RNA **Supplementary materials** pages 1–11

Bioinformatics

Unique folding of Precursor MicroRNAs: Quantitative Evidence and Implications for De Novo Identification

Stanley NG Kwang Loong[†,‡,*], Santosh K. MISHRA[†,‡]

[†] Bioinformatics Institute, 30 Biopolis Street, #07-01, Matrix, Singapore 138671

[‡] NUS Graduate School for Integrative Sciences & Engineering, Centre for Life Sciences, #05-01, 28 Medical Drive, Singapore 117456

Email: stanley@bii.a-star.edu.sg*; santosh@bii.a-star.edu.sg

Supplementary materials: http://web.bii.a-star.edu.sg/~stanley/Publications

Key words: Precursor microRNAs; Minimum Free Energy of Folding; Shannon entropy; *Z*-scores; Second eigenvalue

1 RNASPECTRAL

1.1 Representing RNA secondary structure as planar tree-graph

The primary structure of a linear RNA chain molecule is the nucleotide sequence $\mathbf{s} = s_1 s_2 \ldots s_i \ldots s_L$, and runs in the direction 5' → 3' terminus. L defines the number of nucleotides and $s_i \in \sum = (A, C, G, U)$ is the biochemical nucleotide at the i^{th} position. The RNA molecule **s** folds upon itself relatively rapid into a two-dimensional RNA secondary structure S [1]. The structure S is stabilized by the canonical Watson-Crick $G{\equiv}C$ and $A{=}U$, and wobble $G{=}U$ base pairings.

(Fig. S1) A planar RNA secondary structure S is mathematically described by a set of base pairings $(i, j) \in S$ connecting bases s_i and s_j, where $i < j$ [2]. Given (i, j) and $(k, l) \in S$, a nucleotide can base pair to at most one other nucleotide i.e., $i = k \Leftrightarrow j = l$. A set of $\Delta \in \mathbf{Z}^+$ consecutive base pairs defines a stem for stabilizing the structure against thermal fluctuations. The number of unpaired nucleotides between paired s_i and s_j should at most be $\theta \in \mathbf{Z}^+$ i.e., $i < j + \theta$; otherwise, the structural motif is considered an unpaired-loop of multi-branch, bulge, hairpin, or internal.

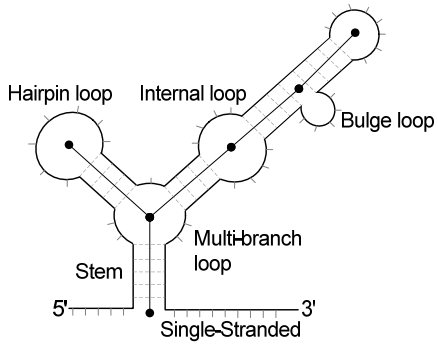

Fig. S1. Planar schematic of RNA secondary structure and its embedded motifs. **Hairpin loop**, folds upon itself; **Internal loop**, an unpaired region between two stems due to mismatched (e.g., AG and CU) or unpaired bases; **Bulge loop**, an asymmetrical internal loop formed from one strand; **Multi-branch loop or junction**, more than two stems coincide with some unpaired bases; **Stem**, a base paired region. *Short and long dashed lines indicate unpaired nucleotides and paired bases.* (•) *and* (—) *represent vertex and edge.*

(Fig. S1) The RNA structure S has two hairpin loops, an internal loop, a bulge loop, a multi-branch loop, and five stems. It is represented as a RNA planar tree-graph $G = (V, E)$ consisting of six vertices (•) and five edges (—) according to the following pair of vertex-edge rules [3,4].

(1) Vertex, V (•) denotes a set of $\theta \geq 1$ mismatched nucleotides or unmatched pairs of bases for hairpin loop, bulge loop, internal loop, the 5' and 3' unpaired termini, and the multi-branch loop. In general, the vertices are arbitrarily labeled in the direction 5' → 3' terminus.

(2) Edge, E (—) denotes a RNA stem having $\Delta \geq 2$ consecutive complementary pairs stabilized by the canonical Watson-Crick $G{\equiv}C$ and $A{=}U$, and wobble $G{=}U$ base pairings

1.2 Converting RNA planar tree-graph to Laplacian matrix

A RNA planar tree-graph $G = (V, E)$ is a mathematical formalism composed of n vertices $v_i \in V$, $i = (1, 2, \ldots, |V|)$ connected by m incident undirected edges $(v_i, v_j) \in E$, each of which is assigned an edge weight E_{ij}. Without loss of generality, edges are unweighted i.e., $E_{ij} = 1$ [5,6]. The tree-graph G in Eq. (1) is uniquely represented by the Laplacian matrix $\mathbf{L}(G)_{n\times n}$.

$$G = (V, E) \leftrightarrow \mathbf{L}(G) = \mathbf{D}(G) - \mathbf{A}(G). \tag{1}$$

Here $\mathbf{D}(G)_{n\times n}$ and $\mathbf{A}(G)_{n\times n}$ are known as the degree and adjacency matrices of the tree-graph G, respectively. The diagonal elements d_{ij} of $\mathbf{D}(G)_{n\times n}$ specify the degree or the minimum number of incident edges that each vertex v_i connects with the other vertices $v_j \neq v_i$, denoted by $\deg(v_i)$. d_{ij} takes on values of $\deg(v_i) = 1$ for hairpin loop, as well as 5' and 3' unpaired termini; $\deg(v_i) = 2$ for internal and bulge loops; and $\deg(v_i) > 2$ for multi-branch loop. The off-diagonal elements a_{ij} of $\mathbf{A}(G)_{n\times n}$ specify whether there exists an incident edge connecting the vertices v_i and v_j. If v_i and v_j are adjacent $a_{ij} = 1$, otherwise $a_{ij} = 0$.

$\mathbf{L}(G)_{n\times n}$ is a symmetric matrix having each of its rows and columns indexed by V, and individually total to zero. The value of element l_{ij} in Eq. (2) is given by the difference between d_{ij} and a_{ij}. It specifies the degree of connectivity between the vertices v_i and v_j of the tree-graph G.

$$l_{ij} = \begin{cases} d_{ij} = \deg(v_i), \text{ if } i = j, \\ -a_{ij} = -1, \text{ if edge } (v_i, v_j) \in E \ \wedge \ i \neq j, \\ 0, \text{ if edge } (v_i, v_j) \notin E. \end{cases} \tag{2}$$

Applying the "Eigen-decomposition theorem" onto $\mathbf{L}(G)_{n\times n}$, as

shown in Eq. (3),

$$\mathbf{L}(G)\mathbf{X} = l\,\mathbf{X} \Leftrightarrow \left[\mathbf{L}(G) - l\,\mathbf{I}\right]\mathbf{X} = \mathbf{O}. \quad (3)$$

Here eigenvalue λ is some scalar of $\mathbf{L}(G)_{n\times n}$ with its corresponding eigenvector $\mathbf{X} \in \Re^n \neq 0$. $\mathbf{I}$ and $\mathbf{O}$ are the identity and null matrices. Equation (3) has non-trivial solutions if and only if the condition in Eq. (4) is satisfied,

$$\det\left|\mathbf{L}(G) - l\,\mathbf{I}\right| = 0. \quad (4)$$

Solving the n^{th}-degree characteristic polynomial in Eq. (4) generates the entire set of ordered eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. This set is the matrix's eigenvalue spectrum quantifying the connectivity as well as characterizing the graph similarity. Generally, $\mathbf{L}(G)$ is always positive semi-definite such that the first eigenvalue $\lambda_1 = 0$ and those of higher orders $\lambda_{k>1} \in \Re^+$ [5,6]. According to the concept of "Spectral Graph Partitioning" that originates from the field of domain decomposition in parallel computing [7], the second (also known as the Fiedler) eigenvalue λ_2 represents mathematically the algebraic connectivity of the tree-graph G [5,6]. In relation to the RNA secondary structure, λ_2 measures the degree of compactness of the RNA topology at the coarsest scale [5,6]. RNA structures having similar values of λ_2 tend to be similar in topologies. Typically, the value of λ_2 increases monotonically with greater compactness in the RNA structure. Large values correspond to vertices of high degree that are in close proximity, while small values for more equally dispersed edge set. Maximum value of λ_2 is either 1 or 2 for an $n > 2$ perfectly connected star-shaped tree-graph or for $n = 2$ linear tree-graph, respectively [5,6].

1.3 RNAspectral Algorithm

The algorithm ***RNAspectral***(*S*) presents our strategy geared towards two tasks. Given a RNA secondary structure *S* described in a Vienna dot-bracket notation containing '.', '(', and ')' [8], it first abstract *S* at the coarsest-scale into a planar tree-graph representation. This transforms uniquely the RNA structural motifs (hairpin loops, internal loops, bulge loops, and multi-branch loops, as well as stems) into a network of vertices connected by incident edges. Next, it computes the Fidler eigenvalue λ_2 from the Laplacian matrix corresponding to the tree-graph.

RNAspectral(*S*) uses two primary functions in *Line* 1−2, whose pseudo-codes are described in the ***optimizeStruct***(*S*) and ***parseStruct***(*S*), respectively. The former returns *S′* and the latter returns the values for five global variables *totalpath, path, stems, ld, ls*, and *hs*. *Line* 3−4, sets the value of adjacency matrix **A** at row *path*[*i*] and column *path*[*i* + 1] to 1; 5−6, sets the value of degree matrix **D** at row *i* and column *i* to *ld*[*i*]; 7, computes the Laplacian matrix **L**; 8, the auxiliary function ***computeEigVals***(**L**) computes the eigenvalue spectrum using the well-established "Eigen-decomposition theorem" and $\det\left|\mathbf{L} - l\,\mathbf{I}\right| = 0$.

Algorithm: ***RNAspectral***(*S*)
Global Vars: *totalpath* ← 0, *path* ← ϕ, *stems* ← 0, *ld* ← ϕ, *ls* ← ϕ, *hs* ← ϕ.
1. **Vars**: *S′* ← ***optimizeStruct***(*S*), **A** ← ϕ, **D** ← ϕ, **L** ← ϕ.
2. (*totalpath, path, stems, ld, ls, hs*) ← ***parseStruct***(*S′*).
3. **For** *i* ← 1 : *totalpath*, **do**
4. **A**[*path*[*i*]][*path*[*i* + 1]] ← 1.
5. **For** *i* ← 1 : *stems* + 1, **do**
6. **D**[*i*][*i*] ← *ld*[*i*].
7. **L** ← **D** – **A**.
8. ***computeEigVals***(**L**).

In ***optimizeStruct***(*S*), it implements the pair of vertex-edge rules described in subsection 1.1. *Line* 1, vector *pt* contains the values returned by the auxiliary function ***makePBTable***(*S*), such that the *pt*[*i*] of nucleotide at position *i* has value of UNPAIRED when that nucleotide is unpaired or denotes the position of the base to which it is paired; 2−8, internal loops with only one pair of mismatches are identified and then paired; 9−12, stems with only one complementary pair are identified and then unpaired; 13−17, bulges having unpaired mono-nucleotide are deleted; 18, the resulting RNA structure *S′* is returned after applying the pair of vertex-edge rules.

