</i>
	<i>rpl22</i>	91,0	<i>Zea</i>
	<i>rpl23</i>	100,0	<i>Zea</i>
	<i>rpl32</i>	77,0	<i>Zea</i>
	<i>rpl33</i>	98,0	<i>Zea</i>
	<i>rpl36</i>	99,0	<i>Zea</i>
	<i>rpoA</i>	98,0	<i>Zea</i>
	<i>rpoB</i>	96,0	<i>Zea</i>
	<i>rpoC1</i>	94,0	<i>Zea</i>
Subunidades da RNA-polimerase	<i>rpoC2</i>	88,0	<i>Zea</i>
	<i>ndhA</i>	95,0	<i>Zea</i>
	<i>ndhB</i>	99,0	<i>Zea</i>
Subunidades da NADH-desidrogenase	<i>ndhC</i>	90,0	<i>Zea</i>
	<i>ndhD</i>	92,0	<i>Zea</i>
	<i>ndhE</i>	61,0	<i>Oryza</i>
	<i>ndhF</i>	89,0	<i>Zea, Sorghum</i>
	<i>ndhG</i>	82,0	<i>Zea</i>
	<i>ndhH</i>	99,0	<i>Zea</i>
	<i>ndhI</i>	98,0	<i>Zea</i>
	<i>ndhJ</i>	100,0	<i>Zea</i>
	<i>ndhK</i>	94,0	<i>Zea</i>
	<i>psaA</i>	96,0	<i>Sorghum</i>
Subunidades do PS I	<i>psaB</i>	95,0	<i>Zea, Oryza</i>
	<i>psaC</i>	100,0	<i>Oryza</i>
	<i>psaI</i>	100,0	<i>Triticum</i>
	<i>psaJ</i>	100,0	<i>Zea</i>
Subunidades do PS II	<i>psbA</i>	95,0	<i>Oryza, Zea</i>
	<i>psbB</i>	97,0	<i>Zea</i>
	<i>psbC</i>	90,0	<i>Zea, Hordeum</i>
	<i>psbD</i>	96,0	<i>Oryza</i>

	<i>psbE</i>	100,0	<i>Zea, Oryza</i>
	<i>psbF</i>	100,0	<i>Zea, Oryza</i>
	<i>psbH</i>	100,0	<i>Zea</i>
	<i>psbI</i>	100,0	<i>Zea</i>
	<i>psbJ</i>	99,0	<i>Zea</i>
	<i>psbK</i>	100,0	<i>Zea</i>
	<i>psbL</i>	100,0	<i>Zea, Oryza</i>
	<i>psbM</i>	99,0	<i>Zea</i>
	<i>psbN</i>	100,0	<i>Zea</i>
	<i>psbT</i>	100,0	<i>Zea</i>
Subunidade grande da rubisco	<i>rbcL</i>	95,0	<i>Zea</i>
Subunidades do citocromo b_6f	<i>petA</i>	95,0	<i>Zea</i>
	<i>petB</i>	93,0	<i>Zea</i>
	<i>petD</i>	99,0	<i>Zea</i>
	<i>petG</i>	100,0	<i>Zea</i>
	<i>petE</i>	96,0	<i>Zea</i>
Subunidades da ATP-sintase	<i>atpA</i>	96,0	<i>Zea</i>
	<i>atpB</i>	99,0	<i>Sorghum, Zea</i>
	<i>atpE</i>	91,0	<i>Zea</i>
	<i>atpF</i>	90,0	<i>Zea</i>
	<i>atpH</i>	100,0	<i>Zea</i>
	<i>atpI</i>	87,0	<i>Zea</i>
Fator de iniciação de tradução I	<i>infA</i>	85,0	<i>Zea</i>
Maturase	<i>matK</i>	98,0	<i>Sorghum</i>
Protease	<i>clpP</i>	99,0	<i>Zea</i>
Proteína da membrana envoltória	<i>cemA</i>	93,0	<i>Zea</i>
ycf's / ORFs	<i>ORF246</i>	97,0	<i>Zea</i>
	<i>ORF139*</i>	96,0	<i>Zea</i>
	<i>ycf3*</i>	95,0	<i>Zea</i>
	<i>ORF46</i>	97,0	<i>Zea</i>
	<i>ycf2*</i>	68,0	<i>Nicotiana</i>
	<i>ORF99</i>	100,0	<i>Zea</i>
	<i>ORF143*</i>	94,0	<i>Glycine</i>
	<i>ORF97*</i>	85,0	<i>Oenothera</i>
	<i>ORF58</i>	100,0	<i>Zea</i>
	<i>ORF85</i>	97,0	<i>Zea</i>
	<i>ORF70b*</i>	56,0	<i>Nicotiana</i>
	<i>ORF133*</i>	83,0	<i>Zea</i>
	<i>ORF67c*</i>	66,0	<i>Pinus</i>
	<i>ORF42*</i>	87,0	<i>Nicotiana</i>
	<i>ORF173</i>	90,0	<i>Zea</i>
	<i>ycf5</i>	84,0	<i>Zea</i>
	<i>ycf4</i>	100,0	<i>Zea</i>
	<i>ycf9</i>	100,0	<i>Triticum</i>
	<i>ORF70</i>	80,0	<i>Oryza</i>
	<i>ORF50*</i>	96,0	<i>Oenothera</i>
	<i>ORF6*</i>	70,0	<i>Nicotiana</i>
	<i>ORF70</i>	80,0	<i>Oryza</i>
	<i>ORF72*</i>	77,0	<i>Spinacia</i>
	<i>ORF63</i>	98,0	<i>Zea</i>
	<i>ORF103*</i>	71,0	<i>Oenothera</i>
	<i>ORF241*</i>	98,0	<i>Zea</i>
	<i>ORF69*</i>	93,0	<i>Zea</i>
	<i>ORF44</i>	75,0	<i>Hordeum</i>
	<i>ORF128*</i>	97,0	<i>Oryza</i>
	<i>ORF31</i>	95,0	<i>Oryza</i>

