

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600



Order Number 9519474

**Structure and expression of ribulose-1,5-bisphosphate
carboxylase/oxygenase small subunit genes in sugarcane**

Tang, Wendong, Ph.D.

University of Hawaii, 1994

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106



STRUCTURE AND EXPRESSION OF
RIBULOSE-1,5-BISPHOSPHATE
CARBOXYLASE/OXYGENASE
SMALL SUBUNIT GENES
IN SUGARCANE

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION
OF THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BOTANICAL SCIENCES (PLANT PHYSIOLOGY)

DECEMBER 1994

By

Wendong Tang

Dissertation Committee:

Samuel S. M. Sun, Chairperson

H. Michael Harrington

Paul Moore

John I. Stiles

Tom Humphreys

ACKNOWLEDGEMENTS

I would like to express my appreciation to my advisor, Dr. Samuel S. M. Sun, for his guidance and support during the development of this dissertation, and to the other members of my committee, Dr. H. Michael Harrington, Dr. Tom Humphreys, Dr. Andy Maretzki, Dr. Paul Moore, and Dr. John Stiles, who have been very generous with their time and assistance.

I want to thank the Hawaiian Sugar Planters' Association (HSPA) for financial support of this thesis. Thanks to Dr. Chifumi Nagai and Dr. Henrik Albert, for their valuable advice and assistance during the past four and half years, and to Dr. Helen Tu, for making of the chimeric gene construct pHT1011. Special thanks to Dr. Piera Sun, for her kind advice in many aspects. Also thanks to all the other members of Dr. Sun's lab and HSPA, and other friends who deserve recognition of their support.

I like to express my appreciation to my parents, Huansheng Tang and Tiande Lei, for their emotional and material support during my graduate study.

I would also like to thank my wife, Yi Huang, for her constant support, understanding, and patience.

ABSTRACT

Sugarcane (*Saccharum* spp. hybrid cultivar H32-8560), a C4 polyploid, is estimated to contain about 16 copies of the ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (*rbcS*) gene. Seven *rbcS* genes were isolated from a sugarcane genomic library, and were partially or completely sequenced. Unlike *rbcS* gene members of other plant species that are heterogeneous in their 3' UTRs, these genes are highly conserved in their coding sequences as well as their 3' and 5' untranslated regions (UTRs). The nested Rapid Amplification of cDNA 3' Ends (nested-RACE) technique was used to clone the 3' UTR of *rbcS* mRNA from sugarcane leaves of different developmental stages. There is no significant difference among the RACE-products from different organs in terms of their overall electrophoretic patterns. The sequences of the 3'-RACE products have more than 95% homology with the 3' UTR of the cloned *rbcS* genes. However, their lengths (from translation stop codon to polyadenylation site) are quite different and can be classified into five major size groups, i.e. Group I, 501bp; II, 351bp; III, 254bp; VI, 184bp; and V, 103bp. Northern analysis of total RNA from mature leaf, S1 mapping, and sequence determination of the *rbcS* cDNA clones isolated from a mature leaf cDNA library confirm the existence of

different lengths of *rbcS* mRNA. The AAUAAA-like motifs for Groups I and II can be identified; however, we fail to detect other sequence motifs that have been suggested to be important in directing correct cleavage and polyadenylation. Differential amplification of sugarcane genomic DNA suggests that the genes encoding Groups I and II mRNAs are similar in length. Groups I and II mRNAs continue to be expressed during development of the sugarcane leaf and the ratio of these two classes of mRNAs in various sugarcane samples is similar. Sugarcane leaf bombardment experiments showed that both the *scrbcS-1* and *scrbcS-3* promoters are able to direct GUS gene expression in mesophyll cells and bundle sheath cells during C3-C4 transition, although at a ratio of 1 to 3 in favor of the bundle sheath cells. These findings support the notion that the *rbcS* genes in C4 plants have a default C3-type expression pattern (Schaffner and Sheen, 1991).

TABLE OF CONTENTS

| | |
|--|------|
| ACKNOWLEDGEMENTS | iii |
| ABSTRACT | iv |
| LIST OF TABLES | viii |
| LIST OF FIGURES | ix |
| LIST OF ABBREVIATIONS | xi |
| CHAPTER I. INTRODUCTION | 1 |
| A. Rubisco | 3 |
| B. C4 Photosynthesis | 5 |
| C. Development of C4 Leaf | 7 |
| D. <i>Rbcs</i> Gene Structure | 9 |
| E. <i>Rbcs</i> Gene Regulation | 12 |
| (1) Light Regulation | 12 |
| (2) Tissue-Specific Regulation | 14 |
| (3) Developmental Regulation | 16 |
| (4) Different Gene Family Members | 21 |
| F. Formation of mRNA 3' End | 28 |
| (1) Transcription Termination | 28 |
| (2) Cleavage and Polyadenylation | 29 |
| G. Gene Regulation by mRNA 3' Untranslated Regions | 31 |
| (1) Messenger RNA Translation | 32 |
| (2) Control of PolyA Tail Length | 32 |
| (3) mRNA Localization | 34 |
| (4) mRNA Degradation | 35 |
| (5) Differential Expression | 37 |
| H. Transient Gene Expression | 38 |
| CHAPTER II. RESEARCH HYPOTHESIS | 42 |
| CHAPTER III. MATERIALS AND METHODS | 44 |
| A. Plant Materials and Growth Conditions | 44 |
| B. Molecular Cloning of <i>Scrbcs</i> Genes | 45 |
| C. Extraction of Genomic DNA from Sugarcane Leaf Sample | 49 |
| D. Copy Number Reconstruction | 49 |
| E. RNA Extraction from Sugarcane Leaves | 50 |
| F. Northern Hybridization | 51 |
| G. Generation of a cDNA Pool by Reverse Transcription | 51 |
| H. Rapid Amplification of cDNA Ends (RACE) | 52 |
| I. Nested-RACE | 54 |
| J. Cloning and Sequencing of the Nested-RACE Products | 55 |

| | | |
|----------------------------------|--|-----|
| K. | cDNA Library Construction and Screening | 55 |
| L. | S1 Mapping | 56 |
| M. | Differential Amplification of Total Sugarcane cDNA | 57 |
| N. | Differential Amplification of Total Sugarcane Genomic DNA | 58 |
| O. | Chimeric Gene Construction | 59 |
| P. | Data Analysis | 63 |
| Q. | Sugarcane Leaf Bombardment | 63 |
| | (1) Coating of Gold Particles with DNA | 64 |
| | (2) Helium Power Setting | 64 |
| | (3) GUS Gene Transient Expression Assay | 65 |
| | (4) Cryostat Sectioning and Examination | 65 |
| R. | PEP Carboxylase Activity | 66 |
| CHAPTER IV. RESULTS | | 68 |
| A. | Sequences of the <i>Scrbcs</i> Genes | 68 |
| B. | Copy Number Reconstruction | 80 |
| C. | Northern Hybridization | 81 |
| D. | RACE Products and cDNA Library Screening | 82 |
| E. | S1 Mapping | 86 |
| F. | Differential Amplification of Total Sugarcane cDNA | 88 |
| G. | Differential Amplification of Total Sugarcane Genomic DNA | 89 |
| H. | Sugarcane Leaf Bombardment | 91 |
| I. | PEP Carboxylase Activity | 97 |
| CHAPTER V. DISCUSSION | | 98 |
| A. | The Highly Conserved <i>Scrbcs</i> Genes and Their mRNAs | 98 |
| B. | Nucleotide Differences among <i>Scrbcs</i> Genes | 99 |
| C. | Length Heterogeneity of the 3' UTRs in <i>Scrbcs</i> mRNAs | 101 |
| D. | Possible Mechanisms of Length Heterogeneity in the 3' UTRs | 105 |
| E. | Multiple Polyadenylation Sites in Developing Sugarcane Leaf | 108 |
| F. | Sugarcane Leaf Bombardment | 111 |
| G. | Evolution of <i>Scrbcs</i> Genes | 116 |
| Chapter VI. CONCLUSION | | 124 |
| REFERENCES | | 126 |

LIST OF TABLES

| | |
|---|-----|
| Table 1. <i>Rbcs</i> Genes from Monocot and Dicot Plants . . . | 10 |
| Table 2. <i>Cis</i> -Elements in Pea <i>Rbcs-3A</i> | 13 |
| Table 3. Oligonucleotide Primer Sequences | 52 |
| Table 4. Number of RACE Products and cDNA Clones within the Different RNA Groups | 86 |
| Table 5. Number of GUS Foci per Shot under 2000psi Helium Pressure | 92 |
| Table 6. Total Number of GUS Foci Located in the Sugarcane Leaf Cells | 95 |
| Table 7. Length of the RACE Products | 103 |
| Table 8. Length of the S1 Mapped mRNAs | 103 |
| Table 9. Summary of the Predicted Length of <i>Scrbcs</i> mRNAs | 104 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1. Chimeric Gene Constructs | 61 |
| Figure 2. Schematic of the 13-1-HE Sequence Assembly | 68 |
| Figure 3. Nucleotide and Deduced Amino Acid Sequences of the <i>Scrbcs-1</i> Gene | 70 |
| Figure 4. Comparison of the Promoter Regions between the <i>Scrbcs-1</i> Gene with the Maize <i>Rbcs</i> Genes | 72 |
| Figure 5. 5' UTR of the Cloned <i>Scrbcs</i> Genes | 75 |
| Figure 6. 3' UTR of the Cloned <i>Scrbcs</i> Genes | 76 |
| Figure 7. Comparison of the Nucleotide Sequences of <i>Scrbcs-1</i> and <i>Scrbcs-3</i> Gene | 78 |
| Figure 8. Copy Number Reconstruction Experiment | 81 |
| Figure 9. Mature Leaf Total RNA Hybridized with <i>Scrbcs-1</i> Gene | 82 |
| Figure 10. Nested RACE Products of RNAs from Sugarcane Leaves at Different Developmental Stages | 83 |
| Figure 11. Example Sequences of 3' UTR of the RACE Clones as Compared with the <i>Scrbcs-1</i> Gene | 84 |
| Figure 12. Schematic of S1 Mapping | 87 |
| Figure 13. S1 Mapping of the 3' UTR of the <i>Scrbcs-1</i> Gene Transcripts | 87 |
| Figure 14. Schematic of Differential Amplification of Total Sugarcane cDNA | 88 |
| Figure 15. Differential PCR Amplification Products of cDNAs from Different Sugarcane Samples | 89 |
| Figure 16. Schematic of Differential Amplification of Total Sugarcane Genomic DNA | 90 |
| Figure 17. Differential PCR Amplification Products Using Genomic DNA and <i>Scrbcs</i> Gene as Templates. | 91 |
| Figure 18. GUS-Stained Sugarcane Leaf after Bombardment with pUbiGUS | 93 |

| | |
|---|-----|
| Figure 19. GUS-Stained Sugarcane Leaf after Bombardment with pWD1 | 93 |
| Figure 20. GUS-Stained Sugarcane Leaf after Bombardment with pWD2 | 94 |
| Figure 21. Transient Expression of GUS Gene in Sugarcane Leaf BS Cells after Bombardment with pWD1 | 95 |
| Figure 22. Transient Expression of GUS Gene in Sugarcane Leaf MS Cells after Bombardment with pWD1 | 96 |
| Figure 23. Transient Expression of GUS Gene in Sugarcane Leaf Epidermis after Bombardment with pWD1 | 96 |
| Figure 24. PEP Carboxylase Activity of Sugarcane Leaf of Different Developmental Stages | 97 |
| Figure 25. Genealogy of <i>Saccharum</i> spp. hybrid cultivar H32-8560. | 119 |

LIST OF ABBREVIATIONS

| | |
|-----------------|---|
| aa | amino acid |
| Ac | acetic |
| ATP | adenosine triphosphate |
| bp | base pair |
| BS | bundle sheath cell |
| BSA | bovine serum albumin |
| CaMV | cauliflower mosaic virus |
| cDNA | complementary DNA |
| CO ₂ | carbon dioxide |
| CPM | counts per minute |
| CTAB | cetyltrimethylammonium bromide |
| cv. | cultivar |
| DEPC | diethylpyrocarbonate |
| dCTP | deoxyl cytidine triphosphate |
| dicot | dicotyledon |
| DMF | N,N-dimethyl formamide |
| DMSO | dimethylsulfoxide |
| dNTP | deoxyribonucleotide triphosphate |
| DTT | dithiothreitol |
| DTE | dithioerythritol |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| EDTA | ethylenedinitrilo-tetraacetic acid |
| EP | epidermic cell |
| ER | endoplasmic reticulum |
| EtoH | ethanol |
| Fig | Figure |
| GC | green callus |
| GL | greening leaf |
| GUS | β -glucuronidase |
| hnRNA | heterogeneous nuclear RNA |
| kb | kilobase |
| kD | kilodalton |
| LB | Luria-Bertani (medium) |
| LHC | light-harvesting chlorophyll-protein complex |
| β -Me | β -mercaptoethanol |
| M | molar |
| ML | mature leaf |
| M-MLV | Moloney murine leukemia virus |
| monocot | monocotyledon |
| mRNA | messenger RNA |
| MS | mesophyll cell or Murashige and Skoog (medium) |
| M.W. | molecular weight |
| NADP-MdH | NADP-malate dehydrogenase |
| NOS | nopaline synthase |
| NPTII | neomycin phosphotransferase II |

| | |
|---------------------|---|
| NTP | ribonucleotide triphosphate |
| O.D. ₆₆₀ | optical density at 660 nm |
| OAA | oxaloacetic acid |
| ORF | open reading frame |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |
| PEG | polyethylene glycol |
| PEP | phosphoenolpyruvate |
| PEPC | phosphoenolpyruvate carboxylase |
| polyA | poly-adenylate |
| PPdK | pyruvate-orthophosphate dikinase |
| PSII | photosystem II |
| RACE | rapid amplification of cDNA ends |
| RB | right border of T-DNA |
| <i>rbcl</i> | Rubisco large subunit |
| <i>rbcs</i> | Rubisco small subunit |
| RER | rough endoplasmic reticulum |
| RFLP | restriction fragment length polymorphism |
| RT | room temperature or reverse transcription |
| RNase | ribonuclease |
| Rubisco | ribulose-1,5-bisphosphate carboxylase/oxygenase |
| RuBp | ribulose-1,5-bisphosphate |
| <i>scrbc</i> | sugarcane <i>rbcs</i> |
| SDS | sodium dodecyl sulfate |
| SH | leaf sheath |
| snRNA | small nuclear RNA |
| T-DNA | transfer DNA |
| TB | Terrific Broth (medium) |
| TEMED | N,N,N',N'-tetramethylethylenediamine |
| Ti plasmid | Tumor-inducing plasmid |
| Tris | Tris[hydroxymethyl]aminomethane |
| U | unit |
| UEL | unexpanded leaf (leaf No. -3) |
| UTR | untranslated region |
| X-gluc | 5-bromo-4-chloro-3-indolyl- β -D-glucuronic acid |
| YL | young leaf |

CHAPTER I. INTRODUCTION

Multicellular eukaryotes are highly organized cell societies. Although derived from a single fertilized egg, eukaryotic cells may differ in structure as well as function. Studying cell differentiation during development, especially at the molecular level, has become one of the most important approaches for understanding the nature of life. Modern molecular biology has revealed that cell differentiation is controlled largely by genetic information stored in the form of DNA sequences. Since almost all cells in an organism contain the same set of genes, differential expression of these genes plays a controlling role in the cell differentiation process.

Several developmental processes have been used as models to study the molecular basis of development in plants. These processes include flower development, storage protein deposition during seed maturation, fruit ripening, and degradation of stored foods during early seed germination (Coen and Meyerowitz, 1991; Boothe and Walden, 1990; Longhurst *et al.*, 1990; Goldberg *et al.*, 1989). Molecular studies on flowering have greatly enhanced our understanding of plant reproductive process, while in photosynthetic studies, the developing C4 plant leaf appears to satisfy the

requirements for the ideal model system in that it is, "an organ with few cell types, with defined patterns of cell origin, and with molecular markers for the differentiation of the individual cell types" (Nelson and Langdale, 1989). Although much has been done towards establishing an understanding of the photosynthetic pattern of mature leaf, leaf development and cell differentiation in the C4 plant leaf (mainly in maize), mechanisms of cell differentiation and differential gene expression are still not clear.

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco, E.C.4.1.1.39) has attracted much interest in plant molecular studies not only because of its abundance in plants, importance in photosynthesis, light-responder characteristics, and coordination between nuclear and chloroplast subunits, but also because of its distinct gene expression pattern in C4 plants.

Sugarcane was chosen as our experimental material for several important reasons. First, it is one of the most important economic crops in Hawaii and the world. Sugarcane improvement through artificial hybridization has been carried out for one century. However, improvement of sugarcane yields has slowed in recent years possibly due to a narrow germplasm base available for conventional hybridization. Application of biotechnology provides new

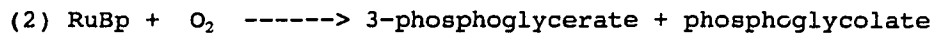
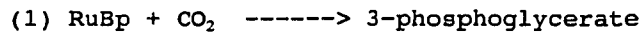
opportunities for sugarcane improvement, however, to realize this potential, it is necessary to understand the physiology and molecular biology of agronomically important processes. Second, sugarcane is perhaps the most efficient converter of solar energy to stored energy (Hunsigi, 1993a). The C₄ photosynthetic feature of sugarcane was first described by Kortschak *et al.* (1965). Molecular studies of the C₄ photosynthesis of sugarcane could eventually lead to improvement of photosynthetic efficiency of other plant species. Third, cell biology studies on sugarcane have been carried out for many years and standard tissue culture methods are well established (for a review, see Maretzki, 1987). A few genes involved in key photosynthetic steps have been cloned (Albert *et al.*, 1992; Tang and Sun, 1993). To increase our knowledge of the molecular physiology of C₄ plants and to generate information for potential sugarcane improvement, it is important to characterize the regulatory patterns of the genes encoding the sugarcane Rubisco enzyme during cell differentiation associated with leaf development and growth.

A. Rubisco

Rubisco is the most abundant protein in the world. It represents about one-half of the total soluble protein in a

photosynthetic leaf, and is responsible for the initial carbon-fixation reaction in the Calvin-Benson pathway of photosynthesis (Calvin and Benson, 1949). The Rubisco enzyme consists of eight 15-kD small subunits and eight 52-kD large subunits. The former, with regulatory function, are encoded in nuclear DNA while the latter, with catalytic activity, are encoded in chloroplast DNA (reviewed by Raines *et al.*, 1991). Under certain conditions (*e.g.* light induction), the nuclear Rubisco small subunit (*rbcs*) genes are transcribed and the resultant mRNAs are translated to make *rbcs* proteins in the cytosol. With the assistance of molecular chaperons, the *rbcs* proteins are transported into the chloroplast stroma, where they combine with the Rubisco large subunit (*rbcl*) proteins to form the whole enzyme (Hemmingsen *et al.*, 1988).

The Rubisco enzyme catalyzes two different types of reactions in the chloroplast in the presence of ribulose-1,5-bisphosphate (RuBp):



The first reaction catalyzes carbon dioxide fixation in photosynthesis and the second reaction (so-called photorespiration) breaks down the substrate of

photosynthesis, *i.e.* RuBp, in the presence of oxygen, thus lowering the rate of photosynthesis (Darnell *et al.*, 1990).

In C3 plants, the Rubisco enzyme is present in mesophyll chloroplasts where oxygen is produced through the photo-electron transport chain (PSI and PSII). The catalytic sites of the Rubisco enzyme are competed for by carbon dioxide and oxygen, lowering the efficiency of photosynthesis.

B. C4 Photosynthesis

Some tropical plants, such as maize and sugarcane, possess special characteristics, including higher efficiency in water use, lower CO₂ compensation points and less discrimination than other plants against ¹³C (Black, 1973; Darnell *et al.*, 1990). These plants are called C4 plants since the primary fixation product of CO₂ is a 4-carbon acid, oxaloacetic acid (OAA). Further studies revealed that, with a few exceptions (Shomer-Ilan *et al.*, 1975; Kennedy, 1976), most of the C4 plants have an obvious "Kranz" anatomical pattern in their mature leaves, in which the bundle sheath cells (BS, surrounding each leaf vein) are surrounded by the mesophyll cells (MS).

MS chloroplasts in C4 leaves contain more PSII in the thylakoids and carry out most of the light-reaction which produces oxygen. BS chloroplasts are found to be "agranaal". They contain significantly fewer thylakoids, with limited PSII activity, and produce almost no oxygen (Laetsch and Price, 1969; Darnell *et al.*, 1990). The Rubisco enzyme is present only in BS and not in MS, while phosphoenolpyruvate carboxylase (PEPC) is found in MS but not in BS (Darnell *et al.*, 1990). Carbon dioxide is first fixed by PEPC in the MS cytosol after entering the leaf through stomata. Then, through transport shuttles, *e.g.* aspartate or malate, depending on the plant species, the CO₂ is concentrated in the BS chloroplasts where it enters the Calvin Cycle for carbon reduction.

Sugarcane has been characterized as highly efficient in photosynthesis (Moore, 1989). Sugarcane (*Saccharum spp. hybrids*) has a higher water use efficiency than other C4 plants and C3 plants. Moore (1989) calculated the efficiency of conservation of light energy for various crops and found that Hawaiian sugarcane has the most efficient use of sunlight. Among different varieties of sugarcane, leaf shape, absorption spectrum, light intensity, and photosynthate mobility were suggested to be important in maximizing the photosynthetic efficiency (Bull, 1969; Irvine, 1975; Alexander, 1990).

C. Development of C4 Leaf

C4 photosynthesis requires Kranz anatomy (Bouton *et al.*, 1986; Brown and Hattersley, 1989). Recent studies suggest that such a special leaf feature is developed from C3 characteristics (Crespo *et al.*, 1979; Bassi and Passera, 1982; Langdale *et al.*, 1987, 1988a and 1988b; Nelson and Langdale, 1989; Wang *et al.*, 1993). At early developmental stages, the whole leaf of a C4 plant exhibits a C3 photosynthetic pattern, e.g. the Rubisco mRNA and protein are found in all photosynthetic cells. Likewise, in early developmental stages of cultured C4 plant tissues, the whole culture consists of unorganized cells and possesses only a C3 pathway, e.g. APO-P cell line of *Amaranthus powellii* (Goldstein and Widholm, 1990) and photomixotrophic cell suspension culture of *Gisekia pharnaceoides* L. (Seeni and Gnanam, 1983). Kortschak and Nickell (1970) established permanent callus cultures from sugarcane stalk parenchyma. When kept under light, these cultures turned green rapidly at the sites of shoot formation. The first stable photosynthetic product found in such "green callus" was glycerate-3-phosphate. They concluded that C3-, not C4-, type of photosynthetic pathway was pursued in the tissue culture.

However, there are reports that C4 photosynthesis may operate in tissue culture. In 1972, Laetsch and Kortschak described C4 photosynthetic tissue cultures of *Froelichia gracilis* (Hook.) Moq., a C4 plant. In these cultures, large parenchyma cells containing chloroplasts make up most of the tissue with islands of meristematic cells and small clusters of lignified cells randomly scattered throughout the callus, indicating considerable differentiation of cell types. However, there was no obvious differentiation of the MS into Kranz arrangements.

Kennedy (1976) reported the photorespiration rate of tissue cultures of *Portulaca oleracea*, a C4 plant, was one-half to one-third that of C3 plant tissue cultures. However, the tissues he used were "similar to mature *P. oleracea* leaves on the basis of cell ultrastructure and ¹⁴CO₂ labelling patterns". Since he did not present the cross section anatomy of the tissues used, it is not clear whether these tissue cultures had differentiated, or whether C4 photosynthesis is not related to Kranz anatomy in the tissue cultures of *P. oleracea*, as he had suggested.

During the development of C4 plants, including those generated from calli, C4 photosynthesis coincides with the development of Kranz anatomy (the differentiation of MS and BS). The expression of genes encoding enzymes involved in

C4 photosynthesis increases dramatically, and the Rubisco mRNA and protein become localized specifically to the BS cells.

D. *Rbcs* Gene Structure

The *rbcs* genes have been cloned from many plants, including both dicotyledonous and monocotyledonous species (reviewed by Dean *et al.*, 1989). Table 1 lists the numbers of *rbcs* genes and cDNAs of monocot and dicot plants that have been deposited into the databanks (search on April 30, 1994).

Table 1. *Rbcs* Genes from Monocot and Dicot Plants.

(Database: GenBank+EMBL)

| PLANT SOURCE | NUMBER OF CDNA SEQUENCES* | NUMBER OF GENE SEQUENCES* |
|---|---------------------------|---------------------------|
| <u>dicotyledonous plants</u> | | |
| <i>Arabidopsis thaliana</i> | 12 | 3 |
| <i>Brassica napus</i> | 1 | 3 |
| <i>Cucumis sativus</i> (Cucumber) | 1 | |
| <i>Glycine max</i> (Soybean) | | 2 |
| <i>Gossypium hirsutum</i> | | 1 |
| <i>Hevea brasiliensis</i> | 1 | |
| <i>Lycopersicon esculentum</i> (Tomato) | 3 | 7 |
| <i>Malus</i> (unidentified species) | 1 | |
| <i>Manihot esculenta</i> | 2 | |
| <i>Mesembryanthemum crystallinum</i> | 3 | 7 |
| <i>Nicotiana glauca</i> | | 1 |
| <i>Nicotiana glauca</i> | 2 | 2 |
| <i>Nicotiana glauca</i> (Tobacco) | | 3** |
| <i>Petunia hybrida</i> (Mitchell) | 1 | 3 |
| <i>Phaseolus vulgaris</i> (French bean) | 3 | |
| <i>Pisum sativum</i> (Pea) | 4 | 3 |
| <i>Pyrus malus</i> | 1 | |
| <i>Pyrus pyrifolia</i> (Japanese pear) | 1 | |
| <i>Raphanus sativus</i> (Radish) | 1 | |
| <i>Silene pratensis</i> (White campion) | 1 | |
| <i>Sinapis alba</i> (Mustard) | 1 | |
| <i>Solanum tuberosum</i> (Potato) | 1 | 5 |
| <i>Trifolium repens</i> (White clover) | | 1 |
| <u>monocotyledonous plants</u> | | |
| <i>Lemna gibba</i> | 2 | 5 |
| <i>Oryza sativum</i> (Rice) | 3 | 1 |
| <i>Saccharum hybrid</i> (Sugarcane) | | 1 |
| <i>Triticum aestivum</i> (Wheat) | 2 | 2 |
| <i>Zea mays</i> (Maize) | 1 | 3 |

* including partial sequences

** including one pseudogene

The *rbcS* genes are highly conserved. Although the sequence and length of the encoded transit peptides may be different among various species, e.g. 47 amino acids in maize and 57 in tobacco, all of the mature *rbcS* proteins contain the hexadecapeptide **YYDGRYWTMWKLPMPFG**, which might be important in the structure and/or function of the small subunits (Dean *et al.*, 1989). Generally, the *rbcS* genes from dicots contain two introns at the same positions except for one of the *rbcS* genes in petunia, tomato, potato, and tobacco, of the *Solanaceae* that have an additional intron downstream of the second (Wolter *et al.*, 1988). The *rbcS* genes in monocots usually contain only one intron, at either position of the two dicot *rbcS* introns.

The *rbcS* genes exist in plant genomes as a gene family of several members. Genes within a species show little divergence of their protein-coding sequences, in contrast to the relatively marked difference among even closely related species. This high degree of sequence homology may be the result of recent gene conversions rather than gene duplications (Dean *et al.*, 1989).

E. *Rbcs* Gene Regulation

Considerable progress has been made toward understanding the regulation of *rbcs* genes. Most of the efforts, however, were based on previous biochemical and physiological studies on the Rubisco enzyme.

(1) Light Regulation

Light has proven to be an important regulatory factor in photosynthesis (Reviewed by Buchanan, 1980). Rubisco activity was found to be greatly enhanced by light during development (Perchorowicz *et al.*, 1981; Berry *et al.*, 1990). This is, at least partially, attributed to the increased transcription of Rubisco genes (Thompson and Meagher, 1990). Increase in *rbcs* mRNA level in different plant species occurs in response to light in etiolated seedlings and dark-adapted plants. Such an increase is mediated by the photoreceptor phytochrome (Gallagher and Ellis, 1982; Silverthorne and Tobin, 1984; Berry-Lowe and Meagher, 1985; Mosinger *et al.*, 1985).

Promoter deletion analysis of *rbcs* genes has been carried out in both dicot and monocot plants using transient gene expression systems (Kuhlemeier *et al.*, 1987; Stephen and Tobin, 1991), and *in vitro* protein binding assays, e.g. gel

retardation and foot-printing. Table 2 shows several *cis*-elements that are responsible for phytochrome-mediated light regulation and their corresponding binding proteins, using pea *rbcS-3A* as a model (reviewed by Gilmartin *et al.*, 1990). However, the role played by each of these individual elements is still unclear, especially in different organs at different developmental stages. Different elements may possess overlapping roles, and combinations of different elements may confer new roles.

