



Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Current Plant Biology

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



Short Communication

A novel comprehensive wheat miRNA database, including related bioinformatics software

Mohamed Amine Remita^{a,1}, Etienne Lord^{a,1}, Zahra Agharbaoui^b, Mickael Leclercq^{a,c}, Mohamed A. Badawi^b, Fathey Sarhan^b, Abdoulaye Baniré Diallo^{a,*}

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

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

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

ARTICLE INFO

Article history:

Received 30 July 2016

Received in revised form 17 October 2016

Accepted 24 October 2016

Keywords:

Abiotic stresses and development

Expressed sequenced tags

miRNA database

Next-generation sequencing

Wheat

ABSTRACT

MicroRNAs (miRNAs) are emerging as important post-transcriptional regulators that may regulate key 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 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>.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

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 [1]. 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, WMP (Wheat MiRNA web-Portal). It is 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 contains the data from 20 different miRNA

studies (references are listed in the database interface). Moreover, the web portal yields and integrates views of the pre-miRNA hairpin structures and the related predicted targets. It also includes two miRNA predictors: *miRdup* [2] and *MIRcheck* [3] tuned for different evolutionary clades (wheat, cereals or plants).

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 [1]. The database also contains a broad description of 5036 published wheat miRNAs. It provides detailed presentation of 199 newly identified wheat miRNAs, 1390 associated target genes, and 1.4 millions ESTs with 127,039 Uniref clusters, collected from seven wheat databases issued from Agharbaoui et al. [4]. Putative target genes were identified for each miRNA using the Tapir program [5]. Gene ontology (GO) associations and enrichments were also identified for associated targets resulting in 2561 biological process, 1616 cellular component, and 1386 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

* Corresponding author.

E-mail address: diallo.abdoulaye@uqam.ca (A.B. Diallo).

¹ These authors contributed equally to this work.

<http://dx.doi.org/10.1016/j.cpb.2016.10.003>

2214-6628/© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

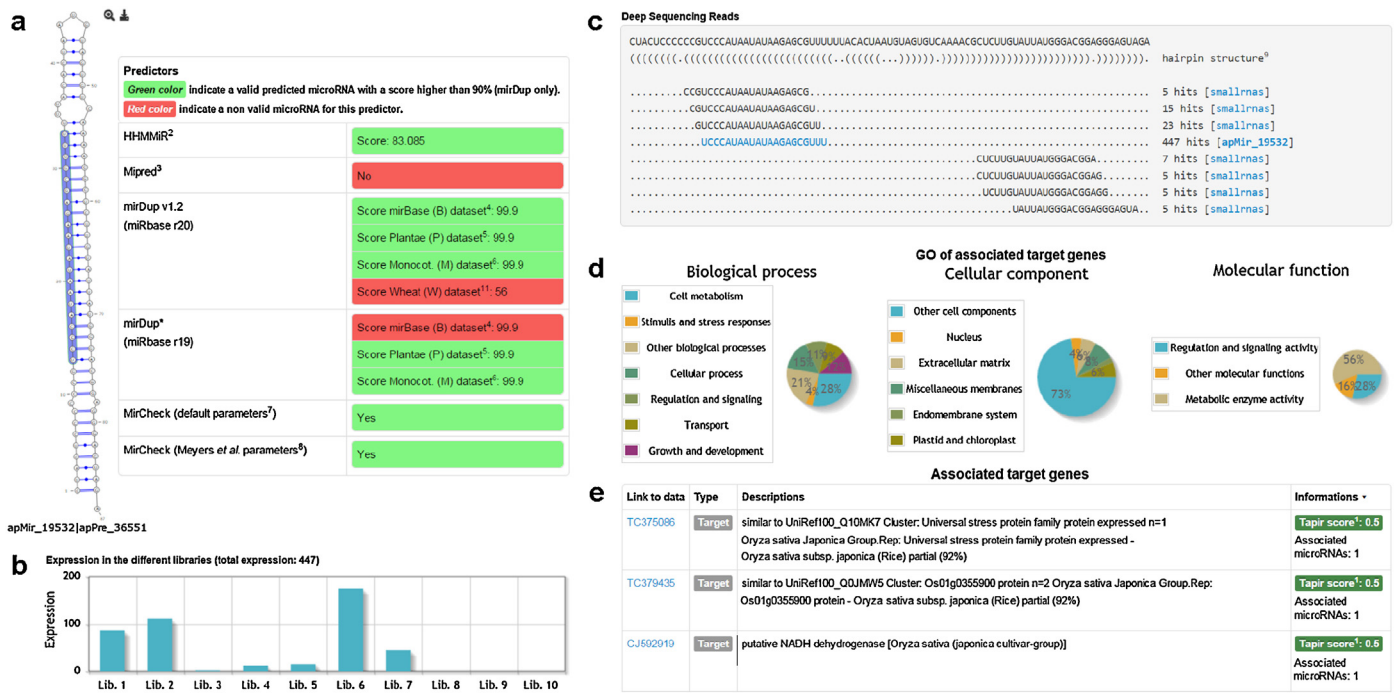


Fig. 1. Overview of the main features of our miRNA database including (a) two-dimensional hairpin structures and the associated prediction scores for the miRNA apMir_19532, (b) expression view of each library, (c) view of small RNA expression and miRNA position against the computed pre-miRNA, (d) Gene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

the Varna package [6]. Fig. 1 highlights an example of graphical and statistical features of miRNA candidates in the database.