ANEXO B - GENES DE RNA TRANSPORTADOR NO PLASTOMA DE CANA-DE-AÇÚCAR.

* %CP=porcentagem de códons utilizados em relação ao total de possíveis códons correspondentes ao mesmo aminoácido.

** análise através de BLAST com homólogos de milho (espécies com e sem introns).

tRNA	Cópias	Códon (s)	Anti-códon (s)	Aminoácido	Nucleotídeos	% CP *
trnP	1	CCU	TGG	prolina	76	25,0
trnW	1	UGG	CCA	triptofano	73	100,0
trnT	3	ACA (1) / ACC (2)	TGT (1) / GGT (2)	treonina	72 / 71	50,0
trnM	1	AUG	CAT	metionina	84	100,0
trnM	4	AUG	CAT	metionina	73	100,0
trnR	3	AGA (1) / CGU (2)	TCT (1) / ACG (2)	arginina	71 / 73	33,3
trnC	1	UGC	GCA	cisteína	70	50,0
trnS	3	UCA (1) / AGC (1) / UCC (1)	TGA (1) / GCT (1) / GGA (1)	serina	87 / 90 / 86	50,0
trnQ	1	CAA	TTG	glutamina	71	50,0
trnH	3	CAC	GTG	histidina	75	50,0
trnV	2	GUC	GAC	valina	71	25,0
trnN	2	AAC	GTT	asparagina	71	50,0
trnL	3	UUG (2) / CUA (1)	CAA (2) / TAG (1)	leucina	80 / 79	33,3
trnG	1	GGC	GCC	glicina	70	25,0
trnE	1	GAA	TTC	glutamato	72	50,0
trnY	1	UAC	GTA	tirosina	83	50,0
trnD	1	GAC	GTC	aspartato	73	50,0
trnA	2	GCG (2 esp.)**	CGC	alanina	71	25,0
trnI	4	AUU (4 esp.)**	GAT	isoleucina	73	33,3
trnK	1	AAG	TTT	lisina	70	50,0
trnF	1	UUC	GAA	fenilalanina	72	50,0