Table 2. *Cis*-Elements in Pea *RbcS-3A*. Numbers denote positions of nucleotide sequences relative to the transcription start site.

| <i>Cis</i> -ELEMENT | BINDING SITE | FACTOR |
|---------------------|-----------------|--------|
| BOXII: | -151 to -138 | GT-1 |
| BOXIII: | -125 to -114 | GT-1 |
| BOXII*: | -224 to -213 | GT-1 |
| BOXIII*: | -257 to -248 | GT-1 |
| BOXII**: | -386 to -372 | GT-1 |
| BOXIII**: | -360 to -347 | GT-1 |
| G-BOX: | -211 to -201(?) | GBF |
| G-BOX: | -247 to -237(?) | GBF |
| 3AF1 SITE: | -51 to -31 | 3AF1 |
| AT-1 SITE: | -288 to -27(?) | AT-1 |
| 3AF1 SITE: | -184 to -176 | GAF-1 |
| 3AF1 SITE: | -236 to -228 | GAF-1 |

Schindler and Cashmore (1990) suggested that photoregulated gene expression may involve ubiquitous DNA binding proteins. This suggestion was based on the discovery that several proteins from *Nicotiana plumbaginifolia* nuclear extracts can interact *in vitro* with promoter regions of different light-regulated genes (e.g.

rbcs, *cab-E*) of different plants (e.g. *Arabidopsis*). Since little homology is found between promoter regions of monocot and dicot *rbcs* genes, it is unlikely that the light-response signal transduction in monocot and dicot plants are very similar, even if the same mechanism has evolved in both kinds of plants.

(2) Tissue-Specific Regulation

As discussed above, C4 plants develop two distinct cell types in their photosynthetic leaves, i.e. BS and MS of Kranz anatomy. Rubisco activity occurs only in BS. A number of experiments, including *in situ* hybridization, northern hybridization and *in vitro* translation of BS and MS mRNAs (using *rbc* genes or Rubisco antibodies as probes), indicate that such a cell-type specific distribution pattern of Rubisco is due to differential expression of Rubisco genes (Sheen and Bogorad, 1987a & 1987b; Nelson and Langdale, 1989).

Although *rbcs* cDNAs and genes from C4 plants (e.g. maize) have been cloned (Lebun et al., 1987), little is known about the mechanism of their tissue-specific, differential expression. Ngerprasisiri et al. (1989) used maize PEPC, *rbcs* and *rbcl* cDNAs as probes to maize (Golden Cross Bantam T-51) BS and MS total DNAs, following digestion with

isoschizomeric pairs of methyl-sensitive and -insensitive restriction enzymes. The experiment was to study the expression of the gene in relation to the DNA methylation state, *i.e.* the methylation of cytosine residues at CpG or CpXpG sites, and the methylation of adenine residues, of PEPC, *rbcs* and *rbcl* genes in BS and MS. These studies reported an inverse correlation between the relative abundance of specific transcripts in a given cell type during greening and the methylation status of the corresponding nuclear or chloroplast gene, that is, more transcription from less methylated genes. These results suggest that DNA methylation state might be one component of the multilevel control of cell type differential expression of maize PEPC gene.

Langdale *et al.* (1991) used the 5' upstream regions of the *rbcs* and PEPC genes as probes to carry out a similar study in another maize line, Pioneer B73. They found that the methylation state of a specific site, more than three kilobase 5'-upstream of the PEPC gene, was correlated with the light-induced and tissue-specific accumulation of the maize PEPC mRNA. However, no correlation was found between methylation state and gene expression for either the *rbcs* or *rbcl* genes. The authors suggested that such contradictory results might be due to the different lines of maize and the different probes used. Thus the role of selective DNA

methylation in relation to the tissue-specific expression of these genes is currently uncertain.

In vitro DNA binding experiments (gel retardation assay) of Yanagisawa and Izui (1990) revealed several tissue-specific nuclear proteins that interact with the promoter of a C4-type PEPC gene of maize. They suggested that the tissue-specific expression of the PEPC gene may be controlled by the combined effects of these nuclear factors, that interact with *cis*-elements in the PEPC promoter *in vivo*. Further studies, such as DNA foot printing and promoter deletion assays, are needed to confirm the role of *trans*-acting factors in the tissue-specific expression of C4 genes.

(3) Developmental Regulation

The evolution of C4 photosynthesis has been studied both ontogenically and phylogenetically. It is generally believed that C4 photosynthesis is polyphyletic in evolutionary origin, having arisen independently several times within monocots and dicots (Cameron *et al.*, 1989). There are several types of C4 photosynthesis, *e.g.* NADP-ME, NAD-ME and PEPCK. Nevertheless, most C4 plants share the Kranz anatomy feature. Moreover, cross hybridization among C4 type and C3 type species within a family, have shown that

the C4 characteristics can be genetically transmitted among species (Holaday *et al.*, 1985; Cameron *et al.*, 1989).

Crespo *et al.* (1979) found that both C3 and C4 photosynthetic characteristics occur in a single maize plant. In three-week old plants, the first leaf appears to be more C3-like than the fourth and fifth leaves in terms of photorespiration, Rubisco and PEPC activity, net photosynthetic rate, and cell ultrastructure. Although all leaves possessed Kranz anatomy, in the first leaf, the BS chloroplasts were not centrifugally arranged and the thylakoid lamellae were more densely concentrated. Leaf position and leaf age were suggested to be factors in controlling the appearance of C4 photosynthetic characteristics, but the mechanisms by which this is accomplished is unknown.

Recently, some of the molecular processes involved in the development of C4 characteristics within a single leaf have been elucidated. Using *in situ* immunolocalization techniques, Langdale *et al.* (1987) found that within a normal developing maize leaf, the C4-specific proteins appear concurrently with the development of Kranz anatomy. The C4 proteins appear first in cells surrounding the median vein, then in cells surrounding other major veins, and finally in cells surrounding minor veins. However, in the

leaf of the temperature sensitive greening mutant *argentina* (*ar*), the appearance of the C4-specific proteins and the development of BS are delayed. In a normal (wild type) maize plant leaf, the C4 mRNAs accumulate longitudinally along the veins, laterally across the leaf and locally around individual veins, and the BS-specific mRNAs accumulate around developing veins before Kranz anatomy is morphologically evident (Langdale et al., 1988a & 1988b). In *ar* mutant maize leaf, accumulation of the C4-specific protein mRNAs is delayed relative to the appearance of Kranz anatomy. These researchers concluded that BS and MS develop in clusters and the development of these clusters is temporally and spatially regulated.

BS and MS do not share the same cell lineage in C4 plant leaves. Dengler et al. (1985) examined several C4 and C3 grass species of *Poaceae*, which are closely related to maize and sugarcane. They found that vascular tissue and BS arise from the meristematic centers in the primordium (procambium), whereas MS arises from the ground meristem at the leaf base.

The mechanism of the "clusterous" development of MS and BS is still not well characterized. Nelson and Langdale (1989) suggested that, like the differentiation of other cell types, the differentiation of BS and MS depends on the

continuous interpretation of positional information. Various leaf types of maize, including the foliar leaf, husk leaf, coleoptile, and glume, which exhibit Kranz anatomy with altered vein spacing patterns (different numbers of MS between two adjacent veins), were studied with *in situ* hybridization technique (Langdale *et al.*, 1988a & 1988b). C4 development occurs adjacent to veins in all leaf types under equivalent illumination, but, beyond a several-cell radius from the vein, the MS develop as conventional C3 chlorenchyma. The genes for Rubisco and the light-harvesting chlorophyll-protein complex (LHC) are expressed in the distal MS cells as in normal C3 plant cells; on the other hand, the genes for PEPC, pyruvate-orthophosphate dikinase (PPdK) and NADP-malate dehydrogenase (NADP-MdH) are expressed in the near-proximal MS cells. It seems that positional control of MS development acts locally within a small radius of each vein and that C3-type photosynthetic development is the default scheme. Physiological measurements of oxygen inhibition of photosynthesis indicated that such gene expression patterns reflected the pathways of carbon dioxide fixation in these cells.

Light is also found to have an essential role in the development of C4 characteristics. Bassi and Passera (1982) discovered that maize plants grown under a high level of

illumination are more C4-like; while those grown under a low level of illumination are more C3-like.

In dark grown plants, accumulation of C4 enzymes and their mRNAs increased greatly in the appropriate cells following illumination (Sheen and Bogorad, 1987a & 1987b; Nelson *et al.*, 1984). Although low levels of Rubisco were detected in both BS and MS of dark-grown maize seedlings, Rubisco accumulated only in BS and decreased only in MS upon illumination (Sheen and Bogorad, 1986).

Maize husk leaves have widely spaced veins, *e.g.* twenty photosynthetic cells between adjacent veins. *In situ* hybridization and immunolocalization studies of C4 enzymes (including Rubisco, ME, PPdK, PEPC and LHCP-II) by Langdale *et al.* (1988a & 1988b) on these leaves showed that, under low light, Rubisco accumulates only in MS while other C4-specific enzymes are absent (like C3 pattern). However, under high light, all the C4-specific enzymes accumulate, in greater levels, principally in cells close to the veins and the Kranz pattern is developed.

Although dark-suppression of Rubisco in the C4 dicot *Amaranthus* occurs at the translational level (Berry *et al.*, 1990), light-regulation of C4-specific enzymes usually occurs at the transcriptional level. Nelson and Langdale

(1989) suggested that light may induce a diffusible substance, probably via phytochrome or one of the blue light receptors. This substance is then transported to MS and BS through the vascular system and interacts with cell-specific factors to regulate the expression of the C4-specific enzymes.

(4) Different Gene Family Members

In C4 plants, enzymes that are involved in C4 photosynthesis, *e.g.* PEPC, PPdK, and Rubisco, are encoded by multigene families. Thus, one might expect that each member of the family should have a specific function in order to withstand the pressure of natural selection.

Many enzymes, *e.g.* PEPC, PPdK, and NADP-malic enzyme, have two different types: the C4-type and the C3-type (Ting and Osmond, 1973; Goatley and Smith, 1974; Grula and Hudspeth, 1987; Hudspeth and Grula, 1989; Albert, 1991). The C4-type enzymes exist specifically in the C4 photosynthetic tissues, while the C3-type, serving a "house-keeping" function, is expressed in both photosynthetic and non-photosynthetic tissues. These two different types of enzymes have been shown to be encoded by different genes in C4 plants, and some of these genes have been isolated from

various C4 plant species (Harpster and Tylor, 1986; Thomas *et al.*, 1990; Sheen, 1991; Albert *et al.*, 1992).

Two predominant PEPC forms have been discovered in C4 *Sorghum*, maize and sugarcane. The C3-"housekeeping" form is present in roots and stems as well as green leaves while the C4-form is expressed exclusively in the MS of C4 photosynthetic leaves. Genes and cDNAs corresponding to these two isoforms have been cloned (Harpster and Taylor, 1986; Thomas *et al.*, 1990; Albert *et al.*, 1992). Sequence analysis showed that the C4 PEPC genes are quite different from the C3-type genes in the promoter region. All C4-type genes characterized thus far share several short upstream sequences, including a TATA-like sequence (TATTTGAA, approximately 30bp upstream of the transcription start site), a 9-bp CCTTATCAT sequence, several copies of direct repeats (CCCTcTCCACATCC and AACAGCACcGAGCCAAGCCAAAAGGAGC), and several copies of CCGCC/G sequence, that are hypothesized to be specific for phytochrome-mediated-regulation and C4-compartmentation (Grob and Stuber, 1987; Matsuoka and Minami, 1989; Hudspeth and Grula, 1989; Yanagisawa and Izui, 1990). From green leaf nuclear extracts, a *trans*-acting factor, PEP-I, binds *in vitro* with one of the direct repeats, CCCTcTCCACATCC (Kano-Murakami *et al.*, 1991).

Recently, Sheen (1991) reported that there are two members of the maize PPdK gene family. One gene, *cyppdkzm2*, encodes the cytosolic PPdK isozyme which serves a housekeeping function in root, stem and leaf. The other gene, *cyppdkzm1*, encodes both chloroplast PPdK (C4-specific isozyme) and cytosolic PPdK. The gene *cyppdkzm1* has two different upstream promoters: the C4 PPdK promoter is located at about 5kb upstream to the cytosolic PPdK promoter. Both genes share the same region from the second exon to the end of the gene. Thus the whole cytosolic PPdK promoter and its first exon occur within the first intron of the C4 PPdK. Sequence analysis and transient gene expression experiments showed that the promoters of the cytosolic *cyppdkzm1* and the *cyppdkzm2* contain motifs of known housekeeping genes. The promoter of the C4 form *cyppdkzm1* also has the C4-specific motifs.

The *rbcS* protein is found to be encoded by a family of nuclear genes in all plant species studied so far (Manzara and Gruissem, 1988; Dean *et al.*, 1989; Silverthorne *et al.*, 1990). The copy numbers of *rbcS* genes in different genomes range from 4 (*Arabidopsis*) to more than 12 (wheat). Different members of the *rbcS* gene family are differentially light- or development-regulated in C3 plant species (Coruzzi *et al.*, 1984; Dean *et al.*, 1987; Silverthorne *et al.*, 1990). All eight members of the *rbcS* gene family in *Petunia*

(Mitchell), a C3 plant, are expressed, albeit to very different levels, in leaves (Dean *et al.*, 1985). The relative expression of the genes, as assayed using gene-specific probes, was measured as: SSU301, 47.3%; SSU611, 23.2%; SSU491, 7%; SSU112, 5.4%; SSU911, 0.5%; SSU231 and SSU511, 15.2%; and SSU211, 1.9%. Each member has its own expression characteristic; however, no significant differential expression was detected among them in different organs including leaf, sepal, petal, stem, root and stigma/anther. SSU301 shows the highest level of expression in each of the organs, while SSU911 always shows the weakest expression (Dean *et al.*, 1987).

Lemna gibba, a C3 monocot plant, contains approximately twelve members of the *rbcS* family. Silverthorne *et al.* (1990) isolated six of them (SSU5A, SSU5B, SSU13, SSU25, SSU40A, SSU40B). They used the 3'-untranslated regions of the six genes and a cDNA clone (pLgSSU1), which represented the seventh gene, as specific probes to detect their expression. Several genes were expressed at different levels, with SSU1 showing the highest expression and SSU40B the lowest. However, all the genes tested show the same phytochrome-controlled responses to dark, red light, and red/far-red light changes.

Coruzzi *et al.* (1984) examined the expression of a pea *rbcs* gene, pPS-2.4, in various organs. They found that, compared with green leaves, the level of total *rbcs* mRNA was reduced to about 50% in pericarps, 8% in petals and seeds, and 1-3% in etiolated leaves, stems, and roots. The pPS-2.4 mRNA accounted for approximately 30-35% of total *rbcs* transcripts in green leaves, 5-10% in pericarps, 15-20% in seeds, and below detection in petals and etiolated leaves. The authors concluded that the pPS-2.4 gene is expressed in an organ-specific and light-regulated fashion and that the individual *rbcs* genes are under differential transcriptional control.

Buzby *et al.* (1990) characterized a light-regulated nuclear factor, LRF-1, that interacts with a specific sequence about 150bp upstream of the transcription start site of SSU5B gene in *Lemna gibba*. A higher level of LRF-1 activity was recovered from nuclei of light-grown plants than from dark-treated plants. Footprinting experiments revealed that LRF-1 binds to the Box X region (-149 to -134bp of SSU5B), that is conserved in SSU13, SSU5A and SSU5B and contains a GATAAG motif (GA-1 site in Table 2). Further, promoter deletion analysis by transient gene expression confirmed that the region between -205 to -83bp, containing the Box X, is required for the phytochrome response (Rolfe and Tobin, 1991). Although nuclear extracts

of *Lemna gibba* also have weak binding activity to BoxII of pea *rbcS-3A*, the binding activity does not change in response to light. In addition, no sequence showing greater than 53% identity with BoxII is present from -205bp to the transcription start site (+1bp) of the SSU5B promoter region. It was suggested that the BoxII sequence is not required for phytochrome regulation in *Lemna gibba* SSU5B, even though the role of BoxII in light-dependent regulation has been well characterized in dicot plants (Gilmartin *et al.*, 1990).

Although different isoforms of *rbcS* have been reported in terms of their isoelectric point, peptide fingerprint and protein sequence (Chen *et al.*, 1976; Gray *et al.*, 1976; Strobaek *et al.*, 1976), there is little information on the roles of different members of the *rbcS* family in terms of C3- and C4-type gene expression. Less work has been done regarding the tissue-specific expression and the photosynthetic types (C3- or C4-form) of the *rbcS* family in C4 plants.

Sheen and Bogorad (1986) isolated three *rbcS* cDNA clones, SS1, SS6 and SS7, from maize leaves that represented three different members of the *rbcS* gene family, and made specific probes for each gene (with little or no cross-hybridization). They discovered that, in dark-grown maize

seedling leaves, the transcripts of all three genes are about one-third as abundant in MS as in BS. The transcript ratio of the three genes, SS1:SS6:SS7, was measured at 4:1:3 in BS and 2:1:1 in MS. In 24-hour illuminated dark-grown maize leaves, total *rbcS* mRNA increased about 3-fold in BS. Under illumination, SS1 accounted for about 45% of the total leaf *rbcS* mRNA, SS7 for about 35% and SS6 for about 10%. The total *rbcS* mRNA decreased from the initial low level in MS. Since the pools of RNAs of the three genes were present in a ratio of 4:1:3 in BS at all stages of greening and 2:1:1 in etiolated MS, the authors suggest that there is some degree of differential expression of members of the *rbcS* gene family within each cell type as well as an overall photoregulated suppression in MS and stimulation in BS.

Recently, Schaffner and Sheen (1991) reported the analysis of promoter sequences of two maize *rbcS* genes, *rbcSZm1* and *rbcSZm3*, that correspond to the two highly expressed cDNA clones (Sheen and Bogorad, 1986). Nuclear run-on transcription of maize BS and MS showed that the corresponding transcripts are detected in both cell nuclei, albeit at a lower level in MS. Transient expression of chimeric CAT gene constructs showed that monocot (maize, wheat) but not dicot (pea, tobacco, and *Arabidopsis*) *rbcS* gene promoters are active in maize mesophyll protoplasts.

F. Formation of mRNA 3' End

In eukaryotic cells, mRNA is transcribed by RNA polymerase II, which generally terminates downstream of the functional mature 3' end of the mRNA. The mature mRNA is subsequently generated from the initial transcription product, by post-transcriptional processing, that includes capping of the 5' end, intron removal, and cleavage and polyadenylation of the 3' end.

(1) Transcription Termination

Although some evidence favors the existence of strong transcription termination sites (Kerppola and Kane, 1988), it is generally accepted that intrinsic transcription termination sites are weak, and transcription terminates far downstream of the polyadenylation site (Nevins and Darnell, 1978; Frayne *et al.*, 1984; Birnstiel *et al.*, 1985). Birnstiel *et al.* (1985) suggested that transcription termination depends on weak, serially-repeated (usually AT-rich) termination sequences. There is no direct evidence that multiple polyadenylation sites are generated from different transcription termination sites.

Multiple cleavage and polyadenylation sites of mRNA maturation have been suggested as a novel phenomenon in

plants (Dean *et al.*, 1987). The function and mechanism of such cleavage and polyadenylation is still unclear.

(2) Cleavage and Polyadenylation

Mutational analysis has shown that several sequence motifs are important in directing correct cleavage and polyadenylation of mRNAs (Wahle and Keller, 1992). Three motifs are involved in directing mRNA cleavage: (1) the AAUAAA sequence, a highly conserved sequence about 10 to 33 nucleotides upstream of the polyA site; (2) the G/U cluster (mostly YGUGUUY), approximately 30 nucleotides upstream of the AAUAAA site; and (3) the CAYUG motif, a sequence close to the cleavage and polyadenylation site and may base pair with the U4 snRNA for cleavage. Although proteins such as the cleavage and polyadenylation specificity factor (Wahle and Keller, 1992), cleavage stimulation factor (Takagaki *et al.*, 1989) and polyA-binding proteins (Das, 1993) have been characterized, the mechanisms of cleavage and polyadenylation remain unclear.

Joshi (1987) reported that the 3' regions of 46 plant genes are distinctly different from those of animal genes, and that the sequence motifs among plant genes are more diverse than those among the animal genes. Analysis of the cauliflower mosaic virus (Sanfacon *et al.*, 1991), the T-DNA-

encoded octopine synthase gene (MacDonald *et al.*, 1991), and the pea *rbcS-E9* gene (Hunt, 1988; Mogen *et al.*, 1992) revealed that the 3' regions of all of these genes are important in determining the correct *in vivo* cleavage and polyadenylation sites. Mogen *et al.* (1992) summarized these results and proposed as a general model that the plant polyadenylation signal requires a far-upstream element (containing a series of GU motifs), a near-upstream element (containing AAUAAA-like motifs), and a putative cleavage and polyadenylation site (usually containing the dinucleotide CA or UA).

Cleavage and polyadenylation of plant mRNAs have been studied far less than that of the animal mRNAs. There are few examples in animal mRNAs of multiple polyadenylation sites (Boardman *et al.*, 1985; Birnstiel *et al.*, 1985; Browne and Dodgson, 1993), generally, the cleavage and polyadenylation of animal mRNAs occurs predominantly at one site (Nevins and Darnell, 1978). In contrast to animal mRNAs, many plant mRNAs have multiple polyadenylation sites (Dean *et al.*, 1986; Hernandez-Lucas *et al.*, 1986; Montoliu *et al.*, 1990), suggesting that this may be a common phenomenon for plant genes. The possible function of these multiple polyadenylation sites is not clear (Gallie, 1993).

Dean *et al.* (1986) isolated *rbcS* cDNA clones from petunia including twelve corresponding to the gene SSU301 and thirteen corresponding to SSU511. The cDNA clones corresponding to each gene fell into three groups with different polyadenylation sites (mapping about 30 to 70 nucleotides apart). The middle site was the one predominantly used. The first A residue in the mRNA polyA tail corresponds to an A residue in the homologous genomic sequence. A putative polyadenylation signal, which most closely resembles the animal AAUAAA sequence, was identified in a region 15 to 29 nucleotides upstream of each polyadenylation site. However, these putative signal sequences diverged significantly from the animal sequence. The putative G/U cluster was found immediately downstream of many, but not all, of the polyadenylation sites, and no sequences corresponding to CAYUG were found.

G. Gene Regulation by mRNA 3' Untranslated Regions

Recent studies have greatly enhanced our knowledge about the importance of the mRNA 3' untranslated regions (3' UTR) in gene regulation. The 3' UTR has been found to be the repository site of all signals determining mRNA localization, of almost all signals controlling polyadenylation, of many motifs regulating mRNA stability,

and of signals controlling translation initiation, as well as of the signals of differentiation (Jackson, 1993).

(1) Messenger RNA Translation

During spermatogenesis in mice, protamine mRNAs are activated for translation in the cytoplasm. A 60-nucleotide sequence motif at the extreme 3' end of these mRNAs is necessary and sufficient for activation *in vivo* (Braun, 1991). Proteins that specifically bind to this sequence motif can be found only when the protamine mRNAs are inactivated, but not when the protamine mRNAs are actively translated. *In vitro* experiments revealed that a 18kd binding protein could effectively inhibit translation of mRNAs bearing the 3' UTR of protamine 2 mRNA but not of mRNAs bearing a different 3' UTR.

Recently, the UA-rich motifs (consensus UUAUUUAU) commonly found in the 3' UTR of lymphokine and proto-oncogene mRNAs were found capable of down-regulating translation initiation (Jackson, 1993).

(2) Control of PolyA Tail Length

During *Xenopus* oogenesis, some masked untranslated mRNAs with short polyA tails (30-50 residues) are stored in the

cytoplasm of oocytes. These mRNAs undergo further polyadenylation and translational activation either at maturation or following fertilization. The signals governing such stage-specific changes in polyA tail length were mapped to the specific motifs within the 3' UTR of the mRNAs. These cytoplasmic polyadenylation elements (CPEs) determine both the shortening of the polyA tail during oogenesis and the polyadenylation upon maturation. The canonical nuclear polyadenylation signal AAUAAA is required during the elongation of the polyA tail. Further studies revealed that the variations in the CPE sequence (consensus UUUUUAU or UUUUAAU), the distance of the CPE from the AAUAAA motif, and the neighboring sequences determine the subtle differences among the different mRNAs in the exact timing of adenylation and the degree of lengthening of the polyA tail (Wickens, 1992; Jackson, 1993). Although some of the CPE-binding proteins have been identified through cross-linking experiments, and differential phosphorylation of these proteins has been found during development of the oocytes, the mechanism of the protein-CPE interaction is largely unknown (Richter, 1991; Standart, 1992).

It is still not clear whether the length of the polyA tail can directly control translation initiation. Protection of the polyA tail in mRNA by the 70kd polyA-binding protein (PABP) inhibits the cessation of translation

of the ribosomal protein mRNAs during *Xenopus* oocyte maturation. In all cases, when a messenger RNA with a modest polyA tail, that has a 3'-deoxyadenosine residue added *in vitro*, was injected into the oocyte, it could not be activated for translation at maturation. The addition of a polyA tail *in vitro* followed by injection into the primary oocytes resulted in the premature translation of the B4 mRNA. However, this was not true for some other mRNAs. Simon *et al.* (1992) suggested that the dynamic process of extension of the polyA tail is the cause of translation activation in these cases. Further *in vitro* experiments showed that the translation activation of the ribonucleotide reductase and cyclin A mRNAs could be inhibited by antisense transcripts to a 140-nucleotide segment located centrally in the 3' UTR. The fact that there was no change in the length of polyA tail, suggested that polyadenylation *in vivo* might not be the primary cause of translation activation (Standart, 1992).

(3) mRNA Localization

Subcellular studies have shown that mRNAs are not evenly distributed within the rough endoplasmic reticulum (RER) or the cytoplasm (reviewed by Okita *et al.*, 1994). The 3' UTRs of the mRNAs have been shown to be very important targeting signals for correct localization; such specific localization

of mRNAs might be independent of translation. High resolution *in situ* hybridization revealed that the *chaoptic* mRNAs of the *Drosophila* eye photoreceptor cells are not randomly localized on subdomains of RER, and are uniformly distributed throughout the cell length. Transcripts of the *chp2* mutants that lack the 3' UTR are severely restricted to the apical regions of the cell, whereas *chp1* transcripts containing a point mutation displayed a less pronounced gradient. The motifs that are important in directing mRNA transport and retention at their ultimate sites have not been well characterized. There is little information on the localization of mRNAs that produce proteins targeted to cellular compartments such as chloroplast and mitochondria. These motifs could interact with the cell skeletal system, e.g. microtubules and microfilaments (Hesketh and Pryme, 1991; Sharpless *et al.*, 1993).

(4) mRNA Degradation

The stability of an mRNA molecule could be affected by many different elements including the cap structure, 5' UTR secondary structures, premature termination codons, open reading frame sequences, 3' UTR sequences, as well as the polyA tail (reviewed by Sachs, 1993). Among these elements, the 3' UTR sequences have received the greatest attention.

The 3' UTR of the transferrin receptor mRNA contains five distinct stem-loop structures (iron-responsive elements, IREs) capable of binding the IRE-binding protein (IRE-BP). The affinity of the IRE-BP for these sequences is regulated by cellular iron. Changes in affinity occur through dissociation and reassociation of an iron-sulfur cluster within the IRE-BP. When deficient iron, IRE-BP is in its high affinity state for IRE-binding, which results in decreased translation of ferritin mRNA and increased stability of the transferrin receptor mRNA by binding to the IREs in these transcripts. When iron is abundant, the IRE-BP is in its [4Fe-4S] state, that has high aconitase activity but low affinity for IREs. Binding to the IREs prevents association of destabilizing factors to the destabilizing sequence (Klausner *et al.*, 1993).

AU-rich elements (AREs) were first described in the 3' UTR of the GM-CSF mRNA by Shaw and Kamen (1986). These sequences (consensus loosely defined as AUUUA repeated once or several times within the 3' UTR) stimulated the degradation of the stable β -globin mRNA. Shyu *et al.* (1991) discovered that the ARE from c-fos mRNA mediates decay by first stimulating deadenylation and then providing an element that stimulates the next phase of the degradation process. They made point mutations which maintained the

ability of the ARE to stimulate deadenylation but inhibited its ability to stimulate mRNA decay.

(5) Differential Expression

In animal cells, polyadenylation at multiple sites associate with differential expression of some genes (Amara *et al.*, 1984; Birnstiel *et al.*, 1985; Saez *et al.*, 1990). However, differential expression of mRNAs with different polyadenylation sites has not been reported in plants.

Dietrich *et al.* (1992) found that 3' UTR sequences are required to activate a gene expressed in the root cortex of embryos and seedlings. The mRNA of the AX92 gene is prevalent in the root cortex of oilseed rape seedlings and mature plants, but at very low levels in all other vegetative organs of the plants. The researchers fused the GUS structural gene and the promoter region of the AX92 gene, with the 3' UTRs of the AX92 gene or the nopaline synthase gene (NOS), and then transformed oilseed rape. Transgenic plants with AX92 3' UTR showed GUS activity in both developing cortex cells in the root apical meristems of the seedlings and in cortex cells at the root end of embryonic axes. Transgenic plants with NOS 3' UTR showed GUS activity only in lower hypocotyls, but not in the axes of embryos and roots.

H. Transient Gene Expression

Expression of a gene is generally regulated by the interaction between *cis*-elements of the gene and the *trans*-acting factors within the cell. Promoter regions contain important regulatory elements in most of genes. Thus, the expression of a gene can be studied by introducing a chimeric gene into the target tissues in which a reporter gene is fused with the promoter region of a designated gene. Detection of the reporter gene product would be expected to mimic the expression of the native, designated gene.

The introduction of foreign genes into plant cells for expression studies has been done successfully in many plant species by various methods. These include PEG precipitation, microinjection, electroporation, *Agrobacterium*-mediated transformation, and particle bombardment. Among these, the particle bombardment method is useful for transient gene expression and stable transformation in monocots (Birch and Franks, 1991). The main benefit of this method is that intact plant organs can serve as targets.

