

Base excision repair in sugarcane

Lucymara F. Agnez-Lima*, Silvia R. Batistuzzo de Medeiros, Bruno S. Maggi and Giovanna A.S. Quaresma

Abstract

DNA damage can be induced by a large number of physical and chemical agents from the environment as well as compounds produced by cellular metabolism. This type of damage can interfere with cellular processes such as replication and transcription, resulting in cell death and/or mutations. The low frequency of mutagenesis in cells is due to the presence of enzymatic pathways which repair damaged DNA. Several DNA repair genes (mainly from bacteria, yeasts and mammals) have been cloned and their products characterized. The high conservation, especially in eukaryotes, of the majority of genes related to DNA repair argues for their importance in the maintenance of life on earth. In plants, our understanding of DNA repair pathways is still very poor, the first plant repair genes having only been cloned in 1997 and the mechanisms of their products have not yet been characterized. The objective of our data mining work was to identify genes related to the base excision repair (BER) pathway, which are present in the database of the Sugarcane Expressed Sequence Tag (SUCEST) Project. This search was performed by tblastn program. We identified sugarcane clusters homologous to the majority of BER proteins used in the analysis and a high degree of conservation was observed. The best results were obtained with BER proteins from *Arabidopsis thaliana*. For some sugarcane BER genes, the presence of more than one form of mRNA is possible, as shown by the occurrence of more than one homologous EST cluster.

INTRODUCTION

DNA from all living organisms is able to react with a large number of physical and chemical agents as well as with chemical compounds produced by cellular metabolism. The DNA damage induced by these different agents can interfere with cellular processes such as replication and transcription, resulting in cell death and/or mutations. The low frequency of mutation normally seen in cells is due to the presence of enzymatic pathways, which repair DNA damage and avoid mutation. The fact that DNA repair pathways are involved in protecting the integrity of genetic information explains the great interest in the study of these pathways. In general, cells deficient in DNA repair mechanisms are more susceptible to mutagenesis, which can predispose organisms, which carry these cells, to diseases such as cancers.

DNA repair enzymes are divided into five groups, which are: i) damage reversion repair (DRR), in which only one enzyme is necessary to directly revert the damage; ii) mismatch repair (MMR), in which the enzymes involved are responsible for correction of mismatched base pairs, principally mismatches induced by replication errors; iii) recombination repair or double strand breaks (DSB) repair, mediated by enzymes involved with homologous recombination, which uses the information on the undamaged sister chromatin or homologous regions to bypass lesions, including DNA strand breaks; iv) nucleotide excision repair (NER), involves enzymes which are able to remove several

types of DNA damage, especially damage that induce large double helix distortion. This pathway involves the excision of an oligonucleotide containing the lesion and posterior repair by DNA synthesis; v) base excision repair (BER) involves the action of glycosylases which cleave the glycosylic bond between the specific damaged base and the deoxyribose, with subsequent incision by AP-endonuclease (apurinic/apyrimidinic endonuclease) at the resultant abasic site. A single or a few nucleotides are synthesized using the undamaged strand as a template. A review of DNA pathways can be found in Eisen and Hanawalt (1999).

In plants, our knowledge of DNA repair pathways is still very poor. The first plant repair genes were only cloned in 1997 and the mechanisms of action of their products have yet to be characterized. In general, the few plant genes known to be involved in DNA repair are from *Arabidopsis thaliana*, and have shown homology with genes related to different repair pathways in other eukaryotes. For example, the *AtXPB* (the letters *At* indicate *A. thaliana*) gene is homologous to the human *XPB* gene, which is involved in DNA repair and transcription (Ribeiro *et al.*, 1998). The *A. thaliana thi1* gene is homologous to the yeast *thi4* gene and seems to be involved in thiamine biosynthesis and mitochondrial DNA damage tolerance (Machado *et al.*, 1996; Machado *et al.*, 1997). The *AtMMH* gene is homologous to the *E. coli mutM* gene, which encodes for a glycosylase that recognizes some types of oxidative deoxyguanosine dam-

age (Ohtsubo *et al.*, 1998). The *A. thaliana* *UVR3* gene shows homology with 6-4 photolyase genes from *Drosophila melanogaster* and *Xenopus laevis*, which act in the repair of 6-4 photoproducts (Nakajima *et al.*, 1998). Two others *A. thaliana* genes, homologous to the human *XPF* gene (encoding an endonuclease involved in NER and recombination) and the *MutS* gene (involved in MMR) respectively, have also been described (Gallego *et al.*, 2000; Ade *et al.*, 1999).

