HIDDEN MARKOV MODELLING FOR IDENTIFYING BRONCHIAL ASTHMA BY MUTATED ORMDL3 GENE

K. Senthamarai Kannan And S. D. Jeniffer*

Abstract

Bronchial asthma is a complex respiratory disorder influenced by genetic and environmental factors. Among the genetic factors, the mutated ORMDL3 gene has been associated with an increased susceptibility to asthma. This study explores the potential application of Hidden Markov Modelling (HMM) techniques for identifying bronchial asthma based on the presence of the mutated ORMDL3 gene. HMM is a statistical modelling approach widely used for analysing sequential data with hidden states. In the context of bronchial asthma, HMM can provide a framework to uncover hidden patterns and transitions associated with the mutated ORMDL3 gene, aiding in the identification and classification of asthma cases. This study aims to develop an HMM-based model that integrates genomic data related to the ORMDL3 gene mutation. By considering the hidden states associated with asthma status and incorporating observed variables, such as genetic markers and patient characteristics, the HMM model will identify and classify individuals at risk of bronchial asthma. The proposed HMM model will be trained using available datasets comprising genomic data. The performance of the model will be evaluated using cross-validation techniques to assess its accuracy, sensitivity, and specificity in identifying bronchial asthma cases associated with the mutated ORMDL3 gene.

Keywords: *Gene Mutation, Hidden Markov Model, Markov Chain, Multiple Sequence Alignment, ORMDL3 gene, Performance Accuracy.*

*Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli – 627 012.
Email :senkannan2002@gmail.com , jenifferdavid1996@gmail.com

***Corresponding Author:** K. Senthamarai Kannan And S. D. Jeniffer
*Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli – 627 012.
Email :senkannan2002@gmail.com , jenifferdavid1996@gmail.com

DOI: 10.48047/ecb/2023.12.si5a.0590

1. Introduction :

The Human Genome Project estimates that the human body has between 20,000 and 25,000 genes. Chromosomes include genes, which are DNA segments that encode different types of proteins. Gene nucleotides include Adenine (A), Thymine (T), Guanine (G), and Cytosine (C). In the human body, each gene is copied twice. Each gene is given a distinct name using a mix of alphabets and digits. One of these is the orosomucoid-like 3 gene. The abbreviation for it is ORMDL3.

Asthma is a chronic respiratory condition characterized by inflammation and narrowing of the airways, which results in recurring episodes of breathlessness, wheezing, coughing, and chest tightness. It is a common condition that affects millions of people worldwide, regardless of age or gender.

The exact causes of asthma are not fully understood, but it is believed to result from a combination of genetic and environmental factors. One of the genetic factors that has been extensively studied in relation to asthma is a gene called ORMDL3 (orosomucoid-like 3). Triggers such as allergens (e.g., dust mites, pollen, pet dander), air pollution, respiratory infections, exercise, and certain medications can lead to asthma symptoms in susceptible individuals.

During an asthma attack or exacerbation, the muscles surrounding the airways tighten, the airway linings become swollen, and excess mucus is produced, causing the airways to become narrow. This constriction makes it difficult for air to flow freely, leading to the characteristic symptoms of asthma.

Diagnosis of asthma involves a thorough evaluation of medical history, symptoms, and lung function tests, such as spirometry. Treatment typically includes a combination of long-term control medications, such as inhaled corticosteroids, which help reduce airway inflammation and prevent symptoms, and quick-relief medications, like short-acting bronchodilators, to provide immediate relief during acute attacks.

ORMDL3 is located on chromosome 17q21, and variations or mutations in this gene have been associated with an increased risk of developing asthma. The specific mechanisms by which ORMDL3 gene mutations contribute to asthma are still not fully understood, but research suggests that it may play a role in airway inflammation and immune responses.

Several genome-wide association studies (GWAS) have identified a strong association between single nucleotide polymorphisms (SNPs) in the ORMDL3 gene and asthma susceptibility. These SNPs have been found to be more common in individuals with asthma compared to those without the condition.

Studies have shown that the ORMDL3 gene is involved in the regulation of immune responses and the production of cytokines, which are signalling molecules that play a role in inflammation. Mutations in ORMDL3 may disrupt the normal functioning of these immune pathways, leading to increased inflammation and susceptibility to asthma.