Function: ***optimizeStruct***(*S*)
1. **Vars**: *L* ← ***len***(*S*), *pt* ← ***makePBTable***(*S*), *S′* ← *S*, *j* ← 1.
2. **For** *i* ← 1 : *L* − 1, **do**
3. **If** *pt*[*i*] = UNPAIRED, **then**
4. **If** ***min***(*pt*[*i* − 1], *pt*[*i* + 1]) = UNPAIRED, **then** continue.
5. **If** ***abs***(*pt*[*i* − 1] − *pt*[*i* + 1]) = 2, **then**
6. *pt*[*i*] ← ***max***(*pt*[*i* − 1], *pt*[*i* + 1]) − 1.
7. *pt*[*pt*[*i*]] ← *i*.
8. *S′*[*i*] ← '(', *S′*[*pt*[*i*]] ← ')'.
9. **If** *pt*[*i*] ≠ UNPAIRED, **then**
10. **If** *pt*[*i* − 1] = *pt*[*i* + 1], **then**
11. *S′*[*i*] ← *S′*[*pt*[*i*]] ← '.'.
12. *pt*[*pt*[*i*]] ← *pt*[*i*] ← UNPAIRED.
13. **For** *i* ← 1 : *L* − 2, **do**
14. **If** *pt*[*i*] = UNPAIRED, **then**
15. **If** ***abs***(*pt*[*i* − 1] − *pt*[*i* + 1]) = 1, **then** continue.
16. *S′*[*j*++] ← *S′*[*i*].
17. *S′*[*j*++] ← *S′*[*L* − 1], *S′*[*j*] ← ϕ.
18. **return** *S′*.

Function: ***makePBTable***(*S*)
1. **Vars**: *L* ← ***len***(*S*), *pt* ← ϕ, *stack* ← ϕ, *j* ← 0.
2. **Foreach** *S*[*i*] such that *i* ← 1 : *L* − 1, **do**
3. **case** '.', **do** *pt*[i] ← UNPAIRED.
4. **case** '(', **do** *stack*[*j*++] ← *i*.
5. **case** ')', **do** *pt*[*i*] ← *stack*[--*j*], *pt*[*pt*[*i*]] ← *i*.
6. **return** *pt*.

In ***parseStruct***(*S*), it implements the Eq. (1) and (2) described in subsection 1.2. *Line* 1, *S′* is a RNA secondary structure specified in an extended dot-bracket format with additional symbols '[', and ']', returned by the auxiliary function ***extStruct***(*S*), to track the onset of a helical stem-loop; 2−14 computes the Euclidean *path* transverse from the first to the final (*stems* + 1)$^{\text{th}}$ vertex, in the direction of 5' → 3' terminus; the size of vector *path* is stored in the variable *totalpath*. The size of each vertex and stem measured by the number of unpaired bases and number of pairs, respectively, are tracked by two variables *ls* and *hs*; the degree of each vertex is stored in the variable *ld*.

Function: ***parseStruct***(*S*)
1. **Vars**: *L* ← ***len***(*S*), *S′* ← ***extStruct***(*S*), *loop* ← ϕ, *lp* ← 0, *j* ← 0.
2. **Foreach** *S′*[*i*] such that *i* ← 1 : *L* − 1, **do**
3. **case** '.', **do** *ls*[*loop*[*lp*]]++.
4. **case** '[', **do**
5. *path*[*totalpath*++] ← *loop*[*lp*++],
6. *ld*[++*stems*] ← 1,
7. *loop*[*lp*] ← *stems*.
8. **case** ')', **do** *j*++.
9. **case** ']', **do**

10. $hs[loop[lp]] \leftarrow j + 1$,
11. $j \leftarrow 0$,
12. $path[totalpath{+}{+}] \leftarrow loop[lp]$,
13. $ld[loop[{-}{-}lp]]{+}{+}$.
14. $path[totalpath] \leftarrow 0$.

Function: ***extStruct***(S)
1. **Vars**: $L \leftarrow$ ***len***(S), $mp \leftarrow \phi$, $S' \leftarrow S$, $o \leftarrow 0$, $j \leftarrow 0$.
2. **Foreach** $S'[i]$ such that $i \leftarrow 1 : L - 1$, **do**
3. **case** '(', **do** $mp[{+}{+}o] \leftarrow i$.
4. **case** ')', **do**
5. $j \leftarrow i$.
6. **While** $S'[j + 1] = =$ ')' $\wedge\ mp[o - 1] = = mp[o] - 1$, **do**
7. $j{+}{+}$, $o{-}{-}$.
8. $S'[j] \leftarrow$ ']', $i \leftarrow j$, $S'[mp[o{-}{-}]] \leftarrow$ '['.
9. **return** S'.

1.4 Methodology

Since its introduction in 2003 [9,10], "Spectral Graph Partitioning" has been extensively applied to a variety of bioinformatics problems: the prediction of multiple mutation to disrupt motifs in riboswitches [5], the prediction of RNA conformational switch by mutation [11], the search and analysis of RNA secondary structures [6], the classification of RNA coarse-grained tree-graph structures [3,4], and lastly for systematically partitioning complex RNA structures into simpler fragments with maximal decoupling between them [10]. These applications underscore the potential of "Spectral Graph Partitioning" as an invaluable computational tool to elucidate the topological patterns hidden in the post-genomic sequences and to offer a tremendous opportunity for an enhanced understanding of both functional and structural genomics.

"RNA Matrix Computer Program" [3,4] is the pioneering and only implementation of "Spectral Graph Partitioning" analysis on RNA structural folding. It is available online and provides a user-friendly interface for uploading a "ct file" produced by Zuker's mfold prediction server [12,13] or equivalent. As an attempt to address the high-throughput demands of our in-house projects, we have designed RNAspectral from scratch based on the mathematical formalisms gathered from literature, and iteratively validated against the 'reference' results of "RNA Matrix Computer Program" [3,4].

RNAspectral is an efficient and rapid algorithm, implemented in ANSI C programming language using the development platform Intel Pentium M 2.0 GHz, and 1.0 GB RAM; Cygwin 1.5.19-Windows XP. It provides a user-friendly command-line interface and four user-adjustable parameters: *-v1*, to enable the level of verbosity for obtaining output identical to that of "RNA Matrix Computer Program" [3,4]; *-v2*, to enable detailed debugging and further analysis into RNAspectral internalities; *--noopt*, to disable the pair of vertex-edge rules; *--monitor*, to monitor the execution time. Together, these options and functionalities allow the inexperienced user to integrate the information from "Spectral Graph Partitioning" analysis such as the second eigenvalue λ_2 and the number of vertices as part of their experimental methodologies, in an intuitive manner.

A typical experimental setup using RNAfold and RNAspectral in an automated manner is outlined in Fig. S2A. Given a primary RNA sequence described in FASTA format, (Step **A**) its optimal secondary structure is predicted using RNAfold [8]. The output of RNAfold is a FASTA-like format appended with the optimal structure in Vienna dot-bracket notation with the base pairs and unpaired bases represented by brackets '( )' and dots '.' [8], respectively and the minimum free energy of folding (MFE). In this example, the RNA secondary structure predicted by RNAfold has two hairpin loops, 5' and 3' termini, two internal loops, one bulge loop, and one multi-branch loop - all of these stabilized by six stems. (Step **B**) This is read by RNAspectral that converts the structure in bracket notation into a planar tree-graph consisting of seven arbitrarily labeled vertices (•) connected by six unweighted edges (—). (Step **C**) RNAspectral computes the seven by seven Laplacian matrix and the eigenvalue spectrum. (Step **D**) The output of RNAspectral is described in a tab-delimited ASCII flat format for convenient import into numerical processing applications such as Mathworks® Matlab™ and Microsoft® Excel™. The labeled header shows the following rows of columnated values corresponding to the identifier (**ID** starts at 1 and increases monotically), minimum free energy of folding (**MFE** in kcal/mol), length of sequence (**Len** in nucleotides), number of vertices (**Ver**), number of stems (**Stems**), number of junctions (**Junct**, >2 stems), number of endpoints (**Endpts**, 1 stem), number of midpoints (**Midpts**, 2 stems), and the second eigenvalue λ_2 (**SecEigen**).

(Fig. S2B) The benchmarking platform was an AMD Opteron Processor 850 2.4 GHz and 1.5 GB RAM; GNU compiler v3.4.5 on Linux 2.6.9-5. We computed the average speed of RNAspectral by running it five times on 6,656 sets of 10^4 random RNA sequences. The random sequences were synthesized from each of the 6,656 sequences (each has 113.451 ± 0.803 nucleotides) gathered from miRBase 7.1 [14] and Rfam 7.0 [15]. RNAspectral requires at most ~7.0 seconds or mean 427.8 milliseconds for processing the entire dataset.

REFERENCES

1. Tinoco J, I, Bustamante C (1999) How RNA folds. *Journal of Molecular Biology* **293:** 271-281.
2. Moulton V et al. (2000) Metrics on RNA Secondary Structures. *Journal of Computational Biology* **7:** 277-292.
3. Fera D et al. (2004) RAG: RNA-As-Graphs web resource. *BMC Bioinformatics* **5:** 88.
4. Gan HH et al. (2004) RAG: RNA-As-Graphs database--concepts, analysis, and features. *Bioinformatics* **20:** 1285-1291.
5. Barash D (2003) Deleterious mutation prediction in the secondary structure of RNAs. *Nucl Acids Res* **31:** 6578-6584.
6. Barash D (2004) Spectral Decomposition for the Search and Analysis of RNA Secondary Structure. *Journal of Computational Biology* **11:** 1169-1174.
7. Alex P, Horst DS, Kan-Pu L (1990) Partitioning sparse matrices with eigenvectors of graphs. *SIAM J Matrix Anal Appl* **11:** 430-452.
8. Hofacker IL (2003) Vienna RNA secondary structure server. *Nucl Acids Res* **31:** 3429-3431.
9. Barash D (2003) Spectral decomposition of the Laplacian matrix applied to RNA folding prediction. CSB 2003 602-03.
10. Gan HH, Pasquali S, Schlick T (2003) Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucl Acids Res* **31:** 2926-2943.
11. Barash D (2004) Second eigenvalue of the Laplacian matrix for predicting RNA conformational switch by mutation. *Bioinformatics* **20:** 1861-1869.
12. Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* **9:** 133-148.
13. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucl Acids Res* **31:** 3406-3415.
14. Griffiths-Jones S (2004) The microRNA Registry. *Nucl Acids Res* **32:** D109-D111.
15. Griffiths-Jones S et al. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucl Acids Res* **33:** D121-D124.
16. Sudarsan N, Barrick JE, Breaker RR (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* **9:** 644-647.
17. Griffiths-Jones S et al. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucl Acids Res* **34:** D140-D144.
18. Freyhult E, Gardner P, Moulton V (2005) A comparison of RNA folding measures. *BMC Bioinformatics* **6:** 241.

