Application of Statistical Mirroring in Biological Sequence Analysis

*Kabir Bindawa Abdullahi

Department of Biology, Faculty of Natural and Applied Sciences, Umaru Musa Yar'adua University, P.M.B., 2218 Katsina, Katsina State, Nigeria.

*Correspondence: kabir.abdullahi@umyu.edu.ng ; kabirnamallam@gmail.com (+2348065995423)

Abstract

Sequence alignment and comparison through pairwise, multiple, global and local techniques are the main principles that underpin comparative genomics. However, most of the algorithms used are alignment-based which imposed some limitations on their use and application. In an attempt to provide an alignment-free alternative approaches, a methodology of comparative optinalysis and statistical mirroring was used and adopted to provide a suitable alternative for multiple genomic sequence comparison. In this article, methods comparison with MUSCLE, MUFFT, Clustal Omega, and T-Coffee was designed to assess the suitability and statistical power of statistical mirroring as an alternative method for multiple genomic sequences comparison using different sets of logically generated biological sequence datasets with different problems and computational complications. The results of the comparisons validate that statistical mirroring is a suitable alignment-free alternative approach for multiple genomic sequence comparison. The applied method (statistical mirroring) distinguishes itself over MUSCLE, MUFFT, Clustal Omega, and T-Coffee in specificity to a position-specific changes, specificity to a base-specific changes, cladogram and phylogenetic linearity, alignment independency, computational simplicity, and limit of input capacity.

Keywords: Statistical mirroring; genomic mirrors; comparative optinalysis; multiple comparison; inferences; homology

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

1. Introduction

Over the last three decades, there have been a growing concern on the application of bioinformatics tools for the analysis of biological sequences due to an increasing interest in bioinformatics research. Recently, the emergence of new generation sequencing (NGS) technology have not only improve the sequencing techniques but also impose big data management strategies. The need understand and infer homological inferences in these interesting Big data of biological sequences necessitate for more complex mathematical and statistical tools to analyze them. Biological sequences are generated every day from our environments. Over time, Scientists all over the world have made a tremendous efforts in the development of methodologies, algorithms, and models (such as BLSAT, FAST, Needle, Matcher, Stretcher, Water, Clustal Omega, Kalign, MAFFT, MUSCLE, T-Coffee, etc) for pairwise and multiple analysis of biological sequences. Nearly almost all of these methodologies and tools developed principally relies on sequence alignment and dynamic programming techniques (Lunter, *et al.,* 2008; Russell, 2014).

Despite the extensive applications of these alignment-base algorithms and tools in bioinformatics, they are however computationally challenged with some limitations and problems. Problem of alignment error and the resulting incongruences is the main serious challenge. Limitations in the application (e.g for whole genome comparisons) and in the input capacity or file size are some of the challenges, depending on the algorithm used (Lunter, *et al.,* 2008; Russell, 2014; Dewey, 2019). Recently, alignment-free algorithms are seem to be the alternative approaches.

The development of some of the alignment-free algorithms fall into two broad categories: methods based on k-mer or word frequency, and methods based on match length (Haubold, 2014). Methods based on k-mer or word frequency are quite popular and studied extensively. Alignment-free algorithms have addressed some computational challenges, but are not completely solved solutions because many researchers are unclear about how these methods work, how they compare to alignment-based methods, and what their potential is for use for their research (Zielezinski *et al.,* 2017). Up till now, alignment-base algorithms are still managed as the most used approaches in bioinformatics.

In this article, the application of statistical mirroring and comparative optinalysis methodologies as an alternative approach have found to be more suitable over MUSCLE, MUFFT, Clustal Omega, and T-Coffee, in specificity to a position-specific changes, specificity to a base-specific changes, cladogram and phylogenetic linearity, alignment independency, computational simplicity, and limit of input capacity.

2. Methodology

2.1 Sequence Datasets

Three (3) different datasets of biological sequences were logically generated to present different computational variations and complications. These generated sequences were used to validate the application of statistical mirroring in biological sequence analysis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

First Sequence Dataset:

Presented below is a list of twelve (12) biological sequences each with a length of 12pb. Each sequence differs from the others by just one base mismatch at different specific position along the sequence read.

Biological sequences

$M_1 = {}^{(5')}(A\ T\ G\ A\ C\ T\ G\ A\ G\ C\ C\ T)^{(3')}$
$M_2 = {}^{(5')}(G\ A\ G\ A\ C\ T\ G\ A\ G\ C\ C\ T)^{(3')}$
$M_3 = {}^{(5')}(G\ T\ A\ A\ C\ T\ G\ A\ G\ C\ C\ T)^{(3')}$
$M_4 = {}^{(5')}(G\ T\ G\ T\ C\ T\ G\ A\ G\ C\ C\ T)^{(3')}$
$M_5 = {}^{(5')}(G\ T\ G\ A\ A\ T\ G\ A\ G\ C\ C\ T)^{(3')}$
$M_6 = {}^{(5')}(G\ T\ G\ A\ C\ A\ G\ A\ G\ C\ C\ T)^{(3')}$
$M_7 = {}^{(5')}(G\ T\ G\ A\ C\ T\ A\ A\ G\ C\ C\ T)^{(3')}$
$M_8 = {}^{(5')}(G\ T\ G\ A\ C\ T\ G\ T\ G\ C\ C\ T)^{(3')}$
$M_9 = {}^{(5')}(G\ T\ G\ A\ C\ T\ G\ A\ A\ C\ C\ T)^{(3')}$
$M_{10} = {}^{(5')}(G\ T\ G\ A\ C\ T\ G\ A\ G\ A\ C\ T)^{(3')}$
$M_{11} = {}^{(5')}(G\ T\ G\ A\ C\ T\ G\ A\ G\ C\ A\ T)^{(3')}$
$M_{12} = {}^{(5')}(G\ T\ G\ A\ C\ T\ G\ A\ G\ C\ C\ A)^{(3')}$

Second Sequence Dataset:

Presented below is a list of twelve (12) biological sequences each with a length of 12pb. Each sequence differs from the others by just one base deletion at different specific position along the sequence read, but the position of deletion is conserved by a gab.

