Genetics and Molecular Biology, 28, 3 (suppl), 496-500 (2005)

www.sbg.org.br

Research Article

The post-transcriptional gene silencing pathway in *Eucalyptus*

Flávio Tetsuo Sassaki[1], Tiago Campos-Pereira[2] and Ivan de Godoy Maia[1]

[1]*Universidade Estadual Paulista 'Júlio de Mesquita Filho', Instituto de Biociências, Departamento de Genética, Botucatu, SP, Brazil.*
[2]*Universidade Estadual de Campinas, Departamento de Genética Médica, Campinas, SP, Brazil.*

Abstract

Post-transcriptional gene silencing (PTGS) is a conserved surveillance mechanism that identifies and cleaves double-stranded RNA molecules and their cellular cognate transcripts. The RNA silencing response is actually used as a powerful technique (named RNA interference) for potent and specific inhibition of gene expression in several organisms. To identify gene products in *Eucalyptus* sharing similarities with enzymes involved in the PTGS pathway, we queried the expressed sequence tag database of the Brazilian *Eucalyptus* Genome Sequence Project Consortium (FORESTs) with the amino acid sequences of known PTGS-related proteins. Among twenty-six prospected genes, our search detected fifteen assembled sequences encoding products presenting high level of similarity (E value < 10^{-40}) to proteins involved in PTGS in plants and other organisms. We conclude that most of the genes known to be involved in the PTGS pathway are represented in the FORESTs database.

Key words: gene silencing, PTGS, RNA silencing, EST, *Eucalyptus*.

Received: May 31, 2004; Accepted: March 17, 2005.

Introduction

Post-transcriptional gene silencing (PTGS) is a widely conserved surveillance system that acts at the transcriptome level. It identifies long double stranded RNA molecules (dsRNA) such as those generated during virus replication, transposons mobilization or aberrant RNA synthesis. These molecules act as trigger signals for the PTGS machinery, guiding the cell to an alert state.

Once identified, trigger dsRNAs are cleaved by a cellular dsRNA-specific endonuclease named Dicer (DCL for Dicer-like in plants; Bernstein *et al*., 2001; Zamore *et al*., 2000; Schauer *et al*., 2002) into ~21-23 nucleotide duplexes known as small interfering RNAs (siRNAs; Hamilton and Baulcombe, 1999). Somehow, a Dicer-associated protein (R2D2) "senses" thermodynamic free energy of siRNA 5' ends, binding to its more stable one (Tomari, *et al*., 2004). This process coordinates Dicer (re-) positioning over siRNA and duplex winding by an RNA helicase (SDE3 in plants; Dalmay *et al*., 2001) through its more unstable termini, allowing the loading of one strand into a multicomponent complex termed RISC (for RNA induced silencing complex; Hammond *et al*., 2000).

Send correspondence to Ivan de Godoy Maia. Universidade Estadual Paulista 'Júlio de Mesquita Filho', Instituto de Biociências, Departamento de Genética, 18618-000 Botucatu, SP, Brazil. E-mail: igmaia@ibb.unesp.br.

RISC is composed of several proteins of unknown function (Gemin 3 and 4, VIG and dFXR; Hutvagner and Zamore, 2002; Caudy *et al*., 2002) and by multiple nucleases such as Slicer (or Argonaute 2), the catalytic endonuclease responsible for mRNA target cleavage (Parker *et al*., 2004; Liu, *et al*., 2004; Song, *et al*., 2004; Meister, *et al*., 2004; Rand, *et al*., 2004), Tudor-SN, a micrococcal nuclease homologue (Caudy *et al*., 2003) and probably Wex, a 3'-5' RNase D-like protein (Glazov *et al*., 2003). Loaded siRNA guides RISC to the target RNA promoting its endonucleolytic cleavage and thus allowing 3'-5' (Wex) and 5'-3' (AtXRN4; Souret *et al*., 2004) exoribonucleases to act on the generated unprotected ends (Figure 1).

In plants and worms, the system is self-sustained and amplified by the activity of a siRNA-primed RNA-dependent RNA polymerase (RdRP or RDR; Mourrain *et al*., 2000; Dalmay *et al*., 2000). In certain cases, such as in silencing induced by dsRNA-replicating viruses, the requirement of an RdRP activity is bypassed (Dalmay *et al*., 2000). In *Caenorhabditis elegans*, local introduction of dsRNA leads to systemic silencing that is probably mediated by a transmembrane protein named SID-1 (Winston *et al*., 2002). The activity of a protein of unknown function with coiled-coil domains (SGS3) is also required for gene silencing (Mourrain *et al*., 2000).