It is important to note that the ORMDL3 gene mutation is just one of many genetic and environmental factors that contribute to the development of asthma. Asthma is a multifactorial disease, and the interplay between genetic predisposition and environmental exposures, such as allergens or air pollutants, also plays a significant role.

2. Review of Literature

Rabiner (1991) introduced the adaptation of Hidden Markov Models (HMM) from speech recognition to the analysis of biological sequences. The study discusses both the theoretical and practical aspects of this methodology. Eddy et al. (1995) expanded upon Rabiner's work and proposed the application of HMM in the field of computational biology. Their research specifically covers the use of HMM in multiple sequence alignment, homology detection, and the underlying assumptions.

In another study, Hughey et al. (1996) provided a detailed explanation of the methodology for determining Hidden Markov Models in computational biology. They utilized the SH2 domain as an example to explore the mathematical extensions and heuristics of HMM. The researchers classified and distinguished three types of Markov Models: conventional Markov Model, Hidden Markov Model, and Profile Hidden Markov Model. They also discussed the advantages and disadvantages of each model. Schuster et al. (2007) further investigated these models in their research work, elaborating on the classification, distinctions, and drawbacks of each model.

To address the analysis of biological sequences, Profile Hidden Markov Models were proposed, incorporating three key principles of modelling. Eddy et al. (1998) focused on software packages, libraries, and the implementation of Profile HMM along with the general HMM for better comprehension and application.

A proposed approach called gene-HMM has been introduced to enhance existing resources for modeling pre-mRNA and predicting gene sequences. Stanke et al. (2006) discussed the parametric approach and consensus profile of amino acids in gene-HMM. They also addressed the HMMER and SAM packages utilized for this purpose.

Multiple Sequence Alignment (MSA) is a fundamental requirement for sequence analysis. Simossis et al. (2003) provided a comprehensive description of MSA, emphasizing its significance in sequence analysis. They explained that MSA offers a structured representation that allows for a coherent examination of sequence attributes by aligning corresponding residues across multiple sequences. The authors conducted an extensive analysis of various methodologies and approaches for performing MSA. They also explored the identification of homologies and consistent characteristics in biological sequences. The study encompassed HMM notations, analysis, HMM training, MSA, gene prediction, and genetic mapping as key concepts to be investigated.

For profile-profile alignment and achieving multiple sequence alignments, the researchers employed tools such as HMMER and SAM modules. The profiler comparer software package utilized the profile-profile method. Other researchers, namely Boer (2016) and Eddy (1995), have also conducted investigations into multiple sequence alignment.

Yoon (2009) conducted a visual examination of the modifications in Hidden Markov Models (HMM). They investigated profile HMM, csHMM, pairHMM, and profile-csHMM to address various challenges. Machine learning, a rapidly evolving field in computing, has found applications in computational biology. The study explained the utilization of Hidden Markov Models in biological sequence analysis, specifically employing semi-supervised learning. These prior findings and proposed strategies have laid the foundation for diverse research projects in the field, contributing to the treatment of various illnesses and showcasing the innovative ideas of the authors.

Subsequently, numerous researchers have built upon these works by introducing modifications and implementations. In their research, Emdadi et al. (2019) focused on estimating the parameters of HMM while considering the path as unknown. They adapted the Ant Colony Optimization procedure for parameter estimation. Bioinformatics has recently emerged as a powerful tool for analyzing biological sequences and predicting disease outcomes. Karuppusamy et al. (2021) developed and trained an HMM specifically for random sequences. The study employed a probability analysis path approach to select the optimal path. The authors primarily focused on the mathematical foundations of the hidden states Intron and Exon. Hidden Markov Models were the subject of a systematic review conducted by Bhausa Mor et al. in 2021.

MathFeature, a program developed by Robson et al. (2021), was introduced as a software package designed to extract mathematical descriptors from biological sequences. The focus of the software primarily revolves around generating mathematical features.

In the realm of COVID-19 research, Stuti et al. conducted sequence alignment of gene sequences from COVID-19 patients. Denesh Kumar et al. (2020) also concentrated on sequence alignment, specifically aligning six gene sequences related to diabetes with six gene sequences associated with contagious diseases using the Clustal webtool.