19. Xue C et al. (2005) Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC Bioinformatics* **6:** 310.

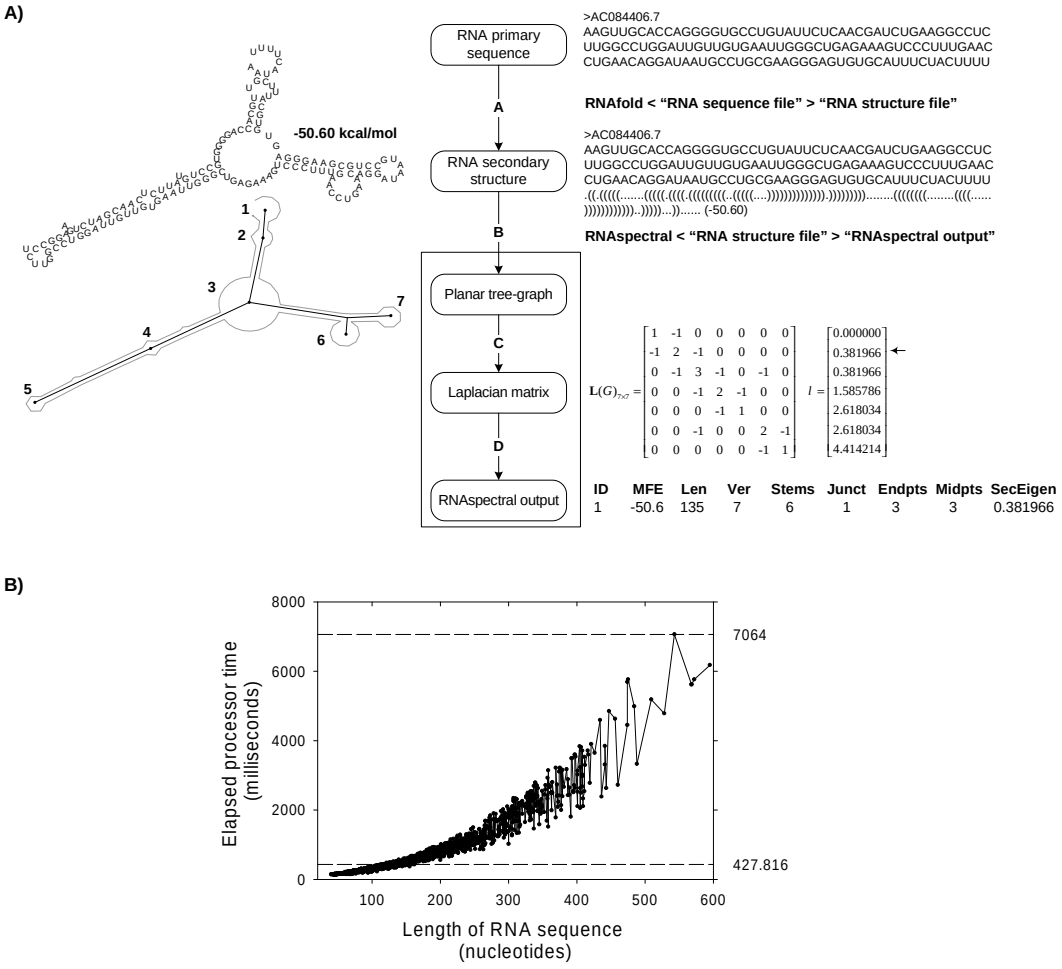

Fig. S2. A) Typical workflow using RNAspectral for "Spectral Graph Partitioning" analysis on RNA structures. Sequence of THI element (AC084406.7; thiamine pyrophosphate riboswitch) [16] was extracted from Rfam 7.0 [15]. ←, computed second eigenvalue λ_2 shows the same results as "RNA Matrix Computer Program" [3,4]; bold, Unix commands. **B**) Average speed performance of RNAspectral. Unlike the actual wall-clock time, elapsed processor time excludes time spent queuing for free I/O or waiting for other processes to complete execution.

Table S1. Statistical comparison between 2241 non-redundant *pre-miRs* [17], 12387 non-redundant ncRNA**s** [15], 31 mRNA sequences [18], and 8494 pseudo hairpins [19] based on nine metrics − *Length*, $MFEI_2$, $MFEI_1$, *%G+C*, *P*(*S*), *MFE*(**s**), *Q*(**s**), *D*(**s**), and *F*(*S*).

Datasets	*Counts*	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*
Arthropoda	171	88.6901 ± 0.8213	-0.0645 ± 0.0016	-0.0089 ± 0.0001	43.3811 ± 0.4752	0.3488 ± 0.0023	-0.3824 ± 0.0050	0.1067 ± 0.0047	0.0403 ± 0.0016	0.2059 ± 0.0067
Nematoda	189	99.0212 ± 0.6723	-0.0556 ± 0.0015	-0.0086 ± 0.0001	44.5725 ± 0.4641	0.3411 ± 0.0025	-0.3831 ± 0.0056	0.1075 ± 0.0059	0.0398 ± 0.0019	0.1577 ± 0.0050
Vertebrata	1203	90.4522 ± 0.4164	-0.0761 ± 0.0013	-0.0091 ± 0.0001	48.3079 ± 0.2504	0.3518 ± 0.0009	-0.4308 ± 0.0025	0.1161 ± 0.0025	0.0431 ± 0.0009	0.2197 ± 0.0042
Viridiplantae	606	137.9175 ± 2.0309	-0.0539 ± 0.0010	-0.0096 ± 0.0001	46.6719 ± 0.3513	0.3545 ± 0.0013	-0.4456 ± 0.0038	0.1424 ± 0.0036	0.0502 ± 0.0011	0.1251 ± 0.0033
Viruses	72	78.8750 ± 1.4665	-0.0780 ± 0.0032	-0.0087 ± 0.0002	53.5111 ± 0.9219	0.3619 ± 0.0029	-0.4615 ± 0.0097	0.0893 ± 0.0051	0.0352 ± 0.0020	0.2059 ± 0.0114
Cis-reg	4002	90.7511 ± 0.8069	-0.0793 ± 0.0017	-0.0065 ± 0.0000	48.9672 ± 0.1188	0.2905 ± 0.0008	-0.3233 ± 0.0017	0.2124 ± 0.0021	0.0689 ± 0.0006	0.3871 ± 0.0064
Cis-reg|frameshift	808	53.2599 ± 0.2543	-0.2210 ± 0.0021	-0.0104 ± 0.0000	46.4785 ± 0.1477	0.3382 ± 0.0010	-0.4814 ± 0.0023	0.1396 ± 0.0024	0.0552 ± 0.0009	0.8865 ± 0.0079
Cis-reg|IRES	1201	276.0841 ± 2.4342	-0.0192 ± 0.0002	-0.0065 ± 0.0000	57.5340 ± 0.1745	0.3039 ± 0.0006	-0.3757 ± 0.0013	0.3702 ± 0.0034	0.1156 ± 0.0010	0.0442 ± 0.0013
Cis-reg|riboswitch	917	138.6358 ± 1.4673	-0.0381 ± 0.0005	-0.0064 ± 0.0000	50.5054 ± 0.3381	0.2877 ± 0.0010	-0.3223 ± 0.0026	0.2515 ± 0.0041	0.0826 ± 0.0012	0.1960 ± 0.0042
Cis-reg|thermoregulator	21	127.0476 ± 4.0447	-0.0330 ± 0.0047	-0.0061 ± 0.0002	42.6490 ± 3.2009	0.2955 ± 0.0075	-0.2713 ± 0.0301	0.2935 ± 0.0269	0.0956 ± 0.0080	0.1312 ± 0.0138
Gene	480	222.2708 ± 5.8445	-0.0372 ± 0.0012	-0.0074 ± 0.0000	51.6146 ± 0.5262	0.3109 ± 0.0012	-0.3808 ± 0.0046	0.2435 ± 0.0060	0.0794 ± 0.0018	0.1258 ± 0.0058
Gene|antisense	147	86.0476 ± 0.8681	-0.0811 ± 0.0030	-0.0083 ± 0.0001	41.7778 ± 0.8673	0.3106 ± 0.0034	-0.3414 ± 0.0076	0.1336 ± 0.0061	0.0468 ± 0.0020	0.3734 ± 0.0133
Gene|ribozyme	561	242.0428 ± 5.4441	-0.0406 ± 0.0017	-0.0070 ± 0.0000	54.4837 ± 0.3930	0.3000 ± 0.0011	-0.3811 ± 0.0040	0.2704 ± 0.0053	0.0863 ± 0.0016	0.2335 ± 0.0145
Gene|rRNA	1010	244.3208 ± 5.8418	-0.0295 ± 0.0005	-0.0066 ± 0.0000	53.8479 ± 0.2508	0.3022 ± 0.0008	-0.3545 ± 0.0023	0.2870 ± 0.0043	0.0921 ± 0.0012	0.0933 ± 0.0020
Gene|snRNA	28	62.0357 ± 0.7024	-0.0764 ± 0.0088	-0.0061 ± 0.0003	41.6782 ± 1.2105	0.2803 ± 0.0064	-0.2631 ± 0.0187	0.2305 ± 0.0260	0.0741 ± 0.0074	0.5372 ± 0.0415
Gene|snRNA|guide|CD-box	1050	91.5867 ± 1.0464	-0.0379 ± 0.0004	-0.0053 ± 0.0000	42.3681 ± 0.2301	0.2764 ± 0.0013	-0.2265 ± 0.0022	0.3174 ± 0.0041	0.1012 ± 0.0012	0.2772 ± 0.0058
Gene|snRNA|guide|HACA-box	419	139.3675 ± 1.2446	-0.0348 ± 0.0005	-0.0068 ± 0.0001	46.3048 ± 0.3160	0.2929 ± 0.0013	-0.3125 ± 0.0029	0.2383 ± 0.0068	0.0783 ± 0.0021	0.1194 ± 0.0028
Gene|snRNA|splicing	250	157.1200 ± 4.4708	-0.0341 ± 0.0006	-0.0068 ± 0.0001	47.6933 ± 0.3731	0.2898 ± 0.0021	-0.3251 ± 0.0042	0.2399 ± 0.0076	0.0781 ± 0.0023	0.1470 ± 0.0043
Gene|sRNA	233	145.6524 ± 4.5117	-0.0432 ± 0.0016	-0.0066 ± 0.0001	46.3963 ± 0.3513	0.2815 ± 0.0024	-0.3036 ± 0.0041	0.2371 ± 0.0077	0.0745 ± 0.0023	0.2531 ± 0.0170
Gene|tRNA	1114	73.4354 ± 0.1529	-0.0676 ± 0.0007	-0.0064 ± 0.0000	48.2725 ± 0.3541	0.2975 ± 0.0010	-0.3138 ± 0.0029	0.2488 ± 0.0035	0.0831 ± 0.0011	0.5333 ± 0.0093
Intron	146	134.4384 ± 8.6225	-0.0604 ± 0.0029	-0.0080 ± 0.0001	44.7871 ± 0.8350	0.3204 ± 0.0024	-0.3551 ± 0.0081	0.1802 ± 0.0089	0.0620 ± 0.0026	0.2200 ± 0.0107
mRNAs	31	332.3226 ± 16.3064	-0.0132 ± 0.0006	-0.0061 ± 0.0001	50.4626 ± 1.4654	0.2881 ± 0.0045	-0.3087 ± 0.0131	0.3828 ± 0.0175	0.1192 ± 0.0049	0.0391 ± 0.0059
Pseudo hairpins	8494	84.7020 ± 0.1268	-0.0476 ± 0.0002	-0.0054 ± 0.0000	56.1466 ± 0.1108	0.2874 ± 0.0003	-0.3070 ± 0.0009	0.3185 ± 0.0016	0.1048 ± 0.0005	0.1818 ± 0.0008

(*Counts*) Number of sequences being investigated. Values are stated as mean ± standard error.