Biological sequences

$D_1 = {}^{(5')}(-\ T\ G\ A\ C\ T\ G\ A\ G\ C\ C\ T)^{(3')}$
$D_2 = {}^{(5')}(G\ -\ G\ A\ C\ T\ G\ A\ G\ C\ C\ T)^{(3')}$
$D_3 = {}^{(5')}(G\ T\ -\ A\ C\ T\ G\ A\ G\ C\ C\ T)^{(3')}$
$D_4 = {}^{(5')}(G\ T\ G\ -\ C\ T\ G\ A\ G\ C\ C\ T)^{(3')}$
$D_5 = {}^{(5')}(G\ T\ G\ A\ -\ T\ G\ A\ G\ C\ C\ T)^{(3')}$
$D_6 = {}^{(5')}(G\ T\ G\ A\ C\ -\ G\ A\ G\ C\ C\ T)^{(3')}$
$D_7 = {}^{(5')}(G\ T\ G\ A\ C\ T\ -\ A\ G\ C\ C\ T)^{(3')}$
$D_8 = {}^{(5')}(G\ T\ G\ A\ C\ T\ G\ -\ G\ C\ C\ T)^{(3')}$
$D_9 = {}^{(5')}(G\ T\ G\ A\ C\ T\ G\ A\ -\ C\ C\ T)^{(3')}$
$D_{10} = {}^{(5')}(G\ T\ G\ A\ C\ T\ G\ A\ G\ -\ C\ T)^{(3')}$
$D_{11} = {}^{(5')}(G\ T\ G\ A\ C\ T\ G\ A\ G\ C\ -\ T)^{(3')}$
$D_{12} = {}^{(5')}(G\ T\ G\ A\ C\ T\ G\ A\ G\ C\ C\ -)^{(3')}$

Third Sequence Dataset:

Presented below is a list of twelve (12) biological sequences with a length range from 42pb to 09bp. Each sequence differs from the others by just a trinucleotide (3) base deletion at one terminal ends of the sequence read.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Biological sequences

$L_1 = {}^{(5')}(AUUAUGGUGUAACACAAAGAUUGUCGCCGGGGGAGGAGUCCA)^{(3')}$

$L_2 = {}^{(5')}(AUUAUGGUGUAACACAAAGAUUGUCGCCGGGGGAGGAGU)^{(3')}$

$L_3 = {}^{(5')}(AUUAUGGUGUAACACAAAGAUUGUCGCCGGGGGAGG)^{(3')}$

$L_4 = {}^{(5')}(AUUAUGGUGUAACACAAAGAUUGUCGCCGGGGG)^{(3')}$

$L_5 = {}^{(5')}(AUUAUGGUGUAACACAAAGAUUGUCGCCGG)^{(3')}$

$L_6 = {}^{(5')}(AUUAUGGUGUAACACAAAGAUUGUCGC)^{(3')}$

$L_7 = {}^{(5')}(AUUAUGGUGUAACACAAAGAUUGU)^{(3')}$

$L_8 = {}^{(5')}(AUUAUGGUGUAACACAAAGAU)^{(3')}$

$L_9 = {}^{(5')}(AUUAUGGUGUAACACAAA)^{(3')}$

$L_{10} = {}^{(5')}(AUUAUGGUGUAACAC)^{(3')}$

$L_{11} = {}^{(5')}(AUUAUGGUGUAA)^{(3')}$

$L_{12} = {}^{(5')}(AUUAUGGUG)^{(3')}$

2.2 Sequence Alignment

Except the designed genomic mirror sequences, all the generated biological sequence datasets (the first, second and third) were also aligned by T-Coffee program on EMBL-EBI website. T-Coffee was chosen here because it shown the greatest spectrum of alignment changes over the other tools. The aligned sequences of each dataset were presented in Appendix A.

2.3 Statistical Mirrors Design

In this application, a maximalic and minimalic statistical mirror were used to infer sequence similarity (homology) between the multiple genomic sequences. Nucleotide base with the highest (Guinine, G) and the lowest (Cytocine, C) molecular masses were expressed as the principal value of the maximal and minimal element for the design of the maximalic and minimalic statistical mirrors respectively. The statistical mirrors below were designed in accordance with Abdullahi (2019b).

Statistical mirror for the first and second sequence datasets:

Maximalic genomic mirror sequence

$G_0 = {}^{(5')}(G\ \ G\ \ G\ \ G\ \ G\ \ G\ \ G\ \ G\ \ G\ \ G\ \ C\ \ G)^{(3')}$

Minimalic genomic mirror sequence

$g_0 = {}^{(5')}(C\ \ C\ \ C\ \ C\ \ C\ \ C\ \ C\ \ C\ \ C\ \ C\ \ C\ \ C)^{(3')}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Statistical mirror for the third sequence dataset:

Maximalic genomic mirror sequence
$G_0 = {}^{(5')}$(G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G)$^{(3')}$

Minimalic genomic mirror sequence
$g_0 = {}^{(5')}$(C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C)$^{(3')}$

2.4 Numerical Sequence Transformation

Since the algorithm of optinalysis works only with numerical values, a numerical or ordinal genomic sequence transformation by molecular mass proposed by Abdullahi (2019b) was adopted to convert the nominal sequences to a numerical values, as presented in Table 1. Both the unaligned and aligned biological sequences were numerically transformed prior for comparative optinalysis. The numerically transformed sequences were presented in appendix B and C.