Jeniffer et al. (2021) utilized the TP53 gene sequence alignment to determine the Transition Probability Matrix. They generated and trained profile Hidden Markov Models (HMM) using the aligned sequences. The EM algorithm was employed for sequence alignment in their study, and they proposed a novel tool for this purpose.
Proctor et al. (2021) introduced JalView as a tool for viewing aligned sequences. It provides a means to visualize aligned sequences effectively.
Bimal Kumar Sankar (2021) employed an entropy-based analysis to investigate the selection of SAR-CoV regions in biological sequences. Regulatory sequence analysis was conducted on the biological sequences, and Roth et al. (2021) utilized the BaMM service to identify motifs.

In the realm of Hidden Markov Models, Jiefu Li et al. (2021) modified the Baum-Welch algorithm to enhance HMM training.

Sasidharan et al. (2021) utilized PHMM (Profile Hidden Markov Models) to detect malware in ProDroid Android apps, achieving an accuracy rate of 94.6 percent. They applied the features of PHMM to identify and analyze potential malware instances.

Meng et al. (2022) employed the features of PHMM to compare homology in Hidden Publishing Services. Their study focused on investigating and measuring the similarities between different instances of Hidden Publishing Services using PHMM as a tool.

The findings and accomplishments from previous research will be leveraged in the current study, which aims to apply them to the task of disease prediction. The study intends to utilize the knowledge gained from earlier research to develop methods for predicting diseases effectively.

3. Methodology :

Preprocessing : Homogeneity testing using Disparity Index

(i) Calculate $E(D_C)$ for each pair of sequences.

$$E(D_C) = \frac{1}{2} \quad E\left(E\left(\left(\sum_{k=1}^{L} \delta_i^{\,k}\right)^2\right)\right)$$

where $D_C = \frac{1}{2}(x_i - y_i)^2$ is the compound difference between the nucleotides of the first and second sequences taken into account. 'i' refers to the nucleotides.

(ii) Disparity Index $I_D = \frac{1}{2}(x_i - y_i)^2 - N_d$, where N_d is the number of sites of nucleotide difference between two sequences.

(iii) Conclude both the sequences are from same organism if $I_D > 0$.

Preparing : Multiple Sequence Alignment

(i) Calculate $\delta(X,Y) = \max_{x_i \in X, x_j \in Y} d(x_i, x_j)$.

(ii) Align the homogeneous sequences using hierarchical clustering according the dissimilarity measure.

(iii) Construct the phylogenetic tree in the order of sequence-sequence, sequence-profile, and profile-profile.

(iv) Using the phylogenetic tree, align the sequences by progressive alignment.

Modelling :

(i) Construct Markov Chain and Transition Probability Matrix.

(ii) Determine Emission Probability Matrix and find Optimal Path using Viterbi Algorithm.

(iii) Estimate the posterior probabilities by Baum-Welch Algorithm.

4. Results :

The study focuses on analyzing the nucleotide sequences of the mutated ORMDL3 gene. A total of 150 biological sequences are included in the analysis, sourced from various published works and GenBank.

As is commonly known, nucleotide sequences consist of combinations of Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). The first step involves examining the homogeneity of the sequences. To assess homogeneity, a disparity index is calculated.

Each sequence is paired with all the other sequences. The first sequence in each pair is denoted as X, while the remaining sequence is denoted as Y. The number of Adenine nucleotides in sequence X and Y is represented by x_A and y_A, respectively. Similar measurements are made for other nucleotides present in the sequences.

For the initial pair of biological sequences, the values obtained are as follows: $x_A = 345$, $y_A = 352$, $x_C = 293$, $y_C = 285$, $x_G = 315$, $y_G = 322$, $x_T = 241$, $y_T = 256$, and $D_C = 1754.6$. The disparity index (I_D) is calculated as 523.27. Since I_D is greater than 0, it indicates that the considered sequences satisfy the homogeneity condition. The remaining sequence pairs undergo a similar verification process.

The dissimilarity measures for the homogeneous biological sequences are determined, and based on these measures, a phylogenetic tree is constructed. The derived phylogenetic guide tree is then utilized for hierarchical clustering. Multiple sequence alignment is performed based on the obtained clusters. The alignment process involves sequence-sequence alignment, sequence-profile alignment, and profile-profile alignment. A portion of the phylogenetic tree, along with its corresponding dissimilarity measure, is presented below.

