Annotation of miRNAs in the COVID-19 Novel Coronavirus
Tie-Ying Yu, Min Chen, and Chun-De Wang


View online: https://doi.org/10.1016/j.jnlest.2020.100060

Articles you may be interested in


Lan-Ju Liang, Zhang Zhang, Xin Yan, Xin Ding, De-Quan Wei, Qi-Li Yang, Zhen-Hua Li, Jian-Quan Yao. Broadband Terahertz Transmission Modulation Based on Hybrid Graphene-Metal Metamaterial[J]. Journal of Electronic Science and Technology, 2018, 16(2): 98-104. doi: 10.11989/JEST.1674-862X.70727018


Follow JEST WeChat public account for more information
Annotation of miRNAs in the COVID-19 Novel Coronavirus

Tie-Ying Yu | Min Chen | Chun-De Wang*

Abstract—The coronavirus disease 2019 (COVID-19) coronavirus is a new strain of coronavirus that had not been previously detected in humans. As its severe pathogenicity is concerned, it is important to study it thoroughly to aid in the discovery of a cure. In this study, the microRNAs (miRNAs) of COVID-19 were annotated to provide a powerful tool for the study of this novel coronavirus. We obtained 16 novel coronavirus genome sequences and the mature sequences of all viruses in the microRNA database (miRbase), and then used the miRNA matures sequences of the virus to perform the Basic Local Alignment Search Tool (BLAST) analysis in the coronavirus genome, extending the matched regions of approximately 20 bp to two segments by 200 bp. Six sequences were obtained after deleting redundant sequences. Then, the hairpin structures of the mature miRNAs were determined using RNAfold. The mature sequence on one hairpin arm was selected into a total of 4 sequences, and finally the relevant miRNA precursor prediction tools were used to verify whether the selected sequences are miRNA precursor sequences of the novel coronavirus. The miRNAs of the novel coronavirus were annotated by our newly developed method, which will lay the foundation for further study of this virus.

Index Terms—microRNA (miRNAs) annotation, novel coronavirus, nucleotide Basic Local Alignment Search Tool (BLASTN), RNAfold.

1. Introduction

The coronavirus disease 2019 (COVID-19), began to erupt in December 2019[1], can cause coughing, fatigue, and diarrhea as well as severe respiratory problems, pneumonia and death[2]. The COVID-19 coronavirus is primarily transmitted by respiratory droplets and is easily spread, and highly infectious[3]. After entering the human body, a series of processes occur, including viral invasion, viral replication, cell apoptosis, and viral fragments dissemination that triggers the human immune response to promote inflammatory reactions, and even inflammatory
storms, leading to pathological damage. Many people have been infected by the COVID-19 coronavirus and died in many countries and regions worldwide. Coronaviruses are coated plus stranded single stranded RNA viruses with a diameter of 80 nm to 120 nm. Their genetic material is the largest of all RNA viruses and infects only humans, mice, pigs, cats, dogs, and poultry vertebrates. Coronaviruses are excreted from the body through respiratory secretions, and is transmitted by oral fluid, spray, contact, and airborne droplets. At present, there are no effective drugs against this virus[4],[5], and little research has been performed, leading to an urgent need to identify effective therapeutics.

The microRNA (miRNA) is non-coding single-stranded RNA molecules with the lengths of approximately 22 nucleotides and regulate gene expression at the posttranscriptional level[6]-[8]. The miRNA can regulate many types of cancers[9],[10] and can be proto-oncogenes or tumour suppressor genes that are closely associated with many diseases[9],[10],[13]-[21]. Viral miRNA is a newly discovered type of miRNA that regulates the expression of host cells and the target genes of the virus by inducing mRNA cleavage degradation, translation, inhibition, or other mechanisms, causing changes in host cell activity or the replication of the virus[22]. To combat and evade host immune surveillance, the virus must protect itself. Finding viral miRNA molecules and identifying their biological functions has become a research hotspot. Thus, it is important to annotate the miRNAs of the novel coronavirus to better understand its mechanism of action and lay the foundation for basic research and drug development[23],[25].

In this study, we obtained 16 novel coronavirus genome sequences and annotated the potential miRNAs. The nucleotide Basic Local Alignment Search Tool (BLASTN) was used to identify the mature miRNA sequences of the novel coronavirus, while RNAfold was used to generate secondary structures and miRNA precursor prediction tools were used to predict the miRNA precursor sequences. To improve the accuracy of the results, 5 miRNA prediction methods were used for each sequence. Our results provide a theoretical basis for the study of the novel coronavirus and lay a foundation for drug development.

2. Materials and Methods

2.1. Datasets

The 16 novel coronavirus sequences used in this study were downloaded from the web site https://bigd.big.ac.cn/ncov/release_genome and all of the viral miRNA mature sequences were obtained from miRBase (http://www.mirbase.org/).

2.2. Scheme of Analysis

As shown in Fig. 1, we first used mature viral miRNAs for the BLAST search in the novel coronavirus genome and then extended the matched regions by 200 bp on either side. The secondary structures were generated by using RNAfold, and the hairpin sequences with a mature miRNA on the arm were observed. Finally, 5 miRNA prediction tools were used to verify whether the selected sequences were miRNA precursor sequences.

