Original Article

OPEN ACCESS

Comparative Genomics Unveils Parallel Recognition of DNA Regulatory Motifs: Insights from Langya Virus

Venu Paritala *[1,2]

[1]Department of Management with Data Analytics, Indiana Wesleyan University, Marion, Indiana, USA
[2]Department of BioTechnology, Vignan's Foundation for Science, Technology and Research, Guntur, Andhra Pradesh, India.

*Corresponding author: Venu Paritala; *vvenuparitala@gmail.com*

Received: 19 May 2024; **Revised:** 29 June 2024; **Accepted:** 05 July 2024; **Published:** 12 July 2024

Abstract

Motivation: The intricate tango between transcription factors and their target genes orchestrates the symphony of cellular life. The functional roles of TFs are intricately tied to the genes under their regulation. Unraveling these roles not only elucidates the genomic and transcriptional landscape of the specific genes and TFs under investigation but also situates them within the comprehensive context of the entire regulatory network. ***Results:*** In this study, we present a novel alignment-free and threshold-independent comparative genomics approach for assigning functional roles to DNA regulatory motifs specifically geared towards prokaryotic gene prediction. This approach, integrated into the Gomo and GeneMarkS algorithms, leverages cross-species information to identify Gene Ontology (GO) terms associated with the target genes of transcription factors (TFs). Incorporating two comparative species for Langya virus (LayV) and Paramyxovirus analysis significantly enhances Gomo's ability to predict GO terms, providing deeper biological insights into TF function. To mitigate false positives, we adjust motif affinity scores based on individual sequence composition through a novel sequence-scoring algorithm that refines thermodynamic binding predictions. Notably, Gomo's accuracy remains robust to promoter definition and requires no parameter tuning due to its threshold-free gene set analysis. This empowers biologists to explore the potential roles of specific DNA regulatory motifs and predicted genes through Gomo (http://meme.nbcr.net) and GeneMark web software (http://opal.biology.gatech.edu/GeneMark/).

Keywords: ***Gene Ontology, Langya Virus, Paramyxovirus, motifs and GeneMark.***

1. Introduction

The intricate landscape of transcription factor (TF) role prediction unveils a multifaceted exploration aimed at refining our understanding of gene regulation. At the forefront is the Gomo algorithm [1], offering valuable insights by reporting significant rank sum P-values for Gene Ontology (GO) terms, meticulously adjusting for multiple tests. Within this context, the PWM-based function used by Gomo in scoring promoter regions comes under scrutiny. Our investigation extends into the delicate balance between computational efficiency and scoring accuracy, specifically addressing the significance of GC content compensation in binding affinity scores. Simultaneously, the integration of comparative genomics emerges as a pivotal approach for advancing TF role prediction algorithms [2]. This strategy leverages the assumption that functional TF binding sites exhibit a degree of conservation in the promoters of orthologous genes across related species. While promising, challenges arise, particularly in accurate multiple alignments of orthologous genes and the conservation of TF binding site characteristics. Despite these challenges, success stories in phylogenetic footprinting and motif modeling underscore the potential of evolutionary information to enhance our understanding of TF regulatory networks.

A deeper layer of sophistication is introduced through the consideration of gene set enrichment analyses and the incorporation of additional information such as chromatin structure and epigenetic modifications. The integration of diverse data sources promises a more comprehensive view of TF binding dynamics and regulatory activities, contributing to the continuous refinement of computational models. In this evolving landscape, advancements in experimental techniques play a crucial role [3]. Cutting-edge approaches like single-cell genomics and chromatin conformation capture technologies provide a detailed characterization of TF binding dynamics and the three-dimensional organization of the genome. The synergy between computational predictions and experimental findings holds the key to unraveling the complexities of TF regulatory networks with unprecedented precision. As the field progresses, interdisciplinary collaborations become paramount. The convergence of computational methodologies, insights from comparative genomics, and experimental advancements promises a more accurate and nuanced understanding of TF functions in cellular processes. The journey to decode the regulatory orchestration of gene expression enters a new phase, marked by innovation, collaboration, and an unwavering commitment to unraveling the intricacies of transcriptional control.

Langya virus (LayV), an enveloped virus, harbors a single-stranded ribonucleic acid (RNA) comprising 18,402 RNA

nucleotides. Its genomic architecture encompasses six crucial structural proteins: nucleocapsid, phosphoprotein, matrix protein, surface glycoprotein, fusion protein, and large viral RNA-dependent RNA polymerase. The intricate genomic composition establishes its closest genetic affiliation with henipaviruses like Mojiang henipavirus [4]. The comprehensive genomic sequences of LayV are meticulously cataloged in GenBank under accession numbers OM101125-OM101130 and OM069567-OM069646. As previously indicated, LayV predominantly disseminates among animals, particularly shrews. RNA from LayV has been extracted from over 25% of nearly 260 scrutinized shrews, suggesting the potential role of this species as a direct or indirect vector for human transmission. Notably, LayV RNA has also been detected in other animals, including dogs and goats [5]. Despite the absence of documented human-to-human transmissions among the 35 identified human infections, the substantial prevalence of LayV in shrew populations underscores the urgency for extensive investigations to elucidate potential transmission routes.

