Acta Scientific Microbiology (ISSN: 2581-3226)

Volume 8 Issue 2 February 2025

Research Article

Comparison of 3 Bacterial Taxonomic Assignment of Pterygium Samples from Mexican Patients by Metagenomic Analysis

Vega-Arce G[1], Sánchez-Vallejo CJ[1], Salas-Lais AG[2], Alarcón-Hernández E[1] and Bautista-de Lucio VM[3]*

[1]*Instituto Politécnico Nacional, Departamento de Bioquímica, Laboratorio de Genética Molecular, CDMX, México*

[2]*Instituto Mexicano del Seguro Social, Banco de Muestras, CDMX, México*

[3]*Instituto de oftalmología "Conde de Valenciana", Unidad de Investigación, Laboratorio de Microbiología y Proteómica Ocular, CDMX, México*

***Corresponding Author:** Bautista de Lucio VM, Instituto de oftalmología "Conde de Valenciana", Unidad de Investigación, Laboratorio de Microbiología y Proteómica Ocular, CDMX, México.

DOI: 10.31080/ASMI.2025.08.1480

Received: December 03, 2024

Published: January 24, 2025

Abstract

Next-generation sequencing has allowed a better understanding of microbiology and microbiomes. While metagenomic analysis has made it possible to identify microorganisms that were not possible with traditional techniques. One of the platforms for the analysis of microbial samples is Galaxy, an open-source platform with that standardized workflow that facilitates metagenomic analysis, allowing the taxonomic assignment of samples with various databases, such as: PlusPF, Greengenes, SILVA, and RDP. One of the platforms to visualize the data obtained through Galaxy is Pavian, which is also open source that allows the visualization of results in a graphical and statistical way. In such a way that at the level of various systems it has been possible to establish the role of microbiota in the predisposition and development of diseases. In this context the respiratory and digestive systems are the ones that have had the most prominence. But there are other microenvironments that have been explored, which could be involved in the development or certain pathologies, for example the ocular surface; where various studies have established and characterized the healthy microbiota, and others point the specific microbiota present in some diseases. Pterygium is one of the diseases, little researched in this context; this being a disease with tumor characteristics, proliferative, fibrovascular and inflammatory; that invades the cornea and can produce alterations such as astigmatism, decreased visual acuity or even blindness. The objective of this study was to determine if there was a significant difference in bacterial identification, using various databases, in pterygium samples form Mexican patients. The database that yielded the highest number of species-level assignments was PlusPF, followed by Greengenes, while SILVA only achieved assignments at the gender level.

Keywords: Pterygium; Mexican Patients; Metagenomic Analysis; Databases; Ocular Surface; Blindness

Abbreviations

RDP, RDP-II, MID.

Introduction

The human body is colonized by abundant bacteria and these play an important role in the health-disease relationship [1]. The identification of disease-associated pathogens therefore provides information that can guide the most appropriate treatment [2] or a better understanding of the disease process. In this context the traditional techniques for detecting microorganisms became insufficient and ineffective so, gene and genomic sequencing now allows the identification of microorganisms in a more precise way than those previously used [3]. Next-generation sequencing, also called high-throughput sequencing or massive sequencing or massive sequencing, is a technology that allows millions of DNA fragments to be sequenced in parallel and individually [4], generating enormous

amounts of data on the human microbiome to understand its taxonomy and functional properties in the human body [5], particularly in non-cultured organisms [6]. The 16S gene continues to be a suitable target for taxonomic assignment because it has hypervariable regions [7]. For taxonomic assignment, 16S sequences readings were aligned to reference sequences, using databases, such as SILVA, Greengenes, and Ribosomal Database Project (RDP) [8]. Galaxy is an open-source platform for accessible, reproducible, and transparent computational research. It has a wealth of tools, datasets, and high-performance computing resources that form a standard web browsing [9]. Into Galaxy one of the tools to the taxonomic assignment is Kraken [10], in which it is possible to choose various databases such as Greengenes, SILVA, RDP [11]and PlusPF, to varying degrees of reliability. To be able to visualize the results of the taxonomic assignment, it is possible to use the Pavian web application [12].

In this context, infectious diseases continue to significantly damage social security and economic development [13] due to the physical disability that they can produce. Pterygium is a common conjunctival disorder [14], that has only been observed in humans [15], which has well-documented risk factors for its predisposition, but its pathogenesis is not clear [14,16-18]. It is a condition of the ocular surface characterized by a fibrovascular proliferation [16,19], triangular in shape [20], which invades the cornea [16,19], overcoming the sclerocorneal barrier [21] in a process called conjunctivalization [15], with tumor characteristics [22], which alters vision secondary to dry eye, astigmatism and occlusion of the visual axis in severe cases [16] is also the result of a local alteration of ocular surface homeostasis [15], a situation in which the microbiota may be involved. Its name derives from the greek pteros which means wing and describes the shape of the lesion [20,23], it is located either in the nasal or temporal pole or both in one or the two eyes [24], with the nasal location being the most frequent [25]. This article shows the comparison of the different metagenomic data databases of pterygium samples from Mexican patients.

