Food and Chemical Toxicology 47 (2009) 353–360

Contents lists available at ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Endogenous small RNAs in grain: Semi-quantification and sequence homology to human and animal genes

Sergey I. Ivashuta, Jay S. Petrick *, Sara E. Heisel, Yuanji Zhang, Liang Guo, Tracey L. Reynolds, James F. Rice, Edwards Allen, James K. Roberts

Monsanto Company, 800 N. Lindbergh Blvd., Mail Code O3F, St. Louis, MO 63167, USA

ARTICLE INFO

Article history:
Received 8 July 2008
Accepted 14 November 2008

Keywords:
RNAi
Small RNA
siRNA
miRNA
Crops
History of safe consumption

ABSTRACT

Small interfering RNAs (siRNAs) and microRNAs (miRNAs) are effector molecules of RNA interference (RNAi), a highly conserved RNA-based gene suppression mechanism in plants, mammals and other eukaryotes. Endogenous RNAi-based gene suppression has been harnessed naturally and through conventional breeding to achieve desired plant phenotypes. The present study demonstrates that endogenous small RNAs, such as siRNAs and miRNAs, are abundant in soybean seeds, corn kernels, and rice grain, plant tissues that are traditionally used for food and feed. Numerous endogenous plant small RNAs were found to have perfect complementarity to human genes as well as those of other mammals. The abundance of endogenous small RNA molecules in grain from safely consumed food and feed crops such as soybean, corn, and rice and the homology of a number of these dietary small RNAs to human and animal genomes and transcriptomes establishes a history of safe consumption for dietary small RNAs.
© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

RNA-mediated gene regulation (RNA interference, RNAi) is a highly conserved endogenous mechanism for regulation of gene expression in eukaryotes that operates through multiple pathways (Di Serio et al., 2001; Bantounas et al., 2004; Mello and Conte, 2004; Brodersen and Voinnet, 2006; Mallory and Vaucheret, 2006). RNAi plays important roles in development, pathogen defense and disease response in mammals, plants, and insects (Chang and Mendell, 2007; Pedersen et al., 2007). RNAi pathways are triggered by small RNAs that are usually 20–26 nucleotides (nt) long and are represented by diverse classes that differ from each other in their biogenesis such as small interfering RNAs (siRNAs), microRNAs (miRNAs), *trans*-acting siRNAs and other classes of small RNAs (Brodersen and Voinnet, 2006; Mallory and Vaucheret, 2006; Peters and Meister, 2007). The function of these various classes of small RNAs in animal and plant RNAi pathways involves sequence-specific recruitment of the RNA silencing complex to mRNA or DNA, leading to target mRNA cleavage, translational inhibition, or DNA modifications (Fig. 1). Small RNA regulatory networks are highly conserved in plants and animals and are an essential part of endogenous gene regulation. For example, it has been predicted that endogenous miRNAs likely regulate expression of at least one third of all human genes (Lewis et al., 2005).

RNAi has been harnessed in the improvement of several conventional crops including soybean, rice and maize. Soybean varieties that are precursors to those currently cultivated have a dark pigmentation due to anthocyanin content. Breeders have selected for soybeans with a yellow seed coat attributed to RNAi-mediated suppression of the chalcone synthase gene (Tuteja et al., 2004). RNAi has also been attributed to a conventional low-glutelin (seed storage protein) rice variety useful for those who must restrict dietary protein levels (Kusaba et al., 2003) and to the green color of conventional maize stalks (Della Vedova et al., 2005). Although RNAi has been harnessed by conventional breeders, crop quality and productivity can also be selectively improved through RNAi by targeted suppression of a specific gene or of a desired group of target genes. There are several biotechnology-derived products in development or that have already been approved for commercial cultivation that utilize RNA-mediated gene suppression. Some of these products, such as the FlavrSavr™ tomato, were designed to suppress target plant genes through antisense RNA, although later studies suggested RNAi as the mode of action (Sheehy et al., 1988; Sanders and Hiatt, 2005; Krieger et al., 2008). Other products utilizing RNA-mediated gene suppression include the papaya ringspot virus resistant papaya (Fuchs and Gonsalves, 2007), potatoes with increased dormancy periods of tubers (Marmiroli et al., 2000), rice and soybean with reduced expression of allergenic proteins

Abbreviations: dsRNA, double stranded RNA; EC, European Commission; EFSA, European Food Safety Authority; FAO, Food and Agriculture Organization of the United Nations; FDA, Food and Drug Administration; ILSI, International Life Sciences Institute; miRNA, Micro RNA; nt, nucleotide; OECD, Organisation for Economic Cooperation and Development; RNAi, RNA interference; siRNA, Small Interfering RNA; WHO, World Health Organization.

* Corresponding author. Tel.: +1 314 694 7292; fax: +1 314 694 5925.
E-mail address: jspetri@monsanto.com (J.S. Petrick).