2.3. BLAST Analysis

Formatdb (version 2.2.20) was used to preprocess the sequences from the genome of the novel coronavirus to ensure that the next search step could be quickly completed. Then BLASTN (version 2.2.20) was used to perform “sequence alignment” of the mature sequences of the virus with the sequences of the novel coronavirus genome. Then the matched sequences of approximately 20 bp were extended on both sides by 200 bp.

2.4. Search for Novel Coronavirus miRNA Precursors

The extended sequences were folded with RNAfold[26],[28] to form secondary structures and to identify a hairpin
structure sequence containing a mature miRNA sequence on one arm for each different secondary structure. Finally, 4 sequences were identified.

2.5. Prediction of Novel Coronavirus miRNA Precursors

We used 5 miRNA precursor prediction tools to independently verify whether the 4 novel coronavirus sequences were true miRNA precursors. The web sites for these 5 tools are as follows:

- MiPred: http://server.malab.cn/MiPred/predict.jsp
- iMiRNA-PseDPC: http://bioinformatics.hitsz.edu.cn/iMiRNA-PseDPC/
- miRNA-dis: http://bioinformatics.hitsz.edu.cn/miRNA-dis/
- iMcRNA: http://bioinformatics.hitsz.edu.cn/iMcRNA/ (This URL contains 2 methods: iMcRNA-PseSSC and iMcRNA-ExPseSSC.)

3. Results

3.1. BLAST Analysis

After the BLAST analysis, we obtained 16 matching regions of approximately 20 bp. When the repeat sequences were removed, 6 sequences remained. The sequence information is shown in Table 1.

Table 1: Matching sequence information

<table>
<thead>
<tr>
<th>Name of the sequences</th>
<th>Sequence coded information</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN938384.1 dev-miR-D7-5p</td>
<td>CTGTACAGGCTATATAATGTTA</td>
</tr>
<tr>
<td>MN938384.1 hvt-miR-H12-5p</td>
<td>TTCTTGAGACCTGTGAAGAAAACA</td>
</tr>
<tr>
<td>MN938384.1 mcpv-miR-P1-3p</td>
<td>TTGGTCCTTCTCTTAGTCGTTGA</td>
</tr>
<tr>
<td>MN938384.1 mcpv-miR-P1-3p</td>
<td>AACAGGAAGAGAATCAGCAACT</td>
</tr>
<tr>
<td>MN938384.1 mcpv-miR-P1-3p</td>
<td>AGTTGCTGATTCTCTCTCCTGTT</td>
</tr>
<tr>
<td>MN938384.1 rlcv-miR-rL1-3p</td>
<td>TTATTTCTTCTACTAGGGT</td>
</tr>
</tbody>
</table>

3.2. Use of RNAfold to Identify miRNA Precursors of Novel Coronavirus

RNAfold was used to fold the extended sequences and form secondary structures. After identifying the hairpin structure sequence of the mature miRNA sequences on one hairpin arm, 4 sequences remained. The secondary structures are shown in Fig. 2, where the hairpin structure with the mature miRNA is in the red frame and the detail information of the precursor sequences is shown in Table 2.
3.3. Prediction of miRNA Precursors of Novel Coronavirus

We used 5 miRNA precursor prediction tools to verify whether the 4 sequences we identified represent the actual miRNA precursor sequences of the novel coronavirus. The 5 tools used in this study were MiPred, iMiRNA-PseDPC, iMcRNA-PseSSC, iMcRNA-ExPseSSC, and miRNA-dis, and the prediction results are shown in Table 3. The results showed that one of the sequences we identified (named MN938384.1 dev-miR-D7-5p MIMAT0028177) was not a miRNA precursor sequence of the novel coronavirus, while the remaining 3 sequences seemed to be miRNA precursor sequences of the novel coronavirus. The location of the 3 novel coronavirus miRNA precursors and the mature sequences are shown in the gff3 format in Table 4.

Table 2: Information of the 4 precursor sequences

<table>
<thead>
<tr>
<th>Name of the sequences</th>
<th>Sequence coded information</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN938384.1 dev-miR-D7-5p MIMAT0028177</td>
<td>CUGUUACAGCGUAUACGCUUACUUCUUCUAAA</td>
</tr>
<tr>
<td>MN938384.1 mcpv-miR-P1-3p MIMAT0010151 (first)</td>
<td>ACACCUGAAGCAUUUUAGAAACCUACGCUUUCGUGG</td>
</tr>
<tr>
<td>MN938384.1 mcpv-miR-P1-3p MIMAT0010151 (second)</td>
<td>UUGGUUCCUCUCUAGCGUUGACACAACCUAAUGCGACAGAAUA</td>
</tr>
<tr>
<td>MN938384.1 mcpv-miR-P1-3p MIMAT0010151 (third)</td>
<td>UUGCGUGUUAUUCGUUCUAAAUAAUCCGCAU</td>
</tr>
</tbody>
</table>

