MicroRNA Prediction Using a Fixed-Order Markov Model Based on the Secondary Structure Pattern

Wei Shen[1], Ming Chen[1], Guo Wei[1], Yan Li[1,2]*

1 Medical Research Center, Southwest Hospital, Third Military Medical University, Chongqing, China, **2** Bioinformatics Laboratory, Chongqing Key Laboratory for Disease Proteomics, Chongqing, China

Abstract

Predicting miRNAs is an arduous task, due to the diversity of the precursors and complexity of enzyme processes. Although several prediction approaches have reached impressive performances, few of them could achieve a full-function recognition of mature miRNA directly from the candidate hairpins across species. Therefore, researchers continue to seek a more powerful model close to biological recognition to miRNA structure. In this report, we describe a novel miRNA prediction algorithm, known as FOMmiR, using a fixed-order Markov model based on the secondary structural pattern. For a training dataset containing 809 human pre-miRNAs and 6441 human pseudo-miRNA hairpins, the model's parameters were defined and evaluated. The results showed that FOMmiR reached 91% accuracy on the human dataset through 5-fold cross-validation. Moreover, for the independent test datasets, the FOMmiR presented an outstanding prediction in human and other species including vertebrates, Drosophila, worms and viruses, even plants, in contrast to the well-known algorithms and models. Especially, the FOMmiR was not only able to distinguish the miRNA precursors from the hairpins, but also locate the position and strand of the mature miRNA. Therefore, this study provides a new generation of miRNA prediction algorithm, which successfully realizes a full-function recognition of the mature miRNAs directly from the hairpin sequences. And it presents a new understanding of the biological recognition based on the strongest signal's location detected by FOMmiR, which might be closely associated with the enzyme cleavage mechanism during the miRNA maturation.

Citation: Shen W, Chen M, Wei G, Li Y (2012) MicroRNA Prediction Using a Fixed-Order Markov Model Based on the Secondary Structure Pattern. PLoS ONE 7(10): e48236. doi:10.1371/journal.pone.0048236

Editor: Raya Khanin, Memorial Sloan Kettering Cancer Center, United States of America

Received February 9, 2012; **Accepted** September 28, 2012; **Published** October 30, 2012

Copyright: © 2012 Shen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This project was supported by the Foundation of Innovative Scientific Research of the Third Military Medical University (2007XG53). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: liyan.com@gmail.com

Introduction

MicroRNAs (miRNAs) are ~22-nucleotide RNAs derived from pri-miRNA transcripts through two important enzyme processes, where the first process is recognized and cut by Drosha and DGCR8 for pre-miRNA formation from pri-mRNA, and the second is by Dicer for miRNA maturation from pre-miRNA [1,2], but the recognition mechanism is still obscure [2–5]. Although many miRNAs have been identified in some species by experimental method, it is believed that there are still a large number of miRNAs uncovered, including those with low expression or in other species [6,7]. Therefore, computational prediction is regarded as a rapid and effective way to solve these problems in contrast to the hard experimental work, however, the diversity of the precursors and complexity of enzyme processes bring challenge for computational approaches to distinguish the real miRNAs from the pseudo-miRNA hairpins with similar stem-loops.

To date, there are mainly four kinds of computational approaches have been tried [8]: (1) A homology-based approach, such as miRNAlign [9], aligns the secondary structure of pre-miRNAs to detect miRNAs. (2) A filter-based approach, such as MIRScan [10] and MiRSeeker [11], picks out pre-miRNAs from an initial set of candidate stem-loops based on GC content, minimum free energy (MFE) and structural filters. (3) A target-centered approach depends on the highly conserved motifs in 3′-UTRs [12]. (4) Machine learning approaches include support vector machine (SVM), hidden Markov model (HMM) and naïve Bayes classifier (NBC), such as Triplet-SVM [6], MiPred [13], miRank [14], CID-miRNA [15], HHMMiR [16], CSHMM [17] and MatureBayes [18]. However, the first three approaches are poor to identify new miRNAs across species lack of homologies. Although the machine leaning approaches achieve satisfactory performance in several species, they are generally limited into a single-function prediction, for instance, either only predicting precursors from hairpins [6,13–17] or miRNAs from precursors [18].

