Vol.65: e22200731, 2022
https://doi.org/10.1590/1678-4324-2022200731
ISSN 1678-4324 Online Edition

Article - Engineering, Technology and Techniques

Wavelet Applied to the Classification of Bacterial Genomes

Leila Maria Ferreira[1]*
https://orcid.org/0000-0003-1723-8253

Thelma Sáfadi[1]
https://orcid.org/0000-0002-4918-300X

Juliano Lino Ferreira[2]
https://orcid.org/0000-0002-8502-4444

[1]Universidade Federal de Lavras, Departamento de Estatística, Lavras, Minas Gerais, Brasil; [2]Empresa Brasileira de Pesquisa Agropecuária, Embrapa Pecuária Sul, Bagé, Rio Grande do Sul, Brasil.

Editor-in-Chief: Alexandre Rasi Aoki
Associate Editor: Andressa Novatski

Received: 19-Nov-2020; Accepted: 01-Nov-2022.

*Correspondence: leilamaria2003@yahoo.com.br; Tel.: 55-35-991740682 (L.M.F.)

HIGHLIGHTS

- Advancement of analyzes using the wavelet technique applied to genome data.
- Analyze the entire genome, that is, there is no loss of information.
- It represents a more detailed technique in which energy (variance) is used.

Abstract: The classifications resulting from phylogenetic analysis are essential tools for evolutionary studies. Phylogenetic is more than a part of evolutionary biology because its underlying philosophy provides a way to see nature, ask questions, and solve problems related to the evolution of organisms. Given the importance of phylogeny, our aim was to devise a method to assess the delimitation of bacterial species. We used the non-decimated discrete wavelet transform. The wavelet function used was Daubechies' with four null moments, considering seven, four and two decomposition levels. For clustering, the energy (variance) obtained at each level of decomposition and the Mahalanobis distance was used to visualize the dendrogram formation process. Through the analysis, we verified that the gram-positive bacteria were classified well into their respective species, but most gram-negative bacteria did not take into account the more significant amount of energy obtained in scenario two. According to the results, the energy plays an important role in the delimitation of groups of bacterial species.

Keywords: Wavelet transform; Decomposition levels; Energy; Bacterial genomes; Species classification.

INTRODUCTION

To understand evolution, the term 'phylogenetic tree' is used to refer to the branching evolutionary history or to a graph that represents this evolutionary history. The phylogenetic tree plays an important role in modern biology because it provides a concise way to visualize evolution from a common ancestor [1]. It is a tool with increasingly recognized utility for various analyses in the field of biological sciences because it helps to

understand biological processes [2]. Conceptually, the phylogenetic tree portrays more than evolution because its underlying philosophy provides a way of seeing nature, asking questions, and solving problems related to the evolution of organisms [3].

Several studies have aimed to improve the methods of phylogeny. Some authors [4] created a bioinformatics pipeline called bcgTree, which uses bacterial genomes gathered from databases or their sequencing. From the genomes input by the user, bcgTree reconstructs their phylogenetic history. Other authors [5] presented a study on the creation of PaHMMTree, a new neighbour joining–based method that estimates pairwise distances without assuming a single alignment. Other authors [6] developed a metric, in the sense of a proper distance function, to be applied to tree shapes, to distinguish trees from random models that produce different tree shapes.

Based on these new methods proposed to structure the phylogenetic tree, the objective of this study is to apply wavelets to the classification of bacterial species from their genomes. Wavelets have been gaining considerable attention in genomics studies. Among the various characteristics of working with wavelets, the degree of detail of the information contained in the data is the most significant benefit of this method, that is, the wavelet works acting as a magnifying glass. The multiresolution technique corresponds to the wavelet transform (which can be discrete decimated, discrete non-decimated, continuous, or complex), which in turn decomposes, for example, the analysed signal into several levels of resolution, in which the first levels carry little information, unlike the last levels, where it is possible to see the information in more detail. In the context of genomes, wavelet decomposition gives information on the ancestry blocks sizes and, primarily, how they are scattered along chromosomes. To highlight the increasing use of wavelets in the study of genomes, we reference the following studies: [7-14].

MATERIAL AND METHODS

For this study, we selected seven species of bacteria, each composed of eight genome sequences. The size of each sequence was approximately two million base pairs. All these sequences were had from the National Center for Biotechnology Information (NCBI) website [15].

The free software R, version 3.6.1 [23] was used to perform the analyses. The packages used were: seqinr, wmtsa, waveslim, cluster. The operating system used was Windows.

The sequences of the selected genomes were transformed into a proportion equal to its GC content, using a window of 10,000 base pairs. That is, the numerator was the sum of the number of guanine (G) and cytosine (C) nucleotides, divided by the denominator, which was the sum of the adenine (A), G, C, and thymine (T).

The GC sequences were analysed using the non-decimated discrete wavelet transform (NDWT), whose main characteristics include working with the same amount of data at all decomposition levels and the fact that the size of the genome does not alter the analysis, i.e., it does not need to be a power of 2. The wavelet function used was Daubechies', with four null moments, using three different decomposition levels, seven, four, and two. With regards to clustering, the energy (variance) of each decomposition level and the Mahalanobis distance was used to visualise the dendrogram formation process.

