*Nucleic Acids Research, 2005, Vol. 33, Database issue* ***D637–D640***
*doi:10.1093/nar/gki127*

# ASRP: the *Arabidopsis* Small RNA Project Database

**Adam M. Gustafson[1,2], Edwards Allen[1,2], Scott Givan[1], Daniel Smith[1,2], James C. Carrington[1,2] and Kristin D. Kasschau[1,2,*]**

[1]Center for Gene Research and Biotechnology and [2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA

Received August 16, 2004; Revised and Accepted October 24, 2004

## ABSTRACT

**Eukaryotes produce functionally diverse classes of small RNAs (20–25 nt). These include microRNAs (miRNAs), which act as regulatory factors during growth and development, and short-interfering RNAs (siRNAs), which function in several epigenetic and post-transcriptional silencing systems. The *Arabidopsis* Small RNA Project (ASRP) seeks to characterize and functionally analyze the major classes of endogenous small RNAs in plants. The ASRP database provides a repository for sequences of small RNAs cloned from various *Arabidopsis* genotypes and tissues. Version 3.0 of the database contains 1920 unique sequences, with tools to assist in miRNA and siRNA identification and analysis. The comprehensive database is publicly available through a web interface at http://asrp.cgrb.oregonstate.edu.**

## INTRODUCTION

Recent studies revealed that plants contain populations of small RNAs (20–25 nt) that belong to two major classes—microRNAs (miRNAs) and endogenous short-interfering RNAs (siRNAs). *MIRNA* gene transcripts adopt imperfect foldback structures and are processed by DICER-LIKE 1 (DCL1), resulting in 20–22 nt miRNAs. Mature miRNAs function as post-transcriptional regulators that guide either site-specific cleavage or non-degradative repression of target mRNAs (1). In many cases, disruption of miRNA-mediated control results in severe developmental abnormalities (2–10).

siRNAs arise from endogenous transcripts that form dsRNA structures, or that are substrates for RNAi pathways. Processing of siRNAs often requires other DCL proteins, such as DCL3 (11). In addition, biogenesis of several classes of endogenous siRNAs requires RNA-dependent RNA polymerases, such as RDR2 (11). siRNA-generating loci often yield multiple, overlapping clusters of small RNAs, in contrast to *MIRNA* loci that generally yield a single miRNA. Endogenous siRNAs arise from repetitive sequences, transposons and retroelements, genomic regions containing inverted duplications, as well as other genic and intergenic regions. A subset of siRNAs also act to guide or assist formation of heterochromatin (12–14). A subclass of siRNAs has been shown to guide cleavage of specific target mRNAs in *trans*, similar to miRNAs. Biogenesis of *trans*-acting siRNAs (ta-siRNAs) requires DCL1 and RDR6 (15,16). In contrast to miRNA genes, ta-siRNA precursor transcripts do not form a foldback structure, but rather both sense and antisense small RNAs are processed from perfectly complementary RNA duplexes.

Several small RNA libraries have been constructed from *Arabidopsis thaliana* plants with the primary goal to identify miRNAs and endogenous siRNAs (17–22). The aim of the *Arabidopsis* Small RNA Project (ASRP) is to analyze small RNAs from different tissues and genotypes of *Arabidopsis*, provide a public database of cloned small RNA sequences and develop web-based tools to assist in analysis of small RNA populations. These resources are intended to aid the identification of miRNAs and *MIRNA* genes, and to enable functional analysis of siRNA-producing regions of the genome.

## DATABASE CONTENT

The ASRP database currently contains 5521 small RNA entries representing 1920 unique sequences. The collection represents small RNA sequences from both in-house cloning projects and sequences deposited in the miRNA registry (23). For sequences derived in-house, multiple small RNA libraries were constructed from *Arabidopsis* (Columbia-0 ecotype) at various developmental stages, including embryos, 3-day post germination seedlings, aerial tissues (including rosette leaves and apical meristems) and inflorescences (stages 1–12). To genetically enrich for miRNA populations, libraries were constructed from *rdr2-1* and *dcl3-1* mutants that have defects in the chromatin siRNA pathway. All unique sequences were given an independent ASRP database (DBE) identifier.

