*D198–D204* *Nucleic Acids Research, 2012, Vol. 40, Database issue* *Published online 1 December 2011*
*doi:10.1093/nar/gkr1159*

# miRNEST database: an integrative approach in microRNA search and annotation

**Michał Wojciech Szcześniak[1,*], Sebastian Deorowicz[2], Jakub Gapski[1], Łukasz Kaczyński[1] and Izabela Makałowska[1,*]**

[1]Laboratory of Bioinformatics, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan and [2]Institute of Informatics, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland

Received August 12, 2011; Revised and Accepted November 10, 2011

## ABSTRACT

**Despite accumulating data on animal and plant microRNAs and their functions, existing public miRNA resources usually collect miRNAs from a very limited number of species. A lot of microRNAs, including those from model organisms, remain undiscovered. As a result there is a continuous need to search for new microRNAs. We present miRNEST (http://mirnest.amu.edu.pl), a comprehensive database of animal, plant and virus microRNAs. The core part of the database is built from our miRNA predictions conducted on Expressed Sequence Tags of 225 animal and 202 plant species. The miRNA search was performed based on sequence similarity and as many as 10 004 miRNA candidates in 221 animal and 199 plant species were discovered. Out of them only 299 have already been deposited in miRBase. Additionally, miRNEST has been integrated with external miRNA data from literature and 13 databases, which includes miRNA sequences, small RNA sequencing data, expression, polymorphisms and targets data as well as links to external miRNA resources, whenever applicable. All this makes miRNEST a considerable miRNA resource in a sense of number of species (544) that integrates a scattered miRNA data into a uniform format with a user-friendly web interface.**

## INTRODUCTION

Animal and plant miRNA genes are transcribed by RNA polymerase II or III into a primary transcript, called pri-miRNA (1). During initial steps of miRNA biogenesis, pri-miRNA is cut and a hairpin-shaped intermediate, called pre-miRNA, is produced. This process is catalyzed by Drosha in animals (2) and DCL1 (DICER-LIKE 1) in plants (3). Subsequently, a pre-miRNA is specifically cut at stem part of the hairpin and a miRNA/miRNA* duplex with 2-nt overhangs at 3′-ends is released. In animals this process is run by Dicer (2) and in plants it is controlled by DCL1 (3). In cytoplasm, one of duplex components, referred to as mature miRNA, gets incorporated into a riboprotein complex, named RISC (RNA-induced silencing complex) (4). RISC contains a functional unit, which allows regulation of the gene expression based on complementarity of the miRNA and the transcript of targeted gene. There are two modes of the regulation: by cleaving transcripts (5) and by inhibiting translation (6). The first one requires high complementarity between miRNA and targeted transcript and is ubiquitously observed in plants, while in animals it is translation inhibition that constitutes a major mechanism of miRNA action.

miRNAs regulate the expression of thousands of genes in plants and animals and are key players in developmental (7), stress-related (8) and signalling processes (9). A number of miRNAs have been associated with diseases in human, e.g. Alzheimer's disease (10), pancreatic cancer (11), or leukemia (12). Hence, identification of miRNAs and subsequent elucidation of their functions, both in plants and animals, became a critical issue not only in molecular biology but also in medical research and agriculture.

Recently, a number of investigations aimed at the identification of miRNAs have been published. Reported miRNAs were discovered either based on computational (13) or experimental approaches (14). Consequently, the growing number of miRNA studies led to accumulation of miRNA databases. However, many of these databases, like miRO (15) or miROrtho (16), are limited to species of high interest. Other resources are focused on selected taxa, e.g. microPC (13) and PMRD (17) contain only

*To whom correspondence should be addressed. Tel: +48 61 829 5836; Email: miszcz@amu.edu.pl
Correspondence may also be addressed to Izabela Makałowska. Tel: +48 61 829 5835; Email: izabel@amu.edu.pl

© The Author(s) 2011. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on February 15, 2012

plant miRNAs, CoGemiR (18) consists of miRNAs from selected animal species and Vir-Mir db (19) is dedicated to viruses only. The main miRNA repository database, miRBase (20), although accommodates data from a wide range of species, contains only already published results. Similarly, databases like miRecords (21) or miRNAMap (22) consist exclusively of experimentally verified miRNAs. Contrary, microPC dataset is based solely on computational methods. Therefore, despite of the ever-growing number of miRNA-related resources, there is a lack of a single universal repository and users need to browse through a number of dispersed data sets to collect information related to specific species or miRNA type.

