Biodiversity in a forest ecosystem in India using environmental DNA sequence analysis of pathogen-hostile eucalyptus trees

Ramsamy Sankar Ram[1,*], Lakshmi Narayanan[2], Santhana Krishnan[3] , Harold Robinson [4]

[1]Anna University, BIT Campus, Tiruchirappalli

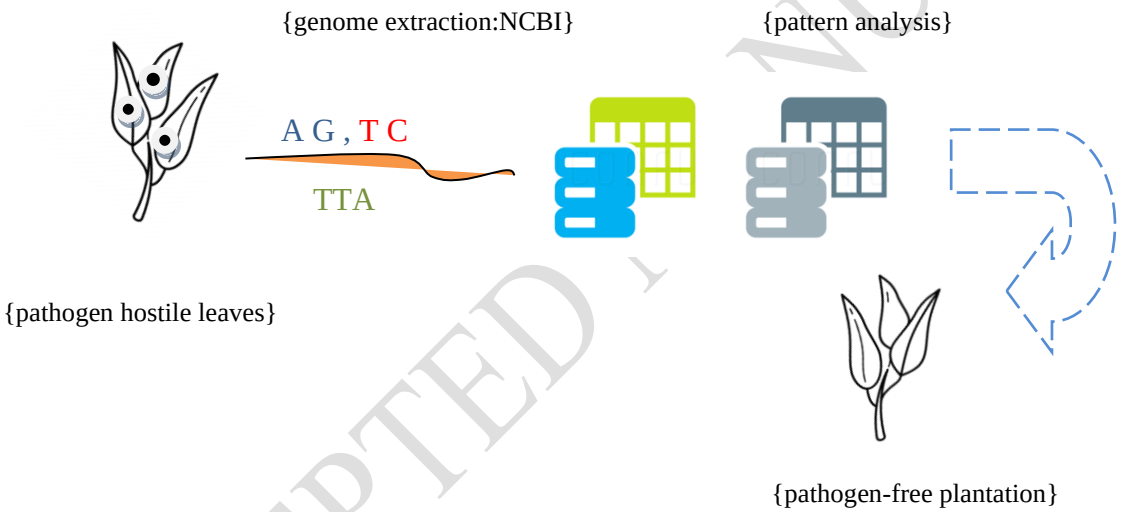
[2]Francis Xavier Engineering College, Tirunelveli

[3]SCAD College of Engineering and Technology, Tirunelveli

[4]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore

*Corresponding author. Email: csankarraam@gmail.com

GRAPHICAL ABSTRACT



ABSTRACT

Trees in the forest are an unprecedented cluster of organisms in ecological, monetary and social importance. With a wide distribution, mostly random spread over and a large population in terms of size, the majority of tree species show considerable variation in genetics both within and between populations. The genus *Phytophthora*, is a most destructive plant pathogen and attacks a wide range of tree hosts, including inexpensively its significant species. Many species of *Phytophthora* are known to be persistent and brought in through nurseries and commercial agriculture. Diseases from various eucalyptus tree species were first reported in India and the symptoms, incidence and leaf damage have been described. As an observation of a larger project to

use genomic data for tree disease diagnosis, pathogen detection and surveillance, in this study the significant analysis of various DNA sequences of *P.meadii 2* on eucalyptus tree species in Indian Forests. The identified outcomes observed that the high frequency of nucleotides and their combinations found in the organisms in trees threatening their lives may be observed and has to be condensed.

Keywords: genus, species, nucleotide, DNA, sequence, pathogen

1. Introduction

A growing number of *Phytophthora* species are posing a major threat to forest trees. There are most species of these funguses, which act like microorganisms, and can attack a variety of host species. This kind of infected woody tissue keeps posing a threat to plantations and native forest ecosystems around the world (Allwright *et. al.*, 2016). From boreal to tropical latitudes, forests can comprise over 30% of the earth's surface and many tree species serve a basic or crucial function in their particular ecosystems while also being a major generator of cost-effective activity in many locations. Forests also have three-quarters of the terrestrial biomass of the earth, which is closely related to the atmospheric carbon balance. Figure 1 shows the states of the largest forest covers in India up to 2022.

