Pre-Cursor microRNAs from Different Species classification based on features extracted from the image

Dr. Rabeb Touati[1*], Dr. Imen Ferchichi[2] , Dr. Imen Messaoudi[3], Dr. Afef Elloumi Oueslati[4], Dr. Zied Lachiri[5], Dr. Maher Kharrat[6]

[1]Post-Doctoral position, University of Tunis El Manar, LR99ES10 Human Genetics Laboratory, Faculty of Medicine of Tunis (FMT), Tunisia, Rabeb.touati.enit@gmail.com

[2]Post-Doctoral position, University of Tunis El Manar, LR99ES10 Human Genetics Laboratory, Faculty of Medicine of Tunis (FMT), Tunisia, imene.ferchichi14@gmail.com

[3]Assiatnt Professor, University of Carthage, Higher Institute of Information Technologies and Communications, Industrial Computing Department, Tunisia imen.messaoudi@enit.rnu.tn

[4] Associate Professor, University of Carthage, National School of Engineers of Cartage, Electrical Engineering Department, Tunisia, Afef.Elloumi@enit.utm.tn

[5] Professor, University Tunis El Manar, SITI Laboratory, National School of Engineers of Tunis, Tunisia, Zied.lachiri@enit.rnu.tn

[6] Associate Professor, University of Tunis El Manar, LR99ES10 Human Genetics Laboratory, Faculty of Medicine of Tunis (FMT), Tunisia, maher.kharrat@fmt.rnu.tn

Abstract: The first MicroRNAs was discovered 27 years ago in the nematode C.elegans genomes. MicroRNAs (miRNAs) sequences are small and are expressed in various genomes to affect the translation or the stability of target mRNAs. These short RNA sequences are involved in targeting post-transcriptional gene regulation. The mature miRNAs are derived from longer sequence precursors (pre-miRNAs). Previous works have shown that pre-miRNAs can be classified by their species of origin using bioinformatics techniques combined with machine learning tools. In this study, we focus on the classification of Precursor microRNAs sequences, from 16 different species ranging from animals, plants, and viruses, based on the combination of the features extracted from images corresponding to DNA sequences and machine learning algorithms. As a result, our classification shows that the system based on features correspond to energy images of pre-miRNAs signals using the PNUC coding technique corresponding to the DNA sequence is very efficient in terms of miRNAs inter-genomics recognition

Keywords: *microRNA, Precursor microRNA, Features, scalogram, wavelet-energy, Classification*

1. Introduction

Protein sequences are tightly regulated on different levels since their dysregulation may frequently lead to disease. One of these regulatory levels is post-transcriptional adjustment governed by microRNAs [1-3] that directly modulates protein abundance. Regulation can also be done at protein stability or the gene level. Mature microRNAs sequences (miRNAs) interact with messenger RNAs (mRNAs) via hybridization. This interaction mechanism leads to modulation of the translation rate [4-5]. A stretch of approximately 22 nucleotides incorporated in the RNA-Induced Silencing Complex (RISC) function as translation repressor or induces mRNAs degradation

Received: Jan 15, 2020 Revised: March 21, 2020 Accepted: April 29, 2020

DOI: 10.5281/zenodo.3786135

This post-transcriptional regulation type has been described for many genomes ranging from viruses [2] to plants [6-7]. MicroRNAs sequences are not functional by themselves, but when co-expressed with their targets they will be involved in modulating the protein abundance. miRNAs are a non-coding RNA (ncRNA), this RNA will not be translated into protein, but participate in various cellular and physiological processes by miRNAs constitute an abundant class of small, and evolutionarily conserved ncRNA molecules. Therefore, many approaches have been developed to detect miRNAs and their target mRNAs and genes [8-11]. Recent studies are based on machine learning. Then, computational prediction of pre-miRNAs has become important and most algorithms employ machine learning [12-13]. Machine learning models have been established for many species; metazoan [14] and plants [6]. These algorithms depend on the parameterization of the folded pre-miRNAs three-dimensional structure [5]. As an example, the K-mers are short nucleotide sequences and Hamming distance features that have been used early on for the machine learning-based ab initio detection of pre-miRNAs [15-17]. Here, we have established a feature, a vector with a size equal to 64, based on scalogram miRNA-image. In our previous work, the time-frequency features were extracted by applying the Wavelet Transform (CWT) to the helitrons' signals [18-22].