Table 1: Numerical sequence transformation of nuclotide bases based on molecular mass (Abdullahi 2019b).

Nucleotide bases and gabs	Molecular mass
Adenine (A)	≈ 135 g/mol
Tymine (T)	≈ 126 g/mol
Cytosine (C)	≈ 111 g/mol
Guinine (G)	≈ 151 g/mol
Uracil (U)	≈ 112 g/mol
All other gabs	0

2.5 Comparative Optinalysis Algorithmic Parameters

Suppose we have the designed genomic (statistical) mirror sequences as (G_0) and (g_0) and sets of subject genomic sequences $(M_{n=1-12})$, $(D_{n=1-12})$, and $(L_{n=1-12})$ as shown in the presented datasets of nucleotide sequences.

The sequence order of the biological sequence was considered as its theoretical sequence order. Head and tail of the sequence were denoted on the sequence datasets left, $(-5')$ to the right $(-3')$ respectively.

Let each of the genomic sequences (G_0) and (g_0) optinalytically reflects head-to-head (H-H or $5'-5'$) with the mutant sequences $(M_n), (D_n)$ and (L_n) with a normalization of a zero unit, such that elements of each of the sequence(G_0) and (g_0) are intermetrically similar to the elements of sequences $(M_n), (D_n)$ and (L_n) with a resultant Kabirain coefficient of x and thus y confidence of similarity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

$$\bigwedge_{B}^{(\pm N=0)} : \int_{c(p)}^{\in(M_n,\, D_n \text{ and } L_n)} \in(G0) = x(y)$$

And also

$$\bigwedge_{B}^{(\pm N=0)} : \int_{c(p)}^{\in(M_n,\, D_n \text{ and } L_n)} \in(g0) = x(y)$$

2.6 Data Analysis

Official website of EMBL-EBI at https://www.ebi.ac.uk was used to access their online bioinformatics tools. MUSCLE, MUFFT, Clustal Omega, and T-Coffee were used to run multiple sequence comparison. The results of similarity matrices of the constructed phylogenetic tree and the aligned sequenced was recorded for each tool used.

Comparative optinalysis and statistical mirroring between the unaligned (assumed unaligned as they were here presented) and aligned (aligned by T-Coffee) biological sequence datasets and their designed statistical mirrors was performed using the method of Abdullahi (2019a and 2019b). Kabirian coefficient of similarity and dissimilarity, and their translated probabilities and percentages were computed using an Excel programed sheets (Seen in the supplementary file sheets). The translated probabilities of percentages of similarity were considered at as the similarity distance from an unrooted evolutionary origin.

The details of the data analysis and the results of comparative optinalysis were presented in Appendix A-D and Excel sheets of the supplementary materials.

2.7 Suitability Assessment Criteria

In this case, three (3) criteria were used to compare the computational power, fitness and suitability of the application of statistical mirroring with some well-known and adopted methods of multiple genomic sequence comparison. These criteria are: (a) specificity to a position-specific changes, (b) specificity to a base-specific changes, (c) cladogram and phylogenetic linearity, (d) alignment independency, (e) computational simplicity, (f) limit of input capacity.

2.8 Results presentation

All results were presented in Tables.

3. Results and Discussion

From Table 2-4, the following explanations were obtained:

Specificity to a position-specific changes: Multiple sequence comparisons, by some bioinformatics tools and statistical mirroring, of the logically generated biological sequences from the first, second and third sequence datasets was presented in Table 2-4. The results shows that all the bioinformatics tools used have failed to quantify any variation and genetic relationship among the multiple set of sequences presented, thus the similarity distances were in some cases uniformly the same (Table 2 and 4). Although a significant variations were observed

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

by some of the used bioinformatics tools in the comparison of the second and third sequence datasets, these computationally detected variations are however not due to position-specific variations but are as a result of alignment effect (Table 3 and 4). This result confirm the earlier finding of Abdullahi (2019b) who showed that all the bioinformatics tools he used for pairwise sequence comparison are not position specific computationally, but algorithms in tools are designed to quantify how much mismatches or variations exist. This however re-validate the used of comparative optinalysis as a most suitable tool for genomic sequence comparisons, especially where positional variations is important.

Specificity to a base-specific changes: Almost all bioinformatics tools are designed to detect and quantify the number of matches, mismatches, and gaps between two or more set of sequences. One advantage of this methodology is that, the numerical sequence transformation does only tell us that a pair of corresponding bases are matched or not matched, but tells how very much alike they are even if they are different. For instance, G (Guanine base) is a same mismatch score if paired with A (Adenine) or T (Tymine) or C (Cytocine), but Guanine is structurally more similar to Adenine than Tymine, and then follows Cytocine. This structural similarities is considered and addressed computationally by molecular mass transformation of biological sequences and the algorithm of comparative optinalysis. Therefore, multiple comparison and phylogenetic reconstruction following this methodology is far better in inferring homology than these well-known and adopted bioinformatics tools used here, since homology is supposed by structure and function.

Cladogram and phylogenetic linearity: In alignment-base and alignment-free statistical mirroring, the trend of genetic relationship by maximalic and minimalic mirror takes almost the same trend (if sorted smallest to largest similarity scores) but are in reverse order except in the third sequence dataset (Table 2-4). Among the used bioinformatics tools, a well-defined and consistent phylogenetic linearity cannot be deduced and assured. Therefore, multiple comparison and phylogenetic tree reconstruction following this methodology is far better in inferring homology than these well-known and adopted bioinformatics tools used here, since the main reason for phylogeny is relationship reconstruction.