0278-6915/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.fct.2008.11.025

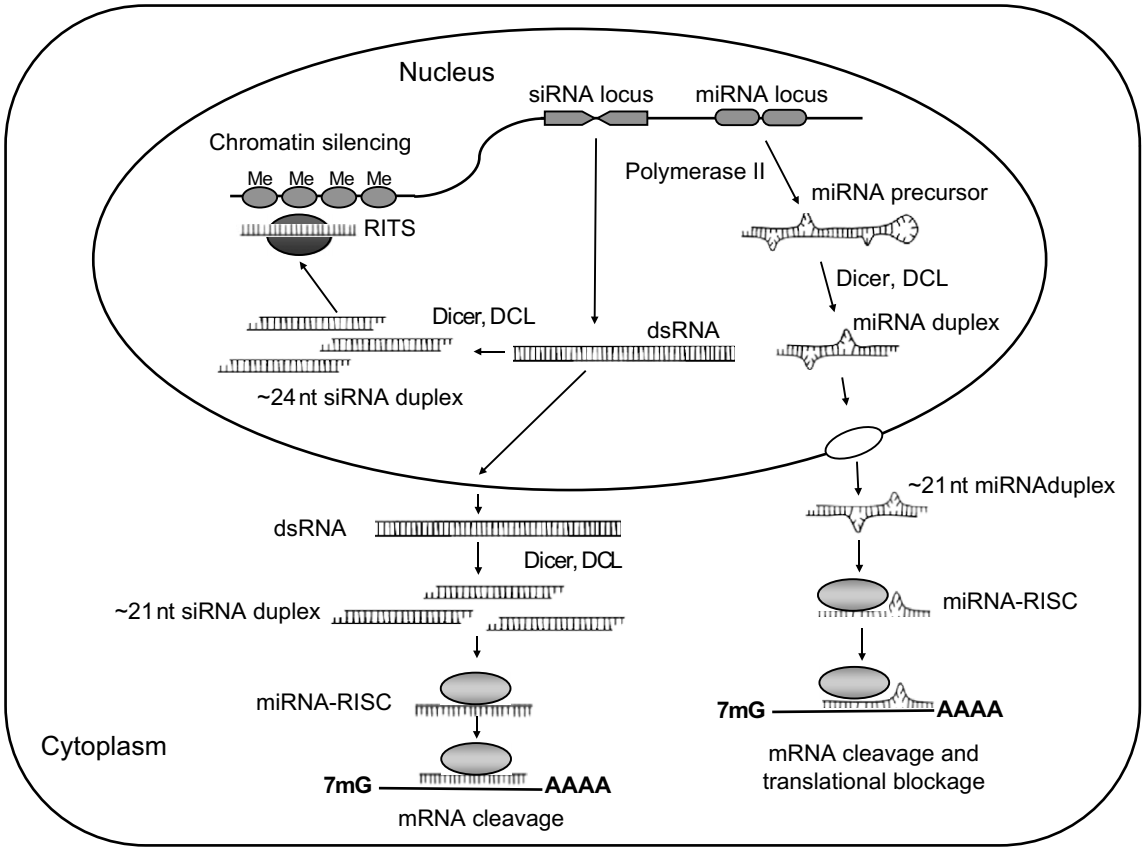


Fig. 1. Multiple pathways for RNA interference-mediated gene suppression. A simplified diagram shows major pathways that are conserved in plants and animals. siRNA or miRNA precursors are processed by multiple DICER RNAse III enzymes (DICER-like enzymes in plants, DCL) into small RNAs (mainly 21–24 nucleotide long) in the cytoplasmic or nuclear compartments. After processing, siRNAs or miRNAs are loaded into an RNA-induced silencing complex (RISC) to drive sequence-specific (antisense-sense sequence interaction with target mRNA) gene suppression by mRNA cleavage or inhibition of protein translation. siRNAs can also been loaded into an RNA-induced transcriptional silencing (RITS) complex in the nucleus that mediates chromatin modification/DNA methylation (Me) processes that may affect transcriptional activity (Jones-Rhoades et al., 2006).

(Herman et al., 2003; Tada et al., 2003) and the amylopectin potato (Hofvander et al., 2004). Recent applications of RNAi in crops include corn and cotton plants resistant to insect pests (Baum et al., 2007; Mao et al., 2007; Price and Gatehouse, 2008) and soybeans resistant to root-knot nematodes (Huang et al., 2006).

Efficient RNA-mediated gene suppression in plants can be achieved by introducing an expression cassette that produces double stranded RNA (dsRNA) with sequence homology to a target gene or by expression of an engineered artificial miRNA (artificial sequence based on a native miRNA precursor, that is processed *in planta* to a mature artificial miRNA) (Smith et al., 2000; Schwab et al., 2006; Ossowski et al., 2008). As illustrated in Fig. 1, expressed dsRNAs or artificial miRNA precursor transcripts are processed by DICER or DICER-like RNAse III enzymes into multiple siRNAs or into single mature miRNAs, respectively (Carmell and Hannon, 2004). These small RNAs are subsequently incorporated into RISC or RISC-like complexes and mediate sequence-specific silencing of plant target genes (Rhoades et al., 2002; Allen et al., 2005; Tang, 2005). Functional small RNAs in plants usually require significant sequence homology to target RNAs within a 'core' sequence; however, mismatches outside of this "core region" may be tolerated (Rhoades et al., 2002; Allen et al., 2005).

Safety assessment of new agricultural biotechnology products is an important process for their regulatory approval, registration and commercial acceptance. The existing safety assessment paradigm for biotechnology-derived crops is a well-defined approach that has been internationally accepted and applied successfully by regulators and regulatory scientists to over 50 crop products (Atherton, 2002; Cockburn, 2002; Codex, 2003a; Codex, 2003b; EC, 2003; ILSI, 2004; Konig et al., 2004). This paradigm utilizes a weight of evidence approach to evaluate the safety of the biotechnology-derived crop relative to its conventional counterpart (Cockburn, 2002). Establishing a well-documented history of safe consumption for RNA molecules including those that mediate RNAi (e.g. miRNAs and siRNAs) will be an important component of this weight of evidence approach for evaluating the safety of crop products developed utilizing RNAi-mediated gene suppression. As RNAi is a highly conserved endogenous mechanism for regulation of gene expression in eukaryotes, there is a long history of safe consumption of foods and feeds that contain siRNAs, miRNAs and longer dsRNAs in human and animal diets. Longer dsRNAs are ubiquitous in plant and animal foods in the diet, in the form of transfer RNAs, double stranded regions of ribosomal RNAs and in the form of endogenous RNA molecules that possess secondary structure. Furthermore, ingested nucleic acids are rapidly degraded through the action of intestinal nucleases (Carver and Walker, 1995) and low gastric pH and are thus unlikely to undergo systemic absorption. Therefore, it is not surprising that the current peer-reviewed literature lacks published studies specifically assessing the safety of consuming endogenous longer dsRNAs, siRNAs or miRNAs in human food or animal feed. However, with more agricultural products on the horizon that will be developed through RNAi-based gene suppression, it will be important to have a well-documented history of safe consumption for small RNAs in order to demonstrate the safety of the RNA molecules involved in this form of gene suppression in plants.

Although not thoroughly documented at present, a history of safe consumption for siRNAs, miRNAs and other small endogenous RNAs in plants can be established in part, through estimation of the abundance of these small RNAs and identification of these small RNAs in commonly consumed food crops. Further evidence for a history of safe consumption of these molecules can also be provided by identifying the homology of some of these widely consumed endogenous small RNAs in grain to the genomes and transcriptomes of consuming organisms. In the present study, an estimate of small RNA abundance was provided through semi-quantification of endogenous small RNAs in grain from soybeans. Corn and rice grain were also found to contain endogenous small RNAs at a level of similar magnitude. Many endogenous siRNAs, miRNAs and other small RNAs were identified in rice, a staple food crop with a long history of safe consumption and many of these small RNAs were complementary to human and animal genomes and transcriptomes. The abundance of these molecules in food and feed and their homology to human and animal genomes and transcriptomes provides evidence for a history of safe consumption of these mediators of RNAi and supports the safety of this technology for use in biotechnology-derived crops.