Particle bombardment transformation is based on particle acceleration for penetration into the cell. Gun powder

explosion or expansion of compressed gas propels gold or tungsten particles, coated with DNA molecules, towards the plant tissue. The particles penetrate the surface of the plant tissue and to release the DNA into the plant cells. Once in the plant cells, the genes might be expressed within the cells transiently (before integrating) and/or be integrated into the plant genome.

The gene encoding β -glucuronidase (GUS) has been widely used as a sensitive and versatile marker gene in higher plants (Jefferson *et al.*, 1987). Most higher plants contain no detectable endogenous GUS activity. This *E. coli* gene product has a monomer molecular weight of about 68kd. The enzyme is very stable; it tolerates many detergents and ionic conditions. Histochemical localization of the GUS activity is accomplished by supplying the samples with 5-bromo-4-chloro-3-indolyl glucuronide (X-Glu). In the presence of the GUS enzyme, an indoxyl derivative is produced from the X-Glu. The colorless product then undergoes an oxidative dimerization to form the insoluble and highly colored indigo blue dye. The dimerization process is stimulated by atmospheric oxygen, and can be greatly enhanced by using an oxidation catalyst such as a potassium ferricyanide/ferrocyanide mixture (Jefferson, 1987).

Bansal et al. (1992) studied the promoter (2.1kb) of the maize *rbcS-m3* genes by fusing it with the GUS gene and the NOS 3' UTR region. They bombarded maize leaves with the chimeric gene constructs, and then histochemically localized the GUS gene product at the cellular level using light microscopy. The chimeric gene was expressed at about twice the level in MS as in BS of dark-grown maize leaves, and about equally in MS and BS upon illumination. Results of *in situ* and northern hybridization showed that the *rbcS-m3* mRNAs are barely detectable in MS but are detectable in unilluminated BS and doubled in illuminated BS. The authors suggested that this 2.1kb fragment of the *rbcS-m3* promoter region may be responsible for most of the light-dependent increase in *rbcS-m3* transcripts observed in the BS of greening leaves, and that transcriptional or posttranscriptional mechanisms are responsible for the lack of *rbcS-m3* transcripts in MS.

Recently, transient gene expression in sugarcane leaves has been achieved by particle bombardment (Gallo-Meagher and Irvine, 1993). Different promoters including rice actin, cauliflower mosaic virus (CaMV) 35S and maize ubiquitin 1, were fused with the GUS gene, along with the NOS 3' UTR. After bombardment, the samples were stained for GUS activity. The total GUS activity and the number of blue foci were counted and compared. All three chimeric genes

were expressed in sugarcane leaves. The maize ubiquitin 1 promoter produced the most GUS foci and the highest GUS activity. However, the authors did not localize the expression of the chimeric genes at the cellular level.

CHAPTER II. RESEARCH HYPOTHESIS

In view of the limited and inconclusive information regarding the photosynthetic types of *rbcS* genes in developing C4 plants, it is of interest to determine if there are *rbcS* genes that are only expressed in C3- or C4-form photosynthetic tissues (like C3- and C4-form PPdK genes). Alternatively, all members of the *rbcS* gene family may be expressed in both C3- and C4-form photosynthetic tissues. Although Schaffner and Sheen (1991) suggested that the C4 maize *rbcS* genes have a default C3-type expression pattern, it remains to be proven since only *in vitro* transient expression of the two genes in MS protoplasts were used in their experiment.

On the basis of the above information, it is proposed to test the hypotheses:

(1) The developmental transition of the Rubisco enzyme activity in sugarcane (C4) plants is primarily regulated at the transcriptional level. Based on this hypothesis, we want to find out whether such a transition is achieved through differential expression of different members of the *rbcS* family (C3-type member turns-off, C4-type member turns-

on at transcriptional level), or differential expression of the same member(s) in both patterns.

(2) The transcriptional regulation of *rbcS* genes involves sequence elements in the promoter region. Based on this, if different types (C3- and C4-type) of *rbcS* genes exist, there should be different sequence elements in the 5' flanking regions. Expression of different *rbcS* genes could be studied by fusing the promoter regions of the *rbcS* genes with the GUS gene, and delivering them into the sugarcane leaf.

CHAPTER III. MATERIALS AND METHODS

A. Plant Materials and Growth Conditions

Sugarcane setts, consisting of stem segments, about 10cm in length containing nodal buds and root primordia from field grown plants, *Saccharum* spp. hybrid cultivar H32-8560, were kindly provided by the Experimental Station of the Hawaiian Sugar Planters' Association (HSPA). The setts were soaked in water at 52°C for 20 minutes, then in 600mg/L Benlate fungicide water-solution (or 10% Clorox) at room temperature for 1 minute, and then planted in moist vermiculite trays in a greenhouse. The germinated shoots were grown until the leaves developed to the desired stage. The setts were also germinated in the dark. After one month, the etiolated plants were exposed to light for 24 hours, and the greening leaves (leaf 1 to 3) were harvested.

Different leaves on the sugarcane plants were identified by their positions, following the method of Kuijper (1915) according to Benda (1978), e.g. the leaf with the top visible dewlap was named Leaf 1. Leaf numbers increase down the stalk. The yellowish leaves, that were still wrapped within the leaf roll, were referred to as unexpanded leaves.

Leaves whose blades were still partially inclosed in the leaf roll were termed expanding leaves. Young leaves were referred to as the just fully expanded leaves (leaf 1 to 3). Mature leaves were those older than leaf 3.

Embryogenic callus, cell suspension culture and green callus (green leaf primordia) of sugarcane (same cultivar as above) were also provided by HSPA (for a review of growth conditions, see Maretzki, 1987).

B. Molecular Cloning of *Scrbcs* Genes

A sugarcane (cv. H32-8560) genomic library (Albert et al., 1991) was the source of sugarcane genes isolated in this work. A near-full-length maize *rbcS* cDNA clone (pJL10; about 470bp insert), obtained from Nelson (Yale University; see Langdale et al., 1988), was used as a heterologous probe for isolating *rbcS* genes.

pJL10 cDNA, labelled with $\alpha^{32}\text{P}$ -dCTP (3000Ci/mmol, DuPont) by the method of random primed DNA labelling kit instruction (Boehringer-Mannheim), was used as a probe to screen the genomic library. Hybridization was by the method of Church and Gilbert (1984). Both prehybridization and hybridization were done at 50°C in a hybridization oven, with the final

wash at 60°C. From one million recombinant phages (about three genome equivalents), one phage clone was selected for further analysis.

Phage DNA from the λ -clone was isolated by the DEAE column method (Helms *et al.*, 1987), except that the plate lysate was diluted with 2 ml water before passing through the column. After digestion with *EcoRI* and *HindIII*, the DNA was electrophoresed, Southern-blotted, and hybridized with the maize *rbcS* cDNA probe. A 2.3kb fragment was identified as a sugarcane *rbcS* genomic clone. This fragment, *scrbcS-1*, was subcloned into the *EcoRI* and *HindIII* sites of pBluescript II KS (+) vector (Stratagene).

Deletions were made from both ends of the *scrbcS-1* by the "Erase-A-Base" (Promega) protocol, so that both strands of the fragment could be sequenced. Plasmid transformation of the *E. coli* Subcloning Efficiency DH5 α TM competent cells (BRL) and color selection of recombinants were done by the BRL protocol.

Plasmid DNA was prepared using a modified boiling procedure (Sambrook *et al.*, 1989). Briefly, a single bacterial colony was grown in 5 ml LB with 100 μ g/ml ampicillin at 37°C overnight, and 1.5 ml culture was centrifuged at maximum speed in an *eppendorf* Centrifuge

5415C for 1 minute. The pellet was washed in 1 ml water, and resuspended in 0.32 ml STET (8% sucrose, 0.5% Triton X-100, 50mM EDTA, pH8, and 10mM Tris, pH8) by vortexing. After 32 μ l fresh lysozyme (10mg/ml in 10mM Tris, pH8) was added, the tube was put in boiling water for 1 minute, quickly cooled on ice, and then centrifuged at maximum speed as above for 30 minutes. The pellet was removed with a toothpick, the aqueous phase was mixed with 170 μ l 7.5M NH₄Ac and 550 μ l cold isopropanol, centrifuged at maximum speed as above for 30 minutes in 4°C, and washed with 70% ethanol. The plasmid DNA was vacuum dried and then resuspended in 200 μ l water with 1 μ l 10mg/ml RNase A and stored at 4°C.

The plasmid DNA was further purified for sequencing. After a 5 minute spin at maximum speed in an *ependorf* Centrifuge 5415C, the aqueous phase of miniprep DNA (200 μ l) was transferred to a new tube, mixed with 40 μ l 5M NaCl and 240 μ l 13% PEG, incubated on ice for at least 20 minutes, and centrifuged at maximum speed as above for 30 minutes at 4°C. The pellet was washed twice with 70% ethanol, air dried, and resuspended in 200 μ l TE (10mM Tris, 1mM EDTA, pH8). Fifty microliters of resuspended DNA solution was removed, mixed with 5 μ l denaturing solution (2M NaOH, 2mM EDTA), incubated at 37°C for 30 minutes, and precipitated with 5.5 μ l 3M NaAc (pH5.2) and 150 μ l cold ethanol (-20°C). After centrifuging at maximum speed as above for 30 minutes at 4°C, the pellet

was resuspended in 1 μ l primer (10ng/ μ l), 2 μ l sequencing reaction buffer (USB) and 7 μ l water. Sequencing reactions were done following the USB sequencing kit instruction (version 2.0).

The sequencing reaction mixtures were analyzed in 6% acrylamide denaturing gel (with 6M urea) using an IBI sequencing apparatus following the manufacturer's recommendation.

The *scrbcS-1* DNA was used as a probe to screen the sugarcane genomic library again as described above. From one million recombinant phages, twenty positive clones were identified. After digesting with *EcoRI* and *HindIII*, the λ DNA from each of the twenty positive genomic clones showed only one hybridization-positive band. Based on their restriction enzyme digestion and hybridization patterns, these clones were grouped into seven groups (*ScrbcS-1* belongs to one of these groups). Within each group, all the λ clones have exactly the same restriction enzyme digestion and hybridization pattern (data not shown). We subcloned *scrbcS* genes from one λ clone of each group. One of them, *scrbcS-3* was completely sequenced. It contains the promoter region and the first exon (total 1,315bp). Other genes were partially sequenced.

C. Extraction of Genomic DNA from Sugarcane Leaf Sample

Young sugarcane leaves from greenhouse-grown plants were cut and frozen on dry ice. Genomic DNA was extracted following the CTAB method (Doyle and Doyle, 1989). The concentration of DNA solution was determined using a fluorometer (Hoefer Scientific Instrument).

D. Copy Number Reconstruction

Ten μg of genomic DNA was digested to completion with *EcoRI* and *HindIII* (Promega), and then electrophoresed, along with 4.0pg, 8.0pg, 16pg, 32pg, 64pg and 124pg of the 2.3kb *scrbcS-1* DNA, in a 0.7% agarose gel in TBE buffer containing 0.5 $\mu\text{g/ml}$ ethidium bromide. Assuming the sugarcane variety we used has a genome size of $5.8 \times 10^9 \text{bp}$ (Albert, 1991), 10 μg genomic DNA would contain 4.0pg of the 2.3kb *scrbcS-1* DNA. Southern blotting of the DNA fragments to Hybond N+ nylon membrane (Amersham) was done according to the manufacturer's recommendations, and hybridization was by the method of Church and Gilbert (1984). Sixty ng of the 2.3kb *scrbcS-1* DNA was labelled with $\alpha^{32}\text{P}$ -dCTP (DuPont) as described above and used as a probe. Prehybridization and hybridization was done at 50°C, with the final wash at 65°C.

E. RNA Extraction from Sugarcane Leaves

Frozen sugarcane leaf samples (100mg) were used for total RNA extraction by the small scale RNA isolation method (Sun, personal communication). After grinding in liquid nitrogen, the frozen leaf powder was mixed with 250 μ l extraction buffer (0.1M LiCl, 0.1M EDTA, 1% SDS, and 0.1M Tris, pH8) and 250 μ l phenol, both prewarmed to 80°C. The samples were vortexed for 30 seconds; mixed with 250 μ l chloroform and isoamylalcohol (24:1); and vortexed again. After a 5 minute centrifugation in an *eppendorf* Centrifuge 5415C at maximum speed (16,000g) in room temperature, the aqueous phase was transferred to a new *eppendorf* tube, mixed with an equal volume of 4M LiCl, and incubated in a dry ice-EtOH bath for 1 hour. The precipitate was pelleted by centrifugation at maximum speed as above in 4°C for 30 minutes, washed with 100% cold EtOH, vacuum dried for 1 minute, and resuspended in 250 μ l DEPC-treated water. RNA was precipitated again with 0.1 volume 3M NaAc, pH 5.2 and 2 volumes of cold EtOH at -20°C for 2 hours, and then centrifuged as above. After washing and drying, the RNA was resuspended in DEPC-treated water. The RNA concentration was determined by measuring the absorption at 260nm, with a UV spectrophotometer (Shimadzu, UV-160). Typical RNA yields were 25-50 μ g for each 100mg sample.

F. Northern Hybridization

RNA was size-separated by gel electrophoresis through 0.66M formaldehyde denaturing gels (1.5% agarose) as described by Fourney *et al.* (1988). Blotting and fixation of RNA to Hybond N+ nylon membrane (Amersham) was done following the manufacturer's recommendations. Probe-labelling and hybridization were the same as described above for Southern blots. Prehybridization and hybridization was done at 50°C, with the final wash at 65°C.

G. Generation of a cDNA Pool by Reverse Transcription

Total RNA from different sugarcane samples was reverse transcribed using M-MLV reverse-transcriptase (BRL) following the manufacturer's recommendations. The 50 μ l reaction mixture contained 50mM Tris (pH8.3), 75mM KCl, 3mM MgCl₂, 10mM DTT, 0.5mM each of the four dNTPs, 100ng RACE-T primer (Table 3), 1 μ g sample RNA and 500U M-MLV reverse-transcriptase. After incubation at 37°C for 60 minutes, the reaction mixture (cDNA pool) was stored at -20°C until used.

Table 3. Oligonucleotide Primer Sequences. All primers were synthesized by National Biosciences, Inc., Plymouth, MN.

| NAME | Nucleotide Sequence |
|-----------------|--|
| RACE-T | 5'-CCGGAATTCTAGAGCTCGAGTTTTTTTTTTTTTTTTTT-3' |
| RACE-adapter | 5'-CCGGAATTCTAGAGCTCGAG-3' |
| <i>rbc</i> s#1- | 5'-AAGGAGAACCAGCTTACAG-3' |
| <i>rbc</i> s#2- | 5'-CCTTGCTCTGTGAATAACC-3' |
| <i>rbc</i> s#3+ | 5'-CCGGAATTCATCGCCTACAAGCCC-3' |
| <i>rbc</i> s#5+ | 5'-ATGTGGAAGCTGCCCATGTTCCG-3' |
| <i>rbc</i> s#6- | 5'-AGAGTCGACCATGGCTGGACTAGCTAG-3' |
| <i>rbc</i> s#9+ | 5'-TCTAGATGATCACCATGGCTCGAGAAGCTTGGAATGTCACTTTC-3' |

H. Rapid Amplification of cDNA Ends (RACE)

Computer analysis of all the available *rbc*s genes in GenBank shows that two stretches of the coding sequences, corresponding to the #1606-#1628 and #1741-#1758 of *scrbc*s-1 gene, are highly (more than 95% identity) conserved. Based on these two conserved sequences, two oligonucleotides, *rbc*s#5+ and *rbc*s#3+ (Table 3), were synthesized in the mRNA-sense direction. The primer *rbc*s#5+ is about 130bp upstream of the *rbc*s#3+. To facilitate cloning, six extra

nucleotides (CCGGAA) were added to the 5' end of the *rbcs#3+* oligomer to create an *EcoRI* site, as shown in Table 3.

The cDNA pool was first amplified with the RACE adapter (Table 3) and *rbcs#5+*, using a modified method of Frohman (1990). The 100- μ l reaction mixture contained:

- 1 μ l cDNA pool
- 10 μ l Taq Buffer (10x, Promega)
- 8 μ l dNTPs (2.5mM each)
- 10 μ l MgCl₂ (25mM)
- 1 μ l RACE adapter (0.5 μ g/ μ l)
- 1 μ l *rbcs#5+* (0.5 μ g/ μ l)
- 0.5 μ l Taq polymerase (5U/ μ l, Promega)
- 68.5 μ l water

After covering with mineral oil, the PCR reaction was carried out in a TempCycler (COY) with the following PCR profile:

94°C 5 min
67°C 2 min
72°C 40 min
(1 cycle)
↓
94°C 1 min
67°C 2 min
72°C 2 min
(30 cycles)
↓
72°C 10 min
(1 cycle)
↓
room temperature

I. Nested-RACE

The first amplification product was further amplified with the internal primer *rbcS#3+* and the RACE adapter. The 100- μ l reaction mixture contained:

- 1 μ l 1st amplification product
- 10 μ l Taq Buffer (10x, Promega)
- 8 μ l dNTPs (2.5mM each)
- 10 μ l MgCl₂ (25mM)
- 1 μ l RACE adapter (0.5 μ g/ μ l)
- 1 μ l *rbcS#3+* (0.5 μ g/ μ l)
- 0.5 μ l Taq polymerase (5U/ μ l, Promega)
- 68.5 μ l water

After covering with mineral oil, the PCR reaction was carried out in the TempCycler (COY) with the following PCR profile:

- 94°C 5 min
- 63°C 2 min
- 72°C 2 min
- (1 cycle)
- ↓
- 94°C 1 min
- 67°C 2 min
- 72°C 2 min
- (30 cycles)
- ↓
- 72°C 10 min
- (1 cycle)
- ↓
- room temperature

J. Cloning and Sequencing of the Nested-RACE Products

Nested-RACE products were purified using "Ultrafree-MC" 30,000 NMWL filter units (Millipore) following the manufacturer's recommendations. After digesting with *EcoRI* and *XbaI*, the DNA was purified again with "Ultrafree-MC" 30,000 NMWL filter units (Millipore), and then ligated to the *EcoRI* and *XbaI* sites of the pBluescript II KS+ vector (Stratagene). Recombinant DNA transformation, preparation and sequencing were done as above.

K. cDNA Library Construction and Screening

PolyA RNA was isolated from sugarcane mature leaf total RNA following the instructions of the "PolyAtract mRNA Isolation System II" (Promega). cDNA synthesis and adapter addition were done using the "Time-Saver cDNA Synthesis Kit" from Pharmacia. T4 DNA ligase (Promega) was used to insert the cDNA into the λ -gt11 vector (Promega). The recombinant λ DNA was packaged using the Gigapack II Gold system (Stratagene). Phage λ plating and titration were done following the "cDNA Synthesis and Cloning" procedure of the *Protocol and Applications Guide* of Promega (2nd ed.). From a total of 5×10^5 unamplified phages, about 5,000 phages were screened using *scrbcS-1* DNA as probe. Probe-labelling and

hybridization were the same as described above for the Southern blots. After three rounds of screening and purification, 20 positive phage λ -candidate clones were identified. λ DNA was isolated according to the "Mini-Prep Isolation of Lambda DNA" procedure in Promega's *Protocol and Applications Guide* (2nd ed.). The DNA was then digested with *EcoRI*. After gel separation, the cDNA inserts were recovered from the gel by the Prep-A-Gene Kit (BioRad), and subcloned into the *EcoRI* site of the pBluescript II KS+ vector (Stratagene). Plasmid transformation, DNA preparation, and sequencing were done as described above.

L. S1 Mapping

After digesting with *EcoRI* and *NheI*, the 550-bp (#1786-#2327) fragment of the *scrbcS-1* gene was purified from agarose and the 3'-end was labelled with $\alpha^{32}\text{P}$ -dCTP (DuPont) by Klenow (Promega) "fill-in" method following manufactory's instruction. The labelled fragment was then hybridized with total RNA extracted from mature sugarcane leaves and subsequently digested with S1 nuclease to remove all the unhybridized sequences according to Sambrook *et al.* (1989). After digestion, the mixture was electrophoresed on a 6% sequencing gel. The gel was exposed to X-ray film.

M. Differential Amplification of Total Sugarcane cDNA

Based on their lengths, the RACE products can be classified into five different groups. Two oligomers, *rbcs#1-* and *rbcs#2-* (Table 3), are antisense to the 3' UTR of the *rbcs* mRNA (#2201-#2219 for *rbcs#1-*, and #2258-#2076 for *rbcs#2-*). Oligomer *rbcs#1-* is specific to Group I mRNA, while *rbcs#2-* is specific to both Group I and Group II mRNA.

cDNA pools of sugarcane leaves at different developmental stages were amplified by primer *rbcs#5+* and either *rbcs#1-* or *rbcs#2-* primer. The 100- μ l reaction mixture contained:

- 2 μ l cDNA pool
- 10 μ l Taq Buffer (10x, Promega)
- 8 μ l dNTPs (2.5mM each)
- 10 μ l MgCl₂ (25mM)
- 1 μ l *rbcs#1-* or *rbcs#2-* (0.5 μ g/ μ l)
- 1 μ l *rbcs#5+* (0.5 μ g/ μ l)
- 0.5 μ l Taq polymerase (5U/ μ l, Promega)
- 67.5 μ l water

After covering with mineral oil, the PCR reaction was carried out in the TempCycler (COY) with the following PCR profile:

94°C 10 min
55°C 2 min
72°C 40 min
(1 cycle)
↓
94°C 2 min
55°C 2 min
72°C 2 min
(30 cycles)
↓
72°C 10 min
(1 cycle)
↓
room temperature

The amplification products were then electrophoresed on a 3% low-melting agarose gel.

N. Differential Amplification of Total Sugarcane Genomic DNA

Total genomic DNA from sugarcane leaves was amplified by the *rbcS#5+* primer and either the *rbcS#1-* or *rbcS#2-* primer. The 100- μ l reaction mixture contained:

1 μ l sugarcane genomic DNA (60ng/ μ l)
10 μ l Taq Buffer (10x, Promega)
8 μ l dNTPs (2.5mM each)
10 μ l MgCl₂ (25mM)
1 μ l *rbcS#1-* or *rbcS#2-* (0.5 μ g/ μ l)
1 μ l *rbcS#5+* (0.5 μ g/ μ l)
0.5 μ l Taq polymerase (5U/ μ l, Promega)
68.5 μ l water

After covering with mineral oil, the PCR reaction was carried out in the TempCycler (COY) with the following PCR profile:

94°C 2 min
55°C 2 min
72°C 2 min
(30 cycles)
↓
72°C 10 min
(1 cycle)
↓
room temperature

The amplification products were then electrophoresed on a 3% low-melting point agarose gel.

O. Chimeric Gene Construction

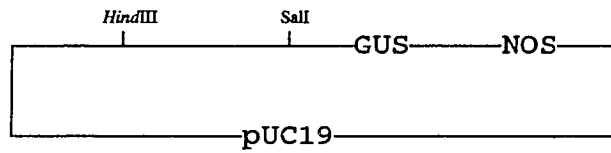
Four chimeric *rbcs* genes, pWD1, pWD2, pWD3, and pHT1011, were constructed for transient expression.

pWD1: The promoter region and the ATG codon of the *scrbcs-1* gene (nt sequence #1 to #1174) was removed by partial digestion with *HindIII* and *NcoI*. This fragment was inserted into the promoterless plasmid construct made by Bayraktaroglu (unpublished results), that contains the GUS structural gene and NOS termination sequences (see pBI vector series, GUS manual, Clontech).

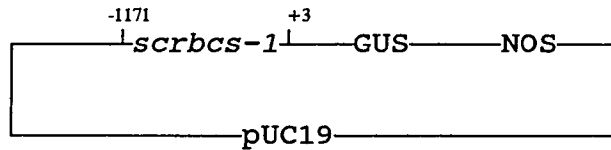
pWD2: This construct is the same as the pWD1, except with a shorter promoter region of *scrbcS-1* (nt sequence #1 to #1044; with no TATA box region, as a result of incorrect partial digestion).

pHT1011: It is the same as the pWD1, except for the 5'-coding region of *scrbcS-1* (nt sequence #1 to #1529), which includes the first exon, intron and part of the second exon. It was constructed by Tu (unpublished results).

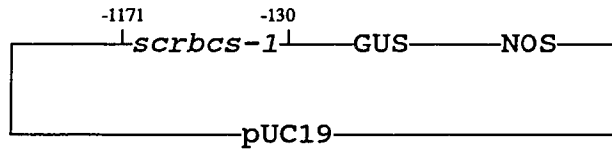
pWD3: The promoter region with the ATG codon of *scrbcS-3* gene were amplified by PCR. Two primers, *rbcs#9+* (#1-#20) and *rbcs#6-* (#1163-#1176), were made based on the 5' and 3' regions of the cloned *scrbcS-3* gene, respectively (Table 3). For ease cloning, restriction enzyme sites were added to the 5'-regions of both primers (underlined; primer *rbcs#9+*, *Xba*I, *Bcl*II, *Nco*I, *Xho*I, and *Hind*III; primer *rbcs#6-*, *Sal*I).



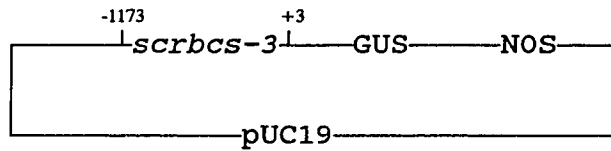
Promoterless Plasmid Construct



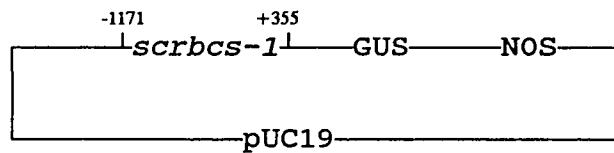
pWD1



pWD2



pWD3



pHT1011

Figure 1. Chimeric Gene Constructs. The maps are not drawn to scale. Numbers are nucleotides relative to translation start codon.

The cloned *scrbc3-3* gene, was used as template for amplification by the primers *rbcs#9+* and *rbcs#6-*. The 100- μ l reaction mixture contained:

1 μ l *scrbc3-3* gene DNA (10ng/ μ l)
10 μ l Taq Buffer (10x, Promega)
8 μ l dNTPs (2.5mM each)
10 μ l MgCl₂ (25mM)
1 μ l *rbcs#6-* (0.5 μ g/ μ l)
1 μ l *rbcs#9+* (1 μ g/ μ l)
0.5 μ l Taq polymerase (5U/ μ l, Promega)
68.5 μ l water

After covering with mineral oil, the PCR reaction was carried out in the TempCycler (COY) with the following PCR profile:

94°C 5 min
55°C 1 min
72°C 2 min
(1 cycle)
↓
94°C 2 min
72°C 3 min
(30 cycles)
↓
72°C 10 min
(1 cycle)
↓
room temperature

The amplification products were purified by centrifugal filtration using "Ultrafree-MC" 30,000 NMWL filter units (Millipore) following the manufacturer's recommendation, and then digested with *Hind*III and *Sal*I. After purification with the Millipore filter units, the DNA was inserted into

the promoterless plasmid construct as described above. The chimeric gene construct (pWD3) was verified by sequencing.

P. Data Analysis

All the sequence data analyses were performed using the GCG program (the Genetics Computer Group "Wisconsin Package", version 7.0, Devereux *et al.*, 1984), and the BLAST program (Altschul *et al.*, 1990) provided by GenBank (Benson *et al.*, 1993).

All the DNA/RNA molecular weights were calculated with the Nucleic Acid Fragment Sizing Program (Schaffer, 1981) according to the positions of the standard DNA/RNA molecules.

Q. Sugarcane Leaf Bombardment

Leaves, both from 3 month old plants placed under continuous darkness for 10 days and from 3 month old greenhouse plants, were excised, surface-sterilized, cut into segments of 2 x 1 cm in size, and placed (adaxial face down) on MS media plates (Murshige and Skoog, 1962). After culturing for 3 days under ambient light illumination, the

leaves were bombarded with DNA-coated particles using the Biolistics Particle Delivery System-1000 (BioRad). The particles were coated with one of the four chimeric plasmids, pWD1, pWD2, pWD3, or pHT1011. The control plasmid was UbiGUS (ubiquitin promoter fused with the GUS gene, see Christensen *et al.*, 1992).

(1) Coating of Gold Particles with DNA

Plasmid DNA was precipitated onto the gold particles (1.6 μ m Gold Microcarrier, BioRad) following the manufacturer's recommendation. Three mg of washed gold particles were mixed with 10 μ l DNA (1 μ g/ μ l), 50 μ l 2.5M CaCl₂ and 20 μ l 0.1M spermidine in an Eppendorf tube. After extensive vortexing (3 min), the DNA/particle mixture was centrifuged at the maximum speed for 10 seconds in an *eppendorf* Centrifuge 5415C, washed with 250 μ l ethanol, and resuspended in 60 μ l ethanol. The resulting suspension was aliquoted (10 μ l/carrier) and dried on five macrocarriers.

(2) Helium Power Setting

DNA carrier was placed at position "2" with a distance of 2cm from the rupture disc. The MS plate containing the target leaf was placed at position "3" (4cm from the DNA

carrier). Various pressures from 1100 to 2200 psi were used.

(3) GUS Gene Transient Expression Assay

After bombardment, the leaves were cultured at room temperature under continuous light for 24 hours and then placed into 1.5ml- Eppendorf tubes containing 1ml GUS staining solution (Sun, personal communication). The GUS staining solution was prepared by dissolving 0.5mg X-Glu in 20 μ l DMF and adding to 1ml of 0.1M NaHPO₄ pH7, 0.1% Triton X-100, 1mM EDTA pH8, 0.164g/L K₃Fe(CN)₆ and 0.212g/L K₄Fe(CN)₆. The leaf sample was infiltrated under house-vacuum twice for 15 min before incubating at 37°C for 24 hours to check the expression of the GUS gene.