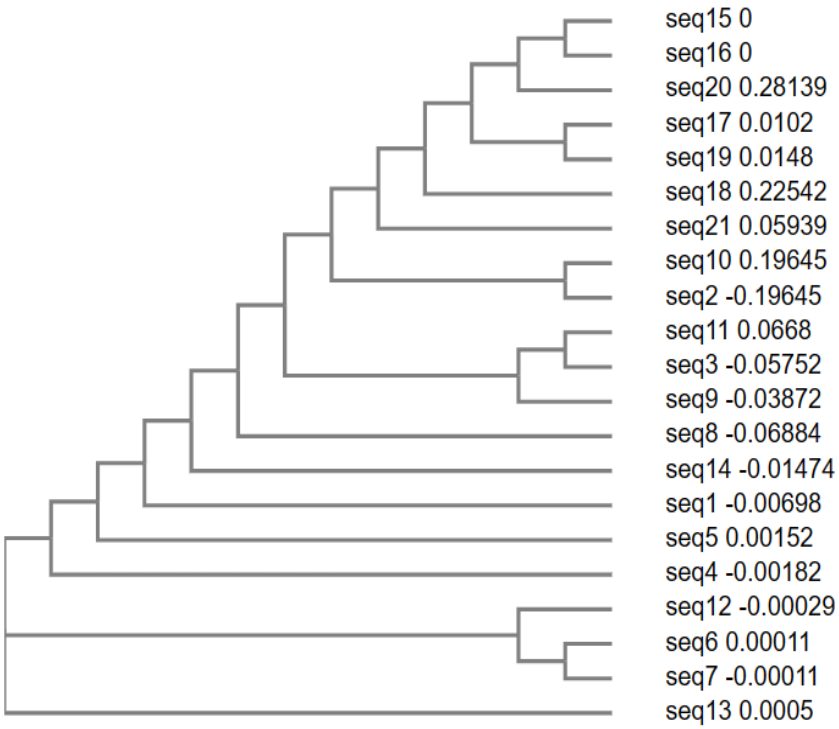

Fig 4.1 : *Phylogenetic tree for multiple sequence alignment*

To represent the 'best fit' and for further analysis, a consensus sequence will be deduced from the aligned sequences. The consensus is mad by the occupancy of nucleotides. It could be visualized, known as consensus logo and shown below.

Fig 4.2 *:Consensus Logo of the aligned sequences*

The Percentage Identity Matrix is as follows.

seq15	100.00	100.00	41.29	34.38	37.70
seq16	100.00	100.00	41.30	34.62	37.46
seq20	41.29	41.30	100.00	36.09	40.54
seq10	34.38	34.62	36.09	100.00	40.12
seq17	37.70	37.46	40.54	40.12	100.00
seq19	37.61	37.39	40.52	38.83	97.50

Construction of Markov Chain :

The procedure of observing and documenting inter-nucleotide and intra-nucleotide transitions enables us to calculate the frequencies of transitions, referred to as Transition frequencies. Following the recognized definition of a Markov Chain, each transition is divided by the total count of transitions in its corresponding row to derive the transition probabilities. The transition probability matrix and the graphical representation of these probabilities are presented in the subsequent sections.

$$P = \begin{pmatrix} 0.2158 & 0.2792 & 0.3842 & 0.1208 \\ 0.2219 & 0.3272 & 0.2587 & 0.1922 \\ 0.1599 & 0.3161 & 0.3382 & 0.1858 \\ 0.1127 & 0.3268 & 0.3725 & 0.1880 \end{pmatrix}$$

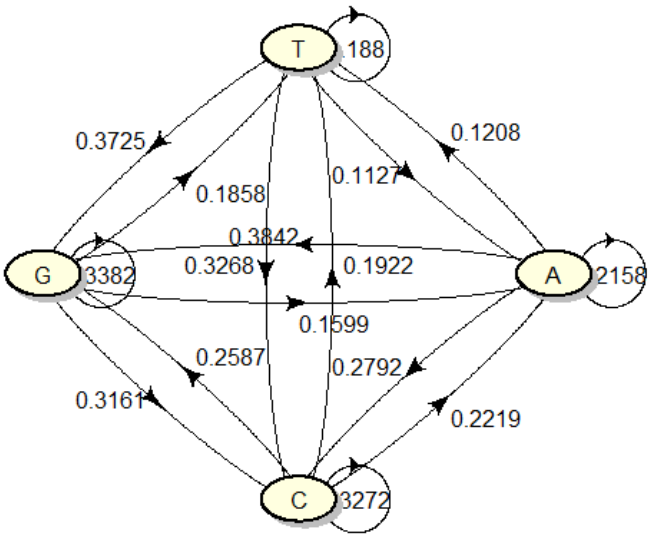

Fig 4.3 : *Transition digraph of Markov Matrix P*

Compositional bias of each nucleotide and dinucleotide for Markov Chain is shown in the below bar diagram.