In this study, based on the secondary structure pattern of miRNA precursors, we try to find out a common structural feature associated with miRNA formation, and describe a new miRNA predictor by using a fixed-order Markov model in order to realize a full-function recognition of mature miRNA directly from the sequence segments with similar stem-loop hairpin across species.

Materials and Methods

Data Preparation

The sequences of miRNAs companied with their precursors were downloaded from miRBase database (release 16) [19,20], containing 1046 sequences from human, 6746 from vertebrates, 580 from worms, 235 from viruses and 3052 from plants. Among them, 809 human miRNA sequences were randomly selected out as the positive training set (D1), the remaining 237 human sequences and all of those from other species as positive test set. On the other hand, there were 8494 human pseudo precursors and 754 ncRNAs obtained from microPred website [6,21], where 5890 pseudo ones and 551 ncRNAs were randomly selected out as negative training set (D2), and the remaining as negative test set. Based on these data, the model's parameters would be trained only in part of human sequences (D1 and D2), but be estimated in human itself and all other species.

Construction of the Stem-bulge-gap Notation

For this model study, we established a stem-bulge-gap notation to describe the secondary structure of hairpin. Figure 1 illustrated the construction process of the notation, at first, the dot-bracket notation was produced by RNAfold [22,23], then converted to a stem-loop structure and finally converted into the stem-bulge-gap notation. Moreover, to avoid the noise from the stem-branches, we appointed the longest stem as the main stem, and treated other stem-branches into loops, bulges or gaps.

The Establishment of a Fixed-order Markov Model for miRNA Recognition

A modified fixed-order Markov model was employed to explore the secondary structure pattern of miRNA on the stem-bulge-gap notation. According to the style of Begleiter [24] and Shmilovici [25], we let Σ be a finite alphabet of size $|\Sigma|$. In the case of this paper $\Sigma=\{o,|,!,:,x,-\}$ and $|\Sigma|=6$. To consider a sequence $x_1^n=x_1x_2\cdots x_n$ where $x_i \in \Sigma$ was the symbol at the position i, with $1\leq i\leq n$ in the sequence and x_ix_{i+1} was the concatenation of x_i and x_{i+1}. Based on the training set x_1^n, a model parameter Pr was assigned as the probability of the next symbol given the position and previous context. For a context $s\in\Sigma^*$ where the Σ^* represents a fixed length of context set, the learner generated a conditional probability distribution $Pr(x_i|s,i)$ for each symbol $x_i\in\Sigma$. For variable-order Markov (VOM) model estimating conditional distribution of the $Pr(x_i|s)$ with a variant context length $|s|\leq D$, we proposed the conditional distribution with a fixed length $|s|=D$, as a special case of the VOM model.

To calculate the probability of the model, the count $N_{x_i}(s,i)$ denoted the number of occurrences in which symbol x in position i following context s in the training sequence. The conditional probability was defined as

$$Pr(x_i|s,i)=\frac{N_{x_i}(s,i)+1/2}{m+|\Sigma|/2}$$

Where m denoted the number of the training sequences. Once the conditional probability distribution was estimated, the probability of a sequence for a given model could be calculated by

$$\begin{aligned}P(x)&=Pr'(x_1\cdots x_D)\prod_{i=D+1}^{n}Pr(x_i|x_{i-D}\cdots x_{i-1},i)\\&=\prod_{i=D+1}^{n}Pr(x_i|x_{i-D}\cdots x_{i-1},i)\end{aligned}$$

Where $Pr'(x_1\cdots x_D)$ was the occurrence probability of the initial context, and we let it be 1. To distinguish pre-miRNAs from other hairpin sequences, a True model was constructed to represent true pre-miRNA and a False model for pseudo pre-miRNA. Then each stem-bulge-gap sequence x was scored by:

$$\begin{aligned}\text{log-odds-ratio}(x)&=\log\frac{P(x|\text{True})}{P(x|\text{False})}\\&=\sum_{i=D+1}^{n}\log\frac{Pr^+(x_i|x_{i-D}\cdots x_{i-1},i)}{Pr^-(x_i|x_{i-D}\cdots x_{i-1},i)}\\&=\sum_{i=D+1}^{n}\log R\end{aligned}$$