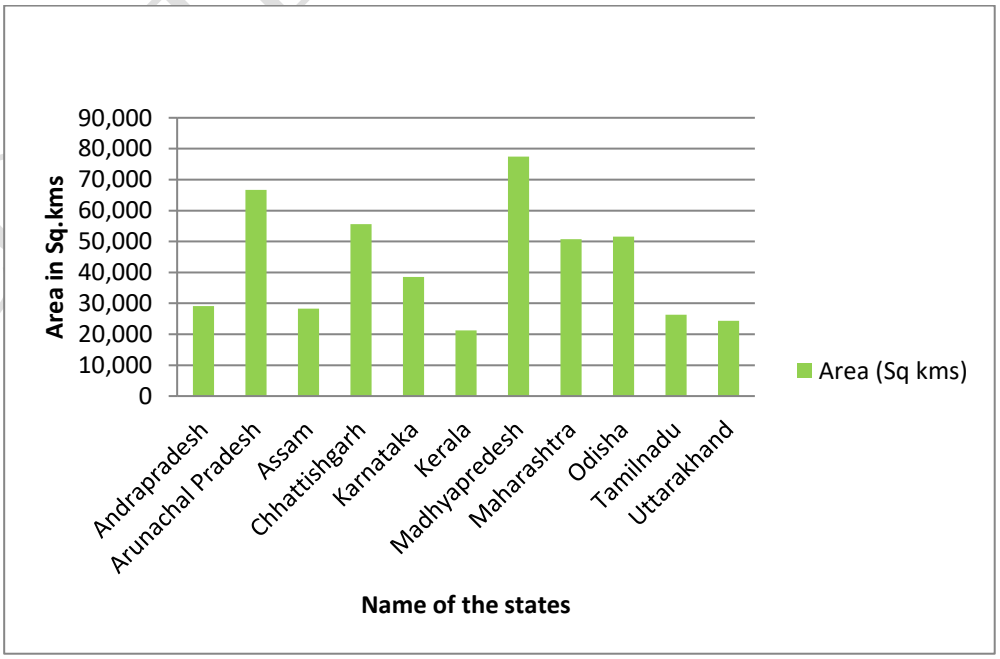


Figure 1. The largest forest covers areas in India

The epidemic of fungal pathogenic diseases has been reported to result in significant economic losses on eucalyptus forest plantations of *Phytophthora meadii 2* (*P. meadii*) observed as abnormal leaf fall, abnormal leaf litter, black stripe, pod rot and stripe canker. Among the above diseases, abnormal leaf litter (ALF) *Phytophthora* is a catastrophic disease that can reduce crop production by up to 40%. Thousands every year Prophylactic sprays on plantations use large amounts of fungicides to prevent the outbreak of large-scale illnesses. The impact of *P.meadii 2* on a eucalyptus leaf and its microscopic reproduced image has given in Figure 2. Early detection, monitoring, and prevention of such occurrences is hampered by the lack of genomics Financial resources. Genome sequencing and comparison should help with further development to predict the pathogenic consequences of monitoring tools and their host interactions with these microorganisms (Aguayo *et. al.*, 2012; Cobb *et. al.*, 2012)

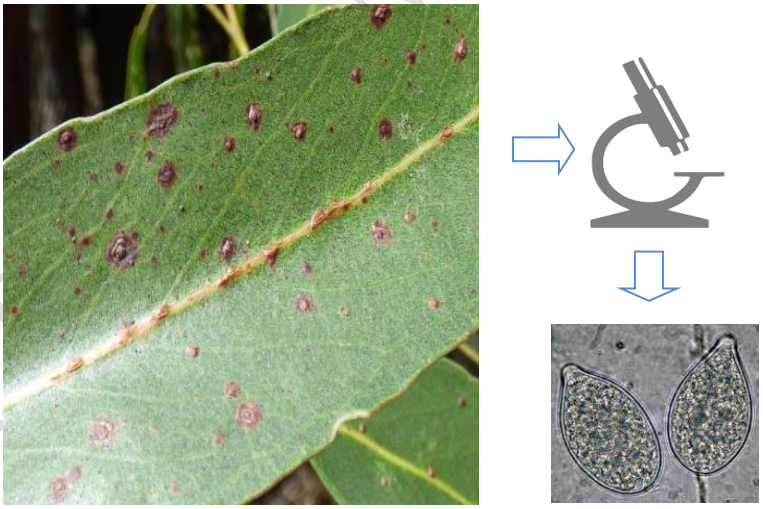


Figure 2. *Phytophthora meadii 2* affected the Eucalyptus leaf

The identification of prototypes of genome sequences is essential ultimately for their recognized behaviors for analysis. In this work, identification of genome sequences and the k-mer has identified using a superior graphical method coined as Chaos Game Representation (CGR) and Frequency Chaos Game Representation (FCGR) (Grattapagila *et. al.*, 2011; Changchuan yin *et.al.*, 2019; Jonas

et. al., 2001). Also, the k-mers are properly evacuated from the order as it has been represented in CGR, encountered tri-mer counts, and fed into deep learning algorithms. It is necessary to recognize the k-mers they exhibit and to be able to deduce them in a biologically significant way and utilized the extracted k-mer patterns by calculating the mono, dinucleotide and trinucleotide frequencies. These kinds of nucleotide frequencies help to use genomic data for detecting tree disease diagnosis; pathogen detection and surveillance, in this study the significant analysis of various DNA sequences of microorganisms affect tree species in Indian Forests. The identified outcomes observed that the high frequency of nucleotides and their combinations found in the organisms in trees threatening their lives may be observed and has to be condensed (Wei Deng *et. al.,* 2013). The key message is that although most of the benefits of genomic selection have come from this proposed modeling of genetic covariance between variable relatives, it is generally believed that the contribution of analysis gives better results.