2. Proposed methods

In this paper, we focus on the classification of miRNAs sequences in various genomes based on the scalogram images extracted from the wavelet transform applied PNUC coding technique. Our methodology (as described in the flowchart below) consists of five steps:

1. Generating a PNUC signal to represent each C. elegans chromosome (2 is the scale of the FCGS representation).
2. Replace the U nucleotide with T nucleotide
3. Generating scalograms images after applying the wavelet analysis to the PNUC signals
4. Calculate the energy value following this formula:

$$E(s) = \sum_{u=1}^{L} \left| W_{(s,u)}[x(t)] \right|^2 \qquad (1)$$

5. Splitting the data into two sets: 70% for training and 30% for testing
6. Extracting statistical features for each data set
7. Applying the Random Forest classification technique

The flowchart describing our methodology is presented in Figure 1.

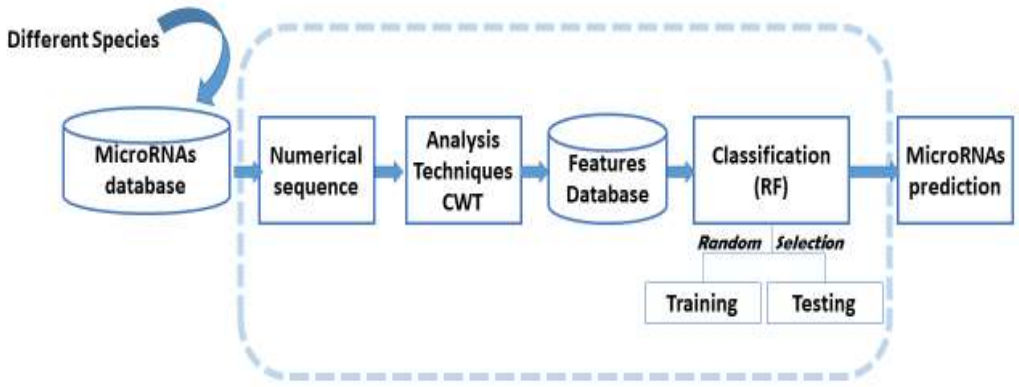

Fig 1: RF- miRNAs recognizer Flowchart

Received: Jan 15, 2020 Revised: March 21, 2020 Accepted: April 29, 2020

DOI: 10.5281/zenodo.3786135

2.1 miRNA database

First of all, we start by extracting the miRNAs data for all species available from the miRTarbase (Release 6.0: Sept.20, 2019), which included 16 spicies; the species Hominidae, Brassicaceae, Hexapoda, Monocotyledons (Liliopsida), Nematoda, Fabaceae, Pisces (Chondricthyes), Virus, Aves, Laurasiatheria, Rodentia, Homo sapiens, Cercopithecidae, Embryophyta, Malvaceae, and Platyhelminthes. The following figure presents the repartition numbers of the precursors miRNAs sequences of each genome available in the miRBase.