To handle events in different level of counts$N_{x_i}(s,i)$, the calculation of R was defined asM

$$R=\begin{cases}1, & N^+{}_{x_i}(s,i)=0\ \&N^-{}_{x_i}(s,i)=0\\(0.5/m^-)/Pr^-(x_i|s,i), & N^+{}_{x_i}(s,i)=0\ \&N^-{}_{x_i}(s,i)>0\\Pr^+(x_i|s,i)/(0.5/m^+), & N^+{}_{x_i}(s,i)>0\ \&N^-{}_{x_i}(s,i)=0\\\dfrac{Pr^+(x_i|s,i)}{Pr^-(x_i|s,i)}, & N^+{}_{x_i}(s,i)>0\ \&N^-{}_{x_i}(s,i)>0\end{cases}$$

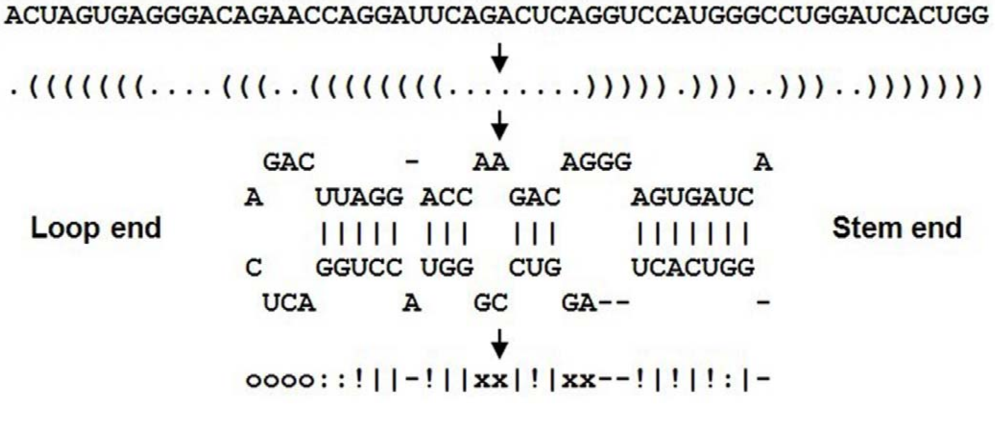

Figure 1. Illustration of the construction of the stem-bulge-gap notation. In the stem-bulge-gap notation at the bottom line, the symbols of '|', '!' and ':' represent respectively the base pair of 'CG', 'AU' and 'GU', the symbols of 'o', 'x' and '-' represent respectively the loop, bulge and gap. In the asymmetric bulges, the symmetric part is indicated with 'x' and the asymmetric part with '-'.
doi:10.1371/journal.pone.0048236.g001