Background

The metagenomic study of the microbiota of the ocular surface has not been focused on pterygium, but for other pathologies of the same anatomical area and for the description of a healthy ocular surface. In 2011 Dong., *et al*. through sequencing and metagenomic analysis and using the Ribosomal Database Project-II (RDP-II) demonstrated the presence of the following Phyla: *Actinobacteria, Bacteroidetes, Cyanobacteria, Firmicutes* and *Proteobacteria*; while the bacterial genera in the conjunctiva of healthy patients were: *Pseudomonas, Bradyrhizobium, Propionibacterium, Acinetobacter, Corynebacterium, Brevundimonas, Staphylococcus, Aquabacterium*and *Streptococcus* [26,27]. In 2016, Huang., *et al*. in the same way, through metagenomic analysis, and through the RDP database, in healthy conjunctivae they tried to establish the normal microbiota, identifying the following Phyla: *Actinobacteria, Firmicutes, Bacteroidetes, Deinococcus-Thermus, Fusobacteria, Cyanobacteria/Chloroplast, Acidobacteria, Candidatus Saccharibacteria* and *Spirochaetes* and the genera identifying were: *Pseudomonas, Staphylococcus, Acinetobacter, Streptococcus, Millisia, Anaerococcus, Finegoldia, Simonsiella* and *Veillonella* [28]. In 2017 Ozkan*., et al*. in a 3 months study demonstrated the stability of the healthy microbiota on the ocular surface, the following Phyla were isolated using traditional techniques in healthy patients: *Actinobacteria, Firmicutes* and *Proteobacteria*; the isolated genera were: *Staphylococcus, Propionibacterium, Micrococcus* and *Corynebacterium*. While, by metagenomic analysis using the RDP database, they identified the following Phyla: *Proteobacteria, Firmicutes* and *Actinobacteria*; the genre identified were: *Corynebacterium, Acinetobacteria, Pseudomonas, Sphingomonas, Streptococcus, Massilia* and *Rothia* [29].

For the specific case of Mexico on the ocular surface, taxonomic assignment results have not yet been reported using traditional techniques where the predominant species in the isolates are: *Estafilococos*, *Estreptococos*, *Pseudomonas*, *Enterococos*, among others [30-32], this related to specific diseases other than pterygium.

Samples

The study group was composed of 6 subjects from Conde de Valenciana Foundation Institute of Ophthalmology with the presence of pterygium in one or both eyes. The samples were taken with sterile cotton swabs exerting light pressure on the conjunctiva with the presence of pterygium, on three sequential occasions to obtain a DNA bank of each subject. The inclusion criteria were subjects with the presence of unilateral, bilateral, monocular or binocular pterygium, pterygiums smaller than 2 mm, age between 30 and 50 years and authorization by informed consent. The exclusion criteria were antecedent of ocular or systemic diseases, use of contact lenses, antecedent of trauma or eye surgery, as well as the use of

systemic or ophthalmic antibiotics in the 3 months prior to sample collection. Finally, the elimination criterium was subjects who did not wish to be included in the study, diagnosis of ocular pathology other than pterygium after sampling, consumption of antibiotics prior to sampling, application of ophthalmic medication prior to sampling.

Materials and Methods

DNA extraction for genomic sequencing

The entire sample of subjects was separately homogenized using the MP instrument FastPrep-24 (MP Biomedicals; speed 6, for 40 seconds). The total DNA was extracted with Nucleospin tissue kit (Macherey-Nagel) and diluted in 70 microliters of molecular biology grade water (DNAse-free).

A fragment of DNA from the initiator codon region was amplified using 11 sets of primers designed in parallel PCRs. A second round of PCRs used the same 11 sets with 454 454 tail-fused initiators and purpose-built multiple identifier (MID) tags.

Each PCR contained 2 microliters of DNA templates, 17.5 microliters of molecular biology grade water, 2.5 microliters 10x of reaction buffer, 1 microliter 50x of magnesium chloride (MgCl2 50 mM), 0.5 ml of deoxyribonucleotide triphosphate mixture (10 mM dNTPs), 0.5 microliters of direct initiator (10 mM), 0.5 microliters of reverse initiator (10 mM), and 0.5 microliters of Invitrogen Platinum Taq polymerase (5U/microliter) in a total volume of 25 microliters.

The PCR conditions were 95°C for 5 minutes, 15 cycles at 94°C for 40 seconds, 46°C for 1 minute and 72°C for 30 seconds after 72°C for 5 minutes. The amplicons were purified with Qiagen´s Minielute PCR purification columns and diluted with 50 microliters of molecular biology grade water. Purified amplicons from the first round of PCR were used as templates in the second round of PCR using initiators with 454 tail-fused primers and purpose-built multiple identifier labels (MIDs) in a 30-cycle amplification regime.

An Eppendorf Mastercycler thermal cycler was used in all PCR reactions.

Illumina MiSeq Amplification and Sequencing of 16S Genetic Regions

Three fragments of 16S genetic region (v3, v4 and v6) were amplified from the homogenized sample using four sets of initiators in parallel PCR reactions for sample. [16SV4F, TGCCAGCAGCCGCGGTAA; 16SV6R, ACGAGCTGACARCCATG; 16sV3F, ACTCCTACGGGAGGCAGCAG; 16SV3R, GGACTACARGGTATCTAAT].

Each PCR contained 2 microliters of DNA templates, 17.5 microliters of molecular biology grade water, 2.5 microliters of 10x reaction buffer, 1 microliter of 50x magnesium chloride (MgCl2 50mM), 0.5 ml of deoxyribonucleotide triphosphate mixture (10 mM dNTPs), 0.5 microliters of direct initiator (10 mM), 0.5 microliters of reverse initiator (10 mM), and 0.5 microliters of Taq Platinum Invitrogen polymerase (5 U/microliter) in a total volume of 24 microliters.

The PCR conditions were: initial denaturation at 95°C for 5 minutes, followed by 15 denaturation cycles at 94°C for 10 seconds, alignment at 46°C for 1 minute and elongation at 72°C for 30 seconds later with a final extension at 71°C for 5 minutes. The amplicons were purified with the Qiagen Minielute PCR reagent set and diluted with 50 microliters of molecular biology grade water. Purified amplicons from the first round of PCR were used as templates in the second round of PCR in which primers fused to Illumina adapters were used in a 10-cycle amplification regimen. An Eppendorf Mastercycler thermal cycler was used in all PCR reactions. Negative controls included in the experiment were included.

