



Universiteit  
Leiden  
The Netherlands

## Ecology and genomics of Actinobacteria and their specialised metabolism

Bergeijk, D.A. van

### Citation


Bergeijk, D. A. van. (2022, October 19). *Ecology and genomics of Actinobacteria and their specialised metabolism*. Retrieved from <https://hdl.handle.net/1887/3484350>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3484350>

**Note:** To cite this publication please use the final published version (if applicable).



# Taxonomic and metabolic diversity of Actinobacteria isolated from faeces of a 28,000-year-old mammoth

Doris A. van Bergeijk\*, Hannah E. Augustijn\*, Somayah S. Elsayed, Joost Willemse, Victor J. Carrión, Mia Urem, Lena V. Grigoreva, Maxim Y. Cheprasov, Bas Wintermans, Andries E. Budding, Herman P. Spaijk, Marnix H. Medema, Gilles P. van Wezel

\* Contributed equally.

## Abstract

The yet unexplored microbial world is a vast resource that can provide us with new evolutionary insights, metabolic pathways, and chemistry. Ancient microbial communities of permafrost soils and frozen animal remains represent an archive that has barely been explored. Here, we show that Actinobacteria revived from a faecal sample from the intestinal tract of a 28,000-year-old Siberian mammoth are phylogenetically and metabolically distinct from their modern siblings. Ancient *Micromonospora*, *Oerskovia*, *Saccharopolyspora*, *Sanguibacter*, and *Streptomyces* species were recovered and subsequently grown up in the laboratory. Surprisingly, the genomes of these ancient ancestors showed a large phylogenetic distance to any known modern strains and harboured many novel biosynthetic gene clusters that may represent uncharacterised biosynthetic potential. The strains produced known molecules like antimycin, conglobatin and macrotetrolides, but the majority of the mass features could not be dereplicated. Thus, our work provides a snapshot into Actinobacteria isolated from a unique ancient animal faeces sample, yielding unexplored genomic information that is not yet present in current databases.

## Introduction

Bacteria are key to life on Earth; they are major players in nitrogen cycling <sup>196</sup>, decompose organic matter <sup>197</sup>, and provide eukaryotic hosts with essential nutrients and protection against pathogens <sup>198,199</sup>. Additionally, bacterial natural products are a major source of the bioactive molecules that find applications in medicine, biotechnology and agriculture <sup>9,18</sup>. It is predicted that we have only uncovered a small percentage of the microbial world; the majority of microorganisms resist cultivation and many bacterial taxa have hardly been explored <sup>200</sup>. One reason is that scientists have so far primarily accessed environments that can be easily reached, while numerous bacteria exist in hard-to-reach places such as deep-sea sediments, caves, and permafrost soils <sup>41</sup>. This uncharacterised microbiology represents an important reservoir of biological information that may be harnessed for drug discovery.

Microbial communities that have been underexplored are those in ancient samples, such as permafrost soils, frozen animal remains, and deep soil layers. These samples not only represent unique environments, but may also allow a glimpse into the past, i.e., may teach us important lessons on the evolution of microbial and chemical diversification. Isolation and metagenome sequencing of ancient bacteria has provided important knowledge on bacterial evolution, changes in microbiome composition, ancient diseases and potential chemical novelty <sup>201-204</sup>. For example, large-scale *de novo* assembly of microbial genomes from human palaeofaeces samples revealed previously undescribed gut microorganisms and a markedly higher abundance of mobile genetic elements in our ancestral gut microbiome compared to modern industrial gut microbiomes <sup>201</sup>. A metagenomics survey of ancient Alaskan soil confirmed that homologues of different resistance genes existed in ancient bacteria, showing that antibiotic resistance predates the modern selective pressure of clinical antibiotic use <sup>202</sup>. Furthermore, ancient Actinobacteria isolated from thousands-year-old Arctic and Antarctic sediments showed promising bioactivity against drug-resistant pathogens, and genome mining showed low similarity to known antibiotics <sup>203,204</sup>. These studies illustrate how ancient sources can lead us to evolutionary insights, as well as novel micro-organisms and chemistry.

The New Siberian Islands, located between the Laptev Sea and the East Siberian Sea, contain permafrost deposits that have been preserved for more than 200,000 years <sup>205</sup>. These islands are considered a time capsule to ancient biology and have been an important source of ancient animal remains, including mammoths <sup>206</sup>. In August 2012, an adult female woolly Mammoth (*Mammuthus primigenius*) was recovered on Maly Lyakhovsky Island (74°07' N, 140°40' E) <sup>207,208</sup>. The mammoth carcass was submerged in permafrost, exposing skull, post-cranial elements, and partial trunk. The lower part of the body was surrounded by almost pure



ice and included lower parts of the head, distal portion of the trunk, chest, abdomen, front legs and distal half of the right hind leg <sup>207-209</sup>. The mammoth was determined to be 28,610 years old and the remains contained exceptionally well-preserved soft tissues. The skin retained its elasticity and mummification of the carcass was minimal <sup>207-209</sup>. In February 2014 the specimen was transported to Yakutsk for investigation.

The finding of this extraordinarily well-conserved ancient specimen provided a unique opportunity to explore an ancient microbiome. Its surprisingly well-preserved soft tissues included the intestinal tract, allowing to extract a faecal sample for isolation of bacteria. We were specifically interested in members of the highly diverse phylum Actinobacteria, well known for their capability to produce an unprecedented diversity of specialised metabolites. Their genomes are full of biosynthetic gene clusters (BGCs) that encode the cellular machinery required for the biosynthesis of natural products. Although many bioactive compounds produced by well-known Actinobacteria have been identified, genomic research shows that a large part of the metabolic diversity of this phylum is still unexploited <sup>27</sup>. In this study we isolated Actinobacteria from this extraordinary ancient sample, compared the genomes to those of their closest modern-day neighbours and analysed their bioactive potential. Sequencing of six isolated Actinobacteria revealed significant phylogenetic distance to currently known strains, with yet uncharacterised biosynthetic potential.

## Results

### Isolation of Actinobacteria from mammoth faeces

The remarkable diversity of Actinobacteria and their specialised metabolites is the result of millions of years of evolution <sup>104,210</sup>. However, the evolutionary drivers that have shaped this metabolic diversity remain largely unknown. Isolation of Actinobacteria from ancient samples may allow a glimpse into the past say 10,000-100,000 of years, and thus provide insights on how natural product biosynthesis evolved over this period of time. Additionally, the microbial communities of ancient samples have rarely been explored, representing unique environments that can yield novel microorganisms and chemistry. Yet, opportunities to study ancient samples do not occur often as most archaeological findings are not in a state that does justice to the biological conditions during this era.

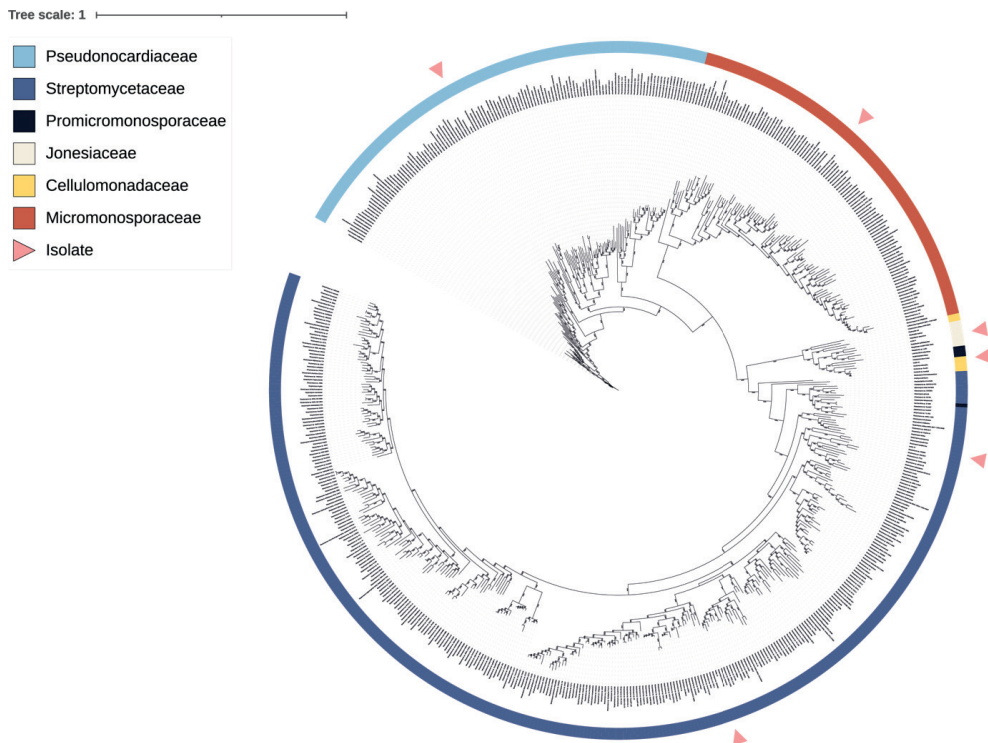
The discovery of an exceptionally well-preserved mammoth on Maly Lyakhovsky Island <sup>207,208</sup> (Fig. S1) provided a unique opportunity to recover ancient Actinobacteria and compare their genomes and biosynthetic potential to their modern descendants. From 10-14 March 2014, we joined an international team of researchers during the dissection of this extraordinary

specimen. During the dissection, every day a deeper layer of the mammoth tissue became accessible, as the specimen gradually thawed. When the thawing had proceeded far enough to explore the abdominal cavity, we found a large part of the intestines fully intact, with the omentum still attached. To extract faecal samples, the intestines were exposed and a 60 cm intestinal specimen was gathered from the remains (Fig. S1). The intestinal lumen was thoroughly inspected for defects and was found to be intact. Two perpendicular incisions were made and the intersection was folded over, exposing the intestinal lumen. Faecal samples were carefully taken from the lumen under sterile conditions, thereby avoiding any cross-contamination (Fig. S1).