To overcome these limitations, we developed miRNEST as a comprehensive online resource for plant, animal and virus miRNAs. We applied a comparative approach to search for new microRNAs in animal and plant EST sequences from dbEST (23). The approach applied by us made it possible to identify 10 004 miRNA candidates in 221 animal and 199 plant species. To the best of our knowledge, for 236 species no miRNAs have been known before. Besides miRNA identification, we also performed miRNA target search in plant candidates and for 29 species we collected small RNA reads from Gene Expression Omnibus (24). Since our goal was not only to identify new miRNAs but also to develop a resource that would integrate miRNA data scattered across literature and databases, we complemented data resulting from our computational analysis with miRNA sequences from three other databases and two publications. In addition, based on availability, we incorporated the data from twelve resources providing further annotations of miRNAs from selected species. This gives the possibility to access, search and browse data from different resources simultaneously. Altogether the miRNEST database contains 39 122 miRNAs from 544 species. All data are presented in the same format via miRNEST interface and are available at http://mirnest.amu.edu.pl.

## Data acquisition

We imported 16 961 known mature miRNA sequences from miRBase (20) and 9212 sequences from PMRD (17), which were used for comparative analysis and identification of conserved miRNAs. EST sequences of 225 animal and 202 plant species were downloaded from dbEST (23). We took into consideration only the species that had at least 10 000 EST sequences. For EST annotation, the UniProtKB/Swiss-Prot protein data set was obtained from UniProt (25), and to remove tRNAs and rRNAs from ESTs and for further miRNA annotation, we obtained ncRNA sequences from RFAM release 9.1 (26). Also, 192 small RNA deep sequencing libraries were downloaded from GEO, Gene Expression Omnibus (24).

In addition to mature miRNA and pre-miRNA data from miRBase and PMRD, we downloaded sequences from microPC (18) and two publications (27,28). The choice of databases and papers was made based on the overlap between resources and the availability of the data. For example, the sets of miRNAs deposited in miRNAMap (22) and RNAdb (29) fully overlapped with miRBase and therefore were not considered for the download. On the other hand, data from resources like miROrtho (16) or GrapeMiRNA (30) could not be downloaded and consequently, were not integrated with miRNEST. Targets data were collected from the largest and most recognized resources as miRDB (31), miRTarBase (32), miRecords (21), PMRD and ASRP (33). We also obtained miRNA expression data from PhenomiR 2.0 (34) and dbDEMC (35), miRNA regulation data and promoters from dPORE-miRNA (36) and PMRD, genomics information from CoGemiR v1.2b (18), miRNA polymorphisms from Patrocles (37) and imprinting data from ncRNAimprint (38). In particular, we took advantage of data available in miRBase, which among others provided us with literature references, links to external databases, genomics data, and served as a source of uniform, non-redundant miRNA nomenclature, which was of great importance when integrating the miRNEST external miRNA data. The list of all databases used for the data assembly is provided in Supplementary Table S1 and the range of miRNA overlap between resources in Supplementary Table S2.

## Prediction of miRNAs

The prediction of animal and plant microRNA candidates was performed according to modified algorithm applied by Zhang *et al.* (39) and Mhuantong *et al.* (13) (Figure 1). The modifications include removal of sequences with high percentage of low-complexity regions, assembling ESTs into contigs using CAP3 (40) and position-based allowance for substitutions in mature miRNAs. Schema of all data processing steps is presented in Figure 1.

In the first step, mature miRNA sequences from miRBase and PMRD were searched against EST sequences using BLASTN (41). We filtered the BLASTN search results, based on pre-computed distribution of mismatches in mature miRNAs within miRNA homologs in animals and plants (Supplementary Figure S1). Then, EST sequences, clustered based on the similarity to the same miRNA, were assembled into contigs using CAP3. This step is important in plant microRNA prediction where pre-miRNA length occasionally exceeds 600 nucleotides (Supplementary Figure S2). Next, we ran BLASTN search against RFAM non-coding RNA sequences to remove rRNAs and tRNAs (with *E*-value $<$ 1e-20). After that we searched for low-complexity regions using Dustmasker (42) and removed sequences containing more than 60% of low complexity regions. Then, RNA secondary structures were predicted using RNAfold (43). We filtered the secondary structures and left only the candidates with mature miRNA located in a stem part of the hairpin structure and with no more than five mismatches and two bulges between mature miRNA and the opposite hairpin arm (Supplementary Figure S3). At the same time, we extracted the hairpin sequences out of longer contig or singleton. After that, we performed a BLASTX search against UniProt to remove sequences with high similarity to proteins (*E*-value $<$ 1e-20). Finally, we removed animal pre-miRNAs with length exceeding 215 nt, which is the