Equimolar amounts of the generated libraries were dually indexed, combined and sequenced on an Illumina MiSeq platform using the MiSeq Reagent kits (300 cycles) followed by the 2x300 base pair terminal sequencing protocol.

Metagenomic analysis

The corresponding files were generated for each of the samples, in FASTQ format, which bore the name assigned from the sequencing; differentiating the aforementioned files using a nomenclature for direct chain and another for reverse chain and these were the input data to the Galaxy platform [33].

Using the Make.contigs tool, the contigs were created using the paired data collection, obtaining a total of 923,934 sequences to be analyzed. Next, the Summary.seqs tool was used to obtain the number of unique representative sequences as the total sequences they represented, having the same number of sequences as the previous step. Subsequently, with the Screen.seqs tool, a data cleaning was performed, eliminating contigs with a length greater than 480 bp, of low quality and those with too many ambiguous bases, leaving a total of 581,795 sequences. Continuing with the analysis and using the Unique.seqs tool, duplicate sequences were eliminated to speed up the analysis, obtaining a total of 268,412 unique sequences, noting the significant difference between the total sequences that were available at the beginning of the analysis. Finally, with the Kraken2 tool, the taxonomic assignment was carried out considering parameters such as the selection of the consulted database (Greengenes, PlusPF, SILVA and RDP), choosing a degree of reliability of 0.3 for each of these databases [33]. The taxonomic mapping report files resulting from the execution of the Kraken2 tool were entered into the Pavian platform (https://fbreitwieser.shinyapps.io/pavian/) to visualize the mappings graphically, entering as a parameter the selection of all taxonomic levels.

Results

The total number of sequences on the Galaxy platform was 268,412 but the Pavian platform took an amount for each case as bacterial sequences, being for PlusPF a total of 260,808, Greengenes 263,008, while SILVA there were 263,008, a significant difference with the sequences reported when performing the taxonomic assignment.

With the PlusPF database, the phylum-level mapping was: *Actinobacteria, Bacteroidetes, Candidatus Saccharibacteria, Deinococcus-Thermus, Firmicutes* and *Proteobacteria* (Figure 1), representing 1.83%, 0.35%, 0.0034%, 0.0026%, 2.17%, 0.013% and 85.5% respectively (Graphic 1). The assigned genres were: *Actinomyces, Porphyromonas, Sphingobacterium, Streptococcus, Veillonella, Finegoldia, Brevundimonas, Methylobacterium, Acinetobacter, Pseudomonas* and *Stenotrophomonas* (Figure 1), representing 0.041%, 0.074%, 0.011%, 0.37%, 0.022%, 0.034%, 0.015%, 0.030%, 0.09%, 0.013% and 0.32 % respectively (Graphic 2). With this database, a greater number of taxonomic assignments were obtained at the species level, these were: *Prophyromona asaccharolytica, Prophyromonaasgingivalis, Porphyromonas sp. oral taxon 275, Dolosigranu-*

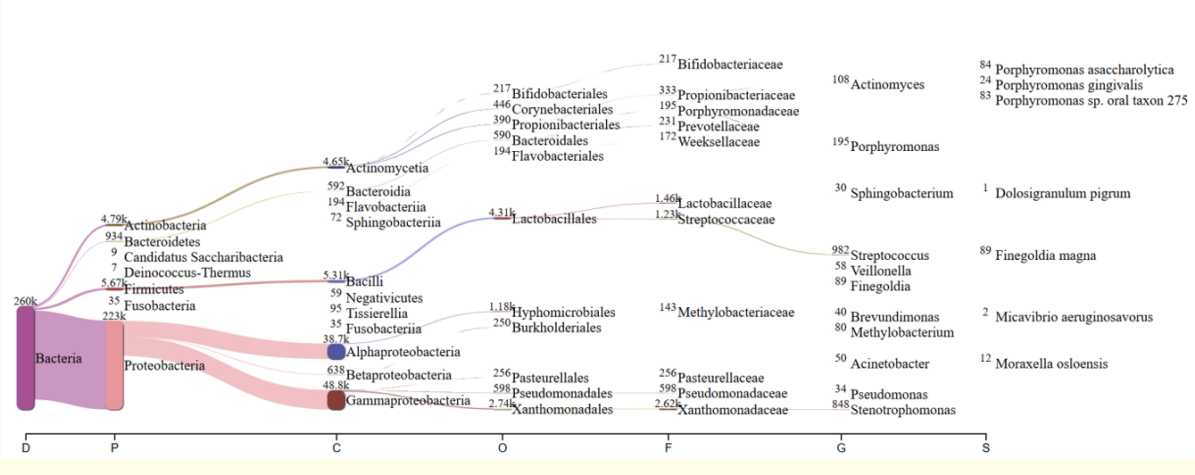
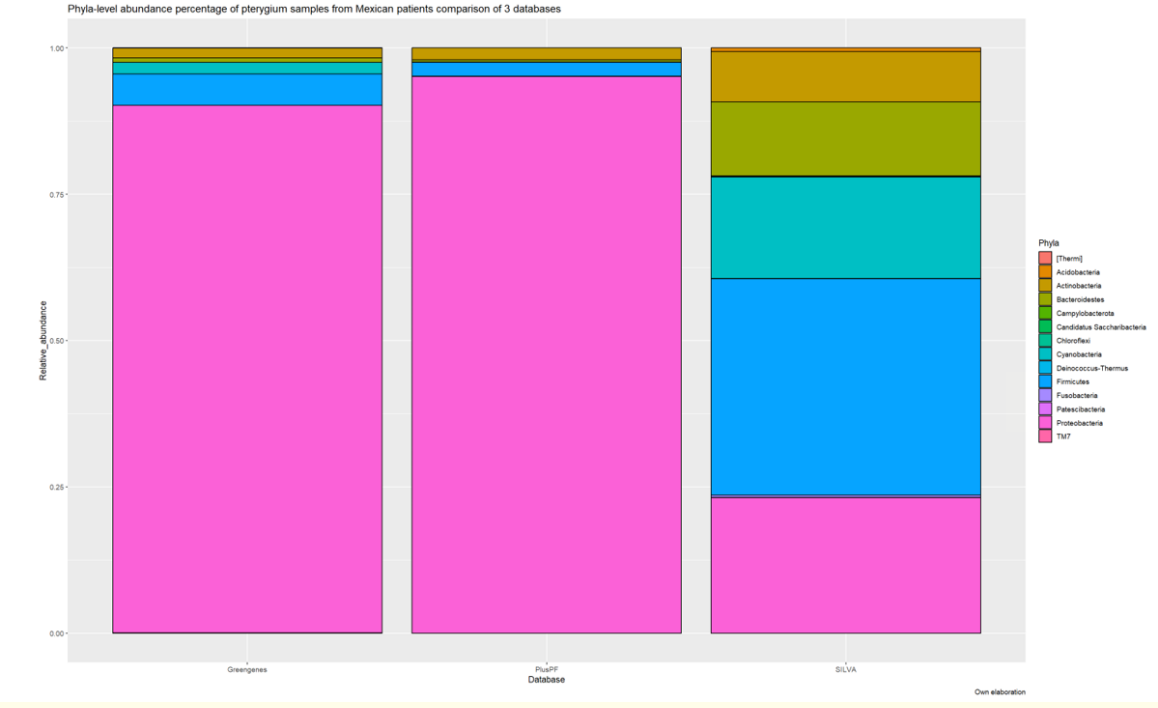


