MRST: A TOOL FOR THE STORAGE OF MICRORNA COMPUTATIONAL RESOURCES

MING CHEN[2,3], ZHONGYANG TAN[1,*], GUANGMING ZENG[2,3,*], MINGFU LI[4]

[1]College of Biology, State Key Laboratory for Chemo/Biosensing and Chemometrics, Hunan University, Changsha 410082, China
[2]College of Environmental Science and Engineering, Hunan University, Changsha 410082, China
[3]Key Laboratory of Environmental Biology and Pollution Control (Hunan University), Ministry of Education, Changsha 410082, China
[4]Chinese Academy of Inspection and Quarantine Beijing, 100029, China
*Corresponding Author: Email- zhongyang@hnu.cn (Z. Tan) and zgming@hnu.cn (G. Zeng)

Abstract- More and more computational resources for microRNAs (miRNAs) have been developed in recent years. That may lead to selective difficulty, mistakes of usage and confusion due to their confined scope of usage. Here, we summarize and classify the knowledge that has been accumulated in the fields and present a novel tool, MRST (miRNA resource store tool), for the storage of miRNA resources.
Key words - microRNA; target gene; computational algorithm.

Introduction

MicroRNAs (miRNAs) are short non-coding RNAs with the length of 22 nt approximately, and sever a functional role that either results in the degradation of mRNAs or suppresses their translation by binding to the 3′-untranslated region (3′-UTR) [1,2]. To date, there are large number of miRNA sequences have been found [3,4]. First, a long primary miRNAs (pri-miRNAs) are processed into 60–70 nt hairpin-like precursor transcripts (pre-miRNAs) by Drosha, and then are cleaved into mature miRNAs by another RNase III, Dicer [5]. It has been demonstrated that miRNAs play important roles in various biological processes, such as cell proliferation, differentiation, development, diseases, transcriptional gene regulatory network, neuronal synapses formation, and cell death [5,6]. Some computational tools for the identification of miRNA gene have been developed, and so far, the two most sensitive tools are MiRscan and miRseeker [7-9]. It is believed that functional understanding of miRNAs will depend heavily on identification of their corresponding target mRNAs [10]. MiRNA target predictions in animals are thought to be more difficult than in plants because miRNAs of animals are short and miRNA-mRNA duplexes are not entirely complementary to one another [11]. Computational algorithms have been developed based on extracted principles such as sequence complementarity, thermodynamic stability calculations and evolutionary conservation among species [5,10]. Moreover, it is also suggested stronger binding at the 3′ region compensate for imperfect base pairing within the seed segment, and a method has considered the contribution of 3′ region in target identification [12].

Due to accumulating knowledge of miRNAs and their target genes, some useful databases have been constructed, of which miRBase serves as a registry of the information of miRNA sequences [3].
Large number of computational tools and databases are developed, leading to selective difficulty for the usage due to the lack of classification and conclusion. Therefore, users must determine appropriate methods or tools according to what they wish to gain from a given analysis. These situations suggest that the performance of currently available algorithms or databases may need to be assembled and sorted. All these encourage us to seek solutions to improve the use effect of miRNA resources.
We present here, a tool for the collection with existed computational tools and databases of miRNAs. In our approach, data consistency and data redundancy are considered. We believe that our work will contribute significantly to facilitate the usage of these tools and databases. Most important, it facilitates the research on the roles of miRNAs.

Materials and methods

There are numerously useful resources and available software tools for analysis of miRNAs, as summarized in Table 1. We conclude the efforts made by different software tools to predict miRNA genes or target genes, their principles of miRNA target recognition, and their supported organisms for computational prediction. This comprehensive collection with tools and databases of miRNAs is expected to provide an overall view of the knowledge that has been

accumulated in the field and be helpful to research of miRNAs.

Data collection with databases of miRNAs and their targets

All data and information were obtained from databases and literature. After finishing data collection, the second step involves mining information from various datasets. Because different databases and computational tools own various limits and scopes of usage. To resolve this problem, a variety of databases and related literature are assessed, and the main features of these resources are investigated. Table 1 presents the data collected which are preprocessed according to the mainly advanced features. The websites of these tools are given in Table 2 and this software. Previous works introduced and reviewed some of these databases, but the knowledge of most databases is not mentioned and sorted out [10,11]. The present work will facilitate the comparisons between different databases.

