

The Chemical Ecology of Geosmin

Liana Zaroubi

A Thesis in
The Department of
Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements
for the Degree of Master Science, Chemistry (Biochemistry)
at Concordia University
Montréal, Québec, Canada

January 2021

© Liana Zaroubi, 2021

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Liana Zaroubi

Entitled: The Chemical Ecology of Geosmin

and submitted in partial fulfillment of the requirements for the degree of

Master of Science, Chemistry (Biochemistry)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Yves Gélinas

_____ Examiner
Dr. Pat Forgione

_____ Examiner
Dr. Vincent Martin

_____ Thesis Supervisor(s)
Dr. Brandon Findlay

_____ Thesis Supervisor(s)

Approved by _____
Dr. Yves Gélinas Chair of Department or Graduate Program Director

Dr. Pascale Sicotte

Dean of Faculty

Abstract

The Chemical Ecology of Geosmin

Liana Zaroubi

Known as the smell of earth after rain, geosmin is an odorous terpene detectable by humans at picomolar concentrations. Geosmin production is heavily conserved in actinobacteria, myxobacteria, cyanobacteria, some fungi and red beets (*Beta vulgaris* L.), suggesting it strongly contributes to the fitness of these organisms. Prokaryotes produce geosmin through the terpene cyclase “geosmin synthase” converting germacradienol into geosmin via isoprenoid pathways (methylerythritol phosphate, mevalonate or leucine-dependent). However, the universality and ecological role of geosmin is poorly understood. I theorized that geosmin is an aposematic signal used to indicate the unpalatability of toxin-producing microbes, discouraging predation by eukaryotes. Consistent with this hypothesis I have found that geosmin reduces predation of *Streptomyces coelicolor* and *Myxococcus xanthus* by the bacteriophagous *Caenorhabditis elegans*. Predation was restored by the removal of both terpene biosynthetic pathways or deletion of the *C. elegans* ASE sensory neuron and resulted in the death of the nematodes. Geosmin itself was non-toxic. This is the first warning chemical to be identified in bacteria or fungi, and suggests molecular signalling affects microbial predator-prey interactions in a manner similar to the well-studied visual markers displayed by poisonous animal prey. In line with geosmin being a warning chemical, I also determined from a bioinformatics analysis that geosmin synthase was acquired from a distant past through horizontal gene transfer with *Actinobacteria* as the ancestral genetic reservoir. In addition, my work also suggests that several genes not directly involved in the biosynthesis of geosmin nevertheless support its activity. The investigation of these genes may enhance our understanding of the biosynthesis and regulation of geosmin and terpenes in prokaryotes.

Acknowledgements

Throughout this thesis I received a lot of guidance, assistance and support from my mentors, colleagues, friends and family. I hope that in this section I will be able to express and recognize the invaluable assistance that you all provided during my study.

First and foremost, I would like to thank Dr. Brandon Findlay for all his guidance during my MSc. His expertise was invaluable in developing the research questions and methodology. You helped me grow as a scientist but also as a person as I became more confident in taking on challenging tasks.

I am grateful for the highly collaborative and friendly environment offered at Concordia University since I was fortunate to work with aspiring scientists. For this, I would like to thank Imge Özügergin, Karina Mastronardi and Dr. Alisa Piekny for all their help with the *C. elegans* experiments. Especially Imge, who not only taught me how to work with nematodes but also came on weekends to perform experiments with me during the COVID-19 pandemic. I wish to show my gratitude to Philippe Prevost for helping me synthesize 2-methylisoborneol in the Forgione Lab and for all his support during my MSc. Additionally, I would like to thank Anic Imfeld and Dr. Yves Gélinas for their assistance in the performed GC-MS analyses, Dr. Chris Law for microscopy support and development of a python script, Dr. Emma Despland for allowing me to use her lab space during the pandemic and Dr. Klas Flärdh for sending me valuable bacterial strains from Sweden.

I would like to thank my committee members, Dr. Pat Forgione and Dr. Vincent Martin for their feedback, advice and guidance during my MSc. You made me realize how ambitious my project was and drove me to set deadlines to finish my MSc and for that I am very grateful.

During my MSc I had the pleasure to supervise Rami Antoun, an amazing CUSRA student who showed great assistance for the bioinformatics side-project. I wish you success in all your future endeavours!

I would also like to pay special regards to all the members in my lab especially Michael Mahdavi and Lydia Rili, who both proofread this manuscript. Thank you everyone for instilling a comfortable and friendly atmosphere in the lab! It has been a great pleasure to work with you!

Finally, I would like to thank my family who always believed in me and strongly supported my career path.

Contributions of Authors

Shown below are the contributions of each of the authors in this thesis.

Liana Zaroubi: Wrote the first draft of this thesis and corrected subsequent versions. Performed all listed experiments and analyses.

Dr. Brandon Findlay: Project supervisor responsible for correcting the manuscript. Designed and ran the python script for the gene co-occurrence for the bioinformatics analysis. The running of the script was performed in the centre for structural and functional genomics (CSFG) at Concordia University with the help of Andrei Wasyluk.

Imge Özügergin: Helped perform the behavioural, predation (*S. coelicolor*) and terpene add-in assays with *C. elegans*. Maintained all wild type and mutant *C. elegans* strains for experiments.

Rami Antoun: Helped on the bacterial genome selection, the environmental isolation categorization of geosmin producers, the phylogenetic, co-occurring and adjacent genes analyses (with a higher focus on myxobacteria).

Karina Mastronardi: Assisted on the chemotaxis experiment using *C. elegans*. Maintained *C. elegans* N2 (wild type) to perform chemotaxis experiment.

Anic Imfeld: Developed GS-MS method for geosmin detection and performed GC-MS analyses for geosmin quantitation by *M. xanthus* DK1622 and determination of media versus cytoplasmic geosmin.

Philippe Prevost: Helped synthesize 2-methylisoborneol in the Forgiere Lab.

Dr. Chris Law: Offered microscopy support and developed the python script to calculate track line percentage to measure linearity of movement of *C. elegans*.

Dr. Yves Gélinas: Ran GC-MS samples for geosmin quantitation of *M. xanthus* DK1622 over 9 days. Provided the information to write the GC-MS method for geosmin detection for this thesis.

Dr. Alisa Piekny: Provided guidance for the design of all experiments with *C. elegans*.

Copyright Information

Sections 2.g-h, and the abstract are adapted from a manuscript submitted for publication on December 19, 2020.

Sections 1.c.ii, 2.i, 3 and the abstract are adapted from a manuscript submitted for publication on January 29, 2021.

Table of Contents

List of Figures	x
List of Tables	xii
List of Abbreviations	xiii
1. Introduction	1
a. Discovery, characterization, and detection of geosmin.....	1
i. Discovery and producers of geosmin	1
ii. Physical and chemical properties of geosmin.....	1
b. The biosynthesis of geosmin	2
i. IPP production via the MVA pathway	2
ii. IPP production via the MEP pathway.....	3
iii. MVA production via the leucine-dependent pathway	4
iv. Production of geosmin from FPP	4
c. The ecological role of geosmin remains to be fully elucidated.	6
i. Geosmin offers an unknown fitness advantage.	6
ii. The gene encoding for geosmin biosynthesis is spread via horizontal gene transfer... 6	
d. Literature review on the potential function of geosmin	8
i. Geosmin as an antibiotic against <i>Salmonella typhimurium</i>	8
ii. Geosmin as a side product of carotenoid synthesis	9
iii. <i>Drosophila</i> is repelled specifically by geosmin in moldy fruit.....	9
iv. Geosmin as an attractant towards arthropods.	11
v. Geosmin as an attractant to various eukaryotes.....	12
e. The life cycle of geosmin producers and their ecology.....	13
i. Growth stage at which geosmin production is activated.	14
1. Actinobacteria (<i>Streptomyces</i> spp. as representative genus)	14
2. Myxobacteria	15
3. Cyanobacteria	16
4. Fungi.....	17
ii. Common life cycle challenges of geosmin producers	18
1. Symbiosis needed for growth and survival.....	18
2. Quorum sensing is important for bacterial communities.....	19
iii. Major stressors for geosmin producers.....	20

1.	Major predators of geosmin producers	21
a.	Amoeba	21
b.	Nematodes.....	22
2.	Results and discussion	23
a.	Effect of geosmin on the lag phase of <i>E. coli</i> cells.....	23
b.	Antibiotic properties of geosmin	23
c.	Effect of geosmin on protein stabilization	25
i.	Geosmin production by <i>M. xanthus</i>	25
ii.	Circular dichroism analysis of the effect of geosmin on bovine serum albumin.	26
d.	Effect of geosmin on protease activity	28
e.	Effect of geosmin on quorum sensing.	29
i.	Violacein assay on 96-well plates	30
ii.	Violacein assay using agar broth dilutions	30
iii.	The function of geosmin as a quorum sensing inhibitor is enantioselective	31
f.	Geosmin as a defense chemical against amoeba predation	31
i.	Predation assay in liquid media	31
ii.	Predation assay in solid media.....	33
iii.	Swarming assay on solid media.....	33
g.	Geosmin as an aposematic signal to <i>C. elegans</i>	34
i.	Chemotaxis experiment with <i>C. elegans</i>	34
ii.	Geosmin alters nematode behaviour.....	36
iii.	Geosmin deters feeding on its producers.....	37
h.	Geosmin is a warning chemical.....	39
i.	Bioinformatics analysis on GS genes among prokaryotes.....	41
i.	Geosmin biosynthetic genes are well dispersed and interrelated in prokaryotes.	41
ii.	Geosmin producers occupy terrestrial, freshwater, and marine ecological niches.....	42
iii.	The phylogeny of GS is distinct of its host genome.....	43
iv.	An uncharacterized cyclic nucleotide-binding protein gene is associated with terpene biosynthesis	43
v.	Several genes co-occur with GS.....	45
vi.	Base-pair and evolutionary analysis of GS.....	46
i.	Geosmin synthase isn't associated with regions of genome plasticity	46

ii.	Geosmin Synthase is Conserved.....	47
vii.	Geosmin synthase is laterally transferred between unrelated genomes.	48
3.	Conclusions	49
4.	Materials and Methods	50
a.	Strains and cultivation	50
i.	Bacteria.....	50
ii.	Amoebae.....	50
iii.	Nematodes, <i>Caenorhabditis elegans</i>	50
b.	Materials	51
c.	Growth phase monitoring experiment	51
d.	Geosmin extraction and GC-MS quantitation	51
e.	Geosmin MIC determination.....	52
f.	CD analysis.....	52
g.	Protease activity assay	53
h.	Violacein assay	53
i.	Using 96-well plates	53
ii.	Violacein agar broth dilutions	53
i.	Predation assays using amoeba cells.....	54
i.	Killing assay in liquid media.....	54
ii.	Killing assay on solid media.....	54
iii.	Swarm assay on solid media.....	54
j.	Chemotaxis assay with <i>C. elegans</i>	55
k.	Behavioural assay with <i>C. elegans</i>	55
l.	Predation assay with <i>C. elegans</i>	55
m.	Terpene add-in assay.	56
n.	<i>C. elegans</i> lethality assay.....	56
o.	Synthesis of 2-MIB.....	56
p.	Statistical analysis.....	56
q.	Bioinformatics analysis of GS.....	57
i.	Bacterial genome selection.....	57
ii.	Phylogenetic analyses of 16S rRNA and GS.....	57

iii.	Genomic analysis and general features	57
iv.	Analysis of GS co-occurrence	57
v.	Evolution selection and codon bias analyses of GS	58
	References	59
	Appendix: Supporting Information.....	71
	List of Supporting Figures and Tables	71
	List of Additional Files and Videos	72
	Supplementary Figures and Tables.....	73

List of Figures

Figure 1: Structure of geosmin and its enantiomers, (-)-geosmin and (+)-geosmin.....	1
Figure 2: MVA pathway for the formation of IPP and DMAPP from acetyl-CoA.....	2
Figure 3: MEP pathway for the formation of IPP and DMAPP from pyruvate and G3P.	3
Figure 5: Mechanism of the formation of FPP from IPP and DMAPP.	5
Figure 6: Mechanism of the formation of geosmin with germacradienol as intermediate.	6
Figure 7: Whole genome and GS phylogenetic analysis of 93 <i>Streptomyces</i> spp.....	7
Figure 9: Effect of light intensity and colour on geosmin and carotenoids.....	9
Figure 10: Geosmin is a repellent to <i>Drosophila melanogaster</i>	10
Figure 11: <i>Folsomia candida</i> springtails detect geosmin and feed on <i>S. coelicolor</i> spores.....	11
Figure 12: Geosmin attracts ants and mosquitos.....	13
Figure 13: <i>Streptomyces</i> life cycle from in the soil.....	14
Figure 14: Geosmin production and growth stages of <i>Streptomyces sampsonii</i> when fermented with <i>Saccharomycopsis fibuligera</i> in potato dextrose medium.....	14
Figure 15: Myxobacteria's vegetative and developmental cycles.....	15
Figure 16: Simplified schematic of cyanobacterial life cycle dependent on season.....	16
Figure 17: Growth (filaments/mL) and geosmin production curve of <i>A. circinalis</i> over 7 days..	17
Figure 18: Simplified fungal life cycle from spore to mushroom.....	17
Figure 19: Geosmin quantitation and kinetics of its production.....	18
Figure 20: Symbiosis of beewolf wasp and <i>Streptomyces philanthi</i> bacteria.....	19
Figure 21: The soil food web, organisms and their interactions as preys/predators.....	21
Figure 22: Schematic of phagocytosis employed by amoeba to engulf bacterial preys.....	21
Figure 23: Adult <i>C. elegans</i> on white background.....	22
Figure 25: MIC determination of (±)-geosmin against 4 bacterial strains.....	24
Figure 26: Growth curve and geosmin production of <i>M. xanthus</i> DK1622 in 1% CTT.	25
Figure 27: CD curves of thermal denaturation of BSA with and without SDS.....	26
Figure 28: CD analyses of the effect of (-)-geosmin on BSA at three constant temperatures.....	27
Figure 29: CD analysis of the effect of (-)-geosmin on BSA at varying temperature.....	28
Figure 30: Schematic of cleavage of FTC-Casein by a protease.....	28
Figure 31: Effect of (±)-geosmin on protease activity using FTC-casein as a fluorescent marker.	29
Figure 32: Violacein inhibition assay of the exposure of VOCs to <i>C. violaceum</i>	29
Figure 33: Violacein assay using agar broth dilutions.....	30
Figure 34: Violacein assay using agar broth dilution and (-)-geosmin.....	31
Figure 35: Liquid assay of amoeba predation on <i>E. coli</i> cells.....	32
Figure 36: Swarming assay of Amoeba <i>N. gruberi</i> ATCC 30223 into <i>E. coli</i> MG1655 lawn.....	34
Figure 37: Schematic of chemotaxis assay plate of the effect of (-)-geosmin on adult <i>C. elegans</i>	35
Figure 38: <i>C. elegans</i> viability in the presence of (-)-geosmin, 2-MIB and bleach.	36
Figure 39: Predation of <i>S. coelicolor</i> and <i>M. xanthus</i> by <i>C. elegans</i>	38

Figure 40: Phylogenetic trees of GS and 16S rRNA among 109 bacterial strains.	42
Figure 41: Phylogenetic tree of Crp-Fnr subfamilies.	45
Figure 42: Analysis of RGP in <i>M. xanthus</i> DK1622.	47
Figure 43: RSCU of GS gene among 23 cyanobacterial, 25 myxobacterial and 257 actinobacterial strains.....	48

List of Tables

Table 1: Comparison of the physical and chemical properties of geosmin and water	1
Table 2: Determination of the effect of (±)-geosmin on the lag phase of <i>E. coli</i> MG1655.....	23
Table 3: MIC of (±)-geosmin against 4 bacterial strains.	24
Table 4: GC-MS results of geosmin quantitation in the cytoplasm (intracellular) and media (extracellular) of <i>M. xanthus</i> DK1622.	25
Table 5: Results of violacein assay and MIC test on 96-well plates of (±)-geosmin, methanol, vanillin, trans-cinnamaldehyde, (-)-limonene and (-)-α-pinene against <i>C. violaceum</i>	30
Table 6: Results of the effect of (-)-geosmin on <i>E. coli</i> 's survival against amoeba predation in liquid media.	32
Table 7: Results of the effect of geosmin on <i>E. coli</i> 's survival against amoeba predation on solid media.....	33
Table 8: Results of the (-)-geosmin chemotaxis assay with adult <i>C. elegans</i>	35
Table 9: Comparison of <i>C. elegans</i> movement on (-)-geosmin (0.54 µg/mL) and non-(-)-geosmin NGM plates using Imaris (Track line) and WormLab (Peristaltic speed, periodicity).....	37
Table 10: Summary of the determination and analysis of genes adjacent to CNBP in 293 actinobacterial genomes.....	44
Table 11: Co-occurrence determination of GS and potential associated genes in 5614 genomes using an orthologue-based approach.....	46

List of Abbreviations

AS	: Amoeba Saline
BLAST	: Basic Local Alignment Search Tool
BSA	: Bovine Serum Albumin
CD	: Circular Dichroism
CFU	: Colony Forming Unit
CNBP	: Cyclic Nucleotide-Binding Protein
Crp	: Cyclic AMP Receptor Protein
CTP	: Cytidine Triphosphate
CTT	: Casitone-Tris
DMA-CoA	: Dimethylacrylyl-CoA
DMAPP	: Dimethylallyl Pyrophosphate
dNS	: Number of Non-Synonymous Substitutions
dS	: Number of Synonymous Substitutions
DXP	: 1-Deoxy-D-Xylulose-5-Phosphate
EtOAc	: Ethyl Acetate
Fnr	: Fumarate Nitrate Reductase Regulator
FPP	: Farnesyl Pyrophosphate
FTC-casein	: Fluorescein Thiocarbamoyl-Casein
GC-MS	: Gas Chromatography – Mass Spectrometry
GPP	: Geranyl Pyrophosphate
GS	: Geosmin Synthase
HGT	: Horizontal Gene Transfer
HMBPP	: 4-Hydroxy-3-Methyl-But-2-Enyl Pyrophosphate
HMG-CoA	: 3-Hydroxy-3-Methyl-Glutaryl-Coenzyme A
Hr(s)	: hour(s)
IPP	: Isopentenyl Pyrophosphate
IV-CoA	: Isovaleryl-CoA
LB	: Luria-Bertani
MEcPP	: 2-C-Methyl-D-Erythritol-2,4-Cyclodiphosphate
MeOH	: Methanol
MEP	: Methylerythritol Phosphate
MG-CoA	: 3-Methylglutaconyl-CoA
2-MIB	: 2-Methylisoborneol
MIC	: Minimal Inhibitory Concentration
MVA	: Mevalonate
NCBI	: National Center for Biotechnology Information
NGM	: Nematode Growth Media
NnrR	: Nitrite and Nitric Oxide Reductase Regulator

OSN	: Olfactory Sensory Neuron
PBS	: Phosphate Buffered Saline
PPi	: Pyrophosphate
QS	: Quorum Sensing
RGP	: Regions of Genome Plasticity
RSCU	: Relative Synonymous Codon Usage
SDS	: Sodium Dodecyl Sulfate
SEM	: Scanning Electron Microscopy
TBS	: Tris-Buffered Saline
TPCK	: n-Tosyl-L-Phenylalanine Chloromethyl Ketone
TSB	: Tryptic Soy Broth
TSS	: Trypsin Stock Solution
VOC	: Volatile Organic Compound
WT	: Wild Type

1. Introduction

a. Discovery, characterization, and detection of geosmin.

i. Discovery and producers of geosmin

Geosmin is a volatile sesquiterpene produced by a diversity of micro-organisms including actinobacteria, myxobacteria, cyanobacteria, fungi (principally *Penicillium*) and red beets (*Beta vulgaris* L.)¹⁻³. Ubiquitous in terrestrial and many aquatic environments, geosmin is responsible for the distinctive smell of earth after the rain^{1,4}. Detected by humans at picomolar concentrations^{1,5}, the odour of geosmin was first described in the 19th century by Berthelot and André⁶, then purified from actinomycetes in 1965 by Gerber and Lechevalier⁷. Geosmin has two enantiomers (+) and (-)-geosmin (Figure 1), the latter of which is naturally produced and can be detected by humans at concentrations as low as 5 ppt⁵.

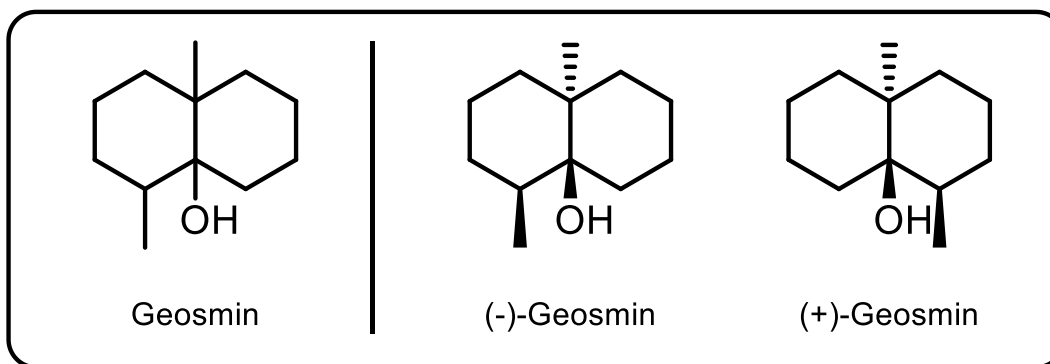


Figure 1: Structure of geosmin and its enantiomers, (-)-geosmin and (+)-geosmin.

ii. Physical and chemical properties of geosmin

Geosmin has a boiling point of 270°C, a vapor pressure of 1.2 mmHg at 25°C and an enthalpy of vaporization of 59.0 ± 6.0 kJ/mol⁸. Compared to water, which has a boiling point of 100°C, a vapor pressure of 23.8 mm at 25°C and an enthalpy of vaporization of 40.65 kJ/mol⁹, geosmin is around 95 times less volatile in terms of vapor pressure and 40 times less volatile in terms of enthalpy of vaporization. It also has a low miscibility in water with a solubility of 157 mg/L at 25°C and a logP of 3.3⁸ (Table 1).

Table 1: Comparison of the physical and chemical properties of geosmin and water^{8,9}.

Properties	Geosmin ⁸	Water ⁹
Boiling point (°C)	270	100
Vapor pressure (mm Hg, 25°C)	1.2	23.8
Enthalpy of vaporization (kJ/mol, 25°C)	59.0 ± 6.0	40.65
Solubility in water (mg/L, 25°C)	157	-
LogP	3.3	-0.5

b. The biosynthesis of geosmin

Geosmin is a sesquiterpene, and like all other terpenes it is biosynthesized using two fundamental terpene building blocks: isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP)¹. These precursors are produced via three different pathways depending on the producing organism: the mevalonate (MVA) pathway, the methylerythritol phosphate pathway (MEP, non-mevalonate pathway) or the leucine-dependent pathway. All three pathways permit the formation of farnesyl pyrophosphate (FPP), which will then undergo a specific cyclization step to form geosmin¹.

i. IPP production via the MVA pathway

The MVA pathway is predominantly used by myxobacteria and *Streptomyces* spp. to produce geosmin¹. Some fungi (*Penicillium* and *Aspergillus* spp.) also use the MVA pathway to produce geosmin. Production of geosmin via the MVA pathway starts with the nucleophilic addition of 2 acetyl-CoA molecules to form acetoacetyl-CoA via the enzyme thiolase (Figure 2)¹⁰. The 2nd carbon of another acetyl-CoA adds to the 3rd carbon of acetoacetyl-CoA using the enzyme 3-hydroxy-3-methyl-glutaryl-coenzyme A (HMG-CoA) synthase to form HMG-CoA. The latter undergoes a reduction to form mevalonate via HMG-CoA reductase. Mevalonate will then be sequentially phosphorylated by mevalonate-5-kinase and phosphomevalonate kinase to form mevalonate-5-pyrophosphate, which will then undergo a decarboxylation and hydroxide elimination to form isopentenyl pyrophosphate (IPP)¹⁰. A head to tail linkage of two IPP units will form geranyl pyrophosphate (GPP) and subsequent addition of a third IPP will form FPP, the precursor of geosmin¹¹. The mechanism of the specific cyclization step forming geosmin will be discussed later.

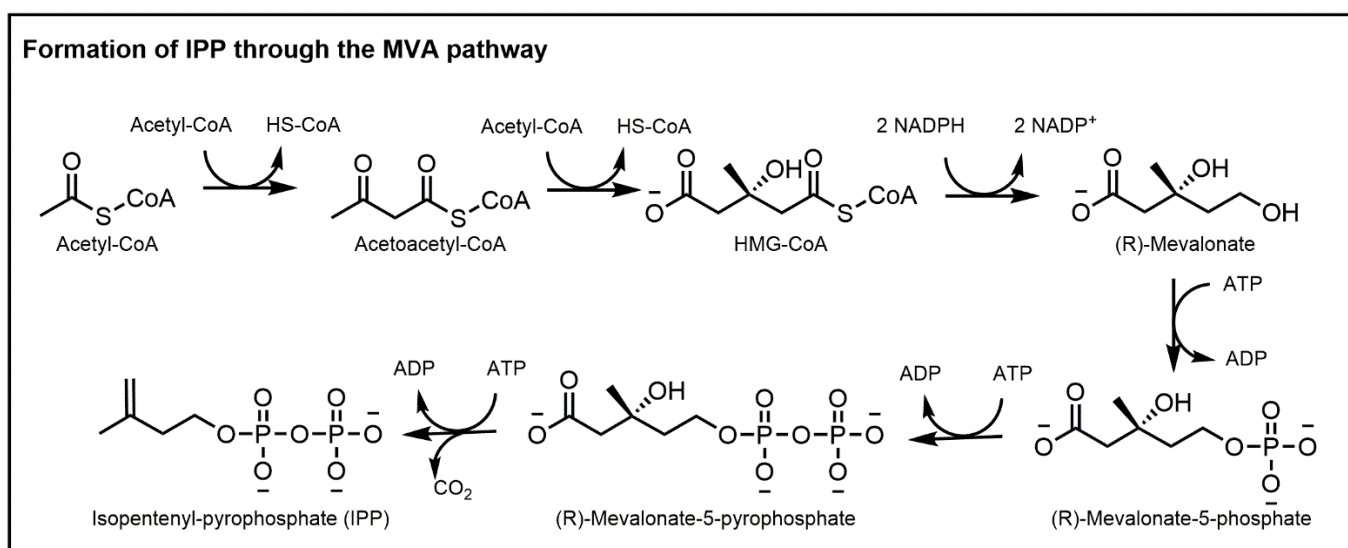


Figure 2: MVA pathway for the formation of IPP and DMAPP from acetyl-CoA.

ii. IPP production via the MEP pathway

The MEP pathway was discovered in the 1990s, and provides an alternative route to IPP in bacteria, algae and higher plants¹².

The first step consists of the condensation of pyruvate with D-glyceraldehyde-3-phosphate to produce 1-deoxy-d-xylulose-5-phosphate (DXP) and carbon dioxide. The next step is a reductive isomerization by the enzyme DXP reducto-isomerase to form MEP. The latter undergoes two phosphorylation steps, initially by cytidine triphosphate (CTP) (at position 4) then ATP (at position 2), forming 2-C-methyl-d-erythritol-2,4-cyclodiphosphate (MEcPP) through the enzyme IspF. The final steps are made by IspG and IspH permitting first a reduction to 4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HMBPP) followed by a dehydration to form IPP¹² (Figure 3), precursors to FPP for the formation of geosmin.

The MEP pathway is used by *Streptomyces* and cyanobacteria to produce terpenoids. Some strains of these micro-organisms can produce IPP through more than one pathway: most *Streptomyces* produce IPP through the MEP pathway during the exponential growth phase, and through the MVA pathway during the stationary growth phase¹.