Alignment independency: The ability and capacity of comparative optinalysis and statistical mirroring to be used for both the aligned and unaligned sequences is another achievement. Comparative optinalysis and statistical mirroring uses the aligned order of biological sequence to deduce homological inferences. For the unaligned sequences, it is a way toward alignment-free approach of sequence comparison. A quite number of studies have shown the advantages of using alignment-free algorithms over alignment-base methods due to their computational limitations such as alignment errors and incongruences, whole genome comparison and etc (Susana and Jonas, 2003; Chan and Ragan, 2013; Zielezinski *et al.*, 2017; Zielezinski *et al.*, 2019; Saw, *et al.*, 2019).

Computational simplicity: Simplicity in computation, and in results interpretation is also another important criteria to assess the suitability of this application. Statistical mirroring avoid the rigorous technique of dynamic programming which are time and memory consuming, especially when dealing with a very lengthy sequences. Lange (2003) reported that it requires about 10^{60} alignments for just two sequence with 100bp length. However, steps for constructing a guide tree and the progressive alignments are not required in this methodology. Therefore, multiple comparison following this methodology is a very easy and simple approach.

 doi:10.20944/preprints201912.0069.v1

Limit of input capacity: application of comparative optinalysis and statistical mirroring in multiple sequence comparison rarely have limit to the input capacity of the sequences. But for most alignment-base tools used in multiple sequence comparison have a limit to the file size and input capacity which varies with the algorithm used (EBI, 2019).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Table 2: Alignment-base, alignment-free and statistically mirrored multiple comparisons of nucleotide base sequences

	Bioinformatics tools				**Alignment-base* and statistically mirrored**		**Alignment-free and statistically Mirrored**	
	Similarity distances				**Similarity values (P_{Sim}-value)**		**Similarity values (P_{Sim}-value)**	
Seq.	**MUSCLE**	**T-Coffee**	**MAFFT**	**Clustal Omega**	**Maximalic mirror**	**Minimalic mirror**	**Maximalic mirror**	**Minimalic mirror**
M_1	0.08333	0.08333	0.08333	0.08333	0.8954	0.8178	0.8954	0.8178
M_2	0.08333	0.08333	0.08333	0.08333	0.9037	0.8097	0.9037	0.8097
M_3	0.08333	0.08333	0.08333	0.08333	0.8819	0.831	0.8819	0.831
M_4	0.08333	0.08333	0.08333	0.08333	0.8885	0.8243	0.8885	0.8243
M_5	0.08333	0.08333	0.08333	0.08333	0.9117	0.8026	0.9117	0.8026
M_6	0.08333	0.08333	0.08333	0.08333	0.9008	0.8128	0.9008	0.8128
M_7	0.08333	0.08333	0.08333	0.08333	0.8871	0.8253	0.8871	0.8253
M_8	0.08333	0.08333	0.08333	0.08333	0.8914	0.8212	0.8914	0.8212
M_9	0.08333	0.08333	0.08333	0.08333	0.8897	0.8225	0.8897	0.8225
M_{10}	0.08333	0.08333	0.08333	0.08333	0.9019	0.8129	0.9019	0.8129
M_{11}	0.08333	0.08333	0.08333	0.08333	0.8999	0.815	0.8999	0.815
M_{12}	0.08333	0.08333	0.08333	0.08333	0.8964	0.8175	0.8964	0.8175

*Aligned by T-Coffee.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Table 3: Alignment-base, alignment-free and statistically mirrored multiple comparisons of nucleotide base sequences

	Bioinformatics tools				**Alignment-base* and statistically mirrored**		**Alignment-free and statistically mirrored**	
	Similarity distances				**Similarity values (P_{Sim}-value)**		**Similarity values (P_{Sim}-value)**	
Seq.	**MUSCLE**	**T-Coffee**	**MAFFT**	**Clustal Omega**	**Maximalic mirror**	**Minimalic mirror**	**Maximalic mirror**	**Minimalic mirror**
D_1	0	-0.00455	0.09091	0.05909	0.7488	0.9901	0.7488	0.9901
D_2	0.09182	0.09545	0	0.03182	0.7705	0.9607	0.7817	0.9448
D_3	-0.00091	0	0	0.05966	0.7709	0.9574	0.7709	0.9574
D_4	-0.00152	0.09091	0	0.03535	0.8025	0.9156	0.7941	0.9269
D_5	-0.00199	-0.00455	0	0.05556	0.8204	0.8948	0.8204	0.8948
D_6	-0.00244	0.09091	0	0.16364	0.8115	0.9047	0.8201	0.8936
D_7	-0.00352	0.09091	0	0.49940	0.8276	0.8817	0.8170	0.8951
D_8	-0.00788	-0.00455	0	0.01818	0.8360	0.8730	0.8360	0.8730
D_9	-0.00227	0.09091	0	0.01648	0.8302	0.8786	0.8409	0.8654
D_{10}	0	-0.00364	0	0	0.8736	0.8314	0.8645	0.8417
D_{11}	0	-0.00227	0	0	0.8736	0.8314	0.8736	0.8314
D_{12}	0	0.09455	0.09091	0.03826	0.8718	0.8320	0.8810	0.8218

*Aligned by T-Coffee.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Table 4: Alignment-base, alignment-free and statistically mirrored multiple comparisons of nucleotide base sequences