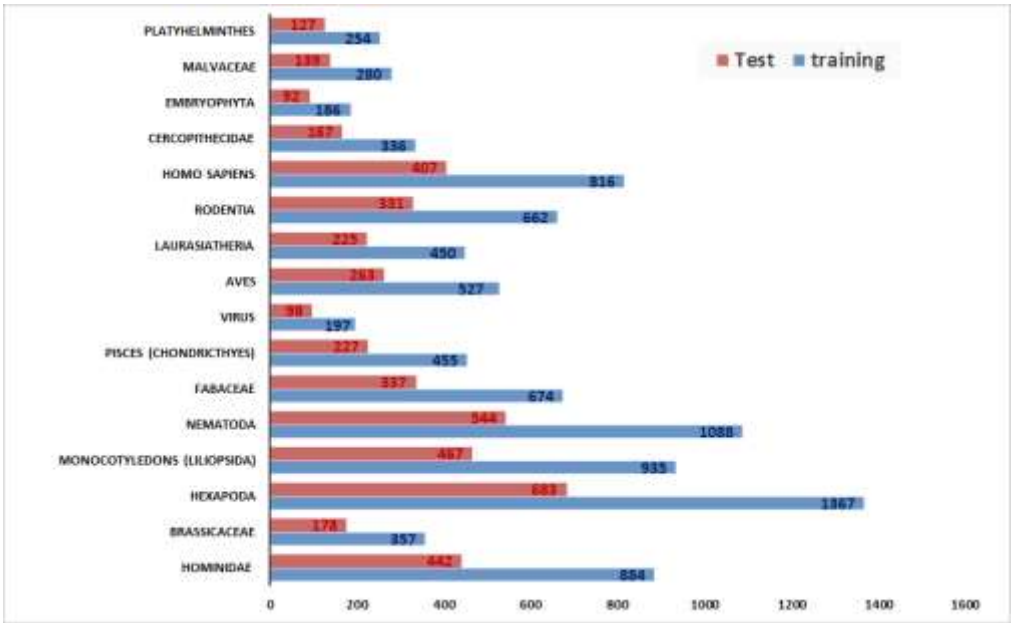

Fig 2: pre-miRNAs list in different species used in the study

2.2 miRNA numerical representation

This step corresponds to the establishment of the helitron's signal database by associating a PNUC [23-24] signal to each miRNA sequence after changing the 'U' by 'T' nucleotide. The PNUC technique is coding based on the local bending and flexibility properties of the double helix. It is based on the scratches curvatures measurement related to 3 nucleotides during nucleosome positioning. The corresponding experimental values are given in Table 1.

$$S_{pnuc} = \sum_i Pnuc_{3nucleotde} \quad (i) \qquad (1)$$

Here, I represent the number of apparitions of the 3 nucleotides in the whole sequence. Here, a new miRNAs database signals are established.

Received: Jan 15, 2020 Revised: March 21, 2020 Accepted: April 29, 2020

DOI: 10.5281/zenodo.3786135

Codon	PNUC	Codon	PNUC
AAA/ TTT	0	CAG/ CTG	4.2
AAC/ GTT	3.7	CCA/TGG	5.4
AAG/ CTT	5.2	CCC/ TGC	6
AAT/ ATT	0.7	CCG/ CGG	4.7
ACA/ TGT	5.2	CGA/ TCG	8.3
ACC/ GGT	5.4	CGC/ GCG	7.5
ACG/ CGT	5.4	CTA/ TAG	2.2
ACT/ AGT	5.8	CTC/GAG	5.4
AGA/ TCT	3.3	GAA/ TTC	3
AGC/ GCT	7.5	GAC/CTG	5.4
AGG/ CCT	5.4	GCA/ GGG	6
ATA/ TAT	2.8	GCC/ GGC	10
ATC/ GAT	5.3	GGA/ TCC	3.8
ATG/ CAT	6.7	GTA/ TAC	3.7
CAA/ TTG	3.3	TTA/ TAA	2
CAC/ GTG	6.5	TCA/TGA	54

Table1: The experimental values associated with each codon (3 successive bases) for the PNUC coding technique

As an example, we represent the signal of miRNA sequences: Seq1 with length equal to 112 bp located in *the Homsapiens* genome and Seq2 length equal to 166 bp located in *Brassicaceae* genome.

- Seq1='GGGCAGCCAGTGAATAGTTAGCTGGTGCAAAAGTAATTGCGGTCTTTGGTATTA CTTTCAGTGGCAAAAACTGCATTACTTTTGCACCAGCCTACTAGAACGCTGAGTTCAG
- Seq2='ATGACAGAGCTACAAGAAGAAAGTCAAATGTCTTCATGAGTTCCCTTTAACGCT TCATTGTTAATATTCAAACTACATTTATTTCCATAAATGTTTGATCGATTTGTATATAAC ACTGAAGTGTTTGGGGGGACTTTTGGTGTCATTCTGGCGTAGTTACGATTAT'

Figure 3 shows miRNA sequences and their correspondent signals using the PNUC coding technique.

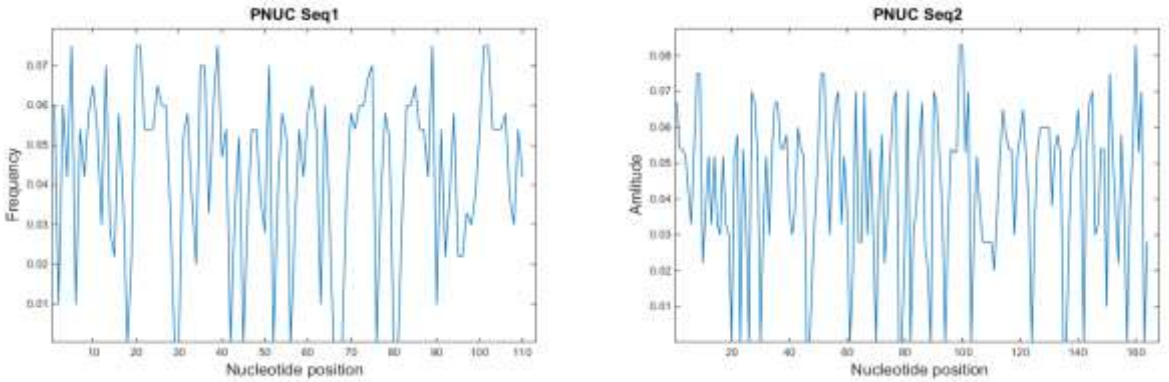

Fig 3: Illustration of PNUC signals of two different pre-miRNAs sequences from *Homsapiens* and *Brassicaceae* genomes