This study focuses on predicting regulatory targets for several transcription factors (TFs), leveraging known sets of regulatory targets. The predictions are validated using the enhanced version of Gomo, a computational tool seamlessly integrated with the Meme motif discovery tool (Bailey et al., 2009; http://meme.nbcr.net). The functionality of Gomo has been augmented, and it is now available for download. To enhance the precision of our predictions, we conduct a false discovery rate (FDR) analysis of the associations between predicted TFs and Gene Ontology (GO) terms. This integrated approach aims to provide a comprehensive understanding of the regulatory landscape influenced by LayV. Researchers and investigators can now easily utilize motifs discovered by Meme, seamlessly transferring them to Gomo for in-depth analysis with a simple mouse click. This integrated computational framework represents a powerful tool for unraveling the intricacies of LayV's regulatory mechanisms and their potential implications in the broader biological context.

2. Materials and Methods

2.1 Genetic Composition through Sequence Analysis and Alignment of Langya Virus

In the initial phase of our analysis, we commenced by redefining the nucleotide sequence of Langya Virus sourced from the National Center for Biotechnology Information (NCBI) database, specifically identified by the LOCUS ID OM101130, encompassing a total of 18,402 base pairs in its complementary RNA (cRNA) sequence [6,7]. As a foundational step, the nucleotide sequence underwent scrutiny through the Basic Local Alignment Search Tool (BLAST) to discern sequence similarities [8]. Notably, our analysis revealed that Paramyxovirus exhibited the highest sequence similarity when compared to Langya Virus (**Table 1, Figure 2**). Subsequently, Paramyxovirus, with GenBank accession number MT063641.1, was selected as the second species for further analysis [9], focusing on gene location and motif exploration. This strategic choice aims to leverage the genetic proximity identified through sequence similarity for a more comprehensive understanding of Langya Virus and its genomic characteristics.

The GeneMark web platform provides a comprehensive suite of software tools, including GeneMark and GeneMark.hmm, specifically designed for gene prediction across prokaryotic, eukaryotic, and viral genomic sequences. This user-friendly interface allows the analysis of nearly 200 prokaryotic and over 10 eukaryotic genomes, offering species-specific versions of the software and precomputed gene models. It facilitates efficient gene prediction, enabling the identification of genes in prokaryotic sequences from novel genomes through on-the-spot model derivation, employing either a heuristic approach or the comprehensive self-training program GeneMarkS. The website also hosts a database featuring reannotations of more than 1000 viral genomes by the GeneMarkS program, with regular updates to ensure access to the latest software versions and gene models. The GeneMark.hmm web interface, tailored for single DNA sequence analysis, accepts input either as an uploaded file or as text pasted into a designated textbox. Providing flexibility, it allows users to select the species name and opt for a model for the Ribosomal Binding Site (RBS) or other genetic codes. This integrated interface streamlines gene prediction, offering customization based on input sequence characteristics and desired analysis parameters (**Figure1**).

2.2 Evaluation methods and datasets

We investigate the efficacy of our prediction method, Gomo, in accurately discerning associations between a transcription factor (TF)'s target genes and Gene Ontology (GO) functional categories. Our focus centers on three distinct species: Langya Virus, Paramyxovirus, and an additional species. For each species, we employ GO annotations, promoter sequences, and promoter sequences from three other species, alongside a species-specific set of TF binding motifs. To assess prediction accuracy, we establish two reference sets of TF-GO associations based on known targets for the motifs in Langya Virus (LayV) and Paramyxovirus, respectively (**see Supplementary Material Figure 3**).

2.1.1 Evaluating Performance through Established TF–GO Term Associations

To establish reference sets of TF-GO associations for E. coli and S. cerevisiae, we implement the methodology outlined in our prior investigation (Reference 12). In this process, we begin by compiling a collection of confirmed gene targets for transcription factors (TFs) in each organism. The known target sequences of TFs are sourced from the National Center for Biotechnology Information (NCBI).

We evaluate the predictive accuracy of Gomo using reference sets of TF-GO associations for E. coli and S. cerevisiae, following a previously established methodology. Designating TF–GO pairs in the reference set as 'positives' and all others as 'negatives,' we employ the AUC50 metric, emphasizing accuracy within the top predictions crucial for biologists. AUC50 calculates the Area Under the Curve up to the 50th false positive, sorting TF–GO terms by increasing St score [13]. The metric's relevance lies in highlighting accuracy nuances in the concise list of the most confident predictions. For each TF in a species with at least one TF-GO pair in the reference set, we compute the AUC50 value, indicating Gomo's performance in distinguishing 'positive' and 'negative' GO terms. The species' overall accuracy is determined by averaging the AUC50 values for TFs with GO terms in the reference set.