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on February 15, 2012

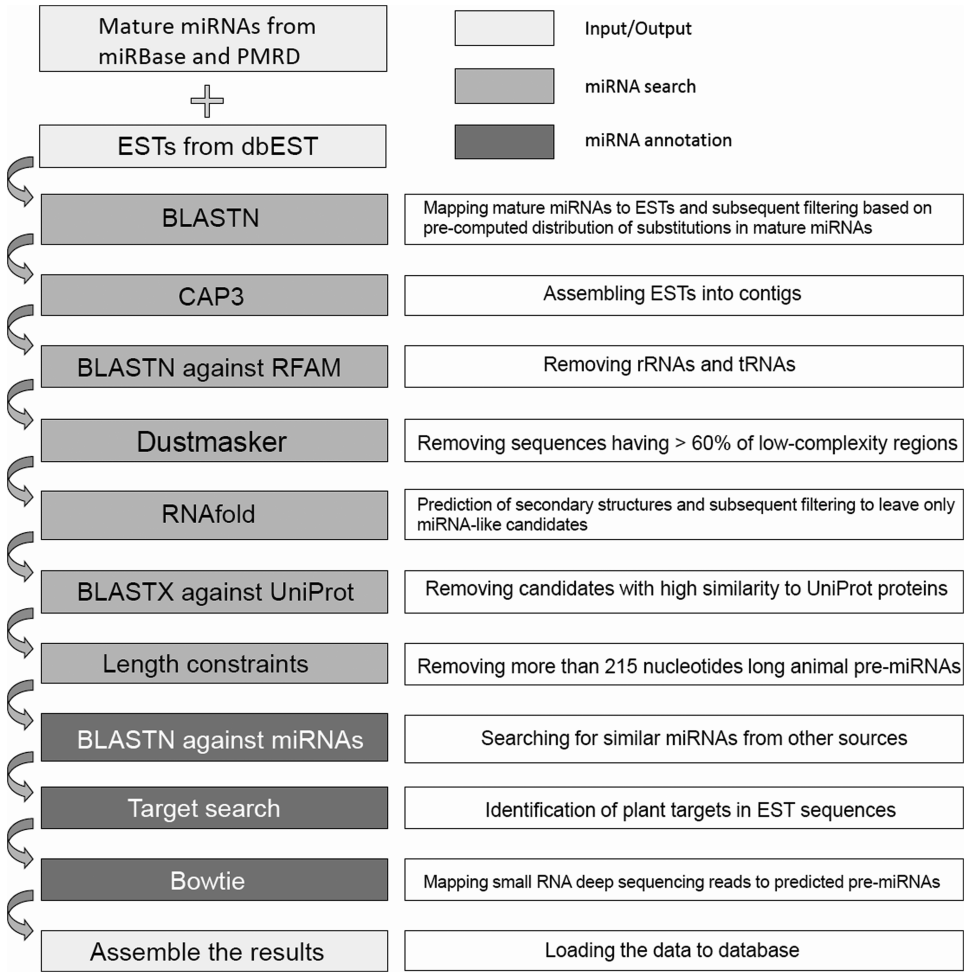


**Figure 1.** A computational pipeline applied for prediction of microRNAs in EST sequences and their annotation. The seven steps in miRNA search part are designed to minimize the false positives rate and provide a high quality set of candidates. In some of them, plant- and animal-specific parameters were applied, as described in the main text. The annotation part serves to provide more data on predicted miRNAs and no candidates are discarded there.

maximum length for animal pre-miRNA in miRBase. As a result, we obtained 10 004 miRNA candidates in 221 animal and 199 plant species.

The ratio of sequences removed from the data set at various filtering steps differed between species; examples for *Arabidopsis thaliana* and *Bos taurus* are presented in Supplementary Table S3.

### miRNA prediction accuracy

*Specificity*. To assess the specificity of miRNEST algorithm we generated one million random sequences that were subsequently subject to the algorithm. The initial number of BLAST hits, i.e. mature miRNAs against generated sequences, was 231 260. After filtering steps, Dustmasker and BLASTN against RFAM, the number of candidates has decreased substantially to 1120 and only three candidates left after secondary structure check point. The last step, BLAST against UniProt, reduced the number of candidates to two, which produces the false positives ratio of 0.0002% (two per million). This result is comparable with the one calculated in a very similar way by authors of microPC database (13). Applying a similar approach they obtained the ratio of 0.00064%. We have also checked how many of our human miRNA predictions are classified as true miRNAs by microPred (44), a tool for miRNA/non-miRNA classification. This tool was specifically designed for human pre-miRNAs and it classified as miRNAs 75.9% of miRNEST predicted human miRNAs. At the same time, 90.16% of miRBase human pre-miRNAs are considered as true miRNAs by this program.