2. Materials and methods

2.1. Plant test materials

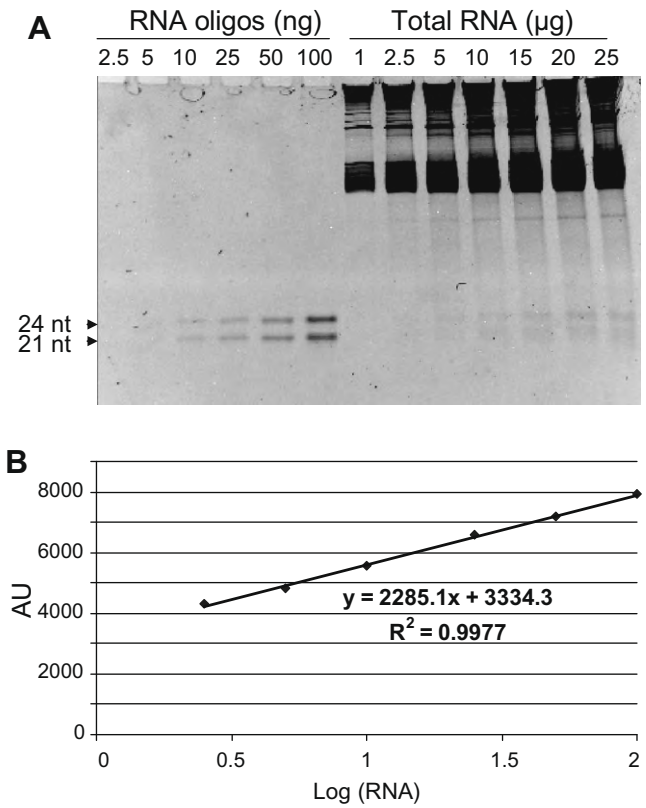
Total RNA was isolated from developing (R5 stage green seeds, about 9 mm diameter) and mature (R8 stage, dry yellow seeds) soybean seeds (*Glycine max* L.) from a conventional cultivar (cv. A3525); from corn kernels obtained from a conventional inbred line (*Zea mays* L., cv. LH244, 39 days after pollination); and from mature conventional rice grain from cv. Nipponbare (*Oryza sativa* L.) as described below. Soy and corn plants were grown in a greenhouse at Monsanto (Chesterfield, MO) under day/night conditions of 16/8 h and 25 °C/18 °C. Mature soy and corn seeds were stored for approximately six months under room temperature before RNA extraction. Developing soy seeds were harvested and immediately frozen in liquid nitrogen and were stored at −80 °C. Mature rice grain was field-grown under contract with the University of Arkansas (Fayetteville, AR). Unless otherwise noted, grain samples were stored at room temperature after harvest and total RNA was isolated as soon as practicable. Approximately 15 soy seeds, 10 corn kernels or 35 rice grains were frozen in liquid nitrogen and ground using a mortar and pestle. Aliquots of this ground material were use for RNA extraction.

2.2. Molecular analysis

Total RNA was isolated using the TRIzol® reagent (Invitrogen, Carlsbad, CA), according to manufacturer's instructions. Total RNA samples for all molecular analyses were separated on TBE-urea acrylamide gels (15 or 17%) stained with SYBR® Gold (Invitrogen). Analysis of the abundance of endogenous soybean small RNAs was accomplished by separating known amounts of soybean total RNA alongside a dilution series of known amounts of synthetic 21 and 24 nt RNA oligonucleotides (Fig. 2A). Fluorescence images of gels were captured using a GelDoc imaging system (BioRad, Hercules, CA) and band quantification of gel images was performed using Quantity One version 4.6 software (BioRad). From fluorescence quantification of the 21 and 24 nucleotide oligo dilution series (log transformed), a standard curve was constructed (Fig. 2B). The equation for this curve was used to calculate small RNA content in multiple samples of soybean total RNA, based on the obtained fluorescence values (Fig. 2C). This method thus provides a relative quantification or "semi-quantification" of endogenous soybean small RNAs in comparison to a known standard, rather than an "absolute" quantification. In this analysis of soybean seed small RNA content, a single calculated value of 3.89 ng small RNA/μg total RNA determined when 20 μg of total RNA was loaded on a gel was found to be more than 6 standard errors from the mean calculated value for soybean small RNA/μg total RNA. When the data set was analyzed using Dixon's Q test (Barnett and Lewis, 1994), this value was identified as a statistical outlier. This data point also lacked reproducibility and was therefore excluded from the analysis. This experiment assessing small RNA abundance was repeated several times and the mean estimated value for soybean small RNA/μg total RNA was reproducible and was consistently of a similar magnitude between experiments (data not shown).

2.3. Sequence analysis of endogenous rice small RNAs

Three independent small RNA libraries were constructed from rice grain (cv. Nipponbare) as described previously (Heisel et al., 2008). Small RNA library construction was performed by extracting the 18–26 nt small RNA fraction from total RNA after separation on a polyacrylamide gel, followed by sequential ligation of 3′ and 5′ cloning adaptors to these extracts (Llave et al., 2002). Approximately three micrograms of cDNA reverse transcribed from each small RNA library was sent to 454 Life Sciences (Branford, CT) for deep sequencing via pyrosequencing. Computer algorithms written in the Perl programming language (Perl scripts) were used to identify small RNA inserts within the raw sequence data through identification of the 5′ and 3′ cloning adaptors. A total of 285,864 small RNA sequence reads were obtained from rice grain small RNA libraries and were used for subsequent analysis.



C

Total RNA Loaded per Well (μg)	Small RNA Detected per Well (ng)	Small RNA ng per μg of Total RNA
2.5	3.64	1.46
5	7.90	1.57
10	16.56	1.66
15	19.80	1.32
25	53.30	2.13
	Mean	1.63
	STD	0.31

Fig. 2. Endogenous small RNA semi-quantification in conventional soybean seeds. (A) Total RNA from mature soybean seeds was separated on a 15% polyacrylamide-urea gel and visualized using SYBR® Gold stain. A dilution series of synthetic RNA oligonucleotides (21 and 24 nt) were used as a reference standard to compare with the relative intensity of fluorescence (given in AU, Arbitrary Units of fluorescence) in the test samples. The inverse image is shown. (B) Standard curve of fluorescence intensity (in Arbitrary Units, AU) from quantification of the dilution series of synthetic oligonucleotides graphed against the log transformed amount of oligonucleotides loaded on the gel. (C) Soybean small RNA semi-quantification in total RNA samples. The equation for the standard curve was used to calculate small RNA content in samples of soybean total RNA-based on the obtained fluorescence values.

2.4. Computational analysis

The megablast computer program (Zhang et al., 2000) was used to match unique 18–26 nt rice small RNAs to various genome and transcriptome sequences

as shown in Table 1 (megablast parameters were "–F F –W 12 –f T –D 3 –X 1 –y 1"). All genome and transcriptome sequences were downloaded from public databases. The versions, dates of compilation, sizes and locations are listed in the data source table found in Supplementary Table 1. Only perfect matches to the entire small RNA sequence were counted. For transcriptome matches, the blastall algorithm (Altschul et al., 1997) was used with parameters that allowed for the detection of mismatches (blastall parameters were "–F F –p blastn –W 9 –m 8 –v 1000 –b 1000 –e 100"). Only those sequences with a perfect match or with one mismatch along the entire small RNA length in reverse complementary orientation were counted in any of the analyses conducted herein.

