Identification of small RNAs associated with Meiotic Silencing by Unpaired DNA

Thomas M. Hammond,*1,2 William G. Spollen,†1 Logan M. Decker,§ Sean M. Blake,§ Gordon K. Springer,** and Patrick K.T. Shiu‡,2

*School of Biological Sciences, Illinois State University, Normal, Illinois 61790
†Informatics Research Core Facility, University of Missouri, Columbia, Missouri 65211
‡Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211
§DNA Core Facility, University of Missouri, Columbia, Missouri 65211
**Department of Computer Science, University of Missouri, Columbia, Missouri 65211
Running head: MSUD-associated small interfering RNAs

Key words: Illumina/Solexa sequencing; Meiotic Silencing by Unpaired DNA (MSUD); Neurospora crassa; RNA interference (RNAi); MSUD-associated small interfering RNAs (masiRNAs)

1These authors contributed equally to this work.

2Corresponding authors:
Thomas M. Hammond
250 Science Laboratory Building
Illinois State University
Normal, IL 61790
Phone: 309-438-2602
Email: tmhammo@ilstu.edu

and

Patrick K. T. Shiu
103 Tucker Hall
University of Missouri
Columbia, MO 65211
Phone: 573-884-0020
Email: shiu@missouri.edu
ABSTRACT

In *Neurospora crassa*, unpaired genes are silenced by a mechanism called meiotic silencing by unpaired DNA (MSUD). Although some RNAi proteins are necessary for this process, its requirement of small RNAs has yet to be formally established. Here we report the characterization of small RNAs targeting an unpaired region using Illumina sequencing.
Introduction

*Neurospora crassa*, a filamentous fungus, is comprised of a network of thread-like cells called hyphae. Like other coenocytic organisms, in which nuclei (and other organelles) share the same cytoplasm, it is especially susceptible to systemic invasion by viruses or transposons. Accordingly, several gene-silencing mechanisms are maintained in this fungus (Catalanotto *et al.* 2006; Dang *et al.* 2011). For example, Quelling, an RNA interference (RNAi) mechanism, silences the expression of hyperhaploid transgenes during the vegetative phase (Romano and Macino 1992). Repeat-induced point mutation (RIP), on the other hand, is a premeiotic process that targets duplicated sequences for G:C to A:T mutations (Cambareri *et al.* 1989).

We are interested in another genome surveillance system known as meiotic silencing by unpaired DNA (MSUD; Aramayo and Metzenberg, 1996; Shiu *et al.* 2001; Kelly and Aramayo, 2007). In this system, genes unpaired during the pairing stage in meiotic prophase I, as well as any homologous copies found in the zygote, are transiently silenced during the sexual phase. Our current model suggests that an unpaired DNA triggers the production of sequence-specific single-stranded aberrant RNAs, which are converted into double strands by the SAD-1 RNA-directed RNA polymerase (RdRP; Shiu and Metzenberg, 2002). A double-stranded RNA (dsRNA) is processed into single-stranded small interfering (si)RNAs (by the DCL-1 Dicer and the QIP exonuclease; Alexander *et al.* 2008; Xiao *et al.* 2010; Lee *et al.* 2010a), which subsequently direct the SMS-2 Argonaute slicer to cleave homologous mRNAs (Lee *et al.* 2003).
Illumina sequencing, previously known as Solexa sequencing, is a high-throughput sequencing method driven by a proprietary parallel sequence-by-synthesis technology (Bennett et al. 2005; Bentley et al. 2008). The standard single-read length, although not ideal for resolving short sequence repeats (Morozova and Marra 2008), is extremely suitable for large-scale analyses of small RNAs (Czech et al. 2008; Qi et al. 2009). A variety of small RNA species have been profiled in N. crassa, including qiRNAs (QDE-2–interacting small RNAs), microRNA-like RNAs (milRNAs), and dicer-independent small interfering RNAs (disiRNAs) (Lee et al. 2009; Lee et al. 2010b). Although MSUD depends on several proteins that are implicated in RNAi (Siomi and Siomi 2009), the involvement of small RNAs in the process has not been formally established. In this work, we have set out to detect and characterize small RNAs targeting an unpaired region using Illumina sequencing.
RESULTS

The set up of a cross containing an unpaired region

Unpaired genes are targeted for silencing during sexual development in Neurospora. The silencing of genes important for sexual development often leads to aberrant ascus (spore sac) and/or ascospore phenotypes (Shiu et al. 2001). For example, when the Round spore (r) gene is unpaired during meiosis, the resulting ascospores are round instead of having the normal spindle shape (Figure 1A). We set up an experimental "unpaired" cross (F2-01 × P12-02; see Supporting Information, Materials and Methods) in which only one of the mating partners contains a deletion at the r locus (Figure 1B). A cross (F2-01 × P9-42) with neither of the mating partners deleted at the r locus (i.e., the two r homologues are paired) was used as a control.