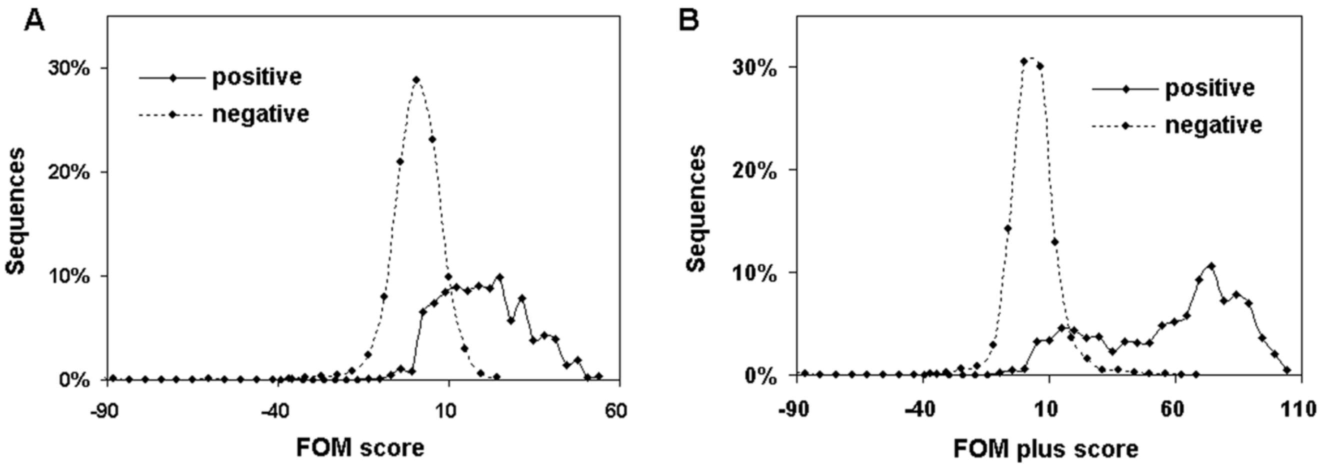

Figure 2. Distribution of the signal scores in positive and negative datasets.
doi:10.1371/journal.pone.0048236.g002

Another similar FOM model were used for mature miRNA strand identification. *StrPr*, like *Pr*, was assigned as the probability of strands given the previous context of stem-bulge-gap sequence and position.

$$StrPr(str_i|s,i)=\frac{N_{str_i}(s,i)}{\sum\limits_{str_i{}'\in STR^*} N_{str_i{}'}(s,i)}$$

Where str was a strand symbol, $STR=\{5',3',both\}$ was alphabet of strands and $str\in STR$, the count $N_{str_i}(s,i)$ denoted the number of occurrences in which strand symbol str in position i following context s. And the strand probabilities of a stem-bulge-gap sequence were calculated by

$$StrP(str|x)=\prod_{i=D+1}^{n} StrPr(str_i|x_{i-D}\cdots x_{i-1},i)$$

The $str\in STR$ with maximum value of $StrP(str|x)$ would be the strand of the sequence x.

Feature Selection

In this study, thirty-six structure features from the previous studies were concerned about as well [21,26–28]. Out of them, only three MFE-related features (MFEI1 [27] MFEI2 [28] MFEI4 [21]) were found be helpful to improve performance of FOM in certain level. Based on a binary logistic regression analysis, the coefficients (MFEI1: −0.209, MFEI2: 0.034, MFEI4: 1.679 and Const: −13.686) were adopted.

Pipeline for the Prediction of miRNA

According to the above model definition, we constructed a miRNA predictor with the pipeline:

(1) **Data preparation.** All the hairpins were converted into the stem-bulge-gap notation for the model computation. The 24 bp stem-bulge-gap segments covering the mature miRNA in precursors of D1 dataset, and the same size segments sliding with 1 bp step size in pseudo miRNA precursors of D2 were used for calculating model parameters.

(2) **Model training.** A set of continuous FOM scores were calculated in a window size of 24 bp sliding on each hairpin from loop to tail with 1 bp step size, and three MFE-related features were added into FOM score as FOM plus score (FOM plus score = FOM score+50×Feature score) for improving the signal. After that, the best FOM plus score was screen out from the first peak followed by a valley of at least 5 bp size. Then one segment with the best score was screened out in each hairpin for judgment, meanwhile, the strand information (5′, 3′ or both) where miRNA located was collected as well. For the training dataset, a threshold of FOM plus score was chosen according to the best classification.

(3) **miRNA judgment.** The screened segment was confirmed as the mature miRNA region, if its FOM plus score reached the threshold. Finally, the miRNA strand was figured out by the highest probability of strand emergence in the FOMmiR predictor.

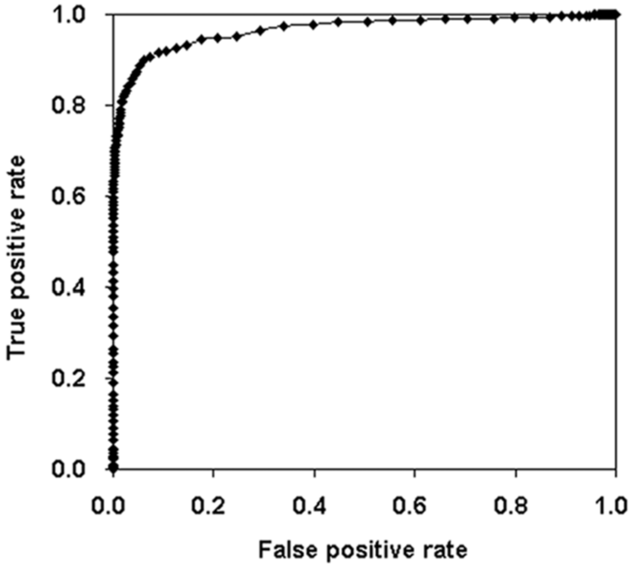

Figure 3. Receiver Operating Characteristic Curve of FOMmiR predictor.
doi:10.1371/journal.pone.0048236.g003