## DATABASE ORGANIZATION

The ASRP database relies on freely available and open-source software. The ASRP graphical user interface (GUI) is

*To whom correspondence should be addressed. Tel: +1 541 737 3679; Fax: +1 541 737 3045; Email: kasschau@cgrb.oregonstate.edu

composed of web pages delivered by an apache HTTP server (http://httpd.apache.org). In addition, the server incorporates mod_perl (http://perl.apache.org) and Mason (http://www.masonhq.org) to dynamically produce web pages based upon user input. The vast majority of the GUI is generated by custom Perl code that increasingly incorporates object-oriented coding practices to improve extensibility and re-usability of the individual software components. Bioperl (24) is used for specific tasks, such as parsing the GenBank files containing the *Arabidopsis* chromosomes. The GUI interacts with a custom database backend utilizing Structured Query Language (SQL) and the open source MySQL (http://www.mysql.com) database engine. Table structures and specific query statements conform to standard SQL language syntax and are portable to other SQL database engines. Currently, the ASRP database resides on a custom-configured server managed by the RedHat Linux AS operating system.

## DATA ACCESS AND WEB INTERFACE

The ASRP database web interface enables users to view and analyze the small RNAs in text and graphical formats. Data for each small RNA is stored in MySQL database tables that are easily sorted and searched. Through the web interface, users may sort and view the small RNA data in the following ways:

(i) *All small RNAs*. This page displays basic information about all unique small RNAs in the database, including, if applicable, the miRNA or ta-siRNA name, number of loci in the *Arabidopsis* genome, number of near predicted loci in the Rice genome, number of potential mRNA targets and number of times isolated. More information about a specific small RNA is available by following the database number (DBE#) link.

(ii) *Small RNA clusters*. Some small RNA loci are clustered in the *Arabidopsis* genome. This page displays clusters containing a minimum of four small RNA loci, with each within 500 nt of the next small RNA loci. From this page, the user can view the sequences and positions of the small RNAs in each cluster in text format or the cluster can be viewed graphically in relation to the *Arabidopsis* genome using an open access genome viewer (25).

(iii) *miRNAs*. All small RNAs characterized as miRNAs are displayed in a similar format as section (i) (Figure 1A). The display page for each individual miRNA is split into four sections; general information, *Arabidopsis MIRNA* genes, predicted and validated target genes, and *Oryza sativa MIRNA* genes (Figure 1B). The general information section includes the sequence and source of the miRNA. The predicted foldback structure for the pre-miRNA, the flanking sequence around the *MIRNA* gene and the graphical genome view are available through links on the *Arabidopsis MIRNA* genes section (Figure 1C). Information about the predicted target genes, the target-miRNA binding site, and the computational or experimental validity of the target-miRNA binding site is displayed in the third section (Figure 1B). The fourth section displays information about the small RNA in *O.sativa*, including the predicted secondary structure of validated precursor miRNAs.

(iv) *ta-siRNAs*. All published ta-siRNAs are in the database. The page is similar in format to the miRNA page. General information, ta-siRNA-generating locus information, and predicted target genes are displayed. The user can view the information about the ta-siRNAs in a manner similar to the miRNA section of the database.

(v) *Annotated small RNAs*. Automated annotation programs such as RepeatMasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html) are used to identify small RNAs that originate from genomic regions of highly repetitive sequences, as well as transposons and retro-elements. The user can display and sort small RNAs by the specific class of annotated repeat element such as MuDR or SINE.

In addition to the sorting features, the web interface provides users with a variety of searching capabilities. Quick searches enable users to locate specific miRNAs based on either the miRNA names or the ASRP database identifiers (DBE#). To search for small RNAs predicted to target specific *Arabidopsis* genes, or that originate from generic sequences, the locus identifiers (e.g. At3g60630) or user-defined FASTA formatted sequences are used, respectively. Finally, users can determine if a small RNA sequence is represented in the ASRP database by searching the sequence against the entire population of small RNAs.

## AVAILABILITY

All small RNAs in the ASRP database are available through the publicly available website (http://asrp.cgrb.oregonstate.edu) or can be downloaded in FASTA format from the website download page (http://asrp.cgrb.oregonstate.edu/downloads/).

## FURTHER DIRECTIONS

The ASRP database was created to serve as a repository and tool to facilitate the analysis of miRNAs and endogenous siRNAs and their targets. To increase accessibility of the database, we are working to more completely integrate the ASRP database with existing *Arabidopsis* resources, such as TAIR. In addition, integration of miRNAs, ta-siRNAs and endogenous siRNAs from the database with other research projects, such as genomic tilling microarrays and chromatin immunoprecipitation arrays (14,26), will enhance the information acquired from these experiments and further expand our understanding of small RNA function.

There are still many unanswered questions concerning miRNAs, ta-siRNAs and endogenous siRNAs. The regulatory roles of miRNA-target gene interaction, the regulation of *MIRNA* gene expression, and the function of siRNAs in the regulation of chromatin structure and gene silencing are just a few questions currently being studied. Future plans include the integration of data from genome-scale microarray projects into the ASRP database (27). The scope of the database may widen with the addition of other plant genomes, libraries or computational analysis. The inclusion of additional plant genomes will enable a more in-depth study of miRNA evolution and conservation and activities of endogenous siRNAs.