Data collection with computational algorithms for identification and prediction of miRNA genes and miRNA targets

Supported organisms of various software tools may be non-identical. These tools are investigated, and are classified as 'any', 'animal', 'vertebrates', 'flies', 'nematodes', 'viruses', and 'plants' according to their supported organisms [10,11]. Main features, websites and references of these tools are shown in Tables 1 and 2.

Construction of software

MRST is composed of three modules: databases of miRNAs and their targets, computational algorithms for miRNA gene identification, and computational algorithms for miRNA target prediction. The proposed procedure interface provides a variety of available websites of reported databases and computational algorithms for both browsing functions and search functions. The detailed introduction is provided in this tool.

Implementation

MRST has been developed in standard Microsoft visual basic language. The program collects most of available databases and computational tools of miRNAs. The program provides four options for users: introduction of computational resources; available online resources for databases of miRNAs; available online resources for computational algorithms of miRNA gene identification; and available online resources for computational algorithms of miRNA target prediction. A sand-alone application with a user's guide is available for free access at http://hudacm11.mysinamail.com/bioinformatics.html. Moreover, this software is available from authors Zhongyang Tan and Guangming Zeng on request (zhongyang@hnu.cn; zgming@hnu.cn)**.**

Results and discussion

MiRNAs have been proved to be involved in a variety of biological processes. It is important to collect widely available algorithms for research of miRNAs. This work will help a lot to broaden our understanding with computational resources of miRNAs, and presents a novel tool for the storage of miRNA resources including databases of miRNAs and their targets, computational algorithms for miRNA gene identification, computational algorithms for miRNA target prediction. We believe that the proposed tool can provide sufficient and effective information for the investigation about availably computational resources of miRNAs. Alternatively, the proposed tool allows users to search databases of miRNAs or find the most suitable software tools for their analysis that is more convenient.

Acknowledgement

References

[1] Megraw M., Sethupathy P., Corda B. and Hatzigeorgiou A.G. (2007) *Nucleic Acids Res* 35, D149-D155.

[2] John B., Enright A.J., Aravin A., Tuschl T., Sander C. and Marks D.S. (2004) *PLoS Biol* 2, e363.

[3] Griffiths-Jones S., Grocock R.J., van Dongen S., Bateman A. and Enright A.J. (2006). *Nucleic Acids Res* 34, D140-D144.

[4] Chen M., Tan Z., Zeng G. and Peng J. (2010) *Mol Biol Evol* 27, 2227-2232.

[5] Bartel D.P. (2004) *Cell,* 116, 281-297.

[6] Hsu P.W. et al. (2006) *Nucleic Acids Res* 34, D135-D139.

[7] Lim L.P., Lau N.C., Weinstein E.G., Abdelhakim A., Yekta S., Rhoades M.W., Burge C.B. and Bartel D.P. (2003) *Genes Dev* 17, 991-1008.

[8] Lim L.P., Glasner M.E., Yekta S., Burge C.B. and Bartel D.P. (2003) *Science* 299, 1540.

[9] Lai E.C., Tomancak P., Williams R.W. and Rubin G.M. (2003) *Genome Biol* 4, R42.

[10] Watanabe Y., Tomita M. and Kanai A. (2007) *Methods Enzymol* 427, 65-86.

[11] Rajewsky N. (2006) *Nat Genet* 38 Suppl, S8-13.

[12] Yousef M., Jung S., Kossenkov A.V., Showe L.C. and Showe M.K. (2007) *Bioinformatics* 23, 2987-2992.

[13] Wang X. (2008) *RNA* 14, 1012-1017.

[14] Xiao F., Zuo Z., Cai G., Kang S., Gao X. and Li T. (2009) *Nucleic Acids Res* 37, D105-D110.

[15] Chiromatzo A.O. et al. (2007) *Genet Mol Res* 6, 859-865.

