Study of Terpenes in Actinobacteria: Diversity and Bioactive Potential

Gabriel Batista Almeida

Received October 17, 2024
Accepted January 31, 2025
Electronic access February 28, 2025

For many years, the same compounds have been employed to treat various diseases, particularly in antibacterial therapies. As a result of the lack of new molecules, microorganisms have developed resistance to these drugs, diminishing their effectiveness. Thus, finding and creating new molecules for medicine is crucial to address the increasing number of resistant organisms. In order to develop these treatments, it is fundamental to understand the genetic mechanisms behind the production of such metabolites. Therefore, this study aims to address the lack of genetic information of terpenes' production - a metabolite produced by bacteria with medical effects - specifically in Actinobacteria - a phylum of bacteria renowned for producing molecules that humans can use in medicine. The data was collected from UniProt and was analyzed in the Cytoscape platform to describe the diversity of families and genera involved in the production of terpenes. The results show that *Streptomyces* is the most diverse genus, having the greatest number of clusters (groups of genes that encode similar proteins), but also underscore the potential of less studied genera. More research should be done with these bacteria in order 1) to obtain more data about terpenes isolated from Actinobacteria and 2) to exploit all the potential of this group of bacteria.

Introduction

Background

For decades, humans have been using the same compounds to treat several diseases, especially when it comes to antibacterial treatments. Consequently, microorganisms acquire resistance against these drugs, which weakens their effect.[1]. Therefore, it is crucial to discover and synthetize novel molecules for medicine to combat the large number of resistant organisms. One genus vastly used for this purpose of producing novel compounds for decades is *Streptomyces*.

Streptomyces is a genus of Actinobacteria that has been a resource of bioactive natural products and medical chemicals for humans since 1947. However, since the 1960s, the number of research with these bacteria and of novel compounds discovered have been decreasing. Nevertheless, the article "Recently Discovered Secondary Metabolites from *Streptomyces* Species"[2] underscore that this genus still has an enormous potential of producing over 150,000 more novel compounds and describe the antibiotic activities of recently discovered molecules.

One of the reasons behind not achieving all the *Streptomyces*' potential may be not understanding completely the reasons behind the production of secondary metabolites. For example, the paper "Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis"[3] claims that ecological and evolutionary forces are related to the production of a high number of metabolites that have not been previously linked to a certain microorganism. In other words, some compounds may only be synthesized when a bacterium is exposed to environmental stresses. One of these stresses could be a competitive environment, which means that microbial interactions (probiotics and prebiotics), may lead to the selection of community members with antagonistic properties against pathogens[4]. The regulation of gene expression in response to the fluctuations in the density of cell-population is known as Quorum sensing, and in bacteria it is used to regulate their physiological activities.[5]. Not only that, but the chemical signaling with other members of the community is believed to prompt the production of natural products as well. The article "Small molecule inducers of Actinobacteria natural product biosynthesis"[6] states that novel compounds were produced by Actinobacteria after they were exposed to different inducers, such as antibiotics - protein synthesis inhibitors, DNA damaging agents, fatty-acid synthesis inhibitors, and disruptor of cell wall integrity - and hormones/autoinducers. Therefore, in order to exploit all the pharmacological potential of Actinobacteria, it is relevant to expose these organisms to populational stresses or to chemical signals.

Moreover, according to the review paper "State-of-the-art methodologies to identify antimicrobial secondary metabolites in soil bacterial communities – A review,"[7] mimicking the environmental conditions and their stimuli when cultivating samples is extremely relevant to the discovery of unknown molecules. Moreover, as important as the environmental data is the genetic data. Experts have been investigating the mechanisms of DNA expression to produce new medicines through prokaryotic cells. One example is the production of human insulin by *Escherichia coli*[8]. Consequently, knowing the relation between protein and gene will be essential to the

production of the recently discovered compounds.

Additionally, extreme habitats are an unexploited source for the discovery of new drugs. For example, novel species of Actinobacteria were isolated from deserts, produced novel specialized metabolites, and contained stress-related genes that encode for properties, such as desiccation, ionizing radiation, lack of carbon source, temperature changes, osmotic stresses, UV radiation, and pH.[9]. Consequently, because the number of new species of Actinobacteria is increasing in biomes with harsh conditions, these habitats - and mainly their environmental aspects - may comprise a rich variety of novel compounds with biological compounds[10]. Therefore, these conditions should be mimicked in order to activate genes related to these specific environmental stresses, and produce compounds that might have pharmacological properties.

Actinobacteria

Actinobacteria are a group of gram-positive filamentous bacteria[11]. They are ubiquitous in both aquatic and terrestrial environments[12]. Actinobacteria can be found on the soil's surface, and their presence depends on various factors, such as pH, temperature and soil moisture. Moreover, Actinobacteria are commonly found in environments with low or medium salt concentrations. Another habitat where Actinobacteria can be isolated from is the rhizosphere, a region in the roots of plants, but also in other plants' tissues (endophytic Actinobacteria). These organisms also inhabit the sea surfaces, usually 10m deep. These microorganisms can also be isolated from inhospitable habitats, such as geothermal and volcanic soils and deserts.[13]. Actinobacteria were also found in ice cores, which shows the adaptations of such phylum in extreme low temperatures[14]. In a nutshell, they are found in marine ecosystems, inland waters, deserts, cryo-environments, geothermal sites, soils, and environments impacted by anthropocentric actions[15].

This phylum is the source of about 45% of molecules obtained from microorganisms.[16]. From the 1950s to 1970s, approximately 60% of new antibiotics were predominantly isolated from Streptomycetes, which represents the most renowned group of Actinobacteria. However, even the non-*Streptomyces* genera (e.g. *Sinomonas*, *Microbacterium*, *Nocardia*) have shown growing potential in discovering novel secondary metabolites"[17]. Hence, studying a great range of genera of this phylum may be crucial to find novel compounds.

Terpenes

One of the compounds produced by Actinobacteria is terpenes[18]. Terpenes are the largest and most diverse class of specialized metabolites on the planet. To date more than 70,000 terpenoids have been described and classified into more than 400 structural families. One crucial aspect of terpenes is that the vast majority of them have been isolated from plants and fungi. Therefore, while the plant and fungal biosynthetic pathways are well studied, the bacterial pathway was studied to a lesser degree, because of the long-held perception that bacteria were not capable of producing complex terpenoids.[19].