PTGS can also affect the genome promoting alterations such as DNA methylation (through MET1; Finnegan

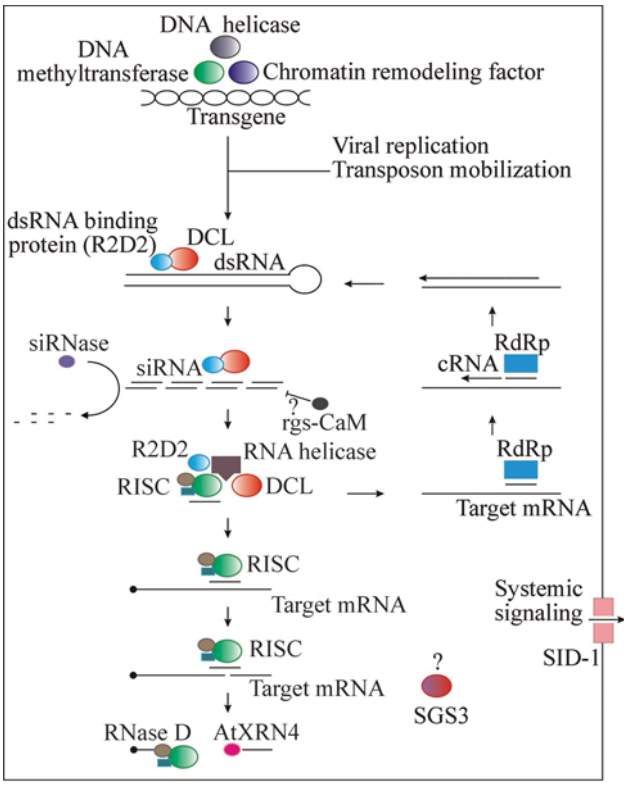

Figure 1 - Schematic representation of the molecular steps in PTGS pathway. Nuclear events at the loci of the target gene or transgene are promoted by MET1, DDM1 and QDE3 (a RecQ DNA helicase). PTGS is triggered by long dsRNA molecules that are converted by the ribonuclease III Dicer (or Dicer-like enzymes) into small interfering dsRNAs of 21-23 nucleotides (siRNA). One strand of siRNA is transferred from DCL to RISC, thus directing target RNA cleavage. Unprotected generated ends are targets for exoribonucleases (Rnase D and AtXRN4). A siRNA-primed RNA-dependent RNA polymerase (RdRP) synthesizes new dsRNA molecules from target RNA thus perpetuating the process. Sid-1 is a transmembrane protein required for systemic signaling in *C. elegans*.

et al., 1996) and chromatin remodeling (through DDM1; Jeddeloh *et al*., 1999) at loci homologous to the target RNA (Morel *et al*., 2000).

PTGS phenomenon, also called RNA silencing, is a potent means to counteract foreign sequences and its natural role in host defenses against transposable elements and viruses has been demonstrated. Consistent with this defense role, a great number of viral suppressors of RNA silencing have been identified so far (Voinnet *et* al., 1999). A gene encoding an endogenous RNA silencing suppressor was discovered in tobacco (named *rgs-CaM*; Anandalakshmi *et al*., 2000) and was proposed to be involved in endogenous gene regulation.

RNA silencing has been observed in plants, protozoans, insects, fungi (named quelling) and mammals (named RNA interference; RNAi), indicating common aspects and functional analogy among pathways. RNAi is actually used as an experimental approach to suppress the expression of specific transcripts, thus becoming a powerful tool to dissect gene function and a promising therapeutic technology.

In this paper we provide an inventory of the putative genes involved in the PTGS pathway of *Eucalyptus* by exploring the expressed sequence tag (EST) information generated in the Brazilian *Eucalyptus* Genome Sequence Project Consortium (FORESTs). Enhanced understanding of this pathway will contribute to the establishment of RNAi-based protocols intending functional genomic studies in this important tree.

Materials and Methods

To identify gene products sharing similarities with enzymes involved in the PTGS pathway (Figure 1), similarity searches against cluster consensus sequences (contigs)

Table 1 - PTGS-related proteins and homologues identified in *Eucalyptus*.

