Gene 522 (2013) 1–7

Contents lists available at SciVerse ScienceDirect

Gene

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

Widespread, abundant, and diverse TE-associated siRNAs in developing wheat grain

Fenglong Sun [1], Weiwei Guo [1], Jinkun Du, Zhongfu Ni, Qixin Sun, Yingyin Yao *

State Key Laboratory for Agrobiotechnology and Key Laboratory of Crop Heterosis and Utilization (MOE) and Key Laboratory of Crop Genomics and Genetic Improvement (MOA), Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing, 100193, P. R. China

ARTICLE INFO

Article history:
Accepted 22 March 2013
Available online 2 April 2013

Keywords:
Wheat
SiRNA
Seed development
Transposon

ABSTRACT

Small RNAs related to RNA interference are key molecules in many developmental processes, in which they can both regulate developmental gene expression and maintain the integrity of the genome and epigenome. In plants, short interfering RNAs (siRNAs) of 24nt in length are an abundant type of small RNA associated with transposable elements (TEs), other repetitive sequences, and viral defense. One means by which TE-associated siRNAs affect genome integrity is by altering chromatin structure through a process called RNA-directed DNA methylation (RdDM). In this paper, we describe a comparative survey of siRNAs from wheat seedling leaves, seedling roots, young spikelets, and grains at 8 and 15 days after pollination (DAP). We find that the general patterns of siRNA distributions are similar across different TEs and within TEs of the same family regardless of tissue, but the relative abundance of 24-nt siRNAs is highest in developing grains. We also find that TEs that are transcriptionally active in endosperm are associated with the highest siRNA abundance not only in grains, but also in other tissues as well. These results suggest that RdDM is an important feature of developing wheat grain and are consistent with the hypothesis that TE expression in endosperm results in increased TE siRNAs, and that RdDM is a conserved feature of plant seed development.

1. Introduction

The majority of eukaryotic organisms employ at least some form of RNA interference (RNAi), in which small RNAs guide Argonaute proteins to their target RNAs by sequence complementary (Carthew and Sontheimer, 2009). In plants, as in other eukaryotes, RNAi-related mechanisms can suppress expression of cellular protein coding genes in response to developmental or environmental cues; however RNAi also plays major roles in defending the genome from foreign nucleic acids such as viruses and transgenes (Carthew and Sontheimer, 2009). In addition, other contexts for RNAi not directly linked to gene silencing have been observed recently, such as repair of DNA double strand breaks (Wei et al., 2012). Related to endogenous gene regulation, small RNAs (sRNAs) of 21 to 24 nucleotides in length silence gene expression by multiple mechanisms and are present in diverse eukaryotic organisms (Jamalkandi and Masoudi-Nejad, 2009; Lu et al., 2005; Sunkar et al., 2005). Plant endogenous sRNAs can be divided into two major classes: microRNAs (miRNAs) and short-interfering RNAs (siRNAs). To date, several subclasses of siRNAs have been studied, including some that function in posttranscriptional gene silencing (PTGS) and some that function in transcriptional gene silencing. PTGS-associated siRNAs include trans-acting siRNAs (ta-siRNAs) (Gasciolli et al., 2005; Xie et al., 2005; Yoshikawa et al., 2005), and natural antisense siRNAs (nat-siRNAs) (Borsani et al., 2005; Katiyar-Agarwal et al., 2006), both of which can regulate gene expression in response to environmental or developmental cues. In contrast, the more abundant siRNAs in plants are associated with transcriptional silencing of repetitive or foreign DNA by RNA-directed DNA methylation (RdDM) (Gehring et al., 2009; Hsieh et al., 2009; Levy and Walbot, 1990; Slotkin et al., 2009; Wassenegger et al., 1994).