The organization of this article followed by the survey of various articles in Section 2 and Section 3 provides Material and methods. The techniques applied are under sections CGR and FCGR gave in Section 4. It contains the observation of the evaluation. The results given in Section 5 followed by the section conclusion have finished this analysis.

2. Survey

The expansion of population genome research was generated by a variety of species, population dynamics, and ecological situations. We have addressed two key questions about how local adaptation occurs in widely distributed long-lived species like trees. The trees are currently underestimated in black among the available plant genomic sequences (Tuskan *et al.,* 2006) and *Eucalyptus grandis* as primary models. The species with reference sequences, and new tree genome sequencing projects implemented from the observation. Given this, genomic clusters are sequenced in a similar phylogenetic branch, which makes this possible. Here the comparative study is showing strong colinearity even between distantly related species for the sake of observation. A tree's local adaptation reflects changes in its environment. Parameters like photoperiod, temperature, and

pathogen pressure are considered. Among them, changes in photoperiod and temperature showed the finest signal in diverse selection among many species.

The species of the *Phytophthora* genus is known as a plant pathogen in general, some of which have a host range in narrow and others are capable of infecting many plant species (Echt *et al.,* 2011). *Phytophthora* species are fungi-like but mycelial oomycetes and belong to the stramenophyll (with the same name as Heterokonta, Kingdom Chromista, SAR upper group) together with diatoms and brown algae. From the above findings, the disease also severely affected young *C. tabularis* seedlings less than 20 cm tall, while seedlings over 40 cm tall showed only marginal fallen leaves (210% versus 20-100% for young seedlings less than 20 cm). The disease progressed in a mild form in *H. integrifolia*, and the seedlings fell by 10-60%.

An iterative CGR representation mapping technique processes the biological sequences to find the coordinates of the position of nucleotides or amino acids in a continuous space. This technique is used to evaluate nucleotide, dinucleotide, and trinucleotide frequencies of the genomic sequence. The distribution of nucleotide sequence on the plane is based on the position of the nucleotides from the original sequence in the coordinates of the plane. These positions are generated using the Markov–Chain probability process where the location of the sequence is determined by the position of the previous nucleotide. The frequency of the oligonucleotide combinations can be determined by dividing the CGR plane on the counting occurrence of nucleotides in each quadrant. The frequency of sequences extracted from the CGR space is mapped into a matrix called Frequency Matrix based FCGR (Wei Deng *et. al.*, 2013; Tung Hoang *et. al.*, 2016).

While biotechnology has created the well-organized tools needed to amend genes, genomics has a platform for the most resourceful genetic analysis of forest trees. The two important techniques of genomics are DNA Sequencing and Gene Mapping. Genetic maps were created using, a gene that allows positional associations to function previously. Although isozyme loci were

utilized as markers, there was a limit to the number of loci that could be chosen. The human genome project (HGP) has had a significant impact on forest biotechnology. This (HGP) revolutionary technology aided the study of many species, particularly forest trees, which had previously been considered hard to study due to their huge size and long production times. Forest tree genome sequencing (Tuskan *et al.,* 2006) has led to the emergence of many new techniques ("omics") that allow the study of gene expression for individual genes or huge gene families.

While featuring the new "omics" technology Understanding the relationships between them in the activity of biological systems is more difficult and needs predictive capabilities with the help of mathematical modeling (Wanger *et al.,* 2012;Wang *et al.,* 2014). The increased collaboration between engineering and molecular biology is reflected in the systems and synthetic biology of forest trees. A deep understanding of metabolic and logical pathways may ultimately be important for the survival of natural and reforestation forests and enables innovation in system design, but may ultimately be important for the survival of natural and reforestation forests. Forests all across the world are under threat from an invasion of the most vulnerable pests and illnesses driven in by global trade and travel.

From the observations of the above studies, the proposed algorithm is implemented to analyze the environmental DNA sequence analysis to sustain biodiversity in the forest ecosystem.

3. Materials and Methods

The microorganism found in the eucalyptus tree was collected from NCBI (The National Center for Biotechnology) and the description of the organism is illustrated in Table 1. The main objectives of this work are given based on the impacts of the analysis.