(4) Cryostat Sectioning and Examination

GUS positive leaves showing blue spots were embedded in the Tissue-Tek^R OCT compound (Miles Inc., Elkhart, Indiana) and sectioned into 50 μ m-thick sections at -20°C, using a cryostat (Cryocut 1800, Reichert-Jung, Cambridge Instruments GmbH, Germany). The sections were thawed-mounted on room-temperature slides, mounted with CytoSealTM 280 (Stephens Scientific Cornwell Corporation, NJ) and examined under a microscope to determine the cellular location of expression.

R. PEP Carboxylase Activity

Leaf samples were harvested from greenhouse-grown sugarcane plants. Leaf segments of different developmental stages were studied, includes the base of unexpanded leaf (UEL-B, leaf -3), the tip portion of UEL (UEL-T), UEL-B cultured under fluorescent light for 3 days (UEL-L), and the mature leaf (ML, leaf 4). All steps of protein extraction and gel electrophoresis were carried out in 4°C or on ice. Sugarcane leaf segments (1.0g) were ground with 1.0g polyvinyl-polyrrrolidone in 10ml extraction buffer (25% ethylene glycol and 0.2g/l DTT in 0.1M potassium phosphate buffer, pH7.5) and filtered through four layers of cheese cloth. After centrifugation for 10 min at 27,000g (Sorvall SS34 rotor), ammonium sulfate was added to the supernatant while stirring to 80% saturation (0.516g per ml). After centrifuging for 10 min at 27,000g, the pellet was dissolved in 0.5ml suspension buffer (10% ethylene glycol and 0.1g/l DTT in 20mM potassium phosphate buffer, pH7.8) and centrifuged at maximum speed for 5 min in an *eppendorf* Centrifuge 5415C. The supernatant was purified twice through a 1ml Sephadex G-50 spin column equilibrated with the suspension buffer. The crude protein extract was stored at 4°C. Protein concentration was determined using the Bio-Rad Bradford Protein Assay method.

The BioRad Mini-Protein II Cell vertical gel system was used for electrophoresis. The comb and spacers were 1.5mm thick. Separation gel was made by mixing Solution A, B, C and water at a ratio of 1:2:4:1. After the separation gel was polymerized (about 30 minutes), the stacking gel was prepared by mixing Solution D, E, F and G at a ratio of 1:2:1:4 and layered on the top of the separation gel. After polymerization, the gel was pre-run for 90 min at a current of 10mA. Protein extracts containing equal amounts of protein (50 μ g) were mixed with Loading Dye (Bromophenol Blue in 50% sucrose) at a ratio of 3:1; the samples were electrophoresized at a current of 2mA overnight.

After electrophoresis, the gel was incubated with the staining solution (5mg/ml PEP monocyclohexyl ammonium, 3mg/ml Fast Violet B, 40 μ M NaHCO₃ and 30 μ M MgCl₂ in 0.1M Tris-HCl, pH8.0) at 37°C for 10 minutes. The stained gel was fixed and stored in 7% acetic acid at room temperature.

The gel component solutions are listed below:

Solution A: 0.46% TEMED in 0.57M Tris-HCl, pH7.5
Solution B: 20% Acrylamide and 0.8% Bis-Acrylamide
Solution C: 0.14% Ammonium persulfate
Solution D: 0.46% TEMED in 0.4M Tris-H₃PO₄, pH5.5
Solution E: 10% Acrylamide and 2.5% Bis-Acrylamide
Solution F: 4mg Riboflavin in 100ml water
Solution G: 40% sucrose
Running Buffer: 0.1mM DTE in 8mM Tris-H₃PO₄, pH7.1

CHAPTER IV. RESULTS

A. Sequences of the *Scrbcs* Genes

After digesting with *EcoRI* and *HindIII*, the λ DNA from the positive genomic clone showed only one hybridization-positive band, that was subcloned and named 13-1-HE fragment. The deletion clones of the 13-1-HE DNA fragment were sequenced and compiled to generate the entire sequence of the 13-1-HE fragment (Figure 2).

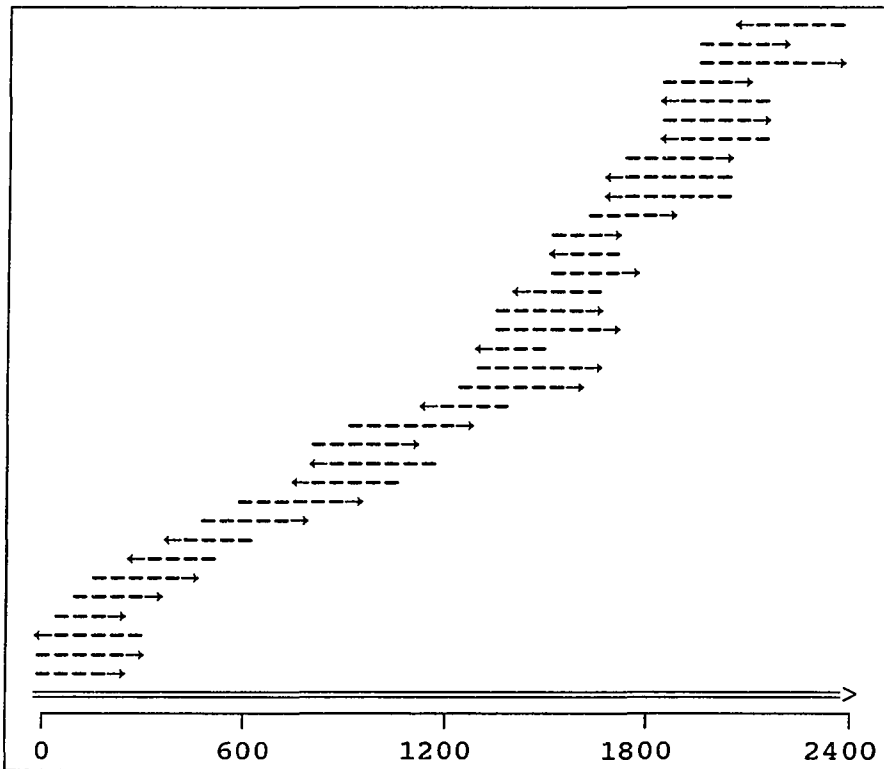


Figure 2. Schematic of the 13-1-HE Sequence Assembly. Arrows represent DNA fragments sequenced and their directions. The consensus sequence is shown in double solid lines. Numbers are in base pairs (bp).

Sequence analysis shows that the 13-1-HE DNA segment contains the full-length coding region (exon 1: 1172-1315bp; exon 2: 1411-1773bp), an intervening sequence (1316-1410bp), 1171bp upstream and 555bp downstream of the *rbcS* coding region. This clone was named *scrbcS-1* (Figure 3).

Results of a databank search using the Blast program revealed that this gene has the highest similarity with the three *rbcS* genes from maize (*Zea mays*). Both maize and sugarcane are C4 plants and belong to the same family *Poaceae*, tribe *Andropogoneae*. Sequence comparison shows that *scrbcS-1* has 88% similarity with the cloned maize *rbcS* genes (*mzrbcS*, see Lebrun *et al.*, 1987; *mzrbcS-1* and *mzrbcS-3*, see Schaffner and Sheen, 1991). Due to their high similarity, *mzrbcS* and *mzrbcS-3* are thought to be derived from the same locus in the maize genome. Figure 4 compares the promoter region of *scrbcS-1* with those of the three maize *rbcS* genes, showing that it contains the I box, G box-like region, GATA box, GC-rich region, and the monocot *rbcS* consensus sequences, that are important for the proper function of the two maize *rbcS* gene promoters (Schaffner and Sheen, 1991).

| | | |
|--|-------------------------------|----------|
| AAGCTTGAATGTCACTTTCCAATTCAATTCAATGCGCACGGAAACAAGCCAGCGACCCA | 60 | |
| TGAAACTGATTTCAATTCAGCCCAATTAATTCTATAGATCCAAACAGGCCGTGATCACTC | 120 | |
| ATCATGGTGGTCCGGACCAGCCGTCCTTGTAGAAATGTAAGTGAACACAAACAACCTGAGAG | 180 | |
| AGACCAGACCACTAGAGGCAAGAGCAAGCAGACGCAGCCGAAGCCCAAGCCCAAACTGT | 240 | |
| TTGTCCTTTGCCTTCCATTGTCGTTCTTAACTCATCCTTTCAAAACCACGACGAGCTGA | 300 | |
| ATGAACTGCTTCTGCTTGAGAGAGACGGCGGATCGGCGCTTACGGCAGAGCTTGTGCGCC | 360 | |
| GAGATTCGAGGAGGAAGGGACGAGTAGAGGAAGGAGTCTTCACCGACCGACGAGACCACG | 420 | |
| ACGACCACCTCGGATCCTGGCTCGTGACGACGAGCTAGCGAGCGAGTCCGGCCACGGCCG | 480 | |
| CGACGTGTCCGGCTCTATACCTCGGTTTGATCGGCGACGGGAGACCGGAGATAGCGAGCG | 540 | |
| AGTCCGGCCACGGCCAGGAAATTTTGAAACAGCTTGTGCGGCGCAGATTGCATGAGGA | 600 | |
| ATGAGACGAGCAGAGGAGTCTTCCTTACCGTCCGAGCAGGACCACCTGGGATCCTCGCT | 660 | |
| CGTGCAGAGCTATCGAGCGAGTCCGGCCACGCCACGGCCACGACGTCTGCTCAATATAT | 720 | |
| ACCTAGGTTTAATCGGAGATAGCGAGTGAGTCCGGGCACGCCAGGAAACTTTGGTTTCA | 780 | |
| GTAGTGTGACTTCACCACTTCTGCTTATCTATCTGCTACCTCCTCTATCTATCTACAAAG | 840 | |
| | | |
| | G box-like | |
| TCCCCGGCACATCACATATAGTCCAACCATGCCGTGGCT <u>CCATGTGG</u> CTGGCTCAGAC | 898 | |
| | | |
| I box | G box-like | GATA box |
| <u>GATAAGG</u> CGCGCCACGGGGACG <u>GACATGTGG</u> CGGCGGACGCGATC <u>AGGATAG</u> CCAGGCT | 959 | |
| | | |
| GC-rich region | monocot <i>rbcS</i> consensus | |
| <u>GGCCGGGTGGGCCAC</u> GGGAG <u>AACGGTGGCCACT</u> CCT <u>CCACAT</u> CCGCTTCGTCTGCCCCG | 1020 | |
| | | |
| | TATA box | |
| AACGACAGCCATCCATCGCCATGGACGCACGCGCTGCCTCTT <u>TATATA</u> TGCCGTCGG | 1080 | |
| TGGGGGAGCCTTACAGGACGACCCAAGCAATAAGCAGCAAAGCAAGGCAGCAAGCTCCCT | 1140 | |

Figure 3. Nucleotide and Deduced Amino Acid Sequences of the *Scrbcs-1* Gene. Putative translation start and stop codons are underlined. Potential regulatory sequences are bolded and underlined.

| | |
|---|------|
| CAGCTCACTACTATTACTAGCTAGTCCAGCCATGGCGCTCACCGTGATGGCCTCGTCCGCC MetAlaLeuThrValMetAlaSerSerAla | 1201 |
| ACCGCCGCTGCTCCGTTCCAGGGGCTCAAGTCCACCGCCAGCCTCCCCGTCGCCCGACGC ThrAlaAlaAlaProPheGlnGlyLeuLysSerThrAlaSerLeuProValAlaArgArg | 1261 |
| TCCACCACCAGCCTCGCCAAAGTCAGCAACGGCGGAAGGATCCGGTGCATGCAGGTAAC SerThrThrSerLeuAlaLysValSerAsnGlyGlyArgIleArgCysMetGln | 1320 |
| ATAACAAGCAACC CGCGCCTTTTCGTTCTTTGCTCCGGCCATCTTTTTGCAGCTGCAGCA | 1380 |
| AATGCAACCCAAACACGTACATCGATGCAGGTGTGGCCGGCGTACGGCAACAAGAAGTTC ValTrpProAlaTyrGlyAsnLysLysPhe | 1440 |
| GAGACGCTGTCGTACCTGCCGCCGCTGACGCAGGAGCAGCTGCTGAAGCAGGTGGACTAC GluThrLeuSerTyrLeuProProLeuThrGlnGluGlnLeuLeuLysGlnValAspTyr | 1500 |
| CTGCTGCGCAACAAC TGGGTGCCCTGCCTCGAGTTCAGCAAGGAAGGCTTCGTGTACCGC LeuLeuArgAsnAsnTrpValProCysLeuGluPheSerLysGluGlyPheValTyrArg | 1560 |
| GAGAACTCCACCTCCCCGTGCTACTACGACGGCCGGTACTGGACCATGTGGAAGCTGCCC GluAsnSerThrSerProCysTyrTyrAspGlyArgTyrTrpThrMetTrpLysLeuPro | 1620 |
| ATGTTCCGGCTGCACCGACGCGTCGCAGGTGTACAAGGAGCTGCAGGAGGCCATCGCGTCC MetPheGlyCysThrAspAlaSerGlnValTyrLysGluLeuGlnGluAlaIleAlaSer | 1680 |
| TACCCGGAGCTACGTGCCATCCTCGGCTTCGACAACATCAGGCAGACGCAGTGGCTCACG TyrProGluLeuArgAlaIleLeuGlyPheAspAsnIleArgGlnThrGlnTrpLeuThr | 1740 |
| TTCATCGCCTACAAGCCCGCGGCAGCGAGTAAAGTGCTAGCTAGCTCCTGCGTGAGCTG PheIleAlaTyrLysProAlaGlySerGluEnd | 1800 |
| CTAATGGCATGCCTGCCTGCCTCTACTGTTTCGGTTTTGCATCCTTCGGGTCACACACCTT | 1860 |
| GTTTCTTTTTTACCTTTTTTCTTATTCCCTTTGCTTCTCTGCTGCTAATGTATCCATTG | 1920 |
| TTGCAAGCATGGCATCATCCATCTCTCCCCCTACATATAACCAGCTAGAAACTACTGCAAC | 1980 |
| TAGCGTTGGGTGAGGAACATGTGAATGCAAGCTCCGGCTATCATATATGTGTAATATATG | 2040 |
| GCATTGGAGTTTGCTCTGGTTATTACAGAGCAAGGAATTACTTTCCCTTCCATATATATA | 2100 |
| TACAACATACTTTAATTTCAATTCATGCAACAATGCTAGAAGCCCTGTTGTAATTTTGCT | 2160 |
| TATTTACTTTTTTCCCCCTTTTTTGTCTCTTGAACTTTCTGTAAGCTGGTTCTCCTTT | 2220 |
| CTTAATAAATAAATGCACAGTAGGGGGAAATCCCTCCTGTTTCCATCAAAAAAGAATCTC | 2280 |
| AACTTCATTCTCATGTCTTTAAGATCATATCTTGAGTTATGAATTC | 2328 |

Figure 3. (Continued) Nucleotide and Deduced Amino Acid Sequences of the *Scrbcs-1* Gene.

```

                                                    -751
zmrbc3-3  TTGAATTCAG CC..CAATTC TGTA.G....   .ATCCAAACA
zmrbc3    GAATTCAG  CC..CAATTC TGTA.G....   .ATCCAAACA
zmrbc3-1  ATTAGTTCAG CC..CAACTC TATA.CGTAC ATA.....  CATCCAAACA
scrbc3-1  CTCGGTTTGA TCGGCGACGG GAGACCGGAG ATAGCGAGCG AGTCCGGCCA

                                                    -701
zmrbc3-3  GGGCCGGCGT CAGT...GCC TCAGGTG...   .
zmrbc3    GGGCCGGCGT CAGT...GCC TCAGGTG...   .
zmrbc3-1  CGCCCTGAGT ACGT...GCC TCAGGTGGTG CGGACCAG.. .CCGGGTAGA
scrbc3-1  CGGCCAGGGA AATTTTGGAA ACAGCTTGTG CGGCGCAGAT TGCATGAGGA

                                                    -651
zmrbc3-3  ..... AGAGAG....   .
zmrbc3    ..... AGAGAG....   .
zmrbc3-1  ATGTGGTAGA ACGGCCAACA GGTGAGTGTG AGAGAGACCA GGGCCA....
scrbc3-1  ATG.AGACGA GCAGAGGAGT CTTCCTTCAC CGTCGGAGCA GGACCACCTG

                                                    -601
zmrbc3-3  .CAGCAGACG ATGCAAAGAG CAAAAGTGG AAG.....CA GACGCAGCCG
zmrbc3    .CAGCAGACG ATGCAAAGAG CAAAAGTGG AAG.....CA GACGCAGCCG
zmrbc3-1  TCAGCAGCAG AGGCAATGAG C.AAA.CAG. ACG.....CA G....AGACG
scrbc3-1  GGATCCTCGC TCGTGCAGAG CTATCGAGCG AGTCCGGCCA CGCCACGGCC

                                                    -501
zmrbc3-3  AAGCCGAAGC CCAAGCCCAA AACTGTTTGT TCTTTGCCCA G.....
zmrbc3    AAGCCGAAGC CCAAGCCCAA AACTGTTTGT TCTTTGCCCA G.....
zmrbc3-1  AGGTCGAAAA CCAAG..... AAGTGTGTTG TCCTTGCCTA G.TTTCCATT
scrbc3-1  ACGACGTCTT GCTCAATATA TA.....   ....CCTA GGTTTAATCG

                                                    -451
zmrbc3-3  ..... AACCGCGACG AGCCTAAAC.   .
zmrbc3    ..... AACCGCGACG AGCCTAAAC.   .
zmrbc3-1  GTCGTACGTT C...TTAATA AACCGCGA.. .GCTGAAC.   .
scrbc3-1  GAGATAGCGA GTGAGTCCGG GCACGCCA.. .GGGAAACT TTGGTTTCAG

                                                    -401
zmrbc3-3  .....TGCC GCTTCCTCCT ATCTACAAGT CCCTGGCACA TCA.....
zmrbc3    .....TGCC GCTTCCTCCT ATCTACAAGT CCCTGGCACA TCA.....
zmrbc3-1  .....TGCT TGTGTTTCGAG AGAGACAGCG GATTGAGATT GCAGGATGCC
scrbc3-1  TAGTGTGACT TCACCATTTC TGCTTATCTA TCTGCTACCT CCTCTATCTA

```

Figure 4. Comparison of the Promoter Regions between the *Scrbcs-1* Gene with the Maize *Rbcs* Genes. Nucleotide numbers denote the distance from the translation start codon. Potential regulatory sequences are shown underlined. The possible transcription start sites of the *zmrbc3-1* and *zmrbc3-3* genes are shadowed.

-351

| | | | | | |
|-----------------|------------|------------|------------|------------|------------|
| <i>zmrbc3-3</i> | | | ...CGCATAG | TCCAA..... |CC |
| <i>zmrbc3</i> | | | ...CGCATAG | TCCAA..... |CC |
| <i>zmrbc3-1</i> | AGGAAGGGGA | CGACCGACTC | GACCGTATAG | TCCATCCGTG | GCGTGGCTCC |
| <i>scrbc3-1</i> | TCTACAAAGT | CCCCGGCACA | TCACATATAG | TCCAACCATG | CCGTGGCTCC |

| | | | | | |
|-----------------|--------------------------|-------------------|--------------------------|------------|--------------------------|
| | <i>G box-like</i> | | <i>I box</i> | | <i>G box-like</i> |
| <i>zmrbc3-3</i> | ATG..GCGCG | ..CAGGCGAT | AAGGCGCGCC | ACGGGGACGC | GACATGTGGT |
| <i>zmrbc3</i> | ATG..GCGCG | ..CAG.CGAT | AAGG.....C | ACGGGGACGC | GACATGTGGT |
| <i>zmrbc3-1</i> | ATGTGGCGCG | GACGGACGAT | AAGGC..... | | GACAAGTGGT |
| <i>scrbc3-1</i> | <u>ATGTGGCTGG</u> | CTCAGACGAT | <u>AAGGCGCGCC</u> | ACGGGGACGC | <u>GACATGTGGC</u> |

| | | | | | |
|-----------------|---------------|-----------------|--------------------------|-----------------------|--------------------------|
| | | <i>GATA box</i> | | <i>GC-rich region</i> | |
| <i>zmrbc3-3</i> | GGCGGAC... .. |GCGAT | CAGGATAGGG | CCAGGCTGGC | CGGGCGCGGC |
| <i>zmrbc3</i> | .GCGGAC... .. |GCGAT | CAGGATAGGG | CCAGGCT..C | CGGGCGCGGC |
| <i>zmrbc3-1</i> | GGCGGACGCG | CGCGTGCGAT | CAGGATAGG. | CCAGGCTGGC | CGGGTGCGGC |
| <i>scrbc3-1</i> | GGCGGAC... .. |GCGAT | <u>CAGGATAGG.</u> | CCAGGCTGGC | <u>CGGGTGCGGC</u> |

| | | | | | |
|-----------------|--------------------------|-------------------------------|--------------------------|------------|-------------|
| | | <i>monocot rbc3 consensus</i> | | | -201 |
| <i>zmrbc3-3</i> | CACGGGAGAA | CGGTGGCCAC | TCGTCCCACA | TCCGCTTCGT | CCTGTCTCTGT |
| <i>zmrbc3</i> | CACGGGAGAA | CGGTGGCCAC | TCGTCCCACA | TCCGCTTCGT | CCTGTCTCTGT |
| <i>zmrbc3-1</i> | CACGGGAGAA | CGGTGGCCAC | TCGTCCCACA | TCCGCTTCGT | CACG..... |
| <i>scrbc3-1</i> | <u>CACGGGAGAA</u> | <u>CGGTGGCCAC</u> | <u>TCCTCCCACA</u> | TCCGCTTCGT | CCTG..... |

| | | | | | |
|-----------------|------------|------------|------------|------------|------------|
| | | | | | -151 |
| <i>zmrbc3-3</i> | ACTGCGTCCT | GCCCCCAACG | AGAGCCGGAG | CCGGCCATCC | CGTCGCACAC |
| <i>zmrbc3</i> | ACTGCGTCCT | GCCCCCAACG | A.....GAG | CCGGCCATCC | CGTCGCACAC |
| <i>zmrbc3-1</i> | | ..CCCCAACG | AGAGGCGGAC | GCGG..ATCC | AGCGACATGG |
| <i>scrbc3-1</i> | | ..CCCGAACG | ACAGCCATCC | ATCGCCATGG | ACGCACGCGC |

| | | | | | |
|-----------------|------------|--------------------------|------------|------------|-------------------|
| | | <i>TATA box</i> | | | -101 |
| <i>zmrbc3-3</i> | TCTCCCCC.T | CTATATATGC | CGTCGGTGTG | GGGGAGCCTA | CTACAGGACG |
| <i>zmrbc3</i> | TCTCCCCC.T | CTATATATGC | CGTCGGTGTG | GGGGAGCCTA | CTACAGGACG |
| <i>zmrbc3-1</i> | ACATGGC..T | CTATATATGC | CGTCGGTG.. | GGGGAGCCCC | .TACAGGACG |
| <i>scrbc3-1</i> | GCTGCCTCTT | <u>CTATATATGC</u> | CGTCGGTG.. | GGGGAGCCT. | .TACAGGACG |

| | | | | | |
|-----------------|------------|------------|------------|------------|------------|
| | | | | | -51 |
| <i>zmrbc3-3</i> | ACCCAAGCA. | |AG | CAAG..... | |
| <i>zmrbc3</i> | ACCCAAGCA. | |AG | CAAG..... | |
| <i>zmrbc3-1</i> | ACCCAAGCA. | |AG | CAAGCTCGAT | C...TACTA |
| <i>scrbc3-1</i> | ACCCAAGCAA | TAAGCAGCAA | AGCAAGGCAG | CAAGCTCCCT | CAGCTCACTA |

| | | | | | |
|-----------------|------------|------------|------------|------------|------------|
| | | | | | -1 |
| <i>zmrbc3-3</i> | ...CAAGCAG | CGAGTACATA | CATACTAGGC | AGCCAGGCA. |GCC |
| <i>zmrbc3</i> | ...CAAGCAG | CGAGTACATA | CATACTAGGC | AGCCAGGCA. |GCC |
| <i>zmrbc3-1</i> | CTACTACTAG | CTGGTAC..A | CATACTAGCC | AGCCTGCCAG | CCAGCTTGCC |
| <i>scrbc3-1</i> | CTATTACTAG | CTAGTCCA.. | | |GCC |

Figure 4. (Continued) Comparison of the Promoter Regions between the *Scrbc3-1* Gene with the Maize *Rbc3* Genes

All of the genomic clones share very high homology (more than 95%) with the *scrbcS-1* gene at both the 5' and 3' regions (Figure 5, 6 & 7).

The *scrbcS-3* clone is a partial genomic clone, containing only the promoter region and the first exon (total 1315bp). This clone shows the greatest nucleotide difference from other *scrbcS* genes, although it still shares more than 90% identity with the *scrbcS-1* gene and is almost identical in the coding region and the proximal promoter region including the I box, G box-like region, GATA box, GC-rich region, and monocot *rbcS* consensus sequences (Figure 7).

| | | | | | | |
|-------------------|------------|------------|------------|-------------|-------------|-----|
| | | | | | | 50 |
| <i>scrbc</i> s-1 | AAGCTTGGAA | TGTCACTTTC | CAATTCAATT | CAATGCGCAC | GGAAACAAGC | |
| <i>scrbc</i> s-2 | AAGCTTGGAA | TGTCACTTTC | CAATTCAATT | CAATGCGCAC | GGAAACAAGC | |
| <i>scrbc</i> s-5 | AAGCTTGGAA | TGTCACTTTC | CAATTCAATT | CAATGCGCAC | GGAAACAAGC | |
| <i>scrbc</i> s-7 | AAGCTTGGAA | TGTCACTTTC | CAATTCAATT | CAATGCGCAC | GGAAACAAGC | |
| <i>scrbc</i> s-19 | AAGCTTGGAA | TGTCACTTTC | CAATTCAATT | GAATGCGCAC | GGAAACAAGC | |
| <i>scrbc</i> s-23 | AAGCTTGGAA | TGTCACTTTC | CAATTCAATT | CAATGCGCAC | GGAAACAAGC | |
| | | | | | | * |
| | | | | | | 100 |
| <i>scrbc</i> s-1 | CAGCGACCCA | TGAAACTGAT | TTCAATTCAG | CCCAATTAAT | TCTATAGATC | |
| <i>scrbc</i> s-2 | CAGCGACCCA | TGAAACTGAT | TTCAATTCAG | CCCAATTAAT | TCTATAGATC | |
| <i>scrbc</i> s-5 | CAGCGACCCA | TGAAACTGAT | TTCAATTCAG | CCCAATTAAT | TCTATAGATC | |
| <i>scrbc</i> s-7 | CAGCGACCCA | TGAAACTGAT | TTCAATTCAG | CCCAATTAAT | TCTATAGATC | |
| <i>scrbc</i> s-19 | CAACGACCCA | TGAAACTGAT | TTCAATTCAG | CCCAATTAAT | TCTATAGATC | |
| <i>scrbc</i> s-23 | CAGCGACCCA | TGAAACTGAT | TTCAATTCAG | CCCAATTAAT | TCTATAGATC | |
| | | | | | | 150 |
| <i>scrbc</i> s-1 | CAAACAGGCC | GTGATCACTC | ATCATGGTGG | TCCGGACCAG | CCGTCCTTGT | |
| <i>scrbc</i> s-2 | CAAACAGGCC | GTGATCACTC | ATCATGGTGG | TCCGGACCAG | CCGTCCTTGT | |
| <i>scrbc</i> s-5 | CAAACAGGCC | GTGATCACTC | ATCATGGTGG | TCCGGACCAG | CCGTCCTTGT | |
| <i>scrbc</i> s-7 | CAAACAGGCC | GTGATCACTC | ATCATGGTGG | TCCGGACCAG | CCGTCCTTGT | |
| <i>scrbc</i> s-19 | CAAACAGGCC | GTGATCACTC | ATCATGGTGG | TCCGGACCAG | CCGTCCTTGT | |
| <i>scrbc</i> s-23 | CAAACAGGCC | GTGATCACTC | ATCATGGTGG | TCCGGACCAG | CCGTCCTTGT | |
| | | | | | | * |
| | | | | | | 200 |
| <i>scrbc</i> s-1 | AGAATGTAAG | TGAACACAAA | CAACTGAGAG | AGACCAGACC | ACTAGAGGCCA | |
| <i>scrbc</i> s-2 | AGAATGTAAG | TGAACACAAA | CAACTGAGAG | AGACCAGACC | ACTAGAGGCCA | |
| <i>scrbc</i> s-5 | AGAATGTAAG | TGAACACAAA | CAACTGAGAG | AGACCAGACC | ACTAGAGGCCA | |
| <i>scrbc</i> s-7 | AGAATGTAAG | TGAACACAAA | CAACTGAGAG | AGACCAGACC | ACTAGAGGCCA | |
| <i>scrbc</i> s-19 | AGAATGTAGG | TGAACACAAA | CAACTGAGAG | AGACCAGACC | ACTAGAGGCCA | |
| <i>scrbc</i> s-23 | AGAATGTAAG | TGAACACAAA | CAACTGAGAG | AGACCAGACC | ACTAGAGGCCA | |
| | | | | | | * |
| | | | | | | 250 |
| <i>scrbc</i> s-1 | AGAGCAAGCA | GACGCAGCCG | AAGCCCAAGC | CCAAAACCTGT | TTGTCCTTTG | |
| <i>scrbc</i> s-2 | AGAGCAAGCA | GACGCAGCCG | AAGCCCAAGC | CCAAAACCTGT | TTGTCCTTTG | |
| <i>scrbc</i> s-5 | AGAGCAAGCA | GACGCAGCCG | AAGCCCAAGC | CCAAAACCTGT | TTGTCCTTTG | |
| <i>scrbc</i> s-7 | AGAGCAAGCA | GACGCAGCCG | AAGCCCAAGC | CCAAAACCTGT | TTGTCCTTTG | |
| <i>scrbc</i> s-19 | AGAGCAAGCA | GACGCAGCCG | AAGCCCAAGC | CCAAAACCTGT | TTGTCCTTTG | |
| <i>scrbc</i> s-23 | AGAGCAAGCA | GACGCAGCCG | AAGCCCAAGC | CCAAAACCTGT | TTGTCCTTTG | |
| | | | | | | 300 |
| <i>scrbc</i> s-1 | CCTTCCATTG | TCGTTCTTAA | CTCATCCTTT | CACAAACCAC | GACGAGCTGA | |
| <i>scrbc</i> s-2 | CCTTCCATTG | TCGTTCTTAA | CTCATCCTTT | CACAAACCAC | GACGAGCTGA | |
| <i>scrbc</i> s-5 | CCTTCCATTG | TCGTTCTTAA | CTCATCCTTT | CACAAACCAC | GACGAGCTGA | |
| <i>scrbc</i> s-7 | CCTTCCATTG | TCGTTCTTAA | CTCATCCTTT | CACAAACCAC | GACGAGCTGA | |
| <i>scrbc</i> s-19 | CCTTCCATTG | TCGTTCTTAA | CTCATCCTTT | CACAAACCAC | GACGAGCTGA | |
| <i>scrbc</i> s-23 | CCTTCCATTG | TCGTTCTTAA | CTCATCCTTT | CACAAACCAC | GACGAGCTGA | |