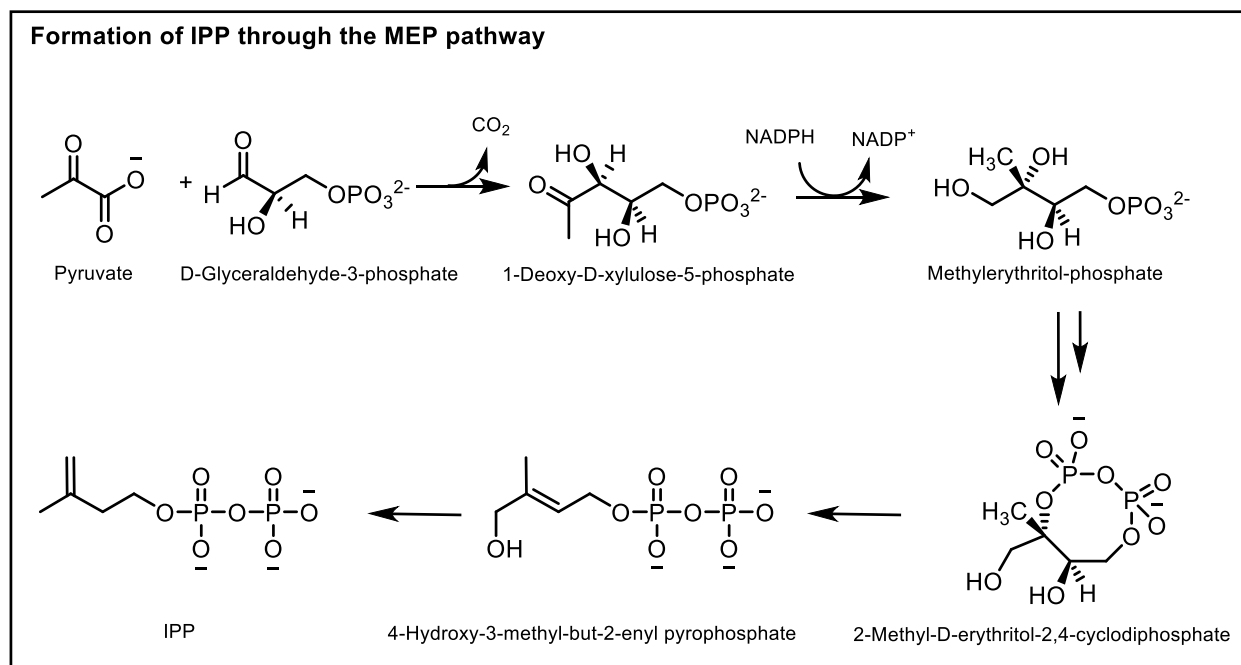


Figure 3: MEP pathway for the formation of IPP and DMAPP from pyruvate and G3P.

iii. MVA production via the leucine-dependent pathway

Mevalonate can be produced via the leucine-dependent pathway in myxobacteria, then used in the production of geosmin. This discovery was made by Dickchat et al. in 2005 using radiolabelled [$^2\text{H}_{10}$]-leucine fed to *M. xanthus*¹¹.

In the leucine dependent pathway, leucine first undergoes transamination and oxidative decarboxylation to form isovaleryl-CoA (IV-CoA). Further oxidation of IV-CoA at the C2-C3 positions forms dimethylacrylyl-CoA (DMA-CoA), which undergoes carboxylation to form 3-methylglutaconyl-CoA (MG-CoA) followed by hydration to form HMG-CoA (Figure 4). The latter is an important intermediate seen in the MVA pathway, which will undergo a reduction to form MVA. As in Figure 2, the MVA undergoes two phosphorylation steps, a decarboxylation step, and a dehydration step to produce IPP and DMAPP. These universal terpene precursors can then permit the formation of FPP, which will then allow the biosynthesis of geosmin¹¹.

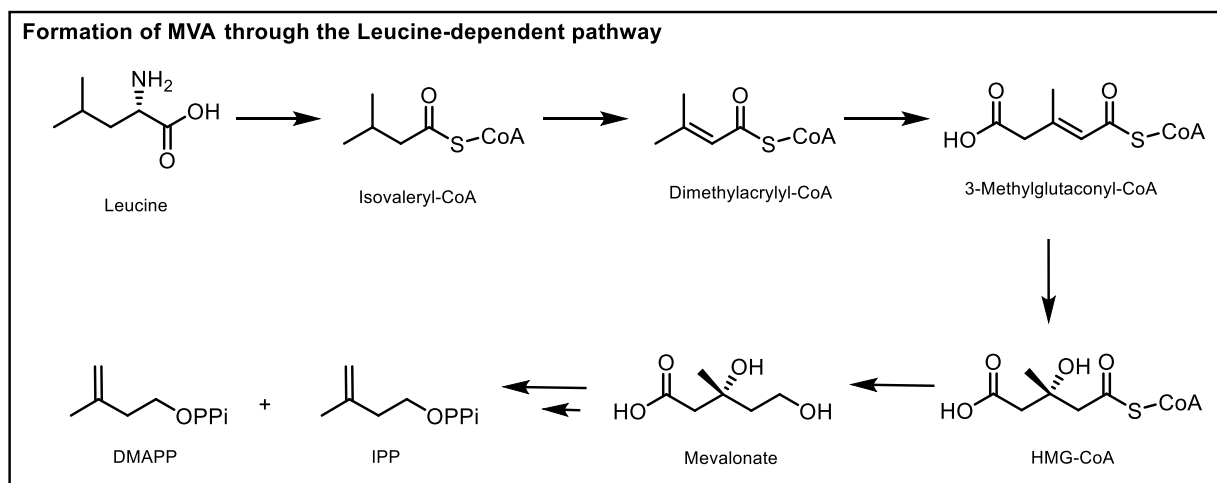


Figure 4: IPP biosynthesis through a leucine-dependent pathway.

iv. Production of geosmin from FPP

The MVA, MEP and leucine-dependent pathways all permit the production of IPP and DMAPP, the essential and universal building blocks for terpene biosynthesis. As shown in Figure 5, DMAPP and IPP merge through a head to tail condensation reaction with pyrophosphate (PP_i) as a leaving group, resulting in the formation of GPP through the enzyme GPP synthase¹³. Addition of another IPP molecule by FPP synthase leads to the formation of the sesquiterpene FPP. (Figure 5)^{11,14}.

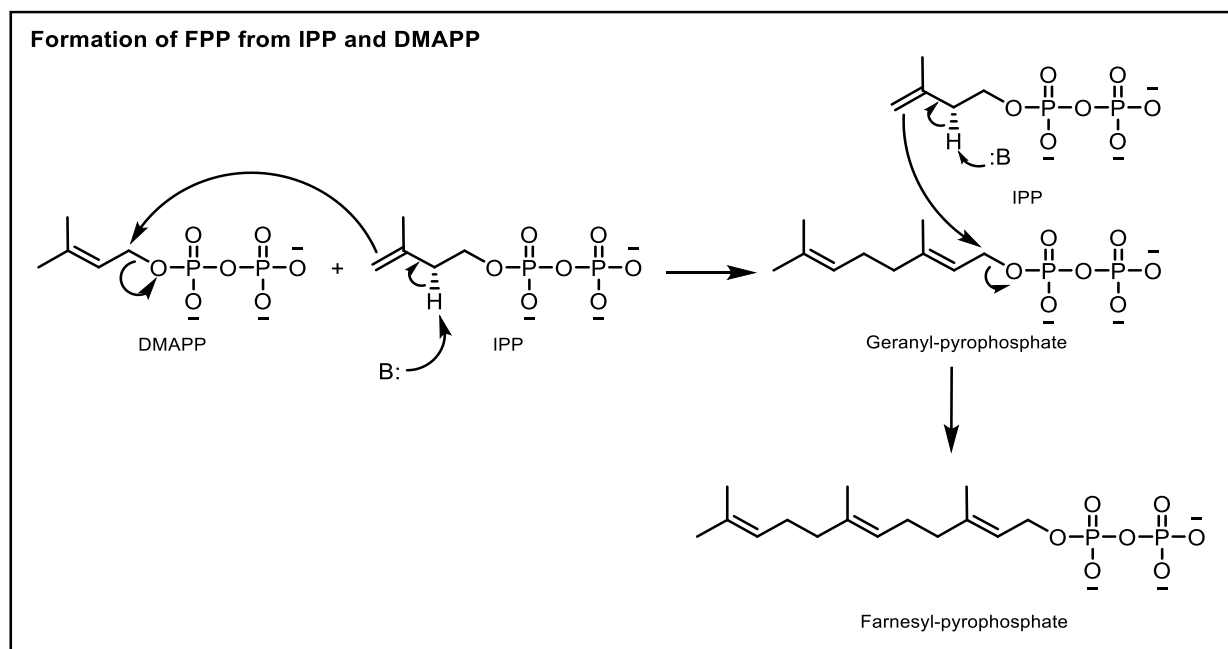


Figure 5: Mechanism of the formation of FPP from IPP and DMAPP.

The mechanism of the cyclization of FPP to geosmin was previously hypothesized to be catalyzed by multiple enzymes conducting several redox reactions. However, in 2006, Jiang, J. et al. reported that this cyclization step is specifically conducted by a single enzyme, a terpene synthase, without the intervention of any co-factors¹⁵. This enzyme was later named geosmin synthase (GS)¹⁶.

The formation of geosmin from FPP by GS is made in two fundamental steps: a proton-catalyzed cyclization of FPP to generate germacradienol, followed by a retro-Prins-type fragmentation with acetone extrusion and hydration to produce geosmin¹⁶. Figure 6 shows this proposed mechanism. The first step is an intramolecular cyclization forming a bond between carbon 1 and 10, releasing OPPi as a leaving group and generating a carbocation intermediate. Step 2 is the formation of a 3-membered ring at the carbocation position. Step 3 consists of the attack of a water molecule to relieve the ring strain, forming germacradienol. The next step follows the retro-Prins-type fragmentation, releasing the 2-propanol as acetone and forming the bicyclic moiety of geosmin. Step 4 consists of a 1,2-hydride shift revealing a stable carbonation intermediate at the most substituted carbon. The last step is the addition of a hydroxyl group at the carbocation position resulting in (-)-geosmin¹⁵.

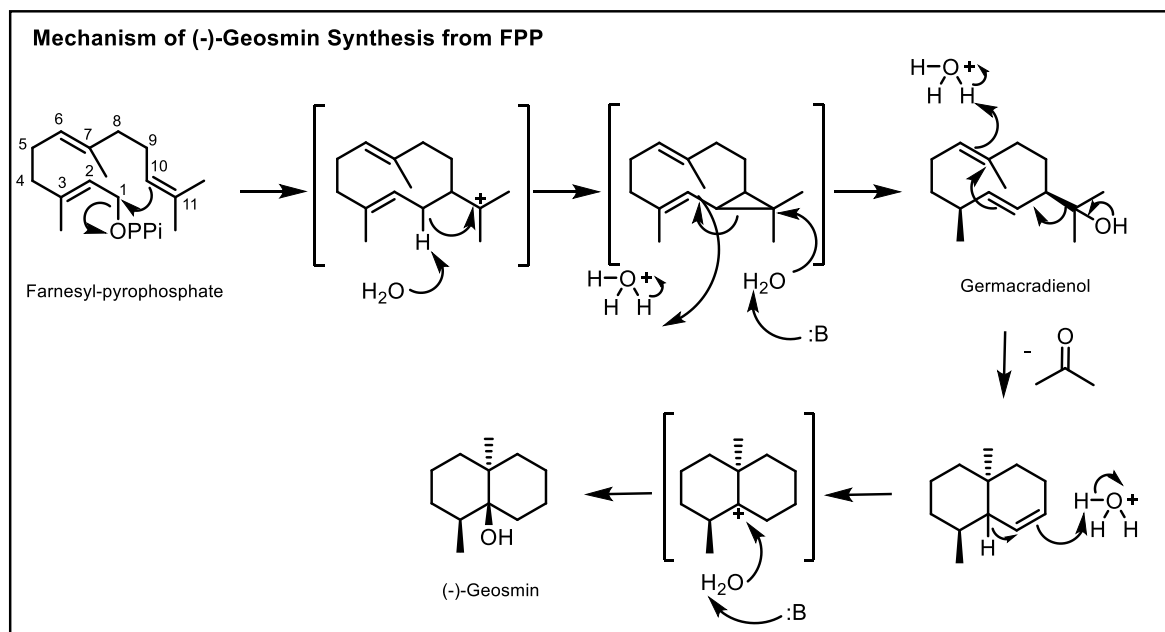


Figure 6: Mechanism of the formation of geosmin with germacradienol as intermediate.

c. The ecological role of geosmin remains to be fully elucidated.

i. Geosmin offers an unknown fitness advantage.

Geosmin was discovered and structurally characterized during the 20th century⁶, and its biosynthesis has been well studied since^{1,10–12,15,16}. However, while we know that geosmin is produced by multiple micro-organisms from all major terrestrial and aquatic environments, to this day its ecological role is not well understood^{1,17}.

The biosynthesis of secondary metabolites for bacteria and fungi is an expensive process (in terms of metabolite building blocks and energy) imposing a strong selective pressure against their production^{18,19}. The ubiquity of geosmin in its producing species, and its presence across several unrelated phyla suggest that this small molecule is an essential natural product with a specific ecological role.

ii. The gene encoding for geosmin biosynthesis is spread via horizontal gene transfer.

A study by Martín-Sánchez, L. et al. showed that among 93 *Streptomyces* strains, GSs genetic sequences were present in all strains except one (*S. pactum* KLBMP 5084), showing that geosmin production is evolutionary conserved (Figure 7)¹⁶. Whole genome-based phylogenetic analysis of *Streptomyces* spp. was generated to verify whether GS co-evolved with the *Streptomyces* spp. This analysis categorized the *Streptomyces* spp. into three different clades, colour-coded as blue, red and green (Figure 7A)¹⁶.

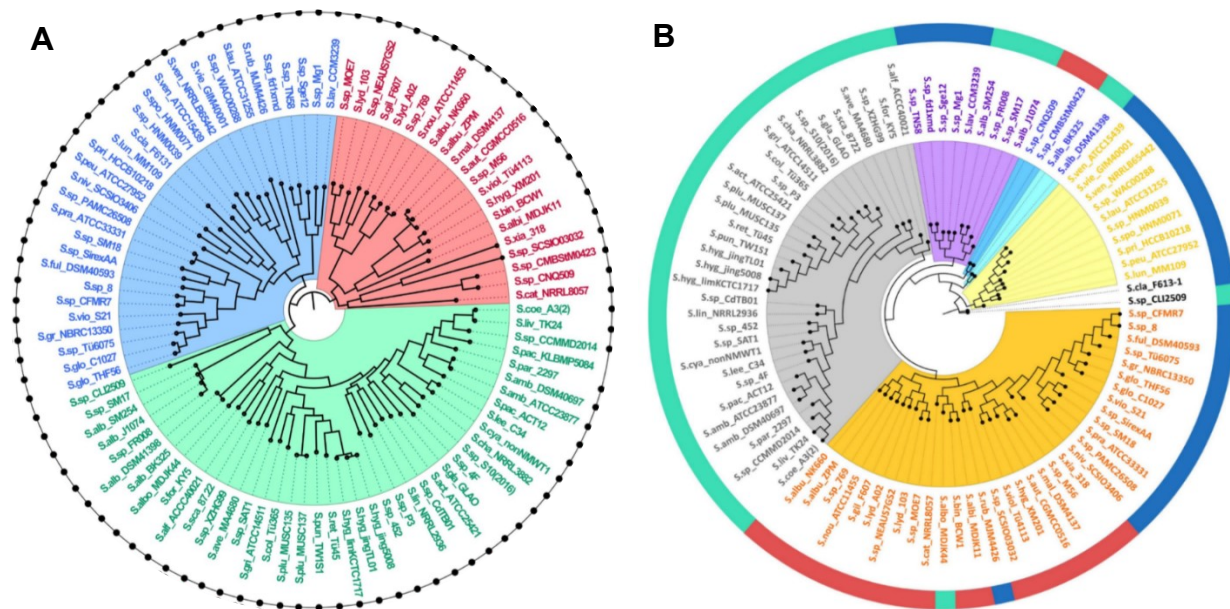


Figure 7: Whole genome and GS phylogenetic analysis of 93 *Streptomyces* spp. (Modified from Martín-Sánchez, L. et al)¹⁶.

(A) Phylogenomic analysis of *Streptomyces* spp. with the blue, red and green showing their three different clades and the black dots representing the presence of GS genes.

(B) Phylogenetic analysis of GS genes showing their distinctive clades in the inner circle; and the outer-circle colours correspond to the clades from figure 7A¹⁶.

Phylogenetic analysis of the GSs from these species showed that GSs of different evolutionary origin were interspersed within each of the streptomyces clades. For example, GSs from most of the species present in the green, blue and red clade form one clade (orange) in the phylogenetic tree of GS (Figure 7B). These results suggest the frequent transfer of GSs through horizontal gene transfer (HGT) in *Streptomyces* spp.¹⁶ and further emphasizes the potential fitness advantage conferred by geosmin, as it is conserved through evolution. These GS genes may also have been transferred to *Streptomyces* from external species.

Wang, Z. et al. conducted a phylogenetic analysis using the GS genes from actinobacteria, cyanobacteria and myxobacteria, clearly showing occurrence of HGT among the three phyla²⁰. Through their evolutionary analysis using various selection tests and a codon bias analysis, they determined cyanobacteria to be the originator of the GS gene due to the strong purifying selection and structural conservation of GS observed in these strains²⁰. However, strong conservation does not necessarily mean that cyanobacteria is the probable ancestral genetic reservoir of GS as the older the gene is the higher the chance of potential mutation rates, leading to positive selection. Nonetheless, GS is highlighted to be evolutionary important among different bacterial families.

d. Literature review on the potential function of geosmin

A review article by Patrick Fink elaborates on the potential ecological function of volatile organic compounds (VOCs). These compounds are infochemicals, which are usually in the form of terpenes⁴. Terpenoids can have semiochemical roles as short-lived chemicals, active at a very low concentration for communication purposes⁴. They could also behave as allelopathy, which are more resistant to degradation and are synthesized on a regular basis for potential protection to the producing organism⁴. Figure 8 shows potential ecological functions of VOCs: they could be involved in allelopathy against competitors, attract or repel predators, be a pheromone against con-specific, or have antibiosis properties against pathogens⁴.

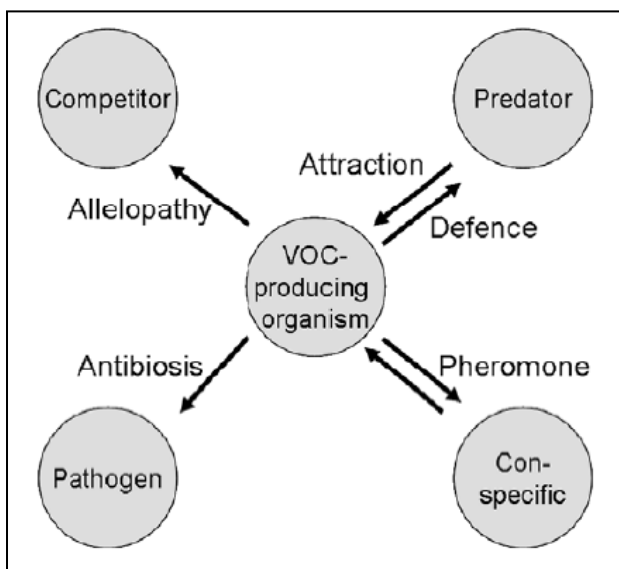


Figure 8: Schematic overview of potential ecological functions of VOCs (Fink, P.)⁴.

Geosmin has been studied previously to elucidate its ecological role, being a VOC produced by a diversity of micro-organisms¹. There is evidence of it behaving as an antibiotic against *Salmonella typhimurium*²¹, a potential side product of carotenoid synthesis²² and affecting the behavior of eukaryotes such as fruit flies²³, arthropods¹⁷, mosquitos²⁴, ants²⁵ and eels²⁶.

i. *Geosmin as an antibiotic against Salmonella typhimurium.*

A study by Diongi, C. P. et al. showed that concentrations of geosmin exceeding 18.1 ppm inhibited the growth of *Salmonella typhimurim*, a Gram-negative bacterium, specifically TA98 and TA100 strains. They conclude that these concentrations of geosmin do not reflect the ones in aquatic environments (two orders of magnitude higher), thus may not indicate the ecological role of geosmin. Furthermore, they report that the minimal inhibitory concentration (MIC) of geosmin increases to 36.2 ppm in the presence of microsomal enzymes (Rat liver S9)²¹.

However, *S. typhimurium* is not a soil or aquatic bacteria. It is found in the intestines of animals and human without the presence of geosmin producers^{1,11,16,27,28}. It is unlikely then that the ecological role of geosmin is to inhibit the growth of *S. typhimurium*.

ii. *Geosmin as a side product of carotenoid synthesis*

A study by Utkilen, H. C. et al. shows that geosmin and carotenoid synthesis are linked. They used two strains of cyanobacteria, *Oscillatoria brevis* NIVA CYA 7 and *Oscillatoria bornetti* NIVA CYA 33/1 as geosmin producers²². Their first experiment was to measure the effect of light on both geosmin and carotenoid production. Their results showed similar trends with a decreased production of both terpenes with exposure to increasing light intensity, showing that the pathway for geosmin and carotenoid synthesis could be connected (Figure 9A)²². Their production behaved similarly when exposed to different coloured lights (Figure 9B). They conclude that this parallelism in response between both terpenes suggest that geosmin could be a by-product of pigment synthesis²². However, both carotenoids and geosmin are biosynthesized through isoprenoid pathway with IPP and DMAPP as precursors^{1,10,15,29,30}. This parallelism could be observed for all terpenes as these environmental changes could affect the isoprenoid pathway. Furthermore, experimental evidence proves that geosmin is produced via a specific cyclization step¹⁶, thus it is unlikely that geosmin could be a by-product of carotenoid synthesis.

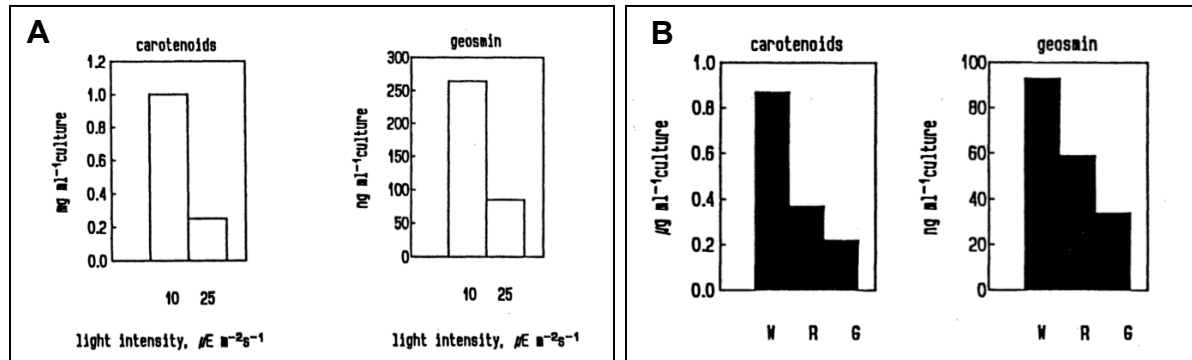


Figure 9: Effect of light intensity and colour on geosmin and carotenoids (Utkilen et al.)²².

(A) The effect of increasing light intensity on geosmin and carotenoids production by *O. bornetti*.

(B) The effect of white, red and green light on geosmin and carotenoids production by *O. brevis*.

iii. *Drosophila is repelled specifically by geosmin in moldy fruit*

A study by Stensmyr, M. C. et al. identified a functionally segregated olfactory sensory neuron (OSN) specific for geosmin detection in *Drosophila melanogaster*²³. The latter uses yeast as a food source and needs to differentiate between toxic and safe mold growing on fruit. As previously mentioned, geosmin is produced by a diversity of microorganisms including *Penicillium* spp. and *Streptomyces coelicolor*, which produce toxic mold on fruit (Figure 10A). Geosmin is seen to act as a repellent against *Drosophila*, thus indicating the presence of toxic mold²³. They first performed a T-maze assay (Figure 10B) to test the behavioral significance of geosmin on *D. melanogaster*. On one end they placed the odorant (geosmin) and on the other end a control (no odor). Geosmin elicited avoidance at a dilution of 10^{-6} (Figure 10B). Benzaldehyde is used as a repellent control and required a 1,000-fold dose increase compared to geosmin. Vinegar was used as an attractant control at a 10,000 fold dilution²³. Geosmin is also seen to

override attraction as adding it to vinegar triggers an aversive effect. This proves that geosmin is a very powerful repellent causing innate avoidance²³.

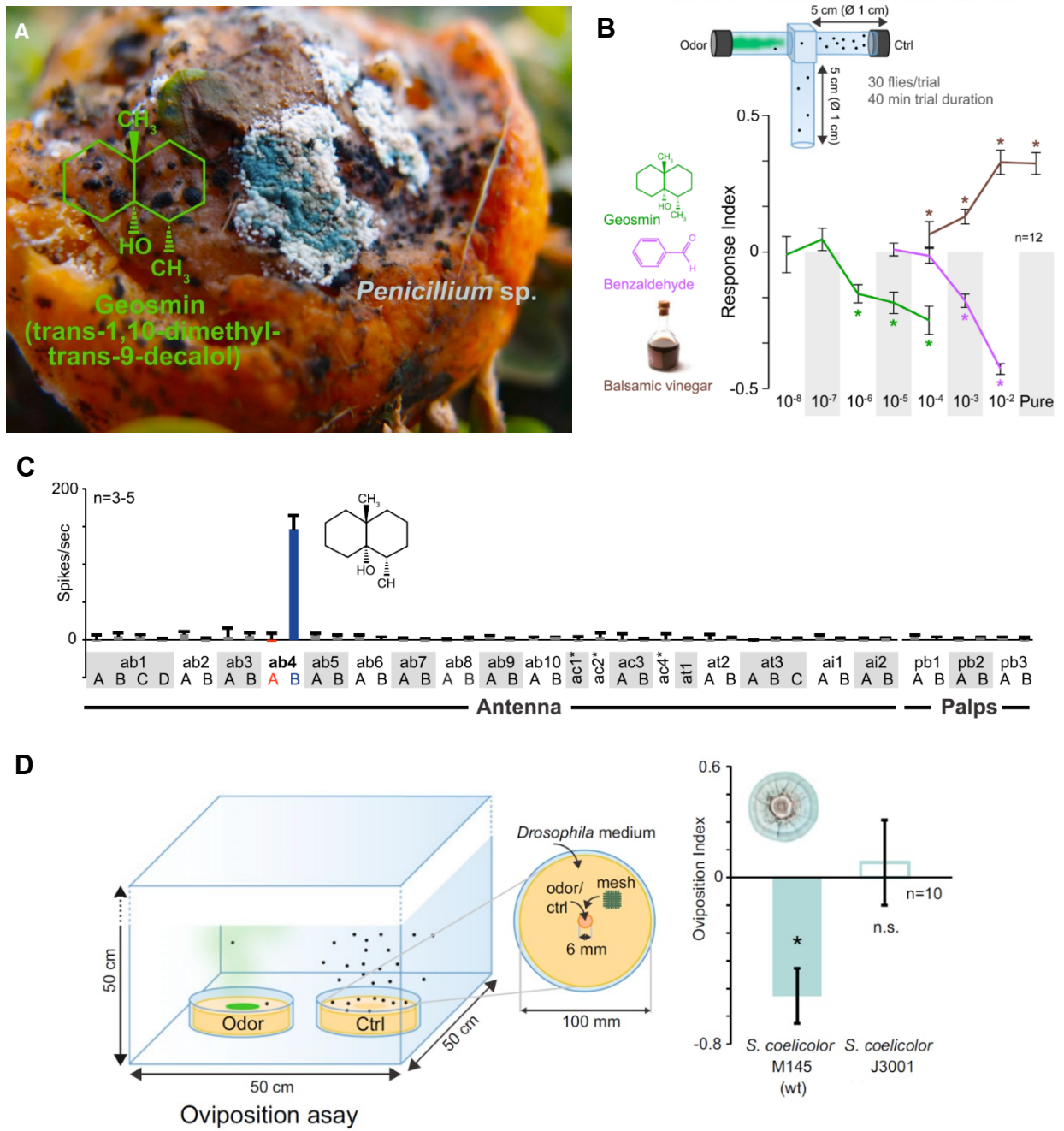


Figure 10: Geosmin is a repellent to *Drosophila melanogaster* (Stensmyr, M. C. et al.)²³.

(A) Geosmin structure and *Penicillium sp.* molds on rotten fruit.

(B) T-maze assay schematic and response index curve of *D. melanogaster* against geosmin, benzaldehyde and balsamic vinegar.

(C) Identification of ab4B OSN required for geosmin repellence in *D. melanogaster*.

(D) Oviposition choice assay showing geosmin as a specific repellent to *D. melanogaster*.