Table 1. *Phytophthora cf. meadii 2* origin reference

Current scientific name:	*Phytophthora cf. meadii 2*
Taxonomic rank:	species
NCBI Taxonomy ID:	2691827
Gen Bank:	MN866098.1

Reference:	*Phytophthora cf. meadii 2* isolate VN810 cytochrome oxidase subunit I (cox1) gene partial cds mitochondrial

- Analysis of the occurrences and diversity of pathogens throughout most of the Indian forests covers by environmental DNA sequence manipulation for eucalyptus species
- To develop a bio-monitoring capable proposed algorithm to predict the outcome of pathogenic interactions between microorganisms and their hosts with genome sequencing, analysis and comparisons
- To implement the algorithm, which provides the frequency or rarity in genomes of microorganisms to reduce the impacts
- Prevent the affection of *P.meadii 2* on eucalyptus trees at an early stage which helps pathogen-free tree seedlings and propagates awareness about pathogens in domestic mass tree plantations through the outcomes
- Increase the treatments against *P.meadii* 2 that can amplify the crop production which helps the nurseries and incorporated with eucalyptus bases

In this analysis, as the genomic tools are used to improve tree health and productivity, genomics results based on tree breeding and conservation of forest by providing more knowledge about ecosystem adaptations and demographic processes, the development of strategies that consider both sustainable use and genetic conservation is typically subjective. As global climate change disrupts the better adaptable paradigm for reforestation, calls for changes to forest management plans (long-term), and demands ever-changing public needs for acquiring the interminable and preferable forest products and services, methods and goals (Fady *et al*,. 2016).

The draught genome sequence of *P.meadii 2* species threatening trees is presented here. These species were chosen for their potential to cause major economic losses and large-scale damage to forest ecosystems. Sequences of *P.meadii 2* gathered from NCBI produced significant alignments based on the query given. (Plomion *et. al.,* 2001; Cunniffe *et. al.,* 2016; Holliday *et. al.,*

2017)

3.1 Markov Chain Model

An analytical method for modeling the data sets of whole-genome markers with genetic control of phenotypic differences was proposed and the potential application of this approach to forest tree breeding is significant (Grattapaglia *et. al.,* 2011).

A Markov chain is a model describing the sequence of possible events whose occurrence in continuous time depends on the preceding states of the event. The first coordinate of the nucleotide sequence is plotted as half the distance between the center of the quadrant and the corner relative to the first nucleotide of the sequence. Similarly, the following points are arranged half the distance between the preceding point and the corner relating to the successive sequence. Hence the location of the sequence in the CGR has been obtained.

The distance measures between the nucleotide sequence is calculated based on the below equations

$$\mathbf{a_i = 0.5(a_{i-1} + g_a(i))} \qquad \mathbf{(1)}$$

$$\mathbf{b_i = 0.5(b_{i-1} + g_b(i))} \qquad \mathbf{(2)}$$

where a_i and b_i are the points to be obtained,

a_{i-1} and b_{i-1} are the immediate previous points,

$g_a(i)$ is the coordinate of the vertex to the nucleotide at position i,

$g_b(i)$ is the b coordinate of that vertex.

Considering the nucleotide sequence AGCTTATTACG and applying the Markov Chain Model for the sequence. The CGR of the basic genome is represented as given below (Fillatti *et. al.,* 1987; Hansen *et. al.,* 2008; Tung Hoang *et. al.,* 2016)

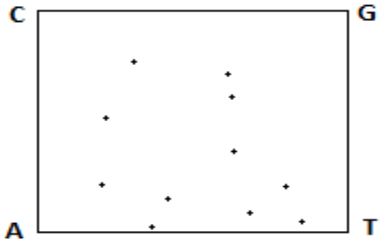


Figure 3: Sequence plotted using distance measure Markov-chain mode

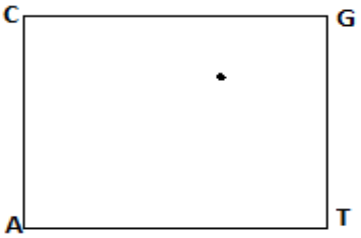


Figure 4: Position of the above nucleotide the sequence in the CGR plane

Figure 3 depicts the CGR that displays the Markov chain model for the desired sequence using the distance measures. The results obtained for the sequence in Figure 4 are precisely defined.