The detection of small RNAs targeting the unpaired Round spore (r) locus

To determine if small RNAs matching a meiotically silenced locus are produced, we have isolated RNAs from fruiting bodies of an r-unpaired cross (as well as an r-paired control) and subjected the preparations to Illumina sequencing. Our data show a 10-fold spike in small RNAs targeting the unpaired region (hereafter, the r-specific RNAs) when the r gene is deleted in one of the mating partners (Figure 2A). Although the DCL-1 Dicer has been shown to be important for MSUD (Alexander et al. 2008) and a small RNA pathway is implicated in its mechanism, this is the first demonstration that the presence of small RNAs correlates with meiotic silencing. These small RNAs are referred to as MSUD-associated small interfering RNAs (masiRNAs) hereafter.
The length distribution of masiRNAs

Previously, diced RNA products from *N. crassa* vegetative cell-free extracts were estimated to be 21–26 nt long (peaking at ~23 nt; Refalo and Sachs 2004). Endogenous small RNAs from an unquelled strain, including milRNAs and disiRNAs, are mostly 17–27 nt long (Lee *et al.* 2010b). Here, the majority of masiRNAs targeting the *r* locus are found to be 21–27 nt in length (93.9%), with 25 nt being the dominant species (30.4%) (Figure 3).

The lack of a dicing preference for GC rich regions

Previous evidence suggests that Dicer proteins from certain plants (*e.g.*, *Dactylis glomerata*) show a preference for GC rich regions (Ho *et al.* 2008). Our nucleotide analysis shows that, on average, the *r*-specific masiRNAs have similar GC content to the targeted DNA region (51.3% vs. 51.5%; Figure S1). We therefore do not have any evidence suggesting that masiRNAs are preferentially produced from GC-rich regions of dsRNAs.

5' and 3' nucleotide biases

Functional small RNAs exhibit a preference for A/U at the 5' end of the molecule (Schwarz *et al.* 2003). An siRNA strand with the 5' end loosely bound to the complementary strand may be more likely to be incorporated into the RNA-induced silencing complex (RISC) and be preserved within the cell. Endogenous small RNAs from *Neurospora* vegetative tissues have a bias for 5' U (82.2%; Lee *et al.* 2010b). For *r*-specific masiRNAs, we have observed a preference for 5' U (74.9%), followed by A (13.3%), C (8.7%), and G (3.0%) (Figure S2A). The 5' nucleotide bias (for all four bases) is similar to the one observed in endogenous small RNAs by Lee *et al.*
There is a slight preference for G (41.9%) and a bias against U (7.6%) in position 2 of the 5' end. There are no obvious biases at the 3' end (Figure S2B).

**Sense vs. antisense sequences**

Of the r-specific masiRNAs identified, those that are antisense (*i.e.*, complementary to the sense r mRNA) account for 62.6% of the reads. The bias for antisense reads is consistent over three sequencing runs (recording 63.8%, 61.2%, and 62.5%). It is unclear whether this is the result of a real biological bias, and if so, what the underlying driving mechanism may be. One possibility is that nonfunctional sense masiRNAs (*i.e.*, those that are non-mRNA complementary and are not actively used by the RISC) are preferentially degraded. The majority (86.3%) of the r-specific small RNAs isolated from the “paired” control cross are sense RNAs and are likely the result of mRNA degradation.

**Secondary small RNAs**

RdRP can potentially amplify the silencing signal by using the mRNA target as a template through *de novo* and/or primer-dependent mechanisms (transitive RNAi; Voinnet 2008). For example, *Neurospora* QDE-1 RdRP is known to produce 9–21-nt RNAs from single-stranded RNA templates *in vitro* (Makeyev and Bamford 2002). A possible outcome of this amplification effect is the production of secondary siRNAs targeting sequences upstream and/or downstream of the initial trigger region. Although small RNAs matching the upstream and downstream paired regions (Figure 1B) can be detected, their numbers (0.0883 and 0.0579 small RNAs/bp per sequencing, respectively; Table S1) do not approach those that have originated from the (unpaired) target region (0.4661) (*P* < 0.05), nor are they significantly higher than the
background levels seen in the controls (0.0363 and 0.0313, respectively) \((P > 0.05)\). In other words, we do not have strong evidence for the presence of a robust transitive RNAi mechanism during MSUD.