Figure 5. 5' UTR of the Cloned *Scrbc*s Genes. Differences in nucleotide are marked with "*".

| | | | | | |
|----------------|------------|-------------|------------|------------|------------|
| | | | | | 1848 |
| <i>scrbc1</i> | GCTAATGGCA | TGCCTGCCTG | CCTCTA.CTG | TTCGGTTTTG | CATCCTTCGG |
| <i>scrbc2</i> | GCTAATGGCA | TGCCTGCCTT | CCTCAACCTG | TTCGGTTTTG | CATCCTTCGG |
| <i>scrbc5</i> | GCTAATGGCA | TGCCTGCCTG | CCTCTACCTG | TTCGGTTTTG | CATCCTTCGG |
| <i>scrbc7</i> | GCTAATGGCA | TGCCTGCCTG | CCTCTACCTG | TTCGGTTTTG | CATCCTTCGG |
| <i>scrbc19</i> | GCTAATGGCA | TGCCTGCCTG | CCTCTACCTG | TTCGGTTTTG | CATCCTTCGG |
| <i>scrbc23</i> | GCTAATGGCA | TGCCTGCCTG | CCTCTACCTG | TTCGGTTTTG | CATCCTTCGG |
| | | * | * * | | |
| 1894 | | | | | |
| <i>scrbc1</i> | GTCACACACC | TT.GTTTCTT | TTTCACCTTT | TTCTTC...T | TATTCCTTTG |
| <i>scrbc2</i> | GTCACACACC | TTGGTTTCTT | TTTTACCTTT | TTCTCCTGTT | GGTTCCTTTG |
| <i>scrbc5</i> | GTCACACACC | TT.GTTTCTT | TTTCACCTTT | TTCTTC...T | TATTCCTTTG |
| <i>scrbc7</i> | GTCACACACC | TT.GTTTCTT | TTTCACCTTT | TTCTTC...T | TATTCCTTTG |
| <i>scrbc19</i> | GTCACACACC | TTGGTTTCTT | TTTTACCTTT | TTCTCC...T | GGTTCCTTTG |
| <i>scrbc23</i> | GTCACACACC | TT.GTTTCTT | TTTCACCTTT | TTCTTC...T | TATTCCTTTG |
| | | * | * | * *** | ** |
| 1944 | | | | | |
| <i>scrbc1</i> | CTTCTCTGCT | GCTAATGTAT | CCATTGTTGC | AAGCATGGCA | TCATCCATCT |
| <i>scrbc2</i> | CTTCTCTGCT | GCTGATGTAT | CCATTGTTGC | AAGCATGGCA | TTATCCATCT |
| <i>scrbc5</i> | CTTCTCTGCT | GCTAATGTAT | CCATTGTTGC | AAGCATGGCA | TCATCCATCT |
| <i>scrbc7</i> | CTTCTCTGCT | GCTAATGTAT | CCATTGTTGC | AAGCATGGCA | TCATCCATCT |
| <i>scrbc19</i> | CTTCTCTGCT | GCTGATGTAT | CCATTGTTGC | AAGCATGGCA | TCATCCATCT |
| <i>scrbc23</i> | CTTCTCTGCT | GCTAATGTAT | CCATTGTTGC | AAGCATGGCA | TCATCCATCT |
| | | * | | | * |
| 1994 | | | | | |
| <i>scrbc1</i> | CTCCCCCTAC | ATATACCAGC | TAGAAACTAC | TGCAACTAGC | GTTGGGTGAG |
| <i>scrbc2</i> | CTCCCCCTAC | ATATACCCTGC | TAGAAACTAC | TGCAACTAGC | GTTGGGTGAG |
| <i>scrbc5</i> | CTCCCCCTAC | ATATACCAGC | TAGAAACTAC | TGCAACTAGC | GTTGGGTGAG |
| <i>scrbc7</i> | CTCCCCCTAC | ATATACCAGC | TAGAAACTAC | TGCAACTAGC | GTTGGGTGAG |
| <i>scrbc19</i> | CTCCCCCTAC | ATATACCCTGC | TAGAAACTAC | TGCAACTAGC | GTTGGGTGAG |
| <i>scrbc23</i> | CTCCCCCTAC | ATATACCAGC | TAGAAACTAC | TGCAACTAGC | GTTGGGTGAG |
| | | * | * | | |
| 2044 | | | | | |
| <i>scrbc1</i> | GAACATGTGA | ATGCAAGCTC | CGGCTATCAT | ATATGTGTAA | TATATGGCAT |
| <i>scrbc2</i> | GAACATGTGA | ATGCAAGCTC | CGGCTATCAT | ATATGTGTAA | TATATGGCAT |
| <i>scrbc5</i> | GAACATGTGA | ATGCAAGCTC | CGGCTATCAT | ATATGTGTAA | TATATGGCAT |
| <i>scrbc7</i> | GAACATGTGA | ATGCAAGCTC | CGGCTATCAT | ATATGTGTAA | TATATGGCAT |
| <i>scrbc19</i> | GAACATGTGA | ATGCAAGCTC | CGGCTATCAT | ATATGTGTAA | TATATGGCAT |
| <i>scrbc23</i> | GAACATGTGA | ATGCAAGCTC | CGGCTATCAT | ATATGTGTAA | TATATGGCAT |

Figure 6. 3' UTR of the Cloned *Scrbc* Genes. Differences in nucleotide are marked with "*".

| | | | | | | |
|-------------------|------------|------------|-------------|------------|------------|------|
| | | | | | | 2094 |
| <i>scrbc</i> s-1 | TGGAGTTTGC | TCTGGTTATT | CACAGAGCAA | GGAATTACTT | TCCTTCCATA | |
| <i>scrbc</i> s-2 | TGGAGTTTGC | TCTGGTTATT | CACAGAGCAA | GGAAGTACTT | TCCTTCCATA | |
| <i>scrbc</i> s-5 | TGGAGTTTGC | TCTGGTTATT | CACAGAGCAA | GGAATTACTT | TCCTTCCATA | |
| <i>scrbc</i> s-7 | TGGAGTTTGC | TCTGGTTATT | CACAGAGCAA | GGAATTACTT | TCCTTCCATA | |
| <i>scrbc</i> s-19 | TGGAGTTTGC | TCTGGTTATT | CACAGAGCAA | GGAAGTATTT | TCCTTCCATA | |
| <i>scrbc</i> s-23 | TGGAGTTTGC | TCTGGTTATT | CACAGAGCAA | GGAATTACTT | TCCTTCCATA | |
| | | | | * | | |
| | | | | | | 2143 |
| <i>scrbc</i> s-1 | TATATATACA | ACATACTTTA | ATTTTCATTCC | ATGCAACAAT | .GCTAGAAGC | |
| <i>scrbc</i> s-2 | TATATATACA | ACATACTTTA | ATTTTCATTCC | ATGCAACAAT | GGCTAGAAGC | |
| <i>scrbc</i> s-5 | TATATATACA | ACATACTTTA | ATTTTCATTCC | ATGCAACAAT | GGCTAGAAGC | |
| <i>scrbc</i> s-7 | TATATATACA | ACATACTTTA | ATTTTCATTCC | ATGCAACAAT | GGCTAGAAGC | |
| <i>scrbc</i> s-19 | TATA...CA | ACATACTTTA | ATTTTCATTCC | ATGTAACAAT | GGCTAGAAGC | |
| <i>scrbc</i> s-23 | TATATATACA | ACATACTTTA | ATTTTCATTCC | ATGCAACAAT | GGCTAGAAGC | |
| | *** | | | * | * | |
| | | | | | | 2193 |
| <i>scrbc</i> s-1 | CCTGTTGTAA | TTTTGCTTAT | TTACTTTTTT | CCCCCTTTT | TTGTCTCTTG | |
| <i>scrbc</i> s-2 | CCTGTTGTAA | TTTTGCTTAT | TTACTTTTTT | T..CCTTTT | TTGTCTCTTG | |
| <i>scrbc</i> s-5 | CCTGTTGTAA | TTTTGCTTAT | TTACTTTTTT | CCCCCTTTT | TTGTCTCTTG | |
| <i>scrbc</i> s-7 | CCTGTTGTAA | TTTTGCTTAT | TTACTTTTTT | CCCCCTTTT | TTGTCTCTTG | |
| <i>scrbc</i> s-19 | CCTGTTGTAA | TTTTGCTTAT | TTACTTTTTT | TTCCCTTTT | TTGTCTCTTG | |
| <i>scrbc</i> s-23 | CCTGTTGTAA | TTTTGCTTAT | TTACTTTTTT | CCCCCTTTT | TTGTCTCTTG | |
| | | | | *** | * | |
| | | | | | | 2243 |
| <i>scrbc</i> s-1 | AACTTTTCTG | TAAGCTGGTT | CTCCTTTCTT | AATAAATAAA | TGCACAGTAG | |
| <i>scrbc</i> s-2 | AACTTTTCTG | TAAGCTGGTT | CTCCTTTCTT | AATAAATAAA | TGCACAGTAG | |
| <i>scrbc</i> s-5 | AACTTTTCTG | TAAGCTGGTT | CTCCTTTCTT | AATAAATAAA | TGCACAGTAG | |
| <i>scrbc</i> s-7 | AACTTTTCTG | TAAGCTGGTT | CTCCTTTCTT | AATAAATAAA | TGCACAGTAG | |
| <i>scrbc</i> s-19 | AACTTTTCTG | TATGCTGGTT | CTCCTTTCTT | AATAAATAAA | TGCACAGTAG | |
| <i>scrbc</i> s-23 | AACTTTTCTG | TAAGCTGGTT | CTCCTTTCTT | AATAAATAAA | TGCACAGTAG | |
| | | * | | | | |
| | | | | | | 2293 |
| <i>scrbc</i> s-1 | GGGGAAATCC | CTCCTGTTTC | CATCAAAAAA | GAATCTCTGA | ACTTCATTCT | |
| <i>scrbc</i> s-2 | GGGGAAATCC | CTCCTGTTTC | CATCAAAAAA | GAATCTCTGA | ACTTCATTCT | |
| <i>scrbc</i> s-5 | GGGGAAATCC | CTCCTGTTTC | CATCAAAAAA | GAATCTCTGA | ACTTCATTCT | |
| <i>scrbc</i> s-7 | GGGGAAATCC | CTCCTGTTTC | CATCAAAAAA | GAATCTCTGA | ACTTCATTCT | |
| <i>scrbc</i> s-19 | GGGGAAATCC | CTCCTGTTTC | CATCAAAA.. | | ..TTCATTCT | |
| <i>scrbc</i> s-23 | GGGGAAATCC | CTCCTGTTTC | CATCAAAAAA | GAATCTCTGA | ACTTCATTCT | |
| | * | | ** | ***** | ** | |
| | | | | | | 2328 |
| <i>scrbc</i> s-1 | CATGTCTTTA | AGATCATATC | TTGAGTTATG | AATTC | | |
| <i>scrbc</i> s-2 | CATGTCTTCA | AGATCATATC | TTGAGTTGTG | AATTC | | |
| <i>scrbc</i> s-5 | CATGTCTTTA | AGATCATATC | TTGAGTTATG | AATTC | | |
| <i>scrbc</i> s-7 | CATGTCTTTA | AGATCATATC | TTGAGTTATG | AATTC | | |
| <i>scrbc</i> s-19 | CATGTCTTTA | AGATCATATC | TTGAGTTGTG | AATTC | | |
| <i>scrbc</i> s-23 | CATGTCTTTA | AGATCATATC | TTGAGTTATG | AATTC | | |
| | * | | * | | | |

Figure 6. (Continued) 3' UTR of the Cloned *Scrbc*s Genes.

```

scrbc1-1  AAGCTTGGGAATGTCACCTTCCAATTCAATTCAATGCGCACGGAAACAAGC 50
scrbc1-3  AAGCTTGGGAATGTCACCTTCCGATT.....CAATGCGCACGGAAACAAGC 45
scrbc1-1  CAGCGACCCATGAAACTGATTTCAATTCAGCCCAATTAATTCTATAGATC 100
scrbc1-3  ...CGACCCATGAAACTGATTTCAATTCAGTCCAATTAATTCTATAGATC 92
scrbc1-1  CAAACAGGCCGTGATCACTCATCATGGTGGTCCGGACCAGCCGTCCTTGT 150
scrbc1-3  CAAACAGGCCGTGATCACTCATCATGGTGGTCCGGACCAGCCGTCCTTGT 142
scrbc1-1  AGAATGTAAGTGAACACAAACAAGTGAAGAGACCAGACCACTAGAGGCA 200
scrbc1-3  AGAATGTAAGTGAACACAAACAAGTGAAGAGAG.....GCAA 178
scrbc1-1  AGAGCAAGCAGACGCAGCCGAAGCCCAAGCCCAAACTGTTTGTCTTTG 250
scrbc1-3  AGAGCAAGCAGACGCAGCCGAAGCCCAAGCCCAAACTGTTTGTCTTTG 228
scrbc1-1  CCTTCCATTGTCGTTCTTAACTCATCCTTTCACAAACCACGAC..... 293
scrbc1-3  CCTTCCATTGTCGTTCTTAACTCATCCTTTCACAAACCACGACGAGCTGA 278
scrbc1-1  .....GAGCTGAATGAACTGCTTCTGC 315
scrbc1-3  ATTTGTCAAAAAAAAAAAACCCTGACTGAGCTGAATGAACTGCTTCTGC 328
scrbc1-1  TTGAGAGAGACGGCGGA..TCGGCGCTTACGGCAGAGCTTGTGCGCCG.A 362
scrbc1-3  TTGAGAGAGACGGCGGATGTGGCGGCTTACGGCAGAGCTTGTGCGCGCA 378
scrbc1-1  GATT.CGAGGAGGAA.GGGACGAGTAGAGGAAGGAGTCTTCACCGACCGA 410
scrbc1-3  GATTGCAGGGAGGAAGGGGACGAGTAGAGGAAGGAGTCTTCACCGACCGA 428
scrbc1-1  CGAGACCACGACGACCACCTCGGATCCTGGCTCGT.GACGACGAGCTAGC 459
scrbc1-3  GCAG..GACCACGACCACCTCGGATCCTGGCTCGTGGACGACAAGCTAGC 476
scrbc1-1  GAGCGAGTCCGGCCACGGCCGCGACTGTCCGGCTCTATACCTCGGTTG 509
scrbc1-3  GAGCGAGTCCGGCCACGGCCGCGACTGTCCGGCTCTATACCTCGGTTG 526
scrbc1-1  ATCGGCGACGGGAGACCGGAGATAGCGAGCGAGTCCGGCCACGGCCAGGG 559
scrbc1-3  ATCGGCGACGGGAGACCGGAGATAGCGAGCGAGTCCGGCCACGGCCAGGG 576
scrbc1-1  AAATTTTGAAACAGCTTGTGCGGCGCAGATTGCATGAGGAATGAGACGA 609
scrbc1-3  AAGTTTGGAAACAGCTTGTGCGGCGCAGATTGCATGAGGAA.GGGACGA 625

```

Figure 7. Comparison of the Nucleotide Sequences of *Scrbc1-1* and *Scrbc1-3* Gene. Putative translation start codons are in bold and underlined. Potential regulatory sequences are in bold.

```

scrbcs-1 GCAGAGGAGTCTTCCTTCACCGTCGGAGCAGGACCACCTGGGATCCTCGC 659
          |||
scrbcs-3 CTAGAGGAGTGT..CTTCACCG.ATGAGCAGGACCACCTGGGCTCCGTGC 672

scrbcs-1 TCGTGCAGAGCTATCGAGCGAGTCCGGCCACGCCACGGCCACGACGTCT 709
          |
scrbcs-3 AC....GAGCTAGCGAGCGAGTCC.GCCATGCCACGGCCACGACGTCT 716

scrbcs-1 GCTCAATATATACCTAGGTTTAATCGGAGATAGCGAGTGAGTCCGGGCAC 759
          |||
scrbcs-3 GCTCAATATATACCTAGGTTTAATCGGAGATAGCGAGTGAGTCCGGGCAC 766

scrbcs-1 GCCAGGAACTTTGGTTT.CAGTAGTGTGACTTCACCACT.TCTGCTTA 807
          |||
scrbcs-3 GCCAGGAACTTTGGTTTCCAGTAGT.GACTTCACCACTGCCTGCTTA 815

scrbcs-1 TCTATCTGTACCTCCTCTATCTATCTACAAAGTCCCCGGCACATCACAT 857
          |||
scrbcs-3 TCTATCTGTACCTCC...TCTATCTACAAAGT.CCTGGCACGTACAT 860

          G box-like I box
scrbcs-1 ATAGTCCAACCATGCCGTGGCTCCATGTGGCTGGCTCAGACGATAAGGCG 907
          |||
scrbcs-3 ATAGTCCAACCATGCCGTGGCTCCATGTGGCTGGCTCAGACGATAAGGCG 910

          G box-like GATA box
scrbcs-1 CGCCACGGGACGCGACATGTGGCGGGACGCGATCAGGATAGGCCAGG 957
          |||
scrbcs-3 CGCCACGGGTAAGCGACATGTGGCGGGACGCGATCAGGATAGGCCAGG 960

          GC-rich region monocot rbcs consensus
scrbcs-1 CTGGCCGGGTGCGGCCACGGGAGAACGGTGGCCACTCCTCCCACATCCGC 1007
          |||
scrbcs-3 CTGGCCGGGTGCGGCCACGGGAGAACGGTGGCCACTCCTCCCACATCCGC 1010

scrbcs-1 TTCGTCCTGCCCGAACGACAGCCATCCATCGCCATGGACGCACGCGGCT 1057
          |||
scrbcs-3 TTCGTCCTGCCCGAACGACAGCCATCCATCGCCATGGACGCACGCGGCT 1060

          TATA box
scrbcs-1 GCCTCTTCTATATATGCCGTGCGTGGGGGAGCCTTACAGGACGACCCAAG 1107
          |||
scrbcs-3 GCCTCTTCTATATATGCCGTGCGTGGGGGAGCC.TACAGGACGACCCAAG 1109

scrbcs-1 CAATAAGCAGCAAAGCAAGGCAGCAAGCTCCCTCAGCTCACTACTATTAC 1157
          |||
scrbcs-3 CAATAAGCAGCAAAGCAAGGCAGCAAGCTCCCTCAGCTCACTACTATTAC 1159

scrbcs-1 TAGCTAGTCCAGCCATG 1174
          |||
scrbcs-3 TAGCTAGTCCAGCCATG 1176

```

Figure 7. (Continued) Comparison of the Nucleotide Sequences of *Scrbc*s-1 and *Scrbc*s-3 Gene.

B. Copy Number Reconstruction

There is no existing data on the genome size of the sugarcane cv. H32-8560. Our best estimation comes from Bennett and Smith (1976), who reported that the *Saccharum* spp. hybrid cv. CP52-68, has a chromosome number of 116-117 and a genome size of 6.4pg DNA, or 5.8×10^9 bp. Price (1963) determined that the cv. H32-8560 has a chromosome number of 115-116. Therefore 5.8×10^9 bp was used as the genome size of the cv. H32-8560 for the current study.

If each genome contains 1 copy of the *rbcS* gene, $10 \mu\text{g}$ of genomic DNA (equal to 1.6×10^6 genomes) would contain 1.6×10^6 copies of the *rbcS* gene, which is equivalent to $1.6 \times 10^6 \times 2.3 \times 6.6 \times 10^5 / 6.02 / 10^{23} \text{g}$, or 4.0pg of the cloned *scrbcS-1* gene (2.3kb). Thus, the intensity of the hybridization signal of $10 \mu\text{g}$ of genomic DNA and 4.0pg *scrbcS-1* would be the same, assuming equal hybridization capability for both DNAs. Figure 8 shows the two major hybridization bands of 4.5kb and 2.3kb, representing about 4-6 and 8-10 copies, respectively, as compared with the control *scrbcS-1* DNA lanes. In addition, some minor bands at ca. 1.3kb and 2.0kb are visible in the blot.

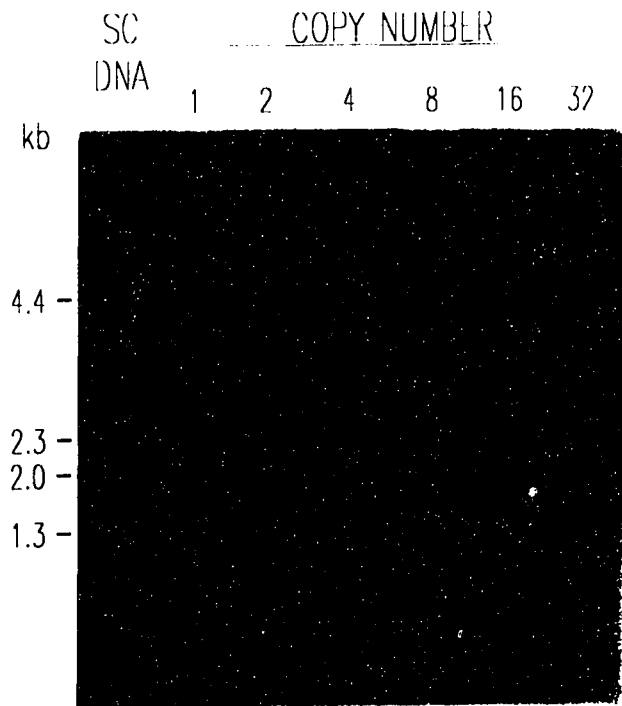


Figure 8. Copy Number Reconstruction Experiment. Lanes SC-DNA, 10 μ g *EcoRI-HindIII* digested sugarcane genomic DNA; 1, 2, 4, 8, 16, and 32, representing copy numbers of the *rbcS* gene in sugarcane, contain 4, 8, 16, 32, 64, and 128pg cloned *scrbcS-1* gene DNA, respectively.

C. Northern Hybridization

As shown in Figure 9, different sizes of RNA hybridized with the *scrbcS-1* gene probe. They are 1.20, 1.06, 0.96, 0.84, and 0.75kb, as calculated with the Nucleic Acid Fragment Sizing Program (Schaffer, 1981) according to the

positions of RNA Ladder (BRL). The bands of smaller sizes might be degradation products.

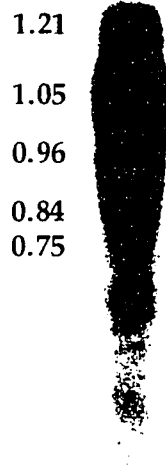


Figure 9. Mature Leaf Total RNA Hybridized with *Scrbcs-1* Gene Probe. Numbers shown are in kilobase (kb).

D. RACE Products and cDNA Library Screening

Nested-RACE products of RNAs from sugarcane leaf at different developmental stages showed, in general, similar electrophoretic patterns (Figure 10). Discrete lengths of products, at 580bp, 430bp, 330bp, 260bp and 190bp, are produced by all the samples studied, including leaf sheath, green callus, greening leaf, young leaf and mature leaf. The relative abundance of these products is generally

similar among the samples; however, there are some exceptions. For example, the 430bp product is most abundant in greening leaf (GL), while the 330bp product is least abundant in the green callus (GC) and leaf sheath (SH).

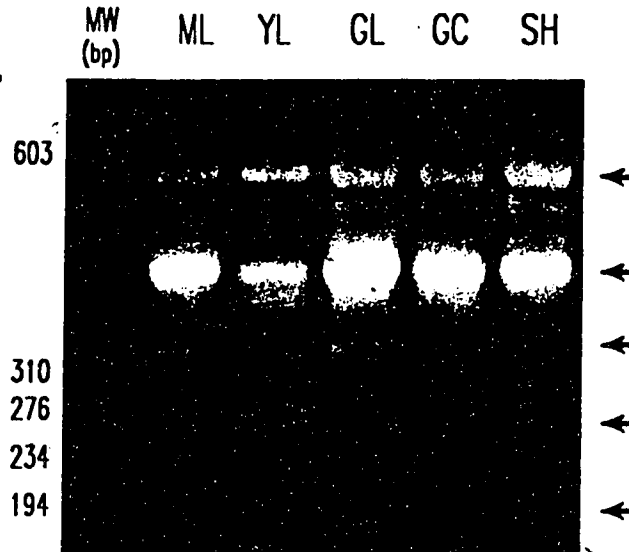


Figure 10. Nested RACE Products of RNAs from Sugarcane Leaves at Different Developmental Stages. MW, molecular weight marker; SH, leaf sheath; GC, green callus; GL, greening leaf; YL, young leaf; ML, mature leaf.

Thirty RACE products and 10 cDNAs from various leaf samples were cloned and sequenced. There are no significant differences in the RACE sequences from the different samples, nor in the cDNAs from RACE sequences (data not shown). All have very high homology (more than 90%) with the cloned *scrbc-1* gene. However, they can be divided into 5 major groups based on their lengths, i.e. Group I (580bp), II (430bp), III (330bp), IV (260bp) and V (190bp). Figure 11 shows some example sequences of different groups compared to the *scrbc-1* gene sequence.