The structural diversity of terpenoids reflects the range of their functional roles, which vary from cholesterol, vitamins A and D, carotenoids, and steroid hormones, pheromones, fragrances, and defense metabolites. Many of the more specific compounds possess significant biological activities, like the anticancer agent Taxol or the antimalarial agent Artemisinin.[19]

Based on the number of carbon atoms, these molecules can be classified as: Hemiterpenes (C5H8), Monoterpenes (C10H16), Sesquiterpenes (C15H24), Diterpenes (C20H32), Sesterpenes (C25H40), Triterpenes (C30H48), and Tetraterpenes (C40H64)[20].

terpenes isolated from Actinobacteria and their activities

The UniProt database has 17 families (*Streptomycetaceae, Thermomonosporaceae, Streptosporangiaceae, Pseudonocardiaceae, Micromonosporaceae, Nocardiopsaceae, Kribbellaceae, Nocardiaceae, Frankiaceae, Actinopolymorphaceae, Dermacoccaceae, Geodermatophilaceae, Thermoactinomycetaceae, Gaiellaceae, Glycomycetaceae, Kineosporiaceae, Nocardioidaceae*) and 48 genera (*Streptomyces, Actinomadura, Nonomuraea, Micromonospora, Kitasatospora, Saccharopolyspora, Streptosporangium, Amycolatopsis, Saccharothrix, Crossiella, Sphaerisporangium, Nocardiopsis, Kribbella, Thermomonospora, Yinghuangia, Actinocorallia, Kutzneria, Microbispora, Nocardia, Actinoplanes, Allocatelliglobosispora, Longispora, Asanoa, Frankia, Prauserella, Streptomonospora, Actinoallomurus, Actinopolymorpha, Allostreptomyces, Halosaccharopolyspora, Plantactinospora, Saccharomonospora, Salinispora, Shimazuella, Allobranchiibius, Flexivirga, Gaiella, Haloactinospora, Kineosporia, Lipingzhangella, Modestobacter, Petropleomorpha, Pseudonocardia, Spinactinospora, Stackebrandtia, Thermasporomyces, Thermopolyspora*) of Actinobacteria related to the production of terpenes.

The terpenes isolated from Actinobacteria are extremely relevant to medicine, as described in table 1:

It is interesting to note the quantity of terpenes that had an activity against other microorganisms, such as bacteria and fungi. This is an evidence that the production of these terpenes is essential to the survival and competitiveness of those organisms in their environments.

However, despite the great potential of terpenes found in Actinobacteria, it is observed from online databases, specifically UniProt, that these compounds are not as extensively studied as

Table 1: Few examples of terpenes isolated from Actinobacteria and their biological activities.

Compound	Activity	Organism	Reference
11, 12, 13-trinor-1,5-eudesmanediol	It demonstrated a moderate activity against Candida	*Streptomyces* spp	[21]
Albaavenone	It demonstrated antibiotic activity against Bacillus subtilis	*Streptomyces abidoavus*	[21]
(+)-caryolan-1-ol	It demonstrated moderate antifungal activity against Botrytis cinerea	*Streptomyces* spp	[21]
Brasilicardin A	It demonstrated cytotoxic activity against cancer cells	*Nocardia terpenica*	[21]
Oxaloterpins A	It showed antibacterial activity	*Sterptomyces* sp KO-3988	[21]
1,8-Cineole	It showed anti-inflammatory, and antioxidant activities	*Streptomyces clavuligerus* ATCC 27064	[22]
Linalool	It showed anticancer and antimicrobial activities	*Streptomyces clavuligerus* ATCC 27064 and *Streptomyces* sp. GWS-BW-H5	[22]
Bicyclogermacrene	It showed antibacterial and antifungal activities	*Streptomyces Xinghaiensis* S187	[22]

other molecules, and there are gaps in the current understanding of their production. Many of the compounds produced by bacteria (antibiotics e.g.) are believed to be important for their survival, acting as mediators of resource competition in a competitive environment. Therefore, microbial interactions have a key role in their activation. Hence, researchers need to elucidate the triggers and cues that activate the expression of these genes to increase the success of natural product-based drug discovery.[23]

Therefore, this study aims 1) to catalog the presence and absence of information about these compounds, and 2) describe the diversity of terpenes produced in the five families with the most number of entries. This will make the process of discovering new terpenes more efficient, by not only making a roadmap showing the gaps in available data, but also demonstrating important families for future research.

Methods

The data was obtained by searching for "Terpene synthase Actinobacteria"in the UniProt database. Then, through Excel, 1) the organisms described were divided into families and genera, 2) the researchers counted the number of entries for each genus and family, 3) calculated Shannon Diversity Index to evaluate the biodiversity of genera related to the production of terpenes 4) provided the mean, standard deviation, and coefficient of variation of each family, 5) made a linear regression graphic relating the number of genera in a family and the number of terpenoids produced by this family, and 6) used Pearson's correlation to calculate how strong the correlation presented in the graphic is.

With the processed data, the "Entry"column was copied and input into the Enzyme Similarity Tool, which grouped the proteins based on similarity. To create cohesive groups, a minimum of 50% similarity between the proteins (score of 79) and 95% identity was used. This percentage of similarity was used, as proteins that are more than 40% are likely to share the same functional similarities[24], and as it formed more cohesive groupings, which helped the study for similar proteins. Also, the percentage of identity was used to assure that duplicate proteins would not be present in those groups. In addition to downloading the data to Cytoscape, it was also transferred to the Genome Neighborhood Tool within the Enzyme Similarity Tool. The Cytoscape application was used to visualize the connection networks between the enzymes. By searching for the IDs found in each group from Cytoscape in the Genome Neighborhood Tool, it was possible to analyze the genes behind these clusters (groups of genes that encode similar proteins). In the end, to discover what was the compound being produced by each cluster, the authors copied its protein sequence and then pasted it on the Protein Blast Tool.

Results and Discussion

As of July 4th 2024, there were 924 protein entries related to terpene synthase in Actinobacteria cataloged on UniProt's database. After analyzing the data using Excel, it was observed that the organisms related to these terpenes were spread into 17 families and 48 genera. Among the families, the ones with the greatest number of entries were: *Streptomycetaceae* (354), *Thermomonosporaceae* (139), *Streptosporangiaceae* (136), *Pseudonocardiaceae* (127), and *Micromonosporaceae* (92) as shown in Figure 1.

Delving deeper into the data, among the 48 genera, the ones with the highest number of entries were *Streptomyces* (286), *Actinomadura* (122), *Nonomuraea* (82), *Micromonospora* (73), *Kitasatospora* (60) as represented in Figure 2.

Although there are not many articles that state why those genera are more prevalent than the other, there are several studies that state the reasons behind the general interest in these genera. In a nutshell, the genera with most entries are mainly the ones with a great potential of producing secondary metabolites (not limited to terpenes), but also with applications in other fields.