Genome projects such as the Sugarcane Expressed Sequence Tag (SUCEST) project allow the possibility of rapidly identifying plant DNA repair genes. In fact, the high conservation observed between DNA repair genes in eukaryotes has facilitated the identification of genes in different organisms. In the data mining work described in this paper our main interest was to investigate the occurrence of sugarcane BER genes. This DNA repair pathway is mainly involved with the correction of base damage that causes relatively minor distortions to the helical DNA structure, such as deaminated, oxidized, alkylated or even absent bases (for reviews see Krokan *et al.*, 2000; Memisoblu and Samson, 2000; Eisen and Hanawalt, 1999). In general, the inactivation of BER genes leads to a spontaneous mutator phenotype, and their relationships with carcinogenesis and aging have been investigated in several organisms (Boiteux and Radicella, 2000; Boiteux and Radicella, 1999; Kvaloy *et al.*, 2001).

Our search for sugarcane BER genes was performed using the tblastn (blast = basic local alignment search tool) program and the sequences of BER proteins were obtained directly from GenBank and used as probes to identify homologous nucleotide sequences in the SUCEST database. A high degree of similarity was observed for the majority of the investigated proteins, indicating the high conservation of BER genes.

MATERIAL AND METHODS

The sequence of BER proteins, from bacteria, yeast, human and plant genes were obtained directly from the GenBank (<http://www.ncbi.nlm.nih.gov>). The sequences of these proteins were used to identify homologous nucleotide sequences in the SUCEST database (<http://www.fapesp.br>). Initially, the analysis were performed by the tblastn program. The nucleotide sequence of clusters identified in the SUCEST database were submitted to GenBank and analyzed by the blastx program to confirm the homology with BER proteins. For some BER proteins, more than one homologous cluster was identified and in these cases, the blastn program was used to compare the nucleotide sequences of these clusters with the objective of identifying possible alleles or gene duplication.

RESULTS

We extracted from the GenBank database the sequences of several BER proteins, mainly glycosylases and

AP endonucleases, from *Arabidopsis thaliana* (P), *Escherichia coli* (B), *Saccharomyces cerevisiae* (Y), *Schizosaccharomyces pombe* (Y*) and *Homo sapiens* (H). These protein sequences were used to screen clusters and reads in the SUCEST database, and the results of this screening are given in Table I. As expected, the best e-values were obtained with BER proteins from *A. thaliana*, a high conservation (e-value $\leq e^{-15}$) was observed for the majority of the proteins investigated. These results were confirmed when the nucleotide sequences of the clusters were submitted to GenBank and analyzed with the blastx program, from which we obtained similar e-values for all the clusters identified in the SUCEST database (data not shown). No homology for endonuclease IV, UVDE and TDG was found in sugarcane, nor has any been found in *A. thaliana* and other plants, raising the possibility that these proteins are not present in the plant kingdom.

The number of reads that constitute each cluster and their occurrence in different cDNA libraries used in this EST genome project are presented in Table II. These results are important because they imply that these genes are expressed in the tissues from which the libraries were constructed. Some BER proteins (*e.g.* the MAG homologue) are widely distributed while others (*e.g.* the UNG homologue) have a restricted distribution.

Table III shows the results obtained when the nucleotide sequence of homologous clusters to the same BER protein were compared using the blastn program. Some clusters presented high similarity between their nucleotide sequences. For example, for the atMMH1 clusters SCPRRT3027A01.g and SCSGAM2105C07.g we obtained an e-value of 0.0, suggesting the probable occurrence of distinct alleles or alternative splicing products. However, other clusters showed no nucleotide similarity, even presenting overlapping alignment with the BER protein, suggesting the presence of different genes, as observed for the ARP homologues.

DISCUSSION

Glycosylases are enzymes that recognize specific damaged bases. There are at least seven different known glycosylases specialized in recognizing specific substrates that initiate the BER pathway. Some glycosylases act also as 3' AP endonucleases (or β -lyase) (Memisoglu and Samson, 2000). Uracil DNA glycosylase (UNG or UDG) is the main enzyme involved in uracil incision. In DNA, uracil results from cytosine deamination that gives rise to guanine:uracil (G:U) mismatches, which can lead to guanine:cytosine to adenine:thymine (G:C to A:T) transition mutations during replication. Deoxyuridine monophosphate (dUMP) can be incorporated instead of deoxythymidine monophosphate (dTMP), which may alter binding of transcription factors to their targets (Verri *et al.*, 1990). UNG homologues have been characterized in many bacterial and eukaryotic species and these proteins have

Table I - Similarity obtained between BER proteins and clusters from SUCEST bank.