Next, they used electrophysiology to identify the OSN activated by the presence of geosmin from a total of 46 functionally distinct OSN classes in *Drosophila*. ab4B OSN is found to be necessary, sufficient and exclusive for aversion by geosmin (Figure 10C)²³. Six years later, another study by Chin, S. G. et al. demonstrated that geosmin binds to a specific receptor in *Drosophila*, Or56a receptor³¹. Furthermore, they conducted experiments to verify possible ecological and evolutionary reason for geosmin repellence in *Drosophila*. First, they confirmed that flies were not able to survive in the presence of either *Penicillium expansum* or *Streptomyces coelicolor*, due to their capacity to produce natural products with insecticidal activity. Then observed that flies also avoided laying eggs on plates containing *S. coelicolor*. Interestingly, they did not avoid a plate containing a mutant strain of *S. coelicolor* deficient in geosmin production (J3001). Abolition of geosmin production completely eliminated the repellence and caused flies to lay their eggs in the harmful plate (Figure 10D)²³. Finally, they verified that geosmin repellence is conserved in the genus *Drosophila*. Evolutionary and ecologically, flies rely on geosmin repellence to protect themselves against toxic molds. However, this may be a secondary effect of geosmin as it does not directly affect the fitness of geosmin producers²³.

iv. *Geosmin as an attractant towards arthropods.*

A study by Becher, P. G. et al. indicated that geosmin and 2-methylisoborneol (2-MIB), a monoterpene with a similar earthy odour, could aid the dispersal of *Streptomyces* spores via arthropods—specifically the springtail *Folsomia candida*¹⁷. Dispersal of spores is beneficial for bacteria as it allows them to colonize new environments. Geosmin and 2-MIB produced by *S. coelicolor* attract the springtails, which then feed on the spore-formed bacteria (Figure 11A). The spores adhere to the hydrophobic surface of the arthropods and are then released in new environments, while those that are consumed may survive the digestion process and be evacuated in faecal pellets¹⁷.



Figure 11: *Folsomia candida* springtails detect geosmin and feed on *S. coelicolor* spores (Becher, P. G. et al)¹⁷.

(A) Feeding of springtails onto *S. coelicolor* spores.

(B) SEM image of *S. coelicolor* spores adhering to springtails.

Attraction of springtails towards geosmin and 2-MIB was verified using trapping experiments using *Folsomia candida* and *S. coelicolor*. The springtail capture per trap was significantly higher with the presence of VOCs produced by *S. coelicolor*, showing attraction towards bacteria. More extensive experiments showed that the attraction is emphasized with and specific to the presence of both volatiles, geosmin and 2-MIB¹⁷. They also report that 70.8% of the faecal pellets released by springtail gave viable spores that grew on agar plates, demonstrating spores survive the animal's digestive tract. Scanning electron microscopy (SEM) confirmed the adherence of bacterial spores to the body surface of springtails, specifically at the setae (Figure 11B)¹⁷.

This study is the first to put forward an ecological role for geosmin, which is beneficial for the geosmin producing specie. It shows a symbiosis between arthropods and *Streptomyces* where the arthropods benefit from a food source and the bacteria gains dispersal to new territories¹⁷. However, this study does not address the universality of geosmin production and focuses on species that produce both geosmin and 2-MIB. Furthermore, this ecological role is limited to geosmin produced in the stationary phase of *Streptomyces*, when geosmin is also made during exponential phase in this microbe^{1,32-34}. Finally, this ecological role for geosmin is specific to soil bacteria and potentially fungi such as *Aspergillus* and *Penicillium*; but not for higher organisms such as mushrooms or for aquatic organisms such as marine myxobacteria and cyanobacteria, as currents already permit widespread dispersal of senescent cells.

v. *Geosmin as an attractant to various eukaryotes*

A study by Melo, N. et al. showed that geosmin attracts the yellow fever mosquito *Aedes aegypti* due to the installation of favourable oviposition sites by geosmin producers (Figure 12A)²⁴. The larvae and adult mosquitos detect geosmin through their antenna, thus through their olfactory system²⁴. This study does not clearly elucidate the advantage for mosquitos to lay their eggs in geosmin rich environments, but it could be possible that they detect geosmin producers (the cyanobacterium, *Kamptonema* PCC6506) as potential preys as they could have obtained resistance to the anti-eukaryotic compounds biosynthesized by the geosmin producing bacteria²⁴. This study however puts forward a cheap and reliable mean for mosquito control in developing countries by using the peel of beetroots, which contains geosmin to trap mosquitoes²⁴.

Another study by Huang, H. et al. determined geosmin to be an attractant to *Solenopsis invicta* (invasive ants) as geosmin producers offer protection through production of anti-fungal compounds²⁵. Favored nesting sites showed presence of actinobacteria, which inhibits growth of fungal pathogens such as *Beauveria sp.*, *Metarhizium sp.* and *Aspergillus sp.*, which are detrimental to *S. invicta* (Figure 12B)²⁵. The ant's survival rates dramatically increased with the presence of actinobacteria thus showing a clear advantage for their attractancy to geosmin producers (Figure 12C)²⁵. However, the advantage for geosmin producers was not elucidated in this study.

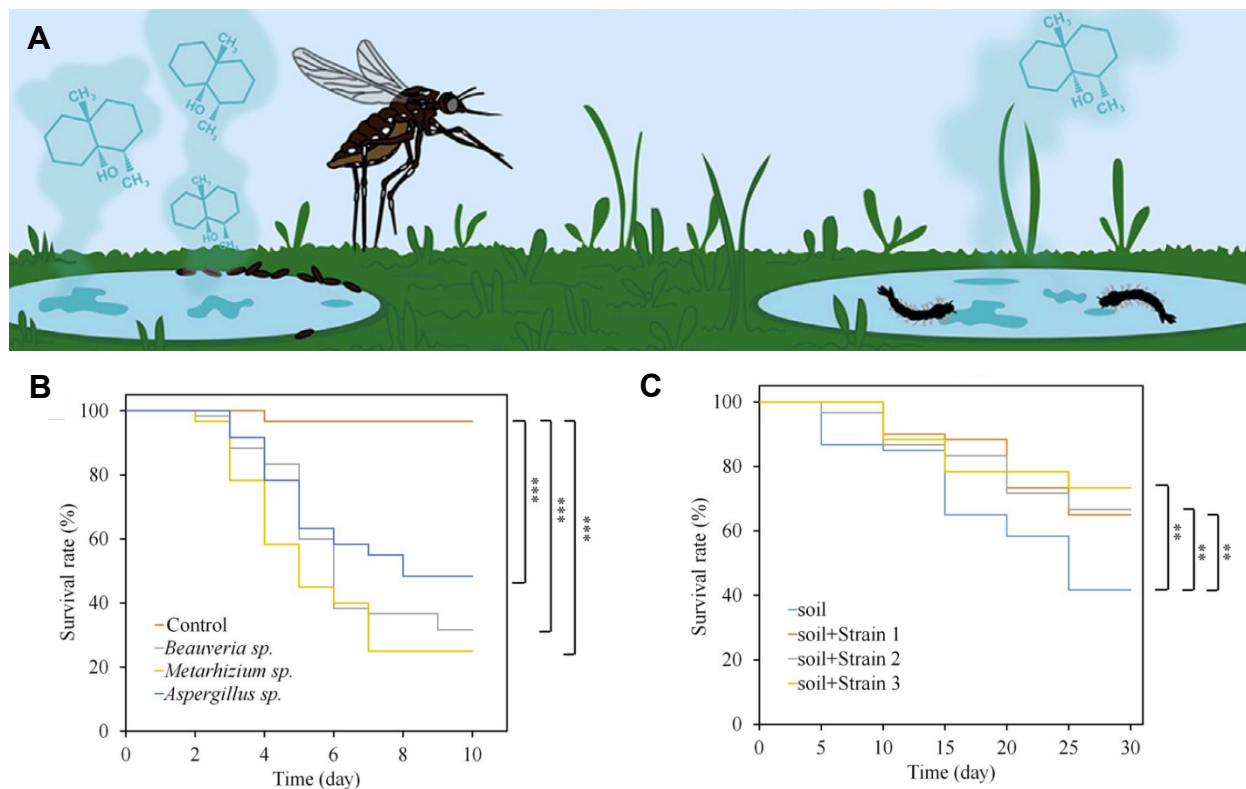


Figure 12: Geosmin attracts ants and mosquitoes^{24,25}.

(A) Graphical abstract of Melo, N. et al. showing that geosmin attracts mosquitos *Aedes aegypti* to favourable oviposition sites²⁴.

(B) Kaplan-Meier plot of survival rate of *S. invicta* when exposed to fungal pathogens over 10 days by Huang H. et al²⁵.

(C) Kaplan-Meier plot of survival rate of *S. invicta* when exposed to fungal pathogens with the presence of three actinobacteria strains over 10 days by Huang H. et al²⁵.

A study conducted in 1993 showed that geosmin is a water inland marker guiding glass eels *Anguilla Anguilla*, (L.) towards freshwater²⁶. It is confirmed that eukaryotes are sensitive for geosmin detection: *Folsomia candida*¹⁷, *Drosophila*²³, *Solenopsis invicta*, *Aedes aegypti*, *Anguilla Anguilla* (L.) all detect geosmin as either an attractant or repellent. However, the importance of this universal interaction for the geosmin producers remains to be explained.

e. The life cycle of geosmin producers and their ecology

Geosmin is produced at different growth stages: the lag phase, the exponential phase or the stationary phase, depending on the producing organism. *Streptomyces*, myxobacteria, cyanobacteria and fungi all have different challenges and stressors during their life cycles and could potentially require the use of geosmin at different points. By studying when each organism produces geosmin we may be able to discern the ecological role of geosmin as a possible tool used to respond to a particular stressor.

i. Growth stage at which geosmin production is activated.

1. Actinobacteria (*Streptomyces* spp. as representative genus)

The life cycle of *Streptomyces* species debuts with a germinating spore in the soil, which branches and forms vegetative mycelium. The vegetative supports the later growth of an aerial hyphae, both by supplying nutrients and the production of secondary metabolites like antibiotics. Once mature, the aerial hyphae segments into a chain of spores (Figure 13)³⁵.

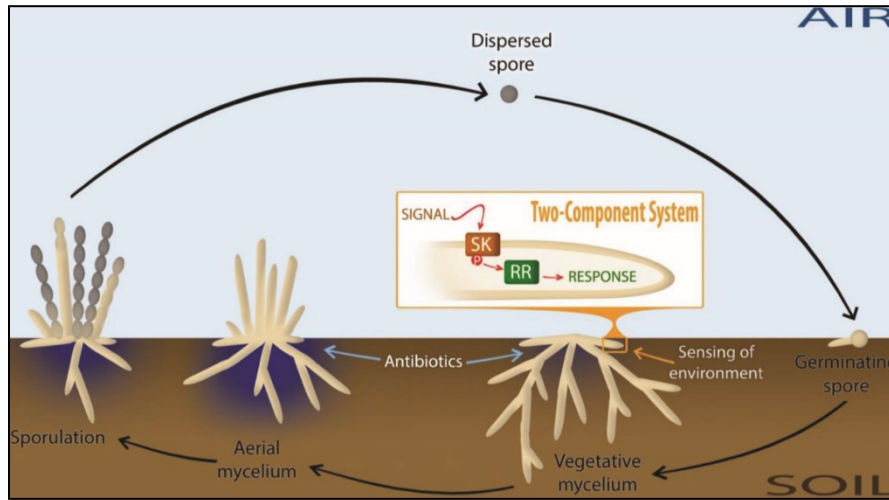


Figure 13: *Streptomyces* life cycle from in the soil (Urem, M. et al.)³⁵.

The duration of each part of the *Streptomyces* growth cycle depends on the presence of specific nutrients, temperature, humidity and pH³⁵. Under favorable conditions, with glucose as a carbon source, the lag phase lasts approximately 0 to 1 day (Figure 14), the exponential phase spans days 1-5 (Figure 14, I-II), the stationary phase days 5-8 (Figure 14, II-III) and the death phase occurs after day 8 (Figure 14)³².

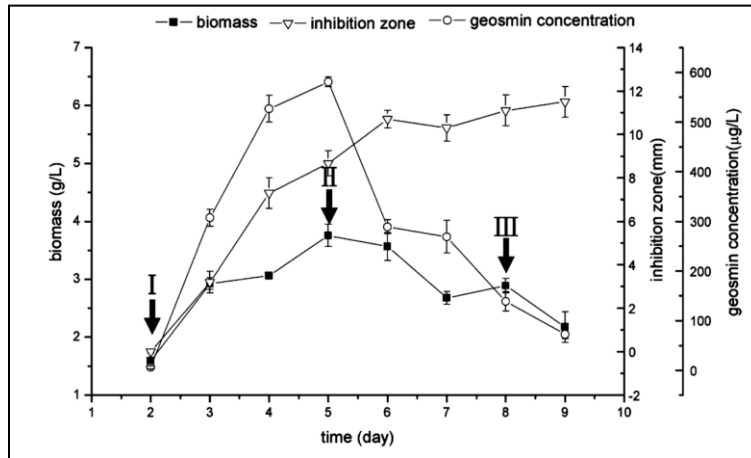


Figure 14: Geosmin production and growth stages of *Streptomyces sampsonii* when fermented with *Saccharomyces fibuligera* in potato dextrose medium (taken from Du, H. et al.)³².

As previously mentioned, *Streptomyces* spp. biosynthesize geosmin through the MEP pathway during exponential phase and through the MVA pathway during stationary phase. For some strains such as *Streptomyces sampsonii*, under laboratory conditions, geosmin is only produced during the exponential phase with a peak concentration at day 5 of 600 $\mu\text{g/L}$ ³². Its concentration is reduced to 80 $\mu\text{g/L}$ by day 9 (Figure 14)³², as it likely degrades upon acidification of the medium³⁶. It can be then hypothesized that *Streptomyces sampsonii* produces geosmin via the MEP pathway as geosmin is predominantly produced during the exponential phase.

Some streptomyces strains, such as *S. aeriouvifer*, use both the MEP and MVA pathways to produce IPP depending on the growth stage leading to the hypothesis that geosmin is produced via both pathways for this strain³⁰. Other *Streptomyces* spp. produce geosmin exclusively during the stationary phase such as C-1W and D-1W, isolated from fish ponds by Schrader, K. K. et al³⁷. Overall geosmin production in *Streptomyces* spp. is dependent on the activity of the MEP and MVA pathways, and is not limited to a particular phase of bacterial growth.

2. Myxobacteria

As previously mentioned, myxobacteria, particularly *Myxococcus xanthus*, produce geosmin via the MVA pathway and the leucine-dependent pathway¹. The life cycle of the bacterium *M. xanthus* follows a predatory social behaviour where cells form “wolf-packs” to swarm over and eat prey bacteria (Figure 15A)³⁸. Under nutrient-poor conditions myxobacteria aggregate and form circular fruiting bodies that are resistant to desiccation and starvation. These “myxospores” can germinate when dispersed in a favorable environment (Figure 15B)³⁸.

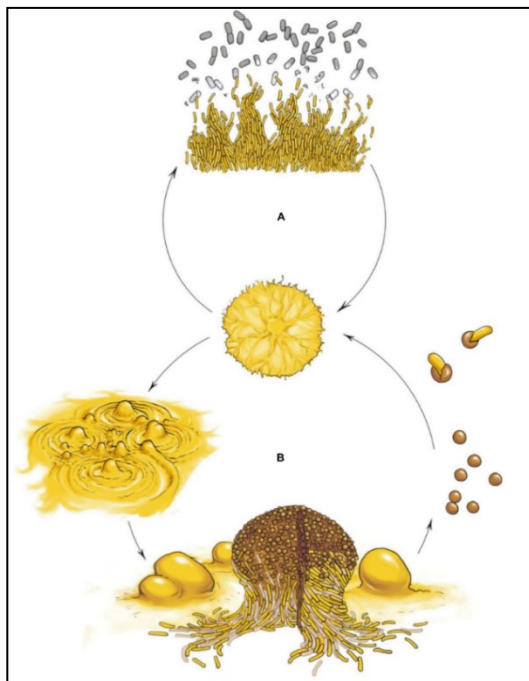


Figure 15: Myxobacteria’s vegetative and developmental cycles (Muñoz-Dorado, J. et al.)³⁸.

(A) Swarm formation of myxobacteria to predate onto bacteria in nutrient rich conditions.

(B) Formation of fruiting bodies by myxobacteria in nutrient-poor conditions.

Most secondary metabolites biosynthesized by myxobacteria, such as althiomycin and the epothilones, are produced during the exponential phase³⁹. These compounds are thought to either incapacitate prey, enhancing predatory behaviour, or ward off predatory eukaryotes like amoebae and nematodes⁴⁰. To my knowledge the growth phase-dependent production of geosmin has not been determined for myxobacteria but given the above description it is likely that geosmin is also produced during the exponential phase.

3. Cyanobacteria

At present we do not have a detailed model of the life cycle of cyanobacteria. It is dependent on many factors (presence of nitrogen, phosphorus, temperature, etc.), and consists of many different stages⁴¹. Figure 16 proposes a simplified model for cyanobacterial life cycle⁴¹.

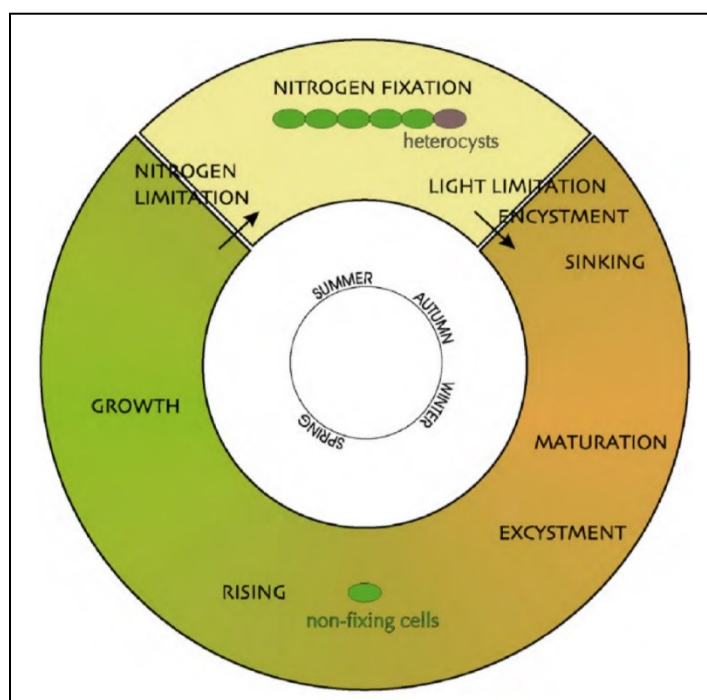


Figure 16: Simplified schematic of cyanobacterial life cycle dependent on season (Hense, I. et al.)⁴¹.

Cyanobacteria uses the MEP pathway to produce terpenoids like geosmin as determined by studies with *Anabaena circinalis*, a geosmin-producing cyanobacteria^{29,33}. From figure 17, the lag phase is from day 0-2, the exponential phase is from day 2-4, the stationary phase is from day 4-7 and the death phase is beyond day 7³³. Geosmin is produced from the lag phase to the end of the exponential phase for *A. circinalis*, suggesting that geosmin is most beneficial for young cells. Geosmin concentrations greatly decrease at the onset of stationary phase (day 4-5)³³.

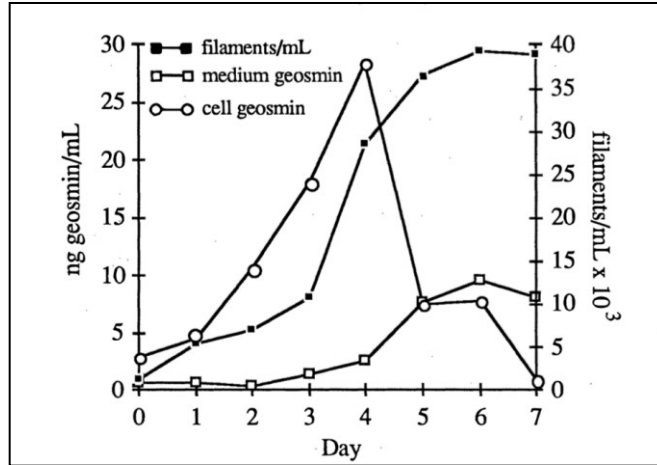


Figure 17: Growth (filaments/mL) and geosmin production curve of *A. circinalis* over 7 days (Rosen, B. H. et al.)³³. It is a blue-green filamentous cyanobacterium where cell densities are obtained as a measure of filaments/mL.

4. Fungi

The general life cycle of fungi starts with a spore that will germinate in the soil to produce hyphae. The latter will feed and grow in the soil to then emerge and produce a young mushroom. The stalk and cap expands to produce a grown mushroom, which will then release spores in the air to restart the cycle (Figure 18)⁴².

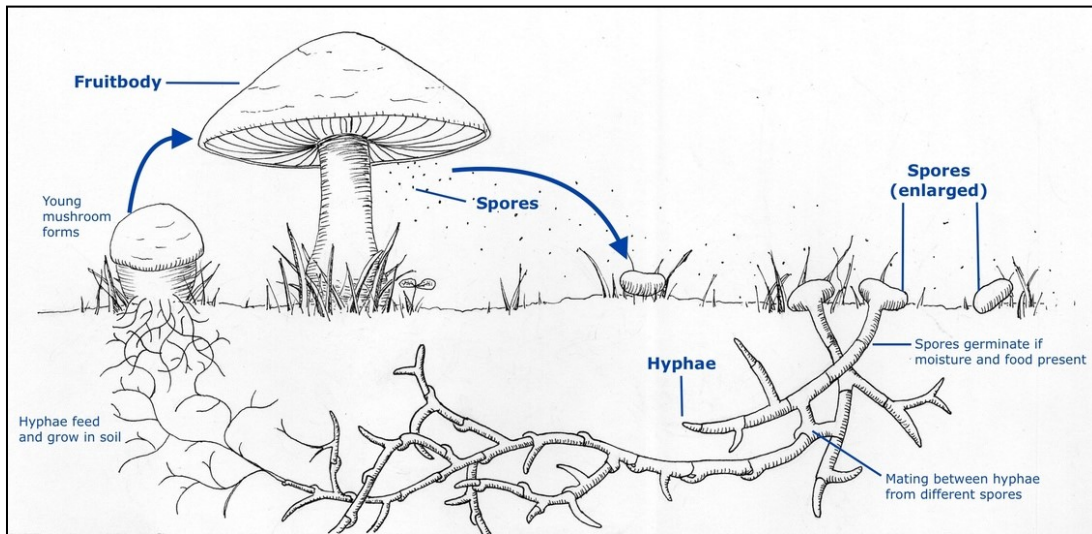


Figure 18: Simplified fungal life cycle from spore to mushroom (taken from Manaaki Whenua)⁴².

Fungi are eukaryotes and produce sterols and terpenoids via the MVA pathway¹. The growth curve of *Penicillium* follows an S shaped curve as with the organisms mentioned previously (sections e.i.1-3). Some strains of fungi produce geosmin, in particular *Penicillium expansum*, *Aspergillus* sp. and larger mushroom producers such as *Curtinarius herculeus* and

Cystoderma amianthinum^{1,43}. However, the biosynthesis of geosmin in fungi has not been extensively studied. *P. expansum* is the most studied fungal geosmin producer.

A study by Mattheis, J. P. et al. collected, in headspace samples, the geosmin produced by *P. expansum* RR89-30 over 4 days. GC-MS analysis of these samples showed varying results with higher geosmin production (1.5-5 ng/L) between day 3 and 4 (Figure 19A), during the exponential/stationary phase of *Penicilium* growth⁴⁴.

Another study on the production of (-)-geosmin by *P. expansum* by La Guerche, S. et al., showed that it is enhanced by the presence of another fungus, *Botrytis cinerea*. A kinetic experiment was made in order to verify the influence of three *B. cinerea* strains, C77-4, UMRSV01M103 and SAS 56 on the quantity of geosmin produced by *P. expansum* over 10 days (Figure 19B). The exposure of *P. expansum* to different *B. cinerea* strains generates different geosmin production curves, each of which start at the onset of the exponential phase. Geosmin concentrations increased throughout the exponential phase, peaking at 35 ng/L on day 7 for the exposure of *P. expansum* to C77-4 (Figure 19B). The authors hypothesize that some activator metabolites produced by *B. cinerea* could trigger geosmin production³⁴

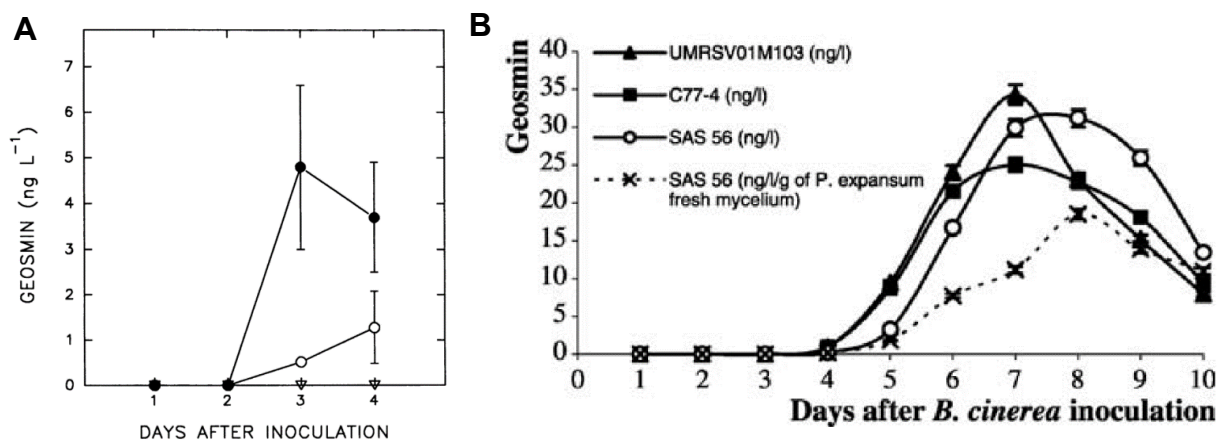


Figure 19: Geosmin quantitation and kinetics of its production^{34,44}.

(A) Geosmin quantitation using GC-MS analysis of headspace samples of *P. expansum* over 4 days. ▽, controls; ○, experiment 1; ●, experiment 2 (Mattheis, J. P. et al.)⁴⁴.

(B) Kinetics of geosmin production by *P. expansum* when exposed to *B. cinerea* SAS 56, C77-4 and UMRSV01M103 strains (La Guerche, S. et al.)³⁴

ii. Common life cycle challenges of geosmin producers

1. Symbiosis needed for growth and survival

Streptomycetes originated approximately 400 million years ago, coinciding with the development of green plants. They played a part in the formation of ancient soil due to their ability to solubilise cell walls and other components of dead insects, plants and fungi leading to the formation of cellulose and carbohydrates⁴⁵.

Most of the antibiotics, anticancer and antifungal drugs known today originate from *Streptomyces spp.*, leading to the belief that these microorganisms are potential predators and tough competitors. However, this is not always the case, as many *Streptomyces spp.* are preyed upon by other microorganisms⁴⁶. They also engage in mutualistic interactions with plants and other organisms for better growth and survival⁴⁷. Several *Streptomyces spp.* are plants endophytes: they promote plant growth and protection from pathogens through secondary metabolite production while plants offers shelter and cellulose to the symbiotic bacteria⁴⁷.

Another example of a beneficial interaction is a symbiosis between beewolf solitary digger wasps (*Philanthus spp.*) and *Candidatus Streptomyces philanthi*. The wasp paralyzes honeybees, then feeds its larvae by placing them in soil brood cells (Figure 20A). Inside the soil resides bacteria and fungi, which can kill the newborn larvae⁴⁷. For protection, the wasp releases *Streptomyces* bacteria from their antennal glands (Figure 20B) into the brood cells containing the larvae. Antibiotics produced by these microbes keep the larvae healthy, and in exchange, the wasps offer food and shelter to the *Streptomyces spp.*⁴⁷.

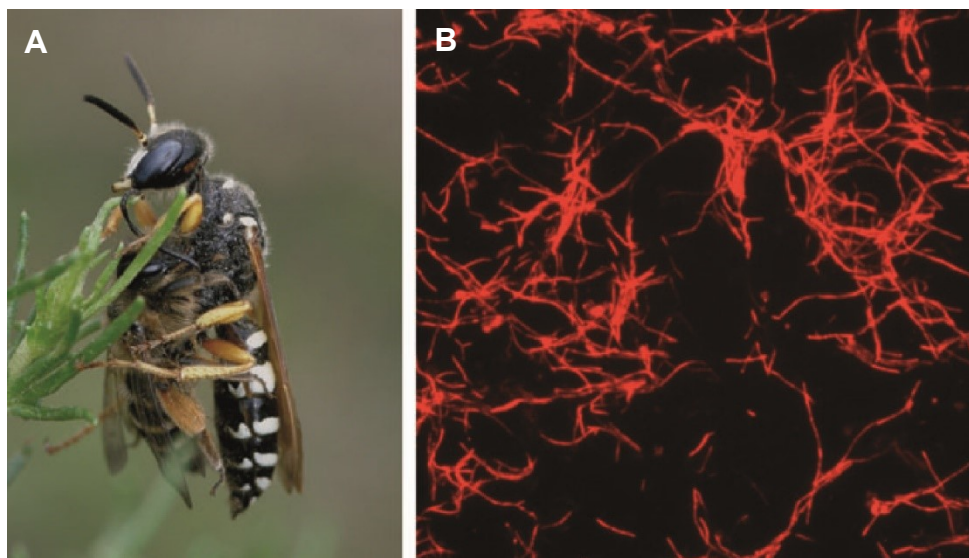


Figure 20: Symbiosis of beewolf wasp and *Streptomyces philanthi* bacteria (Kaltenpoth et al., 2005)⁴⁷.