**Introns vs. exons**

The \(r\) gene contains four introns (61, 51, 62, and 64 bp, respectively). There are 54.7 detectable small RNAs originating from the four introns per sequencing from an unpaired cross. This total is significantly more than the 3 small RNAs detected from an average control cross sequencing \((P < 0.05)\). For some reason, most (72.0\%) of the intron-originated small RNAs were derived from intron 1. It seems likely that introns (or at the very least, intron 1 of \(r\) in this case) are maintained in the dsRNA population and are subsequently represented in the final small RNA counts. The presence of intron-targeting small RNAs suggests that (1) at least some of the introns are not spliced from the sense aberrant RNAs and/or (2) at least some of the aberrant RNAs are antisense (which are complementary to the mRNA and do not encode consensus sequences for splicing).

To determine if aberrant RNAs are subject to splicing at all, we asked whether small RNAs targeting exon-exon (\(i.e.,\) without intron) junction sequences could be detected. Our data show that 14 small RNAs are found to be targeting the exon-exon sequences in an average unpaired cross sequencing, with the vast majority (95.2\%) of them being antisense (\(i.e.,\) they cannot be degradation products of sense \(r\) mRNAs). This is significant considering that less than one (\(i.e.,\) 0.5) exon-exon small RNA is detected in an average control cross sequencing (none of which are antisense) \((P < 0.05)\). The presence of small RNAs targeting exon-exon junctions may
hint that at least some of the aberrant RNAs are processed by the spliceosome. A less likely possibility is that these exon-exon specific RNAs could be secondary small RNAs resulting from the priming of siRNAs to mature mRNAs (which do not have introns).

**Reproducibility of site biases**

Small RNA-generating hot spots have been previously described in the biogenesis of siRNAs (Qi et al. 2009). We have observed reproducible site biases for small RNA detection within the r gene. For example, while the majority (95.63%) of the unpaired DNA (6934 nt when counting both strands) generate one or fewer detectable small RNA per sequencing, 23 sites generate at least 10 on average (Table S2). These sites have high small RNA counts across all three sequencing lanes, suggesting the bias is reproducible. If this is not a deep sequencing artifact, it may signify a yet-to-be identified Dicer/Argonaute preference. Alternatively, the thermodynamics of certain small RNAs may prolong their longevity (Khvororova et al. 2003).

**Endogenous small RNAs from the control cross**

Abundant small RNAs are detected from the control cross (which has no artificially unpaired regions). These endogenous perithecial small RNAs (epsRNAs) can be found in each of the seven linkage groups (chromosomes) as well as the mitochondrial genome of *Neurospora* (Figure S3). Some of the small RNA regions generate RNAs from both strands with similar abundance, while others have an obvious strand bias. The sizes of these epsRNAs are not of normal distribution and they are different from that of the r-specific masiRNAs (e.g., peaking at 20–21 nt instead of 25 nt; Figure S4). The top small RNA-producing regions are listed in Table
S3. Figure S5 illustrates the relative abundance of different types of RNAs found in the epsRNA population.

**Concluding remarks**

Although small RNAs have been implicated in the mechanism of MSUD, this study provides the first direct evidence of their involvement. The Illumina sequence analysis allows us to compare their characteristics with those identified from other RNA silencing systems. In *N. crassa*, vegetative silencing (Quelling) depends on the QDE-2 Argonaute and the redundant activity of DCL-1 and DCL-2 dicers, while MSUD requires the SMS-2 Argonaute and DCL-1 (Catalanotto *et al.* 2002; Catalanotto *et al.* 2004; Lee *et al.* 2003; Alexander *et al.* 2008). Small RNAs from vegetative and sexual stages have similar (but not identical) features. For example, the small RNAs associated with MSUD (masiRNAs) are mostly 21–27 nt in length, which are comparable to diced products from vegetative cells (Refalo and Sachs 2004). Also, both vegetative and sexual small RNAs have a strong bias for 5' U (Lee *et al.* 2010b).

The nucleotide analysis of masiRNAs suggests that MSUD does not have a dicing preference for GC rich regions, unlike the observations previously made in certain plants (Ho *et al.* 2008). Our data also do not support the existence of a robust transitive RNAi mechanism in MSUD, a finding similar to that seen in an *in vivo* QDE-1 assay during vegetative silencing (Lee *et al.* 2009). If a weak transitive RNAi system does exist in *N. crassa*, it is unlikely that it contributes greatly to the production of secondary siRNAs. Other features of masiRNAs offer signposts for future directions in MSUD studies. For example, there is indirect evidence
suggesting that certain aberrant RNA introns are processed. Also, reproducible site biases for small RNA production indicate the presence of a previously unknown MSUD preference.