The genomes used are described below:

Description of genomes

Haemophilus influenzae (HI) is a coccobacillary, gram-negative bacterium with aerobic and anaerobic metabolism. This bacterium causes diseases specifically in humans, usually in young children. *H. influenzae* can cause meningitis, middle ear infections, and respiratory infections (pharyngitis, bronchitis, or pneumonia) [16].

Streptococcus pyogenes (SP) is a species of gram-positive bacteria with a spherical (coccus) morphology belonging to the genus group A Streptococcus. It causes several diseases, from common bacterial pharyngitis to more severe diseases such as scarlet fever [17].

The pathogen *Francisella tularensis* (FT) is an exceptionally infectious gram-negative bacterium. It does not produce spores, is not mobile, and is the disease-causative agent of tularaemia, whose pneumonic form is often lethal and untreatable [18].

Campylobacter jejuni (CJ) is a gram-negative, spiral-shaped bacterium with two flagella at opposite ends that causes diarrhoea in mammals and birds. It is transmitted to animals (including humans) by means of the consumption of unpasteurized dairy products, undercooked meat or water infected with animal faeces. This bacterium is the leading cause of the most common causes of dysentery worldwide [19].

Streptococcus pneumoniae (SPn) also known as pneumococcus, is a species of gram-positive bacteria. It provokes a variety of infectious diseases, for instance, like otitis media, pneumonia, and meningitis [20].

Coxiella burnetii (CB) is a mandatory gram-negative intracellular bacterium that causes Q fever. This bacterium is very resistant facing environmental stresses, for example, osmotic pressure, high temperature, and ultraviolet light [21].

Streptococcus mutans (SM) is a species of gram-positive bacteria with a spherical (coccus) morphology belonging to the genus group A Streptococcus. This bacterium, typically located in the mouth of humans is the main factor in the development of dental caries [22].

Table 1 contains the information for each genome according to the NCBI website.

Table 1. Genome information

Genome	Strain	Sequence Lenght	Accession number
Haemophilus influenza (HI)	Rd KW20	1,830,138	L42023.1
	NCTC8143	1,890,645	LN831035.1
	P642-4396	1,897,311	CP031686.1
	NCTC13377	1,890,469	LS483480.1
	M13034	1,887,933	CP031239.1
	723	1,887,620	CP007472.1
	PittGG	1,887,343	CP044497.1
	5P28H1	1,886,450	CP020008.1
Streptococcus pyogenes (SP)	SF370	1,852,433	AE004092.2
	MGAS10394	1,899,877	CP000003.1
	emm55	1,899,479	CP035430.1
	NCTC8195	1,898,595	LS483351.1
	MGAS6180	1,897,578	CP000056.2
	M28PF1	1,896,976	CP011535.2
	MGAS8232	1,895,017	AE009949.1
	SSI-1	1,894,275	BA000034.2
Francisella tularensis (FT)	SCHU S4	1,892,599	CP073128.1
	FSC147	1,893,886	CP000915.1
	WY-00W4114	1,899,252	CP009753.1
	WY96-3418	1,898,476	CP000608.1
	WY96	1,898,140	CP012037.1
	LVS	1,895,994	AM233362.1
	OSU18	1,895,727	BK006741.1
	425	1,894,186	CP010289.1
Campylobacter jejuni (CJ)	NCTC11951	1,891,235	LR134359.1
	269.97	1,845,106	CP000768.1
	FDAARGOS_295	1,845,051	CP027403.1
	CFSAN054107	1,898,513	CP028185.1
	NCTC13265	1,822,834	LR134498.1
	HF5-4A-4	1,821,520	CP007188.1
	NADC 20827	1,858,258	CP045048.1
	NCTC13268	1,801,392	LR134497.1
Streptococcus pneumonia (SPn)	ATCC 49619	2,096,423	AP018938.1
	M16808	2,096,231	CP031245.1
	INV200	2,093,317	FQ312029.1
	NCTC11902	2,093,242	LS483417.1
	SWU02	2,092,148	CP018347.1
	AUSMDU00010538	2,090,792	CP045931.1
	TCH8431/19A	2,088,772	CP001993.1
	11A	2,079,194	CP018838.1

Cont. Table 1

Coxiella burnetii (CB)	Scurry_Q217	1,973,771	CP014565.1
	2574	2,007,582	CP014555.1
	RSA439	2,006,529	CP040059.1
	Nine Mile phase II	2,024,349	CP035112.1
	Z3055	1,995,457	LK937696.1
	CbuG_Q212	2,008,870	CP001019.1
	RSA 331	2,053,744	CP000890.1
	2014-PE-15890	2,093,382	CP032542.1
Streptococcus mutans (SM)	UA159	2,032,925	AE014133.2
	NCTC10449	2,019,343	LS483349.1
	NBRC 13955	2,018,796	AP019720.1
	LAR01	2,088,369	CP023477.1
	LAB761	2,076,490	CP033199.1
	UA140	2,050,049	CP044495.1
	UA159-FR	2,031,692	CP007016.1
	GS-5	2,027,088	CP003686.1

The details of the wavelet transform and calculation of the scalogram are shown below.