Figure 1: Sankey flow diagram obtained using the Pavian viewer of the taxonomic assignment of the PlusPF database, with 0.3 reliability, of pterygium samples from Mexican patients.



Graphic 1: Stacked bar graph showing the taxonomic assignment at the Phyla level in therms of percentage, of pterygium samples from Mexican patients, from Greengenes, PlusPF and SILVA databases. Graph built using the R programmimg language.

lumpigrum, Finegoldia magna, Micavibrioaeruginosavorus and *Moraxella osloensis*(Figure 1), represented each species a percentage of 0.032%, 0.0092%, 0.031%,0.00038%, 0.034%, 0.00076% and 0.0046% (Graphic 3).

Using the Greengenes database, the following taxonomic mapping were obtained at the phylum level: *Acidobacteria, Actinobacteria, Bacteroidetes, Chloroflexi, Firmicutes, Fusobacteria, Proteobacteria, TM7* and [*Thermi*] (Figure 2), representing 0.035%, 1.3%, 0.56%, 0.0057%, 1.53%, 4.14%, 0.014%, 69.95%, 0.0034% and 0.0038% respectively (Graphic 1); while the assignment at the genera level was: *Actinomyces, Corynebacterium, Porphyromonas, Prevotella, Hymenobacter, Lactobacillus, Anaerococcus, Finegoldia*, [*Eubacterium*], *Leptotrichia, Sneathia, Campylobacter* and *Deinococcus* (Figure 2), representing 0.00038%, 0.029%, 0.042%, 0.18%, 0.0011%, 0.0019%, 0.00076%, 0.022%, 0.0022%, 0.003%, 0.0072%, 0.0019% and 0.00038% respectively (Graphic 2). While the only species-level assignment was: [*Eubacterium*] *biforme* (Figure 2), representing only 0.0015% (Graphic 3).

The SILVA database made the following phylum assignment of the samples in question: *Acidobacteria, Actinobacteriota, Bacteroidota, Campylobacterota, Chloroflexi, Cyanobacteria, Deinococcota, Firmicutes, Fusobacteria, Patescibacteria* and *Proteobacteria* (Figure 3), representing 0.026%, 0.36%, 0.53%, 0.0019%, 0.0057%, 0.72%, 0.003%, 1.55%, 0.014%, 0.0049% and 0.96% respectively (Graphic 1). At the genre level, the assignments were: *Corynebacterium, Porphyromonas, Alloprevotella, Prevotella, Hymenobacter, Campylobacter, Deinococcus, Truepera, Dolosigranulum* and *Leptotrichia*(Figure 3), representing 0.0015%, 0.039%, 0.0011%, 0.00038%, 0.0011%, 0.0019%, 0.00076%, 0.0015%, 0.00038% and 0.003% respectively (Graphic 2). No results were obtained at the species level with this database (Figure 3 and Graphic 3).

The RDP database did not obtain satisfactory results for the taxonomic allocation of pterygium samples from Mexican patients.