2.2.2 Evaluation using FDR

To enhance the assessment of Gomo's predicted TF-GO term associations, we conduct a False Discovery Rate (FDR) analysis. This analysis provides an estimate of the fraction of predicted associations that are statistically significant, though it does not guarantee their biological relevance. FDR analysis has been widely employed in scenarios with incomplete or noisy validation sets, serving as a valuable tool for assessing the accuracy of TF role predictions (Reference 14) and TF binding site predictions (Reference 15). As outlined in Section 2.1, Gomo computes q-values for all TF-GO term associations based on their empirical P-values.

The q-value of a TF-GO pair signifies the minimum FDR at which the association would be deemed significant. We report the number of associations detected at a q-value of 0.05. In computing q-values, we aggregate the P-values of all TF-GO pairs for a single organism across all TFs used as queries to adjust for multiple tests. As an additional validation, we ensure that no significant predictions are generated when input sequences are permuted, providing further confidence in our FDR estimates.

3. Results

The results from GeneMark.hmm are presented in Table 7, offering a comprehensive overview of predicted genes. This includes information on the gene's strand orientation, boundaries, length in nucleotides, and its classification into Typical or Atypical gene models. The gene class designation signifies which of the two Markov chain models demonstrated a higher likelihood for the respective gene sequence. Genes categorized as Typical exhibit codon usage patterns consistent with the majority of genes in the species, while Atypical class genes may deviate and often encompass a substantial number of laterally transferred genes. The output further provides nucleotide sequences of the predicted genes and their translated protein sequences, facilitating subsequent analyses such as BLAST searching. Notably, an option for parallel generation of GeneMark predictions is available, utilizing models derived from the same training data as those applied in the current run of GeneMark.hmm, thereby offering valuable additional insights into the predicted gene landscape.

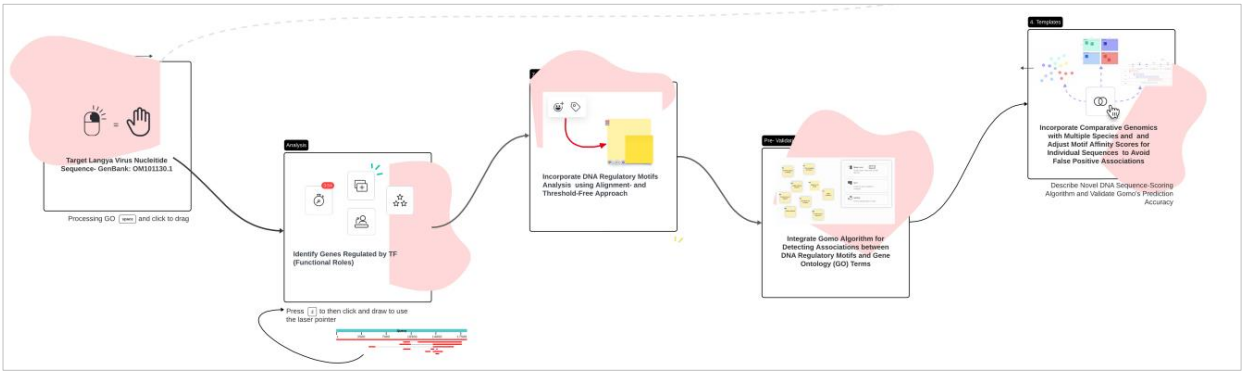


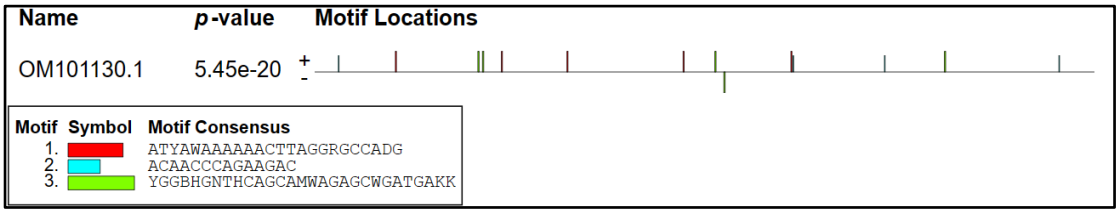
Figure 1: Workflow for the Prediction and Selection of Motifs in Single-Species Gomo (Langya Virus Species)

The GeneMark.hmm and GeneMark algorithms, although distinct, function complementarily, akin to the relationship between the Viterbi algorithm and the posterior decoding algorithm. Their predictions are expected to largely corroborate and validate each other, with discrepancies often signaling potential issues such as sequence errors or deviations in gene organization. The graphical output, available in PDF or PostScript format, illustrates this synergy. Figure 4, a fragment of this output, highlights the advantage of employing multiple Markov chain models representing different gene classes. The coding potential graph, derived from the Typical gene model by GeneMarkS, is depicted by a solid black line, while the Atypical gene model, derived through a heuristic approach, is represented by a dotted line. Notably, the graphical representation demonstrates the ability of the GeneMark programs to detect genes from both Typical and Atypical gene classes. Instances where genes were exclusively detected by the Atypical model, such as the one located between positions 400 to 2000, underscore the utility of this dual-algorithm approach. The graph also provides indications of frameshift positions, often indicative of sequencing errors but occasionally revealing natural and biologically intriguing phenomena. The overall analysis enhances the understanding of gene prediction and organization, particularly when leveraging diverse algorithmic approaches.