Assessment of Prediction System

Several indexes were used to assess the performance of the model: Sensitivity (*Sen*), Specificity (Spe) [7], and Accuracy (*Acc*). Average sensitivity was measured by 5-fold cross-validation on a positive dataset.

$$Sen=\frac{TP}{TP+FN}$$

$$Spe=\frac{TN}{TN+FP}$$

$$Acc=\frac{TP+TN}{TP+FP+TN+FN}$$

Results

Evaluation of the Model Parameters

As described in Methods, we designed a fixed-order Markov model for miRNA identification, known as FOMmiR. Firstly, the fixed context's length D was been estimated based on the performance of positive and negative signal discrimination. The third order (D = 3) was much stable and chosen into FOMmiR predictor due to a consistent classification between the training and test dataset. Although the third-order Markov model achieved a satisfactory performance with independent FOM score (Figure 2A), the three MFE-related features were added as FOM plus score in order to improve the accuracy in certain level (Figure 2B). Then, Receiver operating characteristic curve (ROC) was drawn for threshold judgment (Figure 3). While the threshold value set to 11, the predictor got the best classification in the training dataset with 91.47% sensitivity (Sen), 91.07% specificity (Spe) and 91.11% accuracy (Acc).

Evaluation of the Model Stability Using Five-fold Cross Validation

In order to examine the stability of FOMmiR predictor, the classification performance was evaluated with 5-fold cross validation. The result showed that the FOMmiR still achieved a high performance with 91.47±2.52% sensitivity, 90.84±0.59% specificity and 90.91±0.70% accuracy, and which was very close to the above performance in the whole training dataset. Therefore, it suggested that this predictor was quite stable.

Table 1. The performances of pre-miRNA prediction.

Method	Year	Algorithm	Sen	Spe	Acc
Triplet-SVM	2005	Support vector machine	72.15%	91.09%	89.62%
MiPred	2007	Random Forest	93.25%	6.59%	13.41%
CIDmiRNA	2008	Stochastic context free grammar	75.95%	96.29%	94.71%
CSHMM	2010	Context sensitive HMM	88.19%	71.46%	72.77%
FOMmiR	2012	Fixed order Markov model	89.45%	91.27%	91.13%

doi:10.1371/journal.pone.0048236.t001

Table 2. Comparison of sensitivity across different species.

Method	Vertebrates (6746)	Plants (3052)	Drosophila (1205)	Worms (580)	Viruses (235)
Triplet-SVM	75.26%	65.27%	85.39%	85.00%	65.11%
MiPred	92.48%	47.02%	93.94%	95.52%	96.60%
CIDmiRNA	75.85%	73.23%	85.81%	86.90%	70.64%
CSHMM	93.60%	91.43%	95.68%	97.76%	91.06%
FOMmiR	91.76%	93.55%	97.18%	97.07%	89.79%

doi:10.1371/journal.pone.0048236.t002

Performance of miRNA Precursor Classification

To evaluate the performance of pre-miRNA classification, a comparative test was performed among different approaches against the independent test dataset composed of 273 real pre-miRNAs and 2807 pseudo pre-miRNAs. Despite some approaches not available any longer, we were fortunate to compare FOMmiR with Triplet-SVM [6], MiPred [13], CID-miRNA [15] and CSHMM [17], among which only CSHMM could be retrained with D1 and D2 dataset. Although the different training conditions of other three approaches might bring some small deviations to the comparative evaluation, at least the result displays that FOMmiR was able to achieve a satisfactory prediction as well as them, or even better (Table 1). Furthermore, the FOMmiR not only identified the real pre-miRNA, but also indicate the mature miRNA position that the others were unable to do. Given these, the FOMmiR exhibited an outstanding discriminatory power.

