Genes & Diseases (2023) 10, 668–670

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/en/journals/genes-diseases

RAPID COMMUNICATION

3′-UTR of the SARS-CoV-2 genome as a possible source of piRNAs

SARS-CoV-2 is the causal agent of the COVID-19 pandemic. This is a single-stranded RNA beta-coronavirus that is comprised of a 5′ and 3′-untranslated region (UTR) in its non-coding RNA (ncRNA) region. The P-element-induced wimpy testis (PIWI)-interacting RNAs (piRNAs) are ncRNA sequences that bind to transposons and interfere with the translation of new genes. Here, we search for short sequences in the non-coding regions, corresponding to 28 nucleotides (nt). These start from a uracil and have a conserved region of at least 10 nt of homology with the previously reported piRNAs. The aim of this work was to identify piRNA-like sequences in the 3′-UTR of SARS-CoV-2 from the reported Wuhan Patient's genome sequence (GenBank: MN908947.3).

As a first analysis, we performed alignments of 32,286 piRNAs from piRBase (http://bigdata.ibp.ac.cn/piRBase/, release V2.0), 27,700 from piRNAdb (https://www.pirnadb.org/download/archive), and 51,509 from piRNAQuest (http://bicresources.jcbose.ac.in/zhumur/pirnaquest/), using the 3′-UTR of the SARS-CoV-2 Wuhan patient's genome as a reference. Since we found no homologies within the piRNAs using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi), we used a text editor manually.

Some authors consider piRNAs to be sequences constituting 21 to 35 nt long, while others expect them to have approximately 26–31 nt long. We analyzed piRBase, piRNAdb, and piRNAQuest as shown in the flow diagram (Fig. 1A) and found that, on average, the number of nucleotides corresponds to 28 nt. Zhang et al[1] found that piRNAs, like miRNAs, have a 7 nt "seed region" which needs to interact with their complementary nucleotides. Therefore, there is no consensus criterion to identify a piRNA sequence; nevertheless, it is known that 78%–94% of piRNAs begin with a uracil at their 5′-UTR (position +1),[2] most piRNAs have an A at position +4, and an A or G at position −1.[3] Based on the above, we used the following criteria to identify the possible piRNAs: 1) 28 nt sequences that shared at least 10 consecutive nucleotides with the piRNAs previously reported in databases, 2) a uracil in its 5' (in position +1),[2] 3) an adenine (A) in position +4, 4) an adenine (A) or guanine (G) at position −1.[3]

In addition, for ensuring that the selected sequences were piRNAs, we used the 2L-piRNA server (http://bioinformatics.hitsz.edu.cn/2L-piRNA/#). Furthermore, we looked for 2′-OH methylated sites at the 3′-UTR end using the NmSEER V2.0 server (http://www.rnanut.net/nmseer-v2/). Also, we used the turbofold server to predict the 3′-UTR RNA secondary structure (http://linearfold.org/linearturbofold). By last, we evaluated the identity percentage, E-value, and the Waterman-Eggert scores between the sequences proposed as piRNAs in the 3′-UTR of SARS-COV-2, and their homologous sequences through LALIGN (https://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi), using +5/-4 scoring matrix and changing the gap penalty, to obtain an approximation of similarities and biological relevance.

To confirm that 3′-UTR is a conserved region, we made alignments of the 3′-UTR of the SARS-CoV-2 from *alpha*, *gamma*, *delta*, *lambda*, *mu*, and *omicron* variants, comparing them with the original sequence of the Wuhan patient's genome using Clustal X. Sequences were obtained from GISAID (https://www.gisaid.org/).

Of the fifteen sequences that had 28 nt and matched the characteristics indicated above, two sequences, 29,792–29819 (hsa_piRNA_23,430) and 29,757–29784 (hsa_pRNA_25,334) meet all four criteria, twelve of them showed an A or G at position −1. And only two had an A at position +4, 29,792–29819 and 29,757–29784. Using the 2L-piRNA server, we found six sequences that were positive for piRNAs. Moreover, the server predicted that such sequences have a deadenylation function. Sequence alignment showed more than 10 nt of identity. The following scores were measured between the piRNA sequences in the database and the sequences found of the 3′-UTR of SARS-CoV-2; Waterman-Eggert score: more than 52, E-value: less than 0.1, and identity: between 54.3% and 87.5%. Taking

Peer review under responsibility of Chongqing Medical University.