Streptomyces, for example, are abundant in the soil - participating in the degradation of organic matter -, have a wide phylogenetic spread - and a genetic diversity may suggest a

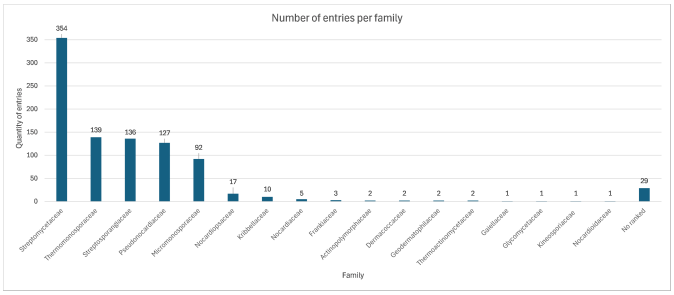

Fig. 1 Numbers of entries per family. The graphic reinforces the idea that only a few families are well explored, while the vast majority is not widely exploited. However - although not as significant as *Streptomycetaceae - Thermomonosporaceae, Streptosporangiaceae, Pseudonocardiaceae*, and *Micromonosporaceae* demonstrate a great potential of producing terpenes, and should be more explored.

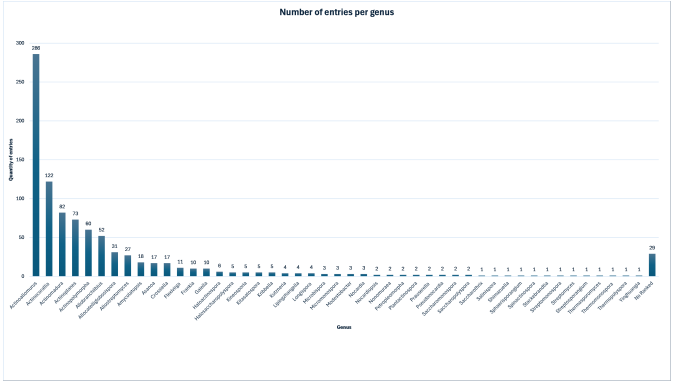

Fig. 2 Numbers of entries per genus. This graphic presents *Streptomyces* as the genus with the most number of entries. This graphic can also be used to identify genera that seem to be productive, but that are not widely explored as *Streptomyces* is, such as *Saccharopolyspora*, *Streptosporangium*, and *Amycolatopsis*.

broad range of clusters-, and produce a large and diverse number of secondary metabolites.[25].

In the case of *Actinomadura*, there are several studies about this group, as this genus inhabits a wide range of environments - such as marine and terrestrial -, and synthesizes numerous diverse bioactive compounds - that are not limited to terpenoids only -, representing an attractive target for research related to drug discovery.[26]. Another particular aspect of this genus is that it was found some individuals that are resistant to some drugs, such as the strain *Actinomadura geliboluensis*, which attracts the attention of researchers that aim to discover the reasons for that, and how it can help to the discovery of more efficient compounds.[27]. Finally, this genus is also related to human diseases, such as Mycetoma, which represents a major target of research that seeks to find treatment for it.[28].

The genus *Nonomuraea* has a great interest, as it is considered a large, and unexplored source for specialized metabolites.[29]. There have also been some studies about the activation of dormant genes in this genus that could lead to novel compounds. This interest is justified, as this genus can be found in extreme environments, such as deep-sea sediments, so the environmental stresses may lead to the activation of these dormant genes, and the production of new proteins.[30].

The interest for *Micromonospora* is explained by their applications in medicine as antibiotic producer, and suppressor of pathogens in vitro and in planta[25]. They are also very well studied because of their applications in agriculture, and on the production of energy, as the genus is considered a rich source for biocontrol agents, biofuels, plant growth products and plant probiotics.[31].

Lastly, the main motivation behind the research with *Kitasatospora* is its unexplored potential of producing numerous specialized second metabolites, which is not limited to terpenoids.[32]

Moreover, with the previous analysis, it was possible to determine how many and what are the genera present in each family, as described in Table 2. The most diverse families in terms of number of genera are *Pseudonocardiaceae* (9 genera), *Micromonosporaceae* (7 genera), *Nocardiopsaceae* and *Streptosporangiaceae* (both with 5 genera).

However, this previous diversity only accounts for genera richness (the number of genera), and does not take in count genera evenness (the relative abundance of the genus in the family). Therefore, the Shannon Diversity Index was calculated to evaluate these two aspects together (equation 1). According to this formula, a higher value for H' indicates that this family is more diverse. The results are shown in table 3.

$$H' = -\sum PiLN(Pi)$$

Equation 1: Shannon Diversity equation quantifies the biodiversity of a population. In this case, "H'" stands for biodiversity; "Pi" relates to the relative population (in this case,

Table 2: Overview of the number of genera present in each family, and the number entries related to the production of terpenoids in each family and genus.

Family (number of entries in the family)	Quantity of Genera	Genera (number of entries in the genus)
Pseudonocardiaceae (127)	9	*Amycolatopsis* (27), *Crossiella* (17), *HaloSaccharopolyspora* (2), *Kutzneria* (5), *Prauserella* (3), *Pseudonocardia* (1), *Saccharomonospora* (2), *Saccharopolyspora*(52), *Saccharothrix* (18)
Micromonosporaceae (92)	7	*Actinoplanes* (4), *Allocatelliglobosispora* (4), *Asanoa* (3), Longispora (4), *Micromonospora* (73), *Plantactinospora* (2), Salinispora (2)
Nocardiopsaceae (17)	5	*Haloactinospora* (1), *Lipingzhangella* (1), *Nocardiopsis* (11), *Spinactinospora* (1), *Streptomonospora* (3)
Streptosporangiaceae (136)	5	*Microbispora* (5), *Nonomuraea* (82), *Sphaerisporangium* (17), *Streptosporangium*(31), *Thermopolyspora* (1)
Streptomycetaceae (354)	4	*Kitasatospora* (60), *Streptomyces* (286), *Yinghuangia* (6), *Allostreptomyces* (2)
Thermomonosporaceae (139)	4	*Actinoallomurus* (2), *Actinocorallia* (5), *Actinomadura* (122), *Thermomonospora* (10)
Dermacoccaceae (2)	2	*Allobranchiibius* (1), *Flexivirga* (1)
Geodermatophilaceae (2)	2	*Modestobacter* (1), *Petropleomorpha* (1)
Actinopolymorphaceae (2)	1	*Actinopolymorpha* (2)
Frankiaceae (3)	1	*Frankia* (3)
Gaiellaceae (1)	1	*Gaiella* (1)
Glycomycetaceae (1)	1	*Stackebrandtia* (1)
Kineosporiaceae (1)	1	*Kineosporia* (1)
Kribbellaceae (10)	1	*Kribbella* (10)
Nocardiaceae (5)	1	*Nocardia* (5)
Nocardioidaceae (1)	1	*Thermasporomyces* (1)
Thermoactinomycetaceae (2)	1	*Shimazuella* (2)

number of entries/total), times "LN(Pi)", which is the natural log for "Pi".