Table 1: Presents the results of Basic Local Alignment Search Tool (BLAST) analysis conducted on the Langya Virus. The purpose of this analysis is to identify the closest species for further investigation of the MEMO motif and gene location prediction in the complete genome sequence of isolate SDQD_S1801.

Scientific Name	Taxid	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Langya virus	2971765	33983	33983	100%	0	100	18402	OM101130.1
Wenzhou shrew henipavirus 1	3084931	2523	3017	38%	0	74.32	18426	OQ715593.1
Wenzhou Apodemus agrarius henipavirus 1	2877509	2067	2583	30%	0	75.94	18309	MZ328275.1
melian virus	2940995	1441	1944	21%	0	75.73	19944	OK623353.1
Langya virus	2971765	944	944	2%	0	100	511	OM069599.1
Jingmen Crocidura shantungensis henipavirus 1	2928971	472	472	7%	2E-126	72.97	18535	OM030314.1
Wenzhou shrew henipavirus 1	3084931	425	425	5%	2E-112	74.08	18425	OQ715594.1
Crocidura tanakae henipavirus	3049973	320	320	4%	8.00E-81	74.67	18480	OQ970176.1
Paramyxovirus PREDICT_PMV-168	2711084	243	243	2%	2.00E-57	75.43	530	MT063669.1
Paramyxovirus PREDICT_PMV-168	2711084	237	237	2%	9.00E-56	75.24	530	MT063674.1
Paramyxovirus PREDICT_PMV-13	1647213	134	134	1%	1.00E-24	79.58	529	MT063566.1
Paramyxovirus PREDICT_PMV-13	1647213	134	134	1%	1.00E-24	79.58	511	MT063532.1
Paramyxovirus PREDICT_PMV-13	1647213	134	134	1%	1.00E-24	79.58	478	MT063529.1
Paramyxovirus PREDICT_PMV-13	1647213	134	134	1%	1.00E-24	79.58	449	MT063508.1
Paramyxovirus PREDICT_PMV-13	1647213	134	134	1%	1.00E-24	79.58	485	MT063492.1

Figure 2: Depicts the sequence OM101130.1 of Langya virus, specifically the complete genome isolate SDQD_S1801. The total length of the matched sequence for Langya Virus is 18,402 nucleotides.



A)

B)

Figure 3: Motif location with *p*-value of Langya and Paramyxoviruses Could Pose adjectives to make it more engaging.

Table 2: LayV motif ATYAWAAAAAACTTAGGRGCCADG MEME-1 sites sorted by position p-value

Sequence name	Strand	Start	P-value	Site
OM101130.1	+	4407	1.98E-14	GATTACTGGT
OM101130.1	+	5955	8.78E-13	TTTAAAGATT
OM101130.1	+	1905	1.23E-12	ATACTAATTT
OM101130.1	+	8697	1.52E-12	CTTCACACTG
OM101130.1	+	11242	1.55E-11	TCATTGATCA

Table 3: LayV Motif ACAACCCAGAAGAC MEME-2 sites sorted by position p-value

Sequence name	Strand	Start	P-value	Site
OM101130.1	+	17569	1.26E-08	CATAGTGTAG
OM101130.1	+	13451	1.26E-08	TGCAAAACAA
OM101130.1	+	11292	1.26E-08	CTCGGAAACA
OM101130. 1	+	559	1.26E-08	AATCTCTTCT

Table 4: LayV Motif YGGBHGNTHCAGCAMWAGAGCWGATGAKK MEME-3 sites sorted by position p-value

Sequence name	Strand	Start	P - value	Site
OM101130.1 TAATTGACAT	+	3961	3.17e - 15 AGCCTGTCAT	CGGGCGGTCCAGCACAAGAGCAGAAGAGG
OM101130.1 GTTACACTAA	+	9446	5.48e - 13 TACCCAACCT	TGGTCCATCCAGCTCTAGTGCTGATGAGT
OM101130.1 GCATCTATGA	+	3852	6.81e - 12 AGATTAGGGC	TGGCTGTTTCATCAATGGAGCAGATGCTG
OM101130.1 TTTGGATTTG	-	9659	7.29e - 12 GTCTCGTCTC	CAGCTGCGACAGCACACGAGTTGATGATT
OM101130.1 ACCGGTCTGT	+	14866	9.51e - 12 ATATCAGGAA	TGCTAGATACAACAAAAGGGCTGATAAGG

Table 5: Paramyxovirus Motif MTTVCMTCAAMATTNGC MEME-1 sites sorted by position p-value