(A) A beewolf wasp hunting a honeybee.

(B) *Candidatus Streptomyces philanthi* contained in the beewolf wasp antennal glands stained with the Cy3-labelled probe SPT177.

Some bacterial endosymbionts (supposedly actinomycetes) produce geosmin within the amoeba *Vannella*. The amoeba offers the microbes food and shelter, but the fitness advantage provided by geosmin is unknown⁴⁸.

2. Quorum sensing is important for bacterial communities

Quorum sensing (QS) is a form of bacteria communication essential for the maintenance and protection of their community. Intraspecific QS with the use of specific compounds can be used to communicate with only same-specie cells⁴⁹. QS defectors can then be eliminated thus

allowing to separate a bacterial population from cells that disrupt normal community behavior like cheater cells³⁸. However, cheater cells can also have the required receptors for QS making it not the best method for discrimination⁴⁹. QS can also be involved in interspecies communication and extend relationships with other bacteria, eukaryotes and mediate host-pathogen interactions⁵⁰.

Terpenes like pinene, limonene, vanillin and cinnamaldehyde, have been previously found to inhibit or enhance QS and disrupt bacterial communities^{51,52}. Geosmin as a sesquiterpene may be involved in QS between bacterial communities.

iii. Major stressors for geosmin producers

In the soil, micro-organisms are exposed to multiple stressors and develop secondary metabolism to response to these threats. These stressors include changes in temperature, humidity, pH, absence of nutrients, predation and even stochastic events such as storms, earthquakes, and forest fires^{46,53-55}.

When temperature increase in the soil, such as in the spring or summer, bacteria and fungi undergo early sporulation to become heat resistant⁵⁴. The rise in temperature can also accelerate germination from dormant fungi and bacteria cells. During low temperatures, micro-organisms protect themselves by inducing dormancy, which is characterized as a hardy, non-replicative state permitting resistance against unfavorable environments⁵⁵. These unfavorable environments include dryness, drastic change in pH (most micro-organisms are neutrophils), presence of bacteriostatic antibiotic or antifungals, or an infection by a pathogen⁵³⁻⁵⁵. The slow-growing micro-organism could then tolerate these harsh conditions⁵⁵. To exit dormancy, the detection of favorable conditions is needed. Some studies have shown that fragments of peptidoglycan (muropeptides) triggers germination in dormant *Bacillus subtilis* spores^{56,57}. Geosmin may be a small molecule that could accelerate growth of micro-organisms as it is seen to be produced during exponential phase.

The two main predators of geosmin producers are amoeba and nematodes (Figure 21), which are much larger and can easily engulf them^{46,58}. In response, geosmin generating micro-organisms produce amebicides and toxins to survive their predators⁵⁹⁻⁶². The stress caused by these predators is exacerbated upon precipitations as amoeba alone can be responsible for 60% of bacterial death⁶³.

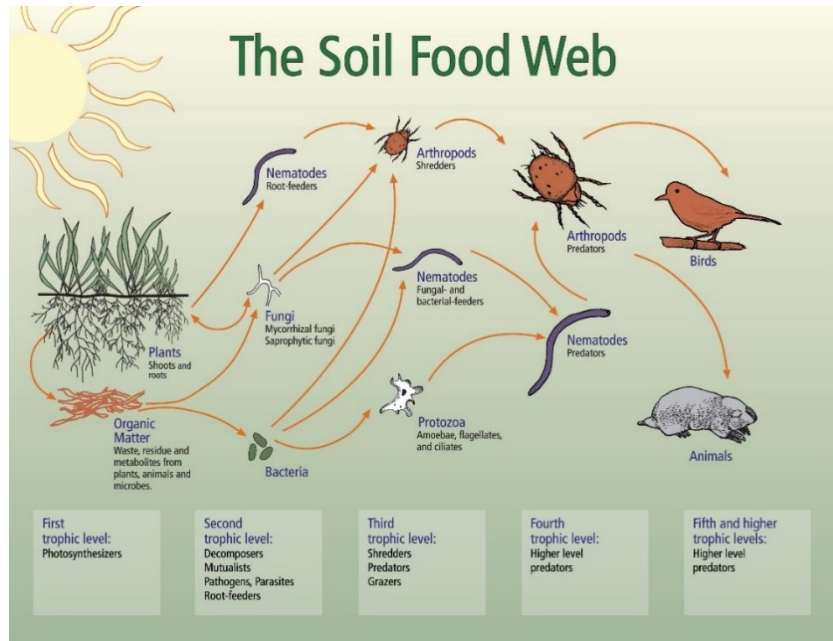


Figure 21: The soil food web, organisms and their interactions as preys/predators⁴⁶.

1. Major predators of geosmin producers
 - a. Amoeba

Amoebae are unicellular protozoan eukaryotes. They are 20-75 times bigger than bacteria and use phagocytosis to engulf their food (mostly bacteria)⁶⁴. There are four steps during phagocytosis: (1) entrapment of preys using the plasma membrane, (2) formation of a vacuole to contain the food inside the cell, (3) fusion of lysosome with vacuole and (4) lysosome digestive enzymes break down and digest the engulfed food (Figure 22)⁶⁴.

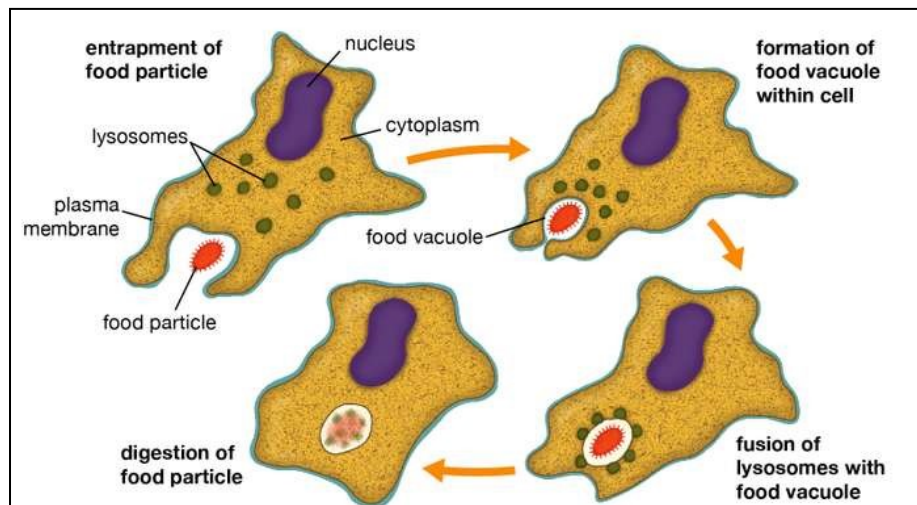


Figure 22: Schematic of phagocytosis employed by amoeba to engulf bacterial preys⁶⁴.

Protists activity have previously been seen to be affected by VOCs⁶⁵. However, this is strongly dependent on the interacting bacteria and amoeba demonstrating an involvement of

terpenes on predator-prey relations. In most cases, amoeba appears to have evolved to recognize terpenes as attractants guiding them to potential preys from a distance⁶⁵. These terpenes can also be bacteria-specific, permitting the amoeba to choose their preys. In other cases, VOCs may even repel protists, providing an ecological advantage to their producers⁶⁵.

b. Nematodes

Nematodes like the well-known species *Caenorhabditis elegans*, are soil predators, which threaten the survival of geosmin producers (Figure 23). Adult *C. elegans* are around 1-2 mm in length, approximately 1000 times bigger than bacteria⁶⁶. Nematodes readily engulf bacteria due to their significant size over their preys. However, bacteria have developed secondary metabolite as defence chemicals against nematode predation⁶⁷. Like amoeba, nematodes also have sensory neuron receptors capable of recognizing VOCs, which could indicate presence of food (isoamyl alcohol) or danger (2-butanone)⁶⁶. It is possible that geosmin acts as a repellent to these worms, preventing them from eating the geosmin producers.



Figure 23: Adult *C. elegans* on white background (picture from Utrecht University).

Nematodes also permit dispersal of bacteria and fungi: live and dormant preys can stick to the surfaces of *C. elegans* and then be released to colonise new environments⁶⁸. There is also evidence that some bacteria can survive in the digestive tract of nematodes and be released as fecal pellets⁶⁸. Symbiosis between *C. elegans* and microbes present in the worm digestive tract also occurs. The nematode provides food, shelter and potential dispersal of the bacteria, and the microbes provides protection against pathogens, better immune system responses and development and synthesis of vitamins⁶⁸. It is possible that geosmin is an attractive signal for the development of mutualistic interactions with nematodes.

Actinobacteria, myxobacteria, cyanobacteria and fungi all produce geosmin during the exponential and stationary phase, suggesting its importance during their feeding/hunting behaviour, their growth, and their transition to senescence^{22,32,34,37,69}. I first examined the effect of geosmin on the lag phase of *Escherichia coli* to determine if it permits an earlier exit from dormancy. Subsequently, I measured the effect of geosmin on the growth of various bacterial strains to rule out any potential antibiotic activity. I investigated the role of geosmin on the feeding behaviour of *M. xanthus* as a protease adjuvant and observed its effects on QS. Finally, I performed predation assays using *Amoeba* and *C. elegans* as well as bioinformatics analyses on GS to elucidate the ecological function and importance of geosmin. I propose that geosmin is a warning chemical advertising the production of toxic secondary metabolites.

2. Results and discussion

a. Effect of geosmin on the lag phase of *E. coli* cells

Upon rainfall, dormant cells stochastically germinate^{53,55}. This germination coincides with an increase in aerosolized geosmin, due to it being released into the air as rain droplets hit the ground⁷⁰. While geosmin appeared unlikely to penetrate the hard outer coat of spores, we hypothesized that it could serve as an early growth signal and set out to investigate its effect on shortening the lag phase of *E. coli* cells.

To do this *E. coli* MG1655 was incubated in liquid culture at 37°C and 225 rpm, before dilution to 0.5 McFarland standard (1.5×10^8 CFU/mL), then subsequent dilution to 1.5×10^{10} CFU/mL in minimal media, with or without (\pm)-geosmin. The concentration of (\pm)-geosmin used was 0.1 ppm, the average concentration detected from GC-MS analysis of *M. xanthus* cultures (*vide infra*). Cells were incubated for 4 hrs, to allow them to reach exponential phase, and plated onto agar. Following overnight incubation, colonies were counted. The results are shown in Table 2.

Table 2: Determination of the effect of (\pm)-geosmin on the lag phase of *E. coli* MG1655.

Conditions	*10 ⁵ CFU/mL	Average (*10 ⁵ CFU/mL)	Standard Deviation (*10 ⁵ CFU/mL)
<i>E. coli</i> + Methanol ^a	447 870 752	689	218
<i>E. coli</i> + Geosmin ^b	628 712 698	679	45

^a Methanol volume is 10 μ L.

^b Geosmin dissolved in methanol, 0.1 ppm concentration

A higher number of quantified cells would equate to an earlier access to the exponential phase, which would translate to a shortening of the lag phase. However, the average CFU/mL count for both assays was equivalent to well within standard deviation ($P=.94$, student T test)⁷¹. (\pm)-Geosmin does not appear to reduce the lag phase of *E. coli* MG1655.

b. Antibiotic properties of geosmin

As stated in section 1.d.i., geosmin was determined to inhibit the growth of the Gram-negative bacterium *S. Typhimurium* at 18.1 ppm (18.1 μ g/mL)²¹. Gram-negative bacteria differ from Gram-positives with the presence of an outer-membrane and a thinner peptidoglycan layer, making them more resilient to small molecule toxins⁷². As *S. Typhimurium* is not present in the soil and is not likely to interact with geosmin producers²⁷, I decided to expose geosmin to soil

Gram-positive bacteria and another Gram-negative bacterium to verify geosmin's antibiotic properties.

To verify this hypothesis, I exposed (\pm)-geosmin to 3 different soil Gram-positive bacteria in order to determine its MIC (Table 3, Figure 25). The 3 soil Gram-positive bacteria are *M. luteus* DSM 20030, *B. subtilis* DSM 10 and *B. thailandensis* E264. *K. aerogenes* ATCC 13048 was also exposed to (\pm)-geosmin for MIC determination as a Gram-negative bacteria control. Methanol (MeOH) was used as a control due to the (\pm)-geosmin standard being diluted in methanol.

Table 3: MIC of (\pm)-geosmin against 4 bacterial strains. This includes 3 Gram-positive bacteria (*M. luteus*, *B. subtilis*, *B. thailandensis*) and 1 Gram-negative bacteria (*K. aerogenes*).

Bacteria Strain	MIC (\pm)-Geosmin ($\mu\text{g/mL}$)	MIC Methanol (μL) ^a
<i>M. luteus</i> DSM 20030	250	50
<i>B. subtilis</i> DSM 10	250	>50
<i>B. thailandensis</i> E 264	62.5	6.25
<i>K. aerogenes</i> ATCC 13048	500	>50

^a The (\pm)-geosmin sample was dissolved in methanol at a concentration of 2 mg/mL.

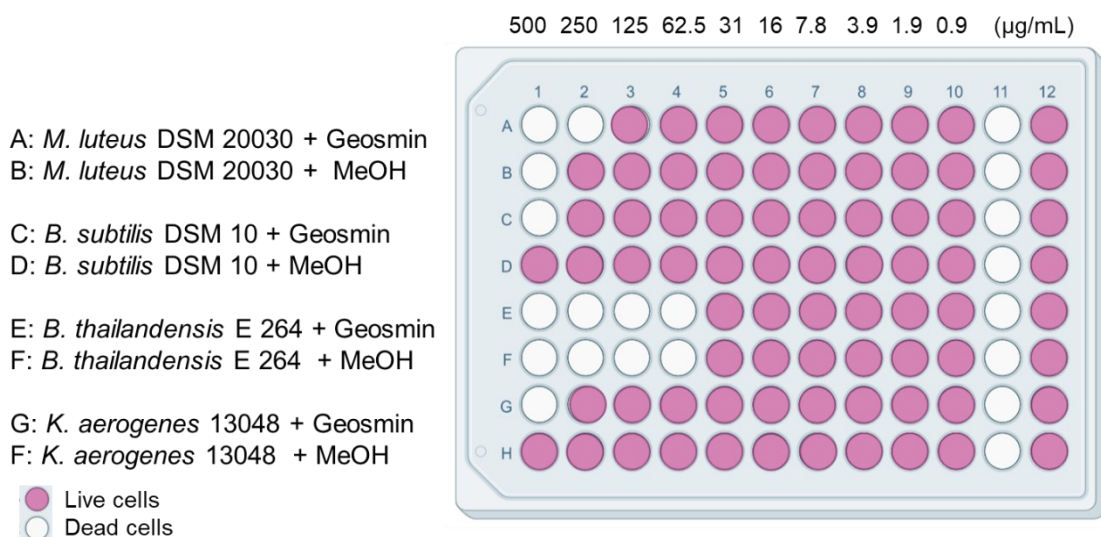


Figure 25: MIC determination of (\pm)-geosmin against 4 bacterial strains. This includes 3 Gram-positive bacteria (*M. luteus*, *B. subtilis*, *B. thailandensis*) and 1 Gram-negative bacteria (*K. aerogenes*). The pink and white circles reveal live and dead cells respectively. Column 11 is a sterile control and column 12 is a growth control.

From Table 3 and Figure 25, the MIC of (\pm)-geosmin against the Gram-positive *M. luteus* is around 250 $\mu\text{g/mL}$, for *B. subtilis* it is 500 $\mu\text{g/mL}$ and for *B. thailandensis* the MIC could not be determined due to inhibition from the methanol control. These are relatively high concentrations of (\pm)-geosmin, much higher than the literature value of 18.1 $\mu\text{g/mL}$. Additionally, the MIC for (\pm)-geosmin against a Gram-negative bacterium, *K. aerogenes* is 500 $\mu\text{g/mL}$, which is as high as the MIC value against *B. subtilis*. (\pm)-Geosmin does not seem to have an antimicrobial function.

c. Effect of geosmin on protein stabilization

i. Geosmin production by *M. xanthus*

The hypothesis that geosmin could be a protein solubilizer was generated from the observation of the growth and geosmin production curves by *M. xanthus* DK1622. The growth curve was determined using OD₆₀₀ measurements over 9 days of a culture of DK1622 in 1% CTT incubated at 30°C and 225 rpm. From Figure 26, the lag phase of DK1622 is from day 0-4, the exponential phase is from day 4-6, the stationary phase is from day 6-7, the death phase is beyond day 9. Clumping began at day 7 causing unreliable data.

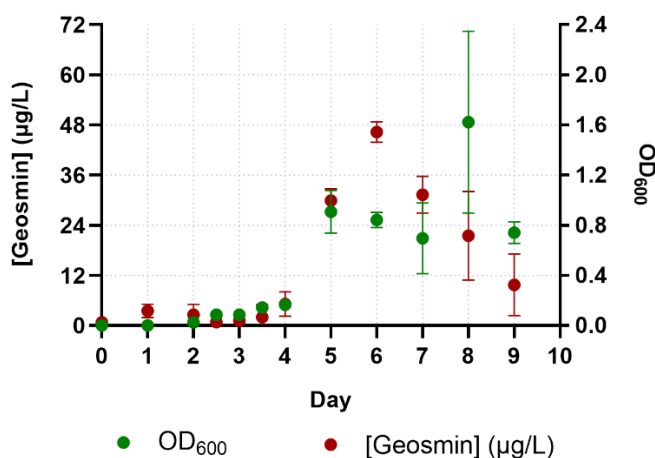


Figure 26: Growth curve and geosmin production of *M. xanthus* DK1622 in 1% CTT.

Geosmin concentrations were determined via GC-MS, with the aid of a calibration curve derived from solutions of authentic geosmin. † Cells began to clump at day 7, complicating OD₆₀₀ measurements.

Table 4: GC-MS results of geosmin quantitation in the cytoplasm (intracellular) and media (extracellular) of *M. xanthus* DK1622.

Extraction Location	[Geosmin] (ppm)	
	Day 2.5	Day 4
Intracellular ^a	0.015353	0.1151
Extracellular ^b	0.043	-
Full ^b	0.07	0.4105

^a1:4 extraction with EtOAc.

^b1:1 extraction with EtOAc.

The geosmin production curve was generated using GC-MS analysis of extracted bacterial cultures. This analysis was made with the help of Anic Imfeld, a PhD student in the Gélinas lab at Concordia University, Montreal, QC. As shown in Figure 26, geosmin production starts at the onset of the exponential phase at day 4. Further GC-MS analysis showed that most

geosmin is present in the media rather than inside the cytosol of DK1622 cells (Table 4). At day 2.5, 61.4% of the detected geosmin is present extracellularly, in the media. This could mean that the ecological role of geosmin is outside of the cells: it could potentially solubilize or stabilize extracellular proteins such as digestive enzymes, thus aiding in the feeding behavior of *M. xanthus*.

ii. *Circular dichroism analysis of the effect of geosmin on bovine serum albumin.*

Bovine serum albumin (BSA) is a hydrophobic protein mainly constituted of alpha-helices⁷³. Geosmin is a hydrophobic molecule with a single hydroxyl group, which can permit hydrophilic interactions. Geosmin could potentially coat BSA and solubilise it in aqueous environments. To address this theory, I performed circular dichroism (CD) experiment to determine if geosmin stabilizes the secondary structure of BSA during thermal denaturation.

As heat denatures BSA the molar ellipticity of the protein increases, particularly at temperatures higher than 80°C (Figure 27, A)⁷³. Sodium dodecyl sulfate (SDS) is a surfactant that has a protective effect on proteins upon thermal denaturation and was used as a positive control⁷³. SDS coats and denatures proteins through its long aliphatic tail, exposing the hydrophobic and buried amino acids in aqueous solution permitting solubilization and stabilization of the protein. A study by Moriyama et al. showed that a concentration of 0.75 mM of SDS permits retention of the helicity of BSA (Figure 27, B). This retention of the secondary structure of BSA shows a protective effect, characteristic of a protein stabilizer⁷³.

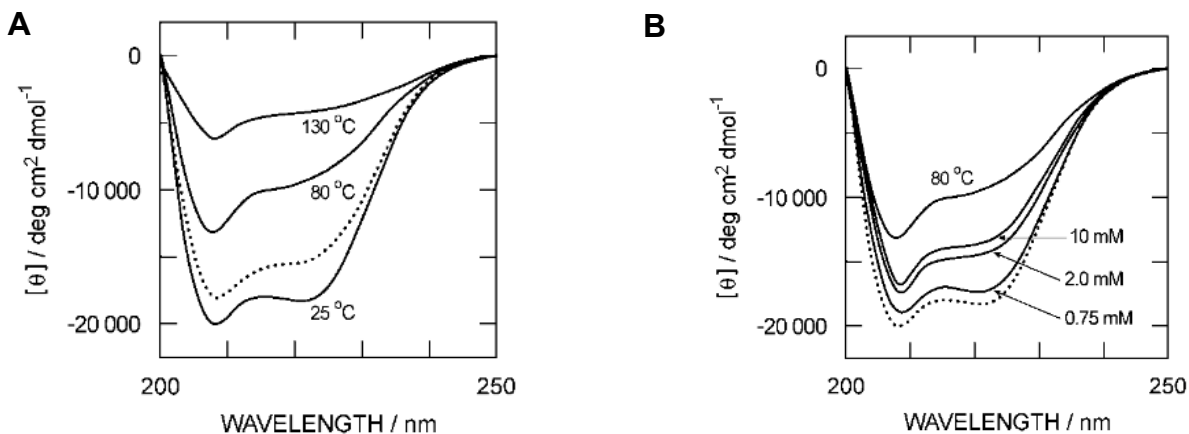


Figure 27: CD curves of thermal denaturation of BSA with and without SDS.

(A) CD curves of BSA at 25, 80 and 130°C. Thermal denaturation is observed at 80 and 130°C. (B) CD curves of BSA at 80°C with 0.75, 2.0 and 10 mM of SDS (B). Retention of secondary structure of BSA is observed with 0.75 mM of SDS (Moriyama et al.)⁷³.

CD analyses were made at three constant temperatures: 25, 50 and 80°C (Figure 28), with 1.26 μM BSA and either 0.19 mM of (-)-geosmin or 0.75 mM of SDS. The stabilizing effect of SDS is seen at 80°C (Figure 28, C), but not by (-)-geosmin at any temperature; the BSA control curves and (-)-geosmin experimental curves overlap (Figure 28). (-)-Geosmin does not seem to stabilize BSA at 80°C.

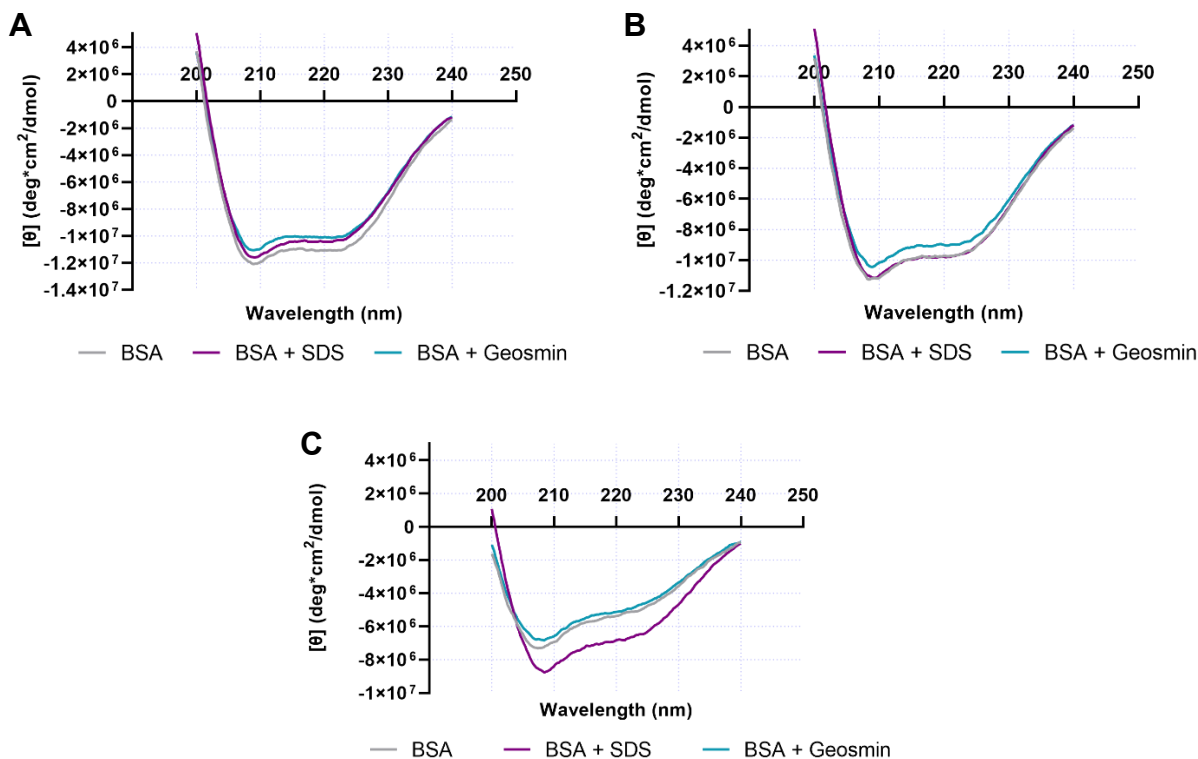


Figure 28: CD analyses of the effect of (-)-geosmin on BSA at three constant temperatures. [BSA] is 1.25 μM , [(-)-geosmin] is 0.18 mM and [SDS] is 0.75 mM. The wavelength window is from 200-240 nm.

(A) At 25°C.

(B) At 50°C.

(C) At 80°C.

The next CD experiment was made at varying temperatures. The sample was heated from 25 to 80°C, kept at 80°C for 30 min, then cooled from 80-25°C at a rate of 1°C/min. As seen in Figure 29, the curves of both BSA control and BSA with (-)-geosmin experimental overlap, suggesting that (-)-geosmin does not appreciably stabilize BSA during thermal denaturation to 80°C and cooling to 25°C. From the results from Figure 28 and 29, (-)-geosmin does not seem to stabilize proteins.

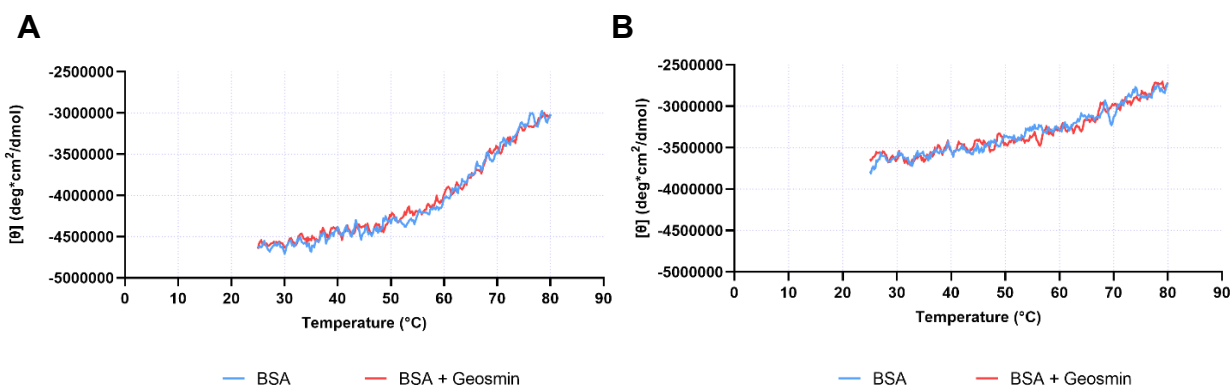


Figure 29: CD analysis of the effect of (-)-geosmin on BSA at varying temperature. [BSA] is 1.25 μM , [(-)-geosmin] is 0.18 mM and [SDS] is 0.75 mM. The wavelength used is 208.6 nm. Heating was made at a rate of 1°C/min.

(A) CD analysis of the effect of (-)-geosmin on BSA upon heating from 25-80°C.

(B) CD analysis of the effect of (-)-geosmin on BSA with a 30min hold at 80°C, followed by cooling to 25°C.

d. Effect of geosmin on protease activity

To orthogonally verify interactions between geosmin and hydrophobic proteins, I analyzed the effect of geosmin on protease activity using a fluorescein thiocarbonyl-casein (FTC-casein) fluorescent marker. For this assay, proteases from *M. xanthus* DK1622 and trypsin were the proteases used. During the experiment, FTC-casein is cleaved, and the environment around the fluorophore changes (Figure 30)^{74,75}. An increase in fluorescent signal correlates to an increase in protease activity^{74,75}.