A fraction of the collected faeces was homogenised in sterile dH<sub>2</sub>O and plated onto various media selective for Actinobacteria<sup>37</sup>. Plates were incubated aerobically at 4 °C and 30 °C, and anaerobically at room temperature. We were able to recover some Actinobacteria from the long-frozen sample. Strains were selected based on their filamentous morphology and grown on different media for phenotypic discrimination, resulting in the isolation of six morphologically distinct strains. The majority of the strains were isolated using selective humic acid agar and/or glucose agar plates incubated at 30 °C. No filamentous bacteria were observed after anaerobic incubation.

### Taxonomic and phenotypic profiling of the isolated Actinobacteria

To determine the taxonomic origin of the isolates and gain insights into the relatedness to current known bacterial species, we obtained the full genome sequences of the isolates using a combination of Nanopore and Illumina sequencing. Comparing the 16S rRNA sequence to those within EzBioCloud database revealed that the isolates belonged to five genera of Actinobacteria, namely *Sanguibacter* (M9), *Micromonospora* (M12), *Oerskovia* (M15), *Saccharopolyspora* (M46), and *Streptomyces* (M10 and M19). To obtain a more detailed taxonomic classification, a maximum-likelihood tree was constructed based on the genome sequences of the six isolates and that of 578 Actinobacteria representing six bacterial families and over 40 different genera (Fig. 1). The relatedness of the isolates and neighbouring organisms was determined by calculating the average nucleotide identity (ANI) score in an all-to-all genome comparison. Interestingly, the ANI score revealed a large phylogenetic distance between the mammoth isolates and currently known strains, with similarity scores ranging from 79-90% (Table S1). Of the isolates, the largest phylogenetic distance was found for *Streptomyces* sp. M19 and its closest neighbours with an ANI score <80%. The species delineation threshold typically lies at approximately 95% gene identity<sup>211</sup>, which suggests that the isolates may be novel species.



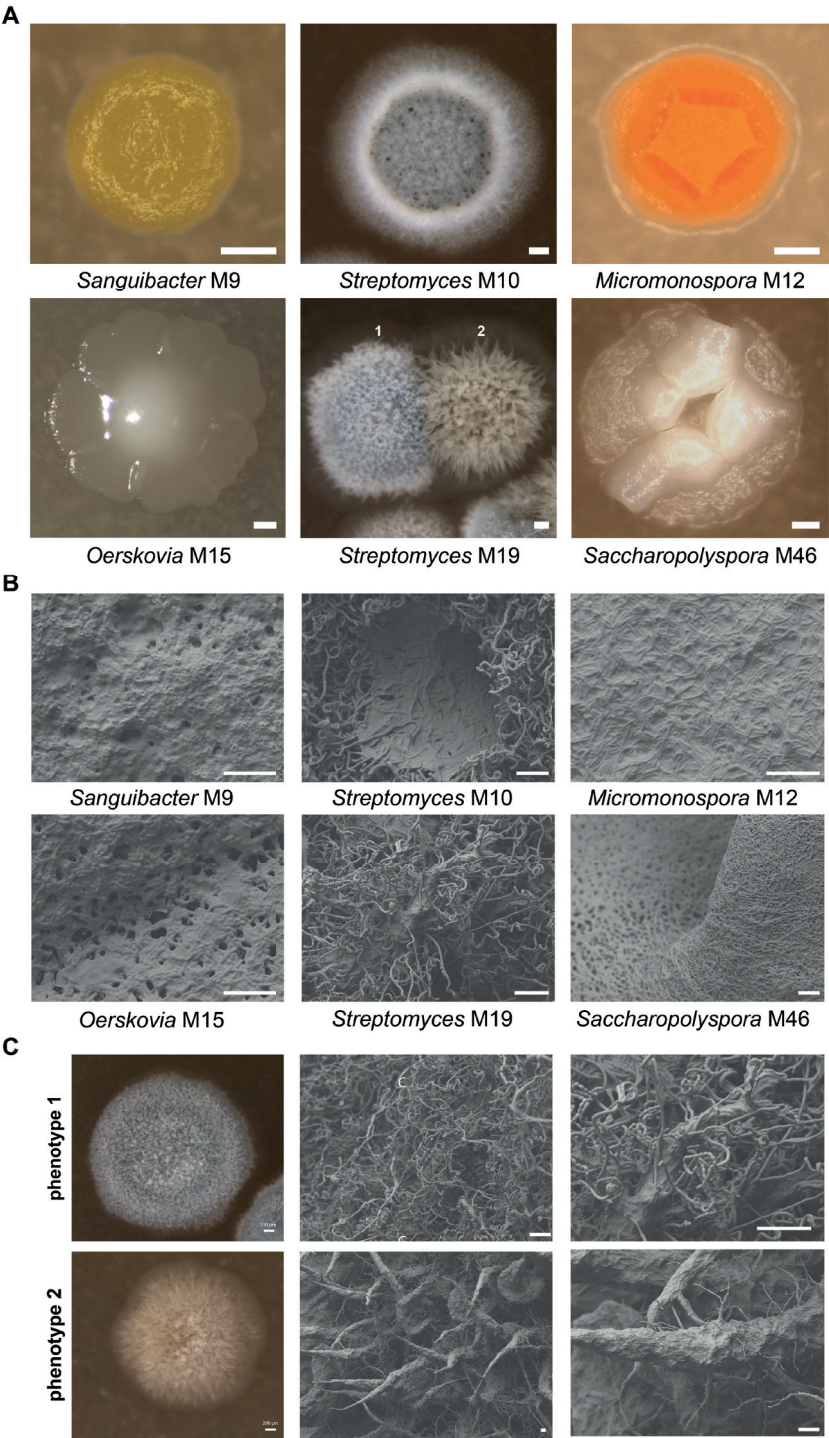
**Figure 1. Phylogeny of mammoth isolates and their closest known neighbours.** Maximum-likelihood tree of the six isolates compared to 578 Actinobacteria of six bacterial families (colour-coded in the outer ring), subdivided into 40 genera. The tree was rooted using *Amycolatopsis vastitatis* as outgroup and the numbers on the tree branches represent the bootstrap values in percentages of a total of 500 bootstraps. Triangles point at the location of the novel isolates. The phylogenetic analysis suggests the following likely classification based on their nearest neighbours: M9, *Sanguibacter*; M12, *Micromonospora*; M15, *Oerskovia*; M46, *Saccharopolyspora*; M10 & M19, *Streptomyces*.

The taxonomic analysis revealed a significant phylogenetic difference between the genomes of the isolates and those of their closest known neighbours. This was especially surprising for the *Streptomyces* isolates, which were compared to more than 300 genomes. We wondered whether the low similarity was a result of the ancientness of the strains or whether this was related to the underexplored environment the strains were sampled from. Therefore, we analysed the phylogenetic distance between other strains isolated from extreme environments and their closest neighbours. For this we selected the deep sea isolates *Streptomyces* sp. NTK 937<sup>212</sup> and *Streptomyces* sp. SCSIO 3032<sup>213</sup>, the desert isolate *Streptomyces jeddahensis*<sup>214</sup>, and *Streptomyces* sp. BF-3 and *Streptomyces* sp. 4F isolated from the Great Salt Plains in Oklahoma<sup>215</sup> (Table S2). In general, the results show higher relatedness to the closest neighbours of these strains (ANI scores > 93%, compared to 79-90% for our isolates), with

the exception of *Streptomyces* sp. SCSIO 3032 (ANI: 85.48%) and *Streptomyces jeddahensis* (ANI: 81.44%). It should be noted that the closest known neighbour of *Streptomyces jeddahensis* is an isolate from an insect, while this strain itself was isolated from a soil sample collected in the desert. Additionally, *Streptomyces* sp. SCSIO 3032 and its neighbour *Streptomyces* sp. MP131-18 both originate from different sampling environments although they both originate from the deep-sea. This suggests that the low level of similarity observed between the genomes of our strains and publicly available genomes may be due to the lack of bacteria isolated from similar environments, rather than the age of the strains themselves. However, we cannot ascertain this at this moment in time.