https://doi.org/10.1016/j.gendis.2022.05.028

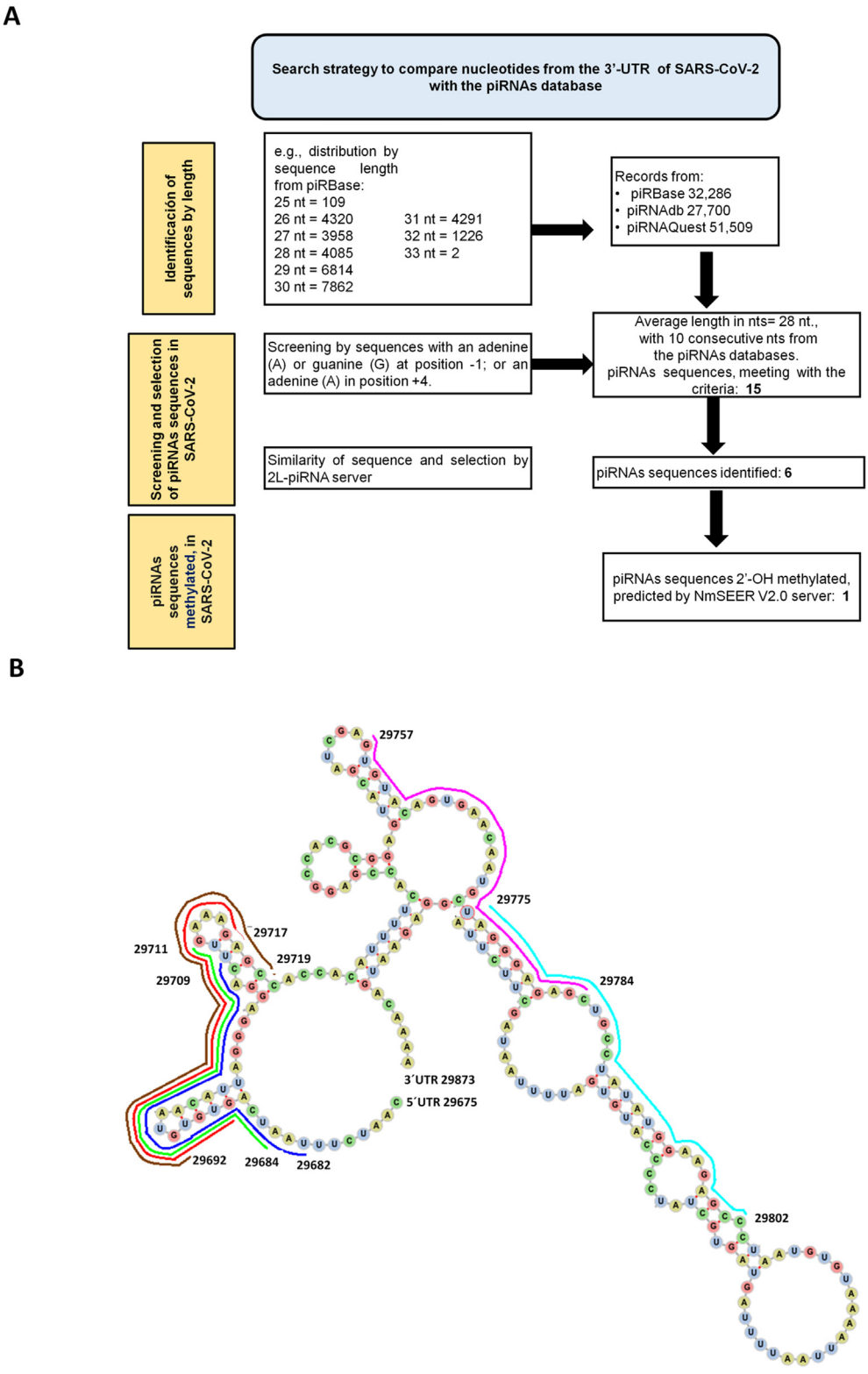

Figure 1 3′-UTR of the SARS-CoV-2 genome as a possible source of piRNAs. **(A)** Flow diagram for identification of piRNAs in 3′-UTR of SARS-CoV-2. **(B)** Localization of piRNA-like sequences in the predicted secondary structure of 3′-UTR of SARS-CoV-2 through LinearFold. The sequence 29,682—29709 (blue), 29,684—29711 (green), 29,688—29717 (red) and 29,692—29719 (brown) are spliced in a common region, while the sequences 29,757—29784 (purple) and 29,775—29802 (cyan) are in a distant region.

into consideration of the % identity, E-value and Waterman-Eggert scores, all sequences selected showed high homology with the piRNAs in the databases (Supplementary data). In addition, RNA secondary structure showed four piRNA-like sequences overlapped in a region and two intersecting in another region, as illustrated in Figure 1B.