Table 3: The most diverse families according to the Shannon Diversity Index

Family	Value for H'
Pseudonocardiaceae	1,63
Nocardiopsaceae	1,09
Streptosporangiaceae	1,06
Micromonosporaceae	0,87
Dermacoccaceae	0,69
Geodermatophilaceae	0,69
Streptomycetaceae	0,57
Thermomonosporaceae	0,48
Frankiaceae	0
Gaiellaceae	0
Glycomycetaceae	0
Kineosporiaceae	0
Kribbellaceae	0
Nocardiaceae	0
Nocardioidaceae	0
Thermoactinomycetaceae	0

It is interesting that, when considering the genera's evenness, there is a change in the list of the most diverse genera. For example, even though *Nocardiopsaceae* and *Streptosporangiaceae* have less genera than *Micromonosporaceae*, they are more diverse according to this metric.

Based on table 2, the authors made statistical analysis in order to compare the productiveness of each family, as well as how homogeneous each family is, or in other words, if the genera inside a family produce a similar or very different number of terpenes. To do so, it was calculated the mean to evaluate how productive a family is (total number of entries in the family / number of genera in the family), the standard deviation[33] to analyze how disperse the number of terpenes of each genus is, and the coefficient of variation (the number for the standard deviation / mean) as a way to unite both informations. In this sense, a high value for the mean indicates that family is very productive, a small value for the standard deviation means that the family's genera produce a similar number of terpenes and a small value for the coefficient of variation suggests that the genera of the family are very productive, but also produce a similar number of terpenes. The results of this analysis are shown in table 4.

The results from this table show that if a researcher is looking for a productive family of Actinobacteria, then they should study *Streptomycetaceae*, as it has the highest mean. On the other hand, if a researcher is looking for a more cohesive group, then they should select *Nocardiopsaceae*, as it has the lowest value for the standard deviation. Finally, if a researcher is looking for a family that unites these two aspects, then they should explore *Streptosporangiaceae*, because it has the lowest value for the coefficient of variation, which indicates that this family is both productive and cohesive.

Table 4: Mean, Standard Deviation and Coefficient of Variation for each family

Family (number of entries in the family)	Mean	Standard Deviation	Coefficient of Variation
Pseudonocardiaceae (127)	14,11	15,97	1,13
Micromonosporaceae (92)	13,14	24,45	1,86
Nocardiopsaceae (17)	3,40	3,88	1,14
Streptosporangiaceae (136)	27,2	29,33	1,08
Streptomycetaceae (354)	88,5	116,30	1,31
Thermomonosporaceae (139)	34,75	50,45	1,45
Dermacoccaceae (2)	1	0	N/a
Geodermatophilaceae (2)	1	0	N/a
Actinopolymorphaceae (2)	2	N/a	N/a
Frankiaceae (3)	3	N/a	N/a
Gaiellaceae (1)	1	N/a	N/a
Glycomycetaceae (1)	1	N/a	N/a
Kineosporiaceae (1)	1	N/a	N/a
Kribbellaceae (10)	10	N/a	N/a
Nocardiaceae (5)	5	N/a	N/a
Nocardioidaceae (1)	1	N/a	N/a
Thermoactinomycetaceae (2)	2	N/a	N/a

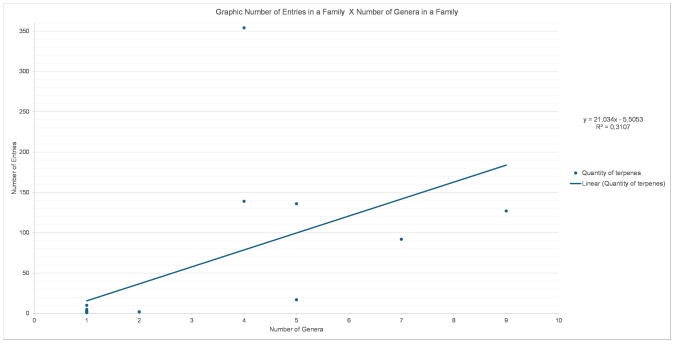

Fig. 3 Graphic Number of Entries in a Family X Number of Genera in a Family. The graphic shows a positive relation between the number of genera in a family and the quantity of terpenes in this family, which means that as the number of genera in a family increases, the quantity of terpenes increases as well.

In order to trace hypotheses about the correlation of the number of genera in a family and the quantity of terpenes being produced by this group, a linear regression graph was done as shown in figure 3.

As expected, the graphic shows that families that comprise more genera tend to be the ones with greater number of entries. In this sense, if a researcher is looking for species that produce a wide variety of terpenes, they should probably look for species whose family comprises a great variety of genera, and avoid the families with few genera.

Based on the data of the graphic, it was done a Pearson's correlation in order to evaluate how strong the number of genes and the quantity of terpenes being produced in a family is correlated. The result for this correlation was of 0,56, which

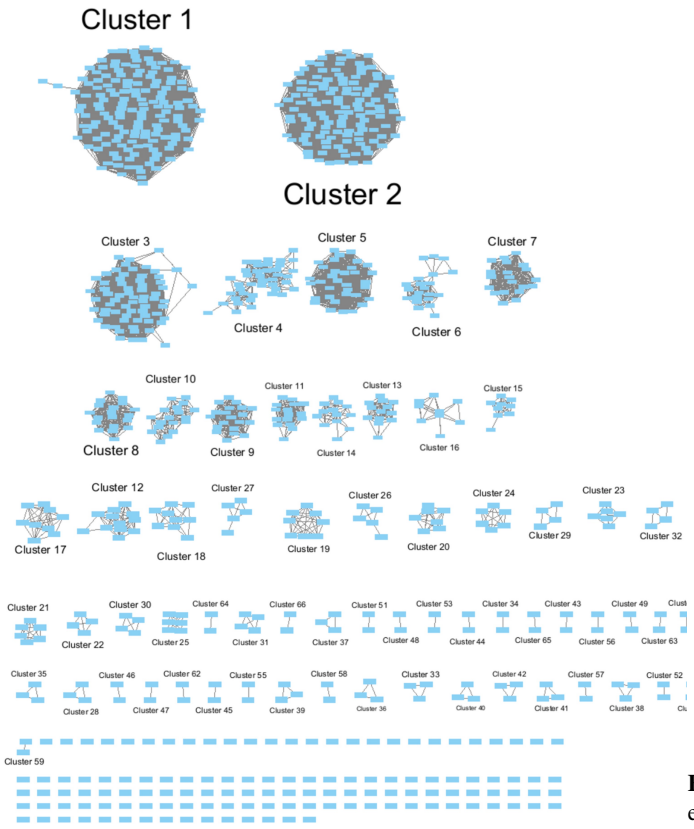

Fig. 4 Picture of the 924 entries of terpene synthases organized in groups based on similarity with each group synthesizing a different protein from each other. It is interesting to note the size of each group, indicating that there are some terpenes that are shared by a lot of organisms, and other terpenoids that are produced by just a few organisms.