Next, we grew the isolates on SFM agar and assessed their morphology using stereomicroscopy, revealing a wide range of phenotypes (Fig. 2A). *Sanguibacter* sp. M9 produced bright yellow round colonies, *Streptomyces* sp. M10 produced cream-coloured substrate mycelia, a grey aerial spore mass, and a dark diffusible pigment, *Micromonospora* sp. M12 produced orange-coloured folded colonies, *Oerskovia* sp. M15 produced white colonies, and *Saccharopolyspora* sp. M46 produced cream-coloured folded colonies. *Streptomyces* sp. M19 displayed a heterogeneous phenotype. When this isolate was grown on SFM agar, two colony phenotypes were observed: a fully developed phenotype, and a variant with strong yellow pigmentation, sparse aerial mycelia, and lack of spores. Morphological heterogeneity was also observed within single colonies (Fig. S2), consistent with a high tendency to genetic heterogeneity<sup>216</sup>. Sequencing of the 16S rRNA strongly suggests that all morphological variants were indeed phenotypes of the same strain.

To obtain more insights into the morphology of the strains at high resolution, colonies were scrutinised using Scanning Electron Microscopy (SEM) (Fig. 2B). Colonies of *Streptomyces* sp. M10 produced hairy spores and the aerial mycelium consisted of both smooth and hairy hyphae, with dark-pigmented droplets on top of the colonies. *Saccharopolyspora* sp. M46 colonies consisted of a thick layer of interwoven mycelium made up of hyphae and extracellular matrix. SEM studies of the two distinct phenotypes of *Streptomyces* M19 revealed spiral spore chains in one variant (phenotype 1), while we failed to identify spores in the other (phenotype 2) (Fig. 2C). Instead, the non-sporulating colony produced large spikes made up of hyphae and extracellular matrix. These spikes could also be found in the fully developed colony, but in low abundance. Colonies of isolates *Sanguibacter* sp. M9, *Micromonospora* sp. M12, and *Oerskovia* sp. M15 were covered by an extracellular matrix and could therefore not be further characterised by SEM.





◀**Figure 2. Phenotypic characterisation of Actinobacteria isolated from a faecal sample of a 28,000-year-old mammoth.** Strains were grown on SFM agar plates for 9 days. **(A)** Stereomicrographs of isolates *Sanguibacter* sp. M9, *Streptomyces* sp. M10, *Micromonospora* sp. M12, *Oerskovia* sp. M15, *Streptomyces* sp. M19, and *Saccharopolyspora* sp. M46 (for phylogenetic analysis see Figure 1). *Streptomyces* sp. M19 showed two distinct phenotypes: fully developed colonies (1) and colonies with a bald appearance (2). Scale bar: 200  $\mu\text{m}$ . **(B)** Scanning electron micrographs of isolates M9, M10, M12, M15, M19, and M46. Scale bar: 10  $\mu\text{m}$ . **(C)** Scanning electron micrographs showing significant differences in morphology between the two colony phenotypes observed when *Streptomyces* sp. M19 is grown on SFM. Images of phenotype 1 show spirals of spore chains. Images of phenotype 2 show large spikes made up of hyphae and extracellular matrix. Also in the fully developed colony, such spikes can be found but in low frequency. 16S rRNA sequencing strongly suggests that these morphological variants are phenotypes of *Streptomyces* sp. M19. Scale bar: 10  $\mu\text{m}$ .

Overall, the taxonomic and phenotypic analyses show that we have isolated a diverse collection of Actinobacteria, of which *Streptomyces* sp. M19 stands out with a highly heterogeneous phenotype and low similarity to its closest neighbours. Moreover, the low ANI scores suggest that all isolates represent novel species yielding unexplored genomic information that is not yet present in current databases.

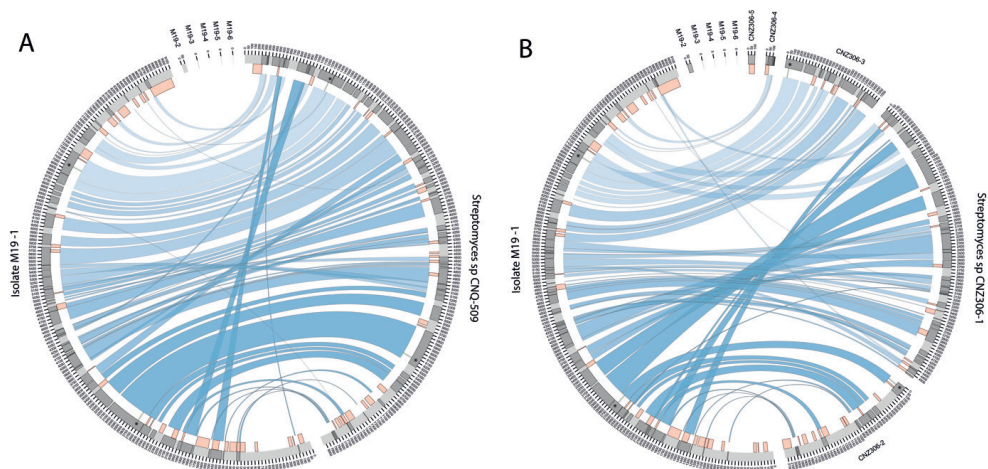
### Potential of the strains in terms of natural product biosynthesis

We then studied the biosynthetic gene clusters (BGCs) of the isolates, to obtain an idea of how related they were to BGCs that are currently available in the databases. For this, the genome sequences were analysed using antiSMASH<sup>217</sup>. This identified a total of 179 putative clusters, namely four in *Oerskovia* sp. M15, six in *Sanguibacter* sp. M9, 22 in *Saccharopolyspora* sp. M46, 43 in *Micromonospora* sp. M12, 50 in *Streptomyces* sp. M10 and 54 in *Streptomyces* sp. M19 (Fig. 3A). Over 70% of the total BGCs shared less than 50% KnownClusterBlast similarity to BGCs within the MIBiG database, with 23% showing no significant similarity to any known BGC. To determine how unique the high level of uncharacterised biosynthetic potential is to ancient actinobacterial isolates, we compared the potential of our isolates to the potential of Actinobacteria isolated from different environments using the antiSMASH database (Table S3). This showed that similarly diverse uncharacterised biosynthetic potential can be found in Actinobacteria isolated from other environments, such as soil. Indeed, even in the well-studied model organism *Streptomyces coelicolor*, several BGCs still show less than 50% KnownClusterBlast similarity to other known BGCs. Therefore, the uniqueness of the biosynthetic potential may not be related to the age of the strains, but may instead be widespread even in contemporary and well-studied *Streptomyces* species. BiG-SCAPE analysis of the genomes showed that the genomes contained relatively high numbers of genes encoding enzymes associated with the biosynthesis of terpenes, in particular for *Micromonospora* sp. M12 (five), *Streptomyces* sp. M19 (four), *Streptomyces* sp. M10 (seven), and *Saccharopolyspora* sp. M46 (six).

The genome of *Streptomyces* sp. M19 showed the largest distance in terms of ANI scores to its nearest neighbours. We therefore subjected the strain to detailed analysis of its biosynthetic potential, to obtain insights into the chemical diversity of its specialised metabolites. The M19 genome had a similar number of predicted BGCs (34) as its nearest neighbours *Streptomyces* sp. CNZ306 (33 BGCs) and *Streptomyces* sp. CNQ-509 (28 BGCs) (Fig. 3B). Surprisingly, *Streptomyces* sp. M19 only shares two BGCs with these nearest neighbours, namely for desferrioxamines and for ectoine. The M19 genome encodes a high percentage of non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs), and also has BGCs for butyrolactones, unusual polyketides / fatty acids (linked to a heterocyst glycolipid synthase-like PKS, and a ladderane-lipid-associated ketosynthase), a likely lasso peptide, an aryl polyene, and an aminoglycoside/aminocyclitol; none of these natural product classes could be found in modern nearest neighbours. Comparative genomic analysis of *Streptomyces* sp. M19 and its neighbours, CNZ306 and CNQ-509, displays the low similarity of approximately 79% and only few corresponding BGCs (Fig. 4). While a high degree of overlap was seen between the genomes of *Streptomyces* sp. CNZ306 and CNQ-509 in terms of overall genome synteny and in shared BGCs (Fig. S3), precious little homology was seen with the genome of *Streptomyces* sp. M19 (Fig. 4).



**Figure 3. Predicted BGC classes for the six actinobacterial isolates. A)** BGC classes for each isolate predicted with antiSMASH v6.0. Known types of BGCs representing <1% of all BGCs were grouped into the “other” category. The large number of BGCs in *Streptomyces* species M10 and M19, *Micromonospora* M12 and *Saccharopolyspora* M46, is particularly noteworthy. **B)** Overview of all predicted BGC classes for isolate M19 and its closest neighbours *Streptomyces* sp. CNZ306 and *Streptomyces* sp. CNQ-509. The comparison shows that only two BGCs are shared between all three isolates, namely BGCs for desferrioxamines and for ectoine. Moreover, the genome of *Streptomyces* sp. M19 contains a greater variety of BGC classes, with a surprisingly high proportion of NRPS BGCs.



**Figure 4. Comparative analysis between isolate M19 and its modern-day closest known neighbours *Streptomyces* sp. CNZ306 and *Streptomyces* sp. CNQ-509. A)** synteny plot comparing *Streptomyces* sp. M19 to *Streptomyces* sp. CNQ-509. The blue lines are obtained by a pairwise BLASTP and collinearity calculation method. Resulting locations represent high similarity regions. In red, the location of predicted BGCs of antiSMASH v6.0 are indicated. Shared BGCs are indicated in green and with an asterisk. **B)** synteny plot of *Streptomyces* sp. M19 and *Streptomyces* sp. CNZ306.