The center of the quadrant contains any sequence. Therefore, it represents the NULL DNA sequence. The adjacent points in the CGR graph do not mean the points are adjacent to each other in the sequence. The path of the points corresponding to each DNA sequence is unique for every sequence (Huang *et. al.*, 1993; Jules *et. al.*, 2002)

1.2. K – Mer Classification

The k-mers are k-length subsequences found inside the biological sequences. Utilized primarily in the sense of sequence analysis where k-mers comprise nucleotides (A, T, C and G), k-mers were gained to assemble sequences of DNA. For the most part, the k-mer refers to a particular n-tuple of nucleic acid or amino acid sequences that can be utilized to categorize certain regions inside biomolecules, for example, DNA (for predicting genes) or proteins. The sequences are collected from the database. The Genbank is a clarified assortment of all DNA sequences of any species and those records are written in FASTA format. In k-mer classification, the DNA sequences are represented in quadrants in terms of reverse order. Before processing the sequence, the count of the k-mers is to be resolved.

$$\alpha = L - k + 1 \qquad (3)$$

where α is the k-mer count,

L is the length of the sequence,

k is the number of mer in which the sequence is divided, and is an integer ranges from

1 to n.

In Chaos Game Representation, the position for the DNA sequences is computed based on the k-mer count. The quadrants for the position of the DNA sequence are constructed as a large dimensional array using the below equation.

$$n_c = 2^k \times 2^k \qquad (4)$$

where n_c is the no. of quadrants in the coordinates.

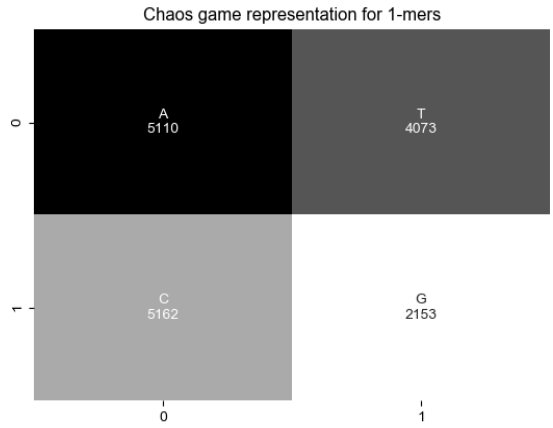


Figure 5: 1-mer count on k-mer of CGR

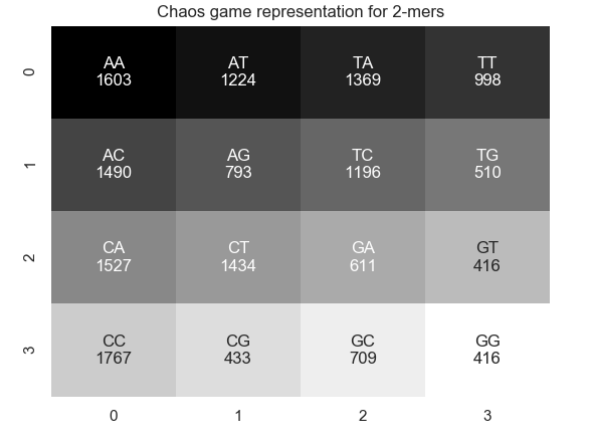


Figure 6: 2-mer count on k-mer of CGR

The sequences can be analyzed by identifying the number of k-mers present in the sequence. The number of k-mers is the number of occurrences of the nucleotides in the sequences (Figure 5 & Figure 6). The DNA sequence is positioned in the quadrants using k-mer based on either by count or by their probabilities. The probability of occurrence in the DNA sequence is calculated using

$$P_{(sequence)} = L_\alpha / L - k + 1 \qquad (5)$$

where $L\alpha$ is the total k-mer count,

$(L - k + 1)$ is the k-mer count.

The arranged DNA sequence in the CGR representation can be visualized as a grayscale image. The color of the image may vary depending on the frequency of the k-mer count. The color white in the image represents the higher frequency and the black represents the lower

frequency of DNA Sequences.

Algorithm

Listing -1

Grouping into destination directory as S

Set n as the file name

For n in S

Apply CGR code

Find k-mers (N-k+1)

N – length of the sequence

k – occurrence

Apply FCGR

Separate trimer, dimer counts among total k-mers

Separate trimer, dimer probabilities among total k-mers

Extract the results

Listing 1 is provided as the proposed algorithm which can produce the similarity results for various genomes of microorganisms found in forest trees. We have utilized the above algorithm to analyze and implementing the matrix frequency matching between various species of *P.meadii 2* (Zobel et. al., 1995; Metz et. al., 2013; Strauss et. al., 2015)

4. Results and Discussion

The key message is that although most of the benefits of genomic selection have come from this proposed modeling of genetic covariance between variable relatives, it is generally believed that the contribution of analysis gives better results. The comparative analysis of the frequent occurrences and remarkable diversity of specific pathogens throughout focused most of the Indian forest covered by environmental DNA sequence manipulation.This implementation has generated a matrix frequency for genetic code analysis of *P.meadii 2*. Here, report the pattern on triplets codon for genetic code analysis of specific pathogen. Also, we developed a bio-monitoring implementation to predict the outcome of

the pathogenic interaction between the microorganisms and their tree hosts with genome sequencing, analysis and comparisons. Designed and implemented the genetic analysis algorithm, which provides the frequency or rarity in genomes of microorganisms to reduce the impacts.