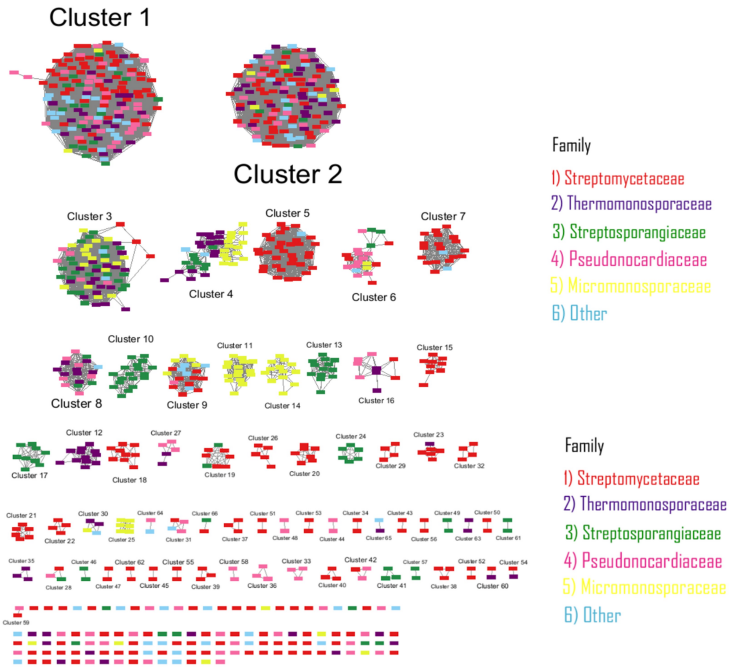

Fig. 5 The entries related to the five families with the most number of entries were colored. The result shows the distribution of the families over the groups, demonstrating clusters that are shared by a wide range of families (e.g. 1, 2 and 3), and clusters that are unique to one family (e.g. 11, 14 and 13).

Figure 5 shows that the most numbered families are present in clusters 1, 2 and 3, while some clusters are unique to certain families. This relation can be better seen in table 5.

indicates a moderate positive correlation between these two variables[34].

Additionally, through Cytoscape, the authors were able to separate these proteins into groups based on their similarity. Consequently, it was observed that there are 66 groups with different clusters (groups of genes that encode similar proteins), each expressing at least two enzymes related to terpene production (Figure 4).

To evaluate the diversity of organisms in each group, the authors filled the nodes based on the 5 most numbered families (*Streptomycetaceae*, Thermomonosoporaceae, *Streptosporangiaceae*, *Pseudonocardiaceae*, and Micromonoporaceae) (Figure 5).

The table demonstrates that the *Streptomycetaceae* family produces a more diverse range of terpenes, while the other families tend to produce more exclusive ones.

The previous steps were done again, but this time to determine

Table 5: Overview of the distribution of families over the clusters

Family	Number of clusters they are present in	Percentage (out of 66)	What clusters they are present it
Streptomycetaceae	39	59%	1, 2, 3, 5, 6, 7, 9, 15, 16, 18, 19, 20, 29, 23, 32, 21, 22, 31, 66, 37, 51, 53, 34, 43, 56, 50, 28, 47, 62, 45, 55, 39, 40, 42, 38, 52, 60, 54, 59
Thermomonosporaceae	15	23%	1, 2, 3, 4, 8, 16, 12, 27, 23, 30, 65, 63, 35, 60, 54
Streptosporangiaceae	17	26%	1, 2, 3, 4, 6, 8, 10, 13, 17, 19, 24, 49, 61, 28, 46, 41, 57
Pseudonocardiaceae	18	27%	1, 2, 3, 6, 8, 9, 16, 27, 64, 31, 48, 44, 28, 58, 36, 33, 42, 59
Micromonosporaceae	10	15%	1, 2, 3, 4, 6, 9, 11, 14, 30, 25

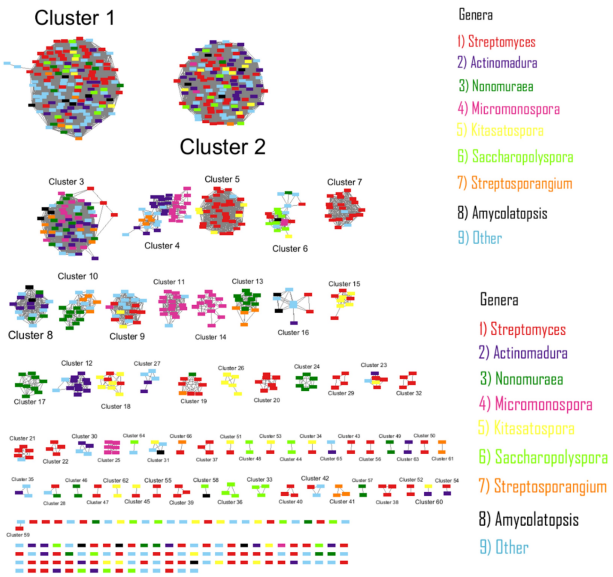

Fig. 6 The entries related to the eight genera with the most number of entries were colored. It is interesting to note that some of the clusters that were unique to a family have a majority of a single genus, which can suggest that the protein being produced relates to the survival of this specific genus.