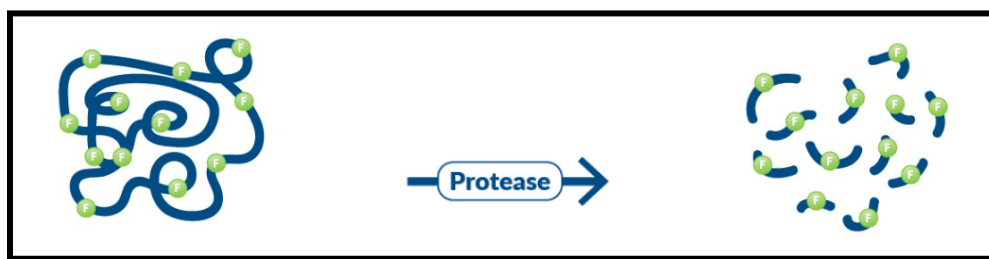


Figure 30: Schematic of cleavage of FTC-Casein by a protease⁷⁶.

For this assay, I added (\pm)-geosmin to determine if it enhances protease's activity by delivering higher fluorescent signals. Methanol was used as a control for this assay. Figure 31A-B shows the results of the activity of *M. xanthus* DK1622 proteases and trypsin respectively, without and with the presence of (\pm)-geosmin and/or methanol. The results with (\pm)-geosmin and/or methanol overlap revealing a slight decrease in the activity of trypsin (Figure 31). This could be due to the methanol interfering with the fluorescein, giving a lower signal. From these results, (\pm)-geosmin does not enhance the activity of proteases and is unlikely to be a digestive enzyme adjuvant.

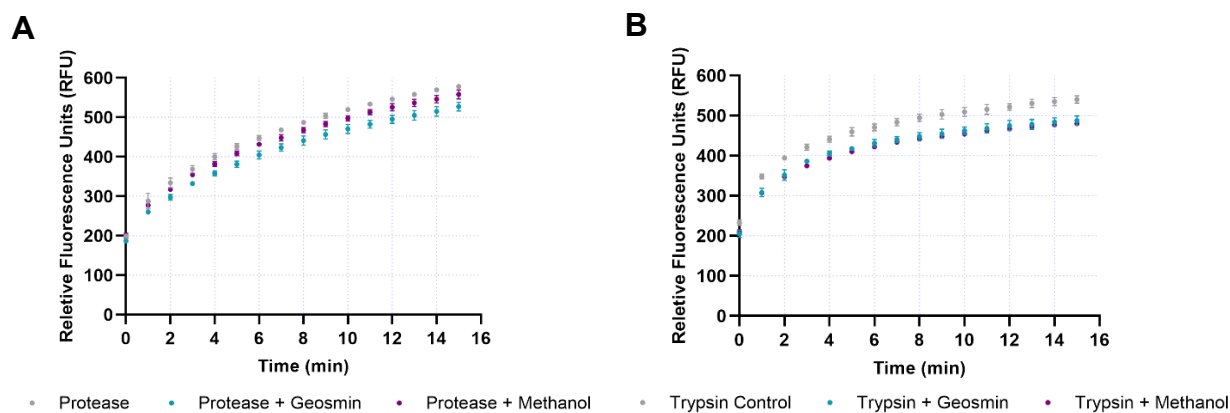


Figure 31: Effect of (\pm)-geosmin on protease activity using FTC-casein as a fluorescent marker. Measurements were made at $\lambda_{\text{excitation}} = 494 \text{ nm}$, and $\lambda_{\text{emission}} = 521 \text{ nm}$. Results of trypsin control are shown in blue, trypsin with methanol in orange and trypsin with (\pm)-geosmin in grey. (A) Determination of the effect of (\pm)-geosmin on *M. xanthus* DK1622 protease activity. (B) Determination of the effect of g(\pm)-eosmin on trypsin's activity.

e. Effect of geosmin on quorum sensing.

A study by Ahmad et al. demonstrated that some VOCs could inhibit QS⁵¹. *Chromobacterium violaceum* produces a purple pigment, violacein, which is under control of a QS system. Inhibition of violacein, of the purple pigmentation correlates to an inhibition of QS. This inhibition is also enantioselective among the terpenes borneol, α -pinene and limonene (Figure 32)⁵¹. (+)-Borneol enhances QS and (-)-borneol inhibits QS (Figure 32)⁵¹. As geosmin is a volatile organic terpene like borneol, I investigated its effects on QS.

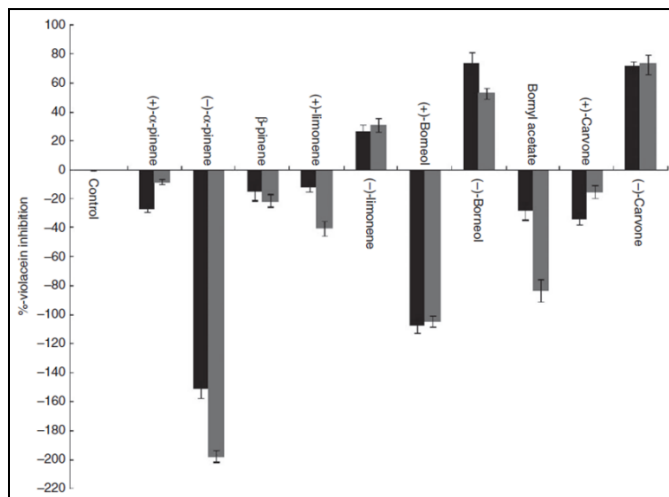


Figure 32: Violacein inhibition assay of the exposure of VOCs to *C. violaceum* (taken from Ahmad et al.)⁵¹. The black bar graphs show the minimal QS inhibitory concentration, and the grey bars show 0.5 x minimal QS inhibitory concentration.

i. *Violacein assay on 96-well plates*

QS inhibition was measured using a 96-well plate using *C. violaceum* and racemic (\pm)-geosmin. Methanol was used as a control as it is the solvent used to dissolve the standard (\pm)-geosmin. Vanillin and trans-cinnamaldehyde are used as strong QS inhibitor positive controls. (-)-limonene is used as a mild QS inhibitor positive control. (-)- α -pinene is used as a negative control as it is a QS enhancer. Table 5 shows the violacein assay results as well as the MIC results of the exposure of (\pm)-geosmin, methanol, vanillin, trans-cinnamaldehyde, (-)-limonene and (-)- α -pinene against *C. violaceum*. The results of (\pm)-geosmin and methanol overlap, suggesting that geosmin has no effect on QS in *C. violaceum*.

Table 5: Results of violacein assay and MIC test on 96-well plates of (\pm)-geosmin, methanol, vanillin, trans-cinnamaldehyde, (-)-limonene and (-)- α -pinene against *C. violaceum*.

Terpene	Violacein Inhibition ($\mu\text{g/mL}$)	MIC ($\mu\text{g/mL}$)
(\pm) Geosmin	125	250
MeOH	12.5 [μL]	25 [μL]
Vanillin	0.03125	0.25
Trans-cinnamaldehyde	0.015625	0.0625
(-)-(<i>S</i>)-limonene	0.25	>0.5
(-)- α -pinene	>0.5	>0.5

ii. *Violacein assay using agar broth dilutions*

The violacein assay was also performed using agar broth dilutions. The results are shown in Figure 33. These results show no anti-QS effect from (\pm)-geosmin at a concentration of 10 $\mu\text{g/mL}$. (-)-*S*-limonene shows mild QS inhibition and trans-cinnamaldehyde shows full QS inhibition as well as antimicrobial activity at 0.5 $\mu\text{g/mL}$.

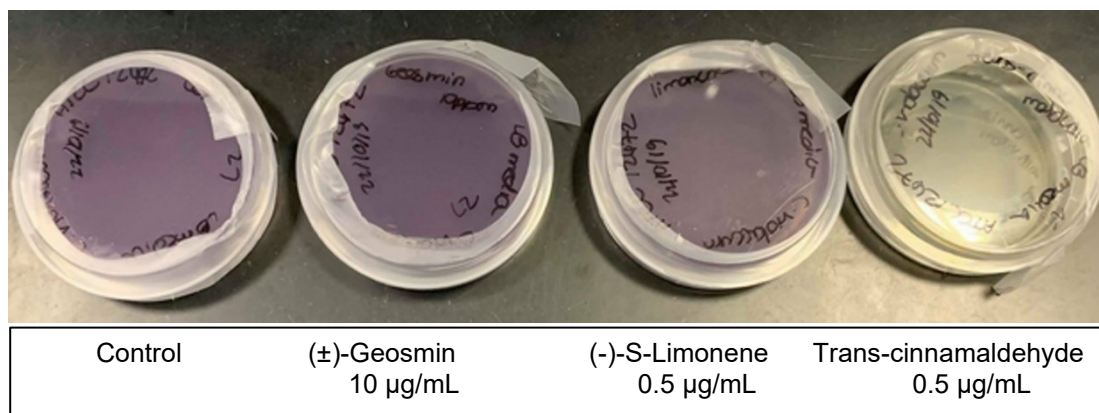


Figure 33: Violacein assay using agar broth dilutions. The concentration of (\pm)-Geosmin used was 10 $\mu\text{g/mL}$, the concentration of of (-)-*S*-limonene and trans-cinnamaldehyde used was 0.5 $\mu\text{g/mL}$.

iii. *The function of geosmin as a quorum sensing inhibitor is enantioselective*

Instead of using a solution dissolved in methanol, which has known effects on QS⁵², I used pure (-)-geosmin, the naturally produced form of the VOC¹ for subsequent experiments. I placed 5 μ L of a high concentration of 10 mg/mL of (-)-geosmin on a *C. violaceum* plate, and observed the result shown in Figure 34. The droplet shows a clear center showing antimicrobial activity, covered by a white outer-layer proving cell survival with violacein inhibition suggesting that (-)-geosmin inhibits QS (Figure 34). Similar effects were not seen with a drop of (\pm)-geosmin on a *C. violaceum* plate.

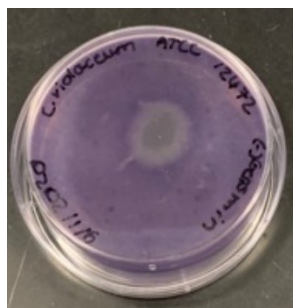


Figure 34: Violacein assay using agar broth dilution and (-)-geosmin. A 5 μ L droplet of [(-)-geosmin] of 10 mg/mL was placed onto the plate containing *C. violaceum*.

f. Geosmin as a defense chemical against amoeba predation

As mentioned previously, amoeba may be responsible for up to 60% of bacterial predation following rainfall⁶³. To determine if the ecological function of geosmin is to guard against amoeba predation bacteria viability assays were adapted from Xiao, Y. et al.⁴⁰. The predator used was the amoeba *Naegleria Gruberi* ATCC 30223 and the prey was *E. coli* MG1655. Geosmin was added in to determine its effect against amoeba predation.

i. *Predation assay in liquid media*

The protocol was adapted from the killing assay from Xiao, Y. et al. *E. coli* (0.5 McFarland standard) was incubated with amoeba cells for 2h at 37°C and 225 rpm⁴⁰ (Figure 35A). The media used is amoeba saline (AS), which is minimal media adapted for amoeba. The amount of *E. coli* cells that survived predation by amoeba are shown in Table 6 (Figure 35B). A higher number of quantified cells equates to a higher resistance against amoeba predation.

From Table 6, a higher number of cells is attributed to the assay containing (-)-geosmin leading to the belief that (-)-geosmin could aid *E. coli* cells against amoeba predation. However, a student T-test revealed a p-value of $P=.21$. From this assay, (-)-geosmin does not seem to aid in the survival of *E. coli* MG1655 against predation by *N. gruberi* ATCC 30223.

Table 6: Results of the effect of (-)-geosmin on *E. coli*'s survival against amoeba predation in liquid media.

Conditions	*10 ⁵ CFU/mL	Average (*10 ⁵ CFU/mL)	Standard Deviation (*10 ⁵ CFU/mL)
<i>E. coli</i> ^a + Amoeba ^b	237	253.6	79.87
	216		
	151		
	329		
	216		
	219		
	422		
	214		
<i>E. coli</i> ^a + Amoeba ^b + (-)-Geosmin ^c	278	433.2	405.3
	228		
	206		
	287		
	78		
	193		
	1431		
	416		
518			
542			

^a *E. coli* strain used is MG1655

^b *Amoeba* strain used is ATCC30223

^c (-)-Geosmin dissolved in nanopure water, 0.1 ppm concentration

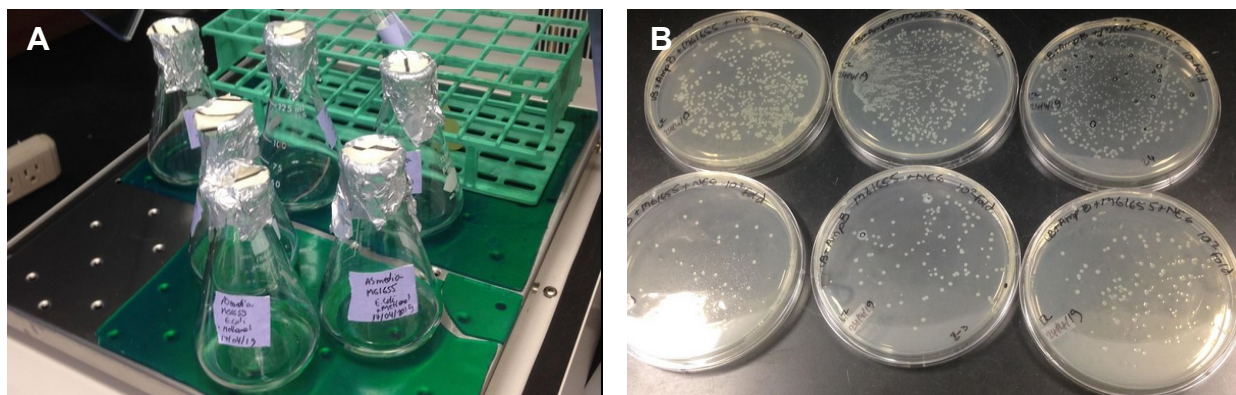


Figure 35: Liquid assay of amoeba predation on *E. coli* cells.

(A) Picture of assay in 37°C incubator.

(B) Agar plates covered with surviving *E. coli* CFU. The agar plate contains epothilone, an anti-eukaryotic metabolite, permitting to selectively kill the amoeba cells, allowing the *E. coli* cells to grow and be quantified.

ii. Predation assay in solid media

Table 7: Results of the effect of geosmin on *E. coli*'s survival against amoeba predation on solid media.

Conditions	CFU/mL	
	10 ³	10 ⁴
<i>E. coli</i> ^a + <i>Amoeba</i> ^b	88	10
	154	11
	186	19
<i>E. coli</i> ^a + <i>Amoeba</i> ^b + 0.1ppm geosmin ^c	129	24
	78	17
	103	18
<i>E. coli</i> ^a + <i>Amoeba</i> ^b + 1 ppm geosmin ^c	91	9
	172	27
	131	20
<i>E. coli</i> ^a + <i>Amoeba</i> ^b + 10 ppm geosmin ^c	107	9
	147	14
	147	20

^a *E. coli* strain used is MG1655

^b Amoeba strain used is *N. gruberi* ATCC30223

^c (-)-Geosmin dissolved in nanopure water

After obtaining the results from the liquid test, I decided to switch to a solid assay using AS agar plates. The liquid media might have prevented the amoeba cells from effectively preying on the *E. coli* cells due the high motility of the latter in liquid media. The results of this study are shown in Table 7. A student T-test revealed the lowest p-value at $P=.15$, for the exposure to 0.1 ppm of (-)-geosmin. However, this value is still higher than the standard of 0.05, depicting insignificant results. There is insufficient evidence from this solid assay to support the hypothesis that (-)-geosmin could reduce predation by amoeba.

iii. Swarming assay on solid media

Swarming assays, adapted from Xiao, Y. et al were made to evaluate the effect of (-)-geosmin on the migration of amoeba into *E. coli* lawn⁴⁰. A 10 µl aliquot of *E. coli* MG1655 of an OD₆₀₀=50 was added to minimal media AS plate. An aliquot of 5µl of amoeba cells of an OD₆₀₀=0.7 was added adjacently to the prey drop (Figure 36A). The migration distance of the amoeba cells into the *E. coli* drop is measured every 7 days for 21 days (Figure 36B). Different concentrations of (-)-geosmin are added to the *E. coli* lawns: 0.1, 1 and 10 ppm. There is a significant ($P=.01145$) decrease in migration distance of the amoeba cells into the *E. coli* lawn with (-)-geosmin at a concentration of 0.1 ppm compared to the negative control; at this concentration (-)-geosmin appears to act like a repellent or deterrent (Figure 36B). At 1 ppm however there is a significant increase ($P=.002768$) compared to the negative control in the

migration distance of the amoeba cells into the *E. coli* lawn, suggesting that (-)-geosmin instead acts as an attractant (Figure 36B).

There is evidence showing that (-)-geosmin may act as an attractant at 1 ppm, and a repellent at 0.1 ppm towards the amoeba *N. gruberi* ATCC 30223.

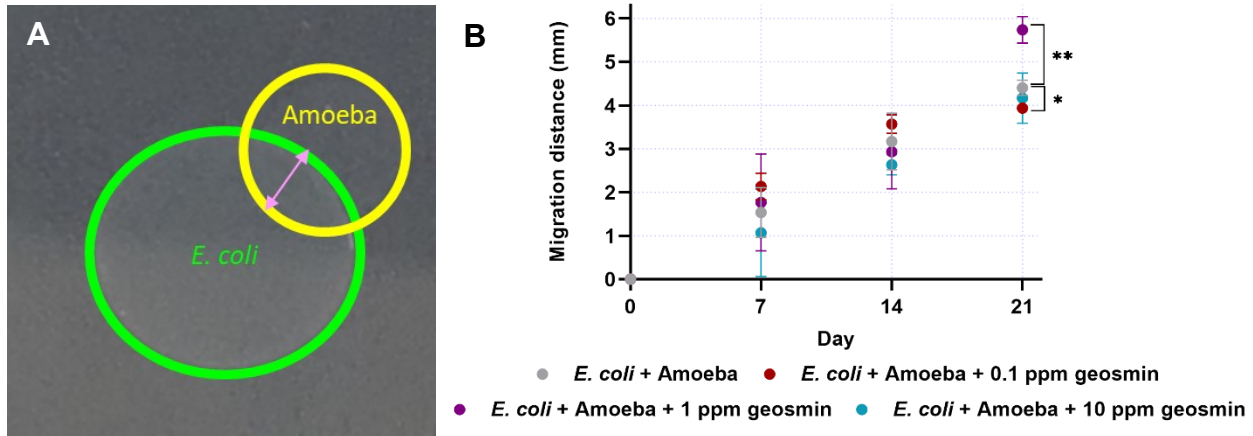


Figure 36: Swarming assay of Amoeba *N. gruberi* ATCC 30223 into *E. coli* MG1655 lawn.

(A) Picture of swarming of amoeba cells into the *E. coli* lawn. The pink double-headed arrow shows the migration distance of the amoeba cells into the *E. coli* lawn.

(B) Results of swarming assay of the effect of (-)-geosmin on amoeba swarming on *E. coli* lawn (n=3). Statistically significant deviations compared to the negative control (*E. coli* + *Amoeba*) are indicated by * (P < 0.05) or ** (P < 0.01).

g. Geosmin as an aposematic signal to *C. elegans*

The previous experiment with the amoeba *N. gruberi* as a predator suggests that geosmin acts as a repellent at 0.1 ppm and as an attractant at 1 ppm. This experiment however, lasted 21 days. Geosmin is unstable in acidic and basic media^{36,70}, and could have differentially evaporated from the agar plates, potentially impacting predation by *N. gruberi*. A faster assay was required, and so I switched to predation assays using the predatory nematode *C. elegans*.

i. Chemotaxis experiment with *C. elegans*

A chemotaxis assay was made to evaluate the effect of concentration of (-)-geosmin on N2 adult *C. elegans*. This experiment was conducted with the help of Karina Mastronardi, a PhD student working in the Piekny lab at Concordia University, Montreal, QC. It was performed on a minimal media plate containing (-)-geosmin (at different concentrations) at one end and nanopure water on the other end as a control (Figure 37)^{77,78}. Each liquid contained levamisole, an anaesthetic that suppressed nematode movement near either drop, allowing future quantification. Fifty worms were starved, washed three times in minimal M9 worm media, and then placed at the center of the plate. After one hour of incubation the chemotaxis index was calculated. This is equal

to the difference of the number of worms at the experiment and control drops divided by the total number of worms. The chemotaxis index can vary from +1, indicating perfect attraction, to -1, indicating perfect repulsion. 2-butanone was used as an attractant control and gave a chemotaxis index of 0.58 towards *C. elegans* (Table 8)^{66,77,78}.

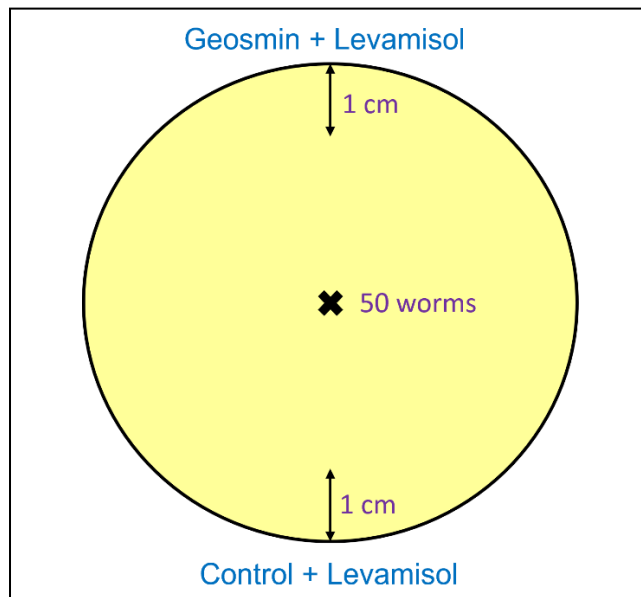


Figure 37: Schematic of chemotaxis assay plate of the effect of (-)-geosmin on adult *C. elegans*. Different concentrations of (-)-geosmin were placed to verify if the effect of (-)-geosmin is concentration dependent over 1 hr. Levamisol was used to paralyse the worms upon contact with either drop (n=50).

Table 8: Results of the (-)-geosmin chemotaxis assay with adult *C. elegans*. Positive chemotaxis index implies attraction, negative chemotaxis index implies aversion or repulsion.

[(-)-Geosmin] ($\mu\text{g/mL}$)	Worms in (-)-Geosmin Spot	Worms in Control Spot	Chemotaxis Index ^b
54000	14	10	0.08
5400	15	7	0.16
540	10	10	0
54	10	5	0.1
5.4	6	12	-0.12
0.54 ^c	4	14	-0.2
0.054	17	8	0.18
0.0054	11	7	0.08
0.00054	11	5	0.12
2-butanone	30	1	0.58

^a Fifty adult *C. elegans* worms per trial. Singlicate experiments, unless otherwise noted.

^b Chemotaxis index calculated as $\frac{\text{Number of worms at geosmin spot} - \text{Number of worms at control spot}}{\text{Total number of worms}}$.

^c Experiment ran in triplicates. The Average chemotaxis index is -0.1, s = 0.11.

The results shown in Table 8, show different chemotaxis index results at varying (-)-geosmin concentrations. At higher concentration (54-54 000 $\mu\text{g}/\text{mL}$), (-)-geosmin acts as a mild attractant, at lower concentration (0.54-5.4 $\mu\text{g}/\text{mL}$) (-)-geosmin act as a mild repellent. However, at very low concentrations (0.00054-0.054 $\mu\text{g}/\text{mL}$), (-)-geosmin acts as an attractant again. The most significant chemotaxis index result from the exposure of (-)-geosmin to *C. elegans* is -0.2, which shows mild repellence at a concentration of 0.54 $\mu\text{g}/\text{mL}$ (with an average chemotaxis index of 0.11). The concentration of (-)-geosmin slightly influences the chemotactic behavior of *C. elegans*. Although this effect does not appear to be significant on chemotaxis, there was a change in behavior observed under the microscope at all concentrations. The worms moved with an erratic behavior in the presence of (-)-geosmin when compared to the control.

ii. *Geosmin alters nematode behaviour*

To determine if (-)-geosmin inhibited bacterial predators, we tested its toxicity against *C. elegans*. (-)-Geosmin had no effect on nematode viability over a 24-hr period (Figure 38A-B, A1). Addition of (-)-geosmin to *E. coli* did not reduce the time required for *C. elegans* to locate prey colonies, or reduce feeding once such colonies were located (Figure 38A-B, A1).

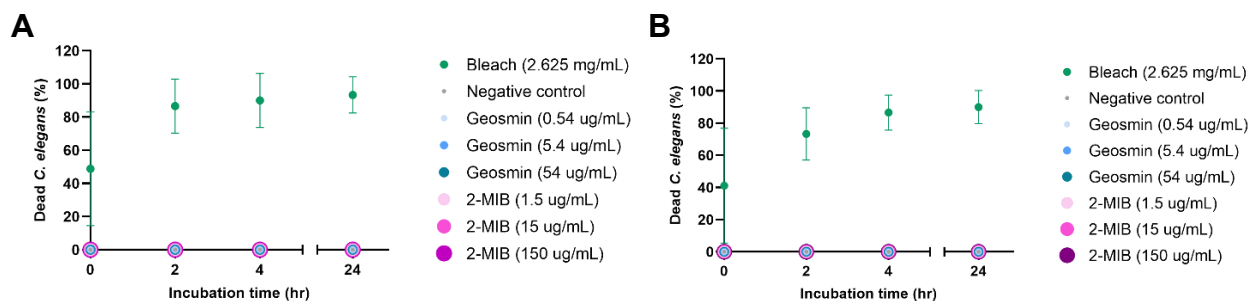


Figure 38: *C. elegans* viability in the presence of (-)-geosmin, 2-MIB and bleach.

Experiments were conducted in triplicates with 5 adult *C. elegans* per well (n=15).

(A) Viability assay on *E. coli* supplemented NGM plates.

(B) Viability assay on plain NGM without *E. coli* as food source.

However, (-)-geosmin strongly altered the worms' behaviour. Nematodes added to (-)-geosmin-laced plates moved more erratically than those on control plates, repeatedly changing direction and favouring more linear movement (Additional file 10-11). To track this change in behaviour we followed the movement of the worms in Imaris and noted significant changes in their track linearity (Table 9). The behaviour of the worms themselves were measured with WormLab, which revealed perturbations in center point and head movement periodicity (Table 9).

To determine how (-)-geosmin was causing these changes in nematode behaviour, we compared the movement of a range of *C. elegans* mutants in the presence or absence of (-)-

geosmin. The behaviour of a mutant defective in chemotaxis, *C. elegans* NL2105, was unaffected by (-)-geosmin, confirming that the terpene's effect was mediated by the host chemosensation machinery (Table 9, Additional file 12-13)⁷⁹. Geosmin's activity was retained in worms deficient in the sensing of volatile attractants and repellents (mutants lacking AWA and AWB neurons) but lost in worms lacking the gustatory neuron ASE. The latter has been previously implicated in the sensing of water-soluble attractants and movement away from food as prey populations decline^{80,81}.

Table 9: Comparison of *C. elegans* movement on (-)-geosmin (0.54 µg/mL) and non-(-)-geosmin NGM plates using Imaris (Track line) and WormLab (Peristaltic speed, periodicity). Significant effects by (-)-geosmin are represented by a red box ($P < 0.05$) and non-significant effects by a green box ($P \geq 0.05$). Experiments were conducted in triplicates with three worms per plate (n=9).

<i>C. elegans</i>	Sensory Deficiency	Track Line (%)	Peristaltic Speed (µm/s)	Periodicity (µm)
N2	Wild type			
BR5514	AFD, ASE and AWC			
CE1248	AWA			
CX2065	AWB and partial AWC			
CX2205	Olfactory system			
CX5893	AWC			
NL2105	Chemosensation			
PR674	ASE			

$P < 0.05$

$P \geq 0.05$

iii. Geosmin deters feeding on its producers.

As geosmin did not limit predation of *E. coli* by *C. elegans*, to test its effect *in situ* *C. elegans* were added to plates containing colonies of *S. coelicolor*, using both the wild type (WT) *S. coelicolor* M145 and mutant strains lacking in the production of geosmin (J3003) and both geosmin and 2-MIB (J2192)^{82,83}. When *C. elegans* N2 were added to *S. coelicolor* M145 or J3003 the majority of worms localized outside of the bacterial colonies over a 4 hr period (Figure 39A). When added to *S. coelicolor* J2192 worms were predominantly found within the bacterial colony at the 2 hr, 4 hr and 24 hr mark (Figure 39A). Worms lacking the ASE neuron, *C. elegans* PR674, localized within the bacterial colony at all time points (Figure 39B). In all experiments, nematodes consumed the bacteria, as noted by the presence of red bacteria within the gut of the nematodes (Figure 39E). The addition of worms lead to rapid sporulation of *S. coelicolor* and the production of the toxic bacterial metabolite actinorhodin (Figure 39F)⁸⁴. The majority of worms within the bacterial colony at the 24 hr mark were coated in white bacterial spores, and either exhibited distress behaviour (Additional file 14), or appeared dead. The movement of worms into and out of the bacterial colony did disperse bacterial spores beyond the boundaries of the colony (Figure

39F), but given the high toll on both bacterial growth and worm viability the overall effect of nematode predation was detrimental to both nematodes and bacteria.