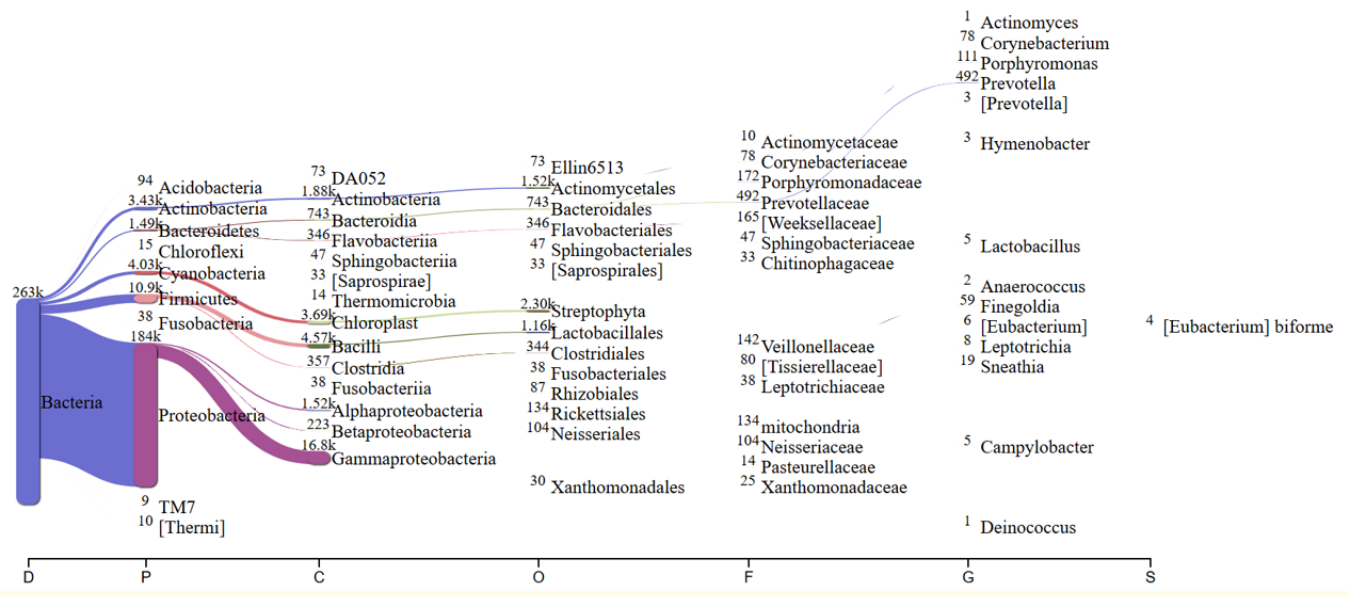
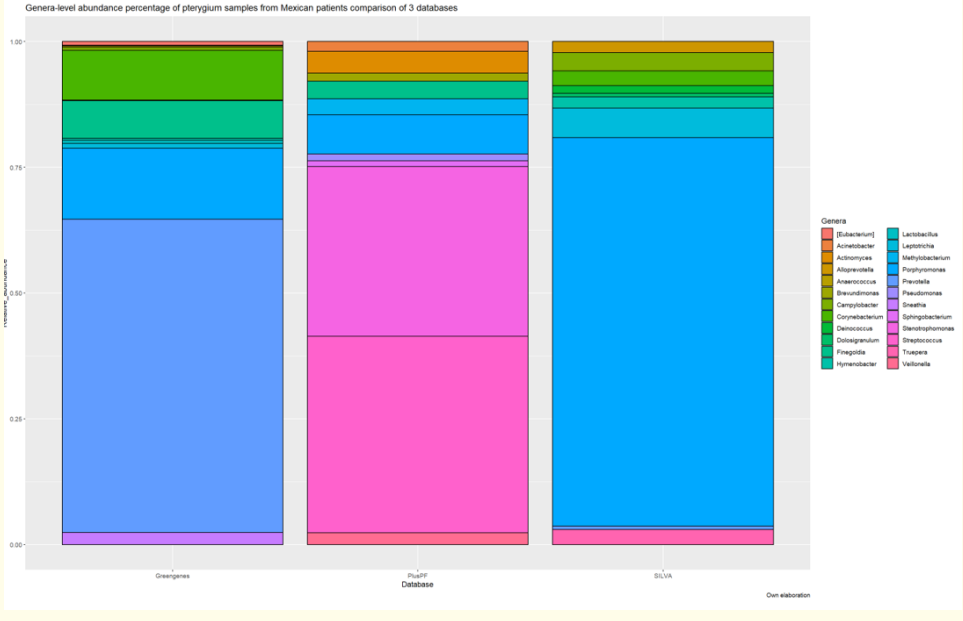


Figure 2: Sankey flow diagram obtained using the Pavian viewer of the taxonomic assignment of the Greengenes database, with 0.3 reliability, of pterygium samples form Mexican patients.



Graphic 2: Stacked bar graph showing the taxonomic assignment at the Gender level in terms of percentage, of pterygium samples from Mexican patients, form the Greengenes, PlusPF and SILVA databases. Graph built using the R programming language.