	Bioinformatics tools				Alignment-base* and statistically mirrored		Alignment-free and statistically mirrored	
	Similarity distances				Similarity values (P_{Sim}-value)		Similarity values (P_{Sim}-value)	
Seq.	**MUSCLE**	**T-Coffee**	**MAFFT**	**Clustal Omega**	**Maximalic mirror**	**Minimalic mirror**	**Maximalic mirror**	**Minimalic mirror**
L_1	0	0.04581	0	0	0.8614	0.8543	0.8614	0.8543
L_2	0	0.01534	0	0	0.8323	0.8777	0.8526	0.8531
L_3	0	0.01244	0	0	0.8028	0.9039	0.8335	0.8646
L_4	0	0.01957	0	0	0.7646	0.9427	0.8022	0.8905
L_5	0	0.04016	0	0	0.7154	0.9977	0.7586	0.9351
L_6	0	0.02827	0	0	0.657	0.9177	0.7090	0.9961
L_7	0	0.05612	0	0	0.598	0.835	0.6552	0.9295
L_8	0	0.03689	0	0	0.5404	0.7544	0.5930	0.8408
L_9	0	0.00700	0	0	0.4378	0.6032	0.5191	0.7335
L_{10}	0	0.08156	0	0	0.3698	0.5075	0.4373	0.6139
L_{11}	0	0.07634	0	0	0.296	0.4021	0.3604	0.5014
L_{12}	0	0.18003	0	0	0.2047	0.272	0.2742	0.3761

*Aligned by T-Coffee.

4. Conclusion

Considering the application and results of comparison here studied, it is concluded that statistical mirroring is a suitable alignment-free alternative approach for multiple genomic sequence comparison. The applied method (statistical mirroring) distinguishes itself over some well-known and adopted methods for multiple genomic sequence comparison (such as MUSCLE, MUFFT, Clustal Omega, and T-Coffee) in specificity to a position-specific changes, specificity to a base-specific changes, cladogram and phylogenetic linearity, alignment independency, computational simplicity, and limit of input capacity.

5. Recommendations

This study recommend further application of statistical mirroring and comparative optinalysis with a real-world examples of biological sequences. Furthermore, a comparison with some alignment-free algorithms should be conducted to further re-validate the methodology.

Acknowledgement

I thank for the motivations and encouragement received from Abubakar Bello (PhD) of Biology Department Murtala Isa Bindawa (PhD) and Aminu Adamu (PhD) of Biochemistry all from Umaru Musa Yar'adua University Katsina.

Funding: This research did not receive any specific grant from funding agencies in the public commercial or not-for-profit sectors.

Conflict of interest

The author declares no conflict of interest.

References

Abdullahi, K. B. (2019a). Optinalysis: A New Approach of Symmetry Detection and Similarity Measurement through a Looking-Glass. *Preprints*, 2019050295, doi: 10.20944/preprints201905.0295.v2

Abdullahi, K.B. (2019b) Optinalytic (Statistical) Mirroring: A New Novel Approach of Measure of Dispersion. *Preprints*, 2019110268, doi: 10.20944/preprints201911.0268.v1.

Chan, C. X., & Ragan, M. A. (2013). Next-generation phylogenomics. *Biology direct*, *8*(1), 3. doi:10.1186/1745-6150-8-3

Dewey C.N. (2019). *Whole-Genome Alignment*. In: Anisimova M. (eds) Evolutionary Genomics. Methods in Molecular Biology, vol 1910. Humana, New York, NY.

European Bioinformatics Institute-EBI, (2019). Bioinformatics Tools FAQ. https://www.ebi.ac.uk. Accessed on 15th November, 2019.

Haubold, B. (2013). Alignment-free phylogenetics and population genetics. *Briefings in bioinformatics*, *15*(3), 407-418.

Lange, K. (2003). *Mathematical and statistical methods for genetic analysis*. Springer Science & Business Media.

Lunter, G., Rocco, A., Mimouni, N., Heger, A., Caldeira, A., and Hein, J. (2008). Uncertainty in homology inferences: Assessing and improving genomic sequence alignment, *Genome Res.*, *18*(2): 298–309. doi: 10.1101/gr.6725608

Saw, A. K., Raj, G., Das, M., Talukdar, N. C., Tripathy, B. C., & Nandi, S. (2019). Alignment-free method for DNA sequence clustering using Fuzzy integral similarity. *Scientific reports*, *9*(1), 3753. doi:10.1038/s41598-019-40452-6

Vinga, S., & Almeida, J. (2003). Alignment-free sequence comparison—a review. *Bioinformatics*, *19*(4), 513-523.

Zielezinski, A., Girgis, H. Z., Bernard, G., Leimeister, C. A., Tang, K., Dencker, T., ... & Comin, M. (2019). Benchmarking of alignment-free sequence comparison methods. *BioRxiv*, 611137.

Zielezinski, A., Vinga, S., Almeida, J., & Karlowski, W. M. (2017). Alignment-free sequence comparison: benefits, applications, and tools. *Genome biology*, *18*(1), 186.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Appendix A

Aligned Biological Sequences of the datasets by T-Coffee

First Sequence Dataset:

The aligned results are the same as the original sequences.