*Sensitivity*. To estimate the sensitivity of applied algorithm known pre-miRNAs of *B. taurus* and *A. thaliana*, from miRBase database, were subjected to miRNA search. Out of 229 Arabidopsis hairpins, 209 (91.3%) have been recovered. In case of *B. taurus*, the result was lower and only 61.1% (392 out of 662) of hairpins have been recognized as miRNA by miRNEST algorithm. This is mostly because of the architecture of animal hairpins: most of the miRNAs that have not been recovered bear more than five mismatches and/or two bulges between mature miRNA and the opposite hairpin arm.

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on February 15, 2012

Nonetheless, we decided to keep this filtering threshold in order to maintain the low false positives ratio. Authors of microPC database, where miRNAs are also predicted based on EST sequences analysis, estimated the sensitivity of their method for all analyzed plant species as 81.3%. We did calculation separately for individual species but on average, the rate of specificity is similar to this obtained in microPC.

### Further annotation of predicted miRNAs

Plant mature miRNAs generally show a high level of complementarity with targeted mRNA sequences, thus target identification usually is not a very complex problem here. On the other hand, effective methods for target search in animals heavily rely on the evolutionary conservation of miRNA target site (45). Since such data are unavailable for a great majority of analyzed by us animal sequences, we decided to focus on target search for plant candidates and we used external target data for animal species whenever possible.

The targets were searched for in EST sequences from corresponding species applying standard rules of plant miRNA–mRNA interactions (46). In the final scoring a mismatch was given a score of 1, a wobble (G:U) was given a score of 0.5 and a bulge was given a score of 2.0. All matches with score above 3.5 were discarded. It was also important that positions 10 and 11 of miRNA perfectly matched to its target and that there was no more than 1 mismatch at positions 2 to 9. As a result, we identified targets for 6963 mature miRNAs in 187 plant species. Sequences of all potentially targeted ESTs were checked against UniProt data for functional annotations.

For 29 species, we downloaded small RNA deep sequencing libraries from GEO. The reads were mapped to predicted pre-miRNAs using Bowtie (47). The mapping was performed against both pre-miRNA strands and only one mismatch was allowed. The choice of species for which deep sequencing data were downloaded was solely based on the availability.

### Processing of external data

The goal of our project was not only to predict miRNAs in EST sequences but also to build a comprehensive miRNA database. Therefore, we supplemented data set from our computational analyses with the sequence data from three external resources: miRBase, PMRD, microPC and two publications (27,28) reporting miRNAs not deposited in miRBase. All imported sequences were run through our pipeline so that all data are deposited and displayed in identical format. However, filtering steps were turned off to ensure that all external miRNAs would be incorporated into the miRNEST database even if they did not match criteria applied in miRNEST algorithm. Finally, we run reciprocal BLASTN search: each miRNEST pre-miRNA against each other to identify similar miRNAs across data sets.

### Web interface

The miRNEST web interface has been divided into five sections that correspond to distinct types of data.

*Browse*. By clicking Browse in main menu, the user gets access to all miRNA sequences: miRNEST predictions and miRNAs from external sources. For each miRNA record we provide miRNA* sequence, coordinates of miRNA and miRNA* in pre-miRNA, guide strand, number of mismatches and bulges between mature miRNA and the opposite hairpin arm, secondary structures of pre-miRNAs, family assignment, experimental evidence and identical miRNEST, miRBase, PMRD or microPC miRNAs, whenever found (Figure 2). There are also links to miRNEST internal resources providing access to results of BLAST search, target predictions in EST sequences as well as target information from external sources, most similar mature miRNAs, source sequences for miRNAs predicted from ESTs, graphical display of deep sequencing reads mapping as well as data collected from external resources: miRNA genomics data, SNPs, promoters, TFBSs, miRNA-disease association, miRNA polymorphism, expression data, literature references, links to external resources and information on imprinted miRNAs. From here users have also opportunity to run BLASTN searches against selected sources of miRNA sequences and to build ClustalW (48) alignments based on BLASTN results. If species is selected, its full NCBI taxonomy is provided (49). By clicking a taxon, all miRNEST species belonging to the taxon are displayed as active links to miRNA data in corresponding species. Users can also browse through species in a taxonomic tree view.