1790

| | | | | | |
|-----------------|------------|------------|------------|-------------------|------------|
| <i>scrbc1-1</i> | TTCATCGCCT | ACAAGCCCGC | GGGCAGCGAG | <u>TAAAGTGCTA</u> | GCTAGCTCCT |
| <i>race-1-1</i> | TTCATCGCCT | ACAAGCCCGC | GGGCAGCGAG | <u>TAAAGTGCTA</u> | GCTAGCTCCT |
| <i>race-2-1</i> | TTCATCGCCT | ACAAGCCCGA | GGGCAGCGAG | <u>TAAAGTGCCA</u> | GCTAGCTCCT |
| <i>race-3-1</i> | TTCATCGCCT | ACAAGCCCGA | GGGCAGCGAG | <u>TAAAGTGCTA</u> | GCTAGCTCCT |
| <i>race-4-1</i> | TTCATCGCCT | ACAAGCCCGC | GGGCAGCGAG | <u>TAAAGTGCTA</u> | GCTAGCTCCT |
| <i>race-5-1</i> | TTCATCGCCT | ACAAGCCCGC | GGGCAGCGAG | <u>TAAAGTGCTA</u> | GCTAGCTCCT |

* * *

1839

| | | | | | |
|-----------------|------------|------------|------------|-------------------|------------|
| <i>scrbc1-1</i> | GCGTGAGCTG | CTAATGGCAT | GCCTGCCTGC | <u>CTCTAC.TGT</u> | TCGGTTTTGC |
| <i>race-1-1</i> | GCGTGAGCTG | CTAATGGCAT | GCCTGCCTGC | <u>CTCTACCTGT</u> | TCGGTTTTGC |
| <i>race-2-1</i> | GCGTGAGCTG | CTAATGGCAT | GCCTGCCTGC | <u>CTCTACCTGT</u> | TCGGTTTTGC |
| <i>race-3-1</i> | GCGTGAGCTG | CTAATGGCAT | GCCTGCCTGC | <u>CTCTACCTGT</u> | TCGGTTTTGC |
| <i>race-4-1</i> | GCGTGAGCTA |AT | GCATGCCTGC | <u>CTCTACCTGT</u> | TCGG.TTTGC |
| <i>race-5-1</i> | GCGTGAGCTG | CTAATGGCAT | GCCTGCCTGC | <u>CTCTACCTGT</u> | TCGGTTTTGC |

* * * * *

1888

| | | | | | |
|-----------------|------------|------------|------------|-------------------|------------|
| <i>scrbc1-1</i> | ATCCTTCGGG | TCACACACCT | T.GTTTCCTT | <u>TTCACCTTTT</u> | TCTTCTTATT |
| <i>race-1-1</i> | ATCCTTCGGG | TCACACACCT | T.GTTTCCTT | <u>TTCACCTTTT</u> | TCTCCTGGCT |
| <i>race-2-1</i> | ATCCTTCGGG | TCACACACCT | TGGTTTCTTT | <u>TTCACCTTTT</u> | TCTCCTGGTT |
| <i>race-3-1</i> | ATCCTTCGGG | TCACACACCT | TGGTTTCTTT | <u>TTCACCTTTT</u> | TCTCCTGGTT |
| <i>race-4-1</i> | AT.CTTCGGG | TCACACACCT | TGGTTTCTTT | <u>TTCACCTTTT</u> | TCTTCTTATT |
| <i>race-5-1</i> | ATCCTTCGGG | TCACGCACCT | T.GTTTCCTT | <u>TTCACCTTTT</u> | TCTTCTTATT |

* * * * *

1938

| | | | | | |
|-----------------|------------|------------|------------|-------------------|------------|
| <i>scrbc1-1</i> | CCTTTGCTTC | TCTGCTGCTA | ATGTATCCAT | <u>TGTTGCAAGC</u> | ATGGCATCAT |
| <i>race-1-1</i> | CCTTTGCTTC | TCTACTGCTG | ATGTATCCAT | <u>TGTTGCAAGC</u> | ATGGCATCAT |
| <i>race-2-1</i> | CCTTTGCTTC | TCTGCTGCTG | ATGTAT.CAT | <u>TGTTGCAAGC</u> | ATGGCATCAT |
| <i>race-3-1</i> | CCTTTGCTTC | TCTGCTGCTG | ATGTATCCAT | <u>TGTTGCAAGC</u> | ATGGCATCAT |
| <i>race-4-1</i> | CCTTTGCTCC | TCTGCTGCTG | ATGTATCCAT | <u>TGTTGCAAGC</u> | ATGGCATCAT |
| <i>race-5-1</i> | CCTTTGCTTC | TCTGCTGCTG | ATGTAAAAAA | AAAAAAAAAA | A |

* * * * *

1988

| | | | | | |
|-----------------|------------|------------|------------|-------------------|------------|
| <i>scrbc1-1</i> | CCATCTCTCC | CCCTACATAT | ACCAGCTAGA | <u>AACTACTGCA</u> | ACTAGCGTTG |
| <i>race-1-1</i> | CCATCTCTCC | CCCTACATAT | ACCTGCTACA | <u>AACTACTGCA</u> | ACTAGCGTTG |
| <i>race-2-1</i> | CCATCTCTCC | CCCTACATAT | ACCTGCTACA | <u>AACTACTGCA</u> | ACTAGCGTTG |
| <i>race-3-1</i> | CCATCTCTCC | CCCTACATAT | ACCTGCTAGA | <u>AACTACTGCA</u> | ACTAGCGTTG |
| <i>race-4-1</i> | CCATCTCTCC | CCCTAAAAAA | AAAAAAAAAA | AA | |
| <i>race-5-1</i> | | | | | |

* * *

Figure 11. Example Sequences of 3' UTR of the RACE Clones as Compared with the *Scrbc1-1* Gene. Dots represent missing nucleotides, while stars represent mismatches. The putative translation stop codon and AAUAAA-like motifs are underlined in the *scrbc1-1* gene. (Continued pp 84-85)

2038
scrbc1-1 GGTGAGGAAC ATGTGAATGC AAGCTCCGGC TATCATATAT GTGTAATATA
race-1-1 GGTGAGGAAC ATGTGAATGC AAGCTCCGGC TATCATATAT GTGTAATATA
race-2-1 GGTGAGGAAC ATGTGAATGC AAGCTCCGGC TATCATATAT GTGTAATATA
race-3-1 GGTGAGGAAC ATGTGAATGC AAGCTCCGGC TATCAAAAAA AAAAAAAAAA
race-4-1
race-5-1

2088
scrbc1-1 TGGCATTGGA GTTTGCTCTG GTTATTCACA GAGCAAGGAA TTACTTTTCCT
race-1-1 TGGCATTGGA GTTTGCTCTG GTTATTCACA GAGCAAGGAA GTACTTTTCCT
race-2-1 TGGCATTGGA GTTTGCTCTG GTTATTCACA GAGCAAGGAA GTACTTTTCCT
race-3-1 AA
race-4-1
race-5-1

* *

2137
scrbc1-1 TCCATATATA TATACAACAT ACTTTAATTT CATTCCATGC AACAAT.GCT
race-1-1 TCC...ATA TATACAACAT ACTTTAATTT CATTCCGTC AACAATGGCT
race-2-1 TCC...ATA TATACAACAT ACTTTAATTT CAAAAAAAAA AAAAAAAAAA
race-3-1
race-4-1
race-5-1

*

*

2187
scrbc1-1 AGAAGCCCTG TTGTAATTTT GCTTATTTAC TTTTTTCCCC CTTTTTTTGT
race-1-1 AGAAGACCTG TTGTAATTTT GCTTATTTAC TTTTTTTTTC CTTTTTTTGT
race-2-1 AAAAAAAAA
race-3-1
race-4-1
race-5-1

*

2237
scrbc1-1 CTCTTGAAC TTTCTGTAAG CTGGTTCTCC TTTCTTAATA AATAAATGCA
race-1-1 CTCTTGAAC TTTCTGTAAG CTGGTTCTCC TTTCTTAATA AATAAATGCA
race-3-1
race-4-1
race-5-1

2286
scrbc1-1 CAGTAGGGGG .AAATCCCTC CTGTTTCCAT CAAAAAGAA TCTCTGAACT
race-1-1 CAGTAGGGGG GAAATCCCTC CTGTTTCCAT CAAAAAAAAA AAAAAA
race-3-1
race-4-1
race-5-1

*

Figure 11. (Continued) Example Sequences of 3' UTR of the RACE Clones as Compared with the *Scrbc1*-1 Gene.

Since the RACE products have RACE-adapters and part of the polyA tail, the sizes of different groups in Table 4 should be adjusted to match the sizes in Figure 11 by adding 76 nucleotides (37 for RACE-adapter and polyT-tract and 39 for the coding region).

Table 4. Number of RACE Products and cDNA Clones within the Different RNA Groups.

| GROUP | POLYADENYLATION SITE FROM TRANSLATION STOP CODON | NUMBER OF CLONES |
|-------|---|------------------|
| I | 501 | 5 |
| II | 351 | 20 |
| III | 254 | 2 |
| IV | 184 | 3 |
| V | 103 | 2 |

E. S1 Mapping

Bands of 480bp, 370bp, 320bp, 300bp, 285bp, and 232bp, were identified in the S1 map, using the labelled Phi X74 HaeIII-digested DNA fragments as standard sizes (Figure 12 & 13).

The sizes of the protected sequences correlate with the sizes of the RACE products as well as the bands of Northern

analysis, confirming the heterogenous nature of the 3' untranslated region of the *scrbc*s gene.

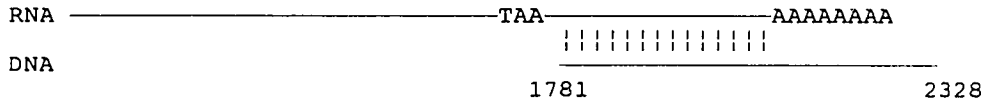


Figure 12. Schematic of S1 Mapping.

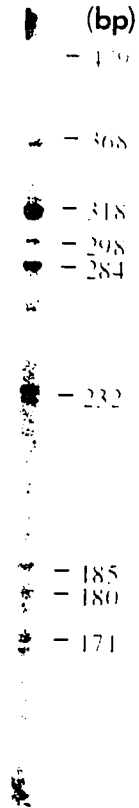


Figure 13. S1 Mapping of the 3' UTR of the *Scrbcs-1* Gene Transcripts. The numbers shown are in bp.

F. Differential Amplification of Total Sugarcane cDNA

As illustrated in Figure 14, in the amplification of total cDNA from a sugarcane leaf sample, primer pair *rbcS#1-* and *rbcS#5+* could amplify only Group I mRNA, while the other pair, *rbcS#2-* and *rbcS#5+*, could amplify both Groups I & II mRNAs. Thus, the ratio of these two PCR products reflects the ratio of Groups I and II mRNAs within the sample.

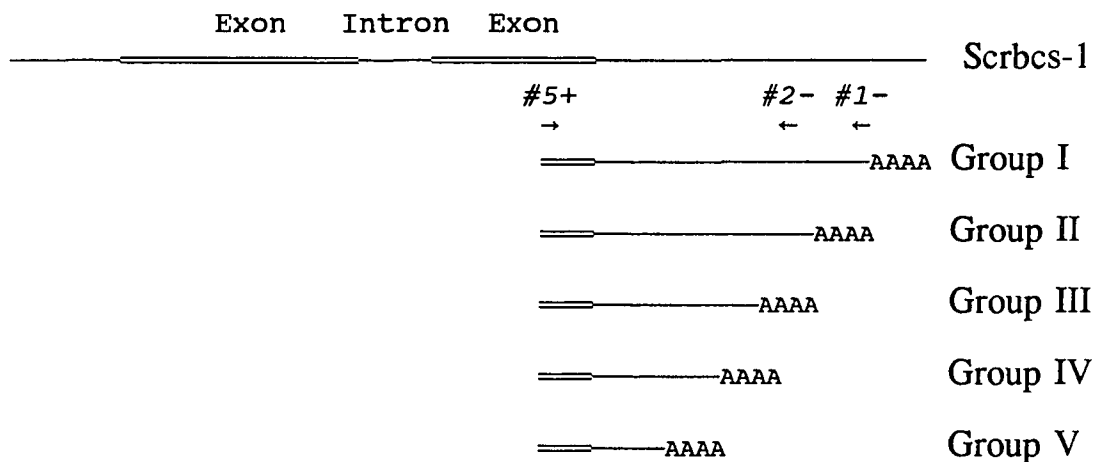


Figure 14. Schematic of Differential Amplification of Total Sugarcane cDNA. The three arrows, from left to right, represent *rbcS#5+*, *rbcS#2-* and *rbcS#1-*, respectively. "AAAA" represents polyA tail of mRNA.

Figure 15 shows the differential amplification products from different sugarcane samples. Although the levels of

mRNA products differ slightly, Groups I and II are expressed throughout the development of the sugarcane leaf, and the ratio of these two groups of mRNAs shows no distinct change, except perhaps in the leaf sheath.

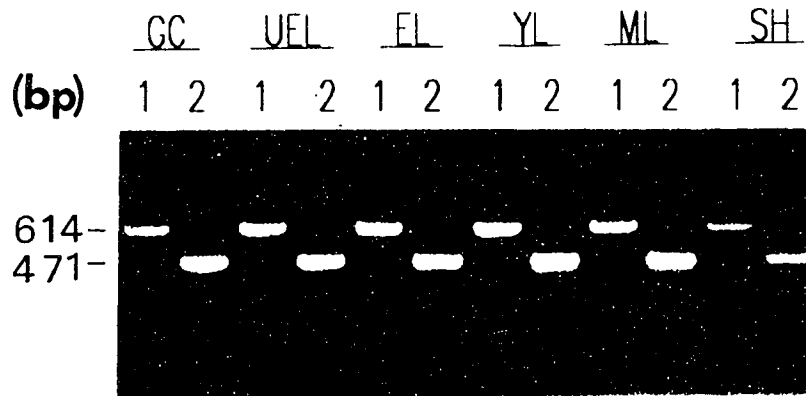


Figure 15. Differential PCR Amplification Products of cDNAs from Different Sugarcane Samples. MW, molecular weight marker; 1, product resulted from *rbcS#1-* and *rbcS#5+* amplification; 2, product resulted from *rbcS#2-* and *rbcS#5+* amplification; GC, green callus; UEL, unexpanded leaf; EL, expanding leaf; YL, young leaf; ML, mature leaf; SH, leaf sheath.

G. Differential Amplification of Total Sugarcane Genomic DNA

If Groups I and II mRNAs are encoded by genes of different lengths, primers *rbcS#1-* and *rbcS#5+* would amplify only the Group I genes when they are used to amplify total genomic DNA from sugarcane (Figure 16). Primers *rbcS#2-* and

rbcS#5+ would amplify both Groups I & II genes. The ratio of these two amplification products would reflect the ratio of Groups I and II genes in sugarcane.

If both Groups I and II mRNAs are encoded by genes of similar lengths, these two pairs of primers will amplify all the *rbcS* genes in the genome, thus the ratio of these two amplification products will be the same as using the cloned *scrbcS* gene as the template.

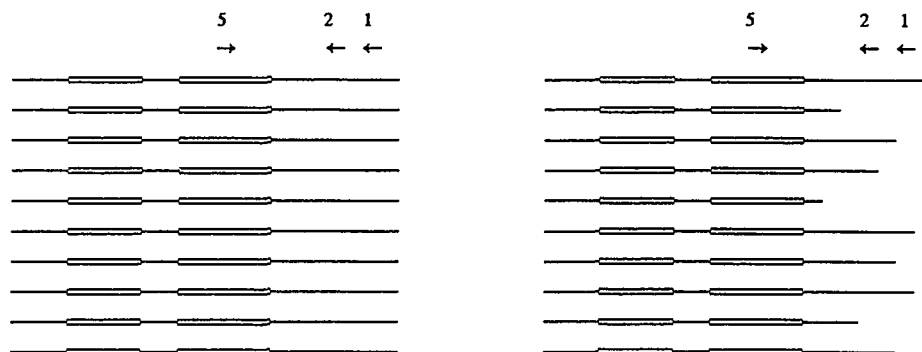


Figure 16. Schematic of Differential Amplification of Total Sugarcane Genomic DNA. The four arrows, from left to right in each panel, represent *rbcS#5+*, *rbcS#2-* and *rbcS#1-*, respectively. Lines represent *rbcS* genes in genome. Left scheme, genes with similar lengths; right, with different lengths.

Since the product ratios have no significant difference between the genomic DNA and *scrbcS* gene (Figure 17), This suggests that Groups I and II mRNAs are encoded by genes of similar lengths.

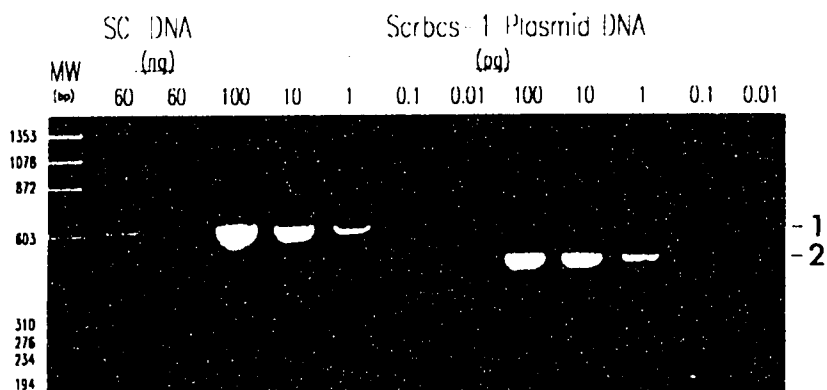


Figure 17. Differential PCR Amplification Products Using Genomic DNA and *Scrbcs* Gene as Templates. MW, molecular weight marker. The product resulting from *rbcS#1-* and *rbcS#5+* amplification is 614bp; product resulting from *rbcS#2-* and *rbcS#5+* amplification is 471bp.

H. Sugarcane Leaf Bombardment

Various conditions were tested to optimize methods for bombardment of sugarcane leaf with DNA-coated particles.

(1) Leaf Competency. For greenhouse-grown sugarcane plants, the green laminae, e.g. leaf 1, were found to have about 10 times fewer GUS foci than the younger, yellow laminae, e.g. leaf -3. The yellow leaves are unexpanded leaves wrapped within the leaf whorl. However, after

culturing for 3 days under light, the number of GUS foci in all leaves increased about ten times.

(2) Pressure. The number of GUS foci increased with increasing helium pressure (from 450psi to 2200psi), reduced distance between the rupture disc and the macrocarrier, and reduced distance between the rupture disc and the leaf sample. At the gunpowder setting, the force was so high that holes were torn in the leaf samples by bombardment. The number of GUS foci was usually about ten times lower than when helium power was used.

(3) pUbiGUS construct produced about 10 times more GUS foci than pWD1. The three constructs, pWD1, pWD3 and pHT1011, gave about the same number of foci. No GUS foci were found after bombardment with pWD2 (Table 5).

Table 5. Number of GUS Foci per Shot under 2000psi Helium Pressure. The distance between the rupture disc and the macrocarrier was 2cm, and the petri dish was placed at position "3" (4cm from the macrocarrier).

| Construct | Cultured for 0 day | | Cultured for 3 days | |
|-----------|--------------------|---------|---------------------|---------|
| | -3 Leaf | +1 Leaf | -3 Leaf | +1 Leaf |
| pUbiGUS | 32±8 | 4±1 | 342±111 | 31±12 |
| pWD1 | 5±2 | 0 | 45±17 | 3±1 |
| pHT1011 | 4±2 | 0 | 52±15 | 5±3 |
| pWD3 | 6±3 | 0 | 57±23 | 2±2 |
| pWD2 | 0 | 0 | 0 | 0 |

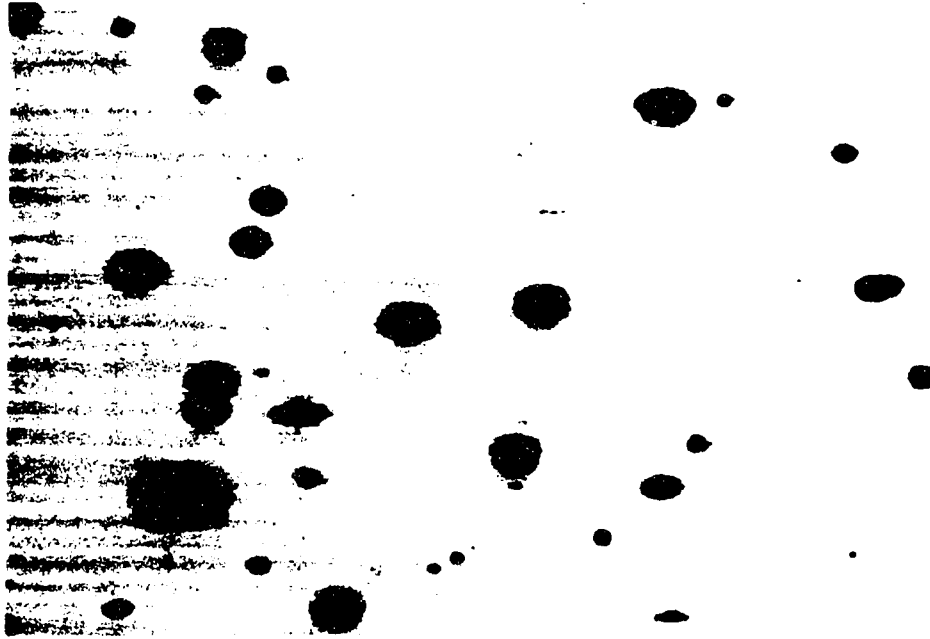


Figure 18. GUS-Stained Sugarcane Leaf after Bombardment with pUbiGUS. The -3 leaf pieces were cultured for three days on MS media and bombarded at 2000psi Helium pressure. The distance between the rupture disc and the macrocarrier was 2cm, and the petri dish containing the leaf was placed at position "3" (4cm from the macrocarrier). Magnification: 160x.

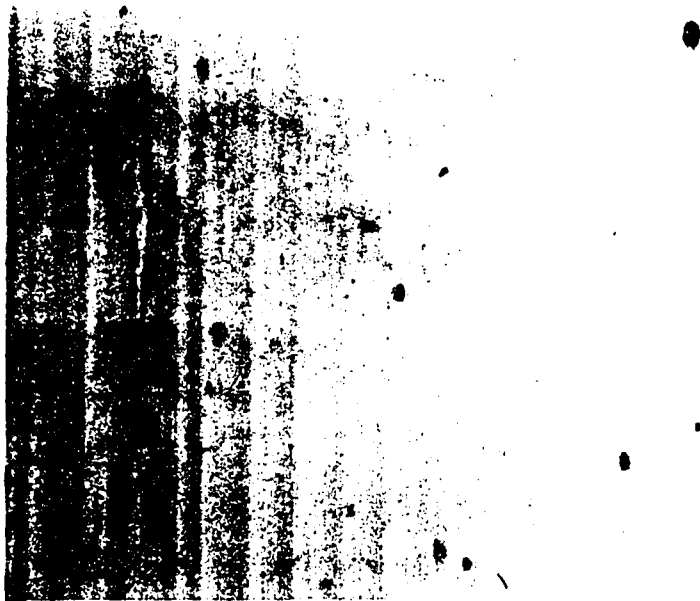


Figure 19. GUS-Stained Sugarcane Leaf after Bombardment with pWD1. Bombardment conditions were the same as in Figure 17.



Figure 20. GUS-Stained Sugarcane Leaf after Bombardment with pWD3. Bombardment conditions were the same as in Figure 17.

In view of these findings, the optimal conditions for bombarding the yellow leaves in subsequent experiments were 2000psi, a distance of 2cm between the rupture disc and the macrocarrier, petri dish at position "3" (4cm from the macrocarrier), and leaves cultured for three days.

After staining for GUS activity, the leaves with blue spots were sectioned to locate the GUS expression at the cellular level. For all plasmid constructs (with either ubiquitin promoter or *scrbc*s promoter) most of the blue spots were in the epidermic cells (EPs) of the sugarcane leaves. Only a few spots were located in the mesophyll and bundle sheath cells (Table 6 & Figure 21-23).

Table 6. Total Number of GUS Foci Located in the Sugarcane Leaf Cells.

| | Total | BS | MS | EP |
|---------|-------|----|----|-----|
| pUbigUS | 771 | 17 | 23 | 731 |
| pWD1 | 178 | 42 | 16 | 118 |
| pWD3 | 339 | 74 | 22 | 243 |

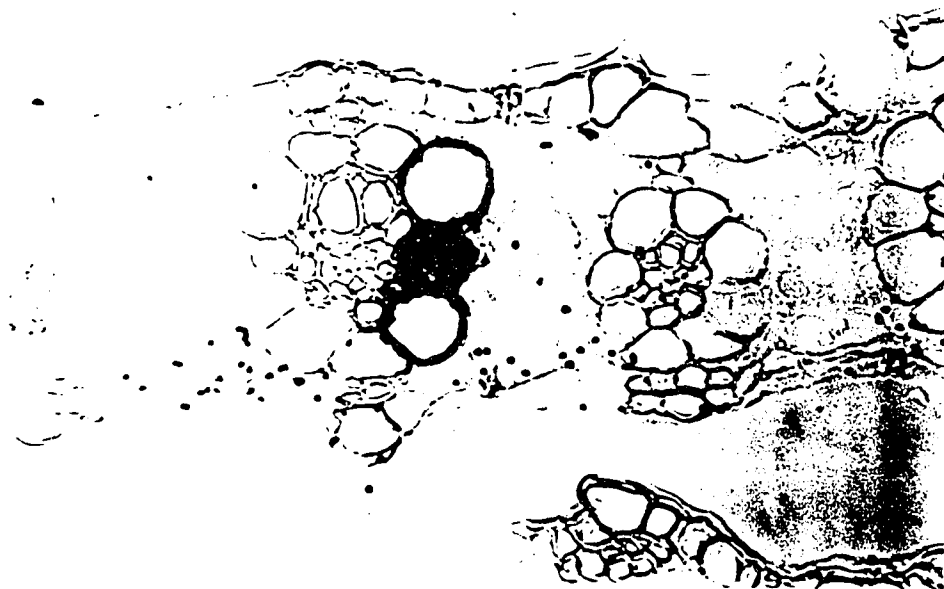


Figure 21. Transient Expression of GUS Gene in Sugarcane Leaf BS Cells after Bombardment with pWD1. (600x)

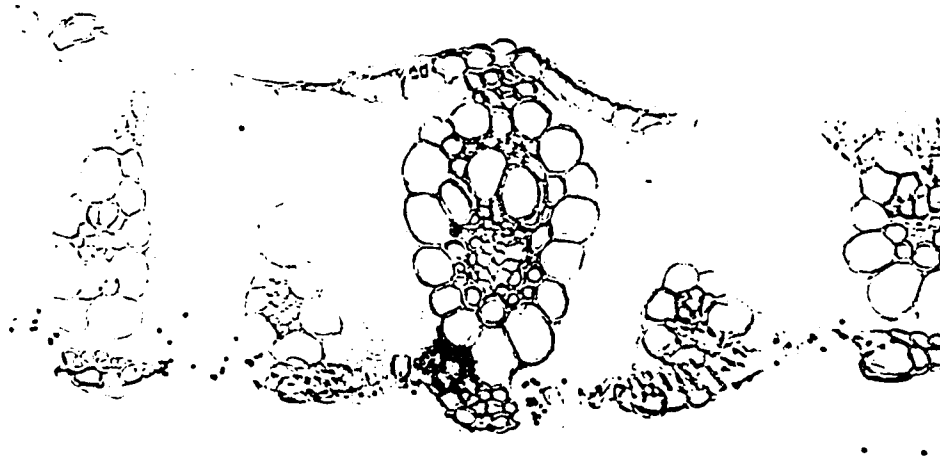


Figure 22. Transient Expression of GUS Gene in Sugarcane Leaf MS Cells after Bombardment with pWD1. (600x)

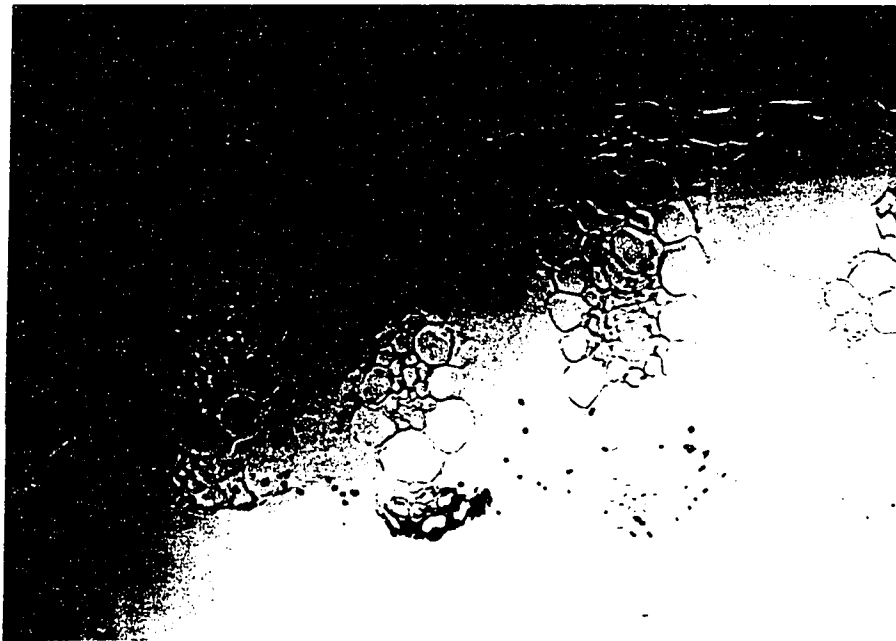


Figure 23. Transient Expression of GUS Gene in Sugarcane Leaf Epidermis after Bombardment with pWD1. (600x)

I. PEP Carboxylase Activity

PEP Carboxylase (PEPC) activity was detected in leaf samples of different developmental stages (Figure 24). The "housekeeping" type ("fast" migrating band, see Albert, 1991) was found in all the four leaf samples. The C4 type ("slow" band) was not found in UEL-T and UEL-B, but the activity appeared in UEL-L. The mature leaf (ML) had the highest activity of the C4-type PEPC.

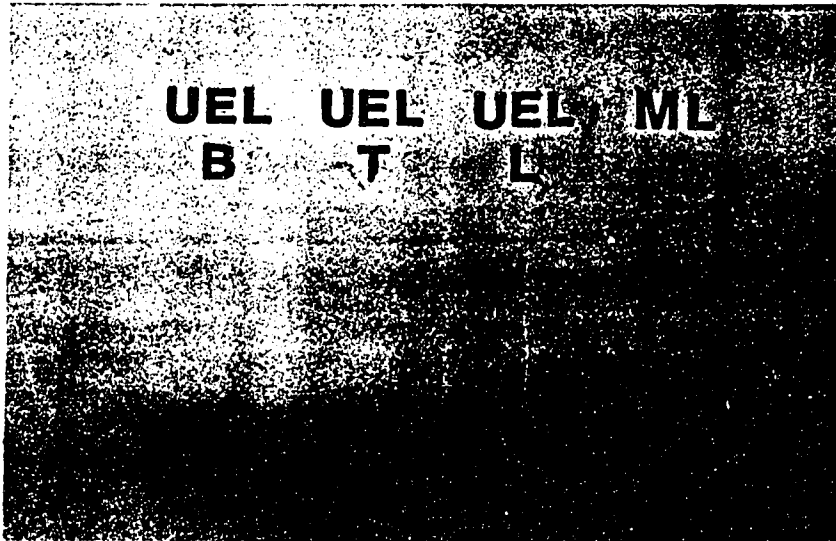


Figure 24. PEP Carboxylase Activity of Sugarcane Leaf of Different Developmental Stages. UEL, unexpanded leaf. T, top. B, base. L, cultured for 3 days under light. ML, mature leaf.