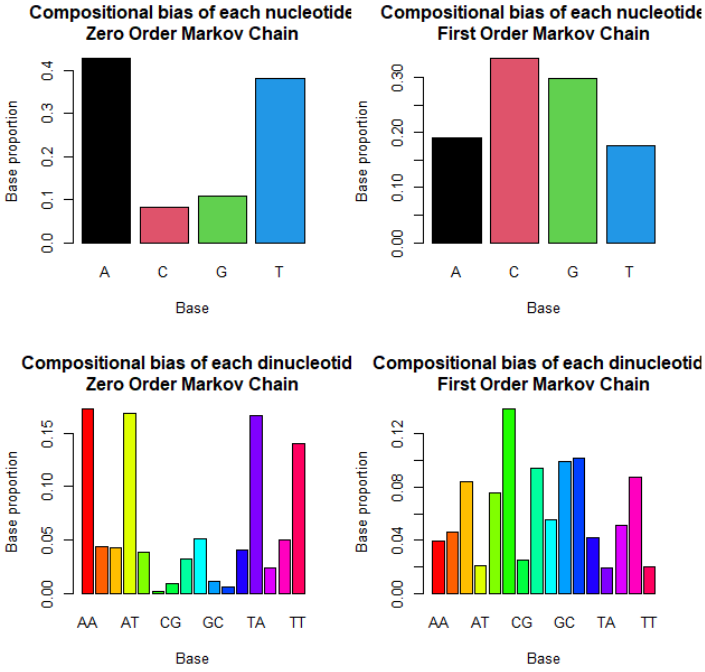

Fig 3.3 *Compositional bias*

Construction of Hidden Markov Model :

In addition to the observable nucleotide state, there exists an undisclosed (latent) state underlying it. These latent states, which include the Match State (M), Insert State (I), and Delete State (D), contribute to the overall sequence. By incorporating both the hidden and visible states, it is possible to create a Hidden Markov Model.

The transition probability matrix according to the hidden states is,

$$T = \begin{array}{c} \\ M \\ I \\ D \end{array} \begin{array}{c} \begin{array}{ccc} M & I & D \end{array} \\ \begin{bmatrix} 0.9532 & 0.0272 & 0.0196 \\ 0.3218 & 0.6285 & 0.0497 \\ 0.0105 & 0.0023 & 0.9872 \end{bmatrix} \end{array}$$

The emission probability matrix for the hidden states and visible states are as follows :

$$E = \begin{array}{c} \\ M \\ I \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{bmatrix} 0.2423 & 0.3215 & 0.2218 & 0.2144 \\ 0.2218 & 0.1215 & 0.3078 & 0.3489 \\ 0.2247 & 0.3280 & 0.2628 & 0.1845 \end{bmatrix} \end{array}$$

The transition diagram of the EP Matrix is,

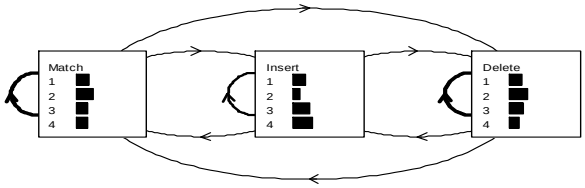

Figure 5. *Transition digraph of the Emission Probability matrix E*

In conjunction with the observable state of nucleotides, there exists a concealed state that accompanies it. These hidden states, namely Match State (M), Insert State (I), and Delete State (D), play a role in forming the sequence. By taking into account both the visible and hidden states, it becomes possible to construct a Hidden Markov Model.