The mathematical characterization of genomic sequences helped to understand the structural relations between different whole genomes of *P.meadii 2*. The degenerate translation of the trinucleotide codon encodes: In the illustration of 20 amino acids, the rest three codons signal indicates the end of the amino acid. (Uddin, M et al, 2022; Hannah FL. Et. al, 2021)

AA (53)	AT (118)	TA (130)	TT (216)
AC (38)	AG (55)	TC (44)	TG (76)
CA (48)	CT (58)	GA (32)	GT (74)
CC (21)	CG (7)	GC (31)	GG (53)

Figure 7: 2-mer count in MN866098.1

AA (0.0503)	AT (0.112)	TA (0.1233)	TT (0.2049)
AC (0.0361)	AG (0.0522)	TC (0.0417)	TG (0.0721)
CA (0.0455)	CT (0.055)	GA (0.0304)	GT (0.0702)
CC (0.0199)	CG (0.0066)	GC (0.0294)	GG (0.0503)

Figure 8: 2-mer probability in MN866098.1

This study generated the first-order Markov chain matrix frequency for 16 (4 x 4) nucleotide 2-mer codon for *P.meadii 2* and its complete genome. The matrix frequency of 2-mer in Figure 7 has shown, the CG codon's least frequency range is <5 probability is 0.0066. Also generated was the second-order Markov chain matrix frequency of 64 (8 x 8) nucleotide (3-mer codon) for *P.meadii 2* and its complete genome.

AAA (0.0152)	AAT (0.0237)	ATA (0.0209)	ATT (0.057)	TAA (0.0199)	TAT (0.0598)	TTA (0.0693)	TTT (0.0931)
AAC (0.0066)	AAG (0.0047)	ATC (0.0133)	ATG (0.0209)	TAC (0.0199)	TAG (0.0237)	TTC (0.0171)	TTG (0.0256)
ACA (0.0123)	ACT (0.0161)	AGA (0.0095)	AGT (0.0123)	TCA (0.0161)	TCT (0.0114)	TGA (0.0104)	TGT (0.0228)
ACC (0.0057)	ACG (0.0019)	AGC (0.0123)	AGG (0.018)	TCC (0.0114)	TCG (0.0028)	TGC (0.0142)	TGG (0.0247)
CAA (0.0095)	CAT (0.0123)	CTA (0.0152)	CTT (0.019)	GAA (0.0057)	GAT (0.0152)	GTA (0.018)	GTT (0.0361)
CAC (0.0066)	CAG (0.0171)	CTC (0.0066)	CTG (0.0142)	GAC (0.0028)	GAG (0.0066)	GTC (0.0047)	GTG (0.0114)

CCA (0.0076)	CCT (0.0095)	CGA (0.000)	CGT (0.0019)	GCA (0.0095)	GCT (0.018)	GGA (0.0104)	GGT (0.0332)
CCC (0.0009)	CCG (0.0019)	CGC (0.0009)	CGG (0.0028)	GCC (0.0019)	GCG (0.0)	GGC (0.0019)	GGG (0.0047)

Figure 9. 3-mer frequency in MN866098.1

AAA (16)	AAT (25)	ATA (22)	ATT (60)	TAA (21)	TAT (63)	TTA (73)	TTT (98)
AAC (7)	AAG (5)	ATC (14)	ATG (22)	TAC (21)	TAG (25)	TTC (18)	TTG (27)
ACA (13)	ACT (17)	AGA (10)	AGT (13)	TCA (17)	TCT (12)	TGA (11)	TGT (24)
ACC (6)	ACG (2)	AGC (13)	AGG (19)	TCC (12)	TCG (3)	TGC (15)	TGG (26)
CAA (10)	CAT (13)	CTA (16)	CTT (20)	GAA (6)	GAT (16)	GTA (19)	GTT (38)
CAC (7)	CAG (18)	CTC (7)	CTG (15)	GAC (3)	GAG (7)	GTC (5)	GTG (12)
CCA (8)	CCT (10)	CGA (0)	CGT (2)	GCA (10)	GCT (19)	GGA (11)	GGT (35)
CCC (1)	CCG (2)	CGC (1)	CGG (3)	GCC (2)	GCG (0)	GGC (2)	GGG (5)

Figure 10. 3-mer frequency probability in MN866098.1

The matrix frequency of 3-mer in Figure 7 has shown, the combinations of AAC, AAG,ACC, ACG, TCG, GAA, CAC, CTC, GAC, GAG, GTC, CCA, CGA, CGT, CCC, CCG, CGC, CGG, GCC, GCG, GGC and GGG have minimum frequency range is <10.