CHAPTER V. DISCUSSION

A. The Highly Conserved *Scrbcs* Genes and Their mRNAs

The results presents here suggest that the *rbc*s genes of sugarcane are extremely conserved, both within the coding region and in the 5' and 3' UTRs. First, all cDNAs, RACE products, and genes so far cloned for the *scrbcs* share more than 95% homology within the coding regions as well as in the 5' and 3' UTRs (Figures 5, 6, 7 & 11). Second, all genes and cDNAs so far cloned were randomly picked from the genomic or cDNA library. Seven *scrbcs* genes and nine cDNAs were identified based on their ability to hybridize with the *scrbcs* probe. It is unlikely that we failed to clone an existing family member with a different 3' UTR. Finally, the two *rbc*s-specific primers (*rbc*s#3+ and *rbc*s#5+) used in nested-RACE were synthesized based on the highly conserved regions of all the available *rbc*s gene sequences in the databank, and all 9 *scrbcs* cDNAs and 6 *scrbcs* genes cloned contained these two sequences. It is unlikely that some *scrbcs* mRNAs (perhaps with different 3' UTRs) were not cloned (because of low homology to these two primer sequences). Nevertheless, we are aware of the fact that,

among the *EcoRI-HindIII*-digested genomic DNA fragments, two major bands (2.3kb and 4.5kb) hybridized with the *scrbc1* gene probe (Figure 8). However, none of the genes we cloned had a 4.5kb size fragment. This might be the result of incomplete digestion of genomic DNA. Further studies are needed to draw a conclusive statement on such a hybridization pattern.

The high conservation of the *rbcs* gene sequences, particularly in the 3' UTRs, is apparently unique to sugarcane. In all plant species studied thus far, that contain more than one member of the *rbcs* gene family, the 3' UTR of at least one member differs from those of the others, and the 3' UTRs could be used as probes to distinguish different members of the *rbcs* gene family (Vallejos *et al.*, 1986; Dean *et al.*, 1989; Schaffner and Sheen, 1991; Galili *et al.*, 1992).

B. Nucleotide Differences among *Scrbc1* Genes

Although all *scrbc1* genes are highly conserved, there are differences in the sequences. Figures 5, 6 & 7 show the nucleotide differences in the 5' and 3' UTRs of different *scrbc1* genes. Figure 11 shows the differences in nucleotide sequences of the RACE products.

RACE products were produced from *scrbc*s mRNAs *in vitro*, that in turn were transcribed from the *scrbc*s genes *in vivo*. The nucleotide differences in the observed RACE products might be the result of: (1) transcriptional and posttranscriptional modifications of the *scrbc*s mRNAs *in vivo*, (2) errors in PCR amplification, (3) errors during reverse transcription *in vitro*, and (4) nucleotide differences among the *scrbc*s genes.

Except capping, tailing and intron removing, there have been no reports on transcriptional and post-transcriptional modifications of nuclear gene-encoded mRNA nucleotide sequences *in vivo* in eukaryotic cells. Even if mechanisms such as RNA editing were to occur in the nucleus, it is unlikely to play a significant role in mRNA processing.

Although PCR is known to produce errors (Saiki *et al.*, 1988; Keohevong and Thilly, 1989; Ennis *et al.*, 1990), the nucleotide differences shown in Figure 11 appear to be not merely PCR artifacts. The rate of nucleotide variations among the RACE products (more than 3%) is much higher than that of the reported PCR errors (2.1×10^{-4} errors base⁻¹ cycle⁻¹, Saiki *et al.*, 1988; Keohevong and Thilly, 1989). The PCR conditions we used are similar to the "high fidelity" conditions (personal communication with Perkin-Elmer technician). Thus, PCR might account for a few

nucleotide differences among the RACE products, but is probably not the major cause.

Reverse transcriptase makes more nucleotide misincorporations during *in vitro* reverse transcription than the Taq polymerase in PCR (personal communication with Perkin-Elmer technician). Sequence comparison of the 3' UTRs of the *rbcs* genes and the RACE products in Figure 6 & 11 shows that the RACE (3' UTR) products contain slightly more nucleotide differences than the 3' UTRs in the *scrbc*s genes. Since there is no difference between the cDNAs and the RACE products (data not shown), it is possible that some of these differences were generated by the M-MLV reverse transcriptase.

Therefore, we conclude that most of the nucleotide differences in the RACE products are the result of nucleotide differences in the *scrbc*s genes.

C. Length Heterogeneity of the 3' UTRs in *Scrbcs* mRNAs

The *scrbc*s mRNAs are length-heterogeneous in their 3' UTR regions, e.g. they may differ in the polyadenylation sites. Based on the lengths of their 3' UTRs, the cDNAs and RACE products so far cloned and sequenced can be grouped into

five major groups. The cDNAs and RACE products within each group have exactly the same polyadenylation sites. The lengths between translation stop site and the polyadenylation sites for Group I, II, III, IV and V are 501, 351, 254, 184, and 103 nucleotides, respectively (Figure 11).

Since the sequences of *scrbc*s genes are closest to those of the maize *rbcs* genes, and the two are almost identical in nucleotide sequence around the transcription start sites (Figure 4, Schaffner and Sheen, 1991), we made the assumption that they have the same transcription start sites. Thus the putative transcription start site of the *scrbc*s-1 gene should be around -77 region. Considering the 507 nucleotides of the coding region plus the nucleotide sequences at the cap site, the lengths of the *scrbc*s mRNAs (exclude polyA tails) represented by the 5 major groups of RACE products would be as indicated in Table 7. Assuming that the polyA tails of *scrbc*s mRNAs are about 100 nucleotides in length, the above deduced lengths for the *scrbc*s mRNAs match with the sizes of the mRNA bands seen on Northern blot (Figure 9).

Table 7. Length of the RACE Products. Numbers are in nucleotides.

| GROUP | POLYADENYLATION SITE from translation stop codon | LENGTH OF mRNA (excluding polyA) |
|-------|---|-------------------------------------|
| I | 501 | 1085 |
| II | 351 | 935 |
| III | 254 | 834 |
| IV | 184 | 768 |
| V | 103 | 687 |

The S1 mapping experiment further documents the heterogeneous nature of the *scrbcs* mRNAs 3' UTRs. If as above, we assume that the *scrbcs* mRNA 5' UTRs are 77 nucleotide in length and the fragment of the *scrbcs* gene used to protect the *scrbcs* mRNAs was from #1786 to #2328 (see Figure 17), then the lengths of the protected *scrbcs* mRNAs (polyA not included) can be calculated by adding $77+507+11=595$ nucleotides to the protected length (Table 8).

Table 8. Length of the S1 Mapped mRNAs.

| PROTECTED LENGTH | LENGTH OF CORRESPONDING mRNA (excluding polyA) |
|------------------|---|
| 479 | 1074 |
| 368 | 963 |
| 318 | 913 |
| 298 | 893 |
| 284 | 879 |
| 232 | 827 |
| 185 | 780 |
| 180 | 775 |
| 170 | 765 |

The strongest bands at 318 and 284bp in S1 mapping correlate with the Group II RACE products, that in turn are the amplified products of the most abundant mRNAs in northern hybridization (1.06kb). The weaker band, 479bp, represents Group I RACE products as well as the band with the highest molecular weight in the northern hybridization (1.20kb).

Nevertheless, there were differences in lengths calculated from the S1 mapping, RACE, and northern hybridization experiments (Table 9).

Table 9. Summary of the Predicted Length of *Scrbcs* mRNAs. The length of polyA tails is not included in the RACE and S1 Mapping prediction. All numbers are in nucleotide.

| Group | RACE Prediction | Northern Bands | S1 Mapping Prediction |
|-------|-----------------|----------------|-----------------------|
| I | 1085 | 1200 | 1074 |
| II | 935 | 1060 | 963,913,893,879 |
| III | 834 | 960 | 827 |
| IV | 768 | 840 | 780,775,765 |
| V | 687 | 750 | |

In northern analysis, the hybridized bands were wide and indistinct; the darkest northern band apparently splits into two strongest bands in the S1 mapping assay. Possible reasons for this are fourfold. First, mRNAs may have been degraded by exo- or endo-nuclease. Second, the lengths of

the 5' UTRs and polyA tails may be different among the messengers. Third, the reverse transcriptase might be more efficient in some regions of the mRNAs than others due to secondary structures. Fourth, the protection may not be efficient in some sequence regions, e.g. lower binding affinity or mis-matches would result in susceptibility to S1 nuclease digestion.

D. Possible Mechanisms of Length Heterogeneity in the 3' UTRs

Among RACE products and cDNA clones, some are identical in nucleotide sequence, while others differ only in the length of the 3' UTRs (data not shown). The latter may be the transcribed products from different genes, or the same gene with different polyadenylation sites. Two pieces of evidence suggest that these messengers of different lengths are produced by the same genes. First, there are only about 12-16 *scrbc*s genes in the whole sugarcane genome (Figure 8), but more than thirty different RACE products and cDNAs have been cloned. This is best explained if a single gene produced messengers of different lengths. However, we can not rule out the possibility that we have underestimated the *scrbc*s gene number, or that errors produced by the M-MLV and Taq polymerase reactions might contribute to the variations among the messengers (see discussion in previous section).

Second, when we differentially amplified the sugarcane genomic DNA with Group I- and II-specific primers, the results (Figure 17) showed that both groups of mRNA are encoded by genes of similar lengths.

Maturation of the 3' UTR of mRNA from hnRNA involves three steps: transcription termination, endonuclease cleavage and polyadenylation. The length of the 3' UTR in a mRNA depends on the site of polyadenylation. Differential polyadenylation sites might be generated by two mechanisms. A number of homologous hnRNAs which have similar transcription termination sites may be cleaved at different sites. Alternatively, transcription may terminate at different sites due to the presence of several weak termination sites along the downstream sequence from the coding region, and the sites for cleavage and polyadenylation mainly depend on the location of these transcription termination sites. At present, it is not clear which mechanism is actually used in the cell. Further experiments, including mutational alternation of 3' end, nuclear run-off assay and isolation of hnRNA, may lead to a clearer picture on whether transcription termination could affect the cleavage and polyadenylation of an mRNA.

In all RACE products and cDNA clones, the first A residue in the mRNA polyA tail corresponds to an A residue in the

genomic sequence, and the nucleotide immediately upstream of the polyA is always C or U (C in Group I, II and V; U in Group III and IV). Such phenomena have been found in animal cells (Birnstiel *et al.*, 1985) as well as plant cells (Mogen *et al.*, 1992), although there are some exceptions in petunia *rbcS* genes (Dean *et al.*, 1986). Although as yet unclear, the conservation of these CA or UA sequences suggests that they may be important in mRNA cleavage.

In sugarcane *rbcS* genes, putative AAUAAA motifs could be identified for Groups I and II mRNAs, but not for the other groups (Figure 11). Since Groups I and II are the most abundant groups, the AAUAAA sequence may be helpful, but not required, in cleavage and polyA addition. We can neither identify the G/U cluster, nor the CAYUG motif, using the method described by Dean *et al.* (1986). We also fail to detect the far-upstream element (a series of GU motifs, especially the UUGUA motif) described by Mogen *et al.* (1992). It seems that different *rbcS* genes might require different sequence motifs in their 3' regions, and/or that the higher-dimensional structures instead of the primary structures of the 3' regions may be important in directing polyadenylation. Functional analysis of the 3' UTR of mRNA, using transgenic sugarcane containing modified *scrbcS* 3' UTRs, might be able to provide a clearer picture of the mechanisms in mRNA cleavage and polyA addition.

E. Multiple Polyadenylation Sites in Developing Sugarcane Leaf

Length heterogeneity of the 3' UTRs has been reported in many plant mRNAs. In most cases, the length differences are less than 100 nucleotides. In sugarcane, the 3' UTRs of the *scrbc*s mRNAs could differ by up to 400 nucleotides.

Studies have shown that the 3' UTRs might be important in mRNA stability, transportation, localization, and translatability. It is of great interest to determine why multiple polyadenylation sites have been preserved during evolution. In animal cells, multiple sites polyadenylation are associated with differential gene expression (Amara et al., 1984; Birnstiel et al., 1985; Saez et al., 1990). Differential expression of mRNAs with different polyadenylation sites has not been studied in plants. As a first attempt, we studied the expression of *rbc*s mRNAs with different lengths of 3' UTRs during development of sugarcane leaf, to test whether differential polyadenylation of mRNAs is involved in the regulation of C3-C4 photosynthesis transition.

In our experiment, we used quantitative PCR, and have the following reasons to show that this method is reliable.

Recent technical developments have greatly enhanced the reliability of quantitative PCR (see Quantitative PCR, *Methods & Applications* Book 3, CLONTECH Laboratories, Inc.). Successful quantitation depends on the use of internal or external controls, reaction conditions, as well as the amount of template. In our experiments, the cloned *scrbcs-1* gene was used as an external control, which showed its fidelity in the series of dilutions of the template DNA when amplified with the two sets of primers. There is no noticeable difference among the central wells of the thermocycler heat block (data not shown). Using the exact conditions as the experimental ones, external control reactions showed that the amount of amplification is quantitatively related to the amount of input templates. The systematic variation due to primer related differences (such as amplification efficiency and product molecular weight) was effectively eliminated by comparing to the ratio of the control products. Since the ratio of the two products from the same template is most important in the experiment, equal amounts of input template DNA in the two reactions is critical. This variation is minimized by making a master mix (without the two primers), and taking equal aliquots to new tubes, followed by addition of primers. Our previous experiments showed that pipetting errors do not affect the amount of product due to excess primers used in the reaction (data not shown).

Caution should be used when comparing the levels of amplification products of different samples. First, we started from 1 μ g total RNA of different samples, which might not represent the same amount of sample. The yields of total RNA from different samples were not the same, ranging from 30 μ g to 75 μ g per 100mg sample. Second, RNAs from different samples may have different efficiency in reverse transcription and PCR amplification, e.g. due to the presence of inhibitor(s) in the RNA preparation.

No *rbcs* gene expression was detected in non-green organs (embryogenic callus, root, stem, and etiolated leaf) by northern hybridization, nested-RACE, or differential amplification (data not shown). In sugarcane, transcription of the *scrbc*s genes is turned on or off in response to light, and that the cleavage and polyadenylation sites are independent of developmental and environmental control. This indicates that multiple polyadenylation is the default nature of the *scrbc*s mRNAs. Thus, the possible biological function of this characteristic is still unclear. However, one might still speculate on the possible function of the multiple polyadenylation. The different lengths of *scrbc*s mRNAs generated (1) may be expressed unevenly between the BS and MS cell types; (2) may respond differentially to environmental changes such as light intensity; or (3) may be differentially translated during development.

F. Sugarcane Leaf Bombardment

All the *scrbc*s genes cloned and analyzed in this study are highly conserved including the promoter regions. They all contain the sequence motifs which are important in directing the maize *rbc*s gene expression. The *scrbc*s-3 gene promoter, however, contains sequences in the distal region quite different from those of the other sugarcane *scrbc*s genes. We thus chose the promoters of *scrbc*s-1 and *scrbc*s-3 genes for our bombardment experiments.

Contrary to the results of experiments in maize (Bansal *et al.*, 1992) that showed very low levels of expression of the GUS gene under the control of different promoters in leaf epidermis, most of the GUS foci in our experiments are located within the leaf epidermal cells. This could be the result of inefficient delivery of DNA to layers below the epidermis, and/or efficient penetration of the enzyme substrate into the sugarcane leaves. Substrate penetration efficiency is unlikely since a similar pattern of expression was obtained when the maize anthocyanin synthesis pathway activators, C1 and R genes, were used as reporter genes (data not shown). The C1 and R proteins activate the sugarcane anthocyanin pathway to synthesize purple anthocyanin, thus no exogenous substrates are needed.

Although it is impossible to locate the synthesized (water-soluble) anthocyanin at the cellular level through cross sectioning of the leaf, many of the faint spots are thought to be below the epidermis (L. Bogorad, personal communication). Therefore, problems in DNA penetration are a more likely explanation. A possible reason for this resistance may be the toughness of the sugarcane leaf. These leaves contain large amounts of silica and have thick-walled epidermal and sclerenchyma cells at the poles of the vascular bundles.

Various means were tried to overcome cell wall toughness, including the use of organic solvents to dissolve the silica, removal of the epidermis by a razor blade or sandpaper, choice of gunpowder setting, and the bombardment of leaf cross-sections. However, no treatment increased GUS expression in the cells below the epidermis. Another group of researchers working on sugarcane also met with similar problems (J. Irvine, personal communication).

Promoters of photosynthetic genes have been reported to direct expression of reporter genes in photosynthetic tissues but not epidermal cells due to the absence of chloroplasts in these cells. Recently, in transgenic tobacco plants, Dupree *et al.* (1991) found that the tobacco *rbcs* promoter directs GUS expression in epidermal cells.

This was explained by demonstration of unusual chloroplasts in the epidermal cells. However, chloroplasts have not been observed in sugarcane epidermal cells. An alternative explanation might be that the *scrbc*s genes have a default expression pattern and the inhibitors are not sufficient to suppress transient expression.

Recently, Hall's group (Texas A & M University) reported that a phaseolin promoter directs seed-specific expression of the GUS gene in transgenic plants. The same construct, after bombarded into the leaf sample, is also expressed transiently in leaf cells (personal communication). The reason may be that in transgenic plants, the integrated reporter genes were prevented from transcription by tightly condensed chromatin in organs other than seeds, and only the genes in the seed were exposed to transcription activators. In the transient gene experiment, the genes were not blocked so they could be activated by the transcription activators. The ubiquitin promoter directs GUS expression in BS and MS cells at a ratio of about 1:1, which means that the BS and MS cells are almost equally competent for transient expression of the GUS gene. It is still unclear why the *scrbc*s directed such a high level of expression of the GUS gene in the sugarcane epidermal cells.

Light is believed to trigger the C3-C4 transition. The yellow leaves used in the experiments were exposed to light when they were placed on MS media. Studies in maize showed that the *rbcS* mRNA begins to be expressed more in the BS cells during this stage (Bansal *et al.*, 1992), and the Kranz anatomy begins to develop. In sugarcane leaves, however, the morphological differentiation of BS and MS cells was detected at an early developmental stage, *e.g.* the yellowish, unexpanded leaves (leaf No. -1 and younger) that have never been exposed to direct light. Since C4 photosynthesis requires high PEPC activity in the MS cells for initial fixation of CO₂, the measurement of C4-specific PEPC activity in leaf samples could be an indirect index of C4 photosynthetic activity. Dr. Sun studied the isozymes of PEPC of different sugarcane leaves and organs. He found that a "fast" migrating band presents in all leaves and organs studied, while a "slow" migrating band can be detected only in green leaves, and is most active in mature photosynthetic leaves. The C3 or "housekeeping" type PEPC is not tissue-specific and not developmentally regulated, while the C4 type is tissue specific and is developmentally and light regulated. Thus, the "slow" band represents the C4-specific PEPC activity, and the "fast" band serves a "housekeeping" function (Albert, 1991). In our experiment, the "slow" band was barely detectable in unexpanded leaves only after culture for 3 days under light. The results

suggest that the leaf samples we used in bombardment experiments (unexpanded leaves after cultured for 3 days under light) are already in the C3-C4 transition stage.

Both *scrbc*s promoters direct GUS expression in the BS and MS cells at a ratio of approximately 4:1, while the ubiquitin promoter gives a ratio of 1:1. These results demonstrate that *rbcs* genes may be expressed in all photosynthetic cells in the sugarcane leaf, although expression in the BS cells is four times more. Our findings further support the suggestion of Schaffner and Sheen (1991) that *rbcs* genes have default C3 character, which means that although all the *scrbc*s genes can be expressed at low level in C3 stage, their expression is restricted in BS cells of the C4 stage. It is necessary to point out that, in our experiment, only the promoters of two *scrbc*s genes with sequence variation in the 5' regions were tested. A more conclusive answer to the different types of *rbcs* genes in C4 plants might be reached by functional analysis of the 3' UTRs as well as the promoter regions of the remaining *scrbc*s genes.

G. Evolution of *Scrbcs* Genes

Sugarcane belongs to the family *Poaceae*, subfamily *Panicoideae*, tribe *Andropogoneae* and subtribe *Saccharininae* (Hunsiqi, 1993b). Modern sugarcane is a complex hybrid of two or more of the five species of the genus *Saccharum* (*S. officinarum*, *S. spontaneum*, *S. barberi*, *S. sinense* and *S. robustum*). Although it is commonly believed that the evolution of sugarcane is a recent event (about several thousand years ago, see Hunsiqi, 1993b), the origin of sugarcane is still controversial. Among different views, the widely-accepted concept of a "*Saccharum* Complex" was proposed by Mukherjee (1957) and revised by Daniels *et al.* (1975). The complex includes the closely related interbreeding genera *Saccharum*, *Erianthus*, *Sclerostachya*, *Narenga* and *Miscanthus*, which represent the most numerous species in the Indo-Myanmar-China border region. Sugarcane is believed to have originated in this region, despite the fact that the diversity center of sugarcane is undisputedly in New Guinea.

S. spontaneum is the most primitive species in the genus. This wild cane has a wide distribution in Asia and Africa, and has contributed substantially toward the development of other species. Daniels and Roach (1987) suggested that *S.*

officinarum, a garden cane that is not found in the wild, evolved from *S. spontaneum* by introgression with other genera, e.g. *Erianthus*, *Sclerostachya*, *Narenga* and *Miscanthus*, possibly through some plausible intermediate steps, the *S. robustum* groups. The wild cane *S. robustum* is not a well-defined species, and it covers a range of introgressions in the *Saccharum* Complex. *S. barberi* and *S. sinense* are believed to be evolved from *S. officinarum* through introgression with *S. spontaneum* and other genera within the complex.

S. officinarum has the commercial qualities of high sucrose content and purity, low fiber and starch, and was first termed as "noble cane" by the Dutch in Java. It was widely planted throughout the world before 1925, when the noble cane era ended due to serious diseases. The noblization process was then used to create new varieties of sugarcane by modified back crossing of *S. officinarum* with the wild cane *S. spontaneum* followed by a repeat back crossing to the noble parent *S. officinarum*. Cytological studies discovered that chromosome transmission during noblization varies with the two parents used, and most of the *S. spontaneum* chromosomes are quickly lost in a few generations (Sreenivasan et al., 1987).

The variety we used in our research (cv. H32-8560) is a fourth generation derivative of *S. officinarum* (2n=80), *S. spontaneum* (2n=40 to 128) and *S. barberi* (2n=82 to 124) within the *Saccharum* complex. Figure 24 shows the genealogy of cv. H32-8560 (Tew, 1987).

Of the 115-116 chromosomes in cv. H32-8560 (Price, 1963), most are from *S. officinarum* and the rest from *S. spontaneum* or *S. barberi*. The genealogy of EK28 is not certain, but it is believed to be a hybrid of POJ100 and EK2 (Jeswiet, 1930), both of which were from *S. officinarum*. Due to the high number of chromosomes in the genome, and the complexity of chromosome behavior, it is not possible to predict the ploidy of this variety. More details on the origin of these chromosomes are not available; thus, the application of new techniques, such as chromosome *in situ* hybridization and high-resolution chromosome banding pattern, would be useful.

Considering the high copy number of *rbcS* genes in other plants, it is not surprising that the whole genome of sugarcane contains about 16 copies of the *rbcS* gene. It is surprising, however, that all the *rbcS* genes in sugarcane are so closely related.

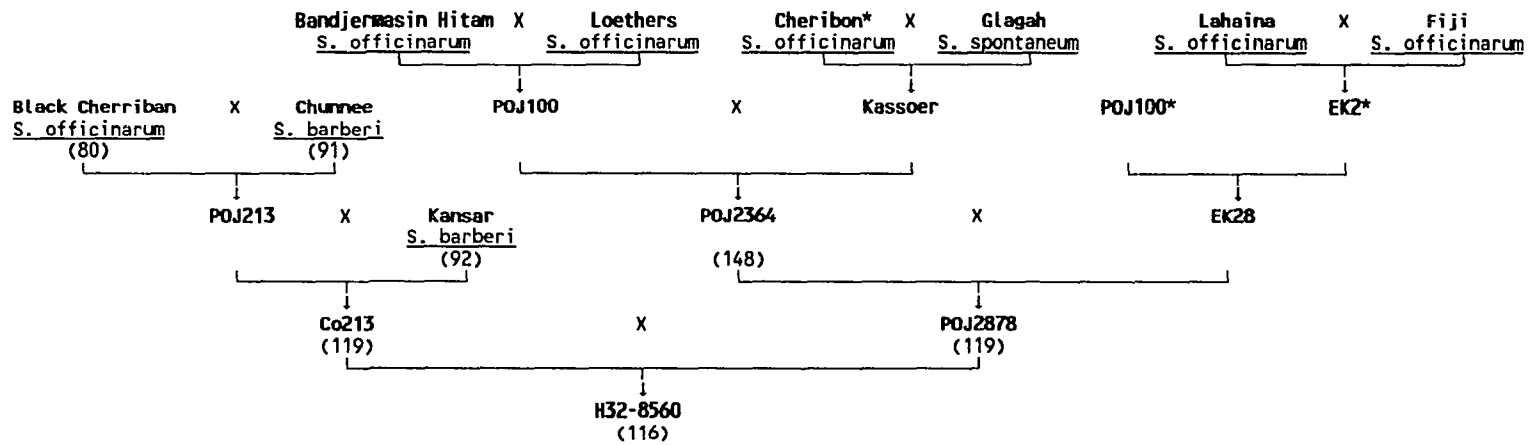


Figure 25. Genealogy of *Saccharum* spp. hybrid cultivar H32-8560. Number of chromosomes (if known) is shown below the variety name. The male parent is shown on the right side of the cross. The uncertain parents were marked with "*".

In all other species studied thus far, *rbcS* genes are highly conserved within their coding region, but far less conserved in the 5' and 3' regions. Based on nucleotide sequence similarity, *rbcS* gene members can be grouped into different subfamilies (all the members in one species are considered to be in one family). Although members of the same subfamily are more similar to each other than to those from different subfamilies (Dean *et al.*, 1989), the homology among their 3'-untranslated regions is generally lower than 85%.

In *Petunia* (Mitchell), the eight *rbcS* gene members are divided into three subfamilies. One subfamily contains six genes, five of which are closely linked in the genome (within 25kb). Within the coding region, there is only 10% nucleotide sequence divergence among different subfamilies and 0-3% within subfamilies (Dean *et al.*, 1989). Computer analysis of their 3'-regions (starting from translation stop codon) showed about 60-75% homology within the subfamily, and about 40-60% homology among different subfamilies.

Three of the five tomato *rbcS* genes belong to one subfamily, and they are closely linked within a 10kb region on chromosome 2 (Vallejos *et al.*, 1986; Sugita *et al.*, 1987). Within the coding region, there is only 2%

nucleotide sequence divergence among these three genes, and about 10-14% among different subfamilies. The 5' UTRs and 3' UTRs of the five genes show far less homology than that of the petunia genes.

Other examples in dicots include: three of the four potato *rbcS* genes cloned by Wolter *et al.* (1988) are arranged in a tandem array within a 10-kb region; in pea two of the five *rbcS* genes have been localized to a single genomic clone (Cashmore, 1983); and three *rbcS* genes of *Arabidopsis* are linked within a 10kb DNA region in the genome (Krebbers *et al.*, 1988).

Linkage of *rbcS* genes in genomic DNA and significant variation in the 3' UTR of *rbcS* genes within a species are also found in monocot plants. Two pairs of the five *rbcS* genes cloned from *Lemna gibba* show linkage; one set is linked in an inverted orientation with a 2.5kb region separating the 5' UTRs of the genes, and the other pair are in tandem within a 6.5kb region (Wimpee *et al.*, 1983). In wheat all the cloned *rbcS* genes have been mapped to the long arm of homoeologous group 5 (containing a single subfamily) and to the short arm of homoeologous group 2 chromosomes (containing three subfamilies) by hybridization of the 3' UTRs of three wheat *rbcS* genes with the various aneuploids of these homoeologous groups (Galili *et al.*, 1992). Maize,

a diploid with a chromosome number of 10, contains about 3 to 4 *rbcS* gene members. The 3' UTRs of these genes were found to be quite different, and were used as probes to distinguish the different gene members (Sheen and Bogorad, 1986; Schaffner and Sheen, 1991).

It is of great interest to determine how such uniformity of sugarcane *rbcS* genes came about. It is very hard to imagine that nature could have such an extremely strong selection force to maintain these nucleotide sequences. There are some possible explanations. First, these different copies of genes might be the results of recent duplication events in the genome. Sugarcane is a newly formed species, and polyploidy might have evolved very recently. As described above, the other four species in the *Saccharum* genus originated from *S. spontaneum* and the other genera within the *Saccharum* Complex. Second, *S. officinarum* was used as the only maternal plant in the genealogy of cv. H32-8560, and therefore the *rbcl* proteins in the variety are only from *S. officinarum* but not *S. spontaneum* or other species. It is reasonable that the interaction between *rbcS* and *rbcl* proteins would favor the selection of *S. officinarum* *rbcS* proteins, thus the *rbcS* genes. Third, each of the positive λ clones (about 20kb) from the genomic library contain only one *scrbcS* gene fragment. It appears unlikely that the *scrbcS* genes exist as clusters in the

chromosomes, as in some plant species. It is possible that each of the *scrbc*s gene is located at the same locus of different homologous chromosomes, just as RFLP fragments are located at the same locus of different homologous chromosomes (Da Silva et al., 1993).

Our finding is consistent with the hypothesis, resulting from recent RFLP experiments, that *S. spontaneum*, and also possibly *S. officinarum*, are an octo-auto-polyploid (Da Silva et al., 1993). Consequently, all the 16 copies of *rbcs* genes might have arisen from a single locus on the chromosome of the *S. officinarum* genome.