*Search*. Using a search option users have the possibility to filter miRNA data by a number of parametres: species, sequence source (miRNEST prediction and/or external sources), mature miRNA sequence or its part, hairpin length, free folding energy, number of allowed mismatches and bulges between mature miRNA and the opposite arm in pre-miRNA hairpin structure, and *E*-values for BLAST search that was pre-run against UniProt, RFAM, miRBase, PMRD, and microPC. We also incorporated the target search option, where user can search for specific targets by typing mature miRNA sequence and selecting a species of interest. Moreover, users can limit their search to the records that have additional data, downloaded from outside resources, like experimental evidence or target sequences.

*Unclassified*. Unclassified section provides users with a list of miRNEST predictions that were not classified as potential miRNAs as they violated at least one of the following criteria: *E*-value for BLASTX search against UniProt >1e-20 or pre-miRNA length for animal candidate ≤215 nucleotides. There are 66 predictions that fall into this section due to the first criterion, 465—the latter and 8 predictions that violence both criteria. *Unclassified* and *Search* possess a very similar interface, however some options are unavailable in *Unclassified* category.

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on February 15, 2012

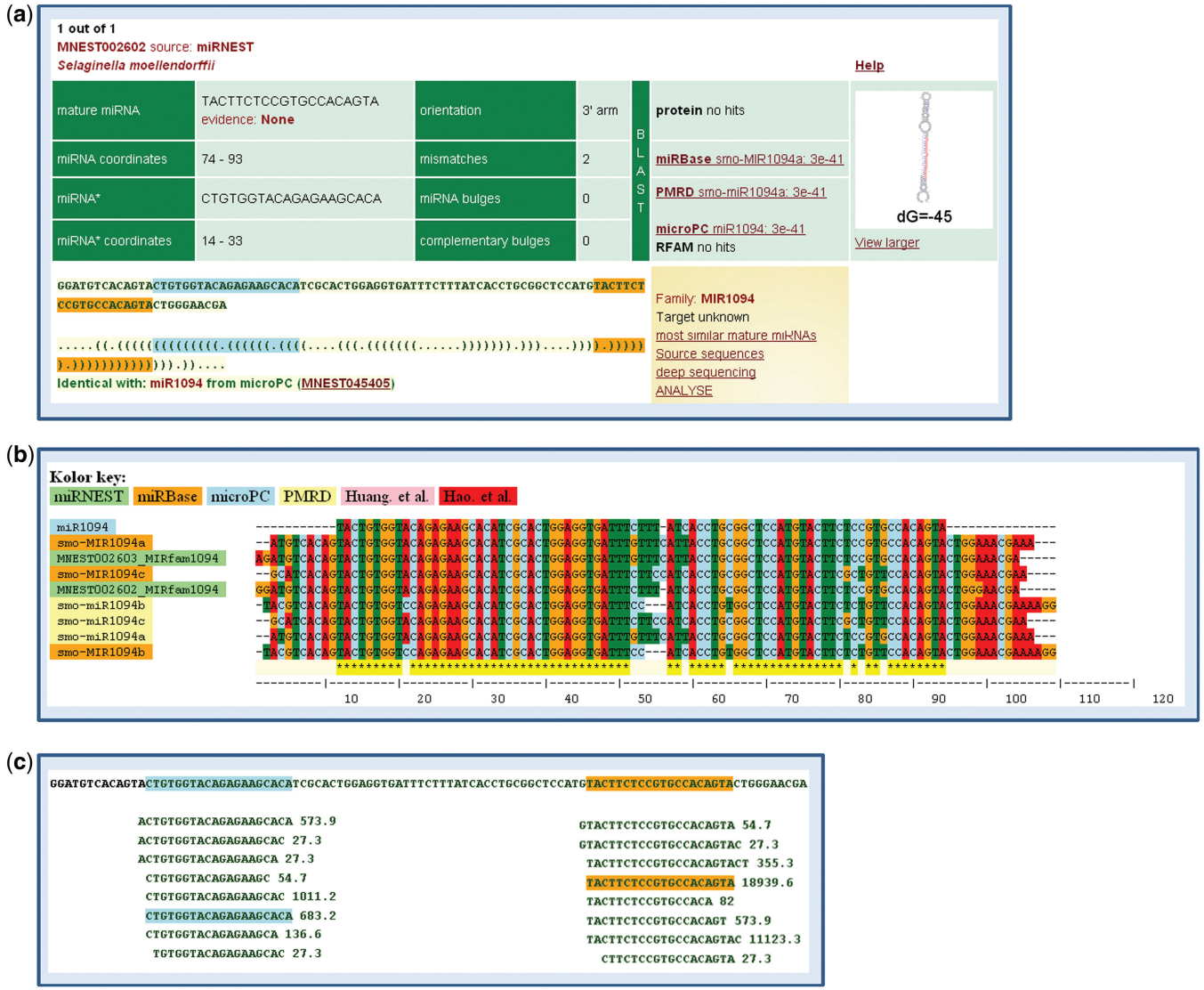


**Figure 2.** (**a**) An example of miRNEST record (MNEST002602) (**b**) by clicking *analyse* link it is possible to perform BLASTN search and ClustalW analysis. (**c**) *deep sequencing* link leads to mapping pattern of small RNA deep sequencing reads.