Second Sequence Dataset:

$D_1 = {}^{(H)}$(-TGACTGAGCCT)$^{(T)}$

$D_2 = {}^{(H)}$(-GGACTGAGCCT)$^{(T)}$

$D_3 = {}^{(H)}$(GT-ACTGAGCCT)$^{(T)}$

$D_4 = {}^{(H)}$(GTGC-TGAGCCT)$^{(T)}$

$D_5 = {}^{(H)}$(GTGA-TGAGCCT)$^{(T)}$

$D_6 = {}^{(H)}$(GTGA-CGAGCCT)$^{(T)}$

$D_7 = {}^{(H)}$(GTGACTA-GCCT)$^{(T)}$

$D_8 = {}^{(H)}$(GTGACTG-GCCT)$^{(T)}$

$D_9 = {}^{(H)}$(GTGACTG-ACCT)$^{(T)}$

$D_{10} = {}^{(H)}$(GTGACTGAGC-T)$^{(T)}$

$D_{11} = {}^{(H)}$(GTGACTGAGC-T)$^{(T)}$

$D_{12} = {}^{(H)}$(GTGACTGAGC-C)$^{(T)}$

Third Sequence Dataset:

$L_1 = {}^{(H)}$(AUUAUGGUGUAACACAAAGAUUGUCGCCGGGGGAGGAGUCCA)$^{(T)}$

$L_2 = {}^{(H)}$(AU-UAUGGUGUA---------A----------CAC)$^{(T)}$

$L_3 = {}^{(H)}$(AU-UAUGGUG----------UAA)$^{(T)}$

$L_4 = {}^{(H)}$(AU-UAU---G----------GUG)$^{(T)}$

$L_5 = {}^{(H)}$(AUUAUGGUGUAACACAAAGAUUGUCGCCGGGG---GAGGAGU)$^{(T)}$

$L_6 = {}^{(H)}$(AUUAUGGUGUAACACAAAGAUUGUCGCCGGG------GGAGG)$^{(T)}$

$L_7 = {}^{(H)}$(AUUAUGGUGUAACACAAAGAUUGUCGCCG---------GGGG)$^{(T)}$

$L_8 = {}^{(H)}$(AUUAUGGUGUAACACAAAGAUUGUCG----------CCGG)$^{(T)}$

$L_9 = {}^{(H)}$(AUUAUGGUGUAACACAAAGAUU---G----------UCGC)$^{(T)}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

$L_{10} = {}^{(H)}$(AUUAUGGUGUAACACAAAGA----------UUGU)$^{(T)}$

$L_{11} = {}^{(H)}$(AUUAUGGUGUAACACAAA----------GAU)$^{(T)}$

$L_{12} = {}^{(H)}$(AU-UAUGGUGUA-----ACAC----------AAA)$^{(T)}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Appendix B

Numerically Transformed Genomic Mirror Sequences

Maximalic genomic mirror sequence

G_0 = (151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151)

Minimalic genomic mirror sequence

g_0 = (111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111)

Maximalic genomic mirror sequence
G_0 = (151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151, 151)

Minimalic genomic mirror sequence
g_0 = (111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111, 111)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Appendix C

Numerically Transformed Biological Sequences of the Datasets

First Sequence Dataset: Unaligned and aligned (Both results are same in this case)

Biological sequences

$M_1 = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$M_2 = {}^{(H)}(151, 135, 151, 135, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$M_3 = {}^{(H)}(151, 126, 135, 135, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$M_4 = {}^{(H)}(151, 126, 151, 126, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$M_5 = {}^{(H)}(151, 126, 151, 135, 135, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$M_6 = {}^{(H)}(151, 126, 151, 135, 111, 135, 151, 135, 151, 111, 111, 126)^{(T)}$
$M_7 = {}^{(H)}(151, 126, 151, 135, 111, 126, 135, 135, 151, 111, 111, 126)^{(T)}$
$M_8 = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 126, 151, 111, 111, 126)^{(T)}$
$M_9 = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 135, 111, 111, 126)^{(T)}$
$M_{10,} = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 151, 135, 111, 126)^{(T)}$
$M_{11} = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 151, 111, 135, 126)^{(T)}$
$M_{12} = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 151, 111, 111, 135)^{(T)}$

Second Sequence Dataset: Unaligned

Biological sequences

$D_1 = {}^{(H)}(0, 126, 151, 135, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_2 = {}^{(H)}(151, 0, 151, 135, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_3 = {}^{(H)}(151, 126, 0, 135, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_4 = {}^{(H)}(151, 126, 151, 0, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_5 = {}^{(H)}(151, 126, 151, 135, 0, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_6 = {}^{(H)}(151, 126, 151, 135, 111, 0, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_7 = {}^{(H)}(151, 126, 151, 135, 111, 126, 0, 135, 151, 111, 111, 126)^{(T)}$
$D_8 = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 0, 151, 111, 111, 126)^{(T)}$
$D_9 = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 0, 111, 111, 126)^{(T)}$
$D_{10,} = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 151, 0, 111, 126)^{(T)}$
$D_{11} = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 151, 111, 0, 126)^{(T)}$
$D_{12} = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 151, 111, 111, 0,{}^{(T)}$

Second Sequence Dataset: Aligned by T-Coffee

Biological sequences

$D_1 = {}^{(H)}(0, 126, 151, 135, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_2 = {}^{(H)}(0, 151, 151, 135, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_3 = {}^{(H)}(151, 126, 0, 135, 111, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_4 = {}^{(H)}(151, 126, 151, 111, 0, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_5 = {}^{(H)}(151, 126, 151, 135, 0, 126, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_6 = {}^{(H)}(151, 126, 151, 135, 0, 111, 151, 135, 151, 111, 111, 126)^{(T)}$
$D_7 = {}^{(H)}(151, 126, 151, 135, 111, 126, 135, 0, 151, 111, 111, 126)^{(T)}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

$D_8 = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 0, 151, 111, 111, 126)^{(T)}$
$D_9 = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 0, 135, 111, 111, 126)^{(T)}$
$D_{10,} = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 151, 111, 0, 126)^{(T)}$
$D_{11} = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 151, 111, 0, 126)^{(T)}$
$D_{12} = {}^{(H)}(151, 126, 151, 135, 111, 126, 151, 135, 151, 111, 0, 111)^{(T)}$

