

WMP: A novel comprehensive wheat miRNA database, including related bioinformatics software

Mohamed Amine Remita¹, Etienne Lord^{1*}, Zahra Agharbaoui², Mickael
Leclercq³, Mohamed A. Badawi², Vladimir Makarenkov¹, Fathey Sarhan², and
Abdoulaye Baniré Diallo^{1**}

¹ Department of Computer Science, Université du Québec à Montréal, P.O. Box 8888
Downtown Station, Montreal, Quebec, H3C 3P8, Canada.

² Department of Biological Sciences, Université du Québec à Montréal, P.O. Box 8888
Downtown Station, Montreal, Quebec, H3C 3P8, Canada.

³ School of Computer Science and McGill Centre for Bioinformatics, McGill University,
Montreal, Quebec, H3A2B2, Canada

Abstract. MicroRNAs (miRNAs) are emerging as important post transcriptional regulators that may regulate key plant genes responsible for agronomic traits such as grain yield and stress tolerance. Several studies identified species and clades specific miRNA families associated with plant stress regulated genes. Here, we propose a novel resource that provides data related to the expression of abiotic stress responsive miRNAs in wheat, one of the most important staple food crops. This database allows

* Mohamed Amine Remita and Etienne Lord contributed equally to this work

** Corresponding author: diallo.abdoulaye@uqam.ca

the query of small RNA libraries, including *in silico* predicted wheat miRNA sequences and the expression profiles of small RNAs identified from those libraries. Our database also provides a direct access to online miRNA prediction software tuned to *de novo* miRNA detection in wheat, in monocotyledon clades, as well as in other plant species. These data and software will facilitate multiple comparative analyses and reproducible studies on small RNAs and miRNA families in plants. Our web-portal is available at: <http://wheat.bioinfo.uqam.ca>.

KEYWORDS: ABIOTIC STRESSES AND DEVELOPMENT, EXPRESSED
SEQUENCED TAGS, MIRNA DATABASE, NEXT-GENERATION
SEQUENCING, WHEAT

1 Introduction

Next-generation sequencing offers interesting tools for generating and searching for microRNAs (miRNAs) that are expressed during growth and development. This approach provides several small RNA libraries and miRNA candidates from hexaploid wheat (*Triticum aestivum* L.) grown under different stress conditions such as heat, cold, and powdery mildew and *fusarium* infection. MiRBase, the current general database of miRNA sequences, is the main resource to access miRNAs data. Unfortunately, only 119 experimentally validated wheat miRNAs are available in the latest release of miRBase (Kozomara and Griffiths-Jones, 2011). Furthermore, the miRBase interface is not suitable for the analysis and assessment of miRNA candidates according to their original biological libraries

and experimental conditions. To circumvent this limitation, we developed a novel comprehensive web-based server dedicated to wheat miRNAs, with emphasis on stress responsive miRNAs from different wheat genotypes and tissues. This database stores and displays differential gene expression data from several small RNA libraries. It also includes the data from 20 different miRNA studies (references are listed in the database interface). Moreover, the web portal yields and integrated views of the pre-miRNA hairpin structures and the related predicted targets. It also includes two miRNA predictors: *miRdup* (Leclercq *et al.*, 2013) and *MIRcheck* (Jones-Rhoades *et al.*, 2006) tuned for different evolutionary clades (wheat, cereals or plants).