*RNA-Seq*. This component of the miRNEST database presents small RNA sequencing data that correspond to miRNEST predictions from 29 species (Supplementary Table S4). The user can select species, miRNA of interest, small RNA deep sequencing library, strand and number of allowed mismatches between reads and pre-miRNA. The results are displayed in a graphical form as an alignment of reads to the pre-miRNA. If no library is specified, then unique reads from all libraries from a given species are displayed. Otherwise, the read counts, as transcripts per million, are supplied. *RNA-Seq* data are also accessible from Browse section through deep sequencing internal link that corresponds to a given miRNA record.

*Taxonomy*. Taxonomy option provides a user with a phylogenetic tree of species that are included in miRNEST predictions. It presents clickable taxon names (kingdom, phylum, subclass, order, family, genus, and species) along with a number of taxon-specific miRNA families. By clicking on the taxon, the user gains access to more detailed information on the taxon-specific miRNA families and links to corresponding miRNEST records.

## Comparison with other miRNA databases

miRNEST is a large collection of miRNA sequences and associated miRNA data (Supplementary Table 1). It encompasses 522 animal and plant species as well as 22 viruses (Supplementary Table S5) and as many as 9429 miRNA predictions can be found uniquely in miRNEST, though a fraction of the predictions overlap with known miRNA data (Supplementary Table S5). For comparison, miRBase (Release 17), the second largest miRNA database collects data from only 153 distinct

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on February 15, 2012

animals, plants and viruses. As many as 236 miRNEST species were not taken into account in any other miRNA database. These include a number of model species as well as the ones of high interest in agriculture and medical research like *Salmo salar*, *Takifugu rubripes*, *Actinidia chinensis*, *Anolis carolinensis*, *Bos indicus* or *Trichinella spiralis*.

Another distinguishing feature of miRNEST is the extent of usage of small RNA deep sequencing reads. We incorporated data from 192 libraries from 29 species retrieved from GEO and mapped the reads to predicted pre-miRNAs. We carefully selected the libraries in order to make sure that they encompass a wide array of tissues, developmental stages and conditions. For comparison, deepBase (50), which is a platform for annotating and discovering small and long ncRNAs from next-generation sequencing, collects data from 185 small RNA libraries from seven species and in miRBase, Release 17, deep sequencing data has been added for seven species, yet this functionality is now being expanded to more species.

miRNEST also gives access to large-scale target search predictions for 187 plant species that were generated using standard rules for plant miRNA target prediction. For more than half of the species this is the first and only miRNA targets prediction.

Finally, miRNEST predictions are complemented with a wide range of external data retrieved from 13 databases. Our goal was to create a resource that would integrate the miRNA data that are currently scattered across multiple resources and to limit existing necessity of searching multiple databases to investigate a single miRNA or miRNAs from a given species.

## AVAILABILITY AND REQUIREMENTS

miRNEST is freely available at http://mirnest.amu.edu.pl. The database was constructed using Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), PHP 5.2.11 (http://www.php.net/), and MySQL 4.0.31 (http://www.mysql.com/). pre-miRNA secondary structures are drawn using Java lightweight applet VARNA (51) which requires the installation of Java plugin.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–3, Supplementary Tables 1–5.

## ACKNOWLEDGEMENTS

The authors thank Professor Wojciech Makałowski for proofreading the article and Joanna Ciomborowska, Michał Kabza and Elżbieta Owczarkowska for helpful suggestions on the database web interface.