Received: Jan 15, 2020 Revised: March 21, 2020 Accepted: April 29, 2020

DOI: 10.5281/zenodo.3786135

2.3 Features Extraction Steps

In this work, we use the features extracted from the wavelet coefficients' matrix (wavelet energy) to classify miRNAs. These matrices are obtained by applying the continuous wavelet transform to the PNUC signals with the Complex Morlet Wavelet is taken as an analysis window. The Complex Morlet Wavelet (CWT) decomposes a given signal into a sum of windows called wavelets. These windows are obtained by expanding and translating a mother wavelet function ψ(t) [25-27].
The daughter wavelets are obtained by:

$$\psi_{s,u}(t) = \frac{1}{\sqrt{s}}\psi^*\left(\frac{t-u}{s}\right),\ s > \ u \in \mathbb{R} \qquad (2)$$

Here, * represents the complex conjugate. The complex Morlet wavelet represents a Gaussian-windowed complex sinusoid that is generated by:

$$\psi_{cmor}(t) = \Pi^{-\frac{1}{4}}\left(e^{i\omega_0 t} - e^{i\omega_0{}^2}\right)e^{-\frac{t^2}{2}}, \qquad (3)$$

Here, ω_0 represents oscillation's number. The CWT of a miRNA-signal *x(t)* is obtained by the given expression:

$$W_{(s,u)}[x(t)] = \frac{1}{\sqrt{s}}\int_{-\infty}^{+\infty} x(t)\,\psi^*\left(\frac{t-u}{s}\right)dt. \qquad (4)$$

The result is a matrix that contains the wavelet coefficients. By considering the magnitude of these coefficients, we generate a 2-D representation in the time-scale or the time-frequency plan: the scalogram. Therefore, we prepare the energy vectors as our features database correspond to miRNAs genomic sequences and to be passed to the RF-classifier.