Sequence name	Strand	Start	P - value	Site
MT063641.1	+	501	6.82e - 09 GGTGGTTTCA	ATTACCTCAACATTTGC AGACTTTATG
MT063641.1	+	345	3.39e - 08 TGGATTGGGT	ATTGCATCAACATTCTC AAGGTACTTG
MT063641.1	+	279	9.96e - 08 ACAAGATCAG	CGTGCAGTAACATTAGC ACTGCCATAG
MT063641.1	+	62	1.82e - 07 GAAGAAATAC	CTTAGCACAAAATTAGC ACAAGAATAT
MT063641.1	+	476	2.39e - 07 TGCTAGCATT	ATTCCATCACAATTGGG TGGTTTCAAT

Table 6: Paramyxovirus Motif CWGCCA MEME-2 sites sorted by position p-value

Sequence name	Strand	Start	P - value	Site
MT063641.1	+	297	1.92e - 04 AACATTAGCA	CTGCCA TAGCAAAATC
MT063641.1	+	157	1.92e - 04 CAATTGTTAG	CAGCCA CTTCTTCGTG

Table 6: Paramyxoviru Motif ACAAGG MEME-3 sites sorted by position p-value

Sequence name	Strand	Start	P - value	Site
MT063641.1	+	319	1.81e - 04 AATCAGTTGA	ACAAGG TTTTTCTAGA
MT063641.1	+	17	1.81e - 04 AATTGCAATA	ACAAGG AGAGTTCATC

Table 7: Gene predictions made by the prokaryotic version of GeneMark.hmm for a fragment of the Langya Virus SDQD_S1801 genome

Gene	Strand	Left End	Right End	Gene Length	Class
1	+	132	1751	1620	2
2	+	1935	2756	822	2
3	+	2753	4273	1521	2
4	+	4530	5492	963	2
5	+	6181	6468	288	2
6	+	6900	8534	1635	2
7	+	8820	10694	1875	2
8	+	11478	18311	6834	2
9	I	18313	> 18402	90	1