### Antibiotic activity and bioactive metabolites produced by the mammoth isolates

Next, the antibiotic-producing potential of the strains was assessed under six different culturing conditions. These were four agar-based media, namely the nutrient-rich Nutrient Agar (NA) and International *Streptomyces* Project 2 (ISP2) and the nutrient-poor Czapek Dox and Minimal Media supplemented with mannitol and glycerol (MM), and the liquid media versions of ISP2 and MM. We used the Gram-positive *Bacillus subtilis* 168 and the Gram-negative *Escherichia coli* ASD19 and *Pseudomonas aeruginosa* PA01 as indicator strains. Different types of growth inhibition were observed, namely complete inhibition, strong reduction in number of colony-forming units (cfu), and impaired growth (Fig. 5). Overall, the isolates displayed the strongest antibacterial activity when grown on NA. Therefore, further metabolic analysis was done on samples isolated from NA-grown cultures.

Besides soluble natural products, also volatile compounds (VCs) may have antibacterial activity<sup>218,219</sup>. To investigate whether (some of) the antimicrobial activity may have been due to the production of antibacterial VCs, the isolates were grown on plates where the Actinobacteria were separated from the indicator strains by an impermeable polystyrene divider. Growth of *E. coli* was completely inhibited by VCs produced by isolates *Streptomyces* sp. M10, *Streptomyces* sp. M19, and *Saccharopolyspora* sp. M46 (Fig. S4). While *Micromonospora* sp. M12 did not display antibacterial activity in the bioactivity assay described above, it partially inhibited growth



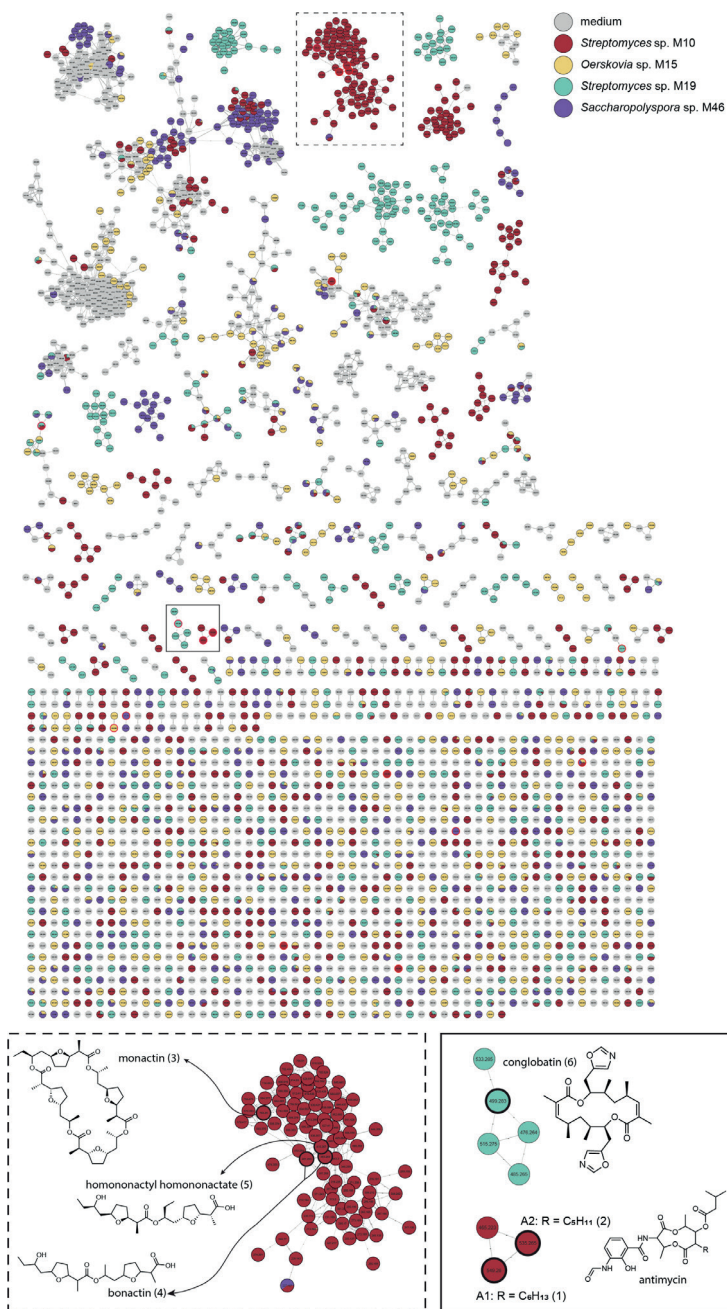
of *E. coli* in the volatile assay. None of the strains produced VCs with activity against *B. subtilis*. These data indicate that the observed antibacterial activity of strains M10, M19, and M46 against *E. coli* was at least in part caused by VCs, while the bioactivity of strains M10, M15, M19, M46 against *B. subtilis* was solely caused by the production of soluble antibiotics.

	<i>B. subtilis</i>						<i>E. coli</i>						<i>P. aeruginosa</i>					
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
M9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%
M10	0.7	1.1	0	0	0.6	1.7	0	0	0	0	0	2.2	0	0	0	0	0	IG
M12	0.6	0.6	0	0	0	0	0.6	0.5	0	0	0	0	0	0	0	0	0	0%
M15	0	0	0	0	0	0.6	0	0	0	0	0	0	0	0	0	0	0	0%
M19	1.9	0	1.8	1.7	0.6	1.1	0	0	0	0.9	0	3.2	0	0	0	0	0	0%
M46	0	0	0	0	0	3.6	0	0	0	0	0	3.3	0	0	0	0	0	IG

**Figure 5. Antimicrobial activity of the actinobacterial isolates.** After 7 days of growth, the bioactivity of the mammoth isolates was assessed against different indicator strains using different growth media and methods: 1 = MM, soft agar overlay; 2 = Czapek Dox, soft agar overlay; 3 = liquid culture ISP2; 4 = liquid culture MM; 5 = ISP2, cross streak, 6 = NA, cross streak. The zone of inhibition (cm) is indicated for each isolate and related to a colour scale ( $n = 3$ ). Most activity was observed in the cross-streak assay on NA. The fields selected with a black border refer to the examples displayed below the table that illustrate the different types of growth inhibition observed. IG: impaired growth.

To gain more insights into the soluble antibiotics produced by isolates *Oerskovia* sp. M15, *Streptomyces* sp. M10, *Streptomyces* sp. M19, and *Saccharopolyspora* sp. M46, the strains were streaked on NA plates and grown for seven days. Metabolites were extracted using ethyl acetate (EtOAc) and tested for bioactivity against the different indicator strains. The crude extracts of *Streptomyces* sp. M10 and M19 showed activity against *B. subtilis*, while *E. coli* and *P. aeruginosa* were not inhibited (Fig. S5). MS/MS data were analysed using Global Natural Products Social molecular networking (GNPS)<sup>220</sup>, resulting in a molecular network containing 2886 nodes clustered in 223 spectral families (Fig. 6). The highest number of unique nodes (491) was attributed to isolate M10, while the lowest number (280) was attributed to isolate M46. 44 nodes were unique to *Streptomyces* isolates M10 and M19. Dereplication based on matching MS/MS spectra against the GNPS spectral library annotated several  $m/z$  values as known natural products that were not present in the medium blank (Table S4), including  $m/z$  value 546.4879  $[M + H]^+$  as the hopanoid aminobacteriohopanetriol. Additionally, several  $m/z$  values were annotated as being known bioactive metabolites: antimycin A1 (1), antimycin A2 (2), and the macrotetrolide monactin (3) and



**Figure 6. Molecular network of the ions detected in the crude extracts of the bioactive isolates revealing the presence of multiple known bioactive compounds.** A pie chart was mapped to the nodes which represents the abundance of each  $m/z$  value in the crude extracts of the different bioactive isolates (M10, M15, M19 and M46) and the medium blank. Nodes highlighted represent dereplicated known bioactive metabolites. The isolates were grown for seven days on NA plates ( $n = 3$ ).

related bonactin (4) and homononactyl homononactate (5) in the extracts of M10, and conglobatin (6) in the extracts of M19 (Fig. 6). To confirm these findings, we analysed the KnownClusterBlast output of antiSMASH for presence of the responsible BGCs. In the genome of M10, antiSMASH identified an antimycin BGC and a BGC with 75% similarity to a macrotetrolide BGC (MIBiG cluster BGC0000243). The macrotetrolide-associated BGC from strain M10 lacks three genes compared to the reference BGC, two encoding hypothetical proteins and one encoding an inositol monophosphatase-like enzyme. M19 harbours a BGC with low (36%) similarity to the conglobatin BGC (Fig. S6). Although the overall match for this cluster is lower, we could detect homologues of all known core conglobatin genes (*congA-E*) with the same domain architecture as in the reference BGC. M10 and M19 contain a hopene cluster with similarity scores of 84% and 76%, to MIBiG cluster BGC0000663, which is primarily due to the absence of two and three hypothetical genes, respectively. Taken together, the isolates produced some known bioactive molecules, but the majority of the mass features could not be dereplicated.