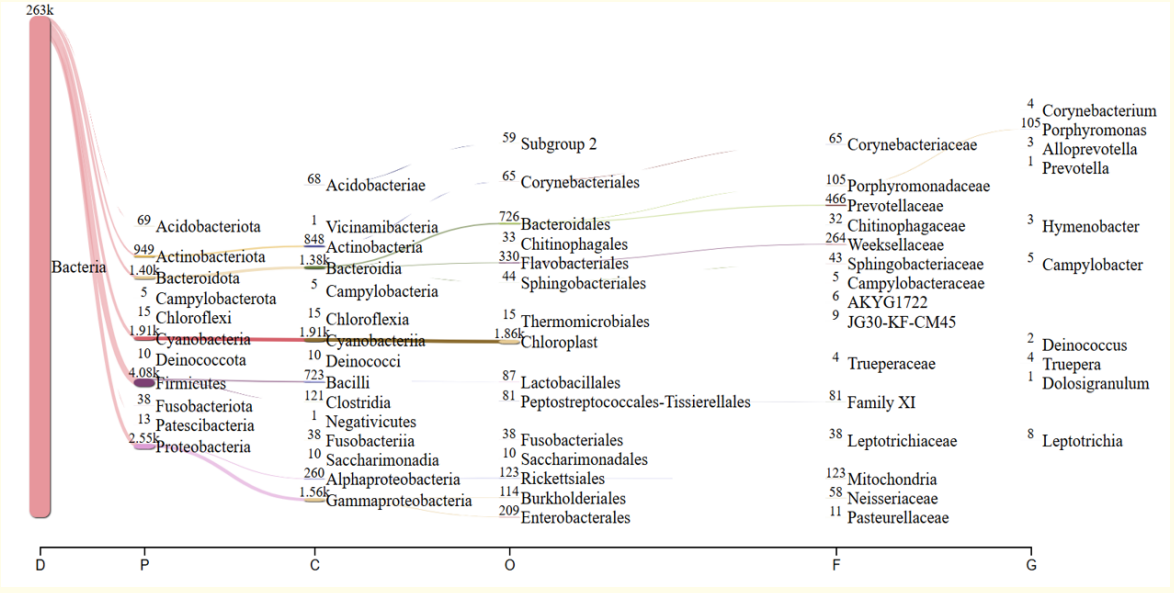
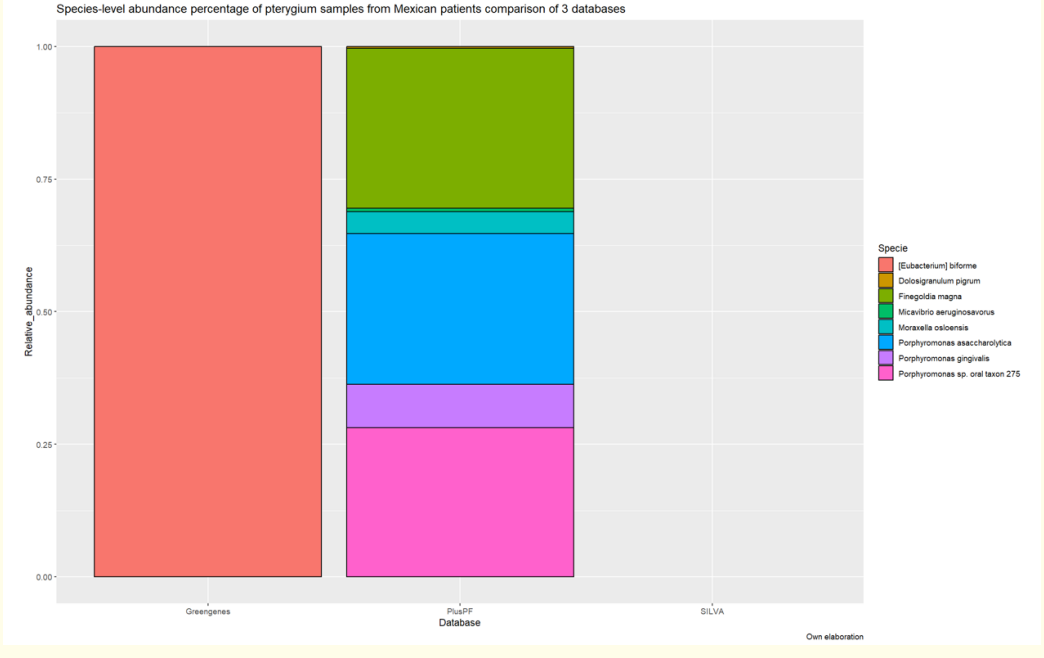


Figure 3: Sankey flow diagram obtained using Pavian viewer of the taxonomic assignment of the SILVA database, with 0.3 reliability, of pterygium samples from Mexican patients.



Graphic 3: Stacked bar graph showing the taxonomic assignment at the Species level in terms of percentage, of pterygium samples from Mexican patients, from the Greengenes, PlusPF, and SILVA databases. Graph built using the R programming language.

Taxonomic assignment of pterygium samples from Mexican patients using 3 databases			
Database	**Taxonomic level**	**Taxonomic assignment**	**Abundance percetage**
PlusPF	Phylum	Actinobacteria	1.836600104
		Bacteroidetes	0.358117849
		Candidatus Saccharibacteria	0.003450814
		Deinococcus-Thermus	0.002683967
		Firmicutes	2.174013067
		Fusobacteria	0.013419834
		Proteobacteria	85.50351216
	Gender	Acinetobacter	0.019171191
		Actinomyces	0.041409773
		Brevundimonas	0.015336953
		Finegoldia	0.03412472
		Methylobaterium	0.030673906
		Porphyromonas	0.074767645
		Pseudomonas	0.01303641
		Sphingobacterium	0.011502715
		Stenotrophomonas	0.325143401
		Streptococcus	0.376522193
		Veilonella	0.022238582
	Specie	*Dolosigranulum pigrum*	0.000383424
		Finegoldia magna	0.03412472
		Micavibrio aeruginosavorus	0.000766848
		Moraxella osloensis	0.004601086
		Porphyromonas asaccharolytica	0.032207601
		Porphyromonas gingivalis	0.009202172
		Porphyromonas sp. oral taxon 275	0.031824177
Green genes	Phylum	Acidobacteria	0.035740358
		Actinobacteria	1.30414284
		Bacteroidetes	0.566522691
		Chloroflexi	0.005703249
		Cyanobacteria	1.532272783
		Firmicutes	4.144360628
		Fusobacteria	0.01444823
		Proteobacteria	69.95984913
		TM7	0.003421949
		[Thermi]	0.003802166