3. Results

3.1. Estimation of endogenous soybean small RNA abundance

Total RNA was extracted from mature soybean grain. The maximal yield of total RNA per gram of mature soybeans was 986.6 μg, with 407.3 μg being the average total RNA yield per gram of mature soybean seeds from seven independent RNA extractions (range 274.7–986.6 μg per gram of seeds). As described in the methods section, the total RNA samples were separated using a SYBR® Gold stained polyacrylamide gel to visualize small RNA bands (Fig. 2A). Distinct bands approximately 21 and 24 nucleotides (nt) long that correspond to the predominant sizes of small RNAs were detected on the gel. Analysis of small RNA abundance was conducted by evaluating intensity of small RNA bands from soybean seeds relative to a reference standard dilution series of synthetic 21 and 24 nt long RNA oligonucleotides run alongside the total RNA samples for visual comparison of fluorescence. The mean estimated amount of small RNA per μg of soybean seed total RNA was 1.63 ± 0.31 ng (Fig. 2B and C). Based upon the maximal total RNA yield, this estimate of small RNA content in soybean seeds was used to calculate that one gram of soybean seeds contains up to 1.61 μg of 21–24 nt small RNAs. Based on average total RNA yield (from seven independent RNA extractions), this estimate would be reduced to 0.66 μg of 21–24 nt small RNAs per one gram of soybean seeds. Similar data were obtained for three independent gels representing independent RNA preparations (data not shown).

Small RNAs were only slightly more abundant in developing seeds than in mature seeds (Fig. 3). Small RNA bands of intensity similar to small RNA bands observed in soybean were also detectable in RNA extracted from mature corn kernels and rice grain (Fig. 4). Gel analysis was repeated three times with independent RNA extractions and similar results were obtained. Thus, corn and rice grain had a comparable amount of total small RNA content per gram of grain to that of soybean. However, unlike soybean RNA samples, these RNA samples had a higher background fluorescence that resulted in a smearing pattern, thus precluding any quantitative analysis of these small RNA fractions with the procedures used in this study.

Table 1
Number of endogenous rice small RNA with matches to publicly available genomes and transcriptomes (perfect match).

Total unique small RNAs	Rice grain	
	285,864	
Species (genome/transcriptome size in MB)	Genome	Transcriptome
Human, *Homo sapiens* (2881/104)	4759	270
Mouse, *Mus musculus* (2567/116)	5361	1313
Pig, *Sus scrofa* (626/34)	1520	297
Cow, *Bos taurus* (2732/52)	4706	1285
Chicken, *Gallus gallus* (1100/39)	4185	164
Soybean, *Glycine max* (925/117)	21,152	10,675
Corn, *Zea mays* (1592/91)	27,156	16,112
Rice, *Oryza sativa* (373/113)	242,459	38,782

3.2. Sequence homology of endogenous plant small RNAs to animal and human genomes and transcriptomes

Computational analysis of small RNA sequences (Margulies et al., 2005) generated as described in the methods section revealed that many endogenous 18–26 mer rice small RNAs had perfect sequence homology to human and mammalian genomes and transcriptomes (Table 1). A total of 285,864 unique 18–26 mer small RNAs were identified from sequencing of three rice grain libraries. Of these unique small RNAs, 4759 had perfect complementarity to sequences from the human genome and 270 of these small RNAs had perfect complementarity to the human transcriptome (Table 1). These numbers are presumed to be underestimates, as these rice small RNA sequencing efforts do not represent a complete sequencing of all endogenous rice grain small RNAs. Because a single unique small RNA can match multiple genes and because multiple RNAs can match a single gene, these 270 small RNAs matching the human transcriptome matched a total of 2035 unique human transcripts. Small RNAs can also regulate expression of transcripts that have less than perfect complementarity between the small RNA and the target site (Rhoades et al., 2002). Using more relaxed analytical criteria, i.e. allowing one mismatch between rice small RNA and human transcripts, the number of unique small RNA hits to human transcripts rose from 270 to 2589 and the number of unique human transcripts matched was 7075 (Supplementary Table 1). Human genes with perfect complementarity to rice grain small RNAs included a diverse list of genes such as those encoding: cell cycle regulators, structural proteins and adhesion molecules, developmental regulators, growth factors, metabolic enzymes/proteins, receptors, signal transduction proteins, transcription factors/transcriptional regulators and transporters (Table 2). A complete list of human genes with perfect complementarity to rice grain small RNAs and those with 1 mismatch relative to rice grain small RNAs can be found in the supplementary data (Supplementary Table 1). Many endogenous rice small RNAs were also found to be complementary to mouse (*Mus musculus*), pig (*Sus scrofa*), cow

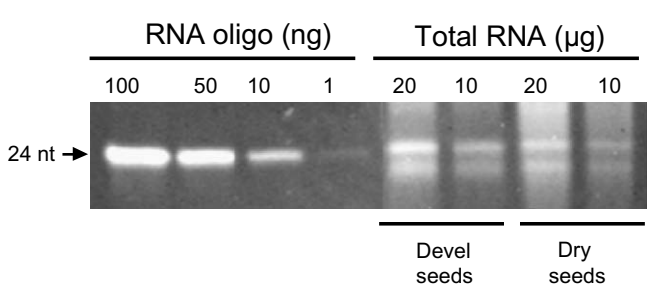


Fig. 3. Endogenous small RNAs accumulate in similar quantities in developing and mature soybean seeds. Total RNA from developing and mature soybean seeds was separated on 15% polyacrylamide-urea gel and visualized using SYBR® Gold stain.

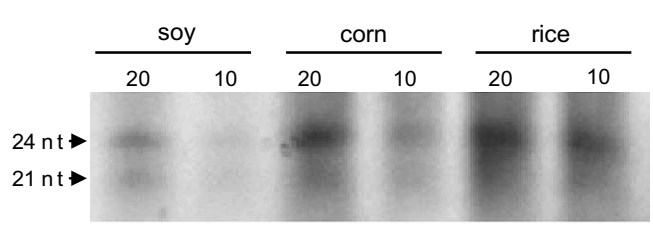


Fig. 4. Visualization of small RNAs in soybean, corn and rice. Total RNA from mature dry soybean seeds, mature dry corn kernels, and rice grain. Total RNA was separated on a 17% polyacrylamide-urea gel and visualized using SYBR® Gold stain. Numbers represent the number of μg of total RNA loaded on the gel. The inverse gel image is shown to increase contrast.

Table 2
List of selected human genes with 100% complementarity to endogenous small RNAs in rice grain.