## Discussion

A huge proportion of the bacterial world remains uncharacterised, representing an important reservoir of biological information and new chemistry as the possible basis for future drugs<sup>41,200,221</sup>. Metagenome sequencing, varying culturing techniques, and isolation of bacteria from rare environments is gradually revealing part of this microbial dark matter, and ancient samples are a promising and underexplored resource<sup>201,202</sup>. We had the unique opportunity to sample faeces from the intestinal tract of an exceptionally well-preserved 28,000-year-old mammoth, which we used to isolate ancient Actinobacteria<sup>207</sup>. These Actinobacteria belong to the well-studied genus *Streptomyces* and to the underexplored genera *Sanguibacter*, *Micromonospora*, *Oerskovia*, and *Saccharopolyspora*. Genomic analysis of the strains revealed a large phylogenetic distance to currently known strains and much uncharacterised biosynthetic potential. With a mammoth, multiple replicates are not readily obtained, to say the least. Therefore, minimizing the contamination risk during the excavation process is particularly important. The specimen was transported and stored in frozen condition. As thawing set in from the outside of the carcass and slowly proceeded inward, each day freshly thawed samples could be taken. As the intestines that we encountered were still intact, we could open them under sterile conditions and extract faeces with a low risk of external contamination. Hence, a lot of attention was paid to the careful removal of the faecal sample, and the specific sample we eventually used for isolation of the Actinobacteria was then extracted from the very core of the large faecal sample, so as to eliminate the risk of contamination to the absolute minimum. The intestinal lumen was thoroughly expected for defects and faeces was taken and stored using sterile materials.

A previous metagenome-based study revealed the presence of actinobacterial reads in different tissues of mammoth remains <sup>222</sup>. In this study, we have now been able to isolate Actinobacteria from such an environment, strengthening these findings and allowing us to further explore their functional role and potential. Studying metagenome data always has the risk for misassembly of (damaged) DNA, which is not the case when isolates are grown up and then sequenced. Taxonomic analysis revealed a significant phylogenetic difference between the isolates and their closest known neighbours, strongly suggesting that these isolates represent novel species. The low similarity to known strains can be attributed to the limited number of publicly available complete genomes for underexplored taxa such as *Oerskovia*, *Sanguibacter* and *Saccharopolyspora*, with three, four and eight available complete genome sequences, respectively, as compared to 323 complete *Streptomyces* genomes. Nevertheless, our *Streptomyces* isolates also showed low similarity scores, while many publicly available complete genomes were available for comparison. The mammoth isolates are therefore a valuable addition to the database and illustrate the power of harvesting microbes from unique environments, even for well-studied genera.

Ancient samples allow a glimpse into the past and may provide evolutionary insights. An interesting question that arises is what the significance is of the major differences that we observe between the isolates and their closest known relatives. The phylogenetic distance between the isolates and their known relatives makes it hard to evaluate evolutionary differences on a genomic level. For better comparison, genome sequences of more closely related strains are required; however, it is yet not known if such strains can be found in the available strain collections.

Genome mining of the isolates revealed a wide variety of BGCs, with 19% encoding PKSs, 14% NRPSs and 14% terpene biosynthetic enzymes. Most of the predicted BGCs showed low or no similarity to any known BGCs. Comparative genome analyses between M19 and its taxonomic neighbours found in publicly available sequence data, *Streptomyces* sp. CNZ306 and *Streptomyces* sp. CNQ-509, revealed a shared conserved internal region and a less conserved region near the ends of the chromosome, holding high numbers of unique and uncharacterised BGCs. This is consistent with results from previous studies showing that common BGCs are often located in the internal regions of the chromosome, while more unique genes are located towards the subtelomeric regions <sup>21,42,115</sup>. Surprisingly, only two BGCs, encoding the desferrioxamine and ectoine pathways, were shared between isolate *Streptomyces* sp. M19 and its neighbours. These specialised metabolites play an important role in survival and their BGCs are well conserved among Actinobacteria <sup>42,223</sup>. The limited amount of shared biosynthetic content between *Streptomyces* sp. M19 and its neighbours is

unexpected as phylogeny has been shown to be an important indicator of BGC distribution<sup>100,224</sup>. This underlines the opportunities offered in terms of the biosynthetic potential of bacteria isolated from underexplored environments.

The majority of the isolates displayed bioactivity against one of the tested indicator strains, both through production of volatile and non-volatile compounds. The antibacterial activity against Gram-positive bacteria by *Oerskovia* sp. M15 is surprising, because to the best of our knowledge, antibacterial activity has never been reported before for members of this genus. Four potential BGCs were identified in the genome sequence of M15, of which three had similarity scores below 12%. However, the crude extracts did not show any activity and chemical dereplication did not give any hits with known bioactive compounds. Future work in our laboratory will focus on identifying the BGC and cognate natural product linked to the bioactivity of *Oerskovia* sp. M15. The crude extracts of *Streptomyces* isolates M10 and M19 exhibited antibiosis against *B. subtilis*. Molecular networking and chemical dereplication using the GNPS platform resulted in the annotation of several known natural products, including antimycin and monactin in the extracts of M10, and conglobatin in the extracts of M19. This was supported by the detection of their BGCs by antiSMASH. Interestingly, antimycin- and conglobatin-related molecules were also detected in the crude extracts of polar Actinobacteria isolated from ancient sediment cores<sup>203</sup>. An evolutionary path of the antimycin BGC has been proposed in which the L-form BGC is appointed as the ancestor of other antimycin BGCs<sup>225</sup>. Therefore, it's not surprising that our detected ancient BGC is the ancestral L-form antimycin cluster. Additionally, analysis of the biogeographical and phylogenetic distribution of the antimycin BGC has revealed that this BGC is widespread among *Streptomyces* spp. and across the globe<sup>225</sup>. While these known metabolites might be responsible for the bioactivity of the crude extracts, other unknown metabolites with antibacterial activity might be produced as well. The GNPS network of the metabolome of these isolates revealed several spectral families unique to the isolates which could not be dereplicated. Proteomics or transcriptomics could be used to analyse whether other BGCs are expressed under these growth conditions.

Much interest has been directed towards the presence of antibiotic-producing Actinobacteria in the mammalian microbiome as this could point towards a protective role against infection by pathogens<sup>99,181</sup>. However, these bacteria have been isolated from a faeces sample and we therefore cannot distinguish whether these strains are part of the mammoth microbiome or whether they are simply passing through as part of digested plant material.

Overall, we have isolated a diverse collection of Actinobacteria from a unique ancient stool sample. The large phylogenetic distance between the isolates and their modern siblings and

the high percentage of uncharacterised biosynthetic potential, even in the well-studied genus *Streptomyces*, illustrate that we have by far captured all microbial diversity. Future studies on the evolutionary differences between the isolates and their modern siblings may allow for a unique glance into history.

## Methods

### Sample acquisition

In February 2014 the specimen was transported by truck to Yakutsk. Dissection of the specimen was done in the autopsy room of the medical faculty of the university of Yakutsk. The samples for microbiological analysis were extracted from 10-14 March 2014, the timespan in which the remains completely thawed. As thawing set on from the outside of the carcass and slowly proceeded inward, each day freshly thawed samples could be taken. Approached from the right side, the intestines were exposed on March 13 after carefully removing soft-tissue and ribs over the previous days. A 60 cm intestinal specimen was taken out of the remains and placed on a sterile surface for further examination. No defects of the intestinal lumen were found. After inspection, two perpendicular (5 cm) incisions were made, and the intersection was folded over, exposing the intestinal lumen. Faecal samples were carefully taken from the lumen using flocked swab collecting tubes (eSwab Copan) or deposited in sterile collection tubes with disposable tweezers with minimal possible cross-contamination. Sterile examination gloves and instruments were used during the whole procedure.

### Isolation of Actinobacteria

Approximately 100 mg of faeces was aseptically placed into a sterile Eppendorf, dissolved in  $\text{dH}_2\text{O}$ , and serially diluted ( $10^{-1}$ – $10^{-6}$ ). Serial dilutions were plated onto different agar media. The media were glucose agar (GA) <sup>138</sup>, humic acid agar (HA) <sup>226</sup>, mannitol soya flour medium (SFM) <sup>227</sup>, modified Starch-Casein agar (MSCA) <sup>228</sup>, minimal medium without carbon sources (MM) <sup>227</sup>, MM + 1% glycerol (w/v) (Y), MM + 1% mannitol (w/v) (A). All media contained nystatin (50  $\mu\text{g/mL}$ ) and nalidixic acid (10  $\mu\text{g/mL}$ ) for the inhibition of fungi and Gram-negative bacteria respectively. Plates were incubated at 30 °C, 4 °C, and anaerobically at room temperature. Single actinomycete colonies were streaked onto SFM agar plates until pure and cryopreserved with glycerol (20 %) and stored at  $-80^\circ\text{C}$ .