Table S2. Statistical comparison between 2241 non-redundant *pre-miRs* [17], 12387 non-redundant ncRNA**s** [15], 31 mRNA sequences [18], and 8494 pseudo hairpins [19] based on *zG*, *zQ*, *zD*, *zP*, and *zF* (normalized forms of *MFE*(**s**), *Q*(**s**), *D*(**s**), *P*(*S*), and *F*(*S*) using the four sequence randomization algorithms).

		zG				*zQ*				*zD*			
Datasets	*Counts*	*MS*	*DS*	*ZM*	*FM*	*MS*	*DS*	*ZM*	*FM*	*MS*	*DS*	*ZM*	*FM*
Arthropoda	171	-4.8894 ± 0.1127	-4.8985 ± 0.1177	-3.5250 ± 0.0839	-3.3032 ± 0.0818	-1.7166 ± 0.0321	-1.6873 ± 0.0316	-1.7259 ± 0.0320	-1.7067 ± 0.0311	-1.6782 ± 0.0377	-1.6526 ± 0.0365	-1.6706 ± 0.0379	-1.6544 ± 0.0364
Nematoda	189	-4.9930 ± 0.1457	-5.0228 ± 0.1443	-3.4481 ± 0.1003	-3.2885 ± 0.0927	-1.7836 ± 0.0399	-1.7548 ± 0.0394	-1.7936 ± 0.0401	-1.7797 ± 0.0390	-1.7598 ± 0.0440	-1.7358 ± 0.0432	-1.7523 ± 0.0443	-1.7433 ± 0.0432
Vertebrata	1203	-5.2058 ± 0.0645	-4.7608 ± 0.0642	-3.6575 ± 0.0463	-3.2317 ± 0.0426	-1.6090 ± 0.0170	-1.5465 ± 0.0164	-1.6252 ± 0.0170	-1.5779 ± 0.0164	-1.5755 ± 0.0196	-1.5209 ± 0.0187	-1.5762 ± 0.0198	-1.5368 ± 0.0188
Viridiplantae	606	-6.9286 ± 0.1033	-6.4395 ± 0.1037	-4.5333 ± 0.0718	-4.1132 ± 0.0693	-1.6602 ± 0.0248	-1.5957 ± 0.0243	-1.6725 ± 0.0248	-1.6211 ± 0.0242	-1.6440 ± 0.0276	-1.5879 ± 0.0267	-1.6422 ± 0.0277	-1.5982 ± 0.0267
Viruses	72	-4.7038 ± 0.1952	-4.5972 ± 0.1908	-3.2593 ± 0.1325	-3.0913 ± 0.1280	-1.6475 ± 0.0414	-1.6214 ± 0.0403	-1.6722 ± 0.0416	-1.6524 ± 0.0405	-1.6088 ± 0.0495	-1.5848 ± 0.0481	-1.6191 ± 0.0498	-1.6016 ± 0.0486
Cis-reg	4002	-2.6887 ± 0.0308	-2.3364 ± 0.0280	-1.9053 ± 0.0203	-1.5172 ± 0.0172	-0.8439 ± 0.0142	-0.7928 ± 0.0139	-0.8336 ± 0.0142	-0.7878 ± 0.0140	-0.8206 ± 0.0147	-0.7788 ± 0.0143	-0.7851 ± 0.0148	-0.7452 ± 0.0145
Cis-reg|frameshift	808	-5.6222 ± 0.0477	-3.7443 ± 0.0357	-4.4470 ± 0.0359	-2.3964 ± 0.0200	-1.1436 ± 0.0158	-1.1970 ± 0.0155	-1.1579 ± 0.0160	-1.1044 ± 0.0146	-0.9865 ± 0.0192	-1.0716 ± 0.0187	-0.9768 ± 0.0197	-0.9303 ± 0.0181
Cis-reg|IRES	1201	-0.7674 ± 0.0353	-1.0895 ± 0.0293	-0.5451 ± 0.0192	-0.6451 ± 0.0149	-0.1924 ± 0.0250	-0.2063 ± 0.0252	-0.2027 ± 0.0250	-0.2300 ± 0.0251	-0.2134 ± 0.0256	-0.2208 ± 0.0259	-0.2121 ± 0.0257	-0.2296 ± 0.0260
Cis-reg|riboswitch	917	-1.5838 ± 0.0452	-1.4806 ± 0.0446	-1.1569 ± 0.0282	-1.0231 ± 0.0261	-0.8469 ± 0.0293	-0.8163 ± 0.0294	-0.8585 ± 0.0294	-0.8513 ± 0.0293	-0.8139 ± 0.0309	-0.7884 ± 0.0309	-0.8086 ± 0.0312	-0.8030 ± 0.0309
Cis-reg|thermoregulator	21	-1.0551 ± 0.2108	-1.0754 ± 0.2211	-0.8443 ± 0.1263	-0.7904 ± 0.1349	-0.5827 ± 0.1521	-0.5791 ± 0.1523	-0.5961 ± 0.1533	-0.6004 ± 0.1542	-0.4561 ± 0.1723	-0.4496 ± 0.1725	-0.4511 ± 0.1754	-0.4460 ± 0.1753
Gene	480	-2.9702 ± 0.0842	-2.8100 ± 0.0851	-1.9501 ± 0.0521	-1.7827 ± 0.0510	-1.0260 ± 0.0391	-1.0098 ± 0.0394	-1.0379 ± 0.0392	-1.0335 ± 0.0395	-1.0127 ± 0.0410	-1.0017 ± 0.0412	-1.0122 ± 0.0412	-1.0082 ± 0.0415
Gene|antisense	147	-4.0852 ± 0.1258	-4.0585 ± 0.1283	-2.9472 ± 0.0900	-2.6765 ± 0.0829	-1.5501 ± 0.0404	-1.5317 ± 0.0409	-1.5473 ± 0.0399	-1.5387 ± 0.0403	-1.5408 ± 0.0466	-1.5220 ± 0.0469	-1.5117 ± 0.0456	-1.4990 ± 0.0460
Gene|ribozyme	561	-3.0964 ± 0.0704	-2.7927 ± 0.0706	-1.9182 ± 0.0392	-1.6665 ± 0.0376	-0.7666 ± 0.0347	-0.7312 ± 0.0346	-0.7737 ± 0.0348	-0.7588 ± 0.0346	-0.7567 ± 0.0361	-0.7347 ± 0.0355	-0.7492 ± 0.0364	-0.7450 ± 0.0357
Gene|rRNA	1010	-2.0655 ± 0.0551	-2.0126 ± 0.0523	-1.3108 ± 0.0298	-1.2051 ± 0.0268	-0.6742 ± 0.0296	-0.6618 ± 0.0296	-0.6858 ± 0.0298	-0.6943 ± 0.0295	-0.6424 ± 0.0302	-0.6329 ± 0.0301	-0.6406 ± 0.0305	-0.6491 ± 0.0302
Gene|snRNA	28	-2.0909 ± 0.2613	-1.3712 ± 0.3083	-1.6055 ± 0.1806	-1.0729 ± 0.2076	-0.6335 ± 0.1771	-0.5674 ± 0.1695	-0.6180 ± 0.1789	-0.5995 ± 0.1699	-0.6270 ± 0.1735	-0.6108 ± 0.1597	-0.5832 ± 0.1759	-0.6090 ± 0.1609
Gene|snRNA|guide|CD-box	1050	-0.8113 ± 0.0397	-0.7089 ± 0.0360	-0.7209 ± 0.0270	-0.6465 ± 0.0244	-0.2189 ± 0.0292	-0.2236 ± 0.0286	-0.2146 ± 0.0295	-0.2497 ± 0.0286	-0.1810 ± 0.0298	-0.1952 ± 0.0291	-0.1512 ± 0.0302	-0.1886 ± 0.0294
Gene|snRNA|guide|HACA-box	419	-2.3694 ± 0.0745	-1.7490 ± 0.0780	-1.6445 ± 0.0499	-1.2434 ± 0.0489	-0.9913 ± 0.0497	-0.9265 ± 0.0494	-0.9997 ± 0.0499	-0.9567 ± 0.0493	-0.9621 ± 0.0532	-0.9169 ± 0.0519	-0.9546 ± 0.0537	-0.9275 ± 0.0523
Gene|snRNA|splicing	250	-2.6848 ± 0.1171	-2.4767 ± 0.1097	-1.7286 ± 0.0667	-1.4502 ± 0.0567	-1.0036 ± 0.0604	-0.9687 ± 0.0610	-1.0112 ± 0.0606	-0.9999 ± 0.0601	-1.0018 ± 0.0632	-0.9681 ± 0.0636	-0.9933 ± 0.0636	-0.9783 ± 0.0631
Gene|sRNA	233	-2.7470 ± 0.1222	-2.7672 ± 0.1234	-1.7773 ± 0.0712	-1.6417 ± 0.0664	-0.9773 ± 0.0589	-0.9675 ± 0.0589	-0.9771 ± 0.0592	-0.9903 ± 0.0584	-1.0182 ± 0.0608	-1.0073 ± 0.0612	-0.9991 ± 0.0611	-1.0064 ± 0.0607
Gene|tRNA	1114	-1.8663 ± 0.0281	-1.7570 ± 0.0289	-1.4794 ± 0.0193	-1.3739 ± 0.0189	-0.5740 ± 0.0237	-0.5524 ± 0.0239	-0.5770 ± 0.0238	-0.5804 ± 0.0238	-0.5223 ± 0.0257	-0.5109 ± 0.0257	-0.5019 ± 0.0260	-0.5050 ± 0.0260
Intron	146	-3.7603 ± 0.1402	-3.6841 ± 0.1513	-2.7426 ± 0.0976	-2.5026 ± 0.0982	-1.3073 ± 0.0531	-1.2842 ± 0.0534	-1.3177 ± 0.0533	-1.3065 ± 0.0530	-1.2483 ± 0.0558	-1.2290 ± 0.0559	-1.2424 ± 0.0564	-1.2335 ± 0.0560
mRNAs	31	-0.7223 ± 0.2089	0.1021 ± 0.1625	-0.4770 ± 0.1098	-0.0830 ± 0.0845	-0.1894 ± 0.1503	-0.1434 ± 0.1486	-0.1907 ± 0.1504	-0.1680 ± 0.1487	-0.1126 ± 0.1518	-0.0994 ± 0.1492	-0.1017 ± 0.1516	-0.1055 ± 0.1496
Pseudo hairpins	8494	-0.6493 ± 0.0121	-0.2347 ± 0.0114	-0.5606 ± 0.0073	-0.3373 ± 0.0067	-0.1058 ± 0.0113	-0.0756 ± 0.0112	-0.1052 ± 0.0114	-0.1044 ± 0.0112	-0.0444 ± 0.0117	-0.0385 ± 0.0114	-0.0208 ± 0.0118	-0.0364 ± 0.0114

(*Counts*) Number of sequences being investigated. Values are stated as mean ± standard error. MS, Mononucleotide Shuffling; DS, Dinucleotide Shuffling; ZM, Zero-order Markov Model; FM, First-order Markov Model.