Table 6: Overview about the distribution of genera over the clusters

Genus	Number of clusters they are present in	Percentage (out of 66)	What clusters they are present in
Streptomyces	37	56%	1, 2, 3, 5, 6, 7, 9, 16, 15, 18, 19, 20, 29, 23, 32, 21, 22, 31, 66, 37, 51, 53, 34, 43, 56, 50, 28, 47, 45, 55, 39, 40, 42, 38, 52, 54, 59
Actinomadura	15	23%	1, 2, 3, 4, 8, 16, 12, 27, 23, 30, 65, 63, 35, 60, 54
Nonomuraea	13	20%	1, 2, 3, 6, 8, 10, 13, 17, 24, 49, 28, 46, 57
Micromonospora	8	12%	2, 3, 4, 6, 9, 11, 14, 25
Kitasatospora	16	24%	1, 2, 5, 6, 9, 15, 18, 26, 23, 31, 51, 53, 34, 62, 45, 60
Saccharopolyspora	9	14%	1, 2, 6, 64, 48, 44, 58, 36, 33
Streptosporangium	10	15%	1, 2, 3, 4, 10, 13, 19, 66, 61, 41
Amycolatopsis	8	12%	1, 2, 3, 6, 8, 16, 31, 58

the variety of genera involved in the production of terpenes in order to understand what clusters are unique to a specific genus, and what clusters are shared among many genera. From this analysis, a possible hypothesis is that the terpenes that are produced by many organisms are essential to the phylum – and that is why this gene was selected in a variety of organisms – and that the compounds produced by only one genus relate to the survival of that specific bacterium in its environment. The results are shown in figure 6.

The image reveals - as expected - that there is a great number of genera related to the production of terpenes. Once again, to determine the degree of this diversity, the authors analyzed what clusters each genus is part of. (table 6)

After that, the data was filtered again to find in what clusters a genus is dominant (the number of entries related to the genus > 50% of the total number of entries)

The researchers decided to filter once again the data by detailing what clusters are produced exclusively by each of the most numbered genera (e.g number of entries of the genera = 100% of the total number of entries of the cluster).

These tables reinforce once more that *Streptomyces* are the most diverse group of bacteria related to the production of terpenes, while other genera tend to produce more restricted compounds. However, they also reveal the potential of less studied genera, since many of them produce unique terpenes that may be related to the survival of this single genera in their habitat. This relation can be better seen in table 9, 10, 11, 12.

Those tables show that, even though those genera present less unique clusters and are less studied, they still produce terpenoids with biological activity against other microorganisms, which implies that these compounds helped these genera live in a competitive environment. Therefore, these genera should be the focus of more studies in order to elucidate more compounds being produced by these unique clusters.

After coloring the clusters based on family criteria, it was possible to delve deeper into some specific clusters. The first analyzed cluster was the number 1, because it has a great number of entries and it is very diverse, which means the protein of this group is being produced by many families. On the other hand, cluster number 7 was also studied, since it has fewer entries, and mainly one family produces it.

Through the Protein Blast Tool and NCBI database, cluster 1 was found to be related to the production of geosmin, a protein that in bacteria may have the function of deterring predators and attracting organisms that disperse their spores [36]. Cluster number 7, on the other hand, may be related to the production of Epi-Isozizane, which catalyzes the formation of the precursor of the Albaflavenone antibiotic [37], and it is an enzyme related to many other compounds with antibacterial, antifungal, antioxidant, and antiproliferative properties [22]. Nevertheless, some clusters that share the same characteristics as those presented previously do not have their compounds' function described in the literature, such as cluster 11.

Table 7: Analysis of the distribution of dominant genera over the clusters

Genus	Number of clusters they are dominant	Percentage (out of the total number of clusters they are present in)	What clusters they are dominant in
Streptomyces	18	49%	5, 7, 19, 20, 29, 32, 21, 22, 37, 43, 56, 50, 47, 55, 39, 40, 38, 52
Actinomadura	2	13%	12, 63
Nonomuraea	6	43%	13, 17, 24, 49, 46, 57
Micromonospora	3	38%	11, 14, 25
Kitasatospora	4	25%	15, 18, 26, 62
Saccharopolyspora	4	44%	48, 44, 36, 33
Streptosporangium	2	20%	61, 41
Amycolatopsis	0	none	none

Table 8: Analysis of exclusive clusters over genera

Genus	Number of clusters that are unique for the genus	Percentage (out of the total number of clusters they are dominant in)	What clusters are unique for the genus
Streptomyces	14	78%	20, 29, 32, 22, 37, 43, 56, 50, 47, 55, 39, 40, 38, 52
Actinomadura	1	50%	63
Nonomuraea	5	83%	17, 24, 49, 46, 57
Micromonospora	2	67%	14, 25
Kitasatospora	2	50%	26, 62
Saccharopolyspora	4	100%	48, 44, 36, 33
Streptosporangium	2	100%	61, 41
Amycolatopsis	0	none	none

Table 9: Few examples of terpenoids isolated and described from the genus

Actinomadura

Compound	Biological activity	Resource
Actinomadurol	Antibacterial against *Staphylococcus aureus, Kocuria rhizophila*, and *Proteus hauseri*	22
k4610422	cytotoxic	22

Table 10: Example of terpenoid isolated and described from the genus

Micromonospora

Compound	Biological activity	Resource
Micromonohalimane B.	Moderate antibacterial activity against methicillin-resistant *Staphylococcus aureus*	35

Table 11: Example of terpenoid isolated and described from the genus

Kitasatospora

Compound	Biological activity	Resource
Terpentecin	Antibacterial against *Staphylococcus aureus, Bacillus subtilis, Corynebacterium bovs, Shigella dysenteriae, Aeromonas salmonicida, Vibrio anguilarum*	22

Table 12: Example of terpenoid isolated and described from the genus

Saccharopolyspora

Biological activity of terpenoids isolated from the genus *Saccharopolyspora*		
Compound	Biological activity	Resource
2, 7, 18 - Dolabellatriene	Antimicrobial against methicillin-resistant *Staphylococcus aureus*	22

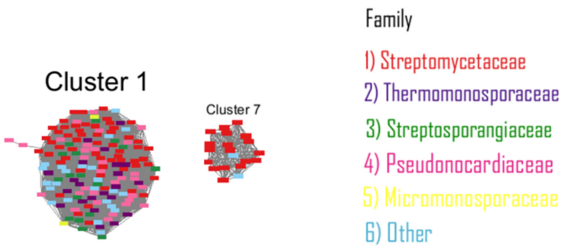

Fig. 7 Comparison of the clusters 1 and 7. Note the difference of diversity and size of these two clusters, which indicates that cluster 1 is widely shared, while cluster 7 is restrictively shared by *Streptomycetaceae*. A possible explanation for these differences is the importance these terpenoids have to the survival of the genera that produce them.

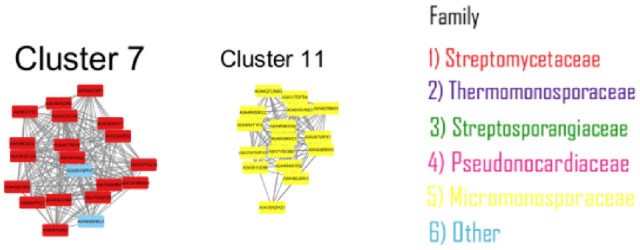

Fig. 8 Clusters 7 and 11 next to each other to compare the similarities between them in terms of number of entries and family diversity. Note that, just like cluster 7, cluster 11 is unique to one family, which can suggest that the protein can also be crucial to the survival of this one family.