The estimated Optimal Viterbi path for the visible and hidden states is as follows :

	Nucleotides	Best_Path
0	A	D
1	T	D
2	C	M
3	G	I
4	A	I
...	...	...
1250	A	I
1251	C	M
1252	A	I
1253	G	M
1254	A	D

The predicted Viterbi path of the hidden and visible states are,

D D D D D D D M D M M I M D M D D D I
"-" "-" "-" "-" "-" "-" "-" "A" "G" "G" "T" "C" "C" "-" "-" "-" "-" "-" "-" "T"
I I M M I M M M M I M M M D D D M I M M
"A" "C" "T" "G" "G" "T" "A" "A" "A" "T" "T" "G" "C" "-" "-" "I" "T" "T" "C" "A"
M I M I I M I I I D I I D I I M I M M D
"C" "C" "C" "C" "C" "C" "G" "A" "G" "A" "G" "T" "G" "T" "T" "T" "-" "-" "T" "G"
D I I I M M M D M M I I I M I D M I M M
"C" "A" "C" "A" "G" "A" "G" "A" "T" "A" "T" "T" "T" "T" "A" "T" "T" "T" "C" "G"
M M D D M M M M D M
"T" "T" "-" "-" "T" "A" "T" "T" "A" "C"

The Baum-Welch Algorithm is employed to estimate the posterior probabilities. The Hidden Markov Model (HMM) delineates the transitions between the hidden and visible states. Multiple sequence alignment plays a crucial role in revealing significant relationships within a family by capturing pairwise alignment. To identify and assess potential matches for new sequences, a profile Hidden Markov Model is constructed based on the multiple sequence alignment. This extension of the HMM allows for the capture of specific information pertaining to each position in the entire family's multiple alignment.

4. Discussion

The constructed Markov Model illustrates the transition probabilities between nucleotides. For example, the transition probability from Adenine Nucleotide to Adenine Nucleotide is 0.2158. Likewise, the transition probabilities from Adenine Nucleotide to Cytosine, Guanine, and Thymine nucleotides are 0.2792, 0.3842, and 0.1208 respectively. In a similar manner, transitions from Cytosine nucleotide to Adenine, Cytosine, Guanine, and Thymine have probabilities of 0.2219, 0.3272, 0.2587, and 0.1922. Transition probabilities from Guanine nucleotides to the aforementioned nucleotides are 0.1599, 0.3161, 0.3382, and 0.1858. Transitions from Thymine nucleotides to other nucleotides have a probability of 0.1127. The sequence data follows a multinomial distribution, while the Markov Chain follows a Geometric distribution. Although both distributions differ, the preferable Markov chain is an Embedded Markov Model, which has been constructed. According to the definition, the transition probabilities from each state to the same

state are found to be zero. As discussed earlier, the transition probability matrix can be described.

In order to investigate the factors contributing to mutated malignant gene sequences, the Hidden Markov Model (HMM) was employed, and an Emission Probability Matrix was constructed. The match state has a 24% probability of emitting the Adenine Nucleotide. Similarly, there is a 32% chance of emitting Cytosine, a 22% chance of emitting Guanine, and a 21% chance of emitting Thymine nucleotides. The insert state has specific probabilities (0.22, 0.12, 0.31, and 0.34) for emitting the corresponding nucleotides at each position. Likewise, the delete state also has defined probabilities for emission.

Conclusion :

This paper presents a dataset comprising 150 ORMDL3 gene sequences with mutations, which were analyzed using hierarchical clustering. Consensus sequences were generated through this process, and both Markov Chain and Hidden Markov Model were constructed and trained using the aligned sequences. The primary objective of this study was to examine the mutated ORMDL3 DNA sequences for the identification and detection of malignant diseases. The findings of this study have important implications for genome analysis and the diagnosis of diseases associated with gene mutations. Early detection of gene mutations can be a valuable and cost-effective strategy for primary cancer prevention and treatment resistance. Moreover, the stochastic approach utilized in this research provides a faster, more cost-effective, and simplified method for testing gene sequences compared to traditional "wholesome genome" selection. The resulting time and cost savings can bring significant benefits to policymakers and society, especially for populations in need of such testing.

Acknowledgement:

The authors would like to express their gratitude to the editor and learned reviewers. There is no conflict of interest as declared by the authors.