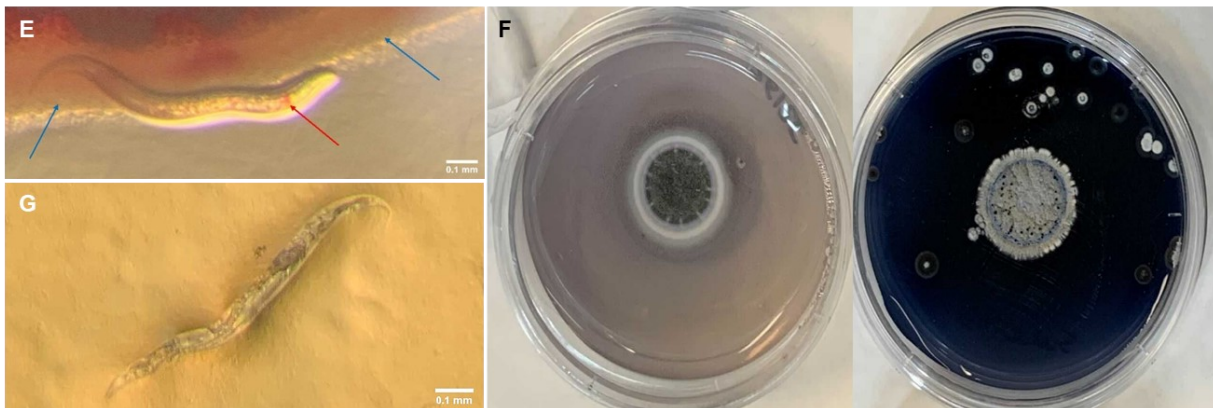
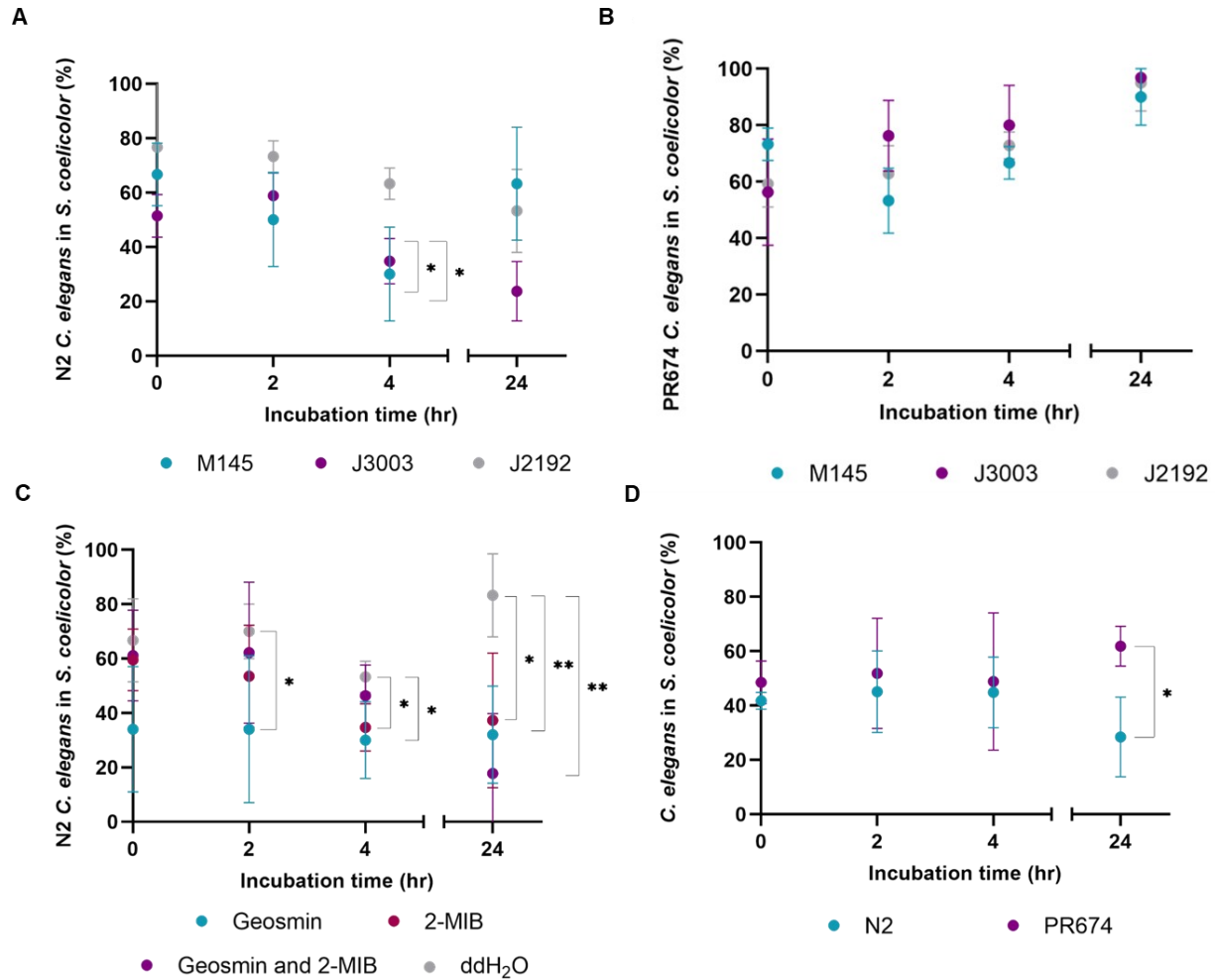


Figure 39: Predation of *S. coelicolor* and *M. xanthus* by *C. elegans*. All worm experiments were run in triplicate, ten worms per study (n=30). Statistically significant deviations from wildtype are indicated by * (P < 0.05) or ** (P<0.01). Representative images are shown. (A) Addition of *C. elegans* N2 to colonies of *S. coelicolor* M145 (WT), J3003 (Δ *GeoA*) and

J2192 (*ΔgeoA ΔmibAB*).

(B) Addition of *C. elegans* PR674 (*che-1(p674)*, ASE deficient) to colonies of *S. coelicolor* M145 (WT), J3003 (*ΔgeoA*) and J2192 (*ΔgeoA ΔmibAB*). Experiments were run in triplicate, ten worms per study (n=30).

(C) Addition of *C. elegans* N2 to colonies of *S. coelicolor* J2912 that were pre-treated with (-)-geosmin, 2-MIB, or distilled, deionized water. Experiments were run in triplicate, ten worms per study (n=30).

(D) Addition of *C. elegans* N2 and PR674 to *M. xanthus* DK1622. Experiments run in triplicate, ten worms per study (n=30).

(E) Consumption of *S. coelicolor* by *C. elegans*. Blue arrows indicate the bacterial colony, red the presence of bacteria in the pharynx.

(F) Production of spores and actinorhodin by *S. coelicolor* in the absence (left) or presence (right) of *C. elegans*. 10 day cultures, 25 °C.

(G) *C. elegans* corpse on *M. xanthus*.

To ensure that the non-terpene metabolite profile of *S. coelicolor* mutants J3003 and J2192 did not meaningfully differ from that of the wildtype we conducted a series of add-in experiments. (-)-Geosmin and 2-MIB were added alone or in combination to colonies of *S. coelicolor* J2192, at concentrations approximating their physiological values, 2.25 μg/mL and 40 μg/mL respectively^{83,85}. (-)-Geosmin or 2-MIB was sufficient to significantly reduce the number of worms in bacterial colonies at the 2 hr, 4 hr, and 24 hr mark (Figure 39C). The effect at 24 hrs with both (-)-geosmin and 2-MIB added together showed a very significant reduction of the worms in the bacteria, in line with the co-occurrence of geosmin and 2-MIB biosynthetic genes in actinobacteria and cyanobacteria^{86,87}.

To determine if the reduction in predation that we observed was limited to *S. coelicolor*, we repeated our assay with *M. xanthus* DK1622. As with the *Streptomyces*, *C. elegans* N2 avoided *M. xanthus*, despite the latter's swarming movement over the agar plate (Figure 39D). *C. elegans* PR674 showed no such aversion, resulting again in the death of both predator and bacterial prey (Figure 39G).

h. Geosmin is a warning chemical.

In the animal kingdom, toxic prey advertise their unpalatability through the use of warning colours⁸⁸. These bright colours make the prey more conspicuous, but when combined with negative stimuli they deter predation through learned responses^{89,90}. To date no warning colours or other aposematic signals have been identified in prokaryotes, though olfactory signals may be used to reduce scavenging of nutrient-rich insects killed by entomopathogenic bacteria⁹¹.

The high prevalence of geosmin synthase genes across a wide range of unrelated bacteria and fungi suggests that geosmin is key to the fitness of a wide range of microorganisms. Here we propose that geosmin acts as a widespread warning chemical, used to advertise the toxicity of its producers and deter predation. Consistent with this role geosmin is produced throughout the

lifecycle of *M. xanthus*, and is actively excreted from the cell (Figure 26, Table 4). Production in *Streptomyces* spp. is more complex, as the geosmin precursor isopentenyl pyrophosphate is produced by both the MEP and mevalonate pathways in this genus⁸⁵, and the two pathways are differentially regulated between strains and growth phases^{92,93}. Cyanobacteria produce both geosmin and 2-MIB during exponential growth and stationary phase, and release the terpenes as they die⁹⁴.

To act as an aposematic signal geosmin must be detected by potential microbial predators. The bacteriophagous *C. elegans* detected geosmin through the gustatory neuron, ASE, and adopted a hesitant hunting pattern, with more frequent changes in direction (Table 9, Table A1, Additional files 10-13). In the presence of *S. coelicolor*, when worms were able to sense geosmin and either geosmin or 2-MIB was present this change in behaviour resulted in significantly fewer worms in the bacterial colonies (Figure 39). Similar results were obtained when geosmin or 2-MIB was added to a *S. coelicolor* mutant deficient in the production of these terpenes, and when worms were added to geosmin-producing *M. xanthus* DK1622. Geosmin was not only non-toxic to these worms (Figure 38), but by altering *C. elegans*'s feeding behaviour geosmin prevented the worms from coating their bodies in bacterial spores or ingesting toxic bacterial metabolites^{84,95}. As with *Drosophila* and *A. aegypti*,^{96,97} this effect was mediated by the host's own chemosensory system.

The use of geosmin as an aposematic signal explains both its reported effects on other eukaryotes and its prevalence across a range of unrelated microbes. Geosmin attracts *Solenopsis invicta* because the terpene reliably indicates the presence of *Streptomyces* spp, and the toxic metabolites produced by these bacteria protect ant colonies from fungal infections⁹⁸. Similarly, geosmin discourages egg laying by *Drosophila*, whose young are susceptible to bacterial toxins⁹⁶, while also signalling the presence of edible cyanobacteria to the more toxin-resistant *A. aegypti*^{97,99}. In principle, acquisition of geosmin synthase by any toxin-producing microbe could recapitulate the aposematic phenotype, favouring lateral gene transfer between evolutionarily unrelated species and the evolution of Müllerian mimics¹⁰⁰⁻¹⁰². Other aposematic signals exhibit positive frequency-dependent selection¹⁰³, and the evolution of Müllerian mimics in bacteria likely favours the further lateral gene transfer of geosmin synthase. The ubiquity of geosmin in natural environments ensures few predators are naïve to the signal, while those that ignore it likely experience consistent fitness penalties from preying on toxic geosmin producers.

Geosmin did not prevent *C. elegans* from feeding on *E. coli* but heavily reduced grazing on *S. coelicolor* or *M. xanthus* (Figure 39), suggesting that aversion requires the terpene and a negative stimuli¹⁰⁴. Microbial predators can discern between adjacent toxin-producing and non-toxic bacteria¹⁰⁵, and the ability of geosmin to deter but not protect against predation may prevent the emergence of Batesian mimics, which express only the terpene. While *C. elegans* reacted to the presence of geosmin prior to the introduction of *S. coelicolor* or *M. xanthus* (Figure 39), it is unclear if the avoidance of toxin-producing bacteria was learned or innate. Higher eukaryotes learn to associate aposematic signals with unpalatability^{89,106}, and while bacterial predators are

significantly less complex, both *C. elegans* and amoeba can associate sensory cues with past events¹⁰⁷⁻¹¹⁰.

i. Bioinformatics analysis on GS genes among prokaryotes.

i. Geosmin biosynthetic genes are well dispersed and interrelated in prokaryotes.

To analyze the distribution of geosmin synthase, a compilation of GS protein and nucleotide sequences were extracted from a list of fully assembled genomes stored in the National Center for Biotechnology Information (NCBI) database (Additional file 1). In total we found 337 sequences with E values $< 1 \times 10^{-20}$ relative to reference GS sequences from *Myxococcus xanthus* DK1622, *Streptomyces coelicolor* A3(2), and *Anabaena ucrainica* CHAB 1434. A representative subsection of these sequences (64 actinobacteria, 23 myxobacteria, 19 cyanobacteria, 1 pseudomonad and 1 ktedonobacteria) were then used to generate a phylogenetic tree (Figure 40A). This phylogenetic analysis showed two dominant clades. The first is composed of sequences from myxobacteria, cyanobacteria, and *Ktedonosporobacter rubrisoli* SCAWS-G2, while the second clade is composed predominantly of actinobacterial GSs. The first clade branches into three distinct sub-clades, separating the GS found in saprophytic myxobacteria from those of the cyanobacteria and predatory myxobacteria. The latter branches into myxobacteria and cyanobacteria (with 1 ktedonobacteria) as separate clusters. The branching pattern of the GS genes showed strong alignment with distantly related species. For example, *Archangium gephyra* DSM 2261, which contained two GS sequences, had one that aligned with *Mellitangium boletus* and *Cystobacter fuscus* (myxobacterial) GSs and another that aligned with *Nocardia terpenica* (actinobacterial) GS. GS genes were also observed in *Pseudomonas agarici* and *Ktedonosporobacter rubrisoli* though geosmin has not been previously characterized in these groups. Overall, there was little correlation between the phylogeny of GS and that of its carriers, strongly suggesting repeated HGT events¹¹¹.

While a number of authors have reported geosmin production in both fungi and plants^{1-3,34,112}, we did not identify any putative eukaryotic GS sequences during our general BLASTp analysis or during antiSMASH analysis of putative eukaryotic producers *Beta vulgaris* L., *Penicillium expansum* and *Aspergillus tubingensis* (Table A2). It is likely that these organisms produce geosmin through an alternative biosynthetic pathway, unrelated to the terpene cyclase found in prokaryotes.

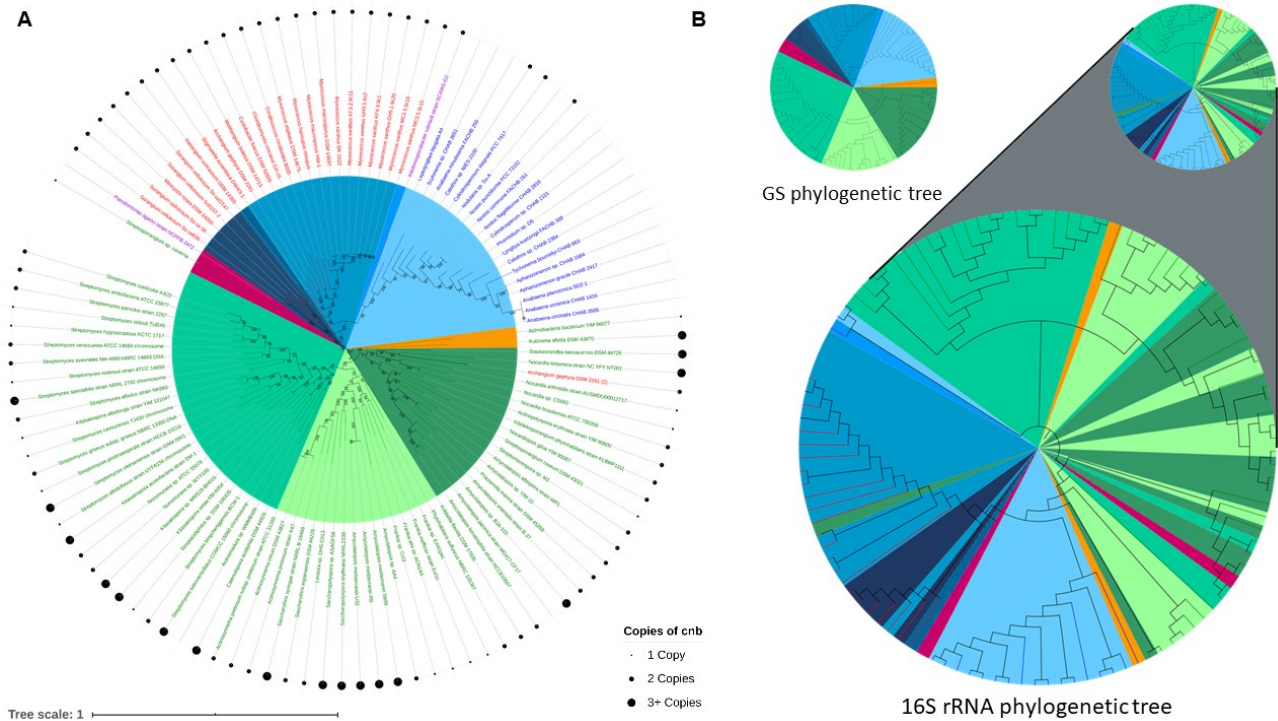


Figure 40: Phylogenetic trees of GS and 16S rRNA among 109 bacterial strains.

(A) GS genes from actinobacteria, myxobacteria, cyanobacteria and other bacteria (*Pseudomonas* and *Ktedonobacteria*) are represented in green, red, blue and purple respectively. The presence of CNBP is represented by black circles of different sizes: small, medium and large, which correspond respectively to 1 copy, 2 copies and more than 3 copies per genome. The tree was generated using the MEGA-X maximum likelihood method with MUSCLE with bootstrap values with 500 replicons following the Tamura-Nei model. The inner circle represents the major clades formed by the phylogenetic tree.

(B) Comparison of 16S rRNA and GS phylogenetic using the colour coding of the clades defined in Figure 40A. The two trees do not correlate to each other as the branching pattern and clade distribution do not align with each other, which is proof of HGT of GS gene among various prokaryotes.

ii. *Geosmin producers occupy terrestrial, freshwater, and marine ecological niches.*

Geosmin is produced by a large number of terrestrial and freshwater organisms, but is thought to be nearly absent in marine organisms²⁸. Analysis of our dataset revealed that GS is present in at least 7 marine Actinobacteria, 4 marine cyanobacteria, and 4 marine myxobacteria (Additional files 2-4). These numbers should be considered an underestimate, as many GS-containing strains do not list their environment of origin. The marine organisms that we identified were dispersed throughout the phylogenetic tree, with their GSs clustering with GSs from terrestrial and aquatic organisms. While the geographically dispersed marine actinobacterial genus

Salinospora^{113–116} lacked GS, GS was found in the ubiquitous picocyanobacterial genus *Synechococcus*, the myxobacterial *Enhygromyxa salina* and the actinobacterial genus *Actinomadura*, suggesting geosmin is broadly prevalent in marine environments^{114–119} (Additional files 2-4).

iii. *The phylogeny of GS is distinct of its host genome*

To determine if GS is laterally transferred, we generated a 16S rRNA phylogenetic tree with the strains used in Figure 40A (Figure 40B). While as in Figure 40A these sequences subdivided into several clades, the branching pattern differed markedly from that of the GS phylogenetic tree. In particular, the GS of *Pseudomonas agarici* NCPPB 2472, a Gram-negative strain, clusters with an actinobacterial GS, *Streptosporangium* sp. caverna, forming one clade in Figure 40A and clusters with myxobacterial 16S rRNA in Figure 40B. *Ktedonosporobacter rubrisoli* strain SCAWS-G2, *Nostoc punctiforme* PCC 73102 and *Cylindrospermum stagnale* PCC 7417 all form one clade in Figure 40B, but their respective GS genes are separated in Figure 40A. Furthermore, the GS genes of *Amycolatopsis* strains do not branch together contrary to their 16S rRNA sequences. This is also observed for the Kitasatospora, *Sorangium cellulosum* and *Minicystis rosea* DSM 24000 and many more strains. Alongside the previously discussed GS variant in *Archangium gephyra* DSM2261, this strongly suggests the transfer of GS genes between prokaryotes.

iv. *An uncharacterized cyclic nucleotide-binding protein gene is associated with terpene biosynthesis*

In both myxobacteria and cyanobacteria the geosmin operon contains genes for two uncharacterized cyclic nucleotide-binding proteins (CNBP)s, but these genes are not part of the GS operon in actinobacteria^{69,120}. Through a preliminary BLASTp analysis, we identified homologues to the *M. xanthus* CNBPs MXAN_6248 and MXAN_6249 in both *S. coelicolor* A3(2) and *S. avermitilis*, over 950 kb from GS. A more extensive BLASTp analysis revealed that 82.57% of all strains in the GS phylogenetic tree have at least one CNBP homologue (Figure 40A). When not adjacent to GS, the CNBP genes are frequently found adjacent to an uncharacterised polyprenyl synthetase (39.35%, Table 10), hypothetical proteins (31.40%, Table 10), an uncharacterized transcriptional regulator (18.28%, Table 10) or other terpene synthases such as 2-methylisoborneol synthase (13.55% , Table 10) and camphene synthase (1.94%, Table 10). All myxobacterial and cyanobacterial strains in our dataset contained two CNBP genes directly adjacent to GS.

Table 10: Summary of the determination and analysis of genes adjacent to CNBP in 293 actinobacterial genomes. Of the 293 genomes, 277 were analyzed, the remaining 16 lacked genome annotations in NCBI, presenting a total of 465 CNPBS.

Adjacent Genes to CNBP	Adjacent Gene Counts	% Adjacent Gene in Total CNBP Count	% Adjacent Gene in Annotated CNBP
Polyprenyl synthetase	183	38.05	39.35
2-MIB synthase	63	13.10	13.55
Geosmin synthase	15	3.12	3.23
Cyclase	43	8.94	9.25
Germacradienol synthase	9	1.87	1.94
Cysteine desulfurase	29	6.03	6.24
Hypothetical protein	146	30.35	31.40
Transcriptional regulator	85	17.67	18.28
Methyltransferase	11	2.29	2.37
N-Acetylmuramoyl-L-alanine amidase	52	10.81	11.18
1-Aminocyclopropane-1-carboxylate deaminase	15	3.12	3.23
Rrf2 family transcriptional regulator	12	2.49	2.58
Camphene synthase	9	1.87	1.94
VOC family protein	7	1.46	1.51
(2E,6E)-Farnesyl diphosphate synthase	10	2.08	2.15
Terpene synthase	17	3.53	3.66
1-Deoxy-D-xylulose-5-phosphate synthase	8	1.66	1.72

At present, the function of these CNBPs is not known, but they are closely related to members of the Crp-Fnr family of transcriptional regulators^{121,122}. To clarify this relationship we constructed a phylogenetic tree containing the *M. xanthus* DK1622 CNBPs MXAN_6248 and MXAN_6249 and representatives from each of the major Crp-Fnr subfamilies (Figure 41)¹²¹. MXAN_6248 and MXAN_6249 clustered with the NnrR subfamily of Crp-Fnr transcriptional regulators, proteins involved in nitric oxide homeostasis¹²². However, only 30-35% of the *M. xanthus* sequences aligned with prototypical NnrR genes (Table A3). MXAN_6248 and MXAN_6249 were also 100-200 amino acids longer than these NnrR proteins, which were

distributed throughout the alignment. Overall, the terpene-associated CNBPs appear to form a distinct subfamily of Crp-Fnr transcriptional regulators.

Tree scale: 1

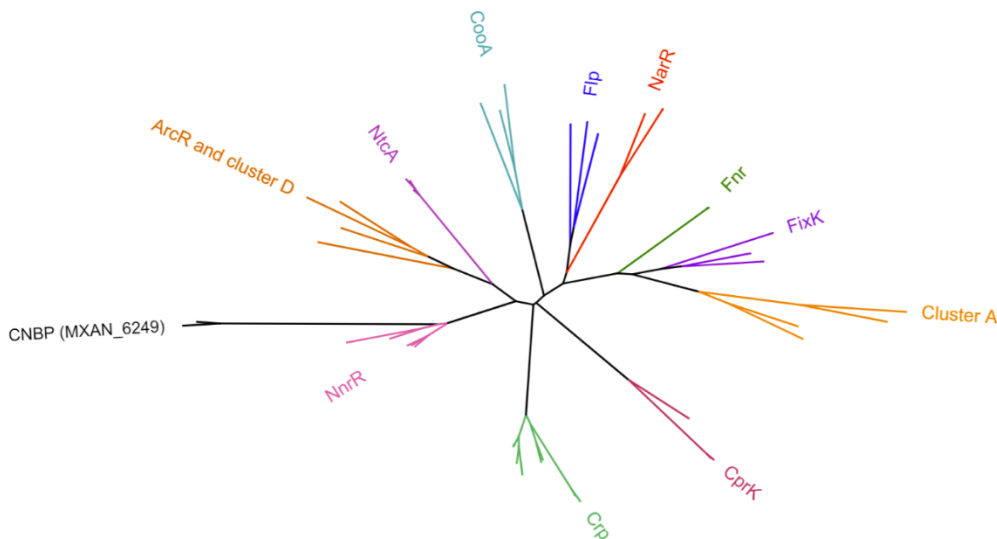


Figure 41: Phylogenetic tree of Crp-Fnr subfamilies. The two CNBPs of *M. xanthus* DK1622 (MXAN_6248 and MXAN_6249) were analysed and seem to be closely related to the NnrR subfamily. The tree was generated using MEGA-X maximum likelihood method.

v. *Several genes co-occur with GS.*

Intrigued by the presence of geosmin-associated genes outside of the geosmin biosynthetic gene cluster, we developed a gene co-occurrence algorithm to expand our investigation to all orthologous genes in these organisms. Working from an orthologue cluster library drawn from 5614 prokaryotic and eukaryotic genomes¹²³, the co-occurrence of geosmin synthase and 2,110,338 orthologue clusters was determined (Additional file 6). Key findings are shown in Table 11. The highest correlation was with OC.1340183, a DUF574 family gene, present in 108 of 114 geosmin synthase-containing genomes (159 genomes overall, Pearson coefficient = 0.798). While this gene is of unknown function, it is hypothesized to encode a SAM methyltransferase. A crystal structure of a homologue from *S. avermitilis* MA-4680 is available in the Protein Data Bank¹²⁴, and the gene for this protein is less than 12 kb from GS in this organism. Other important hits include OC.316564, a polyprenyl synthetase family protein frequently associated with the CNBPs in actinobacteria (Table 11) and OC.1631, a hybrid sensor histidine kinase/response regulator. The CNBP orthologue cluster was found in 113/114 of the GS-containing genomes in this dataset, in line with our initial BLAST analysis (360 genes overall, Pearson coefficient = 0.545). 1-deoxy-D-xylulose-5-phosphate synthase also shows slight co-occurrence with GS at a Pearson coefficient value of 0.380. This protein is a component of the MEP pathway, and produces an early building block in the biosynthesis of terpenes¹². Except for OC.1238530, none of these orthologues have been previously associated with geosmin biosynthesis or function.

Table 11: Co-occurrence determination of GS and potential associated genes in 5614 genomes using an orthologue-based approach. A Pearson coefficient is calculated to determine correlation between GS gene and the listed orthologue clusters: a value of +1 indicates complete correlation and a value of -1 indicates anti-correlation.