GenBank access	Protein ^a (aminoacids)	Function	SUCEST clusters ^b	e-value	Cluster length (bp) ^c	Query alignment (aa) ^d
GI:9294324	UNG-P (330)	Uracil-DNA glycosylase-like protein	SCEQFL5048B07.g SCRFFL5039B01.g	8e ⁻⁴² 1e ⁻²⁴	726 804	4-253 233-311
GI:5903053	AtMMH1-P (397)	Formamidopyrimidina DNA glycosylase	SCCCLR2C01B12.g SCSGAM2105C07.g SCEQRT2094G03.g SCPRRT3027A01.g	e ⁻¹¹² 4e ⁻⁹³ 3e ⁻⁶⁰ 2e ⁻⁵⁴	1380 737 654 563	7-292 8-230 85-229 101-244
GI:5263323	OGG1-P (391)	Putative 8-oxoguanine DNA glycosylase, human OGG1 homolog	SCVPRZ2036B05.g	2e ⁻⁶⁶	715	99-292
GI:11181952	NTH1-P (354)	Thymine glycol glycosylase	SCACSB1036F12.g	1e ⁻⁸⁴	1266	36-275
GI:4586250	MUTY-P (608)	Adenine DNA glycosylase like protein	No hit	-	-	-
GI:2833373	MAG1-P (254)	Methyladenine DNA glycosylase, human MPG homolog	No hit	-	-	-
GI:9758366	MAG2-P (347)	Putative methyladenine DNA glycosylase, <i>E. coli</i> TagI homolog	SCRFL4026A11.g SCACLB1046E02.g SCMCRT2104F02.g SCSGST1072A10.b SCCCRZ3003C09.g SCRLLV1025A10.g	2e ⁻⁷⁷ 3e ⁻⁵⁹ 8e ⁻⁵⁵ 1e ⁻⁴⁸ 3e ⁻⁴⁵ 2e ⁻³⁵	647 1291 1235 711 1294 659	140-319 155-345 112-339 147-273 179-345 163-259
GI:10092274	MAG3-P (428)	Putative methyladenine DNA glycosylase <i>E. coli</i> AlkA homolog	SCJLRT2049G09.g	2e ⁻⁷⁸	1363	1-333
GI:8488963	ARP-P (536)	AP endonuclease redox pro- tein, exonuclease III homologue	SCEZAM2032F04.g SCBFAD1046D12.b SCBFAD1046D12.g SCBFRZ2048B12.g SCEPFL3080F10.g	1e ⁻⁹⁶ 9e ⁻⁵⁴ 5e ⁻¹⁰ 2e ⁻²⁴ 8e ⁻¹²	717 816 750 1158 627	244-466 369-536 244-287 336-534 354-400
GI:2506194	ENDOIV-B (285)	Endonuclease IV	No hit	-	-	-
GI:543825	APN1-Y (367)	AP endonuclease, endonucleas e IV homolog	No hit	-	-	-
GI:1905972	UVDE-Y* (599)	UV damage endonuclease	No hit	-	-	-
GI: 13648121	TDG-H (410)	Thymine DNA glycosylase	No hit	-	-	-

a- P = proteins from *Arabidopsis thaliana*; B = protein from *Escherichia coli*; Y = protein from *Saccharomyces cerevisiae*; Y* = protein from *Schizosaccharomyces pombe*; H = protein from *Homo sapiens*.

b- The letters g and b represent the sequencing from 5' and 3' terminals respectively.

c- bp-base pairs;

d- Query aminoacids region.

strikingly similar structures and functions (Pearl, 2000). In our work, we identified two distinct clusters (Table I) with high similarity to UNG proteins only. One cluster has 726 bp and aligns best with the N-terminal protein region, whereas the other cluster has 804 bp and aligns best with the UNG C-terminal region. Of the UNG genes described in the literature, only mammals have two UNG versions, one nuclear (UNG2) and one mitochondrial (UNG1), which are differentially expressed in tissues and induced during the cell division (Slupphaug *et al.*, 1991; Nilsen *et al.*, 1997; Nilsen *et al.*, 2000). The two forms of human UNG differ in their N-terminals, UNG2 having a specific amino acid sequence which is not present in UNG1

(Otterlei *et al.*, 1998). The smaller sugarcane cluster shows a higher similarity with UNG2 (e-value 3e⁻¹⁹) than UNG1 (e-value 5e⁻¹⁸) and the large cluster has the same e-value (1e⁻²⁰) for both UNG forms. Each cluster has only one read, which are from of the same flower cDNA library (Table II). The sequence of a single *A. thaliana* UNG gene was obtained through its genome project. It thus seems that sugarcane may be a good model to study UNG activity in plants.