### Genome sequencing

Strains were cultured in TSBS at 30 °C with 200 rpm shaking speed. Genomic DNA was isolated by phenol-chloroform extraction as described previously <sup>227</sup> and sent to be commercially sequenced at Future Genomics Technologies, The Netherlands. Genomes were sequenced



using the MinION Nanopore sequencing platform and Illumina NovaSeq6000. Hybrid assembly (both Illumina and ONT reads) was performed for each isolate using Unicycler (v0.4.0.7) <sup>229</sup>. Briefly, Unicycler performs a SPAdes assembly of the Illumina reads and then scaffolds the assembly graph using long reads. Unicycler polishes its final assembly with Illumina reads and uses Pilon <sup>230</sup> to reduce the rate of small base-level errors.

### Phylogenetic analysis

The initial taxonomy of the six isolates was determined using 16S rRNA sequences and BLASTN, with the highest scoring sequencing hits reported. To determine the phylogenetic class of the isolates on a whole genome scale, 578 high quality Actinobacteria genome sequences have been downloaded from the PhyloPhlAn v3.0 <sup>231</sup> database, using the phylophlan\_setup\_database function. Next, the phylophlan\_write\_config\_file script is employed to create a configuration file with DIAMOND as mapping tool, MAFFT for the multiple sequence alignment, trimAl for alignment trimming and IQ-TREE for generating a phylogenetic tree. FastANI v1.32 <sup>232</sup> was used to calculate the relatedness of the isolates and neighbouring strains.

### Bioactive potential and comparative genome analysis

AntiSMASH v6.0 <sup>217</sup> was used under default settings to predict BGCs from the six isolated bacteria and their modern-day closest known neighbours. The resulting predicted BGCs were then used as input for BiG-SCAPE <sup>108</sup> to analyse BGC clustering.

To annotate the genomes and identify coding regions, Prokka v1.14.6 <sup>233</sup> was applied on the assemblies of the isolates. The strain *Streptomyces* sp. CNQ-509 was downloaded from NCBI using accession number GCA\_001011035.1. The EMBL-EBI database <sup>234</sup> was questioned to obtain the sequence information of strain *Streptomyces* sp. CNZ306 (PGEW01000000). Next, BLASTP was used to identify similarity between the isolates and their neighbours, whereafter MCSanX <sup>235</sup> was utilised to predict gene collinearity and synteny. Finally, the coordinates of the collinear regions and predicted BGCs were used as input for the visualizing tool Circos v0.69-8 <sup>236</sup>.

### Stereomicroscopy and Scanning Electron Microscopy (SEM)

Isolates were grown for nine days on SFM. Stereo microscopy was done using a Leica MZ16 FA microscope equipped with a Leica DFC420 C camera. SEM studies were performed using a JEOL JSM-7600F scanning electron microscope. Single colonies were excised from agar plates and the bottom layer of agarose was cut off to minimise the thickness of the sample. The sections were glued upon a Cryo-EM stub and the whole stub was submerged in non-boiling liquid nitrogen and frozen for 20 sec. The stub was then transferred to the gatan Cryo unit, heated to -90 °C for 2 minutes to remove any ice crystals formed during the transfer,

cooled to  $-120^{\circ}\text{C}$ , and coated with gold/palladium (80/20) using a sputter coater. Hereafter, the samples were transferred into the microscope and kept at  $-120^{\circ}\text{C}$  while imaging.

### Antimicrobial activity assays

*Bacillus subtilis* 168, *Escherichia coli* ASD19<sup>237</sup>, and *Pseudomonas aeruginosa* PA01 were used as indicator strains for antimicrobial activity and were cultured in LB media at  $37^{\circ}\text{C}$ . Antimicrobial activity assays were conducted in liquid and on plate, using different methods:

*Liquid cultures:* Strains were grown in ISP2 (DSMZ #987) and NMMP<sup>227</sup> medium for seven days. Wells were performed in soft LB agar (1.8% w/v agar) containing one of the indicator strains pre-grown in liquid LB to exponential phase ( $\text{OD}_{600} = 0.4 - 0.6$ ) and filled with  $100\ \mu\text{L}$  culture supernatant. Plates were incubated overnight at  $37^{\circ}\text{C}$  ( $\pm 18$  hours) and the following day, the zone of inhibition was determined.

*Cross streak method:* each strain was independently inoculated on NA (Difco) and ISP2 agar plates as a single streak in the centre of the plate and incubated for seven days to allow the strains to grow and produce antibiotics. The plates were then seeded with the indicator strains pre-grown in liquid LB to exponential phase ( $\text{OD}_{600} = 0.4 - 0.6$ ) by streaking perpendicular to the line of actinobacterial growth and incubated overnight at  $37^{\circ}\text{C}$  ( $\pm 18$  hours). The following day, the zone of inhibition was determined.

*The double-layer agar method:* Strains were manually spotted ( $2\ \mu\text{L}$ ) on minimal medium agar plates (MM) supplemented with 0.5% mannitol and 1% glycerol (w/v) as non-repressing carbon sources, and Czapek Dox plates. After seven days of incubation, plates were overlaid with soft LB agar (1.8% w/v agar) containing one of the indicator strains pre-grown in liquid LB to exponential phase ( $\text{OD}_{600} = 0.4 - 0.6$ ) and incubated overnight at  $37^{\circ}\text{C}$  ( $\pm 18$  hours). The following day, the zone of inhibition was determined.

*Volatile assay:* The antimicrobial activity of volatile production was assessed using a petri dish with two equally sized compartments, both filled with NA. Mammoth isolates were streaked on one site and plates were incubated for seven days after which, *E. coli* or *B. subtilis* were inoculated on the other side using a concentration of  $10^4$  and  $10^3$  CFU/mL respectively.

### Metabolite profiling

Isolates M10, M15, M19 and M46 were grown confluent on NA plates for seven days. The agar plates were cut into small pieces, soaked overnight in ethyl acetate (EtOAc) to extract the metabolites, evaporated at room temperature, and dissolved in methanol (MeOH) to a concentration of  $1\ \text{mg/mL}$ . LC-MS/MS acquisition was performed using Shimadzu Nexera X2



ultra high-performance liquid chromatography (UPLC) system, with attached photodiode array detector (PDA), coupled to Shimadzu 9030 QTOF mass spectrometer, equipped with a standard electrospray ionisation (ESI) source unit, in which a calibrant delivery system (CDS) is installed. A total of 2  $\mu$ L was injected into a Waters Acquity HSS C18 column (1.8  $\mu$ m, 100 Å, 2.1  $\times$  100 mm). The column was maintained at 30 °C and run at a flow rate of 0.5 mL/min, using 0.1% formic acid in H<sub>2</sub>O, and 0.1% formic acid in acetonitrile (ACN) as solvents A and B, respectively. The gradient used was 5% B for 1 min, 5–85% B for 9 min, 85–100% B for 1 min, and 100% B for 4 min. The column was re-equilibrated to 5% B for 3 min before the next run was started. The PDA acquisition was performed in the range of 200–600 nm, at 4.2 Hz, with 1.2 nm slit width. The flow cell was maintained at 40 °C. All the samples were analysed in positive polarity, using data dependent acquisition mode. In this regard, full scan MS spectra ( $m/z$  100–1700, scan rate 10 Hz, ID enabled) were followed by two data dependent MS/MS spectra ( $m/z$  100–1700, scan rate 10 Hz, ID disabled) for the two most intense ions per scan. The ions were fragmented using collision induced dissociation (CID) with fixed collision energy (CE 20 eV), and excluded for 1 s before being re-selected for fragmentation. The parameters used for the ESI source were: interface voltage 4 kV, interface temperature 300 °C, nebulizing gas flow 3 L/min, and drying gas flow 10 L/min.

### MS/MS-based molecular networking and dereplication

MS/MS raw data (obtained from Shimadzu 9030 QTOF MS) were converted to a 32-bit mzML file using MSConvert (ProteoWizard) and a molecular network was assembled using the online workflow on the Global Natural Product Social Molecular Networking (GNPS) website<sup>220</sup>. Both the precursor ion and the MS/MS fragment ion mass tolerance were set to 0.02 Da. The minimum cosine score was set to 0.7 and the minimum matched peaks set to 6. The MSCluster algorithm was run with a minimum cluster size of 2 spectra. The spectra in the network were searched against the GNPS spectral libraries. For this, the precursor ion and the MS/MS fragment ion mass tolerance were set to 0.5 Da. Matches between network spectra and library spectra required a minimum score of 0.7 and at least 6 matched peaks. Cytoscape 3.8.2 was used for visualization of the generated molecular networks<sup>238</sup>. The edge thickness was set to represent the cosine score, with thicker lines indicating higher similarity between nodes. The molecular networking job in GNPS can be accessed at <https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=e160b564fc7e48e6b82394991bfd79be>.

### Bioactivity crude extracts

The activity of the crude extracts was determined in triplicate. Indicator strains were pre-grown in liquid LB to exponential phase (OD<sub>600</sub> = 0.4 – 0.6). Cultures were diluted to OD<sub>600</sub> = 0.01 in LB and 100  $\mu$ L diluted culture was loaded in wells of a 100-well honeycomb plate. 200

$\mu\text{g}$  crude extract was added to the bioactivity of the crude extract. Additionally, the following controls were added: LB, bacterial dilution (growth control), bacterial cells + 6  $\mu\text{g}$  ampicillin (positive control), and bacterial cells + MeOH (negative control). Subsequently, the optical density at 600 nm was measured every 30 min for 16 hours using a Bioscreen C Reader (Thermo Scientific, Breda, The Netherlands), with continuous shaking. The absorption at 600 nm was plotted against the time.