[16] Lagana A. et al. (2009) Database (Oxford) 2009, bap008.

[17] Bao L., Zhou M., Wu L., Lu L., Goldowitz D., Williams R.W. and Cui Y. (2007) *Nucleic Acids Res* 35, D51-D54.

[18] Li S.C., Shiau C.K. and Lin W.C. (2008) *Nucleic Acids Res* 36, D184-D189.

[19] Thadani R. and Tammi M.T. (2006) *BMC Bioinformatics* 7 Suppl 5, S20.

[20] Hsu P.W., Lin L.Z., Hsu S.D., Hsu J.B. and Huang H.D. (2007) *Nucleic Acids Res* 35, D381-D385.

[21] Bandyopadhyay S. and Mitra R. (2009). *Bioinformatics* 25, 2625-2631

Table 1 - Available online resources for microRNAs:

Database/tool	Description of advantages	Reference (s)
Databases		
miRDB	MiRNA target prediction and functional annotation	[13]
miRecords	An integrated resource for microRNA–target interactions	[14]
miRGen	Relationship between animal microRNA genomic organization and function	[1]
miRNApath	Relationships between the target genes, miRNAs and metabolic pathways	[15]
miRò	A web-based knowledge base that provides users with miRNA–phenotype associations in humans	[16]
PolymiRTS	Linking polymorphisms in microRNA target sites with complex traits	[17]
Vir-Mir db	A database containing predicted viral miRNA candidate hairpins	[18]
Tool for miRNA gene identification		
MiRscan	Relies on the observation that the known miRNAs derive from phylogenetically conserved stem loop precursor RNAs with characteristic features	[7-8]
Tools for miRNA target prediction		
MicroTar	Uses predicted free energies of unbound mRNA and putative mRNA-miRNA heterodimers, implicitly addressing the accessibility of the mRNA 3' untranslated region	[19]
NBmiRTar	Uses machine learning by a naïve Bayes classifierand both the 'seed' and 'out-seed' segments of the miRNA:mRNA duplex for target identification	[12]
ViTa	Curate the known virus miRNA genes and the known/putative target sites of human, mice, rat and chicken miRNAs	[20]
TargetMiner	MicroRNA target prediction with systematic identification of tissue-specific negative examples	[21]

Note: For other miRNA databases (Argonaute, ASRP, miRBase, miRNAMap and Tarbase) and other miRNA gene and target prediction tools (MicroInspector, miTarget, RNAhybrid, GUUGle, MovingTargets, miRanda, FastCompare, TargetBoost, PicTar, rna22, miRU, TargetScanS and DIANA-microT), see [9,10]. The websites of these databases listed in this table are given in Table 2.

Table 2 - Websites of available online resources for microRNAs:

Database/tool	Website
Databases	
miRDB	http://mirdb.org/miRDB/
miRecords	http://mirecords.biolead.org/
miRGen	http://www.diana.pcbi.upenn.edu/miRGen.html
miRNApath	http://lgmb.fmrp.usp.br/mirnapath/
miRò	http://ferrolab.dmi.unict.it/miro/
PolymiRTS	http://compbio.utmem.edu/miRSNP/
Vir-Mir db	http://alk.ibms.sinica.edu.tw/cgi-bin/miRNA/miRNA.cgi
Tool for miRNA gene identification	
MiRscan	http://genes.mit.edu/mirscan/
Tools for miRNA target prediction	
MicroTar	http://tiger.dbs.nus.edu.sg/microtar/
NBmiRTar	http://wotan.wistar.upenn.edu/NBmiRTar/
ViTa	http://vita.mbc.nctu.edu.tw/#
TargetMiner	http://www.isical.ac.in/~bioinfo_miu/

Note: For other miRNA databases (Argonaute, ASRP, miRBase, miRNAMap and Tarbase) and other miRNA gene and target prediction tools (MicroInspector, miTarget, RNAhybrid, GUUGle, MovingTargets, miRanda, FastCompare, TargetBoost, PicTar, rna22, miRU, TargetScanS and DIANA-microT), see [9,10].