Chapter VI. CONCLUSION

In this thesis research, we have determined the structure and expression patterns of the sugarcane ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit genes. The results show that:

1. there are about 12-16 copies of genes encoding the ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit proteins in sugarcane;
2. these genes are highly conserved, even in the 5' and 3' untranslated regions;
3. these genes are expressed throughout the development of sugarcane leaf, at both C3 and C4 stages, but more specifically in the BS cells upon the formation of Kranz anatomy;
4. the messengers of these genes have multiple polyadenylation sites, which are not associated with the differential expression of the genes in different cell types of leaves during development.

The above results suggest that:

1. all the *scrbc*s genes have a default C3 expression pattern;
2. the regulation of differential expression of the *scrbc*s genes might be at different levels, e.g. post-transcriptional;
3. all *scrbc*s genes may have recently evolved from a single locus on the chromosome of the *S. officinarum* genome.

REFERENCES

- Albert, H H (1991). Structure and expression of a sugarcane gene encoding a housekeeping phosphoenolpyruvate carboxylase. PhD thesis. University of Hawaii at manoa.
- Albert, H H, T Martin and S S M Sun (1992). Structure and expression of a sugarcane gene encoding a housekeeping phosphoenolpyruvate carboxylase. *Plant Mol Biol.* 20:663-671.
- Alexander, A G (1990). High growth sugarcane as a natural conservation system for crop land, atmosphere and solar radiation. In *XVI Symp Natural Resources*, Sept 13, 1990, Puerto Rico. pp29.
- Altschul, S F, W Gish, W Miller, E W Myers and D J Lipman (1990). Basic local alignment search tool. *J Mol Biol.* 215:403-410.
- Amara, S G, R M Evans and M G Rosenfeld (1984). Calcitonin/calcitonin gene related peptide transcription unit: tissue-specific expression involves selective use of alternative polyadenylation sites. *Mol Cell Biol.* 4:2151-2160.
- Bansal, K C, J-F Viret, J Haley, B M Khan, R Schantz and L Bogorad (1992). Transient expression from *cab-m1* and *rbcS-m3* promoter sequences is different in mesophyll and bundle sheath cells in maize leaves. *Proc Natl Acad Sci USA.* 89:3654-3658.
- Bassi, R and C Passera (1982). Effect of growth conditions on carboxylating enzymes of *Zea mays* plants. *Photosyn Res.* 3:53-58.
- Benda, G T A (1978). Leaf sheaths and the inhibition of germination of young axillary buds in sugarcane. *Proc Int Soc Sugar Cane Technol.* 16:1599-1603.
- Bennet, M D and J B Smith (1976). Nuclear DNA amounts in angiosperms. *Philos Trans Royal Soc Series B.* 274:227-274.

- Benson, D, D J Lipman and J Ostell (1993). Genbank. Nucl Acid Res. 21:2963-2965.
- Berry, J O, D E Breiding and D F Klessig (1990). Light-mediated control of translational initiation of ribulose-1,5-bisphosphate carboxylase in *Amaranth* cotyledons. Plant Cell. 2:795-803.
- Berry-Lowe, S L and R B Meagher (1985). Transcriptional regulation of a gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in soybean tissue is linked to the phytochrome response. Mol Cell Biol. 5:1910-1917.
- Birch, R G and T Franks (1991). Development and optimisation of microprojectile systems for plant genetic transformation. Aust J Plant Physiol. 18:453-469.
- Birnstiel, M L, M Busslinger and K Strub (1985). Transcription termination and 3' procession: the end is in site! Cell. 41:349-359.
- Black, C C (1973). Photosynthetic carbon fixation in relation to net CO₂ uptake. Ann Rev Plant Physiol. 24:253-286.
- Boardman, M, G S Basi and R V Sorti (1985). Multiple polyadenylation sites in a *Drosophila* tropomyosin gene are used to generate functional mRNAs. Nucl Acid Res. 13:1763-1776.
- Boothe, J G and D B Walden (1990). Gene expression in embryos and seedlings of maize. Maydica. 35:187-194.
- Bouton, J H, R H Brown, P T Evans and J A Jernstedt (1986). Photosynthesis, leaf anatomy, and morphology of progeny from hybrids between C3 and C3/C4 *Panicum* species. Plant Physiol. 80:487-492.
- Braun, R E (1991). Temporal translational regulation of the protamine 1 gene during mouse spermatogenesis. Enzyme. 44:120-128.
- Brown, R H and P W Hattersley (1989). Leaf anatomy of C3-C4 species as related to evolution of C4 photosynthesis. Plant Physiol. 91:1543-1550.
- Browne, D L and J B Dodgson (1993). The gene encoding chicken chromosomal protein HMG-14a is transcribed into multiple mRNAs. Gene. 124:199-206.

- Buchanan, B B (1980). Role of light in the regulation of chloroplast enzymes. *Ann Rev Plant Physiol.* 31:341-374.
- Bull, T A (1969). Photosynthetic efficiencies and photorespiration in Calvin cycle and C4 dicarboxylic acid plants. *Crop Sci.* 9:726-729.
- Burr, G O (1969). A note on the efficiency of sugarcane. *Taiwan Sugar Exp Stn Rep.* 3:1-4.
- Buzby, J S, T Yamada and E M Tobin (1990). A light-regulated binding activity interacts with a conserved region of a *Lemna gibba rbcS* promoter. *Plant Cell.* 2:805-814.
- Calvin, M and A A Benson (1949). The path of carbon in photosynthesis IV: the identity and sequence of the intermediates in sucrose synthesis. *Science.* 109:140-142.
- Cameron, R G, C L Bassett, J H Bouton and R H Brown (1989). Transfer of C4 photosynthetic characters through hybridization of *Flaveria* species. *Plant Physiol.* 90:1538-1545.
- Cashmore, A R (1983). Nuclear genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. In *Genetic Engineering of Plants.* Edited by M Kosuge and A Hollander. Plenum, New York. pp29-38.
- Chen, K, S Johal and S G Wildman (1976). Role of chloroplast and nuclear DNA genes during evolution of Fraction I protein. In *Genetics and Biogenesis of Chloroplasts and Mitochondria.* Edited by Th. Bucher. Elsevier/North-Holland Biomedical Press, Amsterdam, The Netherlands. p3-11.
- Christensen, A H, R A Sharrock and P H Quail (1992). Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Mol Biol.* 18:675-689.
- Church, G M and W Gilbert (1984). Genomic sequencing. *Proc Natl Acad Sci USA.* 81:1991-1995.
- Coen, E S and E M Meyerowitz (1991). The war of the whorls: genetic interactions controlling flower development. *Nature.* 353:31-37.

- Coruzzi, G, R Broglie, C Edwards and N-H Chua (1984).
Tissue-specific and light-regulated expression of a pea
nuclear gene encoding the small subunit of ribulose-
1,5-bisphosphate carboxylase. *EMBO J.* 3:1671-1679.
- Crespo, H M, M Frean, C F Cresswell and J Tew (1979). The
occurrence of both C3 and C4 photosynthetic
characteristics in a single *Zea mays* plant. *Planta.*
147:257-263.
- Da Silva, J A G, M E Sorrells, W L Burnquist and S D
Tanksley (1993). RFLP linkage map and genome analysis
of *Saccharum spontaneum*. *Genome.* 36:782-791.
- Daniels, J, P Smith, N Paton and C A Williams (1975). The
origin of the genus *Saccharum*. *Sugarcane Breed Newsl.*
36:24-39.
- Daniels, J and B T Roach (1987). Taxonomy and evolution.
In *Sugarcane improvement through breeding, Developments
in Crop Sciences II.* Edited by D J Heinz. Elsevier
Science Publishers B. V., Amsterdam, the Netherlands.
pp7-84.
- Darnell, J, H Lodish and D Baltimore (1990).
Photosynthesis. In *Molecular Cell Biology.* 2 Ed.
Scientific American Books. pp631-634.
- Das, A (1993). Control of transcription termination by RNA-
binding proteins. *Ann Rev Biochem.* 62:893-930.
- Dean, C, P van den Elzen, S Tamaki, P Dunsmuir and J
Bedbrook (1985). Differential expression of the eight
genes of the petunia ribulose bisphosphate carboxylase
small subunit multigene family. *EMBO J.* 4:3055-3061.
- Dean, C, S Tamaki, P Dunsmuir, M Favreau, C Katayama, H
Dooner and J Bedbrook (1986). mRNA transcripts of
several plant genes are polyadenylated at multiple
sites *in vivo*. *Nucl Acid Res.* 14:2229-2240.
- Dean, C, Favreau, M, P Dunsmuir and J Bedbrook (1987).
Confirmation of the relative expression levels of the
petunia (Mitchell) *rbcS* genes. *Nucl Acid Res.*
15:4655-4668.
- Dean, C, E Pichersky and P Dunsmuir (1989). Structure,
evolution, and regulation of *rbcS* genes in higher
plants. *Ann Rev Plant Physiol.* 40:415-539.
- Dengler, N G, R E Dengler and P W Hattersley (1985).
Differing ontogenetic origins of PCR (Kranz) sheath in

- leaf blades of C4 grasses (*Poaceae*). *Am J Bot.* 72:284-302.
- Devereux, J R, P Haerberli and O Smithies (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl Acid Res.* 12:387-395.
- Dietrich, R A, S E Radke and J J Harada (1992). Downstream DNA sequences are required to activate a gene expressed in the root cortex of embryos and seedlings. *Plant Cell.* 4:1371-1382.
- Doyle, J J and J L Doyle (1989). Isolation of plant DNA from fresh tissue. *Focus.* 12:13-15.
- Dupree, P, K-H Pwee and J C Gray (1991). Expression of photosynthesis gene-promoter fusions in leaf epidermal cells of transgenic tobacco plants. *Plant J.* 1:115-120.
- Ennis, P D, J Zemmour, R D Salter and P Parham (1990). Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. *Proc Natl Acad Sci USA.* 87:2833-2837.
- Fourney, R M, J Miyakoshi, R S Day and M C Paterson (1988). Northern Blotting: efficient RNA staining and transfer. *Focus* 10:5-6.
- Frayne, E G, E J Leys, G F Crouse, A G Hook and R E Kellems (1984). Transcription of the mouse dihydrofolate reductase gene proceeds unabated through seven polyadenylation sites and terminates near a region of repeated DNA. *Mol Cell Biol.* 4:2921-2924.
- Frohman, M A (1990). Rapid amplification of cDNA ends (RACE): user-friendly cDNA cloning. *Amplifications.* 5:11-15.
- Galili, S, G Galili, Y Avivi and M Feldman (1992). Identification and chromosomal location of four subfamilies of the rubisco small subunit genes in common wheat. *Theor Appl Genet.* 83:385-391.
- Gallagher, T F and R J Ellis (1982). Light-stimulated transcription of genes for two chloroplast polypeptides in isolated pea leaf nuclei. *EMBO J.* 1:1493-1498.
- Gallie, D R (1993). Posttranscriptional regulation of gene expression in plants. *Ann Rev Plant Physiol Plant Mol Biol.* 44:77-105.

- Gallo-Meagher, M and J E Irvine (1993). Effects of tissue type and promoter strength on transient GUS expression in sugarcane following particle bombardment. *Plant Cell Rep.* 12:666-670.
- Gilmartin, P M, L Sarokin, J Memelink and N H Chua (1990). Molecular light switches for plant genes. *Plant Cell.* 2:369-378.
- Goatley, M and H Smith (1974). Differential properties of phosphoenolpyruvate carboxylase from etiolated and green sugar cane. *Planta.* 117:67-73.
- Goldberg, R B, S J Baker and L Perez-Grau (1989). Regulation of gene expression during plant embryogenesis. *Cell.* 56:149-160.
- Goldstein, C S and J M Widholm (1990). Photosynthetic characterization of photoautotrophic cells cultured in a minimal medium. *Plant Physiol.* 94:1641-1646.
- Gray, J C, S D Kung and S G Wildman (1976). Polypeptide chains of the large and small subunits of Fraction I protein. In *Genetics and Biogenesis of Chloroplasts and Mitochondria*. Edited by Th. Bucher. Elsevier/North-Holland Biomedical Press, Amsterdam, The Netherlands. pp13-16.
- Grob, U and K Stuber (1987). Discrimination of phytochrome dependent light inducible from non-light inducible plant genes. *Nucl Acid Res.* 15:9957-9973.
- Gruha, J W and R L Hudspeth (1987). The phosphoenolpyruvate carboxylase gene family of maize. In *Plant Gene Systems and Their Biology*. Edited by Key, J L and L McIntosh. Alan R. Liss, New York. pp207-216.
- Harpster, M H and W C Taylor (1986). Maize phosphoenolpyruvate carboxylase cloning and characterization of mRNAs encoding isozymic forms. *J Biol Chem.* 261:6132-6136.
- Helms, C, J E Dutchik and M V Olson (1987). A λ DNA protocol based on purification of phage on DEAE-cellulose. *Meth Enzymol.* 153:69-82.
- Hemmingsen, S M, C Woolford, S M van de Vies, K Tilly, D T Dennis, C P Georgopoulos, R W Hendrix and R J Ellis (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature.* 333:330-334.

- Hernandez-Lucas, C, J Royo, J Paz-Ares, F Ponz, F Garcia-Olmedo and P Carbonero (1986). Polyadenylation site heterogeneity in mRNA encoding the precursor of the barley toxin β -hordothionin. *FEBS Lett.* 200:103-106.
- Hesketh, J E and I F Pryme (1991). Interaction between mRNA, ribosomes and the cytoskeleton. *Biochem J.* 277:1-10.
- Holaday, A S, S Talkmitt and M E Doohan (1985). Anatomical and enzymic studies of leaves of a C3 x C4 *Flaveria* F1 hybrid exhibiting reduced photorespiration. *Plant Sci.* 41:31-39.
- Hudspeth, R L and J W Grula (1989). Structure and expression of the maize gene encoding phosphoenolpyruvate carboxylase isozyme involved in C4 photosynthesis. *Plant Mol Biol.* 12:579-589.
- Hunsigi, G (1993a). Production potential of sugarcane. In *Production of sugarcane*, Advanced Series in Agricultural Sciences 21. Springer-Verlag. New York. pp7-8.
- Hunsigi, G (1993b). Introduction. In *Production of sugarcane*, Advanced Series in Agricultural Sciences 21. Springer-Verlag. New York. pp12-28.
- Hunt, A G (1988). Identification and characterization of cryptic polyadenylation sites in the 3' region of a pea ribulose-1,5-bisphosphate carboxylase small subunit gene. *DNA.* 7:329-336.
- Irvine, J E (1975). Relations of photosynthetic rates and leaf and canopy characters to sugarcane yield. *Crop Sci.* 15:671-676.
- Jackson, R J (1993). Cytoplasmic regulation of mRNA function: the importance of the 3'untranslated region. *Cell.* 74:9-14.
- Jefferson, R A (1987). Assaying chimeric genes in plants: the GUS fusion system. *Plant Mol Biol Rep.* 387-405.
- Jefferson, R A, T A Kavanagh and M W Bevan (1987). GUS fusions: β -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6:3901-3907.
- Jeswiet, J (1930). The development of selection and breeding of the sugarcane in Java. *Proc Int Soc Sugar Cane Technol.* 3:44-57.

- Joshi, C P (1987). Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. *Nucl Acid Res.* 15:9627-9640.
- Kano-Murakami, Y, I Suzuki, T Sugiyama and M Matsuoka (1991). Sequence-specific interactions of a maize factor with a GC-rich repeat in the phosphoenolpyruvate carboxylase gene. *Mol G Genet.* 225:203-208.
- Kennedy, R A (1976). Photorespiration in C3 and C4 plant tissue cultures. *Plant Physiol.* 58:573-575.
- Keohevong, P and W G Thilly (1989). Fidelity of DNA polymerase in DNA amplification. *Proc Natl Acad Sci USA.* 86:9253-9257.
- Kerppola, T K and C M Kane (1988). Intrinsic sites of transcription termination and pausing in the *c-myc* gene. *Mol Cell Biol.* 8:4389-4394.
- Klausner, R D, T A Rouault and J B Harford (1993). Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell.* 72:19-28.
- Kortschak, H P, C E Hartt and G O Burr (1965). Carbon dioxide fixation in sugarcane leaves. *Plant Physiol.* 40: 209-213.
- Kortschak, H P and L G Nickell (1970). Calvin-type carbon dioxide fixation in sugarcane stalk parenchyma tissue. *Plant Physiol.* 45:515-516.
- Krebbbers, E, J Seurinck, L Herdies, A R Cashmore and M P Timko (1988). Four genes in the two diverged subfamilies encode the ribulose-1,5-bisphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana*. *Plant Mol Biol.* 11:745-759.
- Kuhlemeier, C, P J Green and N H Chua (1987). Regulation of gene expression in higher plants. *Ann Rev Plant Physiol.* 38:221-257.
- Kuijper, J (1915). DeGroei van Bladschijf, Bladscheede en Stengel van het suikerriet. *Arch Suik Ned Indie.* 23:528-556.
- Laetsch, W M and I Price (1969). Development of the dimorphic chloroplasts of sugar cane. *Am J Bot.* 56:77-87.

- Laetsch, W M and H P Kortschak (1972). Chloroplast structure and function in tissue cultures of a C4 plant. *Plant Physiol.* 49:1021-1023.
- Langdale, J A, M C Metzler and T Nelson (1987). The *Argentia* mutation delays normal development of photosynthetic cell-types in *Zea mays*. *Dev Biol.* 122:243-255.
- Langdale, J A, B A Rothermel and T Nelson (1988a). Cellular pattern of photosynthetic gene expression in developing maize leaves. *Gene Dev.* 2:106-115.
- Langdale, J A, I Zelitch, E Miller and T Nelson (1988b). Cell position and light influence C4 versus C3 patterns of photosynthetic gene expression in maize. *EMBO J.* 7:3643-3651.
- Langdale, J A, W C Taylor and T Nelson (1991). Cell-specific accumulation of maize phosphoenolpyruvate carboxylase is correlated with demethylation at a specific site >3Kb upstream of the gene. *Mol Gen Genet.* 225:49-55.
- Lebrun, M, G Waksman and G Freyssinet (1987). Nucleotide sequence of a gene encoding corn ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (rbcs). *Nucl Acid Res.* 15:4360.
- Longhurst, T J, H F Tung and C J Brady (1990). Developmental regulation of the expression of alcohol dehydrogenase in ripening tomato fruits. *J Food Biochem.* 14:421-433.
- MacDonald, M H, B D Mogen and A G Hunt (1991). Characterization of the polyadenylation signal from the T-DNA-encoded octopine synthase gene. *Nucl Acid Res.* 19:5575-5581.
- Manzara, T and W Gruissem (1988). Organization and expression of the genes encoding ribulose-1,5-bisphosphate carboxylase in higher plants. *Photosyn Res.* 16:117-139.
- Maretzki, A (1987). Tissue culture: its prospects and problems. In *Sugarcane improvement through breeding, Developments in Crop Sciences II*. Edited by D J Heinz. Elsevier Science Publishers B. V., Amsterdam, the Netherlands. pp343-384.

- Mason, I J (1992). Rapid and direct sequencing of DNA from bacteriophage plaques using sequential linear and asymmetric PCR. *BioTechniques*. 12:60-61.
- Matsuoka, M and E Minami (1989). Complete structure of the gene for phosphoenolpyruvate carboxylase from maize. *Eur J Biochem*. 181:593-598.
- Mogen, B D, M H MacDonald, G Leggewie and A G Hunt (1992). Several distinct types of sequence elements are required for efficient mRNA 3' end formation in a pea *rbcs* gene. *Mol Cell Biol*. 12:5406-5414.
- Montoliu, L, J Rigau and P Puigdomenech (1990). Multiple polyadenylation sites are active in the α 1-tubulin gene from *Zea mays*. *FEBS Lett*. 277:29-32.
- Moore, P H (1987). Anatomy and morphology. In *Sugarcane improvement through breeding, Developments in Crop Sciences II*. Edited by D J Heinz. Elsevier Science Publishers B. V., Amsterdam, the Netherlands. pp85-142.
- Moore, P H (1989). Physiological basis for varietal improvement. In *Proc Int Symp sugarcane varietal improvement, present status and future thrusts*. Sugarcane Breeding Institute Coimbatore to commemorate its Platinum Jubilee. Sept 3-7, 1987. SBI, Coimbatore. pp19-55.
- Mosinger, E, A Batschauer, E Schafer and K Apel (1985). Phytochrome control of in vitro transcription of specific genes in isolated nuclei of barley (*Hordeum vulgare*). *Eur J Biochem*. 147:137-142.
- Mukherjee, S K (1957). Origin and distribution of *Saccharum*. *Bot Gaz*. 119:55-61.
- Murashige, T and F Skoog (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol Plant*. 15:473-479.
- Nelson, T, M Harpster, S P Mayfield and W C Tylor (1984). Light-regulated gene expression during maize leaf development. *J Cell Biol*. 98:558-564.
- Nelson, T and J A Langdale (1989). Patterns of leaf development in C4 plants. *Plant Cell*. 1:3-13.
- Nevins, J R and J E Darnell, Jr. (1978). Steps in the processing of Ad2 mRNA: poly(A)⁺ nuclear sequences are

- conserved and poly(A) addition precedes splicing. *Cell*. 15:1477-1493.
- Ngernprasirtsiri, J, R Chollet, H Kobayashi, T Sugiyama and T Akazawa (1989). DNA methylation and the differential expression of C4 photosynthesis genes in mesophyll and bundle sheath cells of greening maize leaves. *J Biol Chem*. 264:8241-8248.
- Okita, T W, X Li and M W Roberts (1994). Targeting of mRNAs to domains of the endoplasmic reticulum. *Trends Cell Biol*. 4:91-96.
- Perchorowicz, J T, D A Raynes and R G Jensen (1981). Light limitation of photosynthesis and activation of ribulose biphosphate carboxylase in wheat seedlings. *Proc Natl Acad Sci USA*. 78:2985-2989.
- Price, S (1963). Cytogenetics of modern sugarcane. *Econ Bot*. 17:97-106.
- Raines, C A, J C Lloyd and T A Dyer (1991). Molecular biology of the C3 photosynthetic carbon reduction cycle. *Photosyn Res*. 27:1-14.
- Richter, J D (1991). Translational control during early development. *Bioassay*. 13:179-183.
- Rolfe, S A and E M Tobin (1991) Deletion analysis of a phytochrome-regulated monocot *rbcS* promoter in a transient assay system. *Proc Natl Acad Sci USA*. 88:2683-2686.
- Sachs, A B (1993). Messenger RNA degradation in eukaryotes. *Cell*. 74:413-421.
- Saez, C G, J C Myers, T B Shows and L A Leinwand (1990). Human nonmuscle myosin heavy chain mRNA: generation of diversity through alternative polyadenylation. *Proc Natl Acad Sci USA*. 87:1164-1168.
- Saiki, R K, D H Gelfand, S Stoffel, S J Scharf, R Higuchi, G T Horn, K B Mullis and H A Erlich (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. 239:487-491.
- Sambrook, J, E F Fritsch and T Maniatis (1989). Mapping RNA with S1 Nuclease. In *Molecular cloning, a laboratory manual* Ed 2 Vol 2. Cold Spring Harbor Laboratory, New York. pp7.58-7.83.

- Sanfacon, H, P Brodmann and T Hohn (1991). A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes Dev.* 5:141-149.
- Schaffer, H E (1981). Estimation of DNA fragment length from agarose gels. *Anal Biochem.* 115:113-122.
- Schaffner, A R and J Sheen (1991). Maize *rbcS* promoter activity depends on sequence elements not found in dicot *rbcS* promoters. *Plant Cell.* 3:997-1012.
- Schindler, U and A R Cashmore (1990). Photoregulated gene expression may involve ubiquitous DNA binding proteins. *EMBO J.* 9:3415-3427.
- Seeni, S and A Gnanam (1983). Photosynthesis in cell suspension culture of a C4 plant, *Gisekia pharnaceoides* L.. *Plant & Cell Physiol.* 24:1033-1041.
- Sharpless, K, D Biegel, T Yang and J S Pachter (1993). β -actin mRNA-binding proteins associated with the cytoskeletal framework. *Eur J Biochem.* 212:217-225.
- Shaw, G and R Kamen (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell.* 46:659-667.
- Sheen, J Y and L Bogorad (1986). Expression of the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit gene and three small subunit genes in two cell types of maize leaves. *EMBO J.* 5:3417-3422.
- Sheen, J Y and L Bogorad (1987a). Differential expression of C4 pathway genes in mesophyll and bundle sheath cells of greening maize leaves. *J Biol Chem.* 262:11726-11730.
- Sheen, J Y and L Bogorad (1987b). Regulation of levels of nuclear transcripts for C4 photosynthesis in bundle sheath and mesophyll cells of maize leaves. *Plant Mol Biol.* 8:227-238.
- Sheen, J Y (1991). Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. *Plant Cell.* 3:225-245.
- Shomer-Ilan, A, S Beer and Y Waisel (1975). *Suaeda monoica*, a C4 plant without typical bundle sheaths. *Plant Physiol.* 56:676-679.
- Shyu, A B, J G Belasco and M E Greenberg (1991). Two distinct destabilizing elements in the c-fos message

- trigger deadenylation as a first step in rapid mRNA decay. *Genes Dev.* 5:221-231.
- Silverthorne, J and E M Tobin (1984). Demonstration of transcriptional regulation of specific genes by phytochrome action. *Proc Natl Acad Sci USA.* 81:1112-1116.
- Silverthorne, J, C F Wimpee, T Yamada, S A Rolfe and E M Tobin (1990). Differential expression of individual genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in *Lemna gibba*. *Plant Mol Biol.* 15:49-58.
- Simon, R, J P Tassan and J D Richter (1992). Translational control by poly(A) elongation during *Xenopus* development: differential repression and enhancement by a novel cytoplasmic polyadenylation element. *Genes Dev.* 6:2580-2591.
- Sreenivasan, T V, B S Ahloowalia and D J Heinz (1987). Cytogenetics. In *Sugarcane improvement through breeding*. Developments in Crop Sciences II. Edited by D J Heinz. Elsevier Science Publishers B. V., Amsterdam, the Netherlands. pp211-254.
- Standart, N M (1992). Masking and unmasking of maternal mRNA. *Sem Dev Biol.* 3:367-379.
- Stephen, A R and E M Tobin (1991). Deletion analysis of a phytochrome-regulated monocot *rbcS* promoter in a transient assay system. *Proc Natl Acad Sci USA.* 88:2683-2686.
- Strobaek, S, G C Gibbons, B Haslett, D Boulter and S G Wildman (1976). On the nature of the polymorphism of the small subunit of ribulose-1,5-bisphosphate carboxylase in the amphidiploid *Nicotiana tabacum*. *Carlsberg Res Commun.* 41:335-343.
- Sugita, M, T Manzara, E Pichersky, A Cashmore and W Gruissem (1987). Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase from tomato. *Mol Gen Genet.* 209:247-256.
- Takagaki, Y, L C Ryner and J L Manley (1989). Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev.* 3:1711-1724.

- Tang, W and S S M Sun (1993). Sequence of a sugarcane ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit gene. *Plant Mol Biol.* 21:949-951.
- Tew, T L (1987). New varieties. In *Sugarcane improvement through breeding*, Developments in Crop Sciences II. Edited by D J Heinz. Elsevier Science Publishers B. V., Amsterdam, the Netherlands. pp559-594.
- Thomas, M, C Cretin, J Vidal, E Keryer, P Gadai and E Monsinger (1990). Light-regulation of phosphoenolpyruvate carboxylase mRNA in leaves of C4 plants: evidence for phytochrome control on transcription during greening and for rhythmicity. *Plant Sci.* 69:65-78.
- Thompson, D M and R B Meagher (1990). Transcriptional and post-transcriptional processes regulate expression of RNA encoding the small subunit of ribulose-1,5-bisphosphate carboxylase differently in petunia and in soybean. *Nucl Acid Res.* 18:3621-3629.
- Ting, I P and C B Osmond (1973). Multiple forms of plant phosphoenolpyruvate carboxylase associated with different metabolic pathways. *Plant Physiol.* 51:448-453.
- Vallejos, C E, S D Tanksley and R Bernatzky (1986). Localization in the tomato genome of DNA restriction fragments containing sequences homologous to the rRNA (45S), the major chlorophyll a/b binding polypeptide and the ribulose bisphosphate carboxylase genes. *Genetics.* 112:93-105.
- Wahle, E and W Keller (1992). The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. *Ann Rev Biochem.* 62:419-440.
- Wang, J L, J J Long, T Hotchkiss and J O Berry (1993). C4 photosynthetic gene expression in light- and dark-grown Amaranth Cotyledons. *Plant Physiol.* 102:1085-1093.
- Wickens M (1992). Forward, backward, how much, when: mechanisms of poly(A) addition and removal and their role in early development. *Sem Dev Biol.* 3:399-412.
- Wimpee, C F, W J Stiekema and H M Tobin (1983). Sequence heterogeneity in the RuBP carboxylase small subunit gene family of *Lemna gibba*. In *Plant Molecular Biology*, Plant Molecular Biology, UCLA Symposium on Molecular and Cellular Biology. Edited by RB Goldberg. Liss, New York. pp391-401.

Wolter, F P, C C Fritz, L Willmitzer and J Schell (1988).
Rbcs genes in *Solanum tuberosum*: conservation of
transit peptide and exon shuffling during evolution.
Proc Natl Acad Sci USA. 85:846-850.

Yanagisawa, S and K Izui (1990). Multiple interactions
between tissue-specific nuclear proteins and the
promoter of the phosphoenolpyruvate carboxylase gene
for C4 photosynthesis in *Zea mays*. Mol Gen Genet.
224:325-332.