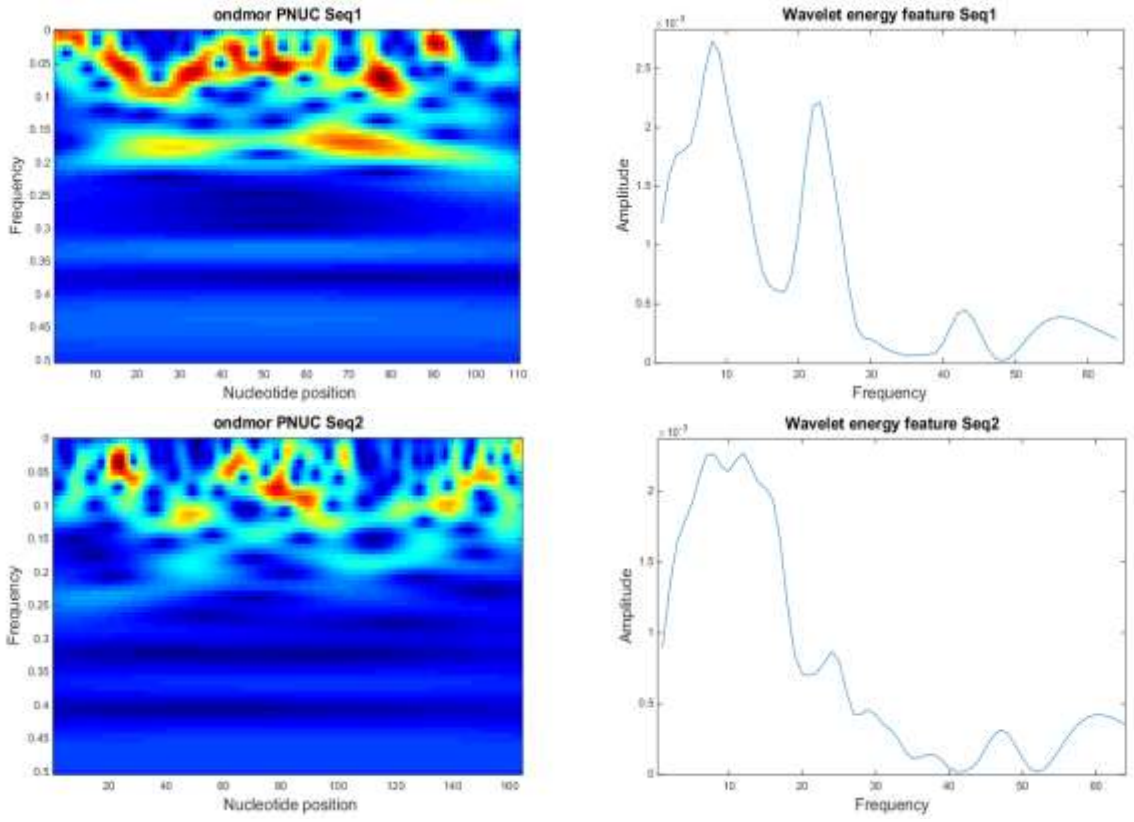

Fig 4: example of miRNAs signatures (scalograms) and their corresponding features vector plots of two different species (*Homsapiens* and *Brassicaceae*) used in the study

Received: Jan 15, 2020 Revised: March 21, 2020 Accepted: April 29, 2020
DOI: 10.5281/zenodo.3786135

2.4 Data categorization

In which data is randomly splitted into two sub-databases mutually exclusive: training set (70%) and testing set (30%). The models are trained and tested using 100 fold Monte Carlo cross-validation [28]. Here we have a total of 14195 vectors as input dataset; among them, 9464 vectors are used as a training set and the remaining 4731 vectors are used for the testing set.

2.5 Classification steps

In this work, we focus on precursor miRNAs sequences classification from various species. For the classification, we choose the Random Forest (RF) technique. The RF is a supervised machine learning technique, which makes use of a technique called feature bootstrap or bagging. Considering the problem of the time running, we build here 100 trees. In this work, we use the RF under the data analytics platform KNIME (version 3.1.2) [29]. The input samples in the RF classification are the energy vectors. For evaluation, we have then calculated the values of precision, sensitivity, specificity, F-measure, Cohen's kappa, Matthew's correlation coefficient and accuracy area under the curve (AUC) (with TP: true positive, FP: false positive, TN: true negative, and FN referring to false-negative classifications).

Sensitivity = TP / (TP + FN)
Specificity = TN / (TN + FP)
Accuracy = (TP + TN) / (TP + TN + FP + FN)

3. Experiments and Results

MicroRNA precursor identification is an important task in bioinformatics. In this study, to predict these sequences the signal processing tools have been used. Based on the training data (energy wavelet vectors), the goal of our work is to predict the target values of the test data given only its attributes. To classify the miRNAs sequences, firstly, we extract the biological information (type, location. Then, we apply the PNUC coding to miRNAs sequence after changing the U letter by T letter. Afterward, we prepare the features from the wavelet images which will be considered as entries to the classifier. Finally, we use RF algorithm to predict the miRNAs. For each pair of species/miRNAs, we trained a classifier using 100-fold MCCV.

First, we evaluated the known k-mer features and how accurately they can categorize the given data into the tested species and clades (Table 2). The average performance of 100-fold MCCV was recorded for all pairs of miRNAs, with one of the miRNAs used as the positive class and the other miRNAs as the negative one. For example, *Embryophyta* vs. *Aves* leads to an average accuracy of 92% using 100-fold MCCV. The average performance of 100-fold MCCV was recorded for all pairs of clades, with one of the clades used as the positive class and the other as the negative one.

In addition, we can draw phylogenetic trees of miRNAs in 16 genomes using these classification results (Table2). This presented figure (Fig.5) was developed using the Past PAleontological Statistics program version 3.23 (free software for scientific data analysis) [30].

Received: Jan 15, 2020 Revised: March 21, 2020 Accepted: April 29, 2020
DOI: 10.5281/zenodo.3786135