The NmSEER V2.0 server predicted four methylation sites at positions 29,719 (C), 29,759 (T), 29,761 (T), and 29,766 (T). However, only one of the piRNA sequences reported here (29,692—29719 homolog to hsa_pRNA_1752) has a 2′-OH methylated site in its penultimate Cytosine (GTAACATTAGGGAGGACTTGAAAGAGCC).

Relating to the alignments comparing the 3′-UTR from Wuhan's SARS-CoV-2 sequence (MN908947) versus the *alpha*, *gamma*, *delta*, *lambda*, *mu*, and *omicron* variants, we found that this entire region is highly conserved. We observed various random mutations in each one of the SARS-CoV-2 variants. However, these mutations do not affect the identification of piRNAs by the 2L-piRNA server.

This is the first report describing the presence of homologous to known piRNA sequences in the 3′-UTR region. We have identified 6 sequences with a high probability of functioning as piRNAs in the 3′-UTR of SARS-CoV-2 from the Wuhan patient sequence. All of them are 28 nt long and have ten or more consecutive nucleotides that are contained in piRNAs already mentioned in different databases such as piRBase, piRNAQuest, and pirnaDB. In addition, they have homologous nucleotides in upstream or downstream positions.

Due to piRNAs could be associated with different pathologies, such as the promotion of tumorigenesis, we suggest that if these sequences are present in high amounts in the human cell after exposure to SARS-CoV-2, as in the case of severe COVID-19; there is a high probability of a mismatch occurring in the homeostasis of the piRNAs already present in the cell, as in carcinomas.

We should note that this is a theoretical study, with no experimental evidence. Another limitation to mention is that the LinearFold server is used to predict the secondary structure, and this depends on the beam size and the number of base pairs, of which only the 3′-UTR was predicted.

We found six piRNA-like sequences in the 3′-UTR of SARS-CoV-2. These meet the criteria indicated above. Their identification has been verified by the 2L-piRNA server and by the similarities in sequences. Further research is needed to verify this finding.

Author contributions

C.R.D. and E.P.C.: conceptualization. M.T.H.H., L.P.C.M., C.A.M.C., C.R.D. and E.P.C.: formal analysis, investigation, methodology, data curation, software, figure preparation, writing original draft, review and editing. E.C.P., E.P.C.M., I.C.M., R.B.H., M.M.C., G.M.A. and M.E.A.V.: investigation, review and editing.

Conflict of interests

The authors declare that they have no conflict of interests.

Funding

This work was supported by Faculty of Medicine of the Autonomous University Benito Juarez of Oaxaca, Mexico, the National Technology of Mexico (TecNM)/IT Oaxaca, and CONACYT.

Acknowledgements

The authors would like to thank the support throughout the work, to the Faculty of Medicine of the Autonomous University Benito Juarez of Oaxaca, Mexico; to the National Technology of Mexico (TecNM)/IT Oaxaca CONACYT- BP-PA-2021050723-4900732-959110 for financial support, and also to Charlotte Grundy for technical review.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gendis.2022.05.028.

References

1. Zhang D, Tu S, Stubna M, et al. The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. *Science*. 2018;359(6375):587—592.
2. Betel D, Sheridan R, Marks DS, et al. Computational analysis of mouse piRNA sequence and biogenesis. *PLoS Comput Biol*. 2007;3(11):e222.
3. Brayet J, Zehraoui F, Jeanson-Leh L, et al. Towards a piRNA prediction using multiple kernel fusion and support vector machine. *Bioinformatics*. 2014;30(17):i364—i370.

María Teresa Hernández-Huerta [a,1], Laura Pérez-Campos Mayoral [b,1], Carlos Alberto Matias-Cervantes [a,1], Carlos Romero Díaz [a,**], Eli Cruz Parada [c], Eduardo Pérez-Campos Mayoral [b], Rafael Baltiérrez-Hoyos [a], Margarito Martínez Cruz [c], Gabriel Mayoral Andrade [b], Eduardo Pérez-Campos [c,*]

[a] *CONACyT, Faculty of Medicine and Surgery, Autonomous University "Benito Juárez" of Oaxaca (UABJO), Oaxaca 68020, Mexico*
[b] *Research Center, Faculty of Medicine UNAM-UABJO, Autonomous University "Benito Juárez" of Oaxaca (UABJO), Oaxaca 8020, Mexico*
[c] *National Technology of Mexico/IT Oaxaca, Oaxaca de Juárez, Oaxaca 68030, Mexico*

*Corresponding author.

**Corresponding author.
E-mail addresses: carlos.rom.74he@gmail.com (C. Romero Díaz), perezcampos@prodigy.net.mx, pcampos@itoaxaca.edu.mx (E. Pérez-Campos)

18 January 2022
Available online 8 June 2022

[1] These authors contributed equally to this work.