Number	Name		Organism	Function	Access	Clusters	e-value	Identity	Coverage
1		MET1	*Zea mays*	DNA methyltransferase	AAG15406	EGSBSL4219H01.g	2e-11	55%	8%
2		DDM1	*Arabidopsis thaliana*	Chromatin remodeling factor	At5g66750	EGMCST7267G07.g	e-150	72%	44%
3		QDE-3	*Neurospora crassa*	RecQ DNA Helicase	AAF31695	EGCCSL4031H03.g	2e-54	50%	11%
4	Argonaute family	AGO.1	*Arabidopsis thaliana*	Involved in siRNA and miRNA pathways	At1g48410	EGEQST6003A06.g EGEQRT3002E10.g	0.0 0.0	86% 76%	43% 78%
5		AGO.2	*Drosophila melanogaster*	RISC component	NP_730054	EGEQST2200C06.g EGEQSL5001E09.g	0.0 4e-89	61% 63%	50% 25%
6		AGO.4	*Arabidopsis thaliana*	Control of siRNA accumulation and DNA methylation	At2g27040	EGJEST2008F12.g EGBMRT6278A01.g EGJFFB1207F01.g EGMCRT6006C01.g EGEQFB1034D02.g	8e-53 e-130 3e-70 e-153 1e-75	44% 67% 83% 73%	23% 35% 16% 38% 23%
7		SGS3	*Arabidopsis thaliana*	Unknown	At5g23570	EGQHST6221A03.g EGJFLV2207H03.g	1e-49 3e-42	41% 41%	46% 33%
8		SDE3	*Arabidopsis thaliana*	RNA helicase	At1g05460	EGEZWD2227F09.g	2e-77	58%	23%
		MUT6	*Chlamydomonas reinhardtii*		AF305070	EGJFRT3163C04.g EGACLV3290A02.g	1e-86 2e-64	71% 61%	16% 10%

Table 1 (cont).

Number	Name	Organism	Function	Access	Clusters	e-value	Identity	Coverage
9	DCL1	*Arabidopsis thaliana*	dsRNA-specific endo-nuclease (RNase III)	At1g01040	EGEQWD2256B05.g	e-139	72%	18%
10	DCL2			At3g03300	EGEZWD2267C04.g	2e-71	56%	18%
11	DCL3			At3g43920	EGEPST6161G12.g	e-139	60%	27%
					EGACSL5244C02.g	2e-50	55%	12%
12	DCL4			At5g20320	EGSBCL1312G08.g	2e-63	57%	13%
13	rgs-CaM	*Nicotiana tabacum*	Supressor of PTGS	AAK11255	EGEZRT5203G11.g	2e-41	62%	72%
14	RDR6	*Arabidopsis thaliana*	RNA-dependent RNA polymerase	At3g49500	EGRFST6265B04.g	2e-92	61%	22%
					EGEZST2003F10.g	2e-77	37%	39%
					EGQHCL1342B05.g	3e-56	54%	18%
15	AtXRN4	*Arabidopsis thaliana*	5'-3' exoribonuclease	At1g54490	EGBGRT6263B05.g	1e-80	60%	26%
					EGEZCL1262C10.g	2e-72		
					EGBGWD2290C10.g	7e-42		

in the FORESTs database (https://forests.esalq.usp.br) were performed using BLAST search programs (Altschul *et al*., 1990). The database was first prospected using query sequences from known PTGS-related proteins (Table 1) and the TBLASTN algorithm. Positive hits (e-values < 10^{-40}) were validated against existing homologous sequences in the GenBank nonredundant protein database by using BLASTX. To further validate annotation, a protein domain analysis was performed using the Reverse Position Specific-BLAST (RPS-BLAST) algorithm at www.ncbi.nlm.nih.gov/structure/cdd/wrpsb.cgi, Pfam (Bateman *et al*., 2000) and SMART (Schultz *et al*., 2000) databases. An estimate of the relative abundance of the identified putative genes was generated based on EST counts per corresponding contig.

Results and Discussion

The FORESTs database was mined for *Eucalyptus* gene products potentially involved in the PTGS pathway (Figure 1). We have searched for genes encoding 26 PTGS-related proteins and have found matches for 15 of them (Table 1). Among the identified EST clusters, the one annotated as MET1 homologue exhibited a low e-value (2e-11) but contained a highly conserved methyltransferase domain, indicating that it probably encodes a DNA methyltransferase. Searches also revealed several clusters with significant sequence and domain similarities to the four *Arabidopsis* homologues of Dicer (DCL1 to 4) and four clusters encoding RDR6 (also known as SDE1/SGS2).