RdDM in endosperm may contribute to genomic imprinting or have other far-reaching effects on chromatin organization (Mosher and Melnyk, 2010). Many studies have suggested that programmed DNA demethylation in endosperm leads to TE activation which in turn leads to production of TE siRNAs. Evidence for such phenomena have been reported in maize (Lauria et al., 2004), rice (Zemach et al., 2010) and *Arabidopsis* (Gehring et al., 2009; Hsieh et al., 2009). Coupled with the loss of DNA methylation in the endosperm in *Arabidopsis* is abundant production of 24-nt siRNAs, which have been speculated to move into the embryo to reinforce TE silencing (Mosher and Melnyk, 2010), similar to the model described by Slotkin et al. (2009), in which 21-nt

Abbreviations: DAP, days after pollination; DCL, DICER-like; DNA, deoxyribonucleic acid; GB, giga base; GRA8, 8-DAP grains; GRA15, 15-DAP grains; LTR, long terminal repeat; mRNA, message RNA; NCBI, National Center for Biotechnology Information; PCR, polymerase chain reaction; PolIV, RNA polymerases IV; PTGS, Post-transcriptional Gene Silencing; RdDM, RNA-directed DNA methylation; RNA, Ribonucleic Acid; RNAi, RNA interference; SDS, Sodium dodecylsulfate; siRNA, short-interferencing RNA; SSC, saline-sodium citrate; SL, seedling leaves; SR, seedling roots; TREP, Triticeae Repeat Sequence Database; TE, transposable elements; YS, young spikelets.

* Corresponding author. Tel.: +86 10 62731452.
E-mail address: yingyin@cau.edu.cn (Y. Yao).
[1] these authors contribute this work equally.

0378-1119/$ – see front matter.
http://dx.doi.org/10.1016/j.gene.2013.03.101

siRNAs from vegetative nuclei move into sperm to silence TEs. Recent studies have shown that DEMETER-dependent demethylation in the central cell and vegetative cell reinforces TE methylation in plant gametes and likely contributes to stable silencing of TEs across generations (Calarco et al., 2012; Ibarra et al., 2012). The logic in these models is that allowing a form of transcription of TEs in the tissue next to the germline or embryo could provide a mobile signal (siRNAs) in order to completely silence TEs in the cells that actually matter for inheritance. Another recent study has lead to the proposal that siRNAs in the endosperm function to modulate parental genome dosage effects (Lu et al., 2012). While precisely defined mechanisms and consequences for these and other RNAi-related phenomena in endosperm and embryo development remain elusive, it is clear that they do play important roles.

Wheat (*Triticum aestivum*, AABBDD, 2n = 42) is the world's most widely grown crop, occupying 17% of all cultivated land and providing the source of approximately 55% of the carbohydrates consumed by humans (Gill et al., 2004). Despite its large size (17 GB) and complicated karyotype, extensive sequence and structural information for the wheat genome has been produced (Brenchley et al., 2012). Like other seeds, wheat grain undergoes a complex series of developmental events leading to production of a mature embryo and endosperm. Despite its agricultural importance, little research has been done on the molecular basis of development of wheat grain. Some topics that have been studied include expression profiles of metabolic proteins in endosperm (Vensel et al., 2005) and of mRNA in whole grain (Wan et al., 2008). Currently, the roles that small RNAs play in grain development in wheat can only be speculated on based on what is known from other plants. Deep sequencing of small RNAs from various wheat tissues including leaves, roots, tillers, and young spikes have revealed a complex pool of small RNAs that attests to the existence of multiple RNAi-related mechanisms in wheat, similar to other plants (Cantu et al., 2010; Kenan-Eichler et al., 2011; Yao et al., 2010). Here we described a comparative study of TE-associated siRNAs from seedling leaves, seedling roots, young spikelets, and grains at 8 and 15 days after pollination (DAP). We found that the general patterns of siRNA distributions were similar across different TEs and within TEs of the same family regardless of tissue, but the relative abundance of 24-nt siRNAs was highest in developing grains. These TE siRNAs in grain corresponded to a diverse set of class I and II TEs. We also found that the TEs that were transcriptionally active in endosperm were associated with the highest siRNA abundance both in grains and in other tissues.

2. Materials and methods

2.1. Plant materials

Five tissues of hexaploid wheat (*T. aestivum* L.) line Chinese Spring were used as the source for small RNA libraries. For collection of seedling leaves and roots, plants were grown in a growth chamber at a relative humidity of 75%, 26/20 °C day/night temperature, and a light intensity of 3000 lx. The tissues were harvested when the third leaf was at least 50% emerged. For collection of young spikelets and developing grains, plants were grown in field conditions. Young spikelets were collected when they reached 10–15 mm in length, and grains were collected at 8 and 15 days after pollination (DAP).