The formamidopyrimidine DNA glycosylase (FPG or MutM) plays a fundamental role in the repair of oxidative DNA damage by recognizing oxidized purines such as 7,8-dihydro-8-deoxyguanine (8-oxo-dG) and imidazol ring-opened purines (FAPY-G). The deficiency of this

Table II - Occurrence of the probable sugarcane BER proteins in the SUCEST cDNA libraries.

Protein ^a	SUCEST clusters	Number of reads	Best read e-value	cDNA libraries
UNG-P	SCEQFL5048B07.g	1	3e ⁻⁴¹	Flower
	SCRFFL5039B01.g	1	4e ⁻²⁴	Flower
AtMMH1-P	SCCCLR2C01B12.g	4	1e ⁻⁸³	Appical Meristem, flower, and root
	SCSGAM2105C07.g	8	2e ⁻⁹²	Appical Meristem, flower, stem and plantlet
	SCEQRT2094G03.g	2	1e ⁻⁶⁰	Leaf-root transition zone and root
	SCPRRT3027A01.g	1	6e ⁻⁵⁴	Root
OGG1-P	SCVPRZ2036B05.g	2	1e ⁻⁷⁷	leaf-root transition zone and flower
NTH1-P	SCACSB1036F12.g	7	1e ⁻⁶⁰	Appical meristem, plantlet, stem bark and leaves
ARP-P	SCEZAM2032F04.g	4	7e ⁻⁹²	Appical meristem and flower
	SCBFAD1046D12.b	1	3e ⁻⁵³	
	SCBFAD1046D12.g	1	2e ⁻⁰⁹	Plantlet
	SCBFRZ2048B12.g	1	3e ⁻¹⁶	Leaves
	SCEPFL3080F10.g	1	3e ⁻¹¹	Flower
MAG2-P	SCRLLV1025A10.g	2	1e ⁻⁸²	Flower
	SCACLB1046E02.g	13	9e ⁻⁴⁸	Appical meristem, plantlet, stem bark, leaves, stem, root, flower, lateral bud and leaf-root transition zone
	SCMCRT2104F02.g	10	1e ⁻⁴⁴	Plantlet, leaves, root, flower, and leaf-root transition zone
	SCSGST1072A10.b	1	5e ⁻⁴⁸	Stem
	SCCCRZ3003C09.g	9	1e ⁻⁴⁴	Plantlet, stem bark, stem, flower, and leaf-root transition zone
	SCRLLV1025A10.g	1	9e ⁻³⁵	Leaves
MAG3-P	SCJLRT2049G09.g	5	2e ⁻⁷¹	Root and flower

Table III - Analyze of the BER SUCEST clusters in the blastn program.

Protein ^a	Cluster x cluster (blastn)	e-value	Comments
UNG-P	SCRFFL5039B01.g x SCEQFL5048B07.g	6e ⁻⁰⁹	Probably two genes?
AtMMH1-P	SCCCLR2C01B12.g x SCSGAM2105C07.g	4e ⁻⁰⁸	Probably two genes with two alleles or splicing products each one
	SCCCLR2C01B12.g x SCEQRT2094G03.g	e ⁻¹⁶⁹	
	SCCCLR2C01B12.g x SCPRRT3027A01.g	2e ⁻⁰⁴	
	SCPRRT3027A01.g x SCSGAM2105C07.g	0.0	
	SCPRRT3027A01.g x SCEQRT2094G03.g	5e ⁻¹⁵	
OGG1-P	Only one cluster	-	Probably one gene
NTH1-P	Only one cluster	-	Probably one gene
ARP-P	SCEPFL3080F10.g x SCEZAM2032F04.g	3e ⁻⁵⁶	The other clusters do not show any similarity, probably four genes
MAG2-P	SCACLB1046E02.g x SCCCZR3003C09.g	0.0	The other clusters do not show any similarity, probably six genes, one of them with two alleles
MAG3-P	Only one cluster	-	Probably one gene

BER enzyme in bacteria causes a mutator phenotype with an increase of transversion mutation involving G:C to T:A transversions due to the ability of 8-oxo-dG to pair with 2-deoxyadenosine (Boiteux and Radicella, 1999). The MutM proteins have been characterized in many bacterial strains and are very conserved. In eukaryotes, the only *mutM* homologous gene known has been found in *A. thalia-*

na (AtMMH). This gene encodes two types of mRNA (AtMMH1 and AtMMH2), formed by alternative splicing of exon 8 (Ohtsubo *et al.*, 1998). In the SUCEST database we found four clusters presenting high similarity with MutM proteins (Table I). The best e-values were observed with AtMMH1. The clusters SCCCLR2C01B12.g and SCSGAM2105C07.g showed a similar alignment with the