REFERÊNCIAS BIBLIOGRÁFICAS

- AL-JANABI, S.M.; McCLELLAND, M.; PETERSEN, C.; SOBRAL, B.W.S. Phylogenetic analysis of organellar DNA sequences in the *Andropogoneae*: *Saccharinae*. **Theoretical and Applied Genetics**, v. 88, p.933-944, 1994.
- ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic Local alignment search tool. **Journal of Molecular Biology**, v.215, n.3, p.403-410, Oct. 1990.
- ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. **Nature**, v.408, p.796-815, 2000.
- ARAGÃO, F.J.L.; RECH, E.L. Isolamento de vetores para transformação direta. In: BRASILEIRO, A.C.M.; CARNEIRO, V.T.C. (Ed.) **Manual de transformação genética de plantas**. Brasília: EMBRAPA, SPI, 1998. cap. 1, p.26-29.
- BAENA-GONZÁLEZ, E.; GRAY, J.C.; TYYSTJARVI, E.; ARO, E-M.; MAENPAA, P. Abnormal regulation of photosynthetic electron transport in a chloroplast *ycf9* inactivation mutant. **Journal of Biological Chemistry**, v. 276, n. 23, p.20795-20802, June 2001.
- BAUER, J.; HILTBRUNNER, A.; KESSLER, F. Molecular biology of chloroplast biogenesis: gene expression, protein import and intraorganellar sorting. **Cell & Molecular Life Sciences**, v.58, n.3, p.420-433, 2001.
- BOCK, R. Analysis of RNA editing in plastids. **Methods: a Companion to Methods in Enzymology**, v.15, p.75-83, 1998.

- BOCK, R.; HERMANN, M.; KOSSEL, H. *In vivo* dissection of *cis*-acting determinants for plastid RNA editing. **EMBO Journal**, v.15, p.5052-5059, 1996.
- BOCK, R.; KOSSEL, H.; MALIGA, P. Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. **EMBO Journal**, v.13, p.4623-4628, 1994.
- BOCK, R.; ALBERTAZZI, F.; FREYER, R.; FUCHS, S.; RUF, S.; ZELTZ, P.; MAIER, R.M.. Transcript editing in chloroplasts of higher plants. In: **International Colloquium on Endocytobiology and Symbiosis proceedings, Eukaryotism and symbiosis**: Berlin: Springer, 1997. p. 123-137.
- BOOKJANS, G.; STUMMANN, B.M.; HENNINGSEN, K.W. Preparation of chloroplast DNA from pea plastids isolated in a medium of high ionic strength. **Analytical Biochemistry**, v.141, p.244-247, 1984.
- 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 Journal**, v.17, p.868-876, 1998.
- CHAUDHURI, S.; MALIGA, P. Sequences directing C to U editing of the plastid *psbL* mRNA are located within a 22 nucleotide segment spanning the editing site. **EMBO Journal**, v.21, p.5958-5964, 1996.
- CHAUDHURI, S.; CARRER, H.; MALIGA, P. Site-specific factors mediate RNA editing of the *psbL* mRNA in tobacco plastids. **EMBO Journal**, v.12, p.2951-2957, 1995.
- CHEN, X.; SCHNELL, D.J. Protein import into chloroplasts. **Trends in Cell Biology**, v.9, p.222-227, 1999.
- CLAYTON, W.D.; RENVOIZE, S.A. **Genera Graminum**: grasses of the world. London: Her Majesty's Stationary Office, 1986. 897p (Kew Bulletin Additional Series, 13).
- CORNEILLE, S.; LUTZ, K.; MALIGA, P. Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? **Molecular and General Genetics**, v.264, p.419-424, 2000.