## FUNDING

Faculty of Biology at Adam Mickiewicz University in Poznan, Poland (PBWB-08/2011); the European Social Fund within the project—'Fellowships for PhD Students in Wielkopolska Region conducting research in strategic areas supporting regional development'; the Ministry of Science and Higher Education (N N301 160935 to I.M. and N N516 441938 to S. D.). Funding for open access charge: Faculty of Biology at Adam Mickiewicz University in Poznan.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Cai,X., Hagedorn,C.H. and Cullen,B.R. (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, **10**, 1957–1966.
2. Lee,Y., Ahn,C., Han,J., Choi,H., Kim,J., Yim,J., Lee,J., Provost,P., Rådmark,O., Kim,S. *et al.* (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415–419.
3. Kurihara,Y., Takashi,Y. and Watanabe,Y. (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA*, **12**, 206–212.
4. Lau,N.C., Lim,L.P., Weinstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
5. Reinhart,B.J., Weinstein,E.G., Rhoades,M.W., Bartel,B. and Bartel,D.P. (2002) MicroRNAs in plants. *Genes Dev.*, **16**, 1616–1626.
6. Lai,E.C. (2002) Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.*, **30**, 363–364.
7. Kedde,M. and Agami,R. (2008) Interplay between microRNAs and RNA-binding proteins determines developmental processes. *Cell Cycle*, **7**, 899–903.
8. Leung,A.K. and Sharp,P.A. (2010) MicroRNA functions in stress responses. *Mol. Cell*, **40**, 205–215.
9. O'Neill,L.A., Sheedy,F.J. and McCoy,C.E. (2011) MicroRNAs: the fine-tuners of Toll-like receptor signalling. *Nat. Rev. Immunol.*, **11**, 163–175.
10. Yao,J., Hennessey,T., Flynt,A., Lai,E., Beal,M.F. and Lin,M.T. (2010) MicroRNA-related cofilin abnormality in Alzheimer's disease. *PLoS One*, **5**, e15546.
11. Brabletz,S., Bajdak,K., Meidhof,S., Burk,U., Niedermann,G., Firat,E., Wellner,U., Dimmler,A., Faller,G., Schubert,J. *et al.* (2011) The ZEB1/miR-200 feedback loop controls Notch signalling in cancer cells. *EMBO J.*, **30**, 770–782.
12. Schotte,D., Pieters,R. and Den Boer,M.L. (2011) MicroRNAs in acute leukemia: from biological players to clinical contributors. *Leukemia*, June 24 (doi:10.1038/leu.2011.151; epub ahead of print).
13. Mhuantong,W. and Wichadakul,D. (2009) MicroPC (microPC): a comprehensive resource for predicting and comparing plant microRNAs. *BMC Genomics*, **10**, 366.
14. Zhang,L., Chia,J.M., Kumari,S., Stein,J.C., Liu,Z., Narechania,A., Maher,C.A., Guill,K., McMullen,M.D. and Ware,D. (2009) A genome-wide characterization of microRNA genes in maize. *PLoS Genet.*, **5**, e1000716.
15. Laganà,A., Forte,S., Giudice,A., Arena,M.R., Puglisi,P.L., Giugno,R., Pulvirenti,A., Shasha,D. and Ferro,A. (2009) miRò: a miRNA knowledge base. *Database*, **2009**, bap008.
16. Gerlach,D., Kriventseva,E.V., Rahman,N., Vejnar,C.E. and Zdobnov,E.M. (2009) miROrtho: computational survey of microRNA genes. *Nucleic Acids Res.*, **37**, D111–7.
17. Zhang,Z., Yu,J., Li,D., Zhang,Z., Liu,F., Zhou,X., Wang,T., Ling,Y. and Su,Z. (2010) PMRD: plant microRNA database. *Nucleic Acids Res.*, **38**, D806–813.
18. Maselli,V., Di Bernardo,D. and Banfi,S. (2008) CoGemiR: a comparative genomics microRNA database. *BMC Genomics*, **9**, 457.
19. Li,S.C., Shiau,C.K. and Lin,W.C. (2008) Vir-Mir db: prediction of viral microRNA candidate hairpins. *Nucleic Acids Res.*, **36**, D184–D189.