Most of the endogenous small RNAs from the paired control cross (epsRNAs) match gene-coding regions (28.2%), presumptive noncoding regions (29.2%), and rDNA regions (35.6%) (Figure S5). While it is easy to deduce the biogenesis mechanism for some of these endogenous RNAs (e.g., milRNAs contain self-complementary structures and could become a Dicer substrate), it is unclear how the others are generated. Some possible explanations include: (1) The presence of regions that are naturally unpaired between the two mating partners due to chromosomal rearrangements (e.g., deletions, duplications, inversions, and translocations). (2) Convergent transcription of both coding and non-coding strands. For example, a significant proportion of the fly genome has convergent transcription (Czech et al. 2008). In Neurospora, the non-coding NTS regions of the rDNA tandem repeats and some of the disiRNA-producing loci are known to generate both sense and antisense transcripts, leading to the production of dsRNAs and siRNAs (Cecere and Cogoni 2009; Lee et al. 2009; Lee et al. 2010b). (3) Other yet-to-be described Neurospora mechanisms that target gene repeats and/or noncoding regions.

Although sharing certain similar features with other small RNAs, masiRNAs appear to have their own unique characteristics and biases. The variation is not surprising as different silencing machineries are known to have their unique preferences (Ho et al. 2008; Takeda et al. 2008). The deep sequencing approach we have utilized is more sensitive and less ambiguous than various hybridization methods, and it will no doubt be useful in future studies of fungal small RNAs as well as the machinery responsible for their biogenesis. For example, in our
companion study, we have shown that the generation of masiRNAs requires certain MSUD proteins (see Hammond et al. 2013 in this issue).

Acknowledgments

We thank James Birchler, Nathan Bivens, Joe Forrester, the Neurospora Functional Genomics Group, the Fungal Genetics Stock Center, members of the Shiu Laboratory, and colleagues from our community for their help. T.M.H. was the recipient of a Life Science fellowship from the University of Missouri and a Ruth L. Kirschstein National Research Service Award from the National Institute of General Medical Sciences. This work was supported by National Science Foundation grants MCB0918937/1157942 (to P.K.T.S.).
Literature Cited


Bennett, S. T., C. Barnes, A. Cox, L. Davies, and C. Brown, 2005  Toward the 1,000 dollars human genome. Pharmacogenomics 6: 373–382.


Cecere, G., and C. Cogoni, 2009  Quelling targets the rDNA locus and functions in rDNA copy number control. BMC Microbiol. 9: 44.


Figure Legends

**Figure 1** The *Round spore* (*r*) gene. (A) A wild-type cross (*r*⁺ × *r*⁺) gives spindle-shaped ascospores (resembling an American football) while a cross containing an unpaired copy of the *r* gene (*r*⁺ × *r*Δ) gives round spores. (B) In the experimental “*r*-unpaired” cross (*r*⁺ × *r*Δ), positions 9281458–9284924 (supercontig 1; *i.e.*, the target region) of the *r* gene are deleted in one of the mating partners. Translated exon sequences are shown as open boxes. Upstream and downstream paired regions are labeled in red.

**Figure 2** MSUD-associated small interfering RNAs (masiRNAs) are produced when a gene is unpaired during meiosis. The total numbers of small RNA (18–30 nt) reads are pooled from (A) three sequencings from an unpaired cross and (B) two sequencings from a paired (control) cross. A forward read (in blue) corresponds to the presence of an RNA matching the + strand sequence listed in the Neurospora Genome database. *r*-specific small RNAs are 10-fold more abundant when the *r* gene is unpaired in a cross (1616 vs. 157 reads per sequencing). The forward reads (which correspond to antisense masiRNAs that are complementary to the sense *r* mRNA) are more abundant than the reverse reads (1011 vs. 605 reads per sequencing) in the unpaired cross. The opposite is true (*e.g.*, 22 forward/antisense vs. 135 reverse/sense reads per sequencing) in the paired (control) cross. Open reading frames are labeled in green.

**Figure 3** The length distribution of masiRNAs. The majority of small RNAs targeting a meiotically-unpaired gene are between 21 and 27 nt and peaking at 25 nt.
A

\( r^+ \times r^+ \)

\( r^+ \times r^\Delta \)

B

\( r^+ \)

\( r^\Delta \)
A

+ Strand

Number of small RNAs

- Strand

NCU02765  r  NCU02763  r+  rΔ

B

r-paired

Number of small RNAs

NCU02765  r  NCU02763  r+  rΔ