- DANIELL, H. GM crops: public perception and scientific solutions. **Trends in Plant Science**, v.4, p.467-469, 1999.
- DANIELL, H.; WIEBE, P.O.; FERNANDEZ-SAN MILLAN, A. Antibiotic-free chloroplast engineering: an environmentally friendly approach. **Trends in Plant Science**, v.6, n.6, p.237-239, June 2001.
- DE COSA, B.; MOAR, W.; LEE, S.B.; MILLER, M.; DANIELL, H. Overexpression of BtCry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. **Nature Biotechnology**, v.19, p.71-74, 2001.
- DENG, X.W.; WING, R.A.; GRUISSEM, M. The chloroplast genome exists in multimeric forms. **Proceedings of the National Academy of Sciences of the USA**, v.86, p.4156-4160, 1989.
- DOYLE, K. Cloning in plasmid vectors. In: DOYLE, K. (Ed.) **Protocols and applications guide**. Madison: Promega, 1996. cap.3, p.41-54.
- DOYLE, K.; DOYLE, J. Plant DNA miniprep. **Focus**, v.12, p.1, 1990.
- DRESCHER, A.; RUF, S.; CALSA Jr., T.; CARRER, H.; BOCK, R. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. **The Plant Journal**, v.22, p.97-104, 2000.
- EWING, B.; HILLIER, L.; WENDL, M.C.; GREEN, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**, v.8, p. 175-185, 1998.
- EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**, v.8, p.186-194, 1998.
- FISHER, N.; STAMPACCHIA, O.; REDDING, K.; ROCHAIX, J.D. Selectable marker recycling in the chloroplast. **Molecular and General Genetics**, v.251, p.373-380, 1996.
- FOJTOVÁ, M.; KOVARIK, A.; MATYÁSEK, R.. Cytosine methylation of plastid genome in higher plants. Fact or artefact? **Plant Science**, v.160, p.585-593, 2001.
- GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: a graphical tool for sequence finishing. **Genome Research**, v.8, p.195-202, 1998.

- HALLICK, R.B.; HONG, L.; DRAGER, R.G.; FAVREAU, M.R.; MONFORT, A.; ORSAT, B.; SPIELMANN, A.; STUTZ, E. Complete sequence of *Euglena gracilis* chloroplast DNA. **Nucleic Acids Research**, v.21, p.3537-3544, 1993.
- HIRATSUKA, J.; SHIMADA, H.; WHITTIER, R.; ISHIBASHI, T.; SAKAMOTO, M.; MORI, M.; KONDO, C.; HONJI, Y.; SUN, C.; MENG, B.; LI, Y.; KANNO, A.; NISHIZAWA, Y.; HIRAI, A.; SHINOZAKI, K.; SUGIURA, M. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. **Molecular and General Genetics**, v.217, p.185-194, 1989.
- IKEO, K.; OGIHARA, Y. Wheat chloroplast genome. JOURNAL Published Only in DataBase (2000). Direct Submission. Mishima: National Institute of Genetics, Center for Information Biology, 2000. /No prelo/
- JACKSON-CONSTAN, D.; KEEGSTRA, K. *Arabidopsis* genes encoding components of the chloroplast protein import apparatus. **Plant Physiology**, v.125, p.1567-1576, Apr. 2001.
- KANEKO, T.; SATO, S.; KOTANI, H. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. **DNA Research**, v.3, p.109-136, 1996.
- KATO, T.; KANEKO, T.; SATO, S.; NAKAMURA, Y.; TABATA, S. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. **DNA Research**, v.7, n.6, p.323-330, 2000.
- KEEGSTRA, K.; CLINE, K. Protein import and routing systems of chloroplasts. **The Plant Cell**, v.11, p.557-570, Apr. 1999.
- KOLODNER, R.; TEWARI, K.K. Chloroplast DNA from higher plants replicates by both the Cairns and rolling circle mechanism. **Nature**, v.236, p.708-711, 1975.

- KOTA, M.; DANIELL, H.; VARMA, S.; GARCZYNSKI, S.F.; GOULD, F.; MOAR, W.J. Overexpression of the *Bacillus thuringiensis* (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. **Proceedings of the National Academy of Sciences of the USA**, v.96, p.1840-1845, 1999.
- LUTZ, K.A.; KNAPP, J.E.; MALIGA, P. Expression of *bar* in the plastid genome confers herbicide resistance. **Plant Physiology**, v.125, p.1585-1590, Apr. 2001.
- MA, C.; TYRAWKA, D.; SMITH, M.A. Isolation and characterization of safflower (*Carthamus tinctorius* L.) chloroplast DNA. **Molecular and Cellular Biochemistry**, v.65, p.181-188, 1984.
- MACKENZIE, S.; McINTOSH, L. Higher plant mitochondria. **The Plant Cell**, v.11, p.571-585, Apr. 1999.
- MAIER, R.M.; NECKERMAN, K.; IGLOI, G.L.; KÖSSEL, H. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. **Journal of Molecular Biology**, v.251, p.614-628, 1995.
- MARIAC, C.; TROUSLOT, P.; POTEAUX, C.; BEZANÇON, G.; RENNO, J.F. Chloroplast DNA extraction from herbaceous and woody plants for direct restriction fragment length polymorphism analysis. **Biotechniques**, v.28, p.110-113, Jan. 2000.
- MARTIN, W.; STOEBE, B.; GOREMYKIN, V.; HANSMANN, S.; HASEGAWA, M.; KOWALLICK, K.V. Gene transfer to the nucleus and the evolution of chloroplasts. **Nature**, v.393, p.162-165, 1998.
- MILLIGAN, B.G. Purification of chloroplast DNA using hexadecyltrimethylammonium bromide. **Plant Molecular Biology Reporter**, v.7, n.2, p.144-149, 1989.
- MOCHIZUKI, N.; BRUSSLAN, J.A.; LARKIN, R.; NAGATANI, A.; CHORY, J. *Arabidopsis* genomes uncoupled 5 (*GUN5*) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. **Proceedings of the National Academy of Sciences of the USA**, v.98, n.4, p.2053-2058, 2001.