Third Sequence Dataset: Unaligned

Biological sequences
$L_1 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112\ 112\ 151, 112\ 111, 151, 111, 111, 151, 151, 151, 151, 151, 135, 151, 151, 135, 151, 112\ 111, 111, 135)^{(T)}$
$L_2 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112\ 112\ 151, 112\ 111, 151, 111, 111, 151, 151, 151, 151, 151, 135, 151, 151, 135, 151, 112\ 0, 0, 0)^{(T)}$
$L_3 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112\ 112\ 151, 112\ 111, 151, 111, 111, 151, 151, 151, 151, 151, 135, 151, 151, 0, 0, 0, 0, 0, 0)^{(T)}$
$L_4 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112\ 112\ 151, 112\ 111, 151, 111, 111, 151, 151, 151, 151, 151, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{(T)}$
$L_5 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112\ 112\ 151, 112\ 111, 151, 111, 111, 151, 151, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{(T)}$
$L_6 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112\ 112\ 151, 112\ 111, 151, 111, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{(T)}$
$L_7 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112\ 112\ 151, 112\ 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{(T)}$
$L_8 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112\ 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{(T)}$
$L_9 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{(T)}$
$L_{10,} = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{(T)}$
$L_{11} = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{(T)}$
$L_{12} = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{(T)}$

Third Sequence Dataset: Aligned by T-Coffee

Biological sequences
$L_1 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112\ 112\ 151, 112\ 111, 151, 111, 111, 151, 151, 151, 151, 151, 135, 151, 151, 135, 151, 112\ 111, 111, 135)^{(T)}$
$L_2 = {}^{(H)}(135, 112\ 112\ 135, 112\ 151, 151, 112\ 151, 112\ 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112\ 112\ 151, 112\ 111, 151, 111, 111, 151, 151, 151, 151, 0, 0, 0, 151, 135, 151, 151, 135, 151, 112)^{(T)}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

$L_3 = {}^{(H)}$(135, 112 112 135, 112 151, 151, 112 151, 112 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112 112 151, 112 111, 151, 111, 111, 151, 151, 151, 0, 0, 0, 0, 0, 0, 151, 151, 135, 151, 151)$^{(T)}$

$L_4 = {}^{(H)}$(135, 112 112 135, 112 151, 151, 112 151, 112 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112 112 151, 112 111, 151, 111, 111, 151, 0, 0, 0, 0, 0, 0, 0, 0, 151, 151, 151, 151)$^{(T)}$

$L_5 = {}^{(H)}$(135, 112 112 135, 112 151, 151, 112 151, 112 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112 112 151, 112 111, 151, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 111, 111, 151, 151)$^{(T)}$

$L_6 = {}^{(H)}$(135, 112 112 135, 112 151, 151, 112 151, 112 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 112 112 0, 0, 0, 151, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 112 111, 151, 111)$^{(T)}$

$L_7 = {}^{(H)}$(135, 112 112 135, 112 151, 151, 112 151, 112 135, 135, 111, 135, 111, 135, 135, 135, 151, 135, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 112 112 151, 112)$^{(T)}$

$L_8 = {}^{(H)}$(135, 112 112 135, 112 151, 151, 112 151, 112 135, 135, 111, 135, 111, 135, 135, 135, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 151, 135, 112)$^{(T)}$

$L_9 = {}^{(H)}$(135, 112 0, 112 135, 112 151, 151, 112 151, 112 135, 0, 0, 0, 0, 0, 135, 111, 135, 111, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 135, 135, 135)$^{(T)}$

$L_{10,} = {}^{(H)}$(135, 112 0, 112 135, 112 151, 151, 112 151, 112 135, 0, 0, 0, 0, 0, 0, 0, 0, 0, 135, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 111, 135, 111)$^{(T)}$

$L_{11} = {}^{(H)}$(135, 112 0, 112 135, 112 151, 151, 112 151, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 112 135, 135)$^{(T)}$

$L_{12} = {}^{(H)}$(135, 112 0, 112 135, 112 0, 0, 0, 151, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 151, 112 151)$^{(T)}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Appendix D

Computational Results of Statistical Mirroring

Table D1: Alignment-free, alignment-base and statistically mirrored multiple comparisons of nucleotide base sequences

	Maximalic Genomic Mirror			Minimalic Genomic Mirror		
Seq.	**K_c-value**	**$P_{Sim.}$**	**$P_{Dsim.}$**	**K_c-value**	**$P_{Sim.}$**	**$P_{Dsim.}$**
M_1	0.973138	0.8954	0.1046	1.052776	0.8178	0.1822
M_2	0.975329	0.9037	0.0963	1.055497	0.8097	0.1903
M_3	0.96958	0.8819	0.1181	1.048399	0.831	0.169
M_4	0.97133	0.8885	0.1115	1.050578	0.8243	0.1757
M_5	0.977429	0.9117	0.0883	1.057914	0.8026	0.1974
M_6	0.974556	0.9008	0.0992	1.054444	0.8128	0.1872
M_7	0.970951	0.8871	0.1129	1.050267	0.8253	0.1747
M_8	0.972102	0.8914	0.1086	1.05163	0.8212	0.1788
M_9	0.971638	0.8897	0.1103	1.051204	0.8225	0.1775
M_{10}	0.974858	0.9019	0.0981	1.054413	0.8129	0.1871
M_{11}	0.974345	0.8999	0.1001	1.053715	0.815	0.185
M_{12}	0.9734	0.8964	0.1036	1.052868	0.8175	0.1825