Conclusion

This study shows the diversity of genera and families related to the production of terpenes in Actinobacteria. *Streptomyces* are the most studied organisms and tend to have more entries in Uniprot's database than other bacteria. Therefore, scientists should keep studying and exploiting the full potential of this renowned genus (by exposing the bacterium to stresses, which can lead to the activation of unknown genes) in order to discover more compounds. However, it does not mean that scientists should not work with other bacteria, because this data also reveals the potential behind less studied bacteria, such as *Actinomadura*, *Micromonospora*, *Saccharopolyspora*, *Streptosporangium*, and *Nonomuraea* that have unique clusters, and are known for having compounds with therapeutic applications. As previously stated, maybe a bacterium is not producing a certain terpene, because it is not being exposed to the natural environmental conditions to do so (such as, osmotic stresses, temperature changes, UV radiation, and chemical signaling). Hence, scientists and pharmaceutical companies can use this article to start exploring new genera and families and varying the stresses related to the production of terpenes, so that they can discover novel terpenoids with pharmacological properties.

This work also presents the method of screening databases as a relevant tool to identify the gaps in the literature, as well as identifying the organisms that have more potential to produce important compounds for medicine. This technique could be used as the first step of an investigation. In order words, before spending time and energy inoculating and exploring bacteria, researchers could identify with these tools what organisms they should work with, what procedures they should do, and what results they should seek. As a result, the process of discovering new compounds could be more efficient.

Another possible assumption that can be explored in the future is that the clusters that are shared by many organisms (e.g. 1 and 3) may be related to the production of essential terpenes to all of these bacteria, while less shared compounds (e.g. clusters 7 and 11) may be unique to the survival of such organisms in their habitat. These analyses can make it easier for specialists to discover future compounds, because it expands humans' knowledge about the evolutionary mechanisms behind the production of compounds in bacteria. In other words, understanding what compounds are produced only by a single organism and what molecules are produced by many individuals enables medical professionals to explore the reasons why nature selected this specific gene (maybe it helps to adapt to the environment or it avoids competition). However, to support this assumption, a new variety of experiments should be done: specialists should first identify and confirm what are the compounds being produced in clusters 1, 3, 7, and 11 and their function; secondly, through Mega software, the

researchers will be able to determine the evolutionary relations between the compounds in a single cluster, which can give a broader idea about the relevancy of that protein to that group of microorganisms that it is being studied. Moreover, researchers could explore other methods for clustering these data, as a way to test the presented groups, verify the hypothesis proposed by the authors, and discover novel compounds. With these data, they will be able to confirm the correlation previously proposed.

A major limitation of this work, nonetheless, is the lack of important data related to terpenes produced by Actinobacteria, such as the description of some compounds produced in the clusters and their function. This gap limits the discovery of new compounds, because scientists can not interpret the patterns of molecules synthetized by each cluster. As a result, it is not possible to revise and expand the current knowledge of the production of terpenes. Moreover, without a vast database, it is impossible to make inferences and comparisons with clusters from other organisms in order to elucidate the triggers behind the production of a specific terpene. Besides that, the most obvious limitation of the lack of description of compounds is that these obscure molecules may be relevant to medicine, but it is not possible to determine that, because there is not enough information supporting this hypothesis. Therefore, it is crucial to purify these unknown compounds, in order to determine what protein is being produced, and its function, because these molecules can be helpful to humankind.

Expanding the current data about other genera of bacteria may lead to the discovery and production of novel compounds, and to a better understanding on the genetic mechanisms behind the synthesis of such proteins.

Acknowledgments

This paper was done with the helpful mentorship of Dr. Lydia Morley, for whom the author is deeply thankful for her precious help, advices, and words of encouragement.

The author also wants to express his most sincere gratitude to University of São Paulo's professor Dr. Gabriel Padilla for giving the opportunity of delving deeper into terpenes in his lab, and to his incredible students for giving all support for this research to be conducted, and for always encouraging teenagers into science.

References

1 C. A. Borda-Roque, P. B. D. Silva and M. C. R. et al., *Challenge in the Discovery of New Drugs: Antimicrobial Peptides against WHO-List of Critical and High-Priority Bacteria*, 2024, https://doi.org/10.3390/pharmaceutics13060773, Accessed: August 13, 2024.

2 H. J. Lacey and J. P. Rutledge, *Recently Discovered Secondary Metabolites from Streptomyces Species*, 2024, https://doi.org/10.3390/molecules27030887, Accessed: August 13, 2024.

3 A. Crits-Christoph, S. Diamond and C. N. B. et al., *Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis*, 2024, https://www.nature.com/articles/s41586-018-0207-y, Accessed: August 13, 2024.

4 R. S. Peixoto, C. R. Voolstra and M. S. et al., *Harnessing the Microbiome to Prevent Global Biodiversity Loss*, 2024, https://www.nature.com/articles/s41564-022-01173-1, Accessed: August 13, 2024.

5 M. B. Miller and B. L. Bassler, *Quorum sensing in bacteria*, 2025, https://doi.org/10.1146/annurev.micro.55.1.165, Accessed: January 3, 2025.

6 A. Y. Alwali and E. I. Parkinson, *Small molecule inducers of actinobacteria natural product biosynthesis*, 2025, https://doi.org/10.1093/jimb/kuad019, Accessed: January 4, 2025.

7 B. Dror, E. Jurkevitch and E. Cytryn, *State-of-the-art methodologies to identify antimicrobial secondary metabolites in soil bacterial communities - A review*, 2024, https://www.sciencedirect.com/science/article/pii/S0038071720301358, Accessed: August 13, 2024.

8 A. D. Riggs, *Making, Cloning, and the Expression of Human Insulin Genes in Bacteria: The Path to Humulin*, 2024, https://doi.org/10.1210/endrev/bnaa029, Accessed: August 13, 2024.

9 A. Sayed, M. Hassan, H. Alhadrami *et al.*, *Extreme environments: microbiology leading to specialized metabolites*, 2025, https://pubmed.ncbi.nlm.nih.gov/31310419/, Accessed on: January 3rd, 2025.

10 F. Xie and W. Pathom-Aree, *Actinobacteria From Desert: Diversity and Biotechnological Applications*, 2025, https://doi.org/10.3389/fmicb.2021.765531, Accessed on: January 3rd, 2025.