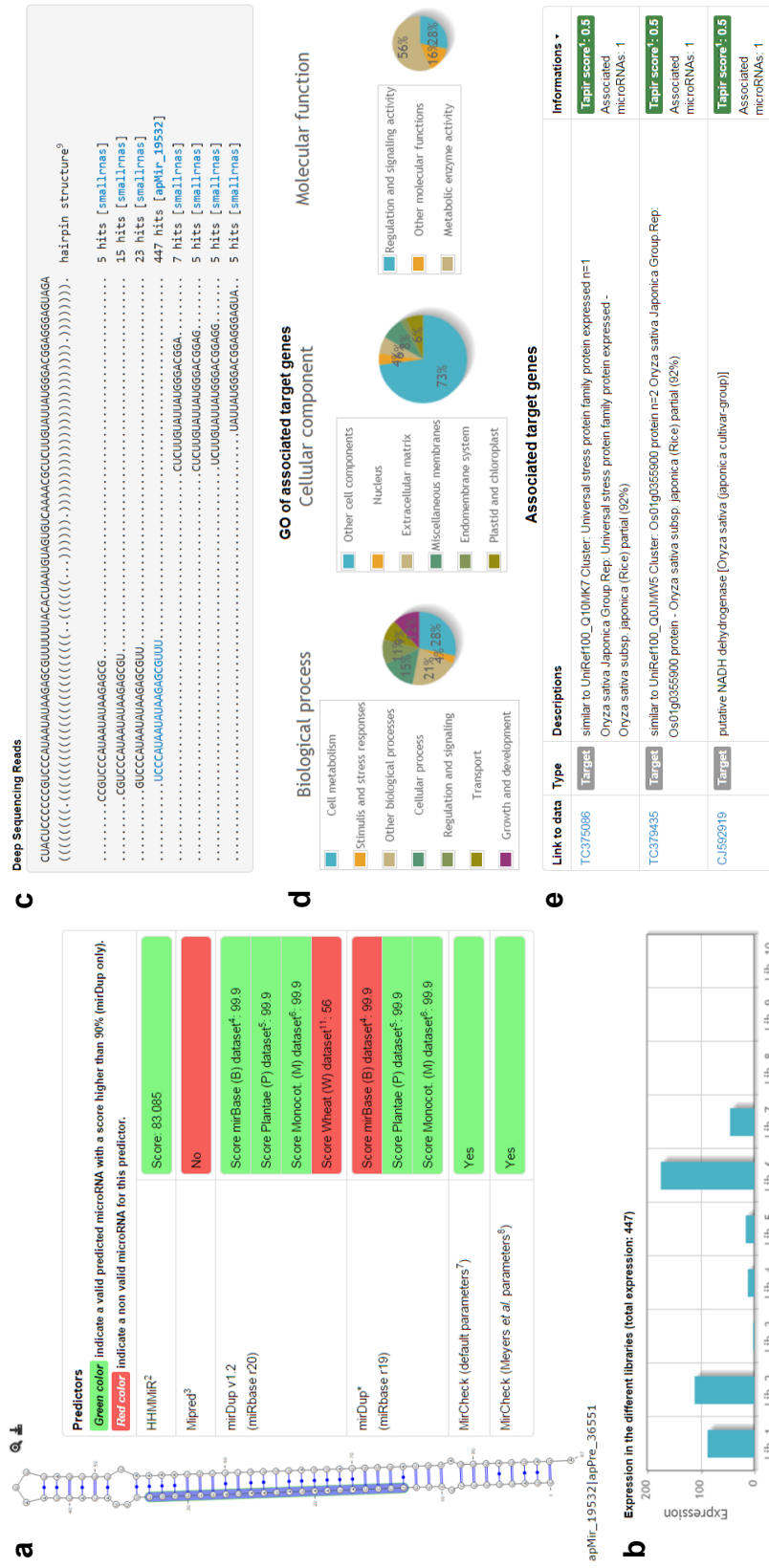


Fig. 1. Overview of the main features of our miRNA database including (a) two-dimensional hairpin structures and the associated prediction scores for the miRNA apMir_19532, (b) expression view of small RNA expression and miRNA position against the computed pre-miRNA, (c) Gene Ontology annotations associated with a set of target genes and (d) description of the obtained target genes.

2 Database content and statistics

The current database includes data of ten small RNA libraries produced from plants grown under different abiotic stress conditions and development stages. From these libraries, a set of 168,834 unique expressed small RNAs are illustrated as well as 267 evolutionary conserved miRNAs according to miRBase (Kozomara and Griffiths-Jones, 2011). The database also contains a broad description of 5,036 published wheat miRNAs. It also includes detailed presentation of 199 newly identified wheat miRNAs, 1,390 associated target genes, and 1.4 millions EST with 127,039 Uniref clusters, collected from seven wheat databases issued from Agharbaoui *et al.* (2015). Putative target genes were identified for each miRNA using the Tapir program (Bonnet *et al.*, 2010). Gene ontology (GO) associations and enrichments were also performed for associated targets resulting of 2,561 biological process, 1,616 cellular components, and 1,386 molecular function associations. To enhance the visualization of miRNA structures, a high quality picture of each miRNA and each pre-miRNA hairpin were generated using the Varna package (Darty *et al.*, 2009). Figure 1 highlights an example of graphical and statistical features of miRNA candidate in our database.

3 User interfaces

The database users have the following four main options: A) *Basic search*: the user can use keywords that correspond to one or multiple entries from the following categories: validated miRNAs, pre-miRNAs or the hairpin-folded structure motifs