Table D2: Alignment-free and statistically mirrored multiple comparisons of nucleotide base sequences

	Maximalic Genomic Mirror			Minimalic Genomic Mirror		
Seq.	**K_c-value**	**$P_{Sim.}$**	**$P_{Dsim.}$**	**K_c-value**	**$P_{Sim.}$**	**$P_{Dsim.}$**
D_1	0.932985	0.7488	0.2512	1.002481	0.9901	0.0099
D_2	0.942275	0.7817	0.2183	1.014398	0.9448	0.0552
D_3	0.939257	0.7709	0.2291	1.010993	0.9574	0.0426
D_4	0.945718	0.7941	0.2059	1.019335	0.9269	0.0731
D_5	0.952995	0.8204	0.1796	1.028563	0.8948	0.1052
D_6	0.952918	0.8201	0.1799	1.028896	0.8936	0.1064
D_7	0.952056	0.817	0.183	1.028459	0.8951	0.1049
D_8	0.957246	0.836	0.164	1.035089	0.873	0.127
D_9	0.958588	0.8409	0.1591	1.03742	0.8654	0.1346
D_{10}	0.964942	0.8645	0.1355	1.044915	0.8417	0.1583
D_{11}	0.967368	0.8736	0.1264	1.048249	0.8314	0.1686
D_{12}	0.96934	0.881	0.119	1.051437	0.8218	0.1782

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Table D3: Alignment-base and statistically mirrored multiple comparisons of nucleotide base sequences

	Maximalic Genomic Mirror			Minimalic Genomic Mirror		
Seq.	**K_c-value**	**$P_{Sim.}$**	**$P_{Dsim.}$**	**K_c-value**	**$P_{Sim.}$**	**$P_{Dsim.}$**
D_1	0.932985	0.7488	0.2512	1.002481	0.9901	0.0099
D_2	0.939133	0.7705	0.2295	1.010134	0.9607	0.0393
D_3	0.939257	0.7709	0.2291	1.010993	0.9574	0.0426
D_4	0.948065	0.8025	0.1975	1.022534	0.9156	0.0844
D_5	0.952995	0.8204	0.1796	1.028563	0.8948	0.1052
D_6	0.950553	0.8115	0.1885	1.025668	0.9047	0.0953
D_7	0.954965	0.8276	0.1724	1.032445	0.8817	0.1183
D_8	0.957246	0.836	0.164	1.035089	0.873	0.127
D_9	0.955657	0.8302	0.1698	1.033395	0.8786	0.1214
D_{10}	0.967368	0.8736	0.1264	1.048249	0.8314	0.1686
D_{11}	0.967368	0.8736	0.1264	1.048249	0.8314	0.1686
D_{12}	0.966893	0.8718	0.1282	1.048066	0.832	0.168

Table D4: Alignment-free and statistically mirrored multiple comparisons of nucleotide base sequences

	Maximalic Genomic Mirror			Minimalic Genomic Mirror		
Seq.	**K_c-value**	**$P_{Sim.}$**	**$P_{Dsim.}$**	**K_c-value**	**$P_{Sim.}$**	**$P_{Dsim.}$**
L_1	0.964108	0.8614	0.1386	1.040904	0.8543	0.1457
L_2	0.961744	0.8526	0.1474	1.041271	0.8531	0.1469
L_3	0.956555	0.8335	0.1665	1.037671	0.8646	0.1354
L_4	0.947989	0.8022	0.1978	1.029839	0.8905	0.1095
L_5	0.935768	0.7586	0.2414	1.017064	0.9351	0.0649
L_6	0.921528	0.709	0.291	1.000976	0.9961	0.0039
L_7	0.905666	0.6552	0.3448	0.982047	0.9295	0.0705
L_8	0.886732	0.593	0.407	0.958555	0.8408	0.1592
L_9	0.863331	0.5191	0.4809	0.928608	0.7335	0.2665
L_{10}	0.836296	0.4373	0.5627	0.893175	0.6139	0.3861
L_{11}	0.809653	0.3604	0.6396	0.857603	0.5014	0.4986
L_{12}	0.77832	0.2742	0.7258	0.815186	0.3761	0.6239

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 December 2019 doi:10.20944/preprints201912.0069.v1

Table D5: Alignment-base and statistically mirrored multiple comparisons of nucleotide base sequences

	Maximalic Genomic Mirror			Minimalic Genomic Mirror		
Seq.	**K_c-value**	**$P_{Sim.}$**	**$P_{Dsim.}$**	**K_c-value**	**$P_{Sim.}$**	**$P_{Dsim.}$**
L_1	0.964108	0.8614	0.1386	1.040904	0.8543	0.1457
L_2	0.956245	0.8323	0.1677	1.033648	0.8777	0.1223
L_3	0.948134	0.8028	0.1972	1.025886	0.9039	0.0961
L_4	0.937459	0.7646	0.2354	1.014955	0.9427	0.0573
L_5	0.9234	0.7154	0.2846	0.999421	0.9977	0.0023
L_6	0.906218	0.657	0.343	0.978995	0.9177	0.0823
L_7	0.888274	0.598	0.402	0.956967	0.835	0.165
L_8	0.870169	0.5404	0.4596	0.934584	0.7544	0.2456
L_9	0.836476	0.4378	0.5622	0.889863	0.6032	0.3968
L_{10}	0.812996	0.3698	0.6302	0.859572	0.5075	0.4925
L_{11}	0.786426	0.296	0.704	0.824251	0.4021	0.5979
L_{12}	0.75183	0.2047	0.7953	0.777512	0.272	0.728