Green genes	Gender	Actinomyces	0.000380217
		Anaerococcus	0.000760433
		Campylobacter	0.001901083
		Corynebacterium	0.029656893
		Deinococcus	0.000380217
		Finegoldia	0.022432778
		Hymenobacter	0.00114065
		Lactobacillus	0.001901083
		Leptotrichia	0.003041733
		Porphyromonas	0.042204039
		Prevotella	0.187066553
		Sneathia	0.007224115
		[*Eubacterium*]	0.002281299
	Specie	[*Eubacterium*] biforme	0.001520866
SILVA	Phylum	Acidobacteriota	0.026234943
		Actinobacteriota	0.360825526
		Bacteroidota	0.5323032
		Campylobacterota	0.001901083
		Chloroflexi	0.005703249
		Cyanobacteria	0.726213651
		Deinococcota	0.003802166
		Firmicutes	1.551283611
		Fusobacteriota	0.01444823
		Patescibacteria	0.004942815
		Proteobacteria	0.969552257
	Gender	Corynebacterium	0.001520866
		Porphyromonas	0.03992274
		Alloprevotella	0.00114065
		Prevotella	0.000380217
		Hymenobacter	0.00114065
		Campylobacter	0.001901083
		Deinococcus	0.000760433
		Truepera	0.001520866
		Dolosigranulum	0.000380217
		Leptotrichia	0.003041733
	Specie	No taxonomic assignment	

Table 1: The percentages of abundance by database of pterygium samples from Mexican patients are shown, as well as the taxonomic assignment and taxonomic level.

Discussion

The phylum with the highest representation for the PlusPF and Greengenes databases was Proteobacteria with an average of 200,000 sequences, while in the case of SILVA it was represented by only 2550 sequences.

Although the PlusPF database obtained results at the species level, this is not the most up-to-date database, that place corresponds to SILVA, a situation that gives it greater credibility when using it to make a taxonomic assignment.

The only database that identified nonbacterial sequences was SILVA where chloroplasts were assignment at the order level.

As can be seen in Table 1, the percentages of taxonomic assignment are low, this is due to the low microbial biomass on the ocular surface; with the consequent low level of DNA; which represents a challenge for the characterization of the ocular microbiome [34], which is secondary to the fact that the ocular surface is a small anatomical region compared to the respiratory tract, to name a few examples.

While the number of unassigned sequences could be related to contamination problems of samples that exceed the biological signal [34], on the other hand, the full execution of the Galaxy workflow was not carried out, since it is designed to make alignment with the SILVA database only. Therefore, the cleaning of missing data could influence taxonomic mapping.

Although there is no cut-off point assignment to determine the representation of taxonomic assignments in terms of percentage with respect to total sequences, there are studies where it is shown as representative from 0.5% representation of the total sequences per sample [26,35].

The greatest similarities at the phylum level of what was reported both in healthy conjunctiva and conjunctiva with some pathology compared to our results are: *Actinobacteria, Firmicutes, Bacteroidetes* and *Proteobacteria*, highlighting the representation of this phylum compared to all the others, as mentioned above.

In the case of the Mexican population, we found at the species level with the PlusPF database a predominance of porphyromonas while in past studies the predominance was given by staphylococci. The comparison with the Greengenes and SILVA databases cannot be carried out due to the null results at the species level obtained through these databases, remembering that the results at this reported in the Mexican population were obtained by isolation with traditional techniques.

The assignment of non-bacterial sequences using the SILVA database demonstrates firstly the richness of the database and secondly the need for more data cleaning for the elimination of unwanted sequences including chloroplasts, mitochondria and chimeric sequences, mainly.

In the case of PlusPF and Greengenes, although they achieved an assignment up to the species level, the former more than the latter. Works published by sequencing and metagenomic analysis do not report results up to the species level only at the genus level, unlike works where traditional techniques were used, where results are reported up to the species level.

Thus, the most reliable database to perform this kind of taxonomic assignments by informatics means is: SILVA because compared to other taxonomies based on the 16S gene it is larger, sharing almost all taxonomic assignments with the National Center for Biotechnology Information (NCBI) [8].

Although the execution of the workflow on the Galaxy platform was not completed, the results obtained are indicative towards the creation of microbiological profile of pterygium samples from Mexican patients, a field that has not been developed to date.

It is evident that the results presented here, although they coincide to a certain degree with what has been reported in healthy conjunctiva and conjunctiva with other diseases, need to be compared with negative control to demonstrate whether there is a significant difference between conjunctiva with the presence of pterygium and conjunctiva without the presence of pterygium, to demonstrate whether the microbiota and in turn the microbiome, they have a leading role in this inflammatory and proliferative process as is pterygium; making possible the following assertion: microbial communities are specific to each region of the human body and changes in them can be a factor in the development of diseases [36].

It is worth noting that we performed microbiological analysis only of bacteria, without considering microorganisms such as viruses, fungi, parasites, etc.

Conclusion

The scientific reports support the SILVA database as the most robust for performing taxonomic assignments in metagenomic analysis and RDP, having greatest representation with history of bacterial identification on the ocular surface, but in this work, the database that performed the best assignment was PlusPF.

The continuation of research using modern molecular biology techniques such as sequencing and metagenomic analysis, to elucidate microenvironments, is essential to understand the health-disease process, describing new mechanisms of interaction and coexistence that allow the development of new diagnostic techniques and therapeutic.

Acknowledgements

Al Consejo Nacional de Humanidades Ciencias y Tecnologías por la beca otorgada para la realización del presente trabajo, así como al Instituto de Oftalmología Fundación "Conde de Valenciana" y a la Escuela Nacional de Ciencias Biológicas (ENCB) del Instituto Politécnico Nacional (IPN) por el apoyo brindado.

Conflict of Interest

The authors declare "not a conflict of interest".

Bibliography

1. Schlunck G., *et al*. "Next-generation sequencing of the human aqueous humour microbiome". *International Journal of Molecular Sciences* 25 (2024): 6128.

2. Watts G and Hurwitz B. "Metagenomic next-generation sequencing in clinical microbiology". *Microbiology Newsletter* 42 (2020): 53-59.