		zP				*zF*			
Datasets	*Counts*	*MS*	*DS*	*ZM*	*FM*	*MS*	*DS*	*ZM*	*FM*
Arthropoda	171	2.4736 ± 0.0653	2.4309 ± 0.0686	2.2904 ± 0.0560	2.2112 ± 0.0579	0.7107 ± 0.0912	0.6432 ± 0.0910	0.5025 ± 0.0791	0.4001 ± 0.0736
Nematoda	189	2.4392 ± 0.0807	2.4022 ± 0.0819	2.1992 ± 0.0673	2.1076 ± 0.0661	1.2007 ± 0.0675	1.1643 ± 0.0660	1.0738 ± 0.0651	1.0329 ± 0.0628
Vertebrata	1203	2.4911 ± 0.0287	2.3364 ± 0.0301	2.3065 ± 0.0246	2.1516 ± 0.0251	0.1902 ± 0.0340	0.1859 ± 0.0333	0.1359 ± 0.0330	0.1427 ± 0.0323
Viridiplantae	606	2.9329 ± 0.0449	2.7807 ± 0.0461	2.6133 ± 0.0376	2.4634 ± 0.0383	0.3538 ± 0.1870	0.5419 ± 0.2134	0.1306 ± 0.1470	0.2984 ± 0.1693
Viruses	72	2.6924 ± 0.0921	2.6297 ± 0.0922	2.4721 ± 0.0760	2.3915 ± 0.0754	-0.0844 ± 0.0335	0.0750 ± 0.0351	-0.1823 ± 0.0289	-0.0562 ± 0.0295
Cis-reg	4002	1.3687 ± 0.0197	1.2727 ± 0.0185	1.2527 ± 0.0160	1.1631 ± 0.0141	-0.1601 ± 0.0469	-0.1078 ± 0.0479	-0.2165 ± 0.0434	-0.1643 ± 0.0441
Cis-reg|frameshift	808	1.8580 ± 0.0237	1.4936 ± 0.0218	1.8881 ± 0.0207	1.4944 ± 0.0157	0.7519 ± 0.0732	0.6479 ± 0.0723	0.6347 ± 0.0692	0.5327 ± 0.0682
Cis-reg|IRES	1201	-0.0329 ± 0.0298	0.1285 ± 0.0291	0.1392 ± 0.0241	0.2594 ± 0.0231	1.1140 ± 0.1503	1.0258 ± 0.1485	0.8554 ± 0.1252	0.7597 ± 0.1218
Cis-reg|riboswitch	917	0.4080 ± 0.0365	0.3818 ± 0.0362	0.5064 ± 0.0292	0.4811 ± 0.0285	1.5069 ± 0.0682	1.5169 ± 0.0692	1.2329 ± 0.0614	1.1410 ± 0.0600
Cis-reg|thermoregulator	21	0.7628 ± 0.2207	0.8579 ± 0.2135	0.8156 ± 0.1827	0.9010 ± 0.1667	0.1460 ± 0.0809	0.0892 ± 0.0815	0.0623 ± 0.0758	0.0149 ± 0.0748
Gene	480	0.8078 ± 0.0495	0.8192 ± 0.0495	0.8420 ± 0.0414	0.8754 ± 0.0403	-0.0795 ± 0.1732	0.0456 ± 0.1812	-0.1270 ± 0.1624	-0.0042 ± 0.1698
Gene|antisense	147	1.4284 ± 0.0751	1.4366 ± 0.0715	1.3817 ± 0.0618	1.3731 ± 0.0582	-0.5938 ± 0.0065	-0.5623 ± 0.0067	-0.6124 ± 0.0057	-0.5726 ± 0.0058
Gene|ribozyme	561	0.8520 ± 0.0431	0.7607 ± 0.0440	0.8343 ± 0.0342	0.7654 ± 0.0337	0.7107 ± 0.0912	0.6432 ± 0.0910	0.5025 ± 0.0791	0.4001 ± 0.0736
Gene|rRNA	1010	0.8805 ± 0.0330	0.8847 ± 0.0329	0.8612 ± 0.0256	0.8387 ± 0.0252	1.2007 ± 0.0675	1.1643 ± 0.0660	1.0738 ± 0.0651	1.0329 ± 0.0628
Gene|snRNA	28	0.8315 ± 0.1820	0.4112 ± 0.1759	0.8631 ± 0.1418	0.5505 ± 0.1294	0.1902 ± 0.0340	0.1859 ± 0.0333	0.1359 ± 0.0330	0.1427 ± 0.0323
Gene|snRNA|guide|CD-box	1050	0.5020 ± 0.0334	0.4050 ± 0.0335	0.6217 ± 0.0278	0.5643 ± 0.0274	0.3538 ± 0.1870	0.5419 ± 0.2134	0.1306 ± 0.1470	0.2984 ± 0.1693
Gene|snRNA|guide|HACA-box	419	0.5749 ± 0.0454	0.3727 ± 0.0479	0.6545 ± 0.0377	0.5069 ± 0.0385	-0.0844 ± 0.0335	0.0750 ± 0.0351	-0.1823 ± 0.0289	-0.0562 ± 0.0295
Gene|snRNA|splicing	250	0.4583 ± 0.0708	0.5303 ± 0.0692	0.5381 ± 0.0574	0.5607 ± 0.0546	-0.1601 ± 0.0469	-0.1078 ± 0.0479	-0.2165 ± 0.0434	-0.1643 ± 0.0441
Gene|sRNA	233	0.7139 ± 0.0700	0.7771 ± 0.0701	0.7326 ± 0.0546	0.7518 ± 0.0529	0.7519 ± 0.0732	0.6479 ± 0.0723	0.6347 ± 0.0692	0.5327 ± 0.0682
Gene|tRNA	1114	0.8293 ± 0.0281	0.8075 ± 0.0281	0.9282 ± 0.0235	0.9367 ± 0.0225	1.1140 ± 0.1503	1.0258 ± 0.1485	0.8554 ± 0.1252	0.7597 ± 0.1218
Intron	146	1.6381 ± 0.0778	1.6300 ± 0.0780	1.5693 ± 0.0661	1.5258 ± 0.0637	1.5069 ± 0.0682	1.5169 ± 0.0692	1.2329 ± 0.0614	1.1410 ± 0.0600
mRNAs	31	0.3438 ± 0.2004	-0.0620 ± 0.1970	0.3700 ± 0.1576	0.0849 ± 0.1532	0.1460 ± 0.0809	0.0892 ± 0.0815	0.0623 ± 0.0758	0.0149 ± 0.0748
Pseudo hairpins	8494	0.5399 ± 0.0103	0.3444 ± 0.0105	0.6197 ± 0.0080	0.4970 ± 0.0079	-0.0795 ± 0.1732	0.0456 ± 0.1812	-0.1270 ± 0.1624	-0.0042 ± 0.1698

(*Counts*) Number of sequences being investigated. Values are stated as mean ± standard error. MS, Mononucleotide Shuffling; DS, Dinucleotide Shuffling; ZM, Zero-order Markov Model; FM, First-order Markov Model.

Table S3. The correlation coefficients, 95th percentile, and *p*-values for 2241 non-redundant *pre-miRs* [17]. Three tables of pearson correlation coefficients C_p, spearman-rank C_s (ranks-based) and Kendall's C_k (relative ranks-based) are shown for each of the four sequence randomization algorithms.

Mononucleotide Shuffling

$C_p(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.3777	-0.0366	-0.0784	-0.0567	0.0394	0.2737	0.2424	-0.4389	-0.2470	0.0148	0.0085	0.0988	-0.0932
$MFEI_2$	6.76E-77	-0.0296	0.5484	-0.0535	-0.2937	0.5478	0.3374	0.3401	-0.8925	0.4418	0.2240	0.2366	-0.2070	-0.8288
$MFEI_1$	8.36E-02	3.76E-176	-0.0064	0.3589	-0.5960	0.5644	0.4323	0.4228	-0.4084	**0.9192**	0.5042	0.4885	-0.5936	-0.4453
%G+C	2.02E-04	1.14E-02	4.28E-69	62.3790	0.0701	-0.5437	0.0166	0.0364	-0.1909	0.2596	0.2028	0.1884	-0.0601	-0.0648
P(S)	7.25E-03	7.98E-46	1.30E-215	8.91E-04	0.4000	-0.6030	-0.3244	-0.2649	0.0515	-0.5436	-0.2878	-0.2377	**0.9013**	0.0538
MFE(s)	6.25E-02	1.11E-175	1.01E-188	1.23E-172	4.89E-222	-0.2937	0.3972	0.3745	-0.1964	0.6065	0.2944	0.2929	-0.4934	-0.3448
Q(s)	8.69E-40	8.85E-61	1.05E-102	4.33E-01	4.43E-56	1.40E-85	0.2885	**0.9829**	-0.2230	0.4257	**0.9444**	**0.9290**	-0.3441	-0.1315
D(s)	2.44E-31	8.42E-62	7.14E-98	8.48E-02	2.70E-37	1.57E-75	0.00E+00	0.0984	-0.2400	0.4251	**0.9396**	**0.9545**	-0.2971	-0.1620
F(S)	3.50E-106	0.00E+00	8.37E-91	7.86E-20	1.47E-02	6.50E-21	1.17E-26	1.01E-30	0.3820	-0.2594	-0.1319	-0.1515	0.0006	0.8292
zG	1.68E-32	1.06E-107	0.00E+00	7.70E-36	1.44E-172	2.69E-225	2.33E-99	4.94E-99	8.88E-36	-1.8302	0.5474	0.5325	-0.6583	-0.4068
zQ	4.85E-01	6.82E-27	7.81E-145	3.20E-22	5.17E-44	4.62E-46	0.00E+00	0.00E+00	3.68E-10	1.99E-175	-0.5625	**0.9844**	-0.3876	-0.1195
zD	6.86E-01	6.85E-30	9.42E-135	2.42E-19	3.70E-30	1.37E-45	0.00E+00	0.00E+00	5.62E-13	2.34E-164	0.00E+00	-0.3737	-0.3316	-0.1508
zP	2.78E-06	4.15E-23	1.87E-213	4.41E-03	0.00E+00	7.75E-138	2.61E-63	6.64E-47	9.77E-01	1.52E-278	3.09E-81	1.13E-58	4.2017	0.0269
zF	9.80E-06	0.00E+00	1.38E-109	2.16E-03	1.09E-02	1.37E-63	4.19E-10	1.21E-14	0.00E+00	4.47E-90	1.38E-08	7.24E-13	2.03E-01	1.3094

$C_s(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.4177	0.0087	-0.0162	-0.0836	0.0175	0.1887	0.1679	-0.5274	-0.1333	-0.0281	-0.0258	0.0209	-0.1057
$MFEI_2$	7.28E-190	-0.0296	0.3772	-0.0689	-0.2149	0.4190	0.3124	0.3092	-0.7060	0.2867	0.1452	0.1614	-0.1283	-0.5533
$MFEI_1$	5.42E-01	8.09E-158	-0.0064	0.2446	-0.4185	0.3975	0.3022	0.2865	-0.2376	0.7732	0.3429	0.3278	-0.4063	-0.2725
%G+C	2.55E-01	1.03E-06	2.25E-67	62.3790	0.0245	-0.3586	0.0258	0.0277	-0.1374	0.1562	0.1775	0.1579	-0.0365	-0.0388
P(S)	4.24E-09	2.66E-52	4.79E-193	8.24E-02	0.4000	-0.4024	-0.2048	-0.1539	0.0334	-0.3727	-0.1560	-0.1190	0.7354	0.0148
MFE(s)	2.19E-01	3.87E-194	6.60E-175	1.70E-142	1.54E-178	-0.2937	0.2502	0.2357	-0.0917	0.4447	0.1470	0.1530	-0.3294	-0.2152
Q(s)	3.22E-40	8.83E-109	6.25E-102	6.76E-02	1.15E-47	1.92E-70	2.88E-01	0.8927	-0.2253	0.2783	0.7257	0.7197	-0.2132	-0.1117
D(s)	3.49E-32	1.59E-106	9.29E-92	4.95E-02	1.29E-27	1.18E-62	0.00E+00	0.0984	-0.2332	0.2681	0.7194	0.7613	-0.1667	-0.1368
F(S)	1.05E-271	0.00E+00	1.40E-57	2.34E-20	2.47E-02	6.70E-10	5.89E-52	1.84E-55	0.3820	-0.1166	-0.0926	-0.1115	-0.0285	0.5982
zG	6.77E-21	5.71E-92	0.00E+00	1.67E-28	1.60E-153	1.84E-218	9.80E-87	1.44E-80	4.08E-15	-1.8302	0.3672	0.3494	-0.4490	-0.2341
zQ	4.83E-02	6.85E-25	8.95E-131	2.57E-36	2.17E-28	1.80E-25	0.00E+00	0.00E+00	4.49E-10	1.21E-149	-0.5625	0.8913	-0.2345	-0.0625
zD	6.96E-02	2.31E-30	1.12E-119	4.31E-29	3.47E-17	1.97E-27	0.00E+00	0.00E+00	6.05E-14	1.16E-135	0.00E+00	-0.3737	-0.1864	-0.0951
zP	1.41E-01	8.84E-20	9.77E-183	9.71E-03	0.00E+00	8.91E-121	1.20E-51	2.99E-32	5.50E-02	1.12E-222	3.70E-62	6.32E-40	4.2017	-0.0313
zF	1.02E-13	0.00E+00	2.76E-83	5.94E-03	2.93E-01	1.31E-52	2.33E-15	3.00E-22	0.00E+00	5.81E-62	9.29E-06	1.48E-11	2.66E-02	1.3094