Table 3: Prediction results of the 4 novel coronavirus miRNA precursors

<table>
<thead>
<tr>
<th>Name of the sequences</th>
<th>MiPred</th>
<th>iMiRNA-PseDPC</th>
<th>iMcRNA-PseSSC</th>
<th>iMcRNA-ExPseSSC</th>
<th>miRNA-dis</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN938384.1 dev-miR-D7-5p MIMAT0028177</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>MN938384.1 mcpv-miR-P1-3p MIMAT0010151 (first)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MN938384.1 mcpv-miR-P1-3p MIMAT0010151 (second)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MN938384.1 mcpv-miR-P1-3p MIMAT0010151 (third)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note: “Yes” means that the tool predicts that the sequence is a miRNA precursor sequence, and “No” means that the tool predicts that the sequence is not a miRNA precursor sequence.
4. Discussion

The virus infection is one of the greatest threats to human health. A virus is a non-cellular organism that is tiny and simple in structure and contains only one kind of nucleic acid (DNA or RNA), which must be parasitized in a living cell and multiplied by replication. Its replication, transcription, and translation capabilities are all carried out in the host cell. When it enters the host cell, it can use the material and energy in the cell to complete life activities.

The COVID-19 coronavirus has caused severe infections recently\cite{29,30}. In just one week, more than eight hundred individuals became infected. People infected with the coronavirus exhibit the common signs of respiratory symptoms, fever, cough, and breathing difficulties\cite{31}. In more severe cases, the infection can lead to dyspnoea, metabolic acidosis, and coagulation disorders as well as organ failure and even death. The speed at which the disease spreads is very fast, occurring over a wide area and involving a large proportion of the population\cite{32,33}. COVID-19 has become a worldwide epidemic in a short time across provincial and international boundaries\cite{34}. At present, there are infectious persons all over the country and even abroad\cite{35}. However, we do not have a deep understanding of the molecular information of the novel coronavirus, which is needed to promote the development of drugs.

The miRNA is a kind of endogenous noncoding RNA with regulatory functions found in eukaryotes. Mature miRNAs are produced by a series of cleavage of primary transcripts by nuclease, and then assembled into RNA induced silencing complex (RISC). The target mRNA is recognized by base complementary pairing, and the translation of the target mRNA is guided by the different degree of complementation\cite{36}. Recent studies have shown that the miRNA is involved in various regulatory pathways, including development, virus defense, hematopoiesis, cell proliferation, and apoptosis. Viral miRNAs have been shown to induce the degradation of mRNAs or regulate the expression of host cells and viral target genes through other mechanisms\cite{37,38}. Therefore, it is important to annotate the miRNA of the novel coronavirus to gain an in-depth understanding of the virus and provide the basis for drug and vaccine development.

5. Conclusions

In this study, we annotated the miRNAs of the novel coronavirus. We used BLASTN to compare the mature sequences of all viruses with the genome sequences of the novel coronavirus, and RNAfold was subsequently used to generate the secondary structures of the extended sequences. Finally, we identified the precursor sequences and verified them using miRNA precursor prediction tools. Through comparisons using a variety of prediction tools, the accuracy of the prediction was improved, increasing the quality of the data. Finally, we
identified 3 miRNA precursor sequences of the novel coronavirus and annotated their mature and precursor sequences. Our results can provide a theoretical basis for use by other researchers to accelerate the study of COVID-19.

Disclosures

The authors declare no conflicts of interest.

References


Tie-Ying Yu was born in Yantai in 1997. She is pursuing her B.S. degree with Shandong Agricultural University, Tai’an. She has been recommended to the Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai for pursuing her Ph.D. degree. From October 2019 to June 2020, she studies bioinformatics at University of Electronic Science and Technology of China, Chengdu. Her current research interests include scallop stock improvement, big data, and biocomputing.

Min Chen was born in Liaocheng in 1991. She received her B.S., M.S., and Ph.D. degrees from Shandong Agricultural University in 2018. Then, she joined the Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences. She is the author of 8 articles. Currently, her research interests include scallop stock improvement, chloroplast development, and chlorophyll synthesis in peaches.

Chun-De Wang was born in Qingdao in 1967. He received his B.S. degree in marine biology from Ocean University of China, Qingdao in 1989, his M.S. degree in malacology from the Institute of Oceanology, Chinese Academy of Sciences, Qingdao in 1992, and his Ph.D. degree from Dalhousie University, Halifax in 2001. He was a research assistant with the Institute of Oceanology, Chinese Academy of Sciences from 1992 to 1995. From 2000 to 2006, he worked as a senior research and development consultant on scallop breeding and aquaculture with Pec-Nord Inc., Quebec. Since 2007, he has joined Qingdao Agricultural University (QAU), Qingdao, and he is now a professor with QAU and a researcher with the Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences. His research interests include scallop stock improvement, seed production, and adult grow-out techniques, quantitative trait locus (QTL) mapping, scallop genome-wide association studies, and genome-selection breeding techniques. He is the author of more than 40 articles, 28 patents, and 3 officially approved new scallop strains. He is a senior member of Chinese Society of Malacology and a member of National Shellfish Association, USA.