	Viruses	Monocotyledons	Fabaceae	Embryophyta	Brassicaceae	Malvaceae	Platyhelminthes	Nematoda	Hexapoda	Pisces	Aves	Laurasiatheria	Rodentia	Hominidae	Homo sapiens	Cercopithecidae
Viruses	0,10	0,85	0,83	0,92	0,90	0,92	0,73	0,68	0,72	0,65	0,69	0,90	0,64	0,61	0,60	0,65
Monocotyledons	0,85	0,10	0,62	0,63	0,65	0,71	0,79	0,79	0,75	0,81	0,78	0,74	0,77	0,81	0,84	0,80
Fabaceae	0,83	0,62	0,10	0,73	0,59	0,64	0,72	0,77	0,69	0,77	0,76	0,79	0,76	0,79	0,82	0,75
Embryophyta	0,92	0,63	0,73	0,10	0,71	0,76	0,89	0,88	0,83	0,90	0,86	0,68	0,85	0,89	0,90	0,89
Brassicaceae	0,90	0,65	0,60	0,71	0,10	0,62	0,81	0,86	0,79	0,86	0,87	0,81	0,85	0,86	0,89	0,86
Malvaceae	0,92	0,70	0,64	0,76	0,62	0,10	0,82	0,87	0,80	0,87	0,88	0,82	0,87	0,87	0,90	0,88
Platyhelminthes	0,73	0,79	0,72	0,89	0,81	0,81	0,10	0,69	0,56	0,71	0,73	0,87	0,74	0,72	0,73	0,70
Nematoda	0,69	0,79	0,77	0,88	0,86	0,86	0,69	0,10	0,66	0,64	0,70	0,86	0,67	0,69	0,70	0,66
Hexapoda	0,71	0,75	0,69	0,83	0,78	0,80	0,56	0,67	0,10	0,68	0,69	0,82	0,72	0,72	0,73	0,67
Pisces	0,66	0,82	0,77	0,89	0,87	0,87	0,70	0,64	0,69	0,10	0,64	0,85	0,62	0,63	0,65	0,58
Aves	0,67	0,78	0,76	0,86	0,86	0,88	0,72	0,70	0,69	0,63	0,10	0,81	0,62	0,64	0,67	0,57
Laurasiatheria	0,90	0,74	0,79	0,68	0,80	0,82	0,87	0,85	0,83	0,85	0,81	0,10	0,81	0,85	0,88	0,85
Rodentia	0,64	0,78	0,76	0,85	0,85	0,87	0,74	0,68	0,72	0,62	0,62	0,82	0,10	0,57	0,60	0,59
Hominidae	0,61	0,81	0,79	0,88	0,86	0,88	0,72	0,69	0,72	0,63	0,65	0,86	0,57	0,10	0,14	0,59
Homo sapiens	0,60	0,84	0,82	0,90	0,89	0,89	0,72	0,70	0,73	0,65	0,67	0,88	0,59	0,14	0,10	0,62
Cercopithecidae	0,65	0,79	0,75	0,89	0,86	0,88	0,70	0,67	0,67	0,58	0,57	0,85	0,59	0,61	0,61	0,10

Table2: The experimental values associated with each codon (3 successive bases) for the PNUC coding technique

The phylogenetic trees figure (Fig.5) confirms the similarities between some r of species/ miRNAs. The results presented in Table 2 for energy-wavelet features confirm the previous observation [31] that were use the k-mer as features. Table 3 presents the average performance (Recall, Precision, specificity, sensitivity, F-measure, MCC, an under the curve, and Cohen's kappa) of all miRNAs species (16 species) using the RF classification model and the energy-wavelet features extracted from the scalogram images.

Recall	Precision	Sensitivity	Specifity
0,8957	0,9116	0,89571429	0,9116
F-measure	Cohen's kappa	MCC	Area Under Curve
0,9027	0,8073	0,8089	0,9643

Table3: Evaluation measures of the RF classification model using the features extracted from the Wavelet Transform.