Performance of Cross-species Classification

Cross-species performance is very important for a model trained in known species to predict new species, so it requests the model should hold a common structure feature for miRNA recognition. Here we tested the FOMmiR prediction rate in four species with the parameters trained only on human dataset. Moreover, a comparative test were performed with the four well-known approaches. The result showed that FOMmiR kept high sensitivities in the specie of vertebrates, worms and viruses, even plants (Table 2). It suggested that the FOMmiR model was reliable across species, and the FOMmiR parameters contained the basic recognition feature of the miRNA.

Performance of Locating Mature miRNA Region

At the same time as the FOMmiR identified the pre-miRNA from the candidate hairpins, the mature miRNA region had been located. we compared its localization performance with that of MatureBayes. The MatureBayes uses naive Bayes algorithm to identify the mature miRNA from pre-miRNA, which has significantly better performance than the two existing approaches with same function, ProMiR and BayesMiRNAfind [18]. For a large number of random test set from different species, the comparative test was performed between FOMmiR and MatureBayes. The distances were calculated between the actual mature miRNA and the predicted mature miRNA. It was obvious in Figure 4 that the localization performance of FOMmiR was close to that of MatureBayes.

Performance of Identifying Mature miRNA Strand

Identifying the mature miRNA strand from the complementary strands seems to be very difficult, few of approaches was reported