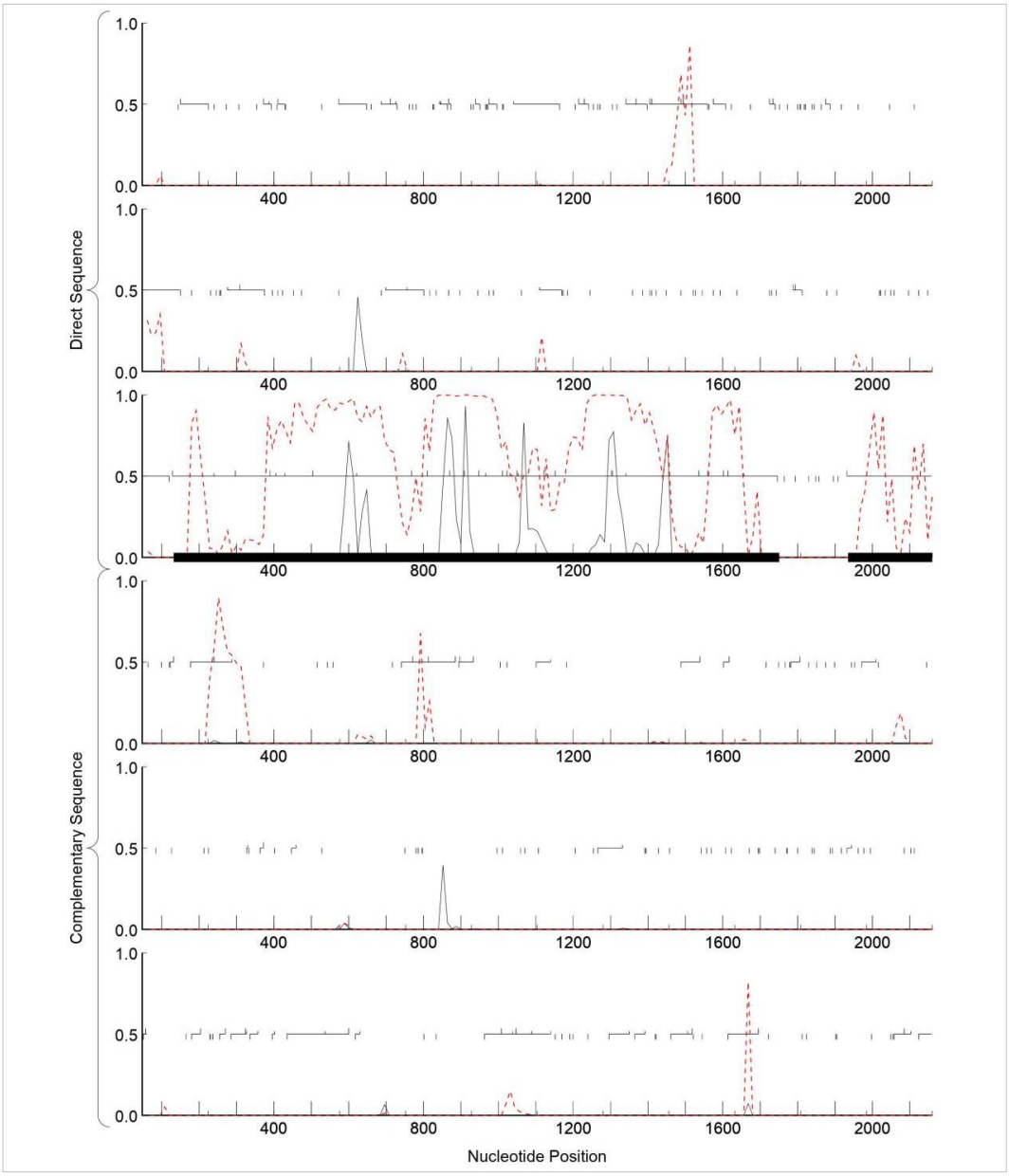


Figure 4: The graphical output resulting from the combination of GeneMark and GeneMark.hmm for a segment of the Langya Virus SDQD_S1801 complete genome is depicted. The coding potential, assessed by the GeneMark program using Typical and Atypical Markov chain models of coding DNA, is represented by solid black and dashed traces, respectively, for the three reading frames on the direct strand. Notably, no genes, either predicted or annotated, are observed on the reverse strand in this specific genome section. The figure also highlights predictions made by GeneMark.hmm with thick black horizontal bars, 'regions of interest' indicated by thick grey bars, and the (longest) Open Reading Frames (ORFs) in each reading frame denoted by thin black horizontal lines, with ticks above and below suggesting potential start and stop codons, respectively.

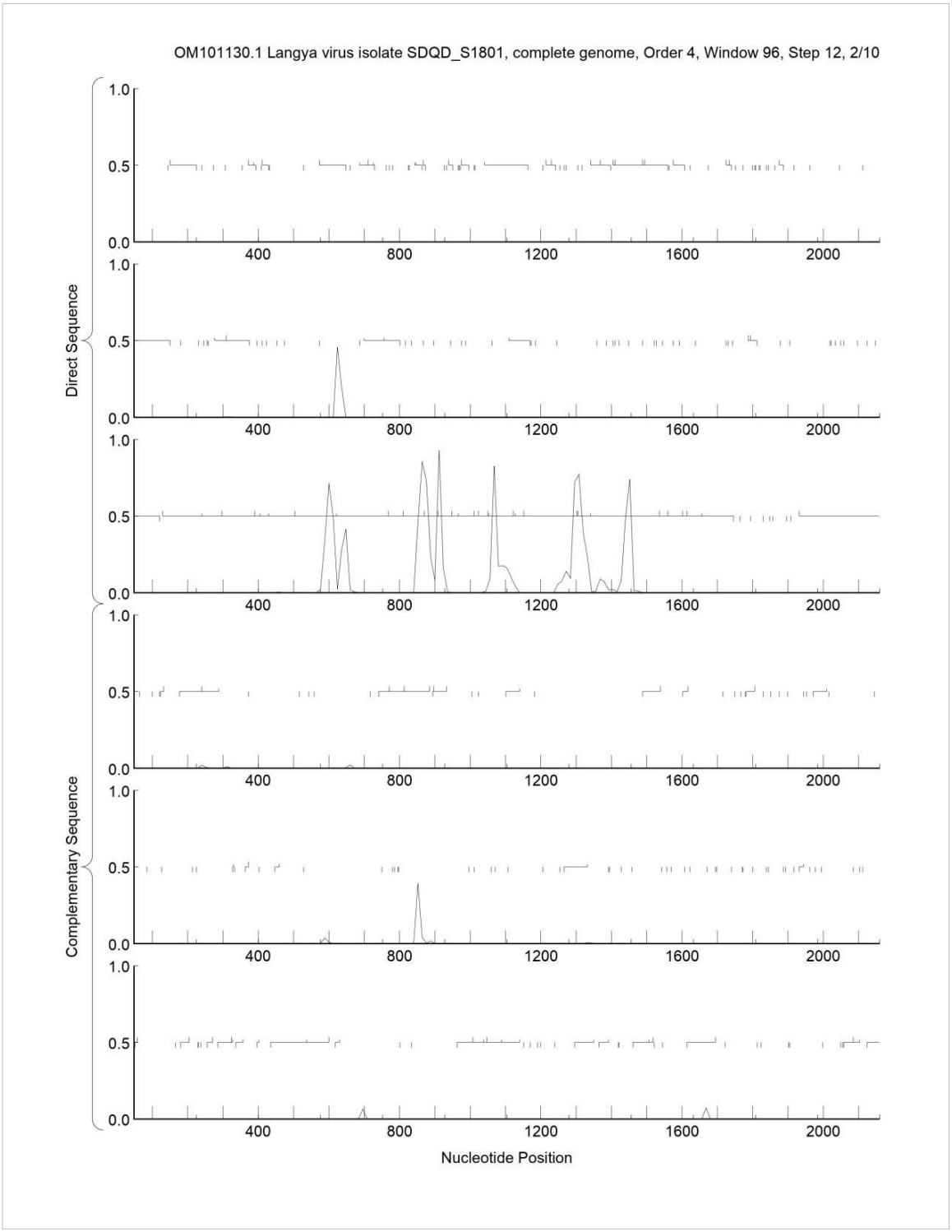


Figure 5: Figure 5 illustrates the representation of coding regions for both strands using Markov chain and Bayes methods. The indicator function obtained through a second-order approach for both strands integrates six charge representations, providing a comprehensive depiction of protein-coding regions in a unified manner.