Received: Jan 15, 2020 Revised: March 21, 2020 Accepted: April 29, 2020

DOI: 10.5281/zenodo.3786135

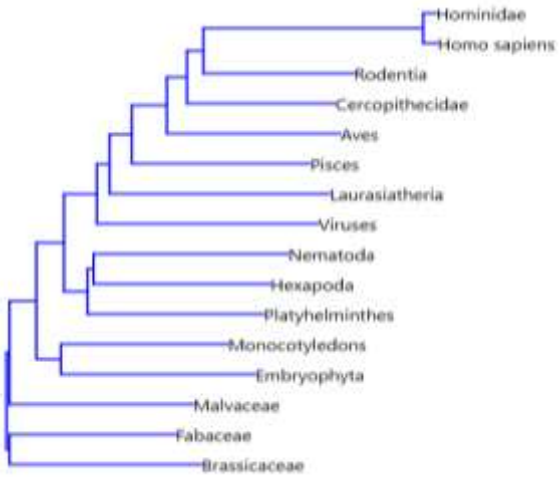

Fig 5: Phylogenetic tree among organisms and clades used in this study

4. Conclusion

Many studies have been conducted in the aim to understand biological phenomena. Different bioinformatics algorithms have been employed to attend this goal. MicroRNAs are involved in post-transcriptional gene regulation. They have been described for many species covering most of the tree of life. MicroRNA evolution is subject to investigations [32]. Nevertheless, this work has addressed miRNAs analysis in the framework of signal processing. A major difficulty in identifying and classifying miRNAs comes from the unspecified length and the unbalanced appearance number of each genome. The purpose of this paper was to characterize and classify miRNAs from different species based on the combination of signal processing tools and machine learning approaches. In this study, the novelty consisted of classifying different microRNAs clades based on features extracted from images, scalograms, (energy-wavelet) resulting from a PNUC coding technique (numerical sequences). Thereafter, a feature vector was extracted from the wavelet coefficient matrix. In this sense, RF was selected as a classifier technique. The model has been trained and tested using a 100-fold Monte Carlo cross-validation, where the dataset was split into 70% training and 30% testing.

References

[1] Erson-Bensan, A. E., "Introduction to microRNAs in biological systems", Methods in molecular biology (Clifton, N.J.), 1107, pp. 1–14. DOI: 10.1007/978-1-62703-748-8_1, 2014

[2] Grey, F. "Role of microRNAs in herpesvirus latency and persistence", The Journal of general virology, 96 (Pt 4), pp. 739–51. DOI: 10.1099/vir.0.070862-0, 2015

[3] Griffiths-Jones, S. (2010) 'miRBase: microRNA sequences and annotation.', Current protocols in bioinformatics/editorial board, Andreas D. Baxevanis ... [et al.], Chapter 12, p. Unit 12.9.1-10. DOI:10.1002/0471250953.bi1209s29

[4] Saçar, M. and Allmer, J. "Machine Learning Methods for MicroRNA Gene Prediction", in Yousef, M., and Allmer, J. (eds) miRNomics: MicroRNA Biology and Computational Analysis SE - 10. Humana Press (Methods in Molecular Biology), pp. 177–187. DOI: 10.1007/978-1-62703-748-8_10, 2014

[5] Sacar, M. D. and Allmer, J. "Data mining for microrna gene prediction: On the impact of class imbalance and feature number for microrna gene prediction", in 2013 8th International Symposium on Health Informatics and Bioinformatics. IEEE, pp. 1–6. DOI: 10.1109/HIBIT.2013.6661685, 2013

Received: Jan 15, 2020 Revised: March 21, 2020 Accepted: April 29, 2020

DOI: 10.5281/zenodo.3786135

[6] Yousef, M., Allmer, J. and Khalifa, W. "Accurate Plant MicroRNA Prediction Can Be Achieved Using Sequence Motif Features", Journal of Intelligent Learning Systems and Applications, 8(1), pp. 9–22. DOI: 10.4236/jilsa.2016.81002. 2016

[7] Yousef, M., Allmer, J. and Khalifa, W., "Feature Selection for MicroRNA Target Prediction – Comparison of One-Class Feature Selection Methodologies", in Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies, pp. 216–225. DOI: 10.5220/0005701602160225, 2016

[8] Ng, K. L. S., & Mishra, S. K. "De novo SVM classification of precursor microRNAs from genomic pseudo hairpins using global and intrinsic folding measures", Bioinformatics, 23(11), 1321-1330, 2007

[9] Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., … Rajewsky, N. " Combinatorial microRNA target predictions", Nature Genetics, 37(5), 495–500. DOI:10.1038/ng1536, 2005

[10] Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., … Bartel, D. P. "The microRNAs of Caenorhabditis elegans". Genes & Development, 17(8), 991–1008. DOI:10.1101/gad.1074403, 2003

[11] Yousef, M., Jung, S., Kossenkov, A. V, Showe, L. C., & Showe, M. K. "Naive Bayes for microRNA target predictions machine learning for microRNA targets". DOI:10.1093/bioinformatics/btm484, 2007

[12] Allmer, J. "Computational and bioinformatics methods for microRNA gene prediction", Methods in molecular biology (Clifton, N.J.), 1107, pp. 157–75. DOI: 10.1007/978-1-62703-748-8_9, 2014

[13] Saçar, M. and Allmer, J. "Machine Learning Methods for MicroRNA Gene Prediction", in Yousef, M., and Allmer, J. (eds) miRNomics: MicroRNA Biology and Computational Analysis SE - 10. Humana Press (Methods in Molecular Biology), pp. 177–187. DOI: 10.1007/978-1-62703-748-8_10, 2014