MutM proteins, but when their nucleotide sequence were compared using the blastn program (Table III), a good e-value was not observed ($4e^{-08}$), which suggests the probable occurrence of two different genes. In contrast, we obtained a high similarity between the clusters SCCCLR2C01B12.g and SCEQRT2094G03.g (e-value = e^{-169}). The same thing occurred with the clusters SCPRT3027A01.g and SCSGAM2105C07.g (e-value = 0.0). From these results we cannot conclude whether or not sugarcane has a similar splicing pattern to that found in *A. thaliana*, or if these clusters represent alleles. But it does seem that sugarcane may have two different expressed mutM homologues.

An 8-oxo-dG glycosylase (OGG1) with functional analog activity to MutM has been characterized in yeasts and mammals (Boiteux and Radicella, 1999). In *A. thaliana*, a nucleotide sequence homologous to the OGG1 gene was found (by *A. thaliana* genome project), although functional assays were not performed. A cluster presenting high similarity with OGG1 has been identified by us in the SUCEST database (Table I). Taken together, these data suggest that both MutM and OGG1 are present in plants. Eisen and Hanawalt (1999) have proposed that OGG1 arose in the common ancestor of eukaryotes and that the MutM found in plants is probably derived from the chloroplast genome. Our results also show a wide distribution of MutM in different cDNA libraries and a high number of reads (Table II), while OGG1 occurred in only two cDNA libraries, with only two reads in one cluster (Table II). In the human genome, seven OGG1 isoforms have been observed due to the alternative splicing, and the expression of this gene is both independent of the cell cycle and is not tissue-specific (Nishioka *et al.*, 1999; Bouziane *et al.*, 2000; Radicella *et al.*, 1997). In sugarcane, it is possible to suggest that the MutM protein could be ubiquitous and the OGG1 gene could be targeted to certain tissues by specific control mechanisms.

Endonuclease III (NTH) is one of the most important enzymes in the repair of oxidative damage in pyrimidines, being its major substrate thymine glycol. NTH has been characterized in several organisms and shows a high degree of conservation (Cadet *et al.*, 2000). After the screening of the SUCEST database for NTH protein, we identified one cluster with a high similarity (Table I). We also observed a large range of distribution in the cDNA libraries (Table II). These data are similar to that obtained for the NTH of *A. thaliana*, which is expressed in different plant tissues (Rodan-Arjona *et al.*, 2000). Plants seem to have different expression patterns for this gene when compared with other eukaryotes. In human cells, the expression of NTH is cell-cycle dependent (Bouziane *et al.*, 2000) and this protein has mitochondrial target sequence and nuclear localization signal (Takao *et al.*, 1998). Two NTH homologous genes have been found in yeasts (NTG1 and NTG2). The NTG1 gene is inducible by exposing cells to DNA-damag-

ing agents and its product is present in the nucleus and mitochondria. The NTG2 gene is constitutively expressed and its product is present only in the nucleus (Alseth *et al.*, 1999). It seems that a single locus of NTH homolog is found in sugarcane.

MutY-glycosylase is another important enzyme for the repair of oxidative DNA damage, removing deoxyadenine when paired with 8-oxo-dG, deoxyguanosine or deoxycytosine. Deficiency in this enzyme leads to an increase in G:C to T:A transversions (Boiteux and Radicella, 1999). Although a nucleotide MutY-like sequence from *A. thaliana* is registered at GenBank, in our work no cluster showing good similarity with MutY proteins was found among the 250.000 ESTs sequences contained in the SUCEST database. This enzyme has been described in several organisms and its absence in sugarcane is unexpected, but a very low level of MutY expression could explain the absence of its cDNA in the SUCEST bank.

The enzyme 3-Methyladenine DNA glycosylase (MAG) recognizes and removes a wide variety of alkylated bases that include 3-methyladenine, 7-methyladenine, 3-methylguanine and 7-methylguanine. MAG genes have been cloned from a variety of bacteria and eukaryotes and can be divided into three gene families, all of which are represented in the *A. thaliana* genome. Sugarcane presents six clusters with homology to the TagI homologue from *A. thaliana* (MAG2 in the Table I), the reads of which are widely distributed in the cDNA libraries (Table II). The blastn analysis showed an e-value of 0.0 for the clusters SCACLB1046E02.g and SCCRZ3003C09.g (Table III), which suggest that they may be alleles. Although all the clusters show alignment overlapping with MAG2 protein, no nucleotide sequence similarity was obtained, suggesting the occurrence of five different genes. One cluster homologous to AlkA-like from *A. thaliana* (MAG3 in the Table I) was also found in sugarcane but its reads had a limited distribution (Table II). Unlike *A. thaliana*, no MPG homologue (MAG1 in the Table I) was found in sugarcane. The presence of multiple copies of MAG genes and their expression in several tissues suggest that plants are frequently exposed to alkylating agents. Functional assays in *A. thaliana* demonstrated that its MAG protein, homologous to MPG from mammals, is preferentially expressed in meristematic tissues, developing embryo and endosperm and organ primordial. This pattern of expression is consistent with a requirement for expression in rapidly dividing tissues. Furthermore, high levels of MAG has also been observed in growing leaves, which undergo a relatively low rate of cell division, suggesting that MAG is required not only for DNA replication, but also for cell growth (Shi *et al.*, 1997).