Cell cycle regulators	
gi|16950654	Cyclin D1
gi|47132608	Cyclin-dependent kinase inhibitor 2B
gi|33946323	Guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1
gi|39812377	RAN binding protein 9
gi|32307123	Nuclear receptor co-activator 3, src-3
Cellular structure and adhesion molecules	
gi|18201922	Collagen, type XII, alpha 1
gi|31317225	Ephrin-B1
gi|52485852	Integrin, alpha 11
gi|14589888	N-cadherin, neuronal
Developmental	
gi|31317225	Ephrin-B1
gi|4503706	Fibroblast growth factor 9
gi|4503694	Fibroblast growth factor 18
gi|23308573	Sprouty 4
Growth factors	
gi|4503706	Fibroblast growth factor 9
gi|4503694	Fibroblast growth factor 18
gi|19923111	Insulin-like growth factor 1
Metabolic enzymes/proteins	
gi|65301138	ATPase, Class II, type 9A
gi|37577154	ATPase, vacuolar, H+ transporting, lysosomal accessory protein 1, ATP6AP1
gi|34335257	ATPase, vacuolar, H+ transporting, lysosomal 38 kDa, V0 subunit d isoform 1 (ATP6V0D1)
gi|51599150	Calpain, small subunit 1
gi|61743919	Cytochrome P450 4F11
gi|30061499	Gamma-glutamyltransferase-like 3
gi|4505610	PARN, Poly(A)-specific ribonuclease (deadenylation nuclease)
gi|38505195	Prostaglandin E synthase
gi|38788121	Serine protease 23
gi|13375784	Steroid 5 alpha-reductase 2-like (SRD5A2L)
gi|32967281	Ubiquitin-conjugating enzyme E2B
gi|58530887	Ubiquitin-conjugating enzyme E2R2
Receptors	
gi|4557266	Adrenergic receptor, beta-3
gi|51988913	Fibroblast growth factor receptor-like 1
gi|61744470	LDL receptor-related protein 8, apolipoprotein e receptor (LRP8, APOE R2)
gi|54792106	Muscarinic cholinergic receptor 2
gi|32307151	Oxytocin receptor
gi|8922178	Serine/threonine/tyrosine kinase 1
gi|65301166	Very low density lipoprotein receptor
Signal transduction	
gi|6138971	Beta adrenergic receptor kinase 1
gi|27477118	Calcium/calmodulin-dependent protein kinase IV
gi|51599150	Calpain, small subunit 1
gi|31317225	Ephrin-B1
gi|4557328	Fas ligand
gi|23065570	GTP binding protein 1
gi|33946323	Guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1
gi|52485852	Integrin, alpha 11
gi|10938013	Jun D proto-oncogene
gi|32481207	Mitogen-activated protein kinase-activated protein kinase 2 (MAPKAPK2)
gi|21735555	Mitogen-activated protein kinase kinase kinase 2 (MAP3K2)
gi|10835172	Nitric oxide synthase 1, neuronal
gi|32307151	Oxytocin receptor
gi|32455247	Phosphoinositide-3-kinase, regulatory subunit 1, p85 alpha
gi|18860871	Protein tyrosine phosphatase receptor type F
gi|88947650	Protein tyrosine phosphatase type IVA, member 2
gi|12232372	RAB GTPase activating protein 1
gi|39812377	RAN binding protein 9
gi|8922178	Serine/threonine/tyrosine kinase 1, STK1
gi|31543197	Serine/threonine kinase 40, STK 40
gi|23308573	Sprouty 4
Transcription factors and transcriptional regulators	
gi|30795241	Aryl hydrocarbon receptor nuclear translocator
gi|19923286	AT-binding transcription factor 1
gi|53749664	COUP-TF1, NR2F1 transcription factor
gi|59938775	cAMP responsive element binding protein 5
gi|10938013	Jun D proto-oncogene
gi|32307123	Nuclear receptor co-activator 3, src-3
gi|32307127	Nuclear receptor co-activator 6
gi|56699487	Nuclear receptor co-repressor 2
gi|61744437	Peroxisome proliferator activated receptor, alpha
gi|58331205	Retinoid X receptor, gamma

(*continued on next page*)

Table 2 (*continued*)

Transporters	
gi|46592914	ABCG1 transporter, cholesterol/phospholipid transport
gi|44680146	Ascorbate/Nucleobase Transporter SVCT2
gi|34335257	ATPase, vacuolar, H+ transporting, lysosomal 38 kDa, V0 subunit *d* isoform 1 (ATP6V0D1)
gi|13386497	Calcium channel, voltage-dependent, P/Q type, alpha 1A subunit
gi|54112391	Calcium channel, voltage-dependent, alpha 2/delta subunit 2
gi|27894377	Chloride intracellular channel 6
gi|40254457	Copper transporter CTR1
gi|9955961	Multidrug resistance-associated Protein 1, MRP1
gi|7706713	Organic anion transporter 3A1
gi|38679889	Organic anion transporter 4C1
gi|13569931	Organic anion transporter 5A1
gi|20143943	Potassium channel KCNK10
gi|24797140	Potassium inwardly-rectifying channel, subfamily J, member 5
gi|5032092	Solute carrier family 1, member 5, neutral amino acid transporter
gi|5032096	Solute carrier family 6, member 8, creatine transporter
gi|38569461	Solute carrier family 12, member 2, sodium/potassium/chloride transporter
gi|31563525	Solute carrier family 24, member 3, sodium/potassium/calcium exchanger
gi|52630414	Solute carrier family 30, member 3, zinc transporter

(*Bos taurus*), and chicken (*Gallus gallus*) genes (Table 1). Table 3 lists some genes that have been previously identified as potential regulators of the human cell cycle through *in vitro* RNAi knockdown experiments (Mukherji et al., 2006; Kittler et al., 2007) and that have perfect complementarity (or only one mismatch) to one or more small RNAs found in rice grain.

4. Discussion

RNAi-based gene suppression in plants is rapidly emerging as a valuable technique for improvement of crop quality via agricultural biotechnology. As with other aspects of agricultural biotechnology, potential issues for RNAi-based gene suppression should be assessed, including those pertaining to food and feed safety. The present studies provide evidence that soybean seeds, rice grain, and maize kernels contain numerous endogenous small RNAs that are detectable through direct staining of separated RNA. Small RNA sequencing and bioinformatics analysis revealed that many endogenous small RNAs in rice grain have perfect or near perfect complementarity to human and animal genes. Based on the ubiquitous nature of the RNAi process, it is presumed that grain from soybean, corn and other widely consumed crops also contains small RNAs with complementarity to human and animal genes. These data can be used to illustrate a history of safe consumption for dietary small RNAs.