[14] Allmer, J. and Yousef, M. "Computational methods for ab initio detection of microRNAs", Frontiers in genetics, 3, p. 209. DOI: 10.3389/fgene.2012.00209, 2012

[15] Lai, E. C., Tomancak, P., Williams, R. W., & Rubin, G. M. "Computational identification of Drosophila microRNA" genes. Genome Biol, 4(7), R42. DOI:10.1186/gb-2003-4-7-r42, 2003

[16] Yousef, M., Khalifa, W., Acar, Ilhan Erkin, & Allmer, J. "MicroRNA categorization using sequence motifs and k-mers. BMC Bioinformatics, 18(1), 170. DOI:10.1186/s12859-017-1584-1, 2017

[17] Yousef, M., Levy, D., & Allmer, J. "Species Categorization via MicroRNAs - Based on 3'UTR Target Sites using Sequence Features". In Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 4: BIOINFORMATICS, (pp. 112–118). SciTePress. DOI:10.5220/0006593301120118, 2018

[18] Touati, R., Messaoudi, I., Oueslati, A. E., & Lachiri, Z. "A combined support vector machine-FCGS classification based on the wavelet transform for Helitrons recognition in C. elegans", Multimedia Tools and Applications, 78(10), 13047-13066, 2019

[19] Touati, R., Oueslati, A. E., Messaoudi, I., & Lachiri, Z. "The Helitron family classification using SVM based on Fourier transform features applied on an unbalanced dataset", Medical & biological engineering & computing, 57(10), 2289-2304, 2019

[20] Touati, R., Messaoudi, I., Oueslati, A. E., & Lachiri, Z. "Distinguishing between intra-genomic helitron families using time-frequency features and random forest approaches. Biomedical Signal Processing and Control, 54, 101579, 2019

[21] Touati, R., Messaoudi, I., Oueslati, A. E., & Lachiri, Z. "Classification of Helitron's Types in the C. elegans Genome based on Features Extracted from Wavelet Transform and SVM Methods", In Bioinformatics , pp. 127-134. 2018

[22] Touati, R., Messaoudi, I., Oueslati, A. E., Lachiri, Z., & Kharrat, M. "New Intraclass Helitrons Classification Using DNA-Image Sequences and Machine Learning Approaches". IRBM. 2020

[23] Parisi, V., De Fonzo, V., & Aluffi-Pentini, F., "STRING: finding tandem repeats in DNA sequences", Bioinformatics, 19(14), 1733-1738, 2003

[24] D.S. Goodsell and R.E. Dickerson, "Bending and curvature calculations in B-DNA", Nucleic Acids Research, vol. 22, n 24, pp: 5497-5503, Oxford University Press, 1994

[25] A. GROSSMANN, J.MORLET, Decomposition of Hardy functions into square integrable wavelets of constant shape, SIAM, J. mathematical analysis. 15 ,pp 723-736, 1984

[26] D.E. Newland, "Time-frequency and time-scale signal analysis by harmonic wavelets", Signal Analysis and Prediction, Birkhäuser, Boston, 3-26, 1998

[27] R.J.E. MERRY, M. STEINBUCH, "Wavelet theory and applications". Literature Study, Eindhoven University of Technology, Department of Mechanical Engineering, Control Systems Technology Group, 2005

[28] XuQS, LiangYZ. "Monte Carlo cross validation", Chemom Intell Lab Syst 2001;56:1–11. https://doi .org /10 .1016 /S0169 -7439(00)00122 -2.

[29] BertholdMR, et al. "KNIME: the Konstanz information mine". In: SIGKDD explo-rations, p.319–26. 2008

[30] HammerØ., HarperDA, RyanPD. "PAST: paleontological statistics software pack-age for education and data analysis", Palaeontol Electronica 2001;4:9. https://folk.uio .no /ohammer /past/.

[31] Yousef, M. "Hamming Distance and K-mer Features for Classification of Pre-cursor microRNAs from Different Species". In International Conference on Smart Innovation, Ergonomics and Applied Human Factors, pp. 180-189. Springer, Cham, 2019

[32] Tanzer, A., & Stadler, P. F. "Evolution of microRNAs". Methods in Molecular Biology (Clifton, N.J.), vol 342, pp 335–50. doi:10.1385/1-59745-123-1:335, 2006

Received: Jan 15, 2020 Revised: March 21, 2020 Accepted: April 29, 2020

DOI: 10.5281/zenodo.3786135