$C_k(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.5843	0.0136	-0.0244	-0.1230	0.0254	0.2765	0.2471	-0.6788	-0.1973	-0.0365	-0.0339	0.0328	-0.1581
$MFEI_2$	2.38E-205	-0.0296	0.5299	-0.1018	-0.3162	0.5720	0.4496	0.4458	-0.8518	0.3971	0.2111	0.2350	-0.1884	-0.7362
$MFEI_1$	5.19E-01	1.76E-162	-0.0064	0.3509	-0.5911	0.5597	0.4326	0.4115	-0.3260	**0.9253**	0.4862	0.4662	-0.5787	-0.3946
%G+C	2.48E-01	1.38E-06	6.04E-66	62.3790	0.0363	-0.5068	0.0376	0.0408	-0.1924	0.2265	0.2567	0.2288	-0.0541	-0.0577
P(S)	5.17E-09	3.11E-53	3.14E-211	8.61E-02	0.4000	-0.5698	-0.3026	-0.2282	0.0475	-0.5352	-0.2325	-0.1776	**0.9049**	0.0221
MFE(s)	2.30E-01	6.67E-195	5.31E-185	1.62E-146	4.43E-193	-0.2937	0.3630	0.3430	-0.1291	0.6129	0.2173	0.2260	-0.4729	-0.3150
Q(s)	1.33E-40	6.12E-112	7.07E-103	7.50E-02	1.13E-48	9.35E-71	0.2885	**0.9837**	-0.3128	0.3967	0.8929	0.8910	-0.3131	-0.1659
D(s)	1.58E-32	6.99E-110	2.52E-92	5.36E-02	7.35E-28	6.60E-63	0.00E+00	0.0984	-0.3237	0.3839	0.8891	**0.9162**	-0.2466	-0.2027
F(S)	1.16E-302	0.00E+00	1.25E-56	3.95E-20	2.45E-02	8.69E-10	4.65E-52	8.18E-56	0.3820	-0.1537	-0.1265	-0.1528	-0.0407	0.7564
zG	4.12E-21	1.64E-85	0.00E+00	1.82E-27	2.81E-166	2.44E-231	2.28E-85	1.28E-79	2.58E-13	-1.8302	0.5136	0.4913	-0.6282	-0.3370
zQ	8.40E-02	5.34E-24	2.63E-133	4.63E-35	6.84E-29	2.32E-25	0.00E+00	0.00E+00	1.86E-09	3.94E-151	-0.5625	**0.9831**	-0.3429	-0.0926
zD	1.08E-01	1.71E-29	2.69E-121	5.24E-28	2.46E-17	2.40E-27	0.00E+00	0.00E+00	3.54E-13	1.62E-136	0.00E+00	-0.3737	-0.2741	-0.1406
zP	1.20E-01	2.37E-19	1.51E-200	1.05E-02	0.00E+00	3.13E-125	3.73E-52	2.07E-32	5.38E-02	2.54E-246	7.45E-63	6.58E-40	4.2017	-0.0459
zF	5.10E-14	0.00E+00	2.22E-84	6.28E-03	2.95E-01	8.61E-53	2.71E-15	3.31E-22	0.00E+00	1.27E-60	1.13E-05	2.33E-11	2.97E-02	1.3094

(*Upper diagonal*) Correlation coefficients $C(f, g)$. $|C| \leq 1.0$, 1.0 for trend identical, -1.0 for perfect opposite, and 0.0 for complete independence. ***Bold, 0.9 ≤ |C|*** strongly correlated, ***0.4 ≤ |C| < 0.9 moderately, and |C| < 0.4 weakly;*** (*Diagonal*) 95th percentile; (*Lower diagonal*) two-tailed *p*-values using the Student's t distribution for C_p. two-tailed *p*-values using the large-sample approximations for C_s and C_k. The pair(s) of variables with $C_p > 0$ ($C_p < 0$) and *p*-value < 0.001 tend to increase together (one variable decreases while the other increases).

Dinucleotide Shuffling

$C_p(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.3777	-0.0366	-0.0784	-0.0567	0.0394	0.2737	0.2424	-0.4389	-0.2180	0.0183	0.0098	0.0814	-0.0884
$MFEI_2$	6.76E-77	-0.0296	0.5484	-0.0535	-0.2937	0.5478	0.3374	0.3401	-0.8925	0.4157	0.2147	0.2279	-0.1888	-0.8092
$MFEI_1$	8.36E-02	3.76E-176	-0.0064	0.3589	-0.5960	0.5644	0.4323	0.4228	-0.4084	0.8833	0.4905	0.4813	-0.5677	-0.4343
%G+C	2.02E-04	1.14E-02	4.28E-69	62.3790	0.0701	-0.5437	0.0166	0.0364	-0.1909	0.2613	0.2059	0.1970	-0.0756	-0.0637
P(S)	7.25E-03	7.98E-46	1.30E-215	8.91E-04	0.4000	-0.6030	-0.3244	-0.2649	0.0515	-0.5410	-0.2862	-0.2379	0.8867	0.0464
MFE(s)	6.25E-02	1.11E-175	1.01E-188	1.23E-172	4.89E-222	-0.2937	0.3972	0.3745	-0.1964	0.5748	0.2794	0.2784	-0.4598	-0.3364
Q(s)	8.69E-40	8.85E-61	1.05E-102	4.33E-01	4.43E-56	1.40E-85	0.2885	**0.9829**	-0.2230	0.4823	**0.9387**	**0.9221**	-0.3730	-0.1336
D(s)	2.44E-31	8.42E-62	7.14E-98	8.48E-02	2.70E-37	1.57E-75	0.00E+00	0.0984	-0.2400	0.4791	**0.9348**	**0.9486**	-0.3237	-0.1640
F(S)	3.50E-106	0.00E+00	8.37E-91	7.86E-20	1.47E-02	6.50E-21	1.17E-26	1.01E-30	0.3820	-0.2459	-0.1279	-0.1486	-0.0074	0.8195
zG	1.59E-25	2.31E-94	0.00E+00	2.59E-36	1.36E-170	2.84E-197	6.51E-131	5.93E-129	3.25E-32	-1.4415	0.5998	0.5862	-0.6668	-0.3767
zQ	3.86E-01	8.99E-25	5.58E-136	7.18E-23	1.62E-43	1.85E-41	0.00E+00	0.00E+00	1.24E-09	4.68E-219	-0.5185	**0.9846**	-0.4207	-0.1125
zD	6.44E-01	8.74E-28	2.72E-130	4.74E-21	3.26E-30	3.71E-41	0.00E+00	0.00E+00	1.52E-12	5.75E-207	0.00E+00	-0.3807	-0.3646	-0.1450
zP	1.14E-04	1.98E-19	2.10E-191	3.39E-04	0.00E+00	1.15E-117	6.87E-75	8.20E-56	7.28E-01	2.83E-288	8.09E-97	2.00E-71	4.1119	0.0030
zF	2.78E-05	0.00E+00	9.35E-104	2.56E-03	2.82E-02	1.95E-60	2.16E-10	5.54E-15	0.00E+00	1.78E-76	9.21E-08	5.33E-12	8.87E-01	1.3811

$C_s(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.4177	0.0087	-0.0162	-0.0836	0.0175	0.1887	0.1679	-0.5274	-0.1114	-0.0212	-0.0211	0.0102	-0.0973
$MFEI_2$	7.28E-190	-0.0296	0.3772	-0.0689	-0.2149	0.4190	0.3124	0.3092	-0.7060	0.2778	0.1371	0.1534	-0.1211	-0.5369
$MFEI_1$	5.42E-01	8.09E-158	-0.0064	0.2446	-0.4185	0.3975	0.3022	0.2865	-0.2376	0.7201	0.3241	0.3121	-0.3907	-0.2646
%G+C	2.55E-01	1.03E-06	2.25E-67	62.3790	0.0245	-0.3586	0.0258	0.0277	-0.1374	0.1604	0.1767	0.1589	-0.0466	-0.0327
P(S)	4.24E-09	2.66E-52	4.79E-193	8.24E-02	0.4000	-0.4024	-0.2048	-0.1539	0.0334	-0.3728	-0.1520	-0.1148	0.7186	0.0143
MFE(s)	2.19E-01	3.87E-194	6.60E-175	1.70E-142	1.54E-178	-0.2937	0.2502	0.2357	-0.0917	0.4153	0.1313	0.1380	-0.3062	-0.2117
Q(s)	3.22E-40	8.83E-109	6.25E-102	6.76E-02	1.15E-47	1.92E-70	0.2885	0.8927	-0.2253	0.3108	0.7139	0.7099	-0.2291	-0.1148
D(s)	3.49E-32	1.59E-106	9.29E-92	4.95E-02	1.29E-27	1.18E-62	0.00E+00	0.0984	-0.2332	0.2986	0.7080	0.7487	-0.1815	-0.1395
F(S)	1.05E-271	0.00E+00	1.40E-57	2.34E-20	2.47E-02	6.70E-10	5.89E-52	1.84E-55	0.3820	-0.1194	-0.0896	-0.1084	-0.0274	0.5785
zG	4.59E-15	1.74E-86	0.00E+00	5.64E-30	1.36E-153	8.33E-191	9.91E-108	1.89E-99	9.00E-16	-1.4415	0.3996	0.3796	-0.4591	-0.2254
zQ	1.36E-01	2.27E-22	5.48E-117	5.42E-36	5.21E-27	1.22E-20	0.00E+00	0.00E+00	1.65E-09	7.65E-177	-0.5185	0.8916	-0.2539	-0.0587
zD	1.38E-01	1.34E-27	1.16E-108	1.91E-29	4.20E-16	1.27E-22	0.00E+00	0.00E+00	2.99E-13	8.26E-160	0.00E+00	-0.3807	-0.2045	-0.0916
zP	4.72E-01	8.41E-18	4.17E-169	9.57E-04	0.00E+00	1.21E-104	2.13E-59	6.97E-38	6.55E-02	8.56E-233	1.54E-72	1.04E-47	4.1119	-0.0365
zF	7.66E-12	0.00E+00	1.31E-78	2.04E-02	3.13E-01	5.72E-51	3.90E-16	4.69E-23	0.00E+00	1.37E-57	3.12E-05	8.21E-11	9.63E-03	1.3811