Currently, data providing even a very rough estimate on quantity of small RNAs present in foods are not available in the scientific literature. This lack of data is probably due in part to technical challenges associated with quantitative analysis of a heterogeneous population of small RNAs within a total RNA preparation. The approach used in the present study allows for the detection and estimation of endogenous grain small RNA abundance using direct visualization on a gel. Whereas we are aware that estimates of small RNA quantity obtained in this study represent a rough estimate of the total small RNA population and that the results may vary depending on plant tissue type, method of RNA extraction, type and sensitivity of stain used and other factors, we believe that these data provide a useful guidance regarding the approximate amount of small RNAs in grain. Based upon an experimentally determined estimate of approximately 1.63 ng of small RNAs per μg of soybean total RNA, it was estimated that up to 1.61 μg of 21–24 nt small RNAs were obtained per gram of soybeans, whereas the average amount of endogenous small RNAs in soybeans was estimated to be 0.66 μg per gram. These numbers indicate that up to 0.1–0.2% of the total RNA population may be composed of small RNAs. With plant mRNA usually representing approximately 0.5–1.5% of total RNA and the average molecular mass of a small RNA being at least 80 times less than the average molecular mass of plant mRNA, the small RNA population is approximately an order of magnitude less abundant than that of mRNA. Furthermore, data from the present studies revealed that small RNAs are only slightly more abundant in developing seeds relative to mature seeds, suggesting that small RNAs either continue to be produced during seed maturation or that they are relatively stable in mature seeds following several months of room temperature storage.

Table 3
List of selected human genes that have been identified through RNAi as potential regulators of the human cell cycle (Mukherji et al., 2006; Kittler et al., 2007) and that have complementary matches to endogenous rice small RNAs.

Gene ID	Annotation/function	Plant small RNA ID*
gi|33946323	Guanine nucleotide binding protein	399664
gi|39812377	RAN binding protein 9	1066113
gi|32307123	Nuclear receptor co-activator 3	146628
gi|6912283	Carbonic anhydrase XIV	1356926
gi|13386499	Calcium channel, alpha 1A subunit	146628
gi|8923822	Potassium inwardly-rectifying Channel	462873
gi|4505664	Phosphodiesterase 4C	1083748
gi|17999536	PRP8 pre-mRNA processing factor	457312
gi|20336472	B-cell CLL/lymphoma 7B	334341, 975870
gi|4758077	c-Src tyrosine kinase	500660
gi|68800039	Small nuclear RNA auxiliary factor 1	397897, 1366463
gi|38327563	Serine/threonine kinase 6	565921

Note: Bioinformatics analysis allowed for 1 mismatch between the endogenous rice grain small RNAs and the genes identified as potential cell cycle regulators.
* See Supplementary Table 1 for sequence of small RNAs.

Sequencing of endogenous small RNAs in rice grain and subsequent computational analysis revealed a number of endogenous small RNAs in rice grain with perfect complementarity to human and animal genomes and transcriptomes. Among the human genes identified as having high homology to small RNA molecules present in rice, several have been previously shown to have an associated phenotype such as cell cycle arrest when suppressed *in vitro* using synthetic siRNAs in cultured human cells (Mukherji et al., 2006; Kittler et al., 2007). In addition, rice small RNAs had perfect complementarity to human genes encoding cell cycle regulators, structural proteins and adhesion molecules, developmental regulators, growth factors, metabolic enzymes/proteins, receptors, signal transduction proteins, transcription factors/transcriptional regulators and transporters. The long history of consumption of rice as human food and the results presented in the present manuscript together illustrate that small RNAs with perfect complementarity to human genes are present in grain from a safely consumed staple crop, therefore implying that dietary small RNAs do not exhibit oral activity to an extent that would have a meaningful impact on human health.

Cooking and processing of foods may destroy endogenous RNA and DNA molecules, although any remaining nucleic acids are likely to be digested after consumption due to low gastric pH and due to pancreatic nucleases secreted into the intestine and those present in the saliva and blood (Carver and Walker, 1995; Park et al., 2006). Less than 5% of nucleic acid catabolites are used for *de novo* synthesis of nucleic acids and the remaining catabolites are utilized in other biochemical reactions or undergo excretion in the urine and feces (Carver and Walker, 1995). In a study conducted in neonatal pigs, no intestinal absorption of RNA was observed (Baintner and Toth, 1986). Furthermore, there does not appear to be any evidence in the scientific literature suggesting that intact RNA is absorbed following ingestion. Digestion of ingested nucleic acids including RNA has been well established and this presumably applies to all dietary RNAs, including siRNAs, miRNAs and other endogenous RNA molecules that contain double stranded regions (e.g. ribosomal RNA, transfer RNA, and double stranded small RNA precursors). The weight of evidence supporting this presumption includes the above references and as described below, also includes the lack of oral bioavailability for nucleic acid therapeutics and the history of safe consumption of dietary RNAs including siRNAs and miRNAs.

There is evidence that horizontal transfer of a specifically engineered gene suppression signal can occur between plants and target pests such as root-knot nematode (Huang et al., 2006; Gheysen and Vanholme, 2007), corn rootworm (Baum et al., 2007), and cotton bollworm (Mao et al., 2007). To work efficiently in target pests, the plant-produced gene suppression trigger must have a high degree of complementarity to the pest target gene(s) and must have the ability to reach the target tissue (Gordon and Waterhouse, 2007). Oral administration of chemically stabilized antisense oligonucleotides to rats, however, resulted in rapid degradation and excretion of the test material (Agrawal et al., 1995). In mice, chemically stabilized siRNAs were digested and/or rapidly eliminated as evidenced by the limited biodistribution and rapid degradation of these siRNAs following intravenous injection (Braasch et al., 2004). Direct intravenous injection of unmodified small RNAs into experimental animals generally requires extremely high doses, specialized delivery routes, or lipid delivery agents to foster small RNA uptake (Soutschek et al., 2004; Behlke, 2006). Based on the general chemical properties of siRNAs and oligonucleotides and the evidence regarding their instability, limited bioavailability, and limited biodistribution in mammalian systems (Agrawal et al., 1995; Zhang et al., 1995; Braasch et al., 2004), systemic absorption of intact siRNAs or miRNAs following dietary exposure to these molecules would be highly improbable.