# Supplementary information for Chapter 3

Table S1. Isolate taxonomic classification and identified closest neighbours.

Isolate ID	Taxonomic classification	Closest known neighbour(s)	Accession number	ANI score %
M9	<i>Sanguibacter</i>	<i>Sanguibacter antarcticus</i>	GCA_002564005.1	84.2
M10	<i>Streptomyces</i>	<i>Streptomyces</i> sp. CB02460	GCA_001905705.1	84.0
		<i>Streptomyces</i> sp. 2131.1	GCA_900105515.1	84.2
M12	<i>Micromonospora</i>	<i>Micromonospora chokoriensis</i>	GCA_900091505.1	89.5
M15	<i>Oerskovia</i>	<i>Oerskovia enterophila</i>	GCA_001624335.1	88.6
		<i>Oerskovia</i> sp. Root22	GCA_001429135.1	88.5
		<i>Oerskovia</i> sp. Root918	GCA_001428945.1	88.6
M19	<i>Streptomyces</i>	<i>Streptomyces</i> sp. CNZ306	PGEW01000000	79.8
		<i>Streptomyces</i> sp. CNQ-509	GCA_001011035.1	80.0
M46	<i>Saccharopolyspora</i>	<i>Saccharopolyspora flava</i>	GCA_900116135.1	88.1

Table S2. Isolates from extreme environments and their similarity scores with closest neighbours.

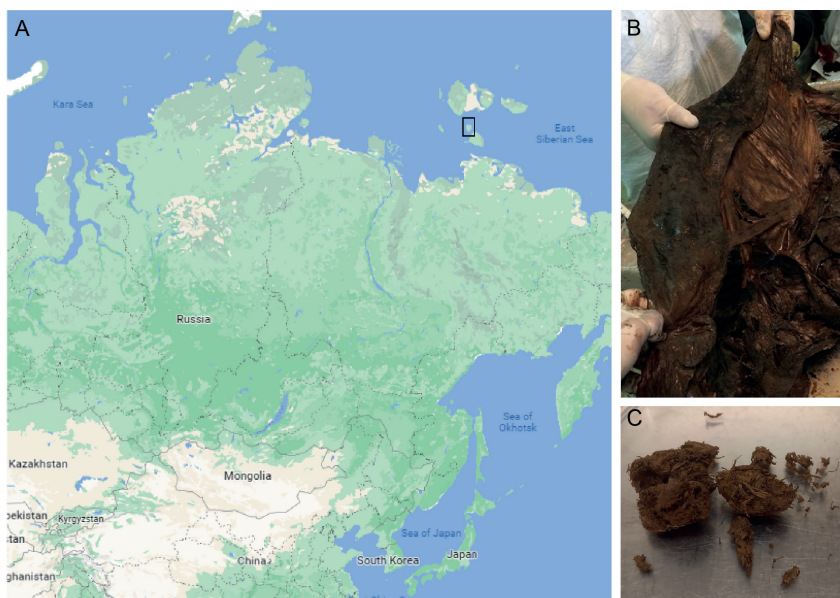
Isolate ID	Origin	Accession number	Closest known neighbour(s)	Origin neighbour(s)	Accession number neighbour(s)	ANI score (%)
<i>Streptomyces</i> sp. NTK 937	Deep-sea, Canary Basin, Canary Islands	GCA_000698495.1	<i>Streptomyces</i> sp. SM18	Deep-sea, Kilkieran Bay, Galway, Ireland	GCA_002910775.2	96.7
<i>Streptomyces</i> sp. SCSIO 3032	Deep-sea, Madeira archipelago, Portugal	GCA_002128305.1	<i>Streptomyces</i> sp. MP131-18	Deep-sea, Trondheim, Norway	GCA_001984575.1	85.5
<i>Streptomyces jeddahensis</i>	Desert, Jeddah, Saudi Arabia	GCA_001642995.1	<i>Streptomyces</i> sp. AmelKG-E11A	Insect	GCA_002705975.1	81.4
<i>Streptomyces</i> sp. BF-3	Great Salt Plains, Oklahoma	GCA_002104865.1	<i>Streptomyces</i> sp. Cmuel-A718b	Unknown	GCA_900092005.1	98.9
<i>Streptomyces</i> sp. 4F	Great Salt Plains, Oklahoma	GCA_002104855.1	<i>Streptomyces</i> sp. Alain-F2R5	Al Ain, Dubai, United Arab Emirates	GCA_002277855.1	93.1
			<i>Streptomyces</i> sp. CB02414	Beach soil, Dubai, United Arab Emirates	GCA_001905385.1	99.2

**Table S3. Percentage of BGCs that share less than 50% KnownClusterBlast similarity to BGCs within the MIBiG database of the mammoth isolates and a selection of Actinobacteria**

Strain	Isolation source	total # BGCs	% BGC < 50% similarity*
<b><i>Sanguibacter</i> sp. M9</b>	mammoth faeces	6	67%
<i>Sanguibacter gelidistatuariae</i> ISLP-3	ice sculpture, Antarctica	3	67%
<i>Sanguibacter keddieii</i> DSM 10542	bovine blood	5	80%
<i>Sanguibacter massiliensis</i>	human stool	2	50%
<b><i>Micromonospora</i> sp. M12</b>	mammoth faeces	28	79%
<i>Micromonospora auratinigra</i> DSM 44815	soil, Thailand	16	88%
<i>Micromonospora chokoriensis</i> DSM 45160	sandy soil	15	73%
<i>Micromonospora inosilota</i> DSM 43819	forest soil	9	89%
<i>Micromonospora narathiwatensis</i> DSM 45248	peat swamp forest soil	18	67%
<i>Micromonospora terminaliae</i> DSM 101760	plant	13	77%
<i>Micromonospora zamorensis</i> DSM 45600	rhizosphere	12	83%
<b><i>Oerskovia</i> sp. M15</b>	mammoth faeces	4	75%
<i>Oerskovia</i> sp. KBS0722	soil	6	83%
<i>Oerskovia</i> sp. Root22	roots of <i>Arabidopsis thaliana</i>	3	33%
<b><i>Saccharopolyspora</i> sp. M46</b>	mammoth faeces	20	80%
<i>Saccharopolyspora coralli</i> E2A	stony coral	10	70%
<i>Saccharopolyspora erythraea</i> NRRL 2338	soil	36	72%
<i>Saccharopolyspora</i> sp. ASAGF58	soil	29	69%
<b><i>Streptomyces</i> sp. M10</b>	mammoth faeces	31	65%
<b><i>Streptomyces</i> sp. M19</b>	mammoth faeces	34	65%
<i>Streptomyces antibioticus</i> DSM 40234	soil	26	50%
<i>Streptomyces</i> sp. MP113-05	sponge	28	79%
<i>Streptomyces cacaoi</i> H2S5	moss soil	31	45%
<i>Streptomyces coelicolor</i> CFB_NBC_0001	soil	27	41%
<i>streptomyces ficellus</i> NRRL 8067	soil	26	73%
<i>streptomyces galilaeus</i> ATCC 14969	soil	24	50%
<i>streptomyces griseus</i> ATCC 13273	soil, Japan	30	60%
<i>Streptomyces leeuwenhoekii</i> C34	desert soil	35	69%
<i>Streptomyces lincolnensis</i> NRRL 2936	soil	29	59%
<i>Streptomyces lunaelactis</i> MM109	moonmilk speleothem	28	71%
<i>streptomyces olivochromogenes</i> DSM 40451	soil	39	72%
<i>streptomyces scabiei</i> 87.22	plant pathogen	34	59%
<i>streptomyces tsukubensis</i> AT3	forest soil, China	36	64%

**Table S4. *m/z* values annotated by GNPS present in the extracts of the isolated bacteria but not in the medium blank.**