$C_k(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.5843	0.0136	-0.0244	-0.1230	0.0254	0.2765	0.2471	-0.6788	-0.1642	-0.0270	-0.0275	0.0162	-0.1457
$MFEI_2$	2.38E-205	-0.0296	0.5299	-0.1018	-0.3162	0.5720	0.4496	0.4458	-0.8518	0.3910	0.1995	0.2235	-0.1775	-0.7185
$MFEI_1$	5.19E-01	1.76E-162	-0.0064	0.3509	-0.5911	0.5597	0.4326	0.4115	-0.3260	0.8869	0.4612	0.4455	-0.5561	-0.3836
%G+C	2.48E-01	1.38E-06	6.04E-66	62.3790	0.0363	-0.5068	0.0376	0.0408	-0.1924	0.2307	0.2555	0.2306	-0.0688	-0.0486
P(S)	5.17E-09	3.11E-53	3.14E-211	8.61E-02	0.4000	-0.5698	-0.3026	-0.2282	0.0475	-0.5337	-0.2261	-0.1716	0.8892	0.0212
MFE(s)	2.30E-01	6.67E-195	5.31E-185	1.62E-146	4.43E-193	-0.2937	0.3630	0.3430	-0.1291	0.5776	0.1948	0.2046	-0.4416	-0.3102
Q(s)	1.33E-40	6.12E-112	7.07E-103	7.50E-02	1.13E-48	9.35E-71	0.2885	**0.9837**	-0.3128	0.4430	0.8842	0.8833	-0.3366	-0.1704
D(s)	1.58E-32	6.99E-110	2.52E-92	5.36E-02	7.35E-28	6.60E-63	0.00E+00	0.0984	-0.3237	0.4276	0.8798	**0.9078**	-0.2683	-0.2065
F(S)	1.16E-302	0.00E+00	1.25E-56	3.95E-20	2.45E-02	8.69E-10	4.65E-52	8.18E-56	0.3820	-0.1603	-0.1223	-0.1485	-0.0391	0.7361
zG	5.13E-15	8.88E-83	0.00E+00	1.87E-28	3.27E-165	1.26E-199	2.30E-108	2.54E-100	2.32E-14	-1.4415	0.5574	0.5335	-0.6398	-0.3270
zQ	2.02E-01	1.47E-21	2.02E-118	1.00E-34	2.25E-27	1.33E-20	0.00E+00	0.00E+00	6.33E-09	3.56E-183	-0.5185	**0.9832**	-0.3708	-0.0860
zD	1.94E-01	9.34E-27	1.04E-109	2.00E-28	2.87E-16	1.31E-22	0.00E+00	0.00E+00	1.59E-12	4.83E-165	0.00E+00	-0.3807	-0.3008	-0.1347
zP	4.45E-01	2.58E-17	3.97E-182	1.12E-03	0.00E+00	1.28E-107	1.70E-60	2.89E-38	6.43E-02	2.75E-258	5.30E-74	4.31E-48	4.1119	-0.0539
zF	4.13E-12	0.00E+00	1.89E-79	2.15E-02	3.15E-01	3.61E-51	4.65E-16	5.32E-23	0.00E+00	5.18E-57	4.56E-05	1.54E-10	1.07E-02	1.3811

(*Upper diagonal*) Correlation coefficients $C(f, g)$. $|C| \leq 1.0$, 1.0 for trend identical, -1.0 for perfect opposite, and 0.0 for complete independence. ***Bold, 0.9*** $\leq |C|$ strongly correlated, ***0.4*** $\leq |C| <$ ***0.9 moderately, and*** $|C| <$ ***0.4 weakly;*** (*Diagonal*) 95th percentile; (*Lower diagonal*) two-tailed p-values using the Student's t distribution for C_p. two-tailed p-values using the large-sample approximations for C_s and C_k. The pair(s) of variables with $C_p > 0$ ($C_p < 0$) and p-value < 0.001 tend to increase together (one variable decreases while the other increases).

Zero-order Markov Model

$C_p(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.3777	-0.0366	-0.0784	-0.0567	0.0394	0.2737	0.2424	-0.4389	-0.1602	0.0172	0.0087	0.0750	-0.0935
$MFEI_2$	6.76E-77	-0.0296	0.5484	-0.0535	-0.2937	0.5478	0.3374	0.3401	-0.8925	0.4701	0.2283	0.2405	-0.2250	-0.8157
$MFEI_1$	8.36E-02	3.76E-176	-0.0064	0.3589	-0.5960	0.5644	0.4323	0.4228	-0.4084	**0.9704**	0.5056	0.4906	-0.6296	-0.4441
%G+C	2.02E-04	1.14E-02	4.28E-69	62.3790	0.0701	-0.5437	0.0166	0.0364	-0.1909	0.3849	0.1922	0.1750	-0.0969	-0.0597
P(S)	7.25E-03	7.98E-46	1.30E-215	8.91E-04	0.4000	-0.6030	-0.3244	-0.2649	0.0515	-0.5565	-0.2953	-0.2478	**0.9384**	0.0583
MFE(s)	6.25E-02	1.11E-175	1.01E-188	1.23E-172	4.89E-222	-0.2937	0.3972	0.3745	-0.1964	0.5294	0.3060	0.3077	-0.4964	-0.3492
Q(s)	8.69E-40	8.85E-61	1.05E-102	4.33E-01	4.43E-56	1.40E-85	0.2885	**0.9829**	-0.2230	0.4231	**0.9475**	**0.9317**	-0.3394	-0.1305
D(s)	2.44E-31	8.42E-62	7.14E-98	8.48E-02	2.70E-37	1.57E-75	0.00E+00	0.0984	-0.2400	0.4194	**0.9423**	**0.9569**	-0.2895	-0.1609
F(S)	3.50E-106	0.00E+00	8.37E-91	7.86E-20	1.47E-02	6.50E-21	1.17E-26	1.01E-30	0.3820	-0.3303	-0.1309	-0.1485	0.0215	0.8123
zG	2.37E-14	1.37E-123	0.00E+00	4.63E-80	2.05E-182	3.99E-162	5.16E-98	3.35E-96	3.38E-58	-1.3010	0.5364	0.5210	-0.6525	-0.4227
zQ	4.15E-01	6.82E-28	9.22E-146	4.41E-20	2.40E-46	8.79E-50	0.00E+00	0.00E+00	5.03E-10	3.25E-167	-0.5731	**0.9840**	-0.3769	-0.1192
zD	6.82E-01	7.57E-31	4.58E-136	7.06E-17	1.07E-32	2.43E-50	0.00E+00	0.00E+00	1.61E-12	3.59E-156	0.00E+00	-0.3648	-0.3234	-0.1509
zP	3.80E-04	4.12E-27	9.33E-248	4.27E-06	0.00E+00	9.14E-140	1.49E-61	1.63E-44	3.10E-01	5.89E-272	1.50E-76	9.97E-56	3.7370	0.0465
zF	9.33E-06	0.00E+00	6.02E-109	4.71E-03	5.78E-03	2.86E-65	5.59E-10	1.81E-14	0.00E+00	8.07E-98	1.53E-08	6.97E-13	2.76E-02	1.0831

$C_s(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.4177	0.0087	-0.0162	-0.0836	0.0175	0.1887	0.1679	-0.5274	-0.0611	-0.0263	-0.0255	0.0000	-0.1074
$MFEI_2$	7.28E-190	-0.0296	0.3772	-0.0689	-0.2149	0.4190	0.3124	0.3092	-0.7060	0.3154	0.1509	0.1672	-0.1454	-0.5530
$MFEI_1$	5.42E-01	8.09E-158	-0.0064	0.2446	-0.4185	0.3975	0.3022	0.2865	-0.2376	0.8793	0.3465	0.3328	-0.4418	-0.2713
%G+C	2.55E-01	1.03E-06	2.25E-67	62.3790	0.0245	-0.3586	0.0258	0.0277	-0.1374	0.2557	0.1683	0.1471	-0.0605	-0.0399
P(S)	4.24E-09	2.66E-52	4.79E-193	8.24E-02	0.4000	-0.4024	-0.2048	-0.1539	0.0334	-0.3840	-0.1629	-0.1276	0.8075	0.0147
MFE(s)	2.19E-01	3.87E-194	6.60E-175	1.70E-142	1.54E-178	-0.2937	0.2502	0.2357	-0.0917	0.3753	0.1580	0.1665	-0.3337	-0.2132
Q(s)	3.22E-40	8.83E-109	6.25E-102	6.76E-02	1.15E-47	1.92E-70	0.2885	0.8927	-0.2253	0.2862	0.7353	0.7268	-0.2079	-0.1117
D(s)	3.49E-32	1.59E-106	9.29E-92	4.95E-02	1.29E-27	1.18E-62	0.00E+00	0.0984	-0.2332	0.2736	0.7277	0.7690	-0.1601	-0.1368
F(S)	1.05E-271	0.00E+00	1.40E-57	2.34E-20	2.47E-02	6.70E-10	5.89E-52	1.84E-55	0.3820	-0.1850	-0.0928	-0.1103	-0.0087	0.6003
zG	1.71E-05	6.65E-111	0.00E+00	1.98E-73	7.11E-163	3.69E-156	1.17E-91	7.56E-84	1.27E-35	-1.3010	0.3676	0.3507	-0.4478	-0.2521
zQ	6.43E-02	9.45E-27	1.96E-133	7.89E-33	8.97E-31	3.69E-29	0.00E+00	0.00E+00	4.25E-10	5.56E-150	-0.5731	0.8887	-0.2234	-0.0655
zD	7.33E-02	1.82E-32	3.07E-123	1.87E-25	1.63E-19	3.49E-32	0.00E+00	0.00E+00	1.12E-13	1.25E-136	0.00E+00	-0.3648	-0.1793	-0.0987
zP	**1.00E+00**	5.85E-25	1.11E-215	1.78E-05	0.00E+00	7.30E-124	3.11E-49	7.52E-30	5.57E-01	1.66E-221	1.43E-56	4.56E-37	3.7370	-0.0107
zF	4.03E-14	0.00E+00	1.46E-82	4.69E-03	2.99E-01	1.07E-51	2.25E-15	3.04E-22	0.00E+00	1.48E-71	3.31E-06	2.51E-12	4.49E-01	1.0831

$C_k(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.5843	0.0136	-0.0244	-0.1230	0.0254	0.2765	0.2471	-0.6788	-0.0914	-0.0339	-0.0336	0.0020	-0.1604
$MFEI_2$	2.38E-205	-0.0296	0.5299	-0.1018	-0.3162	0.5720	0.4496	0.4458	-0.8518	0.4419	0.2192	0.2430	-0.2140	-0.7350
$MFEI_1$	5.19E-01	1.76E-162	-0.0064	0.3509	-0.5911	0.5597	0.4326	0.4115	-0.3260	**0.9793**	0.4912	0.4731	-0.6225	-0.3934
%G+C	2.48E-01	1.38E-06	6.04E-66	62.3790	0.0363	-0.5068	0.0376	0.0408	-0.1924	0.3660	0.2436	0.2132	-0.0898	-0.0591
P(S)	5.17E-09	3.11E-53	3.14E-211	8.61E-02	0.4000	-0.5698	-0.3026	-0.2282	0.0475	-0.5491	-0.2424	-0.1905	**0.9470**	0.0224
MFE(s)	2.30E-01	6.67E-195	5.31E-185	1.62E-146	4.43E-193	-0.2937	0.3630	0.3430	-0.1291	0.5287	0.2331	0.2455	-0.4790	-0.3126
Q(s)	1.33E-40	6.12E-112	7.07E-103	7.50E-02	1.13E-48	9.35E-71	0.2885	**0.9837**	-0.3128	0.4089	0.8997	0.8963	-0.3058	-0.1658
D(s)	1.58E-32	6.99E-110	2.52E-92	5.36E-02	7.35E-28	6.60E-63	0.00E+00	0.0984	-0.3237	0.3924	0.8951	**0.9211**	-0.2371	-0.2023
F(S)	1.16E-302	0.00E+00	1.25E-56	3.95E-20	2.45E-02	8.69E-10	4.65E-52	8.18E-56	0.3820	-0.2495	-0.1265	-0.1508	-0.0129	0.7563
zG	1.47E-05	8.70E-108	0.00E+00	5.29E-72	9.72E-177	1.38E-161	4.76E-91	2.28E-83	3.84E-33	-1.3010	0.5161	0.4944	-0.6279	-0.3634
zQ	1.08E-01	8.97E-26	1.90E-136	1.25E-31	2.44E-31	4.82E-29	0.00E+00	0.00E+00	1.89E-09	8.86E-153	-0.5731	**0.9824**	-0.3276	-0.0966
zD	1.12E-01	1.78E-31	2.30E-125	1.86E-24	9.36E-20	4.01E-32	0.00E+00	0.00E+00	7.16E-13	1.86E-138	0.00E+00	-0.3648	-0.2642	-0.1454
zP	**9.26E-01**	1.27E-24	1.21E-240	2.08E-05	0.00E+00	6.61E-129	9.92E-50	5.32E-30	5.43E-01	5.11E-246	3.39E-57	4.21E-37	3.7370	-0.0152
zF	2.24E-14	0.00E+00	7.79E-84	5.13E-03	2.89E-01	5.51E-52	2.83E-15	3.91E-22	0.00E+00	6.39E-71	4.58E-06	4.60E-12	4.72E-01	1.0831

(*Upper diagonal*) Correlation coefficients $C(f, g)$. $|C| \leq 1.0$, 1.0 for trend identical, -1.0 for perfect opposite, and 0.0 for complete independence. ***Bold, 0.9 ≤ |C|*** strongly correlated, *0.4 ≤ |C| < 0.9 moderately, and |C| < 0.4 weakly;* (*Diagonal*) 95th percentile; (*Lower diagonal*) two-tailed p-values using the Student's t distribution for C_p. two-tailed p-values using the large-sample approximations for C_s and C_k. The pair(s) of variables with $C_p > 0$ ($C_p < 0$) and *p*-value < 0.001 tend to increase together (one variable decreases while the other increases).