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on February 15, 2012

20. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
21. Xiao,F., Zuo,Z., Cai,G., Kang,S., Gao,X. and Li,T. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.*, **37**, D105–D110.
22. Hsu,P.W., Huang,H.D., Hsu,S.D., Lin,L.Z., Tsou,A.P., Tseng,C.P., Stadler,P.F., Washietl,S. and Hofacker,I.L. (2006) miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic Acids Res.*, **34**, D135–D139.
23. Boguski,M.S., Lowe,T.M. and Tolstoshev,C.M. (1993) dbEST—database for "expressed sequence tags". *Nat Genet.*, **4**, 332–333.
24. Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M. *et al.* (2011) NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res.*, **39**, D1005–D1010.
25. The UniProt Consortium. (2011) Ongoing and future developments at the universal protein resource. *Nucleic Acids Res.*, **39**, D214–D219.
26. Griffiths-Jones,S., Moxon,S., Marshall,M., Khanna,A., Eddy,S.R. and Bateman,A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.*, **33**, D121–D124.
27. Huang,J., Hao,P., Chen,H., Hu,W., Yan,Q., Liu,F. and Han,Z.G. (2009) Genome-wide identification of *Schistosoma japonicum* microRNAs using a deep-sequencing approach. *PLoS One*, **4**, e8206.
28. Hao,L., Cai,P., Jiang,N., Wang,H. and Chen,Q. (2010) Identification and characterization of microRNAs and endogenous siRNAs in *Schistosoma japonicum*. *BMC Genomics*, **11**, 55.
29. Pang,K.C., Stephen,S., Dinger,M.E., Engström,P.G., Lenhard,B. and Mattick,J.S. (2007) RNAdb 2.0—an expanded database of mammalian non-coding RNAs. *Nucleic Acids Res.*, **35**, D178–D182.
30. Lazzari,B., Caprera,A., Cestaro,A., Merelli,I., Del Corvo,M., Fontana,P., Milanesi,L., Velasco,R. and Stella,A. (2009) Ontology-oriented retrieval of putative microRNAs in *Vitis vinifera* via GrapeMiRNA: a web database of de novo predicted grape microRNAs. *BMC Plant Biol.*, **9**, 82.
31. Wang,X. (2008) miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA*, **14**, 1012–1017.
32. Hsu,S.D., Lin,F.M., Wu,W.Y., Liang,C., Huang,W.C., Chan,W.L., Tsai,W.T., Chen,G.Z., Lee,C.J., Chiu,C.M. *et al.* (2010) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **39**, D163–D169.
33. Gustafson,A.M., Allen,E., Givan,S., Smith,D., Carrington,J.C. and Kasschau,K.D. (2005) ASRP: the Arabidopsis small RNA project database. *Nucleic Acids Res.*, **33**, D637–D640.
34. Ruepp,A., Kowarsch,A., Schmidl,D., Buggenthin,F., Brauner,B., Dunger,I., Fobo,G., Frishman,G., Montrone,C. and Theis,F.J. (2010) PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.*, **11**, R6.
35. Yang,Z., Ren,F., Liu,C., He,S., Sun,G., Gao,Q., Yao,L., Zhang,Y., Miao,R., Cao,Y. *et al.* (2010) dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics*, **11(Suppl. 4)**, S5.
36. Schmeier,S., Schaefer,U., MacPherson,C.R. and Bajic,V.B. (2011) dPORE-miRNA: polymorphic regulation of microRNA genes. *PLoS One*, **6**, e16657.
37. Hiard,S., Charlier,C., Coppieters,W., Georges,M. and Baurain,D. (2010) Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Res.*, **38**, D640–D651.
38. Zhang,Y., Guan,D.G., Yang,J.H., Shao,P., Zhou,H. and Qu,L.H. (2010) ncRNAimprint: a comprehensive database of mammalian imprinted non-coding RNAs. *RNA*, **16**, 1889–1901.
39. Zhang,B.H., Pan,X.P., Wang,Q.L., Cobb,G.P. and Anderson,T.A. (2005) Identification and characterization of new plant microRNAs using EST analysis. *Cell Res.*, **15**, 336–360.
40. Huang,X. and Madan,A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.*, **9**, 868–877.
41. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
42. Morgulis,A., Gertz,E.M., Schäffer,A.A. and Agarwala,R. (2006) A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.*, **13**, 1028–1240.
43. Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhoeffer,S., Tacker,M. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structures. *Monatshefte f. Chem.*, **125**, 167–188.
44. Batuwita,R. and Palade,V. (2009) microPred: effective classification of pre-miRNAs for human miRNA gene prediction. *Bioinformatics*, **25**, 989–995.
45. Alexiou,P., Maragkakis,M., Papadopoulos,G.L., Reczko,M. and Hatzigeorgiou,A.G. (2009) Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics*, **25**, 3049–3055.
46. Schwab,R., Palatnik,J.F., Riester,M., Schommer,C., Schmid,M. and Weigel,D. (2005) Specific effects of microRNAs on the plant transcriptome. *Dev. Cell*, **8**, 517–527.
47. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
48. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X Version 2.0. *Bioinformatics*, **23**, 2947–2948.
49. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2009) GenBank. *Nucleic Acids Res.*, **37**, D26–D31.
50. Yang,J.H., Shao,P., Zhou,H., Chen,Y.Q. and Qu,L.H. (2010) deepBase: a database for deeply annotating and mining deep sequencing data. *Nucleic Acids Res.*, **38**, D123–D130.
51. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.


Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on February 15, 2012