Financial Support and Funding :

This work was financially supported by DST-INSPIRE, Government of India under the grant DST/INSPIRE Fellowship/2019/IF190881

Reference

1. Böer, J. (2016). Multiple alignment using hidden Markov models. *proteins*, *4*, 14.
2. Bonidia, R. P., Domingues, D. S., Sanches, D. S., & de Carvalho, A. C. (2021). MathFeature: feature extraction package for DNA, RNA and protein sequences based on mathematical descriptors. *Briefings in Bioinformatics*.
3. Deneshkumar, V., Manoprabha, M., & Senthamarai Kannan, K. (2020). Multiple Sequence Alignment with Hidden Markov Model for Diabetic Genome.
4. Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge university press.
5. Eddy, S. R. (1995, July). Multiple alignment using hidden Markov models. In *Ismb* (Vol. 3, pp. 114-120).
6. Gagniuc, P. A. (2017). *Markov chains: from theory to implementation and experimentation*. John Wiley & Sons.
7. Gentleman, J. F., & Mullin, R. C. (1989). The distribution of the frequency of occurrence of nucleotide subsequences, based on their overlap capability. *Biometrics*, 35-52.
8. Jeniffer, S.D., & Senthamarai Kannan, K., (2021). Stochastic modelling for identifying, *Advances and Applications of Mathematical Sciences*, *20*(9), 1923-1936.
9. Karuppusamy, T. (2021). Biological gene sequence stucture analysis using hidden markov model. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(4), 1652-1666.
10. Krumbein, W. C., & Dacey, M. F. (1969). Markov chains and embedded Markov chains in geology. *Journal of the International Association for Mathematical Geology*, *1*(1), 79-96.
11. Kumar, S., & Gadagkar, S. R. (2001). Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. *Genetics*, *158*(3), 1321-1327.
12. Li, J., Lee, J. Y., & Liao, L. (2021). A new algorithm to train hidden Markov models for biological sequences with partial labels. BMC bioinformatics, 22(1), 1- 21.
13. Mandoiu, I., & Zelikovsky, A. (2008). *Bioinformatics algorithms: techniques and applications* (Vol. 3). John Wiley & Sons.
14. Meng, Y., & Fei, J. (2022). Hidden service publishing flow homology comparison using profile-hidden markov model. *International Journal of Intelligent Systems*.
15. Mor, B., Garhwal, S., & Kumar, A. (2021). A Systematic Review of Hidden Markov Models and Their Applications. Archives of computational methods in engineering, 28(3).
16. Procter, J. B., Carstairs, G. M., Soares, B., Mourão, K., Ofoegbu, T. C., Barton, D., ... &

Barton, G. J. (2021). Alignment of biological sequences with Jalview. In Multiple Sequence Alignment (pp. 203-224). Humana, New York, NY.

17. Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257-286.
18. Rastogi, S. C., Mendiratta, N., & Rastogi, P. (2008). Bioinformatics Methods and Applications: Genomics. *Proteomics and Drug Discovery, PHI Learning Private Limited, New Delhi*.
19. Roth, C. (2021). Statistical methods for biological sequence analysis for DNA binding motifs and protein contacts (Doctoral dissertation, Georg-August University).
20. Sarkar, B. K. (2021). Entropy Based Biological Sequence Study. In Entropy and Exergy in Renewable Energy. IntechOpen.
21. Sasidharan, S. K., & Thomas, C. (2021). ProDroid—An Android malware detection framework based on profile hidden Markov model. *Pervasive and Mobile Computing*, *72*, 101336.
22. Singh, G. B. (2015). Introduction to bioinformatics. In *Fundamentals of Bioinformatics and Computational Biology* (pp. 3-10). Springer, Cham.
23. Schuster-Böckler, B., & Bateman, A. (2007). An introduction to hidden Markov models. *Current protocols in bioinformatics*, *18*(1), A-3A.
24. De Fonzo, V., Aluffi-Pentini, F., & Parisi, V. (2007). Hidden Markov models in bioinformatics. *Current Bioinformatics*, *2*(1), 49-61.
25. Simossis, V., Kleinjung, J., & Heringa, J. (2003). An overview of multiple sequence alignment. *Current protocols in bioinformatics, 3*(1), 3-7.
26. Böer, J. (2016). Multiple alignment using hidden Markov models. *proteins*, *4*, 14.
27. Yoon, B. J. (2009). Hidden Markov models and their applications in biological sequence analysis. *Current genomics*, *10*(6), 402-415.