The lack of systemic absorption of dietary small RNAs is further evidenced by the history of safe consumption of dietary siRNAs and miRNAs that match human and animal genes as outlined in the present study and in a recently published study by Heisel and colleagues (2008). The abundance of endogenous small RNAs in staple food crops and presumably in livestock (such as cow, chicken, and pig) and the homology of these small RNAs to the human genome and transcriptome illustrate the history of safe consumption of these molecules. Small RNAs from foods and their dsRNA precursors are therefore consumed safely without evidence of adverse effects, most likely because oral activity of these dietary RNAs would require significant absorption, distribution, and uptake of these molecules. As discussed above, this is an unlikely scenario for dietary RNAs. Therefore, based on the history of safe consumption of dietary small RNAs in staple grains and the homology of some of these endogenous small RNAs to the genomes and transcriptomes of consuming organisms documented in the present manuscript we conclude that RNAi-mediated gene suppression represents a highly specific, safe and effective means to improve crops through agricultural biotechnology. Based on this evidence it can be concluded that RNAi-mediated regulation of gene expression in biotechnology-derived crops is as safe for food and feed use as conventional crops that harness RNAi-based gene regulation as one of several ways to achieve new plant traits. The safety of future crops generated through applications of RNAi should thus be evaluated for safety using the existing comparative safety assessment paradigm, which has been developed for biotechnology-derived crops (FAO/WHO, 1996, 2000; EC, 2003; OECD, 2003, 2005).

Conflict of interest statement

All authors are employees of the Monsanto Company. The Monsanto Company is an agricultural company that produces and sells biotechnology-derived seeds. All funding for this manuscript was provided by the Monsanto Company.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fct.2008.11.025.

References

Agrawal, S., Zhang, X., Lu, Z., Zhao, H., Tamburin, J.M., Yan, J., Cai, H., Diasio, R.B., Habus, I., Jiang, Z., Iyer, R.P., Yu, D., Zhang, R., 1995. Absorption, tissue distribution and in vivo stability in rats of a hybrid antisense oligonucleotide following oral administration. Biochem. Pharmacol. 50, 571–576.

Allen, E., Xie, Z., Gustafson, A.M., Carrington, J.C., 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121, 207–221.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. 25, 3389–3402.

Atherton, K.T., 2002. Safety assessment of genetically modified crops. Toxicology, 421–426.

Baintner, K., Toth, I., 1986. Failure to demonstrate intestinal absorption of RNA in the newborn pig. Preliminary communication. Acta Veter. Hung. 34, 239–241.

Bantounas, I., Phylactou, L.A., Uney, J.B., 2004. RNA interference and the use of small interfering RNA to study gene function in mammalian systems. J. Mol. Endocr. 33, 545–557.

Barnett, V., Lewis, T., 1994. Specific discordancy tests for outliers in univariate samples. In: Outliers in Statistical Data. John Wiley and Sons, Inc., New York, NY, pp. 226–227.

Baum, J.A., Bogaert, T., Clinton, W., Heck, G.R., Feldmann, P., Ilagan, O., Johnson, S., Plaetinck, G., Munyikwa, T., Pleau, M., Vaughn, T., Roberts, J., 2007. Control of coleopteran insect pests through RNA interference. Nat. Biotechnol. 25, 1322–1326.

Behlke, M.A., 2006. Progress towards in vivo use of siRNAs. Mol. Ther. 13, 644–670.

Braasch, D.A., Paroo, Z., Constantinescu, A., Ren, G., Oz, O.K., Mason, R.P., Corey, D.R., 2004. Biodistribution of phosphodiester and phosphorothioate siRNA. Bioorg. Med. Chem. Lett. 14, 1139–1143.

Brodersen, P., Voinnet, O., 2006. The diversity of RNA silencing pathways in plants. Trends Gen. 22, 268–280.