Non-decimated discrete wavelet transform

For an arbitrary sample *n,* the NDWT and the scale coefficients are defined by [24]

$$\widetilde{W}_{j,t} = \sum_{K=0}^{K_j-1} \tilde{h}_{j,K} X_{t-k \bmod \mathrm{n}}$$
$$\tilde{V}_{j,t} = \sum_{K=0}^{K_j-1} \tilde{g}_{j,k} X_{t-k \bmod \mathrm{n}}, \quad t = 0, 1, \ldots, n-1, \tag{1}$$

Where:

$\{\tilde{h}_{j,k}: k = 0, \ldots, K_j - 1\}$ and $\{\tilde{g}_{j,k}: k = 0, \ldots, K_j - 1\}$

are the NDWT, and the filters are defined by

$\tilde{h}_{j,k} \equiv h_{j,k}/2^{j/2}$ and $\tilde{g}_{j,k} \equiv g_{j,k}/2^{j/2}$

respectively. Therefore, $\{h_{j,k}\}$ is the wavelet filter, and $\{g_{j,k}\}$ is the scaling filter

$K_j \equiv (2^j - 1)(K - 1) + 1.$

The NDWT filters are modified in each scale by inserting zeros. That is, in each scale $2^{(j-1)}$, zeros are inserted between each value *K* of the filters $\{\tilde{h}_j\}$ and $\{\tilde{g}_j\}$ of the NDWT. That is,

$$\tilde{h}_0, \underbrace{0, \ldots, 0}_{2^{j-1}}, \tilde{h}_1, \underbrace{0, \ldots, 0}_{2^{j-1}}, \ldots, \tilde{h}_{K-2}, \underbrace{0, \ldots, 0}_{2^{j-1}}, \tilde{h}_{K-1}$$
$$\tilde{g}_0, \underbrace{0, \ldots, 0}_{2^{j-1}}, \tilde{g}_1, \underbrace{0, \ldots, 0}_{2^{j-1}}, \ldots, \tilde{g}_{K-2}, \underbrace{0, \ldots, 0}_{2^{j-1}}, \tilde{g}_{K-1}, \tag{2}$$

which amount to applying a larger sample of size $2^{(j-1)(K-1)+1}$.

The wavelet coefficients $\{\tilde{W}_j\}$ and the scaling coefficients $\{\tilde{V}_j\}$ of the NDWT level *j*, represented in equation (1), can also be calculated using an efficient algorithm based on the scaling coefficients $\{\tilde{V}_{j-1}\}$ of the NDWT level $j - 1$. The advantage of NDWT is that it does not yield a reduced sampling of wavelet and scaling

coefficients and does not require the analysed signal to be a power of 2. At each stage of the pyramid algorithm of the NDWT, the wavelet and scaling filters are expanded, as in equation (2), so that, when performing the convolution of a signal with the filter, *n* coefficients are obtained at each algorithm scale.

The coefficients $\tilde{W}_j$, $\tilde{V}_j$ and $\tilde{V}_{j-1}$ are obtained by circular filtering of X_t with the respective periodized filters

$\{\tilde{h}_{j,k}\}$, $\{\tilde{g}_{j,k}\}$, and $\{\tilde{g}_{j-1,k}\}$,

as in equation (1).

It is possible to obtain $\tilde{W}_j$ and $\tilde{V}_j$ filtering $\tilde{V}_{j-1}$ through the following equation:

$$\tilde{W}_{j,t} = \sum_{K=0}^{K-1} \tilde{h}_k \tilde{V}_{j-1,t-2^{j-1}k \bmod N} \tag{3}$$

$$\tilde{V}_{j,t} = \sum_{K=0}^{K-1} \tilde{g}_k \tilde{V}_{j-1,t-2^{j-1}k \bmod N}, t = 0, 1, \ldots, N-1$$

These two equations constitute the pyramid algorithm of the NDWT.

Daubechies' wavelet

Daubechies' wavelets comprise a family of orthogonal wavelets that are characterized by a maximum number of null moments for a given support. One of the significant advantages of working with wavelets is their multiresolution analysis and their localization in time and frequency [25]. For each type of wavelet of the Daubechies class, lies a scaling function, termed the parent wavelet, which spawns an orthogonal multiresolution analysis [26].

The scaling function $\phi(x)$ and wavelet function $\psi(x)$ of the wavelet Daubechies both satisfy the following relationship [27]:

$$\phi(x) = \sum_{i=0}^{N-1} p_i \phi(2x - i), \tag{4}$$

$$\psi(x) = \sum_{i=2-N}^{1} (-1)^i p_{1-i} \phi(2x - i) \tag{5}$$

Here, ρ_i = (i = 0, 1, …, N – 1) are called filter coefficients, and N is an even integer. The supports of the scaling function and its corresponding wavelet function are given by:

$$\text{supp}\phi_N = [0, N-1], \tag{6}$$

$$\text{supp}\psi_N = [1 - N/2, N/2]. \tag{7}$$

From the Daubechies' wavelet construction process, the scaling function and wavelet function have the following properties:

$$\int_{-\infty}^{\infty} \phi(x)dx = 1; \tag{8}$$

$$\int_{-\infty}^{\infty} \phi(x-j)\phi(x-m)dx = \delta_{j,m},\ j, m \in \mathbb{Z}; \tag{9}$$

$$\int_{-\infty}^{\infty} x^k \psi(x)dx = 0, k = 0, 1, \ldots, N/2 - 1; \tag{10}$$

$$\int_{-\infty}^{\infty} \phi(x)\psi(x-m)dx = 0. \tag{11}$$

Equations (5) and (10) of the scaling function for the Daubechies' wavelet can represent exactly any polynomial whose order is not higher than *N*/2 – 1. That is, for any *f*(*x*) given by

$$f(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_{N/2-1} x^m, \qquad m \leq N/2 - 1 \tag{12}$$

we can represent it as

$$f(x) = \sum_{k=-\infty}^{\infty} c_k \phi(x-k). \tag{13}$$

Scalogram

The scalogram (energy) is a graph of the sum of squares of the wavelet coefficients at the different levels. In the context of discrete transformation, it represents a decomposition of the energy of a function in the time–frequency plane [28].

$$E(j) = \sum_{k=0}^{n} d_{j,k}^2 \quad j = 1, \ldots, J \tag{14}$$

Equation (14) gives the energy calculation in NDWT [29] at level *j*.