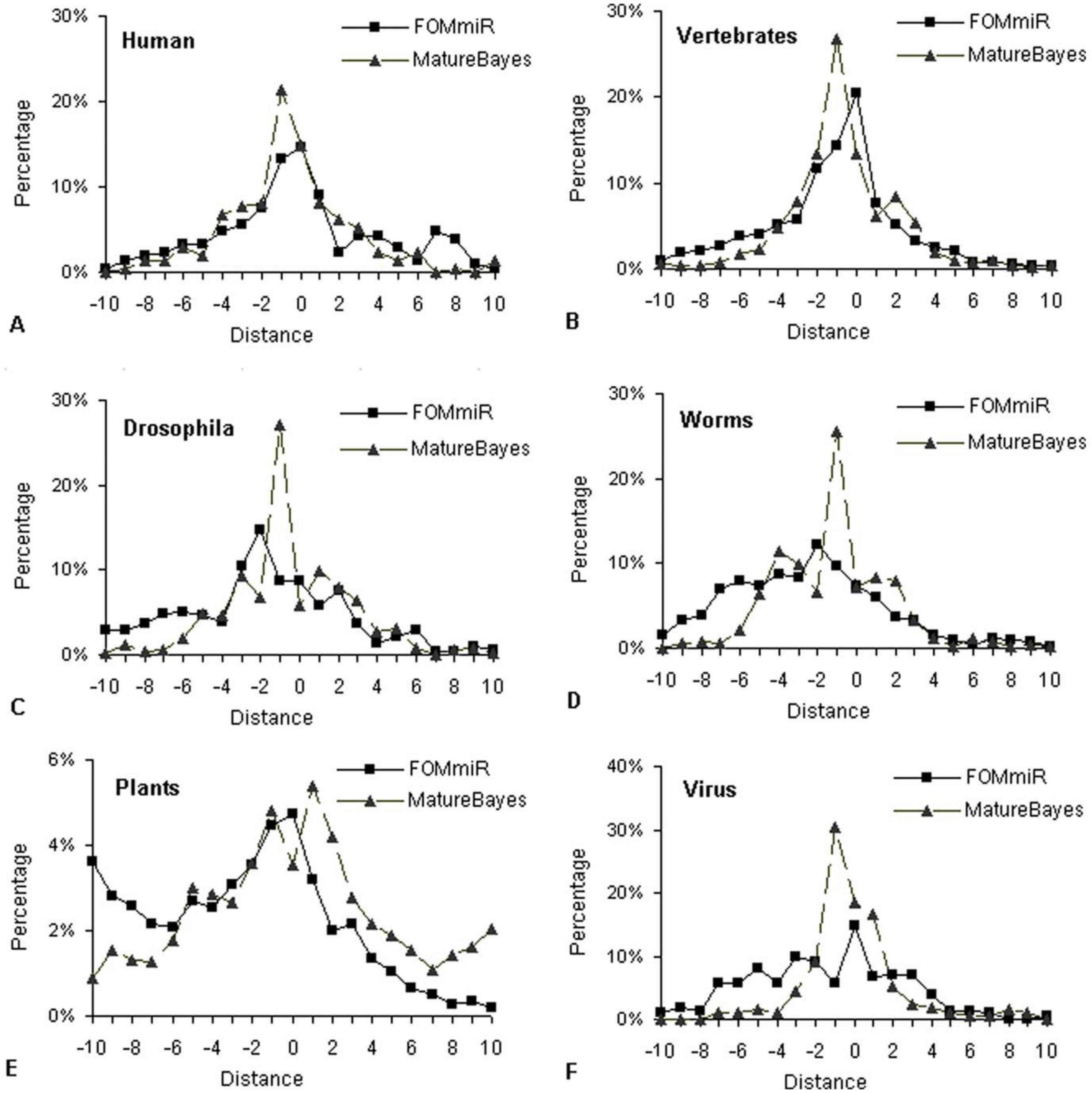

Figure 4. Distribution of distances between the real and predicted mature miRNA region.
doi:10.1371/journal.pone.0048236.g004

Table 3. Quantitative distribution of miRNA strands in positive training dataset.

		Predicted		
	Strand	***5p***	***3p***	***both***
True	***5p***	124	68	25
	3p	0	269	7
	both	0	40	207

doi:10.1371/journal.pone.0048236.t003

Table 4. Quantitative distribution of miRNA strands in positive test dataset.

		Predicted		
	Strand	**5p**	**3p**	**both**
True	**5p**	14	45	17
	3p	0	54	15
	both	1	40	26

doi:10.1371/journal.pone.0048236.t004

to conquer it. Nevertheless, in this study, we extracted the strand information where the mature miRNA located and defined it as another FOM parameter, so the miRNA stand could be checked out from the miRNA region. The result displayed that the strand-check accuracy reached to 86.5% in positive training dataset (Table 3) and 63.7% in the positive test dataset (Table 4).

Discussion and Conclusions

In recent years, a lot of algorithms and models have been tried to predict pre-miRNA or mature miRNA. The excellent ones are always concerned about, not only because it could predict new miRNA, but also because it might interpret the enzyme cleavage mechanism. Although the single-function prediction seems to be close to success, new generation of full-function prediction is very expected.

From the view of system biology, the biological processes always employ the parsimony principle to obtain the best energy efficiency rate. So we suppose that both Drosha/DGCR8 complex and Dicer might focus on a similar structure pattern of miRNA region, despite two independent biological processes needed for the final maturation of miRNA. In this study, we start to find the discriminatory signals in the mature miRNA regions, where the primary structure, secondary structure and their combination have been analyzed respectively in the fixed-order Markov model. But only the pure secondary structure could exhibit the significant signals. It reflects that enzyme recognition is mainly based on the secondary structure.

Based on the secondary structure pattern in the style of stem-bulge-gap notation, the FOMmiR predictor has been built using a fixed-order Markov model and successfully realized a full-function recognition of mature miRNA directly from the hairpins with similar stem-loops. All of the tests displayed that no matter on the classification of precursors, the localization of mature miRNA regions or on the cross-species ability, this approach achieves satisfactory performances in contrast to those well-known ones. Moreover, the FOMmiR experienced a successful trial in identifying the mature miRNA strand, although this function remains to improve.

The secondary structure of pre-miRNAs in plants seems much more complex than those in other species, due to more stem-branches existed in plants. Here we generally focus on the longest stem as the main stem in order to decrease the noise from those stem-branches, so the FOMmiR significantly increased the performance in plants than other algorithms did. On the other hand, with human-trained parameter, both of the FOMmiR and other algorithms obtained a similar result, in which the sensitivity in vertebrates is less than those in Drosophila and worm, even plants (Table 2).

With regard to the model construction, the FOMmiR, as one of machine learning approaches, is much simpler than those of the hidden Markov model, the stochastic context free grammar model and the support vector machine-based methods. Although the actual processes of the biological recognition are obscure, we have got a sense of the potential mechanism during the model construction. Briefly, the quadruple codes on the secondary structure pattern are quite crucial for the miRNA recognition.

Overall, in this study, we provide a new generation of miRNA prediction algorithm, using a fixed-order Markov model based on the secondary structure pattern, which successfully realizes a full-function recognition of the mature miRNAs directly from the hairpin RNA molecules.