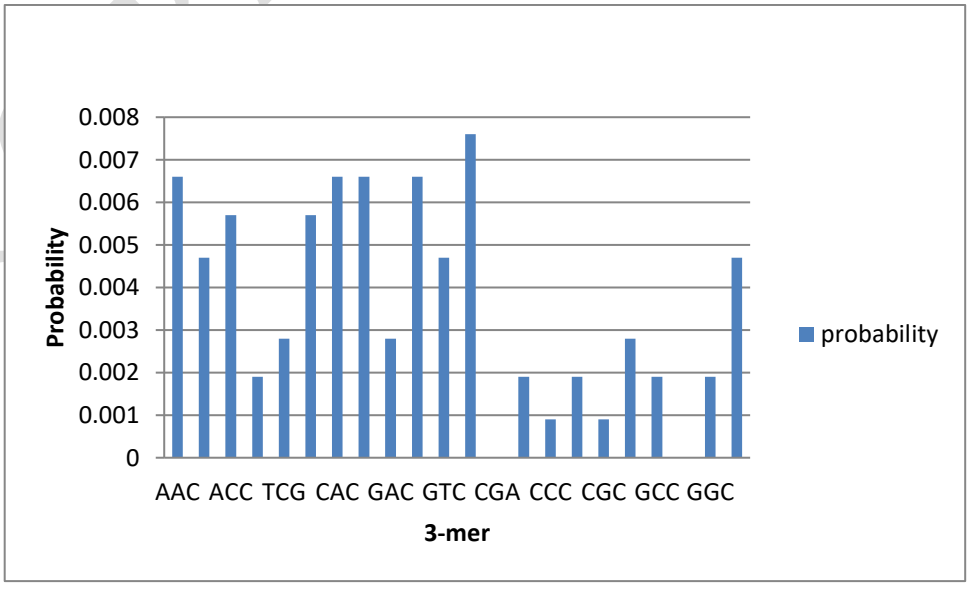


Figure 11. Matrix frequency probability of 3-mer (*P.meadii 2*)

Also, the matrix frequency probabilities of 3-mer in Figure 11 have shown, the combinations of AAC, AAG,ACC, ACG, TCG, GAA, CAC, CTC, GAC, GAG, GTC, CCA, CGA, CGT, CCC, CCG, CGC, CGG, GCC, GCG, GGC and GGG have minimum probability frequency range is <0.009. The consequences observed from the rarity and high frequency of nucleotides trigger the treatment against *P.meadii 2.*

5. Conclusions

The subtle preservation of biodiversity in forest ecosystems these genomic data were ultimately helped and used to proceed bio-surveillance for potential threats. We developed and implemented a bio-monitoring capable proposed algorithm against the threats, and predicted the outcome of the pathogenic interaction between microorganisms and their tree hosts with genome sequencing, analysis and comparisons. The implementation provided the frequency or rarity in the genomes of microorganisms to reduce the impacts. To prevent the affection of *P. meadii 2* on eucalyptus trees at an early stage. The identified outcomes observed that the high frequency of nucleotides and their combinations found in the organisms in trees threatening their lives may be observed and has to be condensed.

References

Uddin, M., Islam, M.K., Hassan, M.R. et al. A fast and efficient algorithm for DNA sequence similarity identification. Complex Intell. Syst., 2022

Hannah FL, Dominik H et. al., Chaos game representation and its applications in bioinformatics, Computational and Structural Biotechnology Journal, Volume 19, 2021, Pages 6263-6271

Aguayo, G.C. Adams, Halkett, F. et al., Strong genetic differentiation between North American and European populations of Phytophthora alni subsp. uniformis. Phytopathology 103 (2012) 190–199

Allwright, M. R., and Taylor, G. (2016). Molecular breeding for improved second generation bioenergy crops. Trends Plant Sci. 21 (1), 43–54. doi: 10.1016/j. tplants.2015.10.002

Botstein, D., White, R. L., Skolnik, M. H., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Amer. J. Hum. Genet. 32, 314–331.

Changchuan yin. Encoding and Decoding DNA Sequences by Integer Chaos Game Representation. J ComputBiol 2019;26:1-9.

Cobb, R.C., Filipe, J.A.N., Meentemeyer, R.K., Gilligan, C.A. & Rizzo, D.M. (2012) Ecosystem transformation by emerging infectious disease: loss of large tanoak from California forests. Journal of Ecology, 100, 712–722.