RESULTS

To form the groups, seeking to identify the phylogenetic relationship of the bacteria analysed, the energy obtained at each decomposition level was used. The maximum number of decomposition levels was calculated from the formula that uses the logarithm of the total length of the analysed signal in base 2. After this process, it was detected that the maximal number of levels to be analysed was seven. Figure 1 illustrates how the energy is distributed at each decomposition level. In Figure 1, we analyse the genome of the SM species, which did not differ significantly in terms of the amount of energy from the other bacterial species analysed in this study.

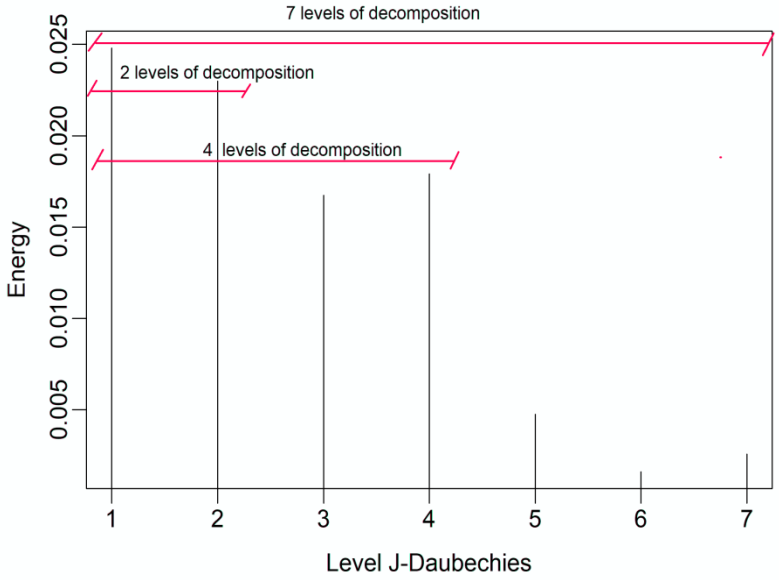

Figure 1. Energy in the Streptococcus mutans genome.

The characteristics shown in Figure 1, referring to levels 1 and 2, show the greatest amount of energy, followed by a slight drop in the amount of energy at levels 3 and 4 and a sharp drop in energy at levels 5, 6, and 7. Among the levels analysed, the highest energy was found at level 1 and the lowest at level 6.

The analysis was performed for the first scenario; take into consideration the information on the amount of energy at each decomposition level, where seven decomposition levels were selected. Figure 2 shows the result obtained from this analysis.