Using the amino acid sequences of AGO-1, AGO-2 and AGO-4 as queries, nine clusters encoding Argonaute proteins were identified. AGO proteins are a common component of all RISC-related complexes and are defined by the presence of two conserved domains: a ~20 kDa N-terminal PAZ domain and a ~40 kDa C-terminal PIWI domain, which is in fact a (cryptic) RNase H domain responsible for RNA target cleavage (Song *et al*., 2004). At least ten AGO homologues have been identified in the *Arabidopsis* genome. Among them, AGO-1 is involved in the siRNA and microRNA (miRNA) pathways (Vaucheret *et al*., 2004) while AGO-4 plays a role in locus-specific siRNA accumulation and RNA-directed DNA methylation (Zilberman, *et al*., 2003; Zilberman, *et al*., 2004). In *Drosophila*, *Homo sapiens* and *Mus musculus*, AGO-2 mediates target mRNA cleavage (Okamura *et al*., 2004, Meister, *et al*., 2004; Liu, *et al*., 2004). Since these proteins are closely related to each other, and biochemical characterization of AGO-2 in plants is still missing, the correct assignment of the annotated *Eucalyptus* clusters proved to be difficult without further analyses.

We were also able to assign three clusters to AtXRN4, two clusters to SGS3, one cluster to SDE3 and a cluster to rgs-CaM (2e-41), a calmodulin-related protein that acts as a cellular suppressor of PTGS (Anandalakshmi *et al*., 2000).

In contrast, BLAST searches revealed no gene product with significant similarity to RDE-4 (R2D2 in *Drosophila*), a dsRNA binding protein that interacts with RNA molecules identical to the trigger dsRNA during RNAi in *C. elegans* (Tabara *et al*., 2002). Likewise no HEN1-related clusters could be found in the FORESTs database. HEN1 is a protein of 942 amino acids playing a role

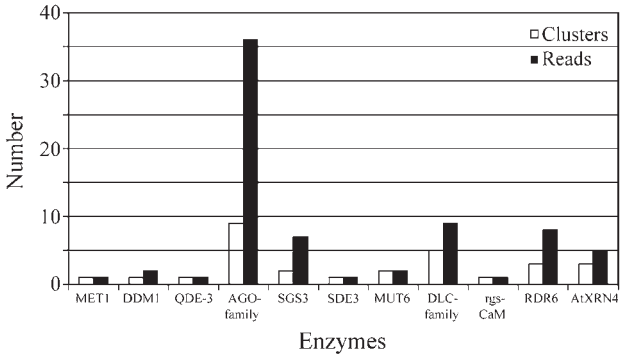

Figure 2 - Cluster distribution and total number of ESTs (in each cluster) encoding enzymes involved in the PTGS pathway.

in siRNA and miRNA accumulation in plants (Boutet *et al*., 2003). We do believe, however, that this protein is present in *Eucalyptus* but is not represented in the FORESTs database. Moreover, we did not find any EST clusters similar to proteins that have been genetically or biochemically linked to the PTGS pathway in humans (Gemin 3 and 4 of unknown function), *Drosophila* (RNA-binding proteins VIG and dFXR associated to RISC) and *C. elegans* (MUT-7 RNase D also known as WEX in *Arabidopsis*; ERI-1, a siRNA-degrading RNase; Kennedy *et al*., 2004).

An estimate of the relative abundance of the identified genes was obtained by comparing the number of times ESTs were assigned to a particular contig. An overall view of the obtained results shows that the enzymes involved in PTGS are poorly represented in the FORESTs database (Figure 2). In general, transcripts encoding proteins of the AGO family, DCL family and RDR6 figured as the most extensively represented.

In summary, several EST clusters could be assigned to proteins of known functions in PTGS in plants and other organisms, covering almost the entire pathway (Figure 1 and Table 1). The identification of components of the PTGS pathway within the generated set of *Eucalyptus* ESTs provides good evidence for the conservation of the RNA silencing mechanism among different plant species and organisms. Recent findings suggest that RNA silencing and related pathways are involved in several cellular processes such as defense, RNA surveillance and development. Elucidation of the molecular mechanisms underlying these processes is an important step towards the full understanding of the PTGS phenomenon. From our analyses, we can conclude that *Eucalyptus* encodes a functional RNA silencing pathway. This collection of PTGS-related ESTs provides an interesting resource for molecular and functional genomic studies in this important tree.

Acknowledgements

The authors are grateful to the *Eucalyptus* Consortium for the computer facilities and to FAPESP for financial support. T.C-P. is supported by FAPESP and I.G.M. is a recipient of a research fellowship from CNPq.

References

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search toll. J Mol Biol 215:403-410.

Anandalakshmi R, Marathe R, Ge X, Herr JM, Mau C, Mallory A, Pruss G, Bowman L and Vance VB (2000) A calmodulin-related protein that suppresses posttranscriptional gene silencing in plants. Science 290:142-144.