Conkle, M. T. (1980). "Amount and distribution of isozyme variation in various conifer species," in Proceedings 17th Meeting of the Canadian Tree Improvement Association, ed. M.A.K. Khalil (Canada: Canadian Forest Service, Environment) 109–117

Cunniffe, N.J., Cobb, R.C., Meentemeyer, R.K., Rizzo, D.M. & Gilligan, C.A. (2016) Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. Proceedings of the National Academy of Sciences, 201602153

Echt, C. S., Saha, S., Krutovsky, K. V., Wimalanathan, K., Erpelding, J. E., Liang, C., et al. (2011). An annotated genetic map of loblolly pine based on microsatellite and cDNA markers. BMC Genet. 12, 17. doi: 10.1186/1471-2156-12-17

Fady B, Aravanopoulos FA, Alizoti P, et al. (2016) Evolution based approach needed for the conservation and silviculture of peripheral forest tree populations. Forest Ecology and Management 375, 66-75.

Fillatti, J. J., Selmer, J., McCown, B., Hassig, B., and Comai, L. (1987). Agrobacterium mediated transformation and regeneration of Populus. Mol. Gen. Genet. 206, 192–199. doi: 10.1007/BF00333574

Grattapaglia, D., and Sederoff, R. (1994). Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137 (4), 1121–1137

Grattapaglia D, Resende MDV (2011) Genomic selection in forest tree breeding. Tree Genetics & Genomes 7, 241-255.

Hansen, E.M., Kanaskie, A., Prospero, S., McWilliams, M., Goheen, E.M., Osterbauer, N., Reeser, P. & Sutton, W. (2008) Epidemiology of Phytophthora ramorum in Oregon tanoak forests. Canadian Journal of Forest Research, 38, 1133–1143.

Holliday, J. A., Aitken, S. N., Cooke, J. E. K., Fady, B., González-Martínez, S. C., Heuertz, M., et al. (2017). Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. Mol. Ecol. 26 (3) 706–717 doi: 10.1111/mec.13963

Huang, H., Karnofski, D. F., and Tauer, C. G. (1993). Applications of biotechnology and molecular genetics to tree improvement. J. Arboriculture 19 (2), 84–98.

Jonas A S, Joao CA, AntonioMaretzek, et al. Analysis of genomi sequences by Chaos Game Representation. Brief Bioinform 2001; 17: 429–437.

Jules, E.S., Kauffman, M.J., Ritts, W.D. & Carroll, A.L. (2002) Spread of an invasive pathogen over a variable landscape: a nonnative root rot on Port Orford cedar. Ecology, 83, 3167–3181.

Leonberger, A.J, et al., A survey of Phytophthora spp. in midwest nurseries, greenhouses, and landscapes. Plant Dis. 97 (2013) 635–640.

Metz, M.R., Varner, J.M., Frangioso, K.M., Meentemeyer, R.K. & Rizzo, D.M. (2013) Unexpected redwood mortality from synergies between wildfire and an emerging infectious disease. Ecology, 94, 2152–2159.

Plomion, C., Leprovost, G., and Stokes, A. (2001). Wood formation in trees. Plant Physiol. 127, 1513–1523. doi: 10.1104/pp.010816

Strauss, S., Costanza, A. and Séguin, A. (2015). Genetically engineered trees: paralysis from good intentions. Science 349 (6250), 794–795. doi: 10.1126/ science.aab0493

Tung Hoang, Changchuan Yin, Stephen S T, et al. Numerical encoding of DNA sequences by chaos game representation with application in similarity comparison.Genomics 2016;108: 134-142.

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313 (5793), 1596–1604. doi: 10.1126/science.1128691

Wagner, A., Donaldson, L., and Ralph, J. (2012). "Lignification and lignin manipulations in conifers," in Advances in Botanical Research, vol. 61 . Eds. L. Jouanin and C. Lapierre (Burlington: Academic Press), 37–76. ISBN:978-0-12- 416023-1.

Wang, J. P., Naik, P. P., Chen, H.-C., Shi, R., Lin, C.-Y., and Liu, J. (2014). Complete proteomic-based enzyme reaction and inhibition kinetics reveal how monolignol biosynthetic enzyme families affect metabolic flux and lignin in Populus trichocarpa. The Plant Cell 26, 894–914.

Wei Deng, Yihui Luan. Analysis of Similarity/Dissimilarity of DNA Sequences Based on Chaos Game Representation. Hindawi Publishing Corporation Abstract and Applied Analysis 2013:pp.1-6

Zobel, B. J., and Jett, J. B. (1995). "Genetics of wood production," in Springer Series in Wood Science. Ed. T. E. Timmel (Berlin, Heidelberg: Springer-Verlag) 337. doi: 10.1007/978-3-642-79514-5