3. User interface

The database provides the following four main options: (A) *Basic search*: allows keywords corresponding 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 profiling of all miRNAs expressed in any given libraries or under any given stress conditions. (D) *Tools menu*: provides five important applications for studying small RNAs libraries and miRNAs. It includes *MIRcheck v1.0* [7], *miRdup v1.2* [2], *Small RNA finder*, *Library comparison* as well as a *Blast* interface [8,9]. *MIRcheck* and *miRdup* are miRNA prediction tools. For both predictors, users can submit 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 [7] and the universal plant miRNA criteria given by Meyers et al. [10]). 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 for

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 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 [11,12]. GSTs have been shown to be miRNA targets in various other plant species exposed to stresses, such as osa-miR1848 in rice [13,14], rsa-miR156 in radish [15], mir168 homolog in sweet potato [16] and ssp-mir169 in sugarcane [17,18]. 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 miRNA gene and 30 gene ontology annotations. To ensure that we did not 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* [2] and *MIRcheck* [7]). 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 of *Arabidopsis thaliana* (ath-miR8175). These results show that in total, 5 miRNAs are associated with GST: 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, 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 also include an evolutionary toolkit to study miRNAs within all cereals. Further development will include option to integrate complex miRNA analysis pipelines using the Armadillo workflow platform [19].

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) to FS and ABD and the Fonds de recherche du Québec – Nature et technologies (FRQNT) to ABD. MAR, EL and ML are FRQNT fellows. MAR is a NSERC fellow.

References

- [1] A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data, *Nucleic Acids Res.* 39 (Suppl. 1) (2011) D152–D157.

- [2] M. Leclercq, A.B. Diallo, M. Blanchette, Computational prediction of the localization of microRNAs within their pre-miRNA, *Nucleic Acids Res.* 41 (15) (2013) 7200–7211.
- [3] M.W. Jones-Rhoades, D.P. Bartel, B. Bartel, MicroRNAs and their regulatory roles in plants, *Annu. Rev. Plant Biol.* 57 (2006) 19–53.
- [4] Z. Agharbaoui, M. Leclercq, M.A. Remita, M.A. Badawi, E. Lord, M. Houde, J. Danyluk, A.B. Diallo, F. Sarhan, An integrative approach to identify hexaploid wheat miRNAome associated with development and tolerance to abiotic stress, *BMC Genomics* 16 (1) (2015) 339.
- [5] E. Bonnet, Y. He, K. Billiau, Y. Van de Peer, TAPIR, a web server for the prediction of plant microRNA targets, including target mimics, *Bioinformatics* 26 (12) (2010) 1566–1568.
- [6] K. Darty, A. Denise, Y. Ponty, VARNA: interactive drawing and editing of the RNA secondary structure, *Bioinformatics* 25 (15) (2009) 1974.
- [7] M.W. Jones-Rhoades, D.P. Bartel, Computational identification of plant microRNAs and their targets, including a stress-induced miRNA, *Mol. Cell* 14 (6) (2004) 787–799.
- [8] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (3) (1990) 403–410.
- [9] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, *BMC Bioinformatics* 10 (1) (2009) 421.
- [10] B.C. Meyers, M.J. Axtell, B. Bartel, D.P. Bartel, D. Baulcombe, J.L. Bowman, X. Cao, J.C. Carrington, X. Chen, P.J. Green, et al., Criteria for annotation of plant MicroRNAs, *Plant Cell* 20 (12) (2008) 3186–3190.
- [11] K.A. Marrs, The functions and regulation of glutathione S-transferases in plants, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47 (1) (1996) 127–158.
- [12] M. Jain, C. Ghanashyam, A. Bhattacharjee, Comprehensive expression analysis suggests overlapping and specific roles of rice glutathione S-transferase genes during development and stress responses, *BMC Genomics* 11 (1) (2010) 73.
- [13] Y.-F. Li, Y. Zheng, C. Addo-Quaye, L. Zhang, A. Saini, G. Jagadeeswaran, M.J. Axtell, W. Zhang, R. Sunkar, Transcriptome-wide identification of microRNA targets in rice, *Plant J.* 62 (5) (2010) 742–759.
- [14] R. Shaik, W. Ramakrishna, Bioinformatic analysis of epigenetic and microRNA mediated regulation of drought responsive genes in rice, *PLoS ONE* 7 (11) (2012) e49331.
- [15] L. Xu, Y. Wang, L. Zhai, Y. Xu, L. Wang, X. Zhu, Y. Gong, R. Yu, C. Limeria, L. Liu, Genome-wide identification and characterization of cadmium-responsive microRNAs and their target genes in radish (*Raphanus sativus* L.) roots, *J. Exp. Bot.* (2013), ert240.
- [16] B. Dehury, D. Panda, J. Sahu, M. Sahu, K. Sarma, M. Barooah, P. Sen, M.K. Modi, In silico identification and characterization of conserved miRNAs and their target genes in sweet potato (*Ipomoea batatas* L.) Expressed Sequence Tags (ESTs), *Plant Signal. Behav.* 8 (12) (2013).
- [17] A. Gentile, T.H. Ferreira, R.S. Mattos, L.I. Dias, A.A. Hoshino, M.S. Carneiro, G.M. Souza, T. Calsa Jr., R.M. Nogueira, L. Endres, et al., Effects of drought on the microtranscriptome of field-grown sugarcane plants, *Planta* 237 (3) (2013) 783–798.
- [18] M.M. Menossi, A. Gentile, T.H. Ferreira, R. Mattos, L.I. Dias, MicroRNAs and drought responses in sugarcane, *Front. Plant Sci.* 6 (2015) 58.
- [19] E. Lord, M. Leclercq, A. Boc, A.B. Diallo, V. Makarenkov, Armadillo 1.1: an original workflow platform for designing and conducting phylogenetic analysis and simulations, *PLoS ONE* 7 (1) (2012) e29903.