11 A. A. Paredes, G. Valdés and N. A. et al., *Microbial Community in the Composting Process and Its Positive Impact on the Soil Biota in Sustainable Agriculture*, 2024, https://doi.org/10.3390/agronomy13020542, Accessed: July 14, 2024.

12 M. L.-Y. Hui, L. T.-H. Tan and V. L. et al., *The Extremophilic Actinobacteria: From Microbes to Medicine*, 2024, https://doi.org/10.3390/antibiotics10060682, Accessed: July 14, 2024.

13 N. Rao and W. J. Li, *Actinobacteria*, Springer, Singapore, 2024.

14 Z.-P. Zhong, F. Tian, S. Roux *et al.*, *Glacier ice archives nearly 15,000-year-old microbes and phages*, 2025, https://doi.org/10.1186/s40168-021-01106-w, Accessed on: January 4th, 2025.

15 A. T. Bull, *Extremophiles Handbook*, Springer, 2011, pp. 1203–1240.

16 S. A. Assou, J. Anissi and K. S. et al., *Diversity and Antimicrobial Activities of Actinobacteria Isolated from Mining Soils in Midelt Region, Morocco*, 2024, https://onlinelibrary.wiley.com/doi/full/10.1155/2023/6106673, Accessed: June 27, 2024.

17 J. W.-F. Law, V. Letchumanan and L. T.-H. T. et al., *The Rising of "Modern Actinobacteria"Era*, 2024, https://journals.hh-publisher.com/index.php/pmmb/article/view/121/150, Accessed: August 4, 2024.

18 R. Salwan and V. Sharma, *Molecular and biotechnological aspects of secondary metabolites in actinobacteria*, 2024, https://doi.org/10.1016/j.micres.2019.126374, Accessed on: July 14th, 2024.

19 E. J. N. Helfrich, G.-M. Lin and C. A. V. et al., *Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering*, 2024, https://doi.org/10.3762/bjoc.15.283, Accessed: June 27, 2024.

20 F. D. Mabou and I. B. N. Yossa, *TERPENES: Structural Classification and Biological Activities*, 2024, https://www.iosrjournals.org/iosr-jpbs/papers/Vol16-issue3/Series-1/D1603012540.pdf, Accessed: June 27, 2024.

21 J. D. Rudolf, T. A. Alsup, B. Xu *et al.*, *Bacterial terpenome*, 2024, https://pubs.rsc.org/en/content/articlelanding/2020/np/d0np00066c/unauth, Accessed on: June 27th, 2024.

22 E. V. Tarasova, N. A. Luchnikova, V. V. Grishko *et al.*, *Actinomycetes as Producers of Biologically Active Terpenoids: Current Trends and Patents*, 2025, https://doi.org/10.3390/ph16060872, Accessed on: January 4th, 2025.

23 D. A. V. Bergeijk, B. R. Terlouw and M. H. M. et al., *Ecology and genomics of Actinobacteria: new concepts for natural product discovery*, 2024, https://www.researchgate.net/publication/341802502_Ecology_and_genomics_of_Actinobacteria_new_concepts_for_natural_product_discovery, Accessed: August 4, 2024.

24 W. R. Pearson, *An Introduction to Sequence Similarity ("Homology") Searching*, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC3820096/, Accessed: January 2, 2025.

25 E. A. Barka, P. Vatsa and L. S. et al., *Taxonomy, Physiology, and Natural Products of Actinobacteria*, 2025, https://doi.org/10.1128/m, Accessed: January 7, 2025.

26 Y. Dashti and J. Errington, *Chemistry and biology of specialized metabolites produced by Actinomadura*, 2025, https://doi.org/10.1039/D3NP00047H, Accessed: January 7, 2025.

27 Y. Yu, G. Yang, Y. Wang *et al.*, *A Rare Strain Actinomadura geliboluensis Was First Isolated from the Bronchoalveolar Fluid of a Patient with Pneumonia*, 2025, https://doi.org/10.2147/IDR.S409701, Accessed on: January 7th, 2025.

28 A. Bonifaz, A. Tirado-Sánchez and D. V.-G. et al., *Actinomycetoma by Actinomadura madurae: Clinical Characteristics and Treatment of 47 Cases*, 2025, 10.4103/idoj.IDOJ_474_20, Accessed: January 7, 2025.

29 O. Yushchuk, A. Andreo-Vidal, G. Marcone *et al.*, *New Molecular Tools for Regulation and Improvement of A40926 Glycopeptide Antibiotic Production in Nonomuraea gerenzanensis ATCC 39727*, 2025, https://doi.org/10.3389/fmicb.2020.00008, Accessed on: January 7th, 2025.

30 A. Talà, G. Wang, M. Zemanova *et al.*, *Activation of Dormant Bacterial Genes by Nonomuraea sp. Strain ATCC 39727 Mutant-Type RNA Polymerase*, 2025, https://doi.org/10.1128/jb.01311-08, Accessed on: January 7th, 2025.

31 L. Carro, I. Nouioui and V. S. et al., *Genome-based classification of micromonosporae with a focus on their biotechnological and ecological potential*, 2025, https://doi.org/10.1038/s41598-017-17392-0, Accessed: January 7, 2025.

32 Y. Wei, G. Wang, Y. Li *et al.*, *The Secondary Metabolites from Genus Kitasatospora: A Promising Source for Drug Discovery*, 2025, https://doi.org/10.1002/cbdv.202401473, Accessed on: January 7th, 2025.

33 H. Hernandez, *Probability Distribution and Bias of the Sample Standard Deviation*, 2025, 10.13140/RG.2.2.22144.51205, Accessed: January 10, 2025.

34 P. Schober, *Correlation Coefficients: Appropriate Use and Interpretation*, 2025, https://doi.org/10.1213/ANE.0000000000002864, Accessed on: January 10th, 2025.

35 M. Gozari, M. Alborz and H. R. E.-S. et al., *Chemistry, biosynthesis and biological activity of terpenoids and meroterpenoids in bacteria and fungi isolated from different marine habitats*, 2025, https://doi.org/10.1016/j.ejmech.2020.112957, Accessed: January 4, 2025.

36 L. Zaroubi, I. Ozugergin, K. Mastronardi *et al.*, *The ubiquitous terpene geosmin is a warning chemical*, 2025, https://doi.org/10.1101/2021.03.09.434661, Accessed on: January 10th, 2025.

37 T. A. Ronnebaum, S. M. Gardner and D. W. Christianson, *An Aromatic Cluster in the Active Site of epi-Isozizaene Synthase is an Electrostatic Toggle for Divergent Terpene Cyclization Pathways*, 2025, https://doi.org/10.1021/acs.biochem.0c00876, Accessed on: January 10th, 2025.