Apurinic/apurimidinic endonucleases cleave the DNA backbone at sites where bases are missing. Two distinct families are known. One includes the XTH protein (exonuclease III) of *E. coli*, the RRP1 protein of *D.*

melanogaster, the HAP1/APE1 protein of mammals, the APN2 protein of yeast and ARP protein of *A. thaliana*. The other family includes the NFO protein (endonuclease IV) of *E. coli* and the APN1 protein of yeast. Some other proteins can serve as AP endonucleases, but usually these activities are included as glycosylases, as MutM and NTH which do not function as AP endonucleases on their own. In eukaryotes, the exonuclease III family has multifunctional roles, acting not only as AP endonuclease but also as a reduction-oxidation (redox) mediator, maintaining transcription factors (e.g. Fos and Jun) in an active reduced state (Memisoglu and Samson, 2000; Babiychuk *et al.*, 1994; Evans *et al.*, 2000). In the SUCEST database, we found four homologue clusters to proteins of the exonuclease III family, distributed in different cDNA libraries (Tables I and II). These were confirmed as exonuclease III family proteins when analyzed at GenBank. In addition, these clusters show alignment overlapping with the ARP protein, but strong divergence when compared between them (Table III). These data suggest that sugarcane may have different forms of exonuclease III-like proteins, a finding consistent with what has been found in other organisms (Hadi and Wilson, 2000; Bennett, 1999).

Sequences homologous to the endonuclease IV family were not found in sugarcane (Table I) nor have they been described in *A. thaliana*. In general, members of the exonuclease III family are found in most species, whereas members of endonuclease IV family have a more limited distribution. It has been postulated (Eisen and Hanawalt, 1999) that these gene families are ancient and that the absence of either gene from a particular species is likely to be due to gene loss. However, all species encode at least a homologue of one of the two AP endonucleases, so the loss of one of these is tolerated, but not the loss of both (Eisen and Hanawalt, 1999).

In conclusion, our results suggest a high degree of conservation for most of the BER genes in sugarcane. Furthermore, sugarcane seems to have different forms for some genes, such as UNG, MutM, MAG and ARP, which could be the result of gene duplication, alternative splicing or differential promoter utilization. The answers to these questions may be obtained through functional analysis. The mechanisms of DNA repair in plants are as yet poorly understood, there being very few functional assays described for plants and no work has been published for genes such as OGG1. Thus, genome projects like the SUCEST and *A. thaliana* projects can provide good models for investigating DNA repair in plants.

ACKNOWLEDGMENTS

We thank Dr. Carlos Frederico Martins Menck for his important contribution to this work, and the Fundação de Amparo a Pesquisa do Estado de São Paulo for the authorization to use the SUCEST database. This work was also

supported by the Secretaria da Indústria, do Comércio, da Ciência e da Tecnologia do Estado do Rio Grande do Norte.

RESUMO

Danos no DNA podem ser induzidos por um grande número de agentes físicos e químicos presentes no ambiente, como também por compostos produzidos pelo próprio metabolismo celular. Estes danos podem interferir com processos celulares como replicação e transcrição, levando a morte celular e/ou mutações. Os baixos níveis de mutação nas células são devidos à presença de vias enzimáticas, que reparam os danos no DNA. Diversos genes de reparo de DNA têm sido clonados e seus produtos caracterizados, principalmente em bactérias, leveduras e mamíferos. O interesse no estudo de mecanismos de reparo de DNA advém de seu envolvimento com a proteção da integridade da informação genética. A alta conservação observada para a maioria dos genes relacionados ao reparo de DNA, especialmente em eucariotos, aponta para sua importância para a manutenção da vida na terra. Em plantas, o conhecimento sobre os mecanismos de reparo de DNA é ainda reduzido. Os primeiros genes de reparo foram recentemente clonados e o mecanismo de ação de seus produtos está por ser caracterizado. Nosso objetivo neste trabalho de *data mining* foi identificar, no banco de dados gerados pelo projeto Genoma da Cana de Açúcar (*Sugarcane Expressed Tag Project-SUCEST*), genes relacionados ao reparo por excisão de bases (BER). Esta busca foi feita através do programa tblastn. Em cana de açúcar, foram identificados clusters homólogos para a maioria das proteínas BER analisadas e um alto grau de conservação foi observado. Os melhores resultados foram obtidos com proteínas BER de *Arabidopsis thaliana*. Para alguns homólogos BER de cana de açúcar, a presença de mais de uma forma de mRNA é possível, como definido pela ocorrência de mais de um cluster homólogo.