2.2. Preparation of small RNA libraries

Total RNA was isolated from frozen materials using Trizol reagent (Invitrogen, USA) according to the manufacturer's instructions. Low molecular weight RNA was enriched by 0.5 M NaCl and 10% PEG8000 precipitation. About 100 μg low molecular weight RNA (15–200 nt) was separated on a denaturing 15% polyacrylamide gel. RNA between 18 and 26 nt was excised, purified from gel, ligated to adaptors, reverse transcribed, PCR amplified, and sequenced using an Illumina GA IIx platform (BGI at Shenzhen).

2.3. Read processing and mapping to TEs

The adaptor sequences were trimmed from the Illumina reads using 'vector strip' in the EMBOSS package. Reads with length 18–26 nt were mapped to the sequences of *Triticum* in TREP (http://wheat.pw.usda.gov/ITMI/Repeats/, Release 10, July 2008) using Bowtie software version one (Langmead et al., 2009). A single mismatch per read was tolerated in the alignments.

2.4. Normalizing RNA read count by TEs abundance

A dataset of 454-sequenced reads with 5× coverage of the wheat genome was downloaded from CerealsDB (http://www.cerealsdb.uk.net/CerealsDB/Documents/DOC_CerealsDB.php). Two sets of DNA tags were generated by extracting the sequences of the first 22 nt and of nucleotides 30 to 51 of each read. These two sets were used as a control for TE abundance in the genome. Bowtie was utilized to map both sets of tags to the *Triticum* TEs in TREP that were represented by at least one sRNA read in at least one of the five tissues. The number of DNA tag reads per TE was obtained by averaging the values from the two sets. The abundance of sRNA reads for each TE was normalized by divided by the averaged number of DNA tags per TE. Finally the values for each TE were summed to produce the cumulative value for each superfamily.

2.5. Identification of TEs with enriched mRNA expression in endosperm or leaves

A dataset of Illumina-sequenced endosperm mRNA was downloaded from the NCBI GEO (GSE38344) (Gillies et al., 2012) and we prepared and sequenced our own mRNA Illumina libraries from leaves (manuscript in preparation). We aligned the reads to the complete set of *T. aestivum* TE sequences in the TREP database using BLASTN, and we identified reads producing fewer than 5 mismatches and at least 95% sequence identity. The read counts from each library for each TE were normalized by the total read count of the library (transcripts per 10 million, TP10M) respectively. The ratio (log base 2) for each TE was calculated using the average values of three endosperm libraries divided by the values from two leaf libraries. TEs with values >1 were categorized as putatively higher expression in endosperm, <−1 as putatively lower expression in endosperm, and values between them as no change between leaf and endosperm.

2.6. Bisulfite sequencing and analysis

Genomic DNA was isolated from leaves, roots, and 8-DAP and 15-DAP grains using a CTAB-based method. The EZ DNA Methylation-Gold kit was used for bisulfite treatment of genomic DNA according to the manufacturer's instructions (ZYMO Research). The bisulfite-treated DNA was amplified using EpiTaqTM HS (TaKaRa). PCR conditions were as follows: 94 °C, 3 min; 40 cycles of (94 °C, 30s; 55 °C, 30s; 72 °C, 30s); 72 °C, 8 min. PCR products were ligated into pGEM-T easy vector (Promega), and at least 20 clones were sequenced. Sequencing data was analyzed using Kismeth software (Gruntman et al., 2008) and percentage methylation (%C) was calculated as 100 × C/(C + T). Primers used for TREP258 were as follows: 5′-GGTGGTAGAAAAY YTGAAGTCGA-3′ and 5′-CAAARCACRRTCTACAAAACTAT-3′. Primers used for TREP3257 were as follows: 5′-GAAGTTGAYYTTATTTAAA ATAYAT-3′ and 5′-TAATCTTCACACAACCCATRTAATACA-3′.