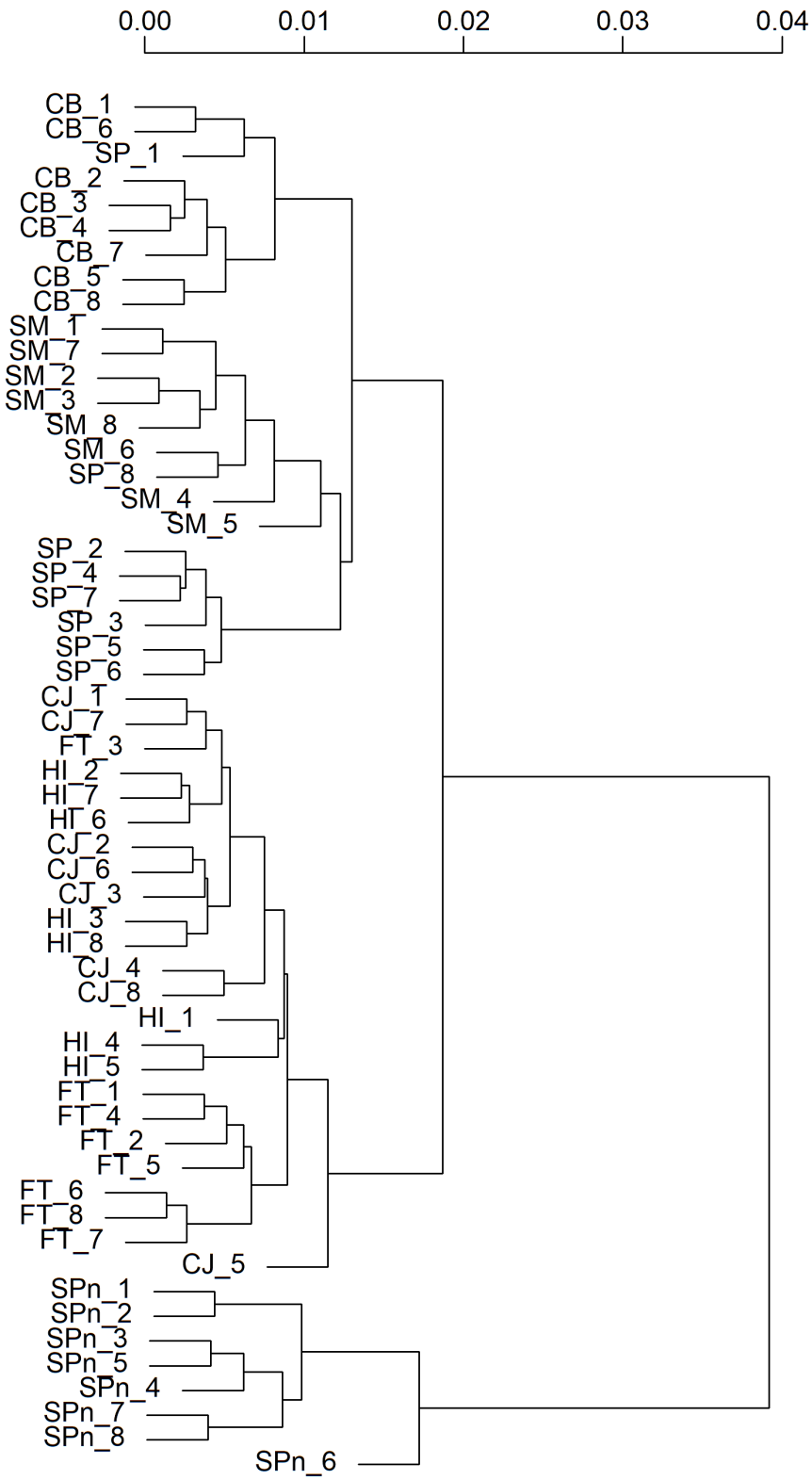

Figure 2. Decomposition with seven levels.

Table 2 (A) shows that the only bacterial species grouped with all of the sequences of the same species and no others were the SPn species. The species of CB was grouped with a sequence of the species of SP.

The species of SM was grouped with a sequence of the species of SP. The species of SP was grouped without any other species but without two sequences of its own species. The species of CJ, FT, and HI were completely mixed with each other.

Table 2. Formation of groups.

(A) Seven decomposition levels	
Species	**Group formed**
C. burnetii (CB)	CB_1,CB_6,SP_1,CB_2,CB_3,CB_4,CB_7,CB_5,CB_8
S. mutans (SM)	SM_1,SM_7,SM_2,SM_3,SM_8,SM_6,SP_8,SM_4,SM_5
S. pyogenes (SP)	SP_2,SP_4,SP_7,SP_3,SP_5,SP_6
C. jejuni (CJ), *F. tularensis* (FT), and *H. influenza* (HI)	CJ_1,CJ_7,FT_3,HI_2,HI_7,HI_6,CJ_2,CJ_6,CJ_3,HI_3,HI_8, CJ_4,CJ_8,HI_1,HI_4,HI_5,FT_1,FT_4,FT_2,FT_5,FT_6, FT_8,FT_7,CJ_5
S. pneumonia (SPn)	SPn_1,SPn_2,SPn_3,SPn_5,SPn_4,SPn_7,SPn_8,SPn_6
(B) Four decomposition levels	
Species	**Group formed**
C. burnetii (CB)	CB_1,CB_6,CB_2,CB_3,CB_4,CB_7,CB_5,CB_8
C. jejuni (CJ), *F. tularensis* (FT), and *H. influenza* (HI)	CJ_1,CJ_7,FT_3,CJ_4,HI_2,HI_6,HI_7,FT_2,CJ_2,CJ_6, HI_3,CJ_3,HI_8,CJ_8,FT_1,HI_5,FT_4,HI_4,FT_5,HI_1,FT6, FT_8,FT_7,CJ_5
S. mutans (SM)	SM_1,SM_7,SM_2,SM_3,SM_8,SM_6,SP_8,SM_4,SM_5
S. pyogenes (SP)	SP_1,SP_2,SP_4,SP_7,SP_3,SP_5,SP_6
S. pneumonia (SPn)	SPn_1,SPn_2,SPn_3,SPn_5,SPn_4,SPn_7,SPn_8,SPn_6
(C) Two decomposition levels	
Species	**Group formed**
S. mutans (SM)	CB_1,SM_6,SM_8,SM_1,SM_7,SP_8,SM_2,SM_4,SM_3, CB_6
C. burnetii (CB)	CB_2,CB_3,CB_4,HI_4,CB_5,CB_8
C. jejuni (CJ), *F. tularensis* (FT), and *H. influenza* (HI)	CB_7,FT_4,FT_6,FT_8,FT_5,FT_1,FT_7,HI_1,FT_2,HI_6, HI_7,HI_5,CJ_1,CJ_7,FT_3,CJ_2,CJ_8,CJ_6,HI_3,CJ_3,HI_8,CJ_4,HI_2,CJ_5
S. pyogenes (SP)	SM_5,SP_1,SP_2,SP_4,SP_3,SP_6,SP_7,SP_5
S. pneumoniae (SPn)	SPn_1,SPn_2,SPn_7,SPn_8,SPn_3,SPn_5,SPn_4,SPn_6

For the second scenario, we chose to use four decomposition levels in the NDWT method. Figure 3 shows the result obtained with this analysis.