- MOURAD, G.; POLACCO, M.L. Mini-preparation of highly purified chloroplast DNA from maize. **Plant Molecular Biology Reporter**, v.7, n.1, p.78-84, 1989.
- MURASHIGE, T.; SKOOG, F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. **Physiologia Plantarum**, v.15, p.473-497, 1962.
- NGERNPRASIRTSIRI, J.; KOBAYASHI, T.; SUGIYAMA, T.; AKAZAWA, T. DNA methylation as a mechanism of transcription regulation in non-photosynthetic plastids in plant cells. **Proceedings of the National Academy of Sciences of the USA**, v.85, p.4750-4754, 1988.
- OHYAMA, K.; FUKUZAWA, H.; KOHCHI, T.; SHIRAI, H.; SANO, T.; SANO, S.; UMESOMO, K.; SHIKI, Y.; TAKEUCHI, M.; CHANG, Z.; AOTA, S-T.; INOKUCHI, H.; OZEKI, H. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. **Nature**, v.322, p.572-574, 1986.
- PALMER, J.D.; STEIN, D.B. Conservation of chloroplast genome structure among vascular plants. **Current Genetics**, v.10, p.823-833, 1986.
- PALMER, J.D. Cell organelles. In: HERRMANN, R.G. (Ed.) **Plant gene research**. Vienna: Springer-Verlag, 1992. p. 99-122.
- PALMER, J.D. Isolation and structural analysis of chloroplast DNA. **Methods in Enzymology**, v.118, p.167-174, 1986.
- PALMER, J.D.; THOMPSON, W.F. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. **Cell**, v.29, p.537-550, 1982.
- PHILLIPS, A.L. Restriction map and clone bank of tomato plastid DNA. **Current Genetics**, v.10, p.147-152, 1985.
- PYKE, K.A. Plastid division and development. **The Plant Cell**, v.11, p.549-556, Apr. 1999.
- RATHINASABAPATHI, B.; MCCUE, K.F.; GAGE, D.A.; HANSON, A.D. Metabolic engineering of glycine betaine synthesis: plant betaine aldehyde dehydrogenases lacking typical transit peptides are targeted to tobacco chloroplasts where they confer aldehyde resistance. **Planta**, v.193, p.155-162, 1994.