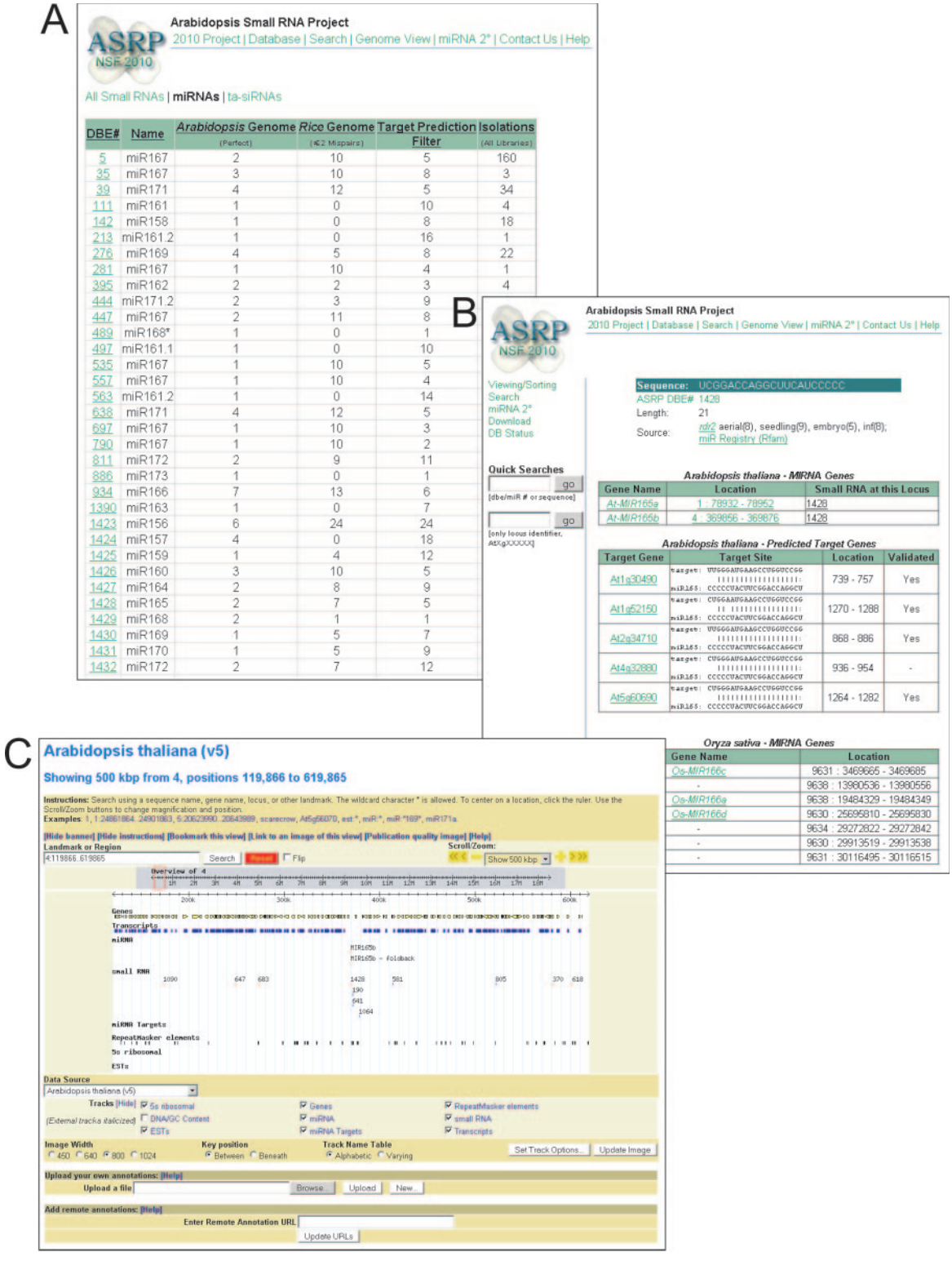


**Figure 1.** Windows from the ASRP database website. (**A**) A partial list of all miRNAs in the database. (**B**) Information specific to a single miRNA. (**C**) Display from the genome browser.

## ACKNOWLEDGEMENTS

We thank Christopher M. Sullivan for developing the infrastructure for the computing cluster and compiling the software technologies used for the database and Heather Fitzgerald for critical reading of the manuscript. The *Arabidopsis* small RNA project database is supported by a 2010 project grant from the National Science Foundation (MCB-0209836).