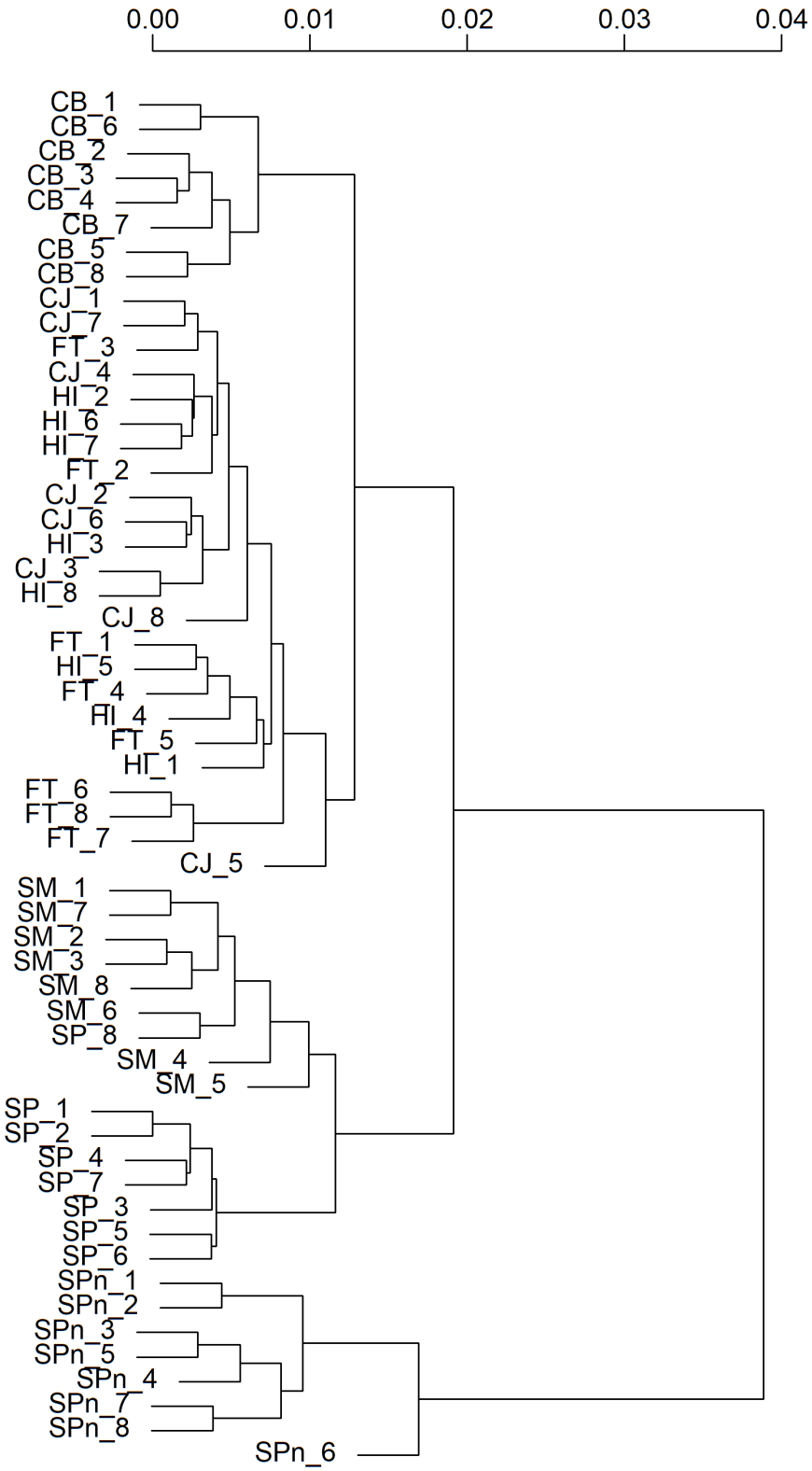

Figure 3. Decomposition with four levels.

For the third scenario, we chose to use two decomposition levels in the NDWT method. Figure 4 shows the result obtained with this analysis.