REFERENCES

- Ade, J., Belzile F., Philippe H. and Doutriaux M.P. (1999) Four mismatch repair paralogues coexist in *Arabidopsis thaliana*: AtMSH2, AtMSH3, AtMSH6-1 and AtMSH6-2. *Mol Gen Genet* 262: 239-249.
- Alseth, I., Eide, L., Pirovano, M., Rognes, E., Seeberg, E. and Bjoras, M. (1999) The *Saccharomyces cerevisiae* homologues of endonuclease III from *Escherichia coli*, Ntg1 and Ntg2, are both required for efficient repair of spontaneous and induced oxidative DNA damage in yeast. *Mol. Cell. Biol.* 19: 3779-3787.
- Babiychuk, E., Kushnir, S., Van Montagu, M. and Inzé, D. (1994) The *Arabidopsis thaliana* apurinic endonuclease Arp reduces human transcription factors Fos and Jun. *Proc. Natl. Acad. USA* 91: 3299-3303.
- Bennett, R.A.O. (1999) The *Saccharomyces cerevisiae* ETH1 gene, an inducible homolog of exonuclease III that provides resistance to DNA-damaging agents and limits spontaneous mutagenesis. *Mol. Cell. Biol.* 19: 1800-1809.

- Boiteux, S. and Radicella, J.P.** (1999) Base excision repair of 8-hydroxyguanine protects DNA from endogenous oxidative stress. *Biochimie* 81: 59-67.
- Boiteux, S. and Radicella, J.P.** (2000) The human OGG1 gene: structure, functions, and its implication in the process of carcinogenesis. *Arch. Biochem. Biophys.* 377: 1-8.
- Bouziane, M., Miao, F., Bates, S.E., Somsouk, L., Sang, B., Denissenko, M. and O'Connor, T.R.** (2000) Promoter structure and cell cycle dependent expression of the human methylpurine-DNA glycosylase gene. *Mutat. Res.* 461: 15-29.
- Cadet, J., Bourdat, A.G., D'Ham, C., Duarte, V., Gasparutto, D., Romieu, A. and Ravanat, J.L.** (2000) Oxidative base damage to DNA: specificity of base excision repair enzymes. *Mutat. Res.* 462: 121-128.
- Eisen, J.A. and Hanawalt, P.C.** (1999) A phylogenetic study of DNA repair genes, proteins, and processes. *Mutat. Res.* 435: 171-213.
- Evans, A.R., Limp-Foster, M. and Kelley, M.R.** (2000) Going APE over ref-1. *Mutat. Res.* 461: 83-108.
- Gallego, F., Fleck, O., Li, A., Wyrzykowska, J. and Tinland, B.** (2000) AtRAD1, a plant homologue of human and yeast nucleotide excision repair endonucleases, is involved in dark repair of UV damages and recombination. *Plant J.* 21: 507-518.
- Hadi, M.Z. and Wilson, D.M.** (2000) Second human protein with homology to the *Escherichia coli* abasic endonuclease exonuclease III. *Environ. Mol. Mutagen.* 36: 312-324.
- Krokan, H.E., Nilsen, H., Skorpen, F., Otterlei, M. and Slupphaug, G.** (2000) Base excision repair of DNA in mammalian cells. *FEBS Lett.* 476: 73-77.
- Kvaloy, K., Nilsen, H., Steinsbekk, K.S., Nedal, A., Montecrotti, B., Alkbari, M. and Krokan, H.E.** (2001) Sequence variation in the human uracil-DNA glycosylase (UNG) gene. *Mutat. Res.* 461: 325-338.
- Lu, A.L. and Fawcett, W.P.** (1998) Characterization of the recombinant MutY homolog, an adenine DNA glycosylase, from yeast *Saccharomyces pombe*. *J. Biol. Chem.* 273: 25098-25105.
- Machado, C.R., de Oliveira, R.C., Boiteux, S., Praekelt, U.M., Meacock, P.A. and Menck, C.F.** (1996) Thi1, a thiamine biosynthesis gene in *Arabidopsis thaliana*, complements bacterial defects in DNA repair. *Plant Mol Biol* 31: 585-593.
- Machado, C.R., Praekelt, U.M., de Oliveira, R.C., Barbosa, A.C., Byrne, K.L., Meacock, P.A. and Menck, C.F.** (1997) Dual role for yeast THI4 gene in thiamine biosynthesis and DNA damage tolerance. *J Mol Biol* 273: 114-121.