Within the GeneMark program, various specific options are available for customization. Default settings for window size (96 nt) and step size (12 nt) determine the dimensions of the sliding window and its incremental movement along the sequence. The threshold parameter sets the minimum average coding potential required for an open reading frame (ORF) to be predicted as a gene. Additional options fine-tune graphical output, and there are features supporting the analysis of eukaryotic DNA sequences, such as the provision of putative splice sites and protein translations for predicted exons. Notably, the GeneMark posterior decoding algorithm doesn't provide precise exon–intron border predictions; for this purpose, the eukaryotic version of GeneMark.hmm, utilizing the generalized Viterbi algorithm, serves as the primary tool. The program output includes a list of predicted genes, each potentially having more than one start, with accompanying data aiding the annotation of the 'true' start. Start probability, derived from sequences surrounding potential starts, and Ribosome Binding Site (RBS) information, including probability score, position, and sequence, are provided. Additionally, GeneMark generates a list of 'regions of interest'—significant spans between in-frame stop codons with noteworthy coding potential spikes, suggesting areas for further analysis even if no genes are automatically predicted based on the set threshold.

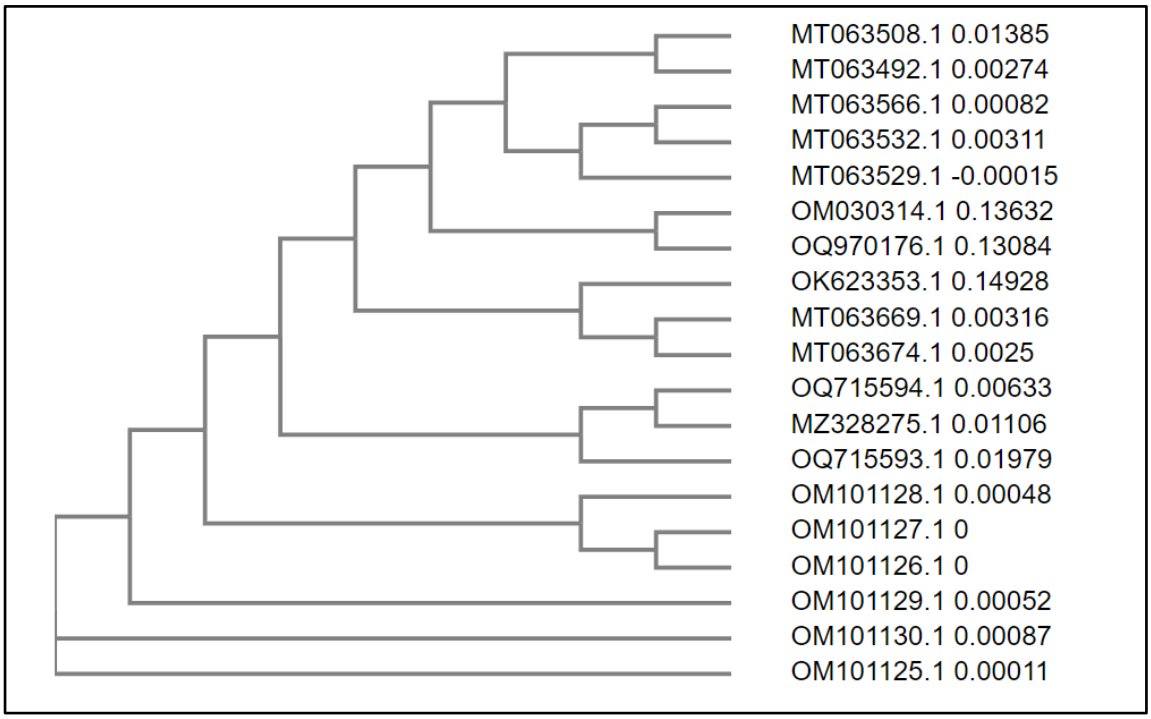


Figure 6. illustrates the representation of Langya Virus

2.2 Successful transfer of GO annotation to related species

The approach to integrating comparative genomics into transcription factor (TF) role prediction relies on the assumption that Gene Ontology (GO) annotations remain reliable when transferred from a gene in the primary species to its ortholog in another species. The validity of this assumption is supported by the results presented in Table 2, 3, 4, 5, and 6, which are derived from Gomo analysis using promoters from a single species, chosen as the key species for LayV and Paramyxovirus. Across various size definitions of promoter regions, the accuracy of TF–GO term associations predicted by Gomo for related species remains comparable to that of the key species. Notably, for LayV, the accuracy is marginally higher when using promoters from its related species (as depicted in Fig. 2). The E. coli reference set, comprising 87 TF–GO term pairs, yields accuracy measurements based on a relatively small sample. In contrast, the S. cerevisiae reference set (503 TF–GO term pairs) demonstrates similar prediction accuracy using Paramyxovirus promoters compared to promoters from two of its related species. Interestingly, employing promoters from Hernipa Virus yields lower accuracy than those from Mojinga Virus, despite Hernipa Virus being evolutionarily closer to Langya Virus based on multiple alignments of all orthologous intergenic regions. It's important to note that the phylogenetic trees in Figure 6 are provided for reference only, as our method does not utilize them. We generated the phylogenetic tree for enterobacter from our 1500 bp promoter sequences using Clustal Omega topology.