Carmell, M.A., Hannon, G.J., 2004. RNase III enzymes and the initiation of gene silencing. Nat. Struct. Mol. Biol. 11, 214–218.
Carver, J., Walker, W.A., 1995. The role of nucleotides in human nutrition. Nutr. Biochem. 6, 58–72.
Chang, T.C., Mendell, J.T., 2007. MicroRNAs in vertebrate physiology and human disease. Ann. Rev. Genomics Hum. Genet. 8, 215–239.
Cockburn, A., 2002. Assuring the safety of genetically modified (GM) foods: the importance of an holistic, integrative approach. J. Biotechnol. 98, 79–106.
Codex, 2003a. Guideline for the Conduct of Food Safety Assessment of Foods Derived from Recombinant-DNA Plants (CAC/GL 45-2003), 7–26. Rome, Italy: Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Food and Agriculture Organisation.
Codex, 2003b. Principles for the Risk Analysis of Foods Derived from Recombinant – DNA Plants (CAC/GL 44-2003), 1–6. Rome, Italy: Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Food and Agriculture Organisation.
Della Vedova, C.B., Lorbiecke, R., Kirsch, H., Schulte, M.B., Scheets, K., Borchert, L.M., Scheffler, B.E., Wienand, U., Cone, K.C., Birchler, J.A., 2005. The dominant inhibitory chalcone synthase allele C2-Idf (inhibitor diffuse) from Zea mays (L.) acts via an endogenous RNA silencing mechanism. Genetics 170, 1989–2002.
Di Serio, F., Schob, H., Iglesias, A., Tarina, C., Bouldoires, E., Meins Jr., F., 2001. Sense- and antisense-mediated gene silencing in tobacco is inhibited by the same viral suppressors and is associated with accumulation of small RNAs. Proc. Natl. Acad. Sci. USA 98, 6506–6510.
EC, 2003. Guidance document for the risk assessment of genetically modified plants and derived food and feed, prepared by the Joint Working Group on Novel Foods and GMOs, 6–7 March 2003. European Commission.
FAO/WHO, 1996. Biotechnology and food safety. Report of a Joint FAO/WHO Consultation, FAO/WHO.
FAO/WHO, 2000. Safety aspects of genetically modified foods of plant origin. Report of a Joint FAO/WHO Consultation, Geneva, Switzerland.
Fuchs, M., Gonsalves, D., 2007. Safety of virus-resistant transgenic plants two decades after their introduction: lessons from realistic field risk assessment studies. Ann. Rev. Phytopath. 45, 173–202.
Gheysen, G., Vanholme, B., 2007. RNAi from plants to nematodes. Trends Biotechnol. 25, 89–92.
Gordon, K.H.J., Waterhouse, P.M., 2007. RNAi for insect-proof plants. Nat. Biotechnol. 25, 1231–1232.
Heisel, S.E., Zhang, Y., Allen, E., Guo, L., Reynolds, T.L., Yang, X., Kovalic, D., Roberts, J.K., 2008. Characterization of unique small RNA populations from rice grain. PLoS One 3, e2871.
Herman, E.M., Helm, R.M., Jung, R., Kinney, A.J., 2003. Genetic Modification Removes an Immunodominant Allergen from Soybean. Plant Physiol. 132, 36–43.
Hofvander, P., Persson, P.T., Tallberg, A., Tallberg, T., Wilkstrom, O., 2004. Genetically engineered modification of potato to form amylopectin starch. vol. 6784338, USA.
Huang, G., Allen, R., Davis, E.L., Baum, T.J., Hussey, R.S., 2006. Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. Proc. Natl. Acad. Sci. USA 103, 14302–14306.
ILSI, 2004. Nutritional and safety assessments of foods and feeds nutritionally improved through biotechnology: an executive summary. J. Food Sci. 69, CRH62–CRH68.
Jones-Rhoades, M.W., Bartel, D.P., Bartel, B., 2006. MicroRNAs and their regulatory roles in plants. Ann. Rev. Plant Biol. 57, 19–53.
Kittler, R., Pelletier, L., Heninger, A.K., Slabicki, M., Theis, M., Miroslaw, L., Poser, I., Lawo, S., Grabner, H., Kozak, K., Wagner, J., Surendranath, V., Richter, C., Bowen, W., Jackson, A.L., Habermann, B., Hyman, A.A., Buchholz, F., 2007. Genome-scale RNAi profiling of cell division in human tissue culture cells. Nat. Cell Biol. 9, 1401–1412.
Konig, A., Cockburn, A., Crevel, R.W., Debruyne, E., Grafstroem, R., Hammerling, U., Kimber, I., Knudsen, I., Kuiper, H.A., Peijnenburg, A.A., Penninks, A.H., Poulsen, M., Schauzu, M., Wal, J.M., 2004. Assessment of the safety of foods derived from genetically modified (GM) crops. Food Chem. Toxicol. 42, 1047–1088.
Krieger, E.K., Allen, E., Gilbertson, L.A., Roberts, J.K., Hiatt, W., Sanders, R.A., 2008. The Flavr Savr tomato, an early example of RNAi technology. HortScience 43, 962–964.
Kusaba, M., Miyahara, K., Iida, S., Fukuoka, H., Takano, T., Sassa, H., Nishimura, M., Nishio, T., 2003. Low glutelin content1: a dominant mutation that suppresses the glutelin multigene family via RNA silencing in rice. Plant Cell 15, 1455–1467.
Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are MicroRNA targets. Cell 120, 15–20.
Llave, C., Kasschau, K.D., Rector, M.A., Carrington, J.C., 2002. Endogenous and silencing-associated small RNAs in plants. Plant Cell 14, 1605–1619.
Mallory, A.C., Vaucheret, H., 2006. Functions of microRNAs and related small RNAs in plants. Nat. Gen. 38 (Suppl.), S31–S36.
Mao, Y.B., Cai, W.J., Wang, J.W., Hong, G.J., Tao, X.Y., Wang, L.J., Huang, Y.P., Chen, X.Y., 2007. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. Nat. Biotechnol. 25, 1307–1313.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.
Marmiroli, N., Agrimonti, C., Visioli, G., Colauzzi, G., Guarda, G., Zuppini, A., 2000. Silencing of G1–1 and A2–1 genes. Effects on general plant phenotype and on tuber dormancy in Solanum tuberosum L.. Potato Res. 43, 313–323.
Mello, C.C., Conte Jr., D., 2004. Revealing the world of RNA interference. Nature 431, 338–342.
Mukherji, M., Bell, R., Supekova, L., Wang, Y., Orth, A.P., Batalov, S., Miraglia, L., Huesken, D., Lange, J., Martin, C., Sahasrabudhe, S., Reinhardt, M., Natt, F., Hall, J., Mickanin, C., Labow, M., Chanda, S.K., Cho, C.Y., Schultz, P.G., 2006. Genome-wide functional analysis of human cell-cycle regulators. Proc. Natl. Acad. Sci. USA 103, 14819–14824.
OECD, 2003. Considerations for the safety assessment of animal feedstuffs derived from genetically modified plants. OECD, ENV/JM/MONO (2003) 10.
OECD, 2005. An introduction to the biosafety consensus documents of OECD's working group for harmonisation in biotechnology. Series on harmonisation of regulatory oversight in biotechnology, 32.
Ossowski, S., Schwab, R., Weigel, D., 2008. Gene silencing in plants using artificial microRNAs and other small RNAs. Plant J. 53, 674–690.
Park, N.J., Li, Y., Yu, T., Brinkman, B.M., Wong, D.T., 2006. Characterization of RNA in saliva. Clin. Chem. 52, 988–994.
Pedersen, I.M., Cheng, G., Wieland, S., Volinia, S., Croce, C.M., Chisari, F.V., David, M., 2007. Interferon modulation of cellular microRNAs as an antiviral mechanism. Nature 449, 919–922.
Peters, L., Meister, G., 2007. Argonaute Proteins: Mediators of RNA Silencing. Mol. Cell 26, 611–623.
Price, D.R., Gatehouse, J.A., 2008. RNAi-mediated crop protection against insects. Trends Biotechnol. 26, 393–400.
Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., Bartel, D.P., 2002. Prediction of plant microRNA targets. Cell 110, 513–520.
Sanders, R.A., Hiatt, W., 2005. Tomato transgene structure and silencing. Nat. Biotechnol. 23, 287–289.
Schwab, R., Ossowski, S., Riester, M., Warthmann, N., Weigel, D., 2006. Highly specific gene silencing by artificial MicroRNAs in arabidopsis. Plant Cell 18, 1121–1133.
Sheehy, R.E., Kramer, M., Hiatt, W.R., 1988. Reduction of polygalacturonase activity in tomato fruit by antisense RNA. Proc. Natl. Acad. Sci. USA 85, 8805–8809.
Smith, N.A., Singh, S.P., Wang, M.B., Stoutjesdijk, P.A., Green, A.G., Waterhouse, P.M., 2000. Gene expression: total silencing by intron-spliced hairpin RNAs. Nature 407, 319–320.
Soutschek, J., Akinc, A., Bramlage, B., Charisse, K., Constien, R., Donoghue, M., Elbashir, S., Geick, A., Hadwiger, P., Harborth, J., John, M., Kesavan, V., Lavine, G., Pandey, R.K., Racie, T., Rajeev, K.G., Rohl, I., Toudjarska, I., Wang, G., Wuschko, S., Bumcrot, D., Koteliansky, V., Limmer, S., Manoharan, M., Vornlocher, H.P., 2004. Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. Nature 432, 173–178.
Tada, Y., Akagi, H., Fujimura, T., Matsuda, T., 2003. Effect of an antisense sequence on rice allergen comprising multigene family. Breed. Sci. 53, 61–67.
Tang, G., 2005. SiRNA and miRNA: an insight into RISCs. Trends Biochem. Sci. 30, 106–114.
Tuteja, J.H., Clough, S.J., Chan, W.C., Vodkin, L.O., 2004. Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in Glycine max. Plant Cell 16, 819–835.
Zhang, R., Lu, Z., Zhao, H., Zhang, X., Diasio, R.B., Habus, I., Jiang, Z., Iyer, R.P., Yu, D., Agrawal, S., 1995. In vivo stability, disposition and metabolism of a "hybrid" oligonucleotide phosphorothioate in rats. Biochem. Pharmacol. 50, 545–556.
Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. J. Computational Biol. 7, 203–214.