- Memisoglu, A. and Samson, L.** (2000) Base excision repair in yeast and mammals. *Mutat. Res.* 451: 39-51.
- Nakajima, S., Sugiyama, M., Iwai, S., Hitomi, K., Otsu, E., Kim, S.T., Jiang, C.Z., Todo, T., Britt, A.B. and Yamamoto, K.** (1998) Cloning and characterization of a gene (UVR3) required for photorepair of 6-4 photoproducts in *Arabidopsis thaliana*. *Nucleic Acids Res* 26: 638-644.
- Nilsen, H., Otterlei, M.; Haug, T., Solumm, K., Negelhus, T.A., Skorpen, F. and Krokan, H.E.** (1997) Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. *Nucleic Acids Res.* 25: 750-755.
- Nilsen, H., Steinsbekk, K.S., Otterlei, M.; Slupphaug, G., Aas, P.A. and Krokan, H.E.** (2000) Analyses of uracil-DNA glycosylases from murine Ung gene reveals differential expression in tissues and in embryonic development and a subcellular sorting pattern that differs from human homologues. *Nucleic Acids Res.* 28: 2277-2285.
- Nishioka, K., Ohtsubo, T., Oda, H., Fujiwara, T., Kang, D., Sugimachi, K. and Nakabeppu, Y.** (1999) Expression and differential intracellular localization of two major forms of human 8-oxoguanine DNA glycosylase encoded by alternative spliced OGG1 mRNAs. *Mol. Biol. Cell* 10: 1637-1652.
- Ohtsubo, T., Matsuda, O., Iba, K., Terashima, I., Sekiguchi, M. and Nakabeppu, Y.** (1998) Molecular cloning of *AtMMH*, an *Arabidopsis thaliana* ortholog of the *Escherichia coli mutM* gene, and analysis of functional domains of its product. *Mol Gen Genet* 259: 577-590.
- Otterlei, M.; Haug, T., Nagelhus, T.A., Slupphaug, G., Lindmo, T. and Krokan, H.E.** (1998) Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localization signal and a strong classical mitochondrial localization signal, respectively. *Nucleic Acids Res.* 26: 4611-4617.
- Pearl, L.** (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.* 460: 165-181.
- Radicella, J.P., Dherin, C., Desmaze, C., Fox, M.S. and Boiteux, S.** (1997) Cloning and characterization of *hOGG1*, a human homolog of *OGG1* gene *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 94: 8010-8015.
- Ribeiro, D.T., Machado, C.R., Costa, R.M., Praekelt, U.M., Van Sluys, M.A. and Menck, C.F.** (1998) Cloning of a cDNA from *Arabidopsis thaliana* homologous to the human XPB gene. *Gene* 208: 207-213.
- Rodan-Arjona, T., Garcia-Ortiz, M.V., Ruiz-Rubio, M. and Ariza, R.R.** (2000) cDNA cloning, expression and functional characterization of *Arabidopsis thaliana* homologue of the *Escherichia coli* DNA repair enzyme endonuclease III. *Plant. Mol. Biol.* 44: 43-52.
- Shi, L., Kent, R., Bence, N. and Britt, A.B.** (1997) Developmental expression of a DNA repair gene in *Arabidopsis*. *Mutat. Res.* 384: 145-156.
- Slupphaug, G., Olsen, L.C., Helland, D., Aaland, R. and Krokan, H.E.** (1991) Cell cycle regulation and *in vitro* hybrid arrest analysis of the major uracil-DNA glycosylase. *Nucleic Acids Res.* 19: 5131-5137.
- Takao, M., Aburatani, H., Kobayashi, K. and Yasui, A.** (1998) Mitochondrial targeting of human DNA glycosylases for repair of oxidative DNA damage. *Nucleic Acids Res.* 26: 2917-2922.
- Verri, A., Mazzarello, P., Biamonti, G., Spadari, S. and Focher, F.** (1990) The specific binding of nuclear proteins to the cAMP responsive element sequence is reduced by misincorporation of uracil and increased by the deamination of cytosine. *Nucleic Acids Res.* 18: 5775-5780.