Author Contributions

Conceived and designed the experiments: YL WS. Performed the experiments: WS. Analyzed the data: YL WS MC. Contributed reagents/materials/analysis tools: GW. Wrote the paper: YL WS.

References

1. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281–297.
2. Faller M, Guo F (2008) MicroRNA biogenesis: there's more than one way to skin a cat. Biochim Biophys Acta 1779: 663–667.
3. Han J, Lee Y, Yeom KH, Nam JW, Heo I, et al. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 125: 887–901.
4. Faller M, Matsunaga M, Yin S, Loo JA, Guo F (2007) Heme is involved in microRNA processing. Nat Struct Mol Biol 14: 23–29.
5. Sohn SY, Bae WJ, Kim JJ, Yeom KH, Kim VN, et al. (2007) Crystal structure of human DGCR8 core. Nat Struct Mol Biol 14: 847–853.
6. Xue C, Li F, He T, Liu GP, Li Y, et al. (2005) Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. BMC Bioinformatics 6: 310.
7. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, et al. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 37: 766–770.
8. Mendes ND, Freitas AT, Sagot MF (2009) Current tools for the identification of miRNA genes and their targets. Nucleic Acids Res 37: 2419–2433.
9. Dezulian T, Remmert M, Palatnik JF, Weigel D, Huson DH (2006) Identification of plant microRNA homologs. Bioinformatics 22: 359–360.
10. Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, et al. (2003) The microRNAs of Caenorhabditis elegans. Genes Dev 17: 991–1008.
11. Lai EC, Tomancak P, Williams RW, Rubin GM (2003) Computational identification of Drosophila microRNA genes. Genome Biol 4: R42.
12. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, et al. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 434: 338–345.
13. Jiang P, Wu H, Wang W, Ma W, Sun X, et al. (2007) MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. Nucleic Acids Res 35: W339–344.
14. Xu Y, Zhou X, Zhang W (2008) MicroRNA prediction with a novel ranking algorithm based on random walks. Bioinformatics 24: i50–58.
15. Tyagi S, Vaz C, Gupta V, Bhatia R, Maheshwari S, et al. (2008) CID-miRNA: a web server for prediction of novel miRNA precursors in human genome. Biochem Biophys Res Commun 372: 831–834.
16. Kadri S, Hinman V, Benos PV (2009) HHMMiR: efficient de novo prediction of microRNAs using hierarchical hidden Markov models. BMC Bioinformatics 10 Suppl 1: S35.
17. Agarwal S, Vaz C, Bhattacharya A, Srinivasan A (2010) Prediction of novel precursor miRNAs using a context-sensitive hidden Markov model (CSHMM). BMC Bioinformatics 11 Suppl 1: S29.
18. Gkirtzou K, Tsamardinos I, Tsakalides P, Poirazi P (2010) MatureBayes: a probabilistic algorithm for identifying the mature miRNA within novel precursors. PLoS One 5: e11843.
19. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34: D140–144.
20. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36: D154–158.
21. Batuwita R, Palade V (2009) microPred: effective classification of pre-miRNAs for human miRNA gene prediction. Bioinformatics 25: 989–995.
22. Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31: 3429–3431.
23. Hofacker IL, Priwitzer B, Stadler PF (2004) Prediction of locally stable RNA secondary structures for genome-wide surveys. Bioinformatics 20: 186–190.
24. Begleiter R, El-Yaniv R, Yona G (2004) On prediction using variable order Markov models. J Artif Int Res 22: 385–421.
25. Shmilovici A, Ben-Gal I (2007) Using a VOM model for reconstructing potential coding regions in EST sequences. 49–69.
26. Ng KL, Mishra SK (2007) De novo SVM classification of precursor microRNAs from genomic pseudo hairpins using global and intrinsic folding measures. Bioinformatics 23: 1321–1330.
27. Zhang BH, Pan XP, Cox SB, Cobb GP, Anderson TA (2006) Evidence that miRNAs are different from other RNAs. Cell Mol Life Sci 63: 246–254.
28. Ng Kwang Loong S, Mishra SK (2007) Unique folding of precursor microRNAs: quantitative evidence and implications for de novo identification. RNA 13: 170–187.