Orthologue Cluster	#GS Genomes	#OC genomes	Shared Genomes	Pearson Coefficient	Putative or Known Function
OC.1344056	114	114	114	1.00	Geosmin synthase
OC.1340183	114	159	108	0.798	Methyltransferase
OC.316564	114	154	96	0.718	Polyprenyl synthetase family protein
OC.158342	114	122	83	0.697	Phosphoribosylamine-glycine ligase
OC.1321757	114	122	82	0.689	Phosphoribosylamine-glycine ligase
OC.671790	114	95	67	0.637	Hypothetical protein / TIGR02452 family protein
OC.253916	114	110	72	0.636	Pyridine nucleotide-disulfide oxidoreductase
OC.1631	114	268	110	0.619	Hybrid sensor histidine kinase/response regulator
OC.288250	114	301	105	0.554	Hypothetical protein / sulfatase-modifying factor 1
OC.1238530	114	360	113	0.545	Crp/Fnr family transcriptional regulator
OC.26486	114	216	86	0.536	Hypothetical protein / polyketide synthase
OC.588757	114	248	91	0.528	Methyltransferase
OC.664721	114	392	114	0.525	Molybdopterin synthase sulfur carrier subunit
OC.1338031	114	81	51	0.523	Hypothetical protein
OC.99118	114	501	96	0.380	1-deoxy-D-xylulose-5-phosphate synthase

vi. Base-pair and evolutionary analysis of GS

i. Geosmin synthase isn't associated with regions of genome plasticity

Given that GS has undergone lateral gene transfer (*vide supra*), we used the IslandViewer 4 interface to examine the genomes of *M. xanthus* DK1622 (Figure 42), *S. coelicolor* A3(2) (Figure A2A), *K. albida* DSM 43870 (Figure A2B), *N. punctiforme* PCC 73102 (Figure A2C), *C. acidiphila* DSM 44928 (Figure A2D), *C. crocatus* Cm c5 (Figure A2E), *M. rosea* DSM 24000 (Figure A2F), *H. ochraceum* DSM 14365 (Figure A2G) and *S. nassauensis* DSM 44728 (Figure A2H) for evidence of genome plasticity. These genomes were selected due to their unique branching in the GS and 16S rRNA phylogenetic trees (Figure 40) or as representative strains (*M. xanthus* DK1622 and *S. coelicolor* A3(2)). The presence of integrases, transposases, unusual GC content, flanking repeats, tDNA, rRNA and tRNA genes (as phage integration sites) and an evaluation of codon usage are all features associated with genome islands that are detected by IslandViewer 4¹²⁵⁻¹²⁸. However, none of these genomes showed evidence of genome plasticity in or around GS (Figure 42). This suggests that acquisition of GS by these organisms occurred in the distant past, allowing for deletion of mobile genetic elements and harmonization of the GC content with the rest of the genome.

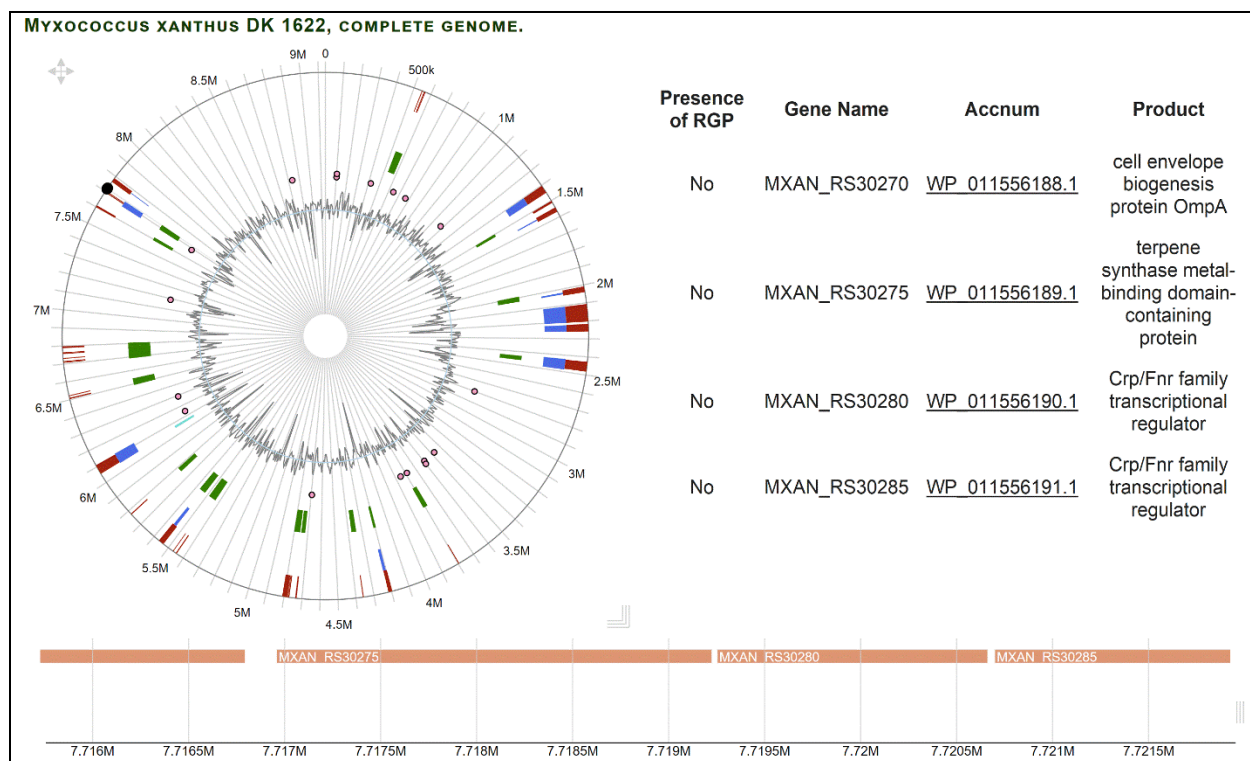


Figure 42: Analysis of RGP in *M. xanthus* DK1622. Full circular genome with RGP indications (upper left), information window of selected region containing GS gene as terpene synthase metal-binding domain-containing protein (upper right) and close-up of linear DNA region containing geosmin synthase as MXAN_RS30275 (lower center) adjacent to two CNBPs. Recent indication of RGP are absent around GS gene indicating HGT that occurred in the distant past.

ii. Geosmin Synthase is Conserved.

Fisher's Exact Test was then used to determine if GS was under positive or purifying selection. This test compares the number of synonymous (dS) and non-synonymous substitutions (dNS) between geosmin producers to determine either positive or purifying selection of GS¹²⁹. All cyanobacterial strains exhibited strong purifying selection (Additional file 9), suggesting GS is a key determinant of bacterial fitness in these organisms. Similarly, 99.25% of actinobacterial and 89.90% of myxobacterial strains revealed a purifying selection as well, with p-values higher than 0.05 (Additional file 7-8). Furthermore, 78.79% of actinobacteria and 36.67% of myxobacteria have p-values higher than 0.95, indicating strong purifying selection. However, 0.75% and 10.10% of actinobacterial and myxobacterial strains respectively, showed p-values lower than 0.05, indicating positive selection of GS. These strains have more than one copy of the GS gene contained in their genome, allowing mutation of one without compromising overall GS function¹³⁰. In line with its broad prevalence in these bacterial clades (Figure 40A), the overall structure of GS is heavily conserved.

The relative synonymous codon usage (RSCU) of GS among cyanobacteria, myxobacteria and actinobacteria was calculated to further understand the evolutionary relationship of the terpene synthase among these families (Figure 43). The RSCU of cyanobacteria is mostly stable around a value of 1, while that of actinobacteria fluctuates from 0-2.8. The RSCU of myxobacteria in our dataset varies between 0-2.3 and has similar values with both cyanobacteria and actinobacteria, in contrast to previous analyses²⁰ (Figure 43).

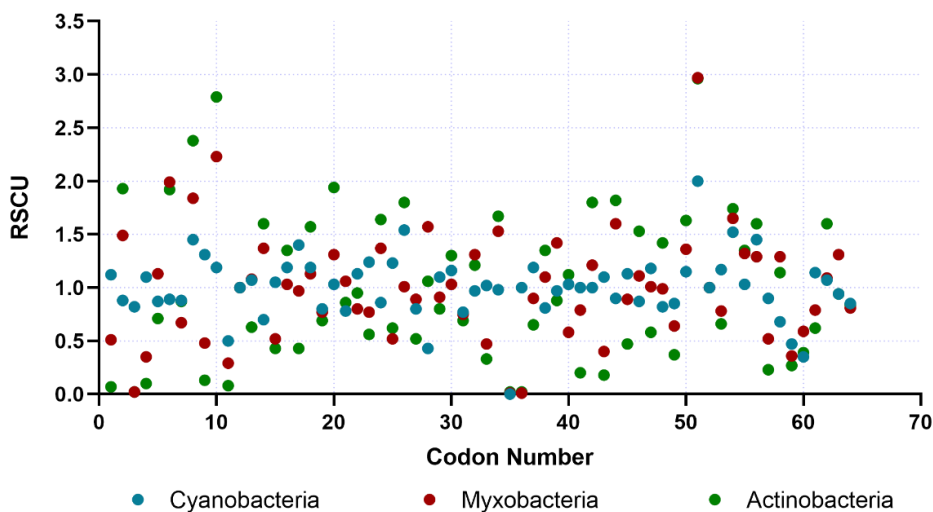


Figure 43: RSCU of GS gene among 23 cyanobacterial, 25 myxobacterial and 257 actinobacterial strains. The analysis was made using MEGA-X using the compute codon usage bias module offered in MEGA-X.

vii. *Geosmin synthase is laterally transferred between unrelated genomes.*

From the Fisher's Exact Test, the codon bias analysis and the GS phylogenetic tree (Figure 40A), significantly less diversity in cyanobacterial and myxobacterial GSs is observed compared to actinobacterial GSs. The high diversity of RSCU of GS among actinobacteria, potentially caused by an expected increase in mutation rates over time, suggests that they could be the originators of geosmin. In the phylogenetic tree, myxobacterial GSs are divided into two sub-clades and *Archangium gephyra* DSM 2261 aligns with the actinobacteria *Nocardia terpenica* NC YFY NT001. However, actinobacterial GSs are divided into three sub-clades (Figure 40A) showing more diversity among their GS genes. When we look at the co-occurrence of GS adjacent genes (Table 10-11), cyanobacteria and myxobacteria have similar and conserved patterns of adjacent genes around GS including 2 CNBPs genes. By contrast, actinobacteria do not frequently have similar co-occurring genes or CNBPs adjacent to GS. The diversity of GS adjacent genes among actinobacteria (Table 10-11), their large number of clades in the phylogenetic tree (Figure 40A) and their presence across subfamily borders suggests that geosmin may have originated from *Actinobacteria*.

3. Conclusions

Geosmin is detected by the predatory nematode *C. elegans* through the gustatory neuron ASE. Geosmin is non-toxic to this species, but nematodes strongly avoid geosmin-producing bacteria. When geosmin production was eliminated or the ASE neuron was disabled, *C. elegans* became coated in bacterial spores and ingested toxic secondary metabolites. Concurrently, bacterial fitness declined through nematode feeding and the conversion of vegetative cells to senescent spores. Geosmin thus acts as an aposematic signal, honestly and reliably advertising the unpalatability of its producers and providing a mutual benefit to predator and prey. Geosmin is the first warning chemical to be identified in bacteria, and it not only shapes bacterial predator-prey interactions but also appears to mediate interactions between eukaryotes and bacteria across the globe.

In agreement with the conclusion that geosmin is a universal warning chemical, GS was found to be broadly distributed in terrestrial, freshwater and marine environments through ancient HGT events, highlighting its strong fitness and evolutionary advantage. Through the performed bioinformatics analysis, their large number of GS clades in the phylogenetic tree along with their diverse co-occurring genes, Actinobacteria seems to be the ancestral genetic reservoir of GS. However, GS BGCs were absent in eukaryotes that produce geosmin^{3,112}, suggesting convergent evolution of the odorous sesquiterpene with an alternative genetic reservoir. While GS was always adjacent to two CNBPs in cyanobacteria and myxobacteria, in actinobacteria we found that CNBPs were majorly adjacent to another terpene synthase, which indicates their importance for terpene biosynthesis. Furthermore, we discovered a vast number of GS co-occurring genes additionally to CNBPs, which infer strong association to the function of geosmin. The study of these genes could be pivotal to our knowledge on the biosynthesis, regulation and importance of terpenes in prokaryotes like geosmin, the warning chemical.

4. Materials and Methods

a. Strains and cultivation

i. Bacteria

Klebsiella aerogenes ATCC13048 and *Chromobacterium violaceum* ATCC 12472 were acquired from the American Type Culture Collection (ATCC). *Escherichia coli* MG1655 and *Burkholderia thailandensis* E264 were gifts of Eric Déziel, INRS. *Myxococcus xanthus* DK1622, *Micrococcus luteus* DSM20030 and *Bacillus subtilis* DSM10 were obtained from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures. *Streptomyces coelicolor* M145, J3003 and J2192 originated from the John Innes Centre, Norwich, UK, and were gifts from Dr. Klas Flärdh. *Escherichia coli* OP50 was obtained from the Caenorhabditis Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

M. luteus, *B. subtilis*, *B. thailandensis*, *Chromobacterium violaceum*, *K. aerogenes* and *E. coli* were grown in Luria-Bertani (LB) media, at 30 °C (37°C for *E. coli* and RT for *C. violaceum*) rotating at 225 rpm for liquid cultures. Bacterial isolates were streaked on 1.5% agar LB plates and placed at 30°C for 24 hrs (37°C for 16 hrs for *E. coli* and RT for 48 hrs for *C. violaceum*) prior to experiments. *M. xanthus* DK1622 was grown in 1% CTT at 30°C and 225 rpm. Bacterial isolates were streaked on 1.5% agar CTT plates and placed at 30°C for 3 days prior to experiments. *Streptomyces coelicolor* M145, J3003 and J2192 were grown at 30 °C in tryptic soy broth (TSB) rotating at 225 rpm and isolates were streaked on 1.5% agar Tryptic Soy Agar (TSA). Growth curves for *M. xanthus* DK1622 were generated by performing daily OD₆₀₀ measurements using a Varian Cary 100 Bio UV-vis spectrophotometer.

ii. Amoebae

Naegleria gruberi ATCC 30223 was acquired from ATCC. *N. gruberi* ATCC 30223 was grown on an ATCC medium 997 (fresh water amoeba) containing an overnight culture of *K. aerogenes* ATCC 13048 at room temperature.

iii. Nematodes, *Caenorhabditis elegans*

The *C. elegans* lineages were maintained on nematode growth medium (NGM) plates with *E. coli* OP50 at 20°C as per standard protocol¹³¹. The wild type N2 and mutant strains BR5514 (*tax-2(p671);tax-4(p678)*), CE1258 (*eat-16(ep273)*), CX2065 (*odr-1(n1936)*), CX2205 (*odr-3(n2150)*), CX5893 (*kyIs140 I; ceh-36(ky646)*), NL2105 (*gpa-3(pk35) odr-3(n1605)*) and PR674 (*che-1(p674)*) were obtained from the Caenorhabditis Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). Gravid nematodes were age synchronized and cleaned from bacterial and fungal contaminants using a bleaching mixture (2.5%

NaClO, 0.5 M NaOH) before each experiment, as previously described¹³¹. All *C. elegans* experiments were conducted using a standard stereomicroscope.

b. Materials

Chemicals. (±)-Geosmin standard was obtained from Sigma-Aldrich, (-)-geosmin was purchased from FUJIFILM WAKO chemicals. Bovine serum albumin, levamisole, (1R)-(+)-camphor and the 3.0 M methylmagnesium bromide solution in diethyl ether (189898), trans-cinnamaldehyde (C80697), vanillic aldehyde (V1104), (S)-(-)-limonene (218367) and (-)- α -pinene (P45702) were all purchased from Sigma-Aldrich. FTC-casein and trypsin proteins were both obtained from ThermoFisher Scientific. Methanol, ethanol, ethyl acetate, tetrahydrofuran (THF), hexane and 2-butanone were acquired from ACS Chemicals and Fisher Scientific.

Media. 1% CTT (1% casitone, 10 mM Tris-HCl [pH 7.6], 8 mM MgSO₄, 1 mM KH₂PO₄) was prepared from scratch. LB (10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract) was prepared from a premix, which was purchased from Bio Basic. TSB (17 g/L casein peptone, 3 g/L soya peptone, 5g/L NaCl, 2.5 g/L K₂HPO₄, glucose 2.5 g/L [pH 7.3]) premix was purchased from Sigma-Aldrich. AS was prepared as per ATCC medium 1323 (Page's amoeba saline)^{132,133}. NGM (3 g NaCl, 2.5 g peptone 20 g agar, 1 mL of 5 mg/mL cholesterol in ethanol, 1 mL of 1M MgSO₄, 25 mL of 1M [pH 6.0] KPO₄ in 1L H₂O) was made from scratch. Worm M9 buffer (3 g/L KH₂PO₄, 6 g/L Na₂HPO₄, 5 g/L NaCl) was made from scratch. The TBS buffer used in the protease assays was made from scratch (25 mM tris, 0.15 M NaCl [pH 7.2]). The phosphate buffered saline (PBS) used in the CD assays as also prepared from scratch (pH 7.0, ionic strength 0.014 M).

c. Growth phase monitoring experiment

E. coli MG1655 was washed 3 times with AS, diluted to 0.5 McFarland standard (1.5 x 10⁸ CFU/mL) before adding 100 μ L of the mixture to a 125 mL Erlenmeyer flask. Either 0.1 ppm (±)-geosmin standard diluted in methanol (100 μ L of a 0.1 mg/mL geosmin standard) or 100 μ L of methanol were added. The mixture was diluted to 10 mL of AS and placed in a 37°C incubator at 225 rpm for 4 hrs. Each experiment was done in triplicate (3 (±)-geosmin experiments and 3 methanol controls). 100 μ L of each flasks was diluted 10⁻⁴ or 10⁻⁵ with AS. 100 μ L from each diluted mixture was then spread on LB agar and placed in a 37°C incubator overnight. CFU counts were made the next day.

d. Geosmin extraction and GC-MS quantitation

A culture of *M. xanthus* DK1622 in stationary phase was diluted to an OD₆₀₀ of 0.125, then diluted 1:100 in fresh 1% CTT. Samples were left shaking at 30 °C and 225 rpm. Aliquots were drawn every 12-24 hrs until day 9. OD₆₀₀ measurements were made on a Varian Cary 100 Bio UV-

vis spectrophotometer. Samples over an OD₆₀₀ of 1.0 were diluted in fresh 1% CTT media to the 0.010-0.99 range. After 7 days incubation clumping was observed and prior to measurements, cells were dispersed via passage through a serological pipette. Geosmin extractions were made using a 1:1 ethyl acetate (EtOAc) extraction with *M. xanthus* DK1622 bacterial culture. Ethyl acetate samples were sonicated, then centrifuged for 1 min at 3000 g to collect clean supernatant before injection for quantification by gas chromatography-mass spectrometry (GC-MS). To quantify media versus cytoplasmic geosmin the bacterial culture was centrifuged at 3000 g for 2 min. The supernatant was used to measure extracellular geosmin as detailed above, while the intracellular/cytoplasmic geosmin concentration was measured by exposing the pellet of *M. xanthus* cells to ethyl acetate, vortexing for 1 min, then sonicating and centrifuging down for 1 min at 3000 g. The GC-MS system (7890B GC coupled to a 5977B MS, Agilent Technologies) was equipped with an autosampler and a split/splitless inlet kept at 300°C. Splitless injections (1.0-3.0 µL) were made on a 60-m DB-EUPAH (0.25 mm ID x 25 µm film thickness; Agilent Technologies) column with the oven kept isothermal (80 °C) for 8 min, ramped to 300 °C at 15 °C/min, and then held at that temperature for 5 min. The inlet was kept at 300 °C throughout. Helium flow rate was 1.2 mL/min, with a He septum purge flow of 5 mL/min. Seven-level external calibration curves between 0.01 and 1.00 mg/L were used for quantitation.

e. Geosmin MIC determination

Following CLSI guidelines for direct colony suspension testing¹³⁴, bacterial cultures were transferred to LB broth and adjusted to a final turbidity equivalent to a 0.5 McFarland standard (1.5 X 10⁸ CFU/mL). Bacteria were then mixed 1:1 with (±)-geosmin in 96-well plates, then incubated at 30 °C for 20-24 hr. Methanol was used as a control. The MIC was defined as the concentration sufficient to inhibit bacterial growth as evaluated by the naked eye. After initial tests at concentrations similar to the level produced by *M. xanthus* DK1622 failed to inhibit growth the quantity was increased, until the methanol used to solubilize the high geosmin concentrations began to impede growth.

f. CD analysis

CD analyses were made using a Jasco J-715 Spectropolarimeter. The cuvette width was 0.2 cm. The BSA concentration was 1.25 µM for all experimental and control assays. Fresh protein samples were prepared before each experiment and kept on ice. The (-)-geosmin concentration was 0.18 mM for the experimental assay and the SDS concentration was 0.75 mM. All samples were diluted in PBS^{73,135}. Subtractions and smoothing were made using Jasco J-715 software using controls of each chemical in PBS exclusively. Elevated temperature analysis was made at 25, 50 and 80°C. The wavelength window measured was from 200-240 nm. Five scans were made for each measurement. Experiments were run in triplicate before data analysis using the Jasco J-715 instrument. Varying temperature analysis was made at a wavelength of 208.6 nm and a heating/cooling rate of 1°C/min.

g. Protease activity assay

A Pierce™ Fluorescent Protease Assay Kit was purchased from Thermo Scientific™. The kit contained FTC-casein, n-tosyl-L-Phenylalanine chloromethyl ketone (TPCK) treated trypsin standard and a Tris-buffered saline (TBS) buffer pack⁷⁶. As per the product sheets instructions⁷⁶, FTC-casein was diluted a 5 mg/mL stock solution with ultrapure water. One 20 µL aliquot of the stock solution was diluted 1:500 in TBS to a final volume of 10 mL and concentration of 0.01 mg/mL. A 20 µL aliquot of trypsin stock solution (TSS) of 50 mg/mL was diluted to 1 mg/mL in TBS. To obtain the extracellular proteases of *M. xanthus* DK1622 absent geosmin the supernatant of a 10-day culture was extracted with EtOAc and the organic layer discarded. A 50% protease solution in TBS buffer was used and was kept in ice with FTC-casein and trypsin. Each fluorometric assay was measured using a Varian Cary Eclipse Fluorescence Spectrophotometer at $\lambda_{\text{excitation}}=494$ nm, and $\lambda_{\text{emission}}=521$ nm. The experiments were made with and without the presence of (±)-geosmin (1 mg/L) to evaluate the effect of (±)-geosmin on the activity of the proteases. Methanol was used as a negative control.

h. Violacein assay

C. violaceum ATCC 12472 was used to measure the presence of violacein as a QS marker as reported by Ahmad, A. et al⁵¹. LB was used as broth and media in the 96-well plates and agar broth dilution violacein assays respectively.

i. Using 96-well plates

For this assay, CLSI guidelines for direct colony suspension testing¹³⁴ were followed, with adaptations for violacein and MIC detection. The positive controls used for anti-QS results were vanillin, trans-cinnamaldehyde and (-)-limonene, the negative control used for QS enhancement was (-)- α -pinene. Stocks of 0.1 mg/mL of each terpene were prepared except for (±)-geosmin. *C. violaceum* ATCC 12472 bacterial culture was adjusted to a turbidity equivalent to a 0.5 McFarland standard (1.5×10^8 CFU/mL) then diluted to 1.5×10^6 CFU/mL and mixed 1:1 with (±)-geosmin (2 mg/mL) for row A, methanol for row B (50 µL), row C, D, E, F were respectively made from 2 µL of vanillin, trans-cinnamaldehyde, (-)-limonene and (-)- α -pinene in 998µL of the bacterial culture, all in 96-well plates. These plates were then incubated at 25 °C for 24-48 hrs, until visualization of purple colour. Inhibition of purple pigmentation with presence of bacterial growth observed by the naked eye is evidence for anti-QS properties⁵¹. Inhibition of bacterial growth shows the MIC¹³⁴.

ii. Violacein agar broth dilutions

Following CLSI guidelines for broth microdilution¹³⁶, 50 µL of *C. violaceum* ATCC 12472 adjusted to a final turbidity equivalent to a 0.5 McFarland standard (1.5×10^8 CFU/ml) was added

to 5 mL of melted 37 °C LB agar media. The mixture was then placed on a 5 cm diameter petri dish and cooled until solidification of the media. Six other plates were made to evaluate the effect of (±)-geosmin and (-)-geosmin on QS by *C. violaceum* each containing respectively: 10 ppm of (±)-geosmin, 25 µL of methanol, 2.5 µg/mL of vanillin, 2.5 µg/mL of trans-cinnamaldehyde, 2.5 µg/mL of (-)-limonene and 2.5 µg/mL of (-)-α-pinene. Violacein inhibition and MIC were detected after 24-48 hr from observation by the naked eye as described above^{51,134,136}.

i. Predation assays using amoeba cells

i. Killing assay in liquid media

The liquid media killing assay was adapted from Xiao, Y. et al⁴⁰. Overnight *E. coli* MG1655 and amoeba *N. gruberi* ATCC 30223 cultures were diluted to 0.5 McFarland standard (1.5 X 10⁸ CFU/ml). 100 µL of the bacterial and amoeba cultures were diluted to 10 mL with AS in a 125 mL erlenmeyer flask, with or without the presence of 0.1 ppm of (±)-geosmin for a total of 3 experimental assays with (±)-geosmin and 3 control assays. The flasks were incubated at 37°C and 225 rpm for 2 h, allowing amoeba cells to predate on *E. coli* cells. 100 µL of each flask was serially diluted 10⁻⁴ or 10⁻⁵ with AS. 100 µL of the diluted mixtures were spread on LB plates containing either epothilone¹³⁷ or amphotericin B¹³⁸ to kill amoeba cells. The plates were wrapped in aluminum foil and placed in a 37°C incubator overnight. CFU counts of surviving *E. coli* cells were made the next day.

ii. Killing assay on solid media

The killing assay on solid media was adapted from Xiao, Y. et al⁴⁰. 10 µL of amoeba *N. gruberi* ATCC 30223 of OD₆₀₀ of 0.7-0.8 in AS broth was added to an AS plate and incubated at 25 °C for 2 hrs before careful addition of 5 µL of *E. coli* MG1655 (OD₆₀₀, 10) directly on top of the amoeba swarm. The same procedure was made with an addition of a concentration of 0.1, 1 or 10 ppm of (±)-geosmin to the *E. coli* culture. All experiments were made in sextuplets for all three (±)-geosmin experiments and the negative control for a total of 24 spots. The plates were placed upright at room temperature for 2 days. After 24 hrs and 48 hrs, 3 spots of each experiment were scraped, washed with AS, and centrifuged for 2 min. The pellets were then resuspended in 1 mL of AS and serially diluted and grown on LB plates at 37°C, containing either epothilone or amphotericin B to quantify surviving *E. coli* prey cells.

iii. Swarm assay on solid media

The swarm assay on solid media was adapted from Xiao, Y. et al⁴⁰. 50 µL *E. coli* MG1655 (OD₆₀₀, 10) containing 0.1, 1 or 10 ppm of geosmin were pipetted onto separate AS agar plates. A control plate without any geosmin was also prepared and each assay was made in triplicates for a total of 12 plates. All plates were incubated overnight at 37°C. 10 µL of a washed culture of amoeba *N. gruberi* ATCC 30223 (OD₆₀₀, 4) in AS was added adjacently to the *E. coli* lawn. Plates

were dried by a flame for 30 min before incubation in a sealed Ziploc containing a humid paper towel. Migration of the amoeba cells into the *E. coli* lawn was measured once a week for four weeks.

j. Chemotaxis assay with *C. elegans*

Chemotaxis experiment was made following Bargmann et al⁷⁸. On an NGM plate 1 μ L of (-)-geosmin at the designated concentration was placed at 1 cm from one end of the plate, and on the other symmetrical opposite end, 1 μ L of control was added (nanopure water). 1 μ L of 1 M levamisole was added to each drop before drying them adjacent to a flame for 20-30min. Pure 2-butanone was used as a positive control⁶⁶. N2 adult *C. elegans* were washed 3 times in worm M9 buffer before addition at the centre of the plates (n=50). The plates were then incubated at room temperature for 1 hr before chemotaxis index calculation.

k. Behavioural assay with *C. elegans*

NGM agar with and without 0.54 μ g/mL (-)-geosmin was poured into 24 well plates. Three adult *C. elegans* were added to each well and were video-taped for 10 min using a system that uses NIS Imaging BR version 3 software hooked up to a DSFi1c camera on a NIKON SMZ1500 microscope. The track line % (TL% = $\frac{\text{Distance from first to current point (D2S)}}{\text{Length from first to current point (Len)}} * 100$) was calculated using Imaris 9.5¹³⁹, data processed with a custom-built python script (Additional file 15). Briefly, the script read the .xls files produced by Imaris particle tracking analysis, split the data into five two-minute segments, then extracted results for the track length and displacement of each worm for track line percentage determination. Videos were analysed using WormLab software 2020.1.1¹⁴⁰ to generate the peristaltic speed (μ m/s) and the head movement periodicity (μ m) of the worms. Peristaltic speed is defined as peristaltic track length, the length of the track made by the worm during its movement, divided over time. Head movement periodicity is the wavelength of the sinusoidal wave created by tracing the head of the worms as they crawl. Worms that lodged in crevices and ceased moving were removed prior to analysis. Data was processed using Microsoft Excel 2011. Experiments were conducted in triplicate.

l. Predation assay with *C. elegans*.