## REFERENCES


1. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
2. Chen,X. (2004) A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science*, **303**, 2022–2025.
3. Mallory,A.C., Dugas,D.V., Bartel,D.P. and Bartel,B. (2004) MicroRNA regulation of NAC-Domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. *Curr. Biol.*, **14**, 1035–1046.
4. Juarez,M.T., Kui,J.S., Thomas,J., Heller,B.A. and Timmermans,M.C. (2004) microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. *Nature*, **428**, 84–88.
5. Emery,J.F., Floyd,S.K., Alvarez,J., Eshed,Y., Hawker,N.P., Izhaki,A., Baum,S.F. and Bowman,J.L. (2003) Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. *Curr. Biol.*, **13**, 1768–1774.
6. Palatnik,J.F., Allen,E., Wu,X., Schommer,C., Schwab,R., Carrington,J.C. and Weigel,D. (2003) Control of leaf morphogenesis by microRNAs. *Nature*, **425**, 257–263.
7. Achard,P., Herr,A., Baulcombe,D.C. and Harberd,N.P. (2004) Modulation of floral development by a gibberellin-regulated microRNA. *Development*, **131**, 3357–3365.
8. Tang,G., Reinhart,B.J., Bartel,D.P. and Zamore,P.D. (2003) A biochemical framework for RNA silencing in plants. *Genes Dev.*, **17**, 49–63.
9. Vaucheret,H., Vazquez,F., Crete,P. and Bartel,D.P. (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.
10. Kidner,C.A. and Martienssen,R.A. (2004) Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. *Nature*, **428**, 81–84.
11. Xie,Z., Johansen,L.K., Gustafson,A.M., Kasschau,K.D., Lellis,A.D., Zilberman,D., Jacobsen,S.E. and Carrington,J.C. (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.*, **2**, 642–652.
12. Lippman,Z., May,B., Yordan,C., Singer,T. and Martienssen,R. (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.*, **1**, 420–428.
13. Zilberman,D., Cao,X. and Jacobsen,S.E. (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*, **299**, 716–719.
14. Lippman,Z., Gendrel,A.V., Black,M., Vaughn,M.W., Dedhia,N., McCombie,W.R., Lavine,K., Mittal,V., May,B., Kasschau,K.D. *et al.* (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature*, **430**, 471–476.
15. Peragine,A., Yoshikawa,M., Wu,G., Albrecht,H.L. and Poethig,R.S. (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. *Genes Dev.*, **18**, 2368–2379.
16. Vazquez,F., Vaucheret,H., Rajagopalan,R., Lepers,C., Gasciolli,V., Mallory,A.C., Hilbert,J.L., Bartel,D.P. and Crete,P. (2004) Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell*, **16**, 69–79.
17. Llave,C., Kasschau,K.D., Rector,M.A. and Carrington,J.C. (2002) Endogenous and silencing-associated small RNAs in plants. *Plant Cell*, **14**, 1605–1619.
18. Reinhart,B.J., Weinstein,E.G., Rhoades,M.W., Bartel,B. and Bartel,D.P. (2002) MicroRNAs in plants. *Genes Dev.*, **16**, 1616–1626.
19. Mette,M.F., van der Winden,J., Matzke,M. and Matzke,A.J. (2002) Short RNAs can identify new candidate transposable element families in *Arabidopsis*. *Plant Physiol.*, **130**, 6–9.
20. Park,W., Li,J., Song,R., Messing,J. and Chen,X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.*, **12**, 1484–1495.
21. Sunkar,R. and Zhu,J.K. (2004) Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell*, **16**, 2001–2019.
22. Jones-Rhoades,M.W. and Bartel,D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell*, **14**, 787–799.
23. Griffiths-Jones,S. (2004) The microRNA Registry. *Nucleic Acids Res.*, **32**, D109–D111.
24. Stajich,J.E., Block,C., Boulez,K., Brenner,S.E., Chervitz,S.A., Dagdigian,C., Fuellen,G., Gilbert,J.G., Korf,I., Lapp,H. *et al.* (2002) The Bioperl Toolkit: Perl modules for the life science. *Genome Res.*, **12**, 1611–1618.
25. Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W., Arva,A. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
26. Yamada,K., Lim,J., Dale,J.M., Chen,H., Shinn,P., Palm,C.J., Southwick,A.M., Wu,H.C., Kim,C., Nguyen,M. *et al.* (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science*, **302**, 842–846.
27. Schmid,M., Uhlenhaut,N.H., Godard,F., Demar,M., Bressan,R., Weigel,D. and Lohmann,J.U. (2003) Dissection of floral induction pathways using global expression analysis. *Development*, **130**, 6001–6012.