4. Discussion

We have introduced a comparative genomics strategy for attributing biological roles to sequence motifs, employing a variant of gene set enrichment analysis. Unlike methods requiring multiple alignments, our approach conducts role predictions independently for each comparative genome, and subsequently consolidates the results across genomes. This design avoids assuming the conservation of binding site location and orientation across species. However, orthologous gene identification in the relevant species is a prerequisite, as it would be for methods employing multiple alignments. Additionally, our approach assumes the conservation of functions for orthologous genes and the DNA-binding affinity of the regulatory molecule in the species under consideration. While these assumptions may occasionally be violated, our comparative genomics approach significantly enhances the sensitivity of transcription factor role predictions.

Our main finding underscores the considerable enhancement achievable in predicting the linkage between a DNA-binding motif and an annotation term through the transfer of annotations from a key species to related species, followed by the amalgamation of association scores for a given motif and term across species. This method only necessitates the computation of the P-value for the motif-term association in each species, and the combination across species is accomplished by computing the geometric mean of the P-values. Significance values for the combined motif-term score are subsequently assigned through a straightforward permutation test. Importantly, our alignment-free approach mitigates issues that might arise from imperfect alignments or 'motif drift' [11].

Declarations

Ethical Approval

Not Applicable

Contributions

Not Applicable

Conflict of Interest

There are no competing interests.

Funding Statement

None

References

[1] Bodén, M. and Bailey, T.L. (2008) Associating transcription factor-binding site motifs with target GO terms and target genes. Nucleic Acids Res., 36, 4108–4117

[2] Ashburner, M. et al. (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet., 25, 25–29.

[3] Buske FA, Bodén M, Bauer DC, Bailey TL. Assigning roles to DNA regulatory motifs using comparative genomics. Bioinformatics. 2010 Apr 1;26(7):860-6. doi: 10.1093/bioinformatics/btq049. Epub 2010 Feb 10. PMID: 20147307; PMCID: PMC2844991.

[4] Zhang XA, Li H, Jiang FC, Zhu F, Zhang YF, Chen JJ, et al. A Zoonotic Henipavirus in Febrile Patients in China. N Engl J Med. 2022; 387:470-2. 10.1056/NEJMc2202705

[5] Choudhary OP, Priyanka, Fahrni ML, Asmaa A, Metwally AA, Saied AA. Spillover zoonotic 'Langya virus': is it a matter of concern? Vet Q. 2022; 42:172-4. 10.1080/01652176.2022.2117874

[6] Chakraborty S, Chandran D, Mohapatra RK, Islam MA, Alagawany M, Bhattacharya M, Chakraborty C, Dhama K. Langya virus, a newly identified Henipavirus in China - Zoonotic pathogen causing febrile illness in humans, and its health concerns: Current knowledge and counteracting strategies - Correspondence. Int J Surg. 2022 Sep; 105:106882. doi: 10.1016/j.ijsu.2022.106882. Epub 2022 Sep 6. PMID: 36075552; PMCID: PMC9444302.

[7] Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022 Jan 7;50(D1): D20-D26. doi: 10.1093/nar/gkab1112. PMID: 34850941; PMCID: PMC8728269.

[8] Mark Johnson, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, Thomas L. Madden, NCBI BLAST: a better web interface, Nucleic Acids Research, Volume 36, Issue suppl_2, 1 July 2008, Pages W5–W9, https://doi.org/10.1093/nar/gkn201

[9] Paritala, Venu, and Harsha Thummala. "Gene Article Analyser: A crucial part of the genome and word data analysis identification from Pub Med articles." International Journal 10.3 (2022).

[10] Moses, A.M. et al. (2006) Large-scale turnover of functional transcription-factor binding sites in Drosophila. PLoS Comput. Biol., 2, e130.

[11] Bodén, M. and Bailey, T.L. (2008) Associating transcription factor-binding site motifs with target GO terms and target genes. Nucleic Acids Res., 36, 4108–4117

[12] Gribskov, M. and Robinson, N.L. (1996) Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching. Comput. Chem., 20, 25–33.

[13] Sinha, S. et al. (2008) Systematic functional characterization of cis-regulatory motifs in human core promoters. Genome Res, 18, 477–488.

[14] Kheradpour, P. et al. (2007) Reliable prediction of regulator targets using 12 Drosophila genomes. Genome Res., 17, 1919–1931