(dotbracket notation), associated target gene accession identifier, Uniref identifier or name, EST sequences or identifications, and GO descriptions or identifiers (e.g. GO:0008152). B) *Advanced search*: the user can search for differentially expressed miRNAs by choosing the growth conditions and selecting a list of miRNA patterns (upregulated, downregulated, not differentially expressed, or not related to any of the conditions). The metrics for statistical significance (*p-values*) and the miRNA expression fold-change between libraries could also be set to filter candidates. C) *Data* menu: provides a direct access to predicted miRNAs, conserved miRNAs, the associated target genes and ESTs. This menu also gives access to *Libraries and conditions* option. This option allows the profiling all miRNAs expressed in any given libraries or under given stress conditions. D) *Tools* menu: provides five important applications for studying small RNAs libraries and miRNAs. It includes *MIRcheck v1.0* (Jones-Rhoades and Bartel, 2004), *miRdup v1.2* (Leclercq *et al.*, 2013), *Small RNA finder*, *Library comparison* as well as a *Blast* interface (Altschul *et al.*, 1990; Camacho *et al.*, 2009). *MIRcheck* and *miRdup* are miRNA prediction tools. For both predictors, the users can submit their candidate miRNAs, miRNA precursors, or hairpin secondary structures for analysis. The *MIRcheck* tool allows two modes of computation (default parameters used by Jones-Rhoades and Bartel (2004) and the universal plant miRNA criteria given by Meyers *et al.* (2008)). The *miRdup* predictor (based on a machine learning approach) allows a selection among four evolutionary clade-training sets. These sets are extracted from experimentally validated miRNA sequences found in miRBase (all miRNAs of miRBase, viridiplantae (plant), monocotyledons and

wheat (*Triticum aestivum L.*). *Small RNA finder* option permits users to search multiple sequences in a single query against putative miRNA and small RNA database. *The library comparison* option allows extracting differential expressed small RNAs between two libraries. In the *Blast* option, users can identify the location of their query sequences (nucleotide and amino acid sequences) against the wheat whole sequenced genome using the *Blast* tool. Further details on our database and a demonstration of some of its features are provided in the following section.

4 Study case: searching for miRNAs regulating glutathione S-transferases

The usefulness of our miRNA database is shown with an example of the search of miRNAs that regulate glutathione S-transferase (GST) enzyme family in wheat. GSTs are multifunctional proteins known for their important roles in both normal cellular metabolisms and the detoxification of a wide variety of products under stress conditions (Marrs, 1996; Jain *et al.*, 2010). GSTs have been shown to be miRNA targets in various other plant species exposed to stresses, such as osa-miR1848 in rice (Li *et al.*, 2010; Shaik and Ramakrishna, 2012), rsa-miR156 in radish (Xu *et al.*, 2013), mir168 homolog in sweet potato (Dehury *et al.*, 2013) and ssp-mir169 in sugarcane (Gentile *et al.*, 2013; Menossi *et al.*, 2015). Using the basic keyword search option, we first looked for the complete name of the GST enzyme: glutathione S-transferase. We found 31 target gene sequences, one microRNA gene and 30 gene ontology annotations. To ensure that we didn't

miss any other related data in the database we searched for alternative keywords related to the enzyme such as glutathione transferase and GST. Clicking on the target sequence identifier CJ893705 (with a good Tapir score of one), allows the user to access further related information such as the alignment with the miRNA apMir_11506, the Uniref clusters where the target belongs and the associated GO annotations. Information related to the found miRNA such as its sequence, length and expression details, can be easily obtained by clicking on its identifier link. In the *in silico* evidence section of the miRNA page, one could find that it is embedded in two precursors (apPre_12405 and apPre_8495) and it is predicted as valid miRNA with two different predictors *miRdup* ((Leclercq *et al.*, 2013) and *MIRcheck* (Jones-Rhoades and Bartel, 2004)). A deep sequencing reads view of precursors and their mapping small RNAs is appended to the evidence section. Interestingly, we noticed that apMir_11506 is regulated by cold and salinity. Furthermore, in miRBase section, we found that it has a 20-nucleotides homolog miRNA in *Arabidopsis thaliana* (ath-miR8175). These results show that in total, 5 miRNAs are associated with glutathione S-transferase: apMir_54471, apMir_15117, apMir_18589, apMir_11506 and apMir_19203. The first four miRNAs target the GST enzyme. apMir_18589 and apMir_19203 are generated by ESTs associated with GST Uniref. Analyzing the expression behavior in the different conditions reveals that miRNAs targeting the GST were regulated in aluminum (Al), salt and cold stresses from sensitive and tolerant wheat genotypes.

5 Conclusion

The WMP database presents a novel resource for the analysis of miRNAs in cereals. This first version offers an access to miRNAs and small RNAs in the context of their association with different growth and development stages under stress conditions in wheat. It constitutes a unique entry for wheat expressed small RNAs and miRNAs for several studies. Furthermore, it provides a direct access to useful miRNA predictors for wheat species, as well as for other cereals and plants. In the future, we plan to update this database regularly by adding new libraries prepared by our team for wheat grown under different stress conditions and newly published small RNAs in wheat and other cereals. We will provide confidence metrics over all the published miRNAs in wheat. We will also include an evolutionary toolkit to study miRNAs within all cereals. Further development will include an option to integrate complex miRNA analysis pipelines using the Armadillo workflow platform (Lord *et al.*, 2012).