3. Matysiak A., *et al*. "Characterization of Ocular Surface Microbial Profiles Revealed Discrepancies between Conjunctival and Corneal Microbiota". *Pathogens (Basel, Switzerland),* 10 (2021): 405.

4. Gu W., *et al*. "Clinical metagenomic next-generation sequencing for pathogen detection". *Annual Review of Pathology* 14 (2019): 319-338.

5. Kim S., *et al*. "Hierarchical structured component analysis for microbiome data using taxonomy assignments". *EEE/ACM Transactions on Computational Biology and Bioinformatics* 19 (2022): 1302-1312.

6. Valenzuela F., *et al*. "El gen ARNr 16S en el estudio de comunidades microbianas marinas". *Ciencias Marinas* 41 (2015): 297-313.

7. Durazzi F., *et al*. "Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gun microbioma". *Scientific Reports* 11 (2021): 3030.

8. Balvočiūtė M and Huson D. "SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare?". *BMC Genomics* 18 (2017): 114.

9. Hiltemann S., *et al*. "Galaxy Training: A powerful framework for teaching!". *PLoS Computational Biology* 19 (2023): e1010752.

10. Lu J., *et al*. "Metagenome analysis using the Kraken software suite". *Nature Protocols* 17 (2022): 2815-2839.

11. Walker M. "Metagenomic Data Analysis". (1st ed.). (S. Mitra, Ed.) Humana (2023): 55-67.

12. Breitwieser F and Salzberg S. "Pavian interactive analysis of metagenomics data for microbiome studies and pathogen identification". *Bioinformatics* 36 (2020): 1303-1304.

13. Wang C., *et al*. "The clinical application of metagenomic next-generation sequencing in infectious diseases at a tertiary hospital in China". *Frontiers in Cellular and Infection Microbiology* 12 (2022): 957073.

14. Chu W., *et al*. "Pterygium: new insights". *Eye* 34 (2020): 1047-1050.

15. Shahraki T., *et al*. "Pterygium: an update on pathophysiology, clinical features, and management". *Therapeutc Advances in Ophthalmology* 13 (2021): 1-21.

16. Tong L., *et al*. "Conjuntivochalasis is the precursor to pterygium". *Medical Hypotheses* 81 (2013): 927-930.

17. Wanzeler AC., *et al*. "Mechanisms and biomarker candidates in pterygium development". *Arquivos brasileiros de oftalmologia* 82 (2019): 528-536.

18. Cárdenas C., *et al*. "Molecular Basis of Pterygium Development". *Seminars in Ophthalmology* 31 (2016): 567-583.

19. Taher N., *et al*. "Amniotic membrane transplantation and conjunctival autograft combined with mitomycin C for the management of primary pterygium: A systematic review and meta-analysis". *Frontiers in Medicine* 9 (2022): 1-13.

20. Cubillo A and Chacón E. "Pterigión: conceptos y manejo actual". *Revista Médica Sinergia* 6 (2021): e740.

21. Lluch S., *et al*. "What biomarkers explain about pterygium OCT pattern". *Graefe´s Archive for Clinical and Experimental Ophthalmology* 254 (2016): 143-148.

22. Zavala J., *et al*. "Expresión de genes del metabolismo de lípidos y dolesterol en fibroblastos de pterigión: análisis comparativo con adipocitos y otros fibroblastos". *Revista Mexicana De Oftalmología* 93 (2019): 7-13.

23. Sati A., *et al*. "Mini-simple limbal epithelial transplantation versus conjunctival autograft fixation with fibrin glue after pterygium excision: a randomized controlled trial". *Cornea* 38 (2019): 1345-1350.

24. García KP., *et al*. "Correlación morfológica del pterigión y su evolución clínica". *Revista del Hospital General de México,* 69 (2006): 205-211.

25. Singh S. "Pterygium: epidemiology prevention and treatment". *Community Eye Healht* 30 (2017): S5-S6.

26. Dong Q., *et al*. "Diversity of bacteria at healthy human conjunctiva". *Investigative Ophthalmology and Visual Sciences* 52 (2011): 5408-5413.

27. Lu L and Liu J. "Human microbiota and ophthalmic disease". *The Yale Journal of Biology and Medicine* 89 (2016): 325-330.

28. Huang Y., *et al*. "Defining the normal core microbiome of conjunctival microbial communities". *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 22 (2016): 643. e7-643.e12.

29. Ozkan J., *et al*. "Temporal stability and composition of the ocular surface microbiome". *Scientific Reports* 7 (2017): 9880.

30. Córdova J., *et al*. "Determinación de la flora bacteriana conjuntival en el preoperatorio de la cirugía de catarata". *Revista Médica del Hospital Genera de México* 71 (2008): 77-82.

31. Hernández JC., *et al*. "Queratitis infecciosa: tendencias microbiológicas y sensibilidad a antibiótios. Segundo reporte anual del grupo de estudio de microbiología ocular del Instituto de Oftalmología "Conde de Valenciana"". *Revista mexicana de oftalmología* 87 (2013): 100-109.

32. Martin V., *et al*. "Caracterización de la microbiota conjuntival transitoria y residente de adultos jóvenes". *Ciencia y Tecnología para la Salud Visual y Ocular* 15 (2017): 37-45.

33. Community T. "The Galaxy platform for accesible, reproducible and collaborative data analyses: 2024 update". *Nuleic Acids Research* (2024): W83-W84.

34. Baim A., *et al*. "The microbiome and ophthlamic disease". *Experimental Biology and Medicine* 244 (2019): 419-429.

35. Petrillo F., *et al*. "Current evidence on the ocular surface microbiota and related diseases". *Microorganisms* 8 (2020): 1033.

36. Zilliox M and Bouchard C. "The microbiome ocular surface, and corneal disorders". *American Journal of Pathology* 193 (2023): 1648-1661.