40 μ L of *S. coelicolor* (OD₆₀₀ = 0.4-0.6) or *M. xanthus* (OD₆₀₀ = 0.2-0.3) was added to the center of a 5 cm diameter NGM plate. The liquid was allowed to dry and the plates were then incubated at 30°C (7 days for *S. coelicolor* and 3 days for *M. xanthus*). Ten adult hermaphrodite *C. elegans* were then added at 2-5 mm away from the bacterial colony. Quantification of the worms inside and outside of the bacteria was made at T= 0, 2, 4 and 24 hrs. Each experiment was conducted in at least triplicate.

m. Terpene add-in assay.

Add-in experiments were made using *S. coelicolor* J2192, following the same procedure as in the predation assays with the following modification. One hr prior to the addition of *C. elegans* 40 μ L of (-)-geosmin (2.25 μ g/mL), (-)-2-MIB (40 μ g/mL), or both was added to the bacterial colony and allowed to dry. Autoclaved nanopure H₂O was used as a negative control. Triplicates of each added compound were performed with the addition of 10 adult hermaphrodite N2 *C. elegans* (n=30). Quantification was performed as per the predation assay.

n. *C. elegans* lethality assay.

Solutions of (-)-geosmin (0.54, 5.4, 54 μ g/mL) and 2-MIB (1.5, 15, 150 μ g/mL) in LB broth were added to 24-well plates containing either NGM or NGM with *E. coli* OP50 as per H. Xiong et al¹⁴¹. LB broth was used as a negative control and a sodium hypochlorite solution (2.625 mg/mL) in LB broth was used as a positive control. The plates were incubated overnight at RT. Age synchronized plates of adult N2 worms were then washed in worm M9 buffer, dried on NGM plates, then added to each well (n = 5). Lack of response to touch stimuli was used to presume death. The number of dead *C. elegans* was quantified at T = 0, 2, 4 and 24 hrs. Experiments were run in triplicate.

o. Synthesis of 2-MIB.

To a solution of (1R)-(+)-camphor (3.3 mmols) in anhydrous tetrahydrofuran (0.2 M) at 0 °C was added methylmagnesium bromide (4.95 mmols, 3 M). The reaction was then heated to reflux for 10 hrs, before being quenched with a saturated solution of ammonium chloride in water. The THF was removed *in vacuo* and the remaining solution was extracted three times with EtOAc. The organic layers were then pooled and extracted with brine, then dried with sodium sulfate. The organic solvent was then removed *in vacuo* to give a white solid. (325.8 mg, 58.70 %). The crude mixture was then purified by flash chromatography (12% EtOAc in hexanes) to give the title compound as a white solid (48.7 mg, 8.77 % yield). ¹H NMR (500 MHz, CDCl₃) δ 2.07 (dt, 1H, *J* = 3.72 Hz, 13.08 Hz), 1.70 (m, 2H), 1.39 (m, 4H), 1.24 (s, 3H), 1.10 (s, 3H), 0.87 (s, 3H), 0.85 (s, 3H). ¹³C NMR (125 MHz, CDCl₃) δ 79.6, 51.9, 48.9, 47.3, 45.4, 31.3, 27.0, 26.8, 21.4, 21.2, 9.9.

p. Statistical analysis.

All statistical analyses were performed using student T-test calculations with two-tailed distribution and unequal variance. Statistical significance of $P < 0.05$ was noted as a red box in Table 9 and as * for $P < 0.05$ or ** for $P < 0.001$ in Figure 39A-D. All replicates are distinct samples (biological replicates). Data was visualized with GraphPad Prism 9.0.

q. Bioinformatics analysis of GS

i. Bacterial genome selection

The full set of genomes with full sequences, which contained GS were retrieved from NCBI using GenBank and the BLAST algorithm (www.blast.ncbi.nlm.nih.gov/Blast.cgi) with an e value cutoff of 10^{-20} . A combination of BLASTn and BLASTp were used to identify the nucleotide and amino acid sequences of GS. Full genomes, GS and 16S rRNA were downloaded as FASTA files. Additional file 1 shows the bacterial strains used in this study along with their respective genome and GS NCBI accession numbers and GS and 16S sequences.

ii. Phylogenetic analyses of 16S rRNA and GS

Phylogenetic analyses of 64 actinobacteria, 23 myxobacteria, 19 cyanobacteria, 1 pseudomonad and 1 ktedonobacteria were performed on both GS and 16S rRNA. Base pair sequences were aligned using default parameters through MEGA-X using MUSCLE. Phylogenetic trees were constructed using the maximum likelihood analysis method through MEGA-X with shown bootstrap values using the bootstrap method with 500 replicons following the Tamura-Nei model. Additional file 1 contains the FASTA files used for the alignments of GS and 16S rRNA. The GS phylogenetic tree was then edited to indicate the presence and quantity of CNBPs in each genome using Interactive Tree Of Life (iTOL) software version 5.7 (<http://itol.embl.de>)¹⁴². Basic figure construction was made using paint.net.

iii. Genomic analysis and general features

The sequences of the two CNBPs in *Myxococcus xanthus* DK1622 (MXAN_6248 and MXAN_6249) were obtained from NCBI. Homologous sequences for these proteins in myxobacteria, actinobacteria, and cyanobacteria were identified and retrieved from GenBank using BLASTp. Strains with CNBPs protein genes were recorded and cross referenced with the list of strains carrying GS. The number of copies of the CNBP was also noted. A BLASTp E value cutoff of 10^{-20} was used. Isolation sites of each strain were obtained from the Bio sample database page of NCBI or noted from the original isolation paper. Environments were categorized as marine, freshwater, or terrestrial/soil. Strains isolated from other organisms were classified as their respectful endosymbionts.

iv. Analysis of GS co-occurrence

The Orthologue Cluster database was downloaded from <https://www.genome.jp/oc/>.¹²³ A python script was then used to scan this database for genomes that contained both GS and each other orthologue cluster in the dataset (Additional file 5). Co-occurrence was assessed through the Pearson correlation coefficient¹⁴³ and the results written to a comma separated variable file (Additional file 6). The average Pearson coefficient for the dataset was 0.00328 (s = 0.0370).

Rstudio was then used to remove all entries with a Pearson coefficient less than 0.25. SPARQL queries (Appendix) and a second python script (Additional file 5) were used to construct a list of all orthologue clusters found in *Myxococcus*, *Nostoc*, and *Kitasatospora* spp., and the GS co-occurrence list was reduced to only the orthologous clusters that were found in all three subspecies (Table 11, Additional file 6). This was done to remove the large number of orthologous clusters that had high Pearson coefficients with GS solely as a result of their ubiquity in one of the three clades. Of the 114 GS-containing organisms in the KEGG OC database, 96 are from the phylum Actinobacteria.

v. *Evolution selection and codon bias analyses of GS*

The relative synonymous codon usage (RSCU) of GS was generated for each bacterial family (25 myxobacterial, 257 actinobacterial and 23 cyanobacterial GS sequences) using the Compute Codon Usage Bias module of MEGA-X. Data was integrated using Microsoft Excel 2011 and visualized using GraphPad Prism 9.0.

A Codon-based Fisher's Exact Test of Selection was made using the selection module equipped in MEGA-X to determine positive or purifying selection of GS (25 myxobacterial, 257 actinobacterial and 23 cyanobacterial GS sequences). The number of non-synonymous substitutions (dN) and synonymous substitutions (dS) is generated. If $dS > dNS$, the Fisher's Exact Test reports a value of $P = 1$ indicating purifying selection where there are no significant changes in the amino acid sequence of the GS gene, thus conserving the overall GS protein structure. If $dNS > dS$ the Fisher's Exact Test reports a value of $P < 0.05$ indicating positive selection where there are significant changes within amino acid sequence of the GS gene, thus changing the overall protein structure.

IslandViewer 4 interface was used to examine the genomes of *M. xanthus* DK1622 and *S. coelicolor* A3(2) for evidence of regions of genome plasticity (<http://www.pathogenomics.sfu.ca/islandviewer/>)¹²⁵.

References

- (1) Juttner, F.; Watson, S. B. Biochemical and Ecological Control of Geosmin and 2-Methylisoborneol in Source Waters. *Applied and Environmental Microbiology* **2007**, *73* (14), 4395–4406. <https://doi.org/10.1128/AEM.02250-06>.
- (2) Polizzi, V.; Adams, A.; De Saeger, S.; Van Peteghem, C.; Moretti, A.; De Kimpe, N. Influence of Various Growth Parameters on Fungal Growth and Volatile Metabolite Production by Indoor Molds. *Science of The Total Environment* **2012**, *414*, 277–286. <https://doi.org/10.1016/j.scitotenv.2011.10.035>.
- (3) Maher, L.; Goldman, I. L. Endogenous Production of Geosmin in Table Beet. *HortScience* **2018**, *53* (1), 67–72. <https://doi.org/10.21273/HORTSCI12488-17>.
- (4) Fink, P. Ecological Functions of Volatile Organic Compounds in Aquatic Systems. *Marine and Freshwater Behaviour and Physiology* **2007**, *40* (3), 155–168. <https://doi.org/10.1080/10236240701602218>.
- (5) Yean-Woong You. Sensitive Detection of 2-MIB and Geosmin in Drinking Water <http://hpst.cz/sites/default/files/attachments/5991-1031en-sensitive-detection-2-mib-and-geosmin-drinking-water.pdf> (accessed Apr 1, 2020).
- (6) Berthelot, M.; André, G. Sur l'odeur Propre de La Terre. *Compt. Rend.* **1891**, *112*, 598–599.
- (7) Gerber, N. N.; Lechevalier, H. A. Geosmin, an Earthy-Smelling Substance Isolated from Actinomycetes. *Appl. Microbiol.* **1965**, *13* (6), 935–938.
- (8) National Center for Biotechnology Information. Geosmin, CID=29746. <https://pubchem.ncbi.nlm.nih.gov/compound/Geosmin> (accessed Apr 1, 2020).
- (9) National Center for Biotechnology Information. Water, CID=962. <https://pubchem.ncbi.nlm.nih.gov/compound/Water> (accessed Apr 1, 2020).
- (10) Miziorko, H. M. Enzymes of the Mevalonate Pathway of Isoprenoid Biosynthesis. *Archives of Biochemistry and Biophysics* **2011**, *505* (2), 131–143. <https://doi.org/10.1016/j.abb.2010.09.028>.
- (11) Dickschat, J. S.; Bode, H. B.; Mahmud, T.; Müller, R.; Schulz, S. A Novel Type of Geosmin Biosynthesis in Myxobacteria. *J. Org. Chem.* **2005**, *70* (13), 5174–5182. <https://doi.org/10.1021/jo050449g>.
- (12) Rohmer, M.; Rohmer, M. The Discovery of a Mevalonate-Independent Pathway for Isoprenoid Biosynthesis in Bacteria, Algae and Higher Plants†. *Nat. Prod. Rep.* **1999**, *16* (5), 565–574. <https://doi.org/10.1039/a709175c>.
- (13) Burke, C. C.; Wildung, M. R.; Croteau, R. Geranyl Diphosphate Synthase: Cloning, Expression, and Characterization of This Prenyltransferase as a Heterodimer. *Proceedings of the National Academy of Sciences* **1999**, *96* (23), 13062–13067. <https://doi.org/10.1073/pnas.96.23.13062>.
- (14) Thulasiram, H. V.; Poulter, C. D. Farnesyl Diphosphate Synthase: The Art of Compromise between Substrate Selectivity and Stereoselectivity. *J. Am. Chem. Soc.* **2006**, *128* (49), 15819–15823. <https://doi.org/10.1021/ja065573b>.

- (15) Jiang, J.; He, X.; Cane, D. E. Biosynthesis of the Earthy Odorant Geosmin by a Bifunctional Streptomyces Coelicolor Enzyme. *Nat Chem Biol* **2007**, *3* (11), 711–715. <https://doi.org/10.1038/nchembio.2007.29>.
- (16) Martín-Sánchez, L.; Singh, K. S.; Avalos, M.; van Wezel, G. P.; Dickschat, J. S.; Garbeva, P. Phylogenomic Analyses and Distribution of Terpene Synthases among *Streptomyces*. *Beilstein J. Org. Chem.* **2019**, *15*, 1181–1193. <https://doi.org/10.3762/bjoc.15.115>.
- (17) Becher, P. G.; Verschut, V.; Bibb, M. J.; Bush, M. J.; Molnár, B. P.; Barane, E.; Al-Bassam, M. M.; Chandra, G.; Song, L.; Challis, G. L.; Buttner, M. J.; Flärdh, K. Developmentally Regulated Volatiles Geosmin and 2-Methylisoborneol Attract a Soil Arthropod to Streptomyces Bacteria Promoting Spore Dispersal. *Nat Microbiol* **2020**. <https://doi.org/10.1038/s41564-020-0697-x>.
- (18) Perlova, O.; Gerth, K.; Kuhlmann, S.; Zhang, Y.; Müller, R. Novel Expression Hosts for Complex Secondary Metabolite Megasyntetases: Production of Myxochromide in the Thermophilic Isolate Corallococcus Macrosporus GT-2. *Microb Cell Fact* **2009**, *8* (1), 1. <https://doi.org/10.1186/1475-2859-8-1>.
- (19) Reyes-Lamothe, R.; Sherratt, D. J. The Bacterial Cell Cycle, Chromosome Inheritance and Cell Growth. *Nat Rev Microbiol* **2019**, *17* (8), 467–478. <https://doi.org/10.1038/s41579-019-0212-7>.
- (20) Wang, Z.; Song, G.; Li, Y.; Yu, G.; Hou, X.; Gan, Z.; Li, R. The Diversity, Origin, and Evolutionary Analysis of Geosmin Synthase Gene in Cyanobacteria. *Science of The Total Environment* **2019**, *689*, 789–796. <https://doi.org/10.1016/j.scitotenv.2019.06.468>.
- (21) Dionigi, C. P.; Lawlor, T. E.; McFarland, J. E.; Johnsen, P. B. Evaluation of Geosmin and 2-Methylisoborneol on the Histidine Dependence of TA98 and TA100 Salmonella Typhimurium Tester Strains. *Water Research* **1993**, *27* (11), 1615–1618. [https://doi.org/10.1016/0043-1354\(93\)90125-2](https://doi.org/10.1016/0043-1354(93)90125-2).
- (22) Utkilen, H. C.; Frøshaug, M. Geosmin Production and Excretion in a Planktonic and Benthic Oscillatoria. *Water Science and Technology* **1992**, *25* (2), 199–206. <https://doi.org/10.2166/wst.1992.0053>.
- (23) Stensmyr, M. C.; Dweck, H. K. M.; Farhan, A.; Ibba, I.; Strutz, A.; Mukunda, L.; Linz, J.; Grabe, V.; Steck, K.; Lavista-Llanos, S.; Wicher, D.; Sachse, S.; Knaden, M.; Becher, P. G.; Seki, Y.; Hansson, B. S. A Conserved Dedicated Olfactory Circuit for Detecting Harmful Microbes in Drosophila. *Cell* **2012**, *151* (6), 1345–1357. <https://doi.org/10.1016/j.cell.2012.09.046>.
- (24) Melo, N.; Wolff, G. H.; Costa-da-Silva, A. L.; Arribas, R.; Triana, M. F.; Gugger, M.; Riffell, J. A.; DeGennaro, M.; Stensmyr, M. C. Geosmin Attracts Aedes Aegypti Mosquitoes to Oviposition Sites. *Current Biology* **2020**, *30* (1), 127-134.e5. <https://doi.org/10.1016/j.cub.2019.11.002>.

- (25) Huang, H.; Ren, L.; Li, H.; Schmidt, A.; Gershenzon, J.; Lu, Y.; Cheng, D. The Nesting Preference of an Invasive Ant Is Associated with the Cues Produced by Actinobacteria in Soil. *PLoS Pathog* **2020**, *16* (9), e1008800. <https://doi.org/10.1371/journal.ppat.1008800>.
- (26) Tosi, L.; Sola, C. Role of Geosmin, a Typical Inland Water Odour, in Guiding Glass Eel *Anguilla Anguilla* (L.) Migration. *Ethology* **2010**, *95* (3), 177–185. <https://doi.org/10.1111/j.1439-0310.1993.tb00468.x>.
- (27) Wales, A. D.; Davies, R. H. A Critical Review of *Salmonella* Typhimurium Infection in Laying Hens. *Avian Pathology* **2011**, *40* (5), 429–436. <https://doi.org/10.1080/03079457.2011.606799>.
- (28) Churro, C.; Semedo-Aguiar, A. P.; Silva, A. D.; Pereira-Leal, J. B.; Leite, R. B. A Novel Cyanobacterial Geosmin Producer, Revising GeoA Distribution and Dispersion Patterns in Bacteria. *Sci Rep* **2020**, *10* (1), 8679. <https://doi.org/10.1038/s41598-020-64774-y>.
- (29) Pattanaik, B.; Lindberg, P. Terpenoids and Their Biosynthesis in Cyanobacteria. *Life* **2015**, *5* (1), 269–293. <https://doi.org/10.3390/life5010269>.
- (30) Seto, H.; Hiroyuki, W.; Furihata, K. Simultaneous Operation of the Mevalonate and Non-Mevalonate Pathways in the Biosynthesis of Isopentenyl Diphosphate in *Streptomyces Aeriouvifer*. *Tetrahedron Letters* **1996**, *37* (44), 7979–7982.
- (31) Neff, E. P. Stop and Smell the Geosmin. *Lab Anim* **2018**, *47* (10), 270–270. <https://doi.org/10.1038/s41684-018-0161-1>.
- (32) Du, H.; Lu, H.; Xu, Y. Influence of Geosmin-Producing *Streptomyces* on the Growth and Volatile Metabolites of Yeasts during Chinese Liquor Fermentation. *J. Agric. Food Chem.* **2015**, *63* (1), 290–296. <https://doi.org/10.1021/jf503351w>.
- (33) Rosen, B. H.; MacLeod, B. W.; Simpson, M. R. ACCUMULATION AND RELEASE OF GEOSMIN DURING THE GROWTH PHASES OF ANABAENA CIRCINALIS (KUTZ.) RABENHORST. *Wal. Sci. Tech.* **1992**, *25* (2), 185–190.
- (34) La Guerche, S.; Chamont, S.; Blancard, D.; Dubourdiou, D.; Darriet, P. Origin of (–)-Geosmin on Grapes: On the Complementary Action of Two Fungi, *Botrytis Cinerea* and *Penicillium Expansum*. *Antonie Van Leeuwenhoek* **2005**, *88* (2), 131–139. <https://doi.org/10.1007/s10482-005-3872-4>.
- (35) Urem, M.; van Rossum, T.; Bucca, G.; Moolenaar, G. F.; Laing, E.; Świątek-Połatyńska, M. A.; Willemse, J.; Tenconi, E.; Rigali, S.; Goosen, N.; Smith, C. P.; van Wezel, G. P. OsdR of *Streptomyces Coelicolor* and the Dormancy Regulator DevR of *Mycobacterium Tuberculosis* Control Overlapping Regulons. *mSystems* **2016**, *1* (3), e00014-16, /msys/1/3/e00014-16.atom. <https://doi.org/10.1128/mSystems.00014-16>.
- (36) Li, Z.; Hobson, P.; An, W.; Burch, M. D.; House, J.; Yang, M. Earthy Odor Compounds Production and Loss in Three Cyanobacterial Cultures. *Water Research* **2012**, *46* (16), 5165–5173. <https://doi.org/10.1016/j.watres.2012.06.008>.
- (37) Schrader, K. K.; Blevins, W. T. Geosmin-Producing Species of *Streptomyces* and *Lyngbya* from Aquaculture Ponds. *Can. J. Microbiol.* **1993**, *39*, 834–840.

- (38) Muñoz-Dorado, J.; Marcos-Torres, F. J.; García-Bravo, E.; Moraleda-Muñoz, A.; Pérez, J. Myxobacteria: Moving, Killing, Feeding, and Surviving Together. *Front. Microbiol.* **2016**, *7*. <https://doi.org/10.3389/fmicb.2016.00781>.
- (39) Diez, J.; Martinez, J. P.; Mestres, J.; Sasse, F.; Frank, R.; Meyerhans, A. Myxobacteria: Natural Pharmaceutical Factories. *Microb Cell Fact* **2012**, *11* (1), 52. <https://doi.org/10.1186/1475-2859-11-52>.
- (40) Xiao, Y.; Wei, X.; Ebright, R.; Wall, D. Antibiotic Production by Myxobacteria Plays a Role in Predation. *Journal of Bacteriology* **2011**, *193* (18), 4626–4633. <https://doi.org/10.1128/JB.05052-11>.
- (41) Hense, I.; Beckmann, A. The Representation of Cyanobacteria Life Cycle Processes in Aquatic Ecosystem Models. *Ecological Modelling* **2010**, *221* (19), 2330–2338. <https://doi.org/10.1016/j.ecolmodel.2010.06.014>.
- (42) Manaaki Whenua. Fungal life cycles - spores and more. <https://www.sciencelearn.org.nz/images/3689-mushroom-life-cycle> (accessed Apr 2, 2020).
- (43) Breheret, S.; Talou, T.; Rapior, S.; Bessière, J.-M. Geosmin, a Sesquiterpenoid Compound Responsible for the Musty-Earthy Odor of *Cortinarius Hercules*, *Cystoderma Amianthinum*, and *Cy. Carcharias*. *Mycologia* **1999**, *91* (1), 117–120. <https://doi.org/10.1080/00275514.1999.12060999>.
- (44) Mattheis, J. P.; Roberts, R. G. Identification of Geosmin as a Volatile Metabolite of *Penicillium Expansum*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* **1992**, *58* (9), 3170–3172.
- (45) Chater, K. F. Recent Advances in Understanding *Streptomyces*. *F1000Res* **2016**, *5*, 2795. <https://doi.org/10.12688/f1000research.9534.1>.
- (46) Ingham, E. R. The Soil Food Web https://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/soils/health/biology/?cid=nrcs142p2_053868 (accessed Apr 9, 2020).
- (47) Seipke, R. F.; Kaltenpoth, M.; Hutchings, M. I. *Streptomyces* as Symbionts: An Emerging and Widespread Theme? *FEMS Microbiol Rev* **2012**, *36* (4), 862–876. <https://doi.org/10.1111/j.1574-6976.2011.00313.x>.
- (48) Hayes, S. J.; Hayes, K. P.; Robinson, B. S. Geosmin as an Odorous Metabolite in Cultures of a Free-Living Amoeba, *Vannella* Species (*Gymnamoebia*, *Vannellidae*). *J. Protozool.* **1991**, *38* (1), 44–47.
- (49) Bruger, E.; Waters, C. Sharing the Sandbox: Evolutionary Mechanisms That Maintain Bacterial Cooperation. *F1000Res* **2015**, *4*, 1504. <https://doi.org/10.12688/f1000research.7363.1>.
- (50) Lowery, C. A.; Dickerson, T. J.; Janda, K. D. Interspecies and Interkingdom Communication Mediated by Bacterial Quorum Sensing. *Chem. Soc. Rev.* **2008**, *37* (7), 1337. <https://doi.org/10.1039/b702781h>.

- (51) Ahmad, A.; Viljoen, A. M.; Chenia, H. Y. The Impact of Plant Volatiles on Bacterial Quorum Sensing. *Lett Appl Microbiol* **2015**, *60* (1), 8–19. <https://doi.org/10.1111/lam.12343>.
- (52) Husain, F. M.; Ahmad, I.; Al-thubiani, A. S.; Abulreesh, H. H.; AlHazza, I. M.; Aqil, F. Leaf Extracts of *Mangifera Indica* L. Inhibit Quorum Sensing – Regulated Production of Virulence Factors and Biofilm in Test Bacteria. *Front. Microbiol.* **2017**, *8*, 727. <https://doi.org/10.3389/fmicb.2017.00727>.
- (53) Dunn, P. H.; Barro, S. C.; Poth, M. Soil Moisture Affects Survival of Microorganisms in Heated Chaparral Soil. *Soil Biology and Biochemistry* **1985**, *17* (2), 143–148. [https://doi.org/10.1016/0038-0717\(85\)90105-1](https://doi.org/10.1016/0038-0717(85)90105-1).
- (54) Kmiha, S.; Aouadhi, C.; Klibi, A.; Jouini, A.; Béjaoui, A.; Mejri, S.; Maaroufi, A. Seasonal and Regional Occurrence of Heat-Resistant Spore-Forming Bacteria in the Course of Ultra-High Temperature Milk Production in Tunisia. *Journal of Dairy Science* **2017**, *100* (8), 6090–6099. <https://doi.org/10.3168/jds.2016-11616>.
- (55) Rittershaus, E. S. C.; Baek, S.-H.; Sasseti, C. M. The Normalcy of Dormancy: Common Themes in Microbial Quiescence. *Cell Host & Microbe* **2013**, *13* (6), 643–651. <https://doi.org/10.1016/j.chom.2013.05.012>.
- (56) Dworkin, J.; Shah, I. M. Exit from Dormancy in Microbial Organisms. *Nat Rev Microbiol* **2010**, *8* (12), 890–896. <https://doi.org/10.1038/nrmicro2453>.
- (57) Shah, I. M.; Laaberki, M.-H.; Popham, D. L.; Dworkin, J. A Eukaryotic-like Ser/Thr Kinase Signals Bacteria to Exit Dormancy in Response to Peptidoglycan Fragments. *Cell* **2008**, *135* (3), 486–496. <https://doi.org/10.1016/j.cell.2008.08.039>.
- (58) van den Hoogen, J.; Geisen, S.; Routh, D.; Ferris, H.; Traunspurger, W.; Wardle, D. A.; de Goede, R. G. M.; Adams, B. J.; Ahmad, W.; Andriuzzi, W. S.; Bardgett, R. D.; Bonkowski, M.; Campos-Herrera, R.; Cares, J. E.; Caruso, T.; de Brito Caixeta, L.; Chen, X.; Costa, S. R.; Creamer, R.; Mauro da Cunha Castro, J.; Dam, M.; Djigal, D.; Escuer, M.; Griffiths, B. S.; Gutiérrez, C.; Hohberg, K.; Kalinkina, D.; Kardol, P.; Kergunteuil, A.; Korthals, G.; Krashevskaya, V.; Kudrin, A. A.; Li, Q.; Liang, W.; Magilton, M.; Marais, M.; Martín, J. A. R.; Matveeva, E.; Mayad, E. H.; Mulder, C.; Mullin, P.; Neilson, R.; Nguyen, T. A. D.; Nielsen, U. N.; Okada, H.; Rius, J. E. P.; Pan, K.; Peneva, V.; Pellissier, L.; Carlos Pereira da Silva, J.; Pitteloud, C.; Powers, T. O.; Powers, K.; Quist, C. W.; Rasmann, S.; Moreno, S. S.; Scheu, S.; Setälä, H.; Sushchuk, A.; Tiunov, A. V.; Trap, J.; van der Putten, W.; Vestergård, M.; Villenave, C.; Waeyenbergh, L.; Wall, D. H.; Wilschut, R.; Wright, D. G.; Yang, J.; Crowther, T. W. Soil Nematode Abundance and Functional Group Composition at a Global Scale. *Nature* **2019**, *572* (7768), 194–198. <https://doi.org/10.1038/s41586-019-1418-6>.
- (59) Anacarso, I.; Bondi, M.; Condo, C. Amoebicidal Effects of Three Bacteriocin like Substances from Lactic Acid Bacteria against *Acanthamoeba Polyphaga*. *J Bacteriol Parasitol* **2015**, *6* (s1). <https://doi.org/10.4172/2155-9597.1000201>.

- (60) Alegado, R. A.; Campbell, M. C.; Chen, W. C.; Slutz, S. S.; Tan, M.-W. Characterization of Mediators of Microbial Virulence and Innate Immunity Using the *Caenorhabditis Elegans* Host-Pathogen Model. *Cell Microbiol* **2003**, *5* (7), 435–444. <https://doi.org/10.1046/j.1462-5822.2003.00287.x>.
- (61) Sifri, C. D.; Begun, J.; Ausubel, F. M. The Worm Has Turned – Microbial Virulence Modeled in *Caenorhabditis Elegans*. *Trends in Microbiology* **2005**, *13* (3), 119–127. <https://doi.org/10.1016/j.tim.2005.01.003>.
- (62) Watkins, A. L.; Ray, A.; Roberts, L.; Caldwell, K. A.; Olson, J. B. The Prevalence and Distribution of Neurodegenerative Compound-Producing Soil *Streptomyces* Spp. *Sci Rep* **2016**, *6* (1), 22566. <https://doi.org/10.1038/srep22566>.
- (63) Clarholm, M. Protozoan Grazing of Bacteria in Soil--Lmpact and Importance. 8.
- (64) The Editors of Encyclopaedia Britannica. Phagocytosis <https://www.britannica.com/science/phagocytosis> (accessed Apr 9, 2020).
- (65) Schulz-Bohm, K.; Geisen, S.; Wubs, E. R. J.; Song, C.; de Boer, W.