- REED, M.L.; PEETERS, N.M.; HANSON, M.R. A single alteration 20 nt 5' to na editing target inhibits chloroplast RNA editing *in vivo*. **Nucleic Acids Research**, v.29, n.7, p.1507-1513, 2001.
- RUF, S.; HERMANN, M.; BERGER, I.J.; CARRER, H.; BOCK R. Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. **Nature Biotechnology**, v.19, n.9, p.870-875, 2001.
- SAKAI, A.; SUZUKI, T.; MIYAZAWA, Y.; KUROIWA, T. Simultaneous isolation of cell-nuclei, plastid-nuclei and mitochondrial-nuclei from cultured tobacco cells; comparative analysis of their transcriptional activities in vitro. **Plant Science**, v.133, p.17-31, 1998.
- SATO, S.; NAKAMURA, Y.; KANEKO, T.; ASAMIZU, E.; TABATA, S. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. **DNA Research**, v.6, n.5, p.283-290, 1999.
- SCHMITZ-LINNEWEBER, C.; MAIER, R.M.; ALCARAZ, J.P.; COTTET, A.; HERRMANN, R.G.; MACHE, R. The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. **Plant Molecular Biology**, v.45, p.307-315, 2001.
- SCHULER, M.A.; ZIELINSKI, R.E. **Methods in plant molecular biology**. USA, Academic Press, 1989. 402p.
- SHINOZAKI, K.; OHME, M.; TANAKA, M.; WAKASUGI, T.; HAYASHIDA, N.; MATSUBAYASHI, T.; ZAITA, N.; CHUNWONGSE, J.; OBOKATA, J.; YAMAGUCHISHINOZAKI, K.; OHTO, C.; TORAZAWA, K.; MENG, B.Y.; SUGITA, M.; DENO, H.; KAMOGASHIRA, T.; YAMADA, K.; KUSUDA, J.; TAKAIWA, F.; KATO, A.; TOHDOH, N.; SHIMADA, H.; SUGIURA, M. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. **EMBO Journal**, v.5, n.9, p.2043-2049, 1986.
- SOBRAL, B.W.S.; BRAGA, D.P.V.; LAHOOD, E.S.; KEIM, P. Phylogenetic analysis of chloroplast restriction enzyme site mutations in the *Saccharinae* Griseb. subtribe of the *Andropogoneae* Dumort. tribe. **Theoretical and Applied Genetics**, v.87, p.843-853, 1994.

- SPRINGER, P.S.; ZIMMER, E.A.; BENNETZEN, J.L. Genomic organization of the ribosomal DNA of sorghum and its close relatives. **Theoretical and Applied Genetics**, v.77, p.844-850, 1989.
- STAUB, J.M.; GARCIA, B.; GRAVES, J.; HAJDUKIEWICZ, P.T.; HUNTER, P.; NEHRA, N.; PARADKAR, V.; SCHLITTLER, M.; CARROL, J.A.; SPATOLA, L.; WARD, D.; YE, G.; RUSSEL, D.A. High-yield production of a human therapeutic protein in tobacco chloroplasts. **Nature Biotechnology**, v.18, p.333-338, 2000.
- SVAB, Z.; HAJDUKIEWICZ, P.; MALIGA, P. Stable transformation of plastids in higher plants. **Proceedings of the National Academy of Sciences of the USA**, v.87, p.8526-8530, 1990.
- SVAB, Z.; MALIGA, P. High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. **Proceedings of the National Academy of Sciences of the USA**, v.90, n.3., p.913-917, 1993.
- TAIZ, L.; ZEIGER, E. **Plant physiology**. 2.ed. Sunderland: Sinauer Associates, 1998. 315p.
- TAYLOR, B.H.; MANHART, J.R.; AMASINO, R.M. Isolation and characterization of plant DNAs. In: THOMPSON, J. ; GLICK, D. **Methods in plant molecular biology and biotechnology**. Quebec: CRC Press, 1993. p.37-44.
- TRIBOUSH, S.O.; DANILENKO, N.G.; DAVYDENKO, O.G. A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower. **Plant Molecular Biology Reporter**, v.16, p.183-189, 1998.
- WAKASUGI, T.; TSUDZUKI, J.; ITO, S.; NAKASHIMA, K.; TSUDZUKI, T.; SUGIURA, M. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. **Proceedings of the National Academy of Sciences of the USA**, v.91, p.9794-9798, 1994.
- 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 genes possibly involved in chloroplast division. **Proceedings of the National Academy of Sciences of the USA**, v.94, p.5967-5972, 1997.

WEIQI, L.; SKINNER, D.Z.; LIANG, G.H. An efficient method of isolating mitochondrial and chloroplast DNA from higher plants. **Current Science**, v.70, n.1, p.78, Jan. 1996.

WHITNEY, S.M.; BALDET, P.; HUDSON, G.S.; ANDREWS, T.J. Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. **The Plant Journal**, v.26, n.5, p.535-547, 2001.

WOLFE, K.H.; SHARP, P.M. Identification of functional open reading frames in chloroplast genomes. **Gene**, v.66, p.215-222, 1988.