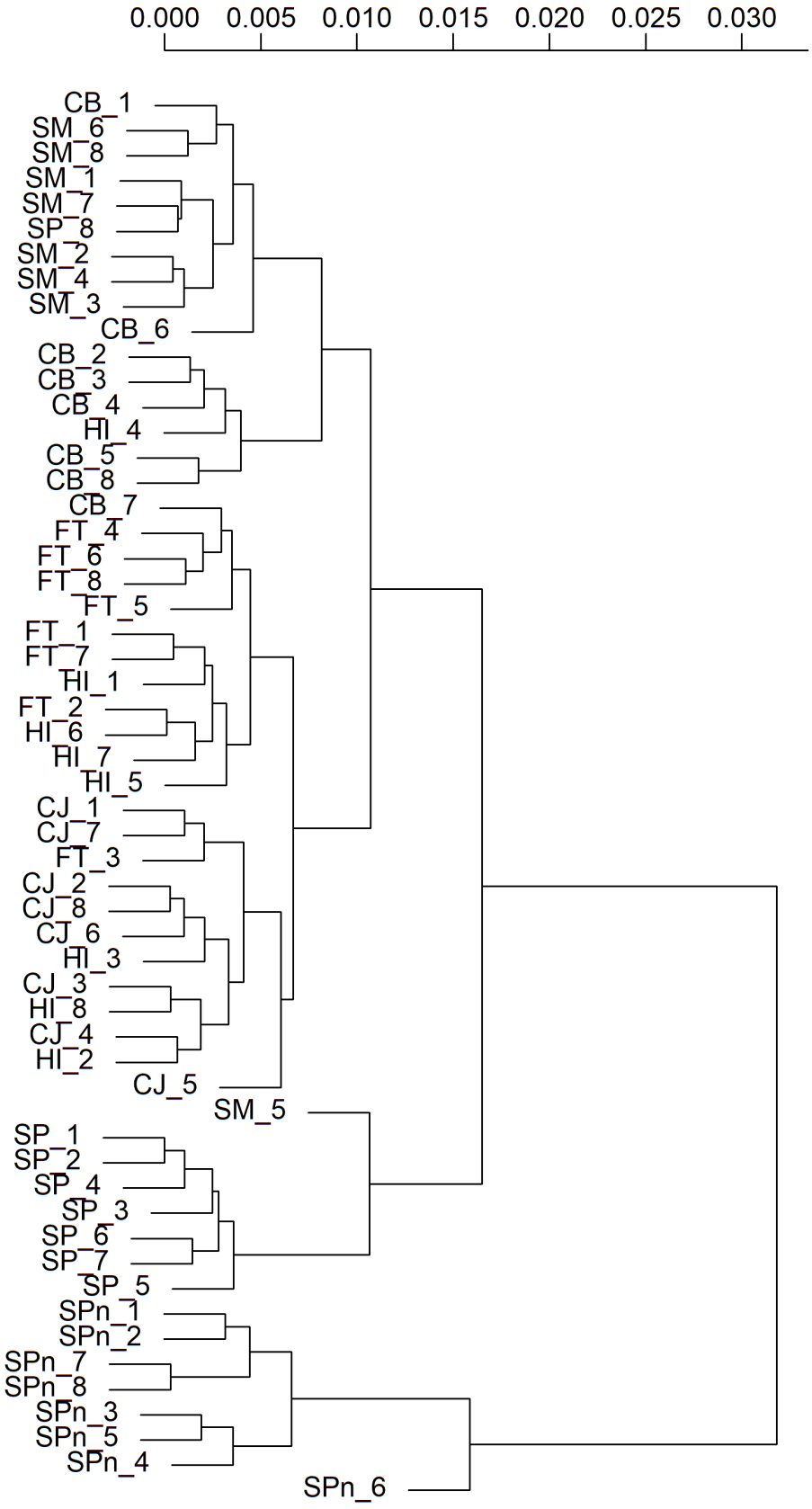

Figure 4. Decomposition with two levels.

Analysing this scenario, we found that the grouping of bacterial species showed more interference from other species in the formation of groups (Table 2 (C)). The species of SM lost a representative of its own kind, was still grouped with a representative of the SP species, and was now grouped also with two representatives of the CB species. The species of CB lost three sequences of its own kind and was grouped with a representative of the species of HI. The species of SP was grouped with a representative of the SM species and without a sequence of its own kind. The species of SPn was grouped again without any other species of bacteria. The species HI, CJ, and FT were still wholly mixed.

DISCUSSION

Analysing the results obtained in the three scenarios, it is quite evident that the amount of energy at each decomposition level has a direct influence on the clustering of bacterial species. Before-mentioned analyses, taking into account the energy at each decomposition level, lead to a gain in the degree of detail of the data [30].

Of the three scenarios explored, the second scenario was where the bacterial species were most accurately grouped, where four decomposition levels were used, corresponding to the levels with the highest amount of energy. In this scenario, the species of CB, SM, SP, and SPn were clearly defined. The species of CB stood out in that group because it was the only gram-negative bacterium that was properly clustered, whereas the other three bacterial species were gram-positive. The other species of gram-negative bacteria in this scenario were not correctly clustered. The species of CJ, FT, and HI were completely mixed. The contrast between gram-negative and gram-positive bacteria lie that the gram-negative bacteria, thanks to their thick peptidoglycan layer, retain the in their cell walls the red colour, while the gram-positive ones, with their thinner peptidoglycan layer, retain the purple colour.

In the first and third scenarios, with seven and two decomposition levels, respectively, the only bacterial species that formed a cluster without any interference from other bacterial sequences was that of SPn. In the second scenario, we had two species of bacteria that were grouped only with their sequences, the species of SPn and CB.

CONCLUSION

The proposed method for the formation of groups of bacterial species showed that the gram-positive bacteria carry more information that helps in their classification, so the wavelet method performed better on them than on gram-negative bacteria.

Concerning the three scenarios tested, the best was the second one, where the four decomposition levels analysed had the most significant amount of energy. Thus, energy plays an imperative role in the delimitation of groups of bacterial species.

Funding: “This research received no external funding”.
Acknowledgments: “The Federal University of Lavras for being able to carry out the post-doctoral training”.
Conflicts of Interest: “The authors declare no conflict of interest”.