6 acknowledgements

Acknowledgements This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) to FS and ABD and the Fond de Recherche du Québec-Nature et Technologie (FRQNT) to ABD. MAR is a NSERC and FRQNT fellow.

Bibliography

- Agharbaoui, Z., Leclercq, M., Remita, M. A., Badawi, M. A., Lord, E., Houde, M., Danyluk, J., Diallo, A. B., and Sarhan, F. 2015. An integrative approach to identify hexaploid wheat miRNAome associated with development and tolerance to abiotic stress. *BMC genomics*, 16(1): 339.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3): 403–410.
- Bonnet, E., He, Y., Billiau, K., and Van de Peer, Y. 2010. Tapir, a web server for the prediction of plant miRNA targets, including target mimics. *Bioinformatics*, 26(12): 1566–1568.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. 2009. Blast+: architecture and applications. *BMC bioinformatics*, 10(1): 421.
- Darty, K., Denise, A., and Ponty, Y. 2009. Varna: Interactive drawing and editing of the rna secondary structure. *Bioinformatics*, 25(15): 1974.
- Dehury, B., Panda, D., Sahu, J., Sahu, M., Sarma, K., Barooah, M., Sen, P., and Modi, M. K. 2013. In silico identification and characterization of conserved miRNAs and their target genes in sweet potato (*Ipomoea batatas* L.) expressed sequence tags (ESTs). *Plant signaling & behavior*, 8(12): e26543.
- Gentile, A., Ferreira, T. H., Mattos, R. S., Dias, L. I., Hoshino, A. A., Carneiro, M. S., Souza, G. M., Calsa Jr, T., Nogueira, R. M., Endres, L., *et al.* 2013. Effects of drought on the microtranscriptome of field-grown sugarcane plants. *Planta*, 237(3): 783–798.
- Jain, M., Ghanashyam, C., and Bhattacharjee, A. 2010. Comprehensive expression analysis suggests overlapping and specific roles of rice glutathione S-transferase genes during development and stress responses. *BMC genomics*, 11(1): 73.
- Jones-Rhoades, M. W. and Bartel, D. P. 2004. Computational identification of plant miRNAs and their targets, including a stress-induced miRNA. *Molecular cell*, 14(6): 787–799.
- Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. 2006. miRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.*, 57: 19–53.
- Kozomara, A. and Griffiths-Jones, S. 2011. miRBase: integrating miRNA annotation and deep-sequencing data. *Nucleic acids research*, 39(suppl 1): D152–D157.

- Leclercq, M., Diallo, A. B., and Blanchette, M. 2013. Computational prediction of the localization of micrnas within their pre-mirna. *Nucleic acids research*, 41(15): 7200–7211.
- Li, Y.-F., Zheng, Y., Addo-Quaye, C., Zhang, L., Saini, A., Jagadeeswaran, G., Axtell, M. J., Zhang, W., and Sunkar, R. 2010. Transcriptome-wide identification of microrna targets in rice. *The Plant Journal*, 62(5): 742–759.
- Lord, E., Leclercq, M., Boc, A., Diallo, A. B., and Makarenkov, V. 2012. Armadillo 1.1: an original workflow platform for designing and conducting phylogenetic analysis and simulations. *PloS One*, 7(1): e29903.
- Marrs, K. A. 1996. The functions and regulation of glutathione s-transferases in plants. *Annual review of plant biology*, 47(1): 127–158.
- Menossi, M. M., Gentile, A., Ferreira, T. H., Mattos, R., and Dias, L. I. 2015. Micrnas and drought responses in sugarcane. *Frontiers in Plant Science*, 6: 58.
- Meyers, B. C., Axtell, M. J., Bartel, B., Bartel, D. P., Baulcombe, D., Bowman, J. L., Cao, X., Carrington, J. C., Chen, X., Green, P. J., *et al.* 2008. Criteria for annotation of plant micrnas. *The Plant Cell Online*, 20(12): 3186–3190.
- Shaik, R. and Ramakrishna, W. 2012. Bioinformatic analysis of epigenetic and microrna mediated regulation of drought responsive genes in rice. *PloS one*, 7(11): e49331.
- Xu, L., Wang, Y., Zhai, L., Xu, Y., Wang, L., Zhu, X., Gong, Y., Yu, R., Limera, C., and Liu, L. 2013. Genome-wide identification and characterization of cadmium-responsive micrnas and their target genes in radish (*raphanus sativus* l.) roots. *Journal of experimental botany*, 64(14): 4271–4287.