Bateman A, Birney E, Durbin R, Eddy SR, Howe KL and Sonnhammer EL (2000) The Pfam protein families database. Nucleic Acids Res 28:263-266.

Bernstein E, Caudy AA, Hammond SM and Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409:363-366.

Boutet S, Vazquez F, Liu J, Beclin C, Fagard M, Gratias A, Morel JB, Crete P, Chen X and Vaucheret H (2003) A genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. Curr Biol 13:843-848.

Caudy AA, Myers M, Hannon GJ and Hammond SM (2002) Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev 16:2491-2496.

Caudy AA, Ketting RF, Hammond SM, Denli AM, Bathoorn AM, Tops BB, Silva JM, Myers MM, Hannon GJ and Plasterk RH (2003) A micrococcal nuclease homologue in RNAi effector complexes. Nature 425:411-414.

Dalmay T, Hamilton A, Rudd S, Angell S and Baulcombe DC (2000) An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101:543-53.

Dalmay T, Horsefield R, Braunstein TH and Baulcombe DC (2001) SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. EMBO J 20:2069-2078.

Finnegan EJ, Peacock WJ and Dennis ES (1996) Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. Proc Natl Acad Sci USA 93:8449-8454.

Glazov E, Phillips K, Budziszewski GJ, Meins F Jr and Levin JZ (2003) A gene encoding an RNase D exonuclease-like protein is required for post-transcriptional silencing in *Arabidopsis*. Plant J 35:342-349.

Hamilton AJ and Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286:950-952.

Hammond SM, Bernstein E, Beach D and Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. Nature 404:293-296.

Hutvagner G and Zamore PD (2002) A microRNA in a multiple-turnover RNAi enzyme complex. Science 297:2056-2060.

Jeddeloh JA, Stokes TL and Richards EJ (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet 22:94-97.

Kennedy S, Wang D and Ruvkun G (2004) A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. Nature 427:645-649.

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L and Hannon GJ (2004) Argonaute2 is the catalytic engine of mammalian RNAi. Science 305:1437-1441.

Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G and Tuschl T (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell 15:185-197.

Morel JB, Mourrain P, Beclin C and Vaucheret H (2000) DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in *Arabidopsis*. Curr Biol 10:1591-1594.

Mourrain P, Beclin C, Elmayan T, Feuerbach F, Godon C, Morel JB, Jouette D, Lacombe AM, Nikic S, Picault N, Remoue K, Sanial M, Vo TA and Vaucheret H (2000) *Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell 101:533-542.

Okamura K, Ishizuka A, Siomi H and Siomi MC (2004) Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev 18:1655-1666.

Parker JS, Roe SM and Barford D (2004) Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J 23:4727-4737.

Rand TA, Ginalski K, Grishin NV and Wang X (2004) Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. Proc Natl Acad Sci USA 101:14385-14389.

Schauer SE, Jacobsen SE, Meinke DW and Ray A (2002) *DICER-LIKE1*: Blind men and elephants in *Arabidopsis* development. Trends Plant Sci 7:487-491.

Schultz J, Copley RR, Doerks T, Ponting CP and Bork P (2000) SMART: A web-based tool for the study of genetically mobile domains. Nucleic Acids Res 28:231-234.

Song JJ, Smith SK, Hannon GJ and Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305:1434-1437.

Souret FF, Kastenmayer JP and Green PJ (2004) AtXRN4 degrades mRNA in *Arabidopsis* and its substrates include selected miRNA targets. Mol Cell 15:173-183.

Tabara H, Yigit E, Siomi H and Mello CC (2002) The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in *C. elegans*. Cell 109:861-871.

Tang G, Reinhart BJ, Bartel DP and Zamore PD (2003) A biochemical framework for RNA silencing in plants. Genes Dev 17:49-63.

Tomari Y, Matranga C, Haley B, Martinez N and Zamore PD (2004) A protein sensor for siRNA asymmetry. Science 306:1377-1380.

Vaucheret H, Vazquez F, Crete P and Bartel DP (2004) The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev 18:1187-1197.

Voinnet O, Pinto VM and Baulcombe DC (1999) Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. Proc Natl Acad Sci USA 96:14147-14152.

Winston WM, Molodowitch C and Hunter CP (2002) Systemic RNAi in *C. elegans* requires the putative transmembrane protein SID-1. Science 295:2456-2459.

Zamore PD, Tuschl T, Sharp PA and Bartel DP (2000) RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101:25-33.

Zilberman D, Cao X and Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science 299:716-719.

Zilberman D, Cao X, Johansen LK, Xie Z, Carrington JC and Jacobsen SE (2004) Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr Biol 14:1214-1220.

Associate Editor: Carlos F. M. Menck