REFERENCES

1. Baum DA, Smith SD. Tree thinking: An introduction to phylogenetic biology. United States: Roberts and Company Publishers; 2013. 476 p.
2. Hall BG. Phylogenetic trees made easy: A how to manual. United States: Sinauer Associates; 2004. 255 p.
3. Wiley EO. Phylogenetics: The theory and practice of phylogenetic systematics. United States: John Wiley & Sons; 1981. 575 p.
4. Ankenbrand MJ, Keller A. bcgTree: automatized phylogenetic tree building from bacterial core genomes. Genome. 2016 May;59(10):783-91.
5. Bogusz M, Whelan S. Phylogenetic tree estimation with and without alignment: new distance methods and benchmarking. Syst Biol. 2017 Mar;66(2):218-31.
6. Colijn C, Plazzotta G. A metric on phylogenetic tree shapes. Syst Biol. 2018 Jan;67(1):113-26.
7. Sato M. An application study of DNA structural properties for promoter prediction with wavelet and support vector machine. Procedia Comput Sci. 2018 Nov;140:292-7.
8. Sanderson J, Sudoyo H, Karafet TM, Hammer MF, Cox MP. Reconstructing past admixture processes from local genomic ancestry using wavelet transformation. Genetics. 2015 Jun;200(2):469-81.

9. Inbamalar TM, Sivakumar R. Improved algorithm for analysis of DNA sequences using multiresolution transformation. Sci. World J. 2015 Apr:1-9.
10. Weighill D, Macaya-Sanz D, DiFazio SP, Joubert W, Shah M, Schmutz J, et al. (2019). Wavelet-based genomic signal processing for centromere identification and hypothesis generation. Front Genet. 2019 May;10:1-17.
11. Huang HH, Girimurugan SB. A novel real-time genome comparison method using discrete wavelet transform. J Comput Biol. 2018 Apr;25(4):405-16.
12. Shim H, Stephens M. Wavelet-based genetic association analysis of functional phenotypes arising from high-throughput sequencing assays. Ann Appl Stat. 2015 Feb;9(2):1-22.
13. Oueslati AE, Messaoudi I, Lachiri Z, Ellouze N. A new way to visualize DNA's base succession: the Caenorhabditis elegans chromosome landscapes. Med Biol Eng Comput. 2015 May;53(11):1165-76.
14. Jia J, Xiao X, Liu B. Prediction of protein–protein interactions with physicochemical descriptors and wavelet transform via random forests. J Lab Autom. 2016 Apr;21(3):368-77.
15. National Center for Biotechnology Information (NCBI) [Internet]. Genoma: Prokaryotic reference genomes [updated 2020 Jan 15; cited 2020 Jan 15]. Available from: https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/refseq_category:reference
16. Butler DF, Myers AL. Changing epidemiology of Haemophilus influenzae in children. Infect Dis Clin North Am. 2018 Mar;32(1):119-28.
17. You Y, Davies MR, Protani M, McIntyre L, Walker MJ, Zhang J. Scarlet fever epidemic in China caused by Streptococcus pyogenes serotype M12: epidemiologic and molecular analysis. EBioMedicine. 2018 Feb;28:128-35.
18. Bröms J, Sjöstedt A, Lavander M. The role of the Francisella tularensis pathogenicity island in type VI secretion, intracellular survival, and modulation of host cell signaling. Front Microbiol. 2010 Dec;1:136.
19. Crofts AA, Poly FM, Ewing CP, Kuroiwa JM, Rimmer JE, Harro C et al. Campylobacter jejuni transcriptional and genetic adaptation during human infection. Nat Microbiol. 2018 Mar;3(4):494-502.
20. Weiser JN, Ferreira DM, Paton JC. Streptococcus pneumoniae: transmission, colonization and invasion. Nat Rev Microbiol. 2018 Mar;16(6):355-67.
21. Klemmer J, Njeru J, Emam A, El-Sayed A, Moawad AA, Henning K et al. Q fever in Egypt: Epidemiological survey of Coxiella burnetii specific antibodies in cattle, buffaloes, sheep, goats and camels. PloS One. 2018 Feb;13(2):1-12.
22. Ahn KB, Baik JE, Park OJ, Yun CH, Han SH. Lactobacillus plantarum lipoteichoic acid inhibits biofilm formation of Streptococcus mutans. PloS One. 2018 Feb;13(2):1-16.
23. R Core Team [Internet]. A Language and environment for statistical computing. Vienna, Austria [updated 2020 Jan; cited 2020 Jan 30]. Available from: https://www.R-project.org/
24. Brassarote GON, Souza EM, Monico JFG. Non-decimated Wavelet Transform for a Shift-invariant Analysis. Tend Mat Apl Comput. 2018 Jan-Apr;19(1):93-110.
25. Liu Y, Qin F, Liu Y, Cen Z. A Daubechies wavelet-based method for elastic problems. Eng Anal Bound Elem. 2010 Feb;34(2):114-21.
26. Daubechies I. Ten lectures on wavelets. Philadelphia: Society for industrial and applied mathematics; 1992. 357 p.
27. Ma J, Xue J, Yang S, He Z. A study of the construction and application of a Daubechies wavelet-based beam element. Finite Elem Anal Des. 2003 Jul;39(10):965-75.
28. Liò P, Vannucci M. Finding pathogenicity islands and gene transfer events in genoma data. Bioinformatics. 2000 Oct;16(10):932-40.
29. Gençay R, Selçuk F, Whitcher BJ. An introduction to wavelets and other filtering methods in finance and economics. Elsevier; 2001. 384 p.
30. Ferreira LM, Sáfadi T, Lima RR. Evaluation of genome similarities using the non-decimated wavelet transform. Genet Mol Res. 2017 Aug;16(3):1-12.