First-order Markov Model

$C_p(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.3777	-0.0366	-0.0784	-0.0567	0.0394	0.2737	0.2424	-0.4389	-0.1578	0.0208	0.0101	0.0641	-0.0880
$MFEI_2$	6.76E-77	-0.0296	0.5484	-0.0535	-0.2937	0.5478	0.3374	0.3401	-0.8925	0.4137	0.2179	0.2309	-0.1919	-0.7784
$MFEI_1$	8.36E-02	3.76E-176	-0.0064	0.3589	-0.5960	0.5644	0.4323	0.4228	-0.4084	**0.9075**	0.4930	0.4841	-0.5859	-0.4301
%G+C	2.02E-04	1.14E-02	4.28E-69	62.3790	0.0701	-0.5437	0.0166	0.0364	-0.1909	0.3349	0.1979	0.1851	-0.0902	-0.0592
P(S)	7.25E-03	7.98E-46	1.30E-215	8.91E-04	0.4000	-0.6030	-0.3244	-0.2649	0.0515	-0.5626	-0.2912	-0.2463	**0.9177**	0.0469
MFE(s)	6.25E-02	1.11E-175	1.01E-188	1.23E-172	4.89E-222	-0.2937	0.3972	0.3745	-0.1964	0.5241	0.2892	0.2922	-0.4662	-0.3372
Q(s)	8.69E-40	8.85E-61	1.05E-102	4.33E-01	4.43E-56	1.40E-85	0.2885	**0.9829**	-0.2230	0.4721	**0.9417**	**0.9252**	-0.3634	-0.1304
D(s)	2.44E-31	8.42E-62	7.14E-98	8.48E-02	2.70E-37	1.57E-75	0.00E+00	0.0984	-0.2400	0.4644	**0.9374**	**0.9514**	-0.3104	-0.1612
F(S)	3.50E-106	0.00E+00	8.37E-91	7.86E-20	1.47E-02	6.50E-21	1.17E-26	1.01E-30	0.3820	-0.2710	-0.1266	-0.1451	-0.0027	0.7941
zG	5.85E-14	2.26E-93	0.00E+00	7.52E-60	2.67E-187	2.28E-158	8.60E-125	2.68E-120	5.21E-39	-1.0477	0.5781	0.5641	-0.6670	-0.3799
zQ	3.24E-01	1.68E-25	1.40E-137	3.17E-21	4.84E-45	2.03E-44	0.00E+00	0.00E+00	1.82E-09	4.85E-200	-0.5572	**0.9841**	-0.4021	-0.1095
zD	6.32E-01	1.71E-28	4.94E-132	1.01E-18	2.54E-32	2.42E-45	0.00E+00	0.00E+00	5.06E-12	1.65E-188	0.00E+00	-0.3779	-0.3481	-0.1437
zP	2.39E-03	4.96E-20	1.08E-206	1.88E-05	0.00E+00	2.71E-121	6.46E-71	2.93E-51	8.98E-01	1.63E-288	7.14E-88	7.73E-65	3.6154	0.0103
zF	2.99E-05	0.00E+00	1.33E-101	5.09E-03	2.64E-02	1.02E-60	5.84E-10	1.65E-14	0.00E+00	7.48E-78	2.04E-07	8.35E-12	6.26E-01	1.0988

$C_s(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.4177	0.0087	-0.0162	-0.0836	0.0175	0.1887	0.1679	-0.5274	-0.0590	-0.0226	-0.0232	-0.0054	-0.0992
$MFEI_2$	7.28E-190	-0.0296	0.3772	-0.0689	-0.2149	0.4190	0.3124	0.3092	-0.7060	0.2965	0.1407	0.1574	-0.1314	-0.5352
$MFEI_1$	5.42E-01	8.09E-158	-0.0064	0.2446	-0.4185	0.3975	0.3022	0.2865	-0.2376	0.7672	0.3276	0.3165	-0.4134	-0.2609
%G+C	2.55E-01	1.03E-06	2.25E-67	62.3790	0.0245	-0.3586	0.0258	0.0277	-0.1374	0.2181	0.1694	0.1485	-0.0576	-0.0328
P(S)	4.24E-09	2.66E-52	4.79E-193	8.24E-02	0.4000	-0.4024	-0.2048	-0.1539	0.0334	-0.3918	-0.1560	-0.1217	0.7789	0.0127
MFE(s)	2.19E-01	3.87E-194	6.60E-175	1.70E-142	1.54E-178	-0.2937	0.2502	0.2357	-0.0917	0.3734	0.1408	0.1509	-0.3131	-0.2083
Q(s)	3.22E-40	8.83E-109	6.25E-102	6.76E-02	1.15E-47	1.92E-70	0.2885	0.8927	-0.2253	0.3044	0.7211	0.7162	-0.2153	-0.1147
D(s)	3.49E-32	1.59E-106	9.29E-92	4.95E-02	1.29E-27	1.18E-62	0.00E+00	0.0984	-0.2332	0.2905	0.7151	0.7564	-0.1666	-0.1393
F(S)	1.05E-271	0.00E+00	1.40E-57	2.34E-20	2.47E-02	6.70E-10	5.89E-52	1.84E-55	0.3820	-0.1599	-0.0877	-0.1050	-0.0181	0.5805
zG	3.28E-05	2.85E-98	0.00E+00	6.40E-54	1.91E-169	1.38E-154	2.06E-103	2.99E-94	5.06E-27	-1.0477	0.3807	0.3634	-0.4622	-0.2390
zQ	1.12E-01	1.78E-23	1.54E-119	3.26E-33	2.23E-28	1.67E-23	0.00E+00	0.00E+00	3.53E-09	1.02E-160	-0.5572	0.8891	-0.2342	-0.0600
zD	1.02E-01	6.02E-29	1.19E-111	6.55E-26	6.96E-18	9.78E-27	0.00E+00	0.00E+00	1.55E-12	1.41E-146	0.00E+00	-0.3779	-0.1891	-0.0937
zP	7.03E-01	1.16E-20	4.54E-189	4.39E-05	0.00E+00	2.70E-109	1.16E-52	3.42E-32	2.22E-01	6.49E-236	5.17E-62	4.76E-41	3.6154	-0.0204
zF	2.94E-12	0.00E+00	1.72E-76	2.01E-02	3.69E-01	2.13E-49	4.03E-16	5.36E-23	0.00E+00	1.64E-64	2.05E-05	2.93E-11	1.48E-01	1.0988

$C_k(f, g)$	*Length*	$MFEI_2$	$MFEI_1$	*%G+C*	*P(S)*	*MFE(s)*	*Q(s)*	*D(s)*	*F(S)*	*zG*	*zQ*	*zD*	*zP*	*zF*
Length	174.4500	0.5843	0.0136	-0.0244	-0.1230	0.0254	0.2765	0.2471	-0.6788	-0.0875	-0.0286	-0.0302	-0.0069	-0.1483
$MFEI_2$	2.38E-205	-0.0296	0.5299	-0.1018	-0.3162	0.5720	0.4496	0.4458	-0.8518	0.4190	0.2048	0.2288	-0.1933	-0.7159
$MFEI_1$	5.19E-01	1.76E-162	-0.0064	0.3509	-0.5911	0.5597	0.4326	0.4115	-0.3260	**0.9157**	0.4660	0.4514	-0.5841	-0.3788
%G+C	2.48E-01	1.38E-06	6.04E-66	62.3790	0.0363	-0.5068	0.0376	0.0408	-0.1924	0.3120	0.2452	0.2158	-0.0854	-0.0488
P(S)	5.17E-09	3.11E-53	3.14E-211	8.61E-02	0.4000	-0.5698	-0.3026	-0.2282	0.0475	-0.5569	-0.2324	-0.1819	**0.9259**	0.0193
MFE(s)	2.30E-01	6.67E-195	5.31E-185	1.62E-146	4.43E-193	-0.2937	0.3630	0.3430	-0.1291	0.5265	0.2087	0.2232	-0.4511	-0.3057
Q(s)	1.33E-40	6.12E-112	7.07E-103	7.50E-02	1.13E-48	9.35E-71	0.2885	**0.9837**	-0.3128	0.4352	0.8896	0.8884	-0.3171	-0.1701
D(s)	1.58E-32	6.99E-110	2.52E-92	5.36E-02	7.35E-28	6.60E-63	0.00E+00	0.0984	-0.3237	0.4170	0.8853	**0.9131**	-0.2472	-0.2059
F(S)	1.16E-302	0.00E+00	1.25E-56	3.95E-20	2.45E-02	8.69E-10	4.65E-52	8.18E-56	0.3820	-0.2171	-0.1201	-0.1439	-0.0258	0.7360
zG	3.34E-05	5.47E-96	0.00E+00	8.92E-52	9.13E-183	4.62E-160	3.25E-104	5.16E-95	2.60E-25	-1.0477	0.5351	0.5135	-0.6441	-0.3468
zQ	1.76E-01	1.21E-22	3.44E-121	4.98E-32	7.47E-29	1.76E-23	0.00E+00	0.00E+00	1.17E-08	3.23E-166	-0.5572	**0.9826**	-0.3437	-0.0882
zD	1.53E-01	5.14E-28	5.89E-113	4.93E-25	4.08E-18	1.08E-26	0.00E+00	0.00E+00	7.62E-12	5.14E-151	0.00E+00	-0.3779	-0.2793	-0.1376
zP	7.42E-01	2.59E-20	3.60E-205	5.14E-05	0.00E+00	9.54E-113	1.54E-53	1.54E-32	2.21E-01	7.85E-263	3.81E-63	1.94E-41	3.6154	-0.0302
zF	1.73E-12	0.00E+00	2.31E-77	2.08E-02	3.61E-01	1.10E-49	5.16E-16	6.88E-23	0.00E+00	2.46E-64	2.93E-05	6.00E-11	1.53E-01	1.0988

(*Upper diagonal*) Correlation coefficients $C(f, g)$. $|C| \leq 1.0$, 1.0 for trend identical, -1.0 for perfect opposite, and 0.0 for complete independence. ***Bold, 0.9 ≤ |C|*** strongly correlated, ***0.4 ≤ |C| < 0.9 moderately, and |C| < 0.4 weakly;*** (*Diagonal*) 95th percentile; (*Lower diagonal*) two-tailed p-values using the Student's t distribution for C_p. two-tailed p-values using the large-sample approximations for C_s and C_k. The pair(s) of variables with $C_p > 0$ ($C_p < 0$) and *p*-value < 0.001 tend to increase together (one variable decreases while the other increases).