<i>m/z</i> value	Adduct	Annotated by GNPS as	Present in extracts of
203.179	[M + H - H <sub>2</sub> O] <sup>+</sup>	Alismol (terpenoid)	<i>Streptomyces</i> sp. M19
415.123	[M + H] <sup>+</sup>	Futalosine	<i>Streptomyces</i> sp. M10
546.488	[M + H] <sup>+</sup>	Aminobacteriohopanetriol	<i>Streptomyces</i> sp. M10, <i>Streptomyces</i> sp. M19
549.281	[M + H] <sup>+</sup>	Antimycin A1	<i>Streptomyces</i> sp. M10
535.266	[M + H] <sup>+</sup>	Antimycin A2	<i>Streptomyces</i> sp. M10
401.253	[M + H] <sup>+</sup>	Bonactin	<i>Streptomyces</i> sp. M10
383.243	[M + H - H <sub>2</sub> O] <sup>+</sup>		
499.281	[M + H] <sup>+</sup>	Conglobatin	<i>Streptomyces</i> sp. M19
415.269	[M + H] <sup>+</sup>	Homononactyl homononactate	<i>Streptomyces</i> sp. M10
454.293	[M + H] <sup>+</sup>	1-palmitoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine	<i>Streptomyces</i> sp. M10
768.489	[M + NH <sub>4</sub> ] <sup>+</sup>	Monactin	<i>Streptomyces</i> sp. M10
237.148	[M + H - H <sub>2</sub> O] <sup>+</sup>	Gliocladic acid (terpenoid)	<i>Saccharopolyspora</i> sp. M46
203.118	[M + H] <sup>+</sup>	Nb-acetyltryptamine	<i>Streptomyces</i> sp. M10, <i>Saccharopolyspora</i> sp. M46, <i>Oerskovia</i> sp. M15
416.265	[M + NH <sub>4</sub> ] <sup>+</sup>	NCGC00381146-01!2-[5-[2-[2-[5-(2-oxopropyl)oxolan-2-yl]propanoyloxy]butyl]oxolan-2-yl]propanoic acid	<i>Streptomyces</i> sp. M10
373.274	[M + H - H <sub>2</sub> O] <sup>+</sup>	12-Ketodeoxycholic acid	<i>Oerskovia</i> sp. M15
402.301	[M + H] <sup>+</sup>	(Z)-N-hexadec-9-enoyl-L-phenylalanine	<i>Streptomyces</i> sp. M10

**Figure S1. Origin and visualization of the mammoth sample.** A) the mammoth carcass discovered in August 2012 on Maly Lyakhovsky Island, Russia. B) extracted intestinal specimen C) faecal sample.

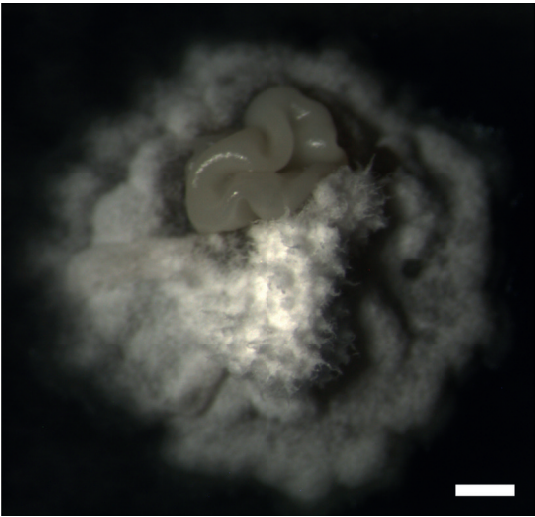


Figure S2. Different morphologies are observed within single colonies of isolate *Streptomyces* sp. M19.

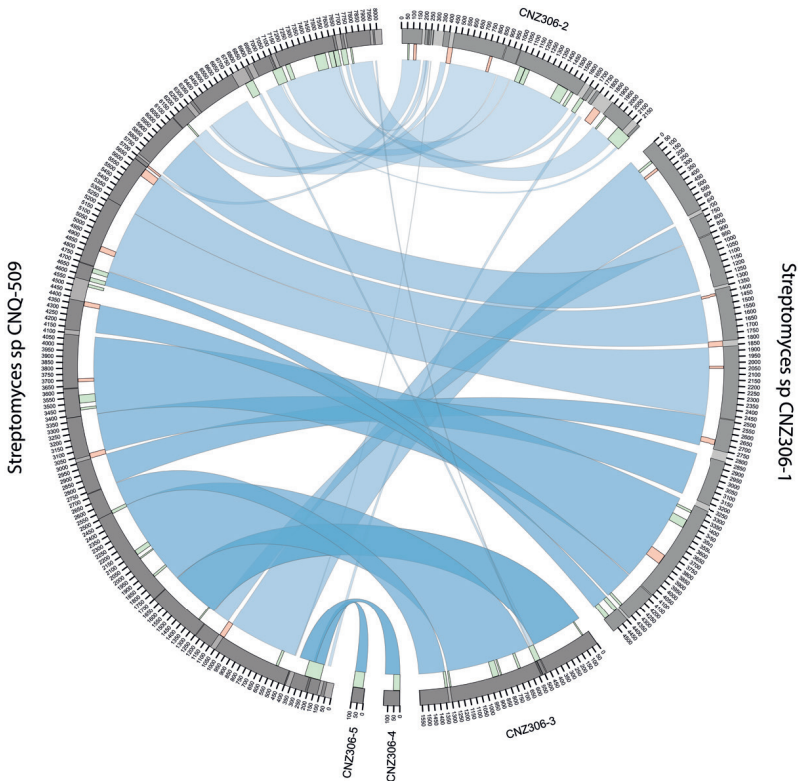
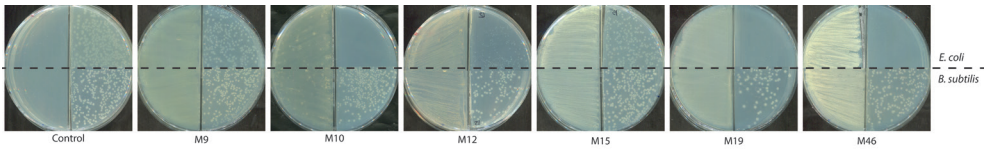
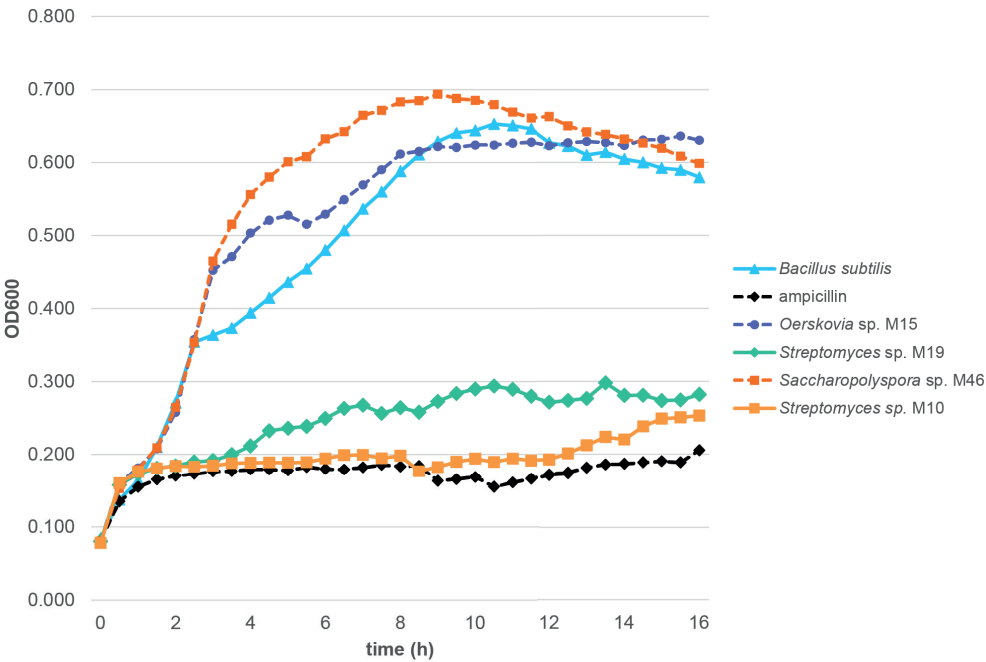


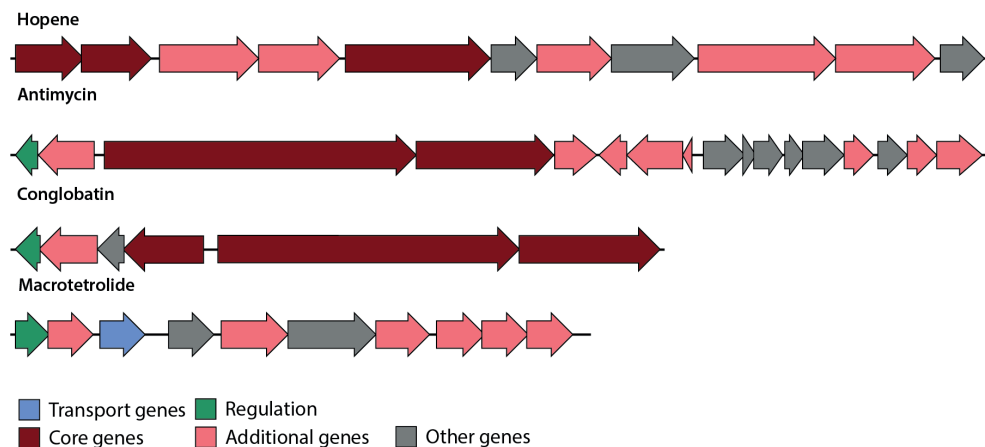
Figure S3. Synteny plot of *Streptomyces* sp. CNQ-509 and *Streptomyces* sp. CNZ306.



**Figure S4. Bioactivity of volatiles released by the mammoth isolates against *E. coli* strain ASD19 and *B. subtilis* 168 after seven days of growth on NA.**



**Figure S5. Antimicrobial activity of crude extracts against *B. subtilis***



**Figure S6. Predicted hopene cluster of M10 & M19, antimycin and macrotetrolide gene cluster of M10 and the conglobatin gene cluster of M19.** The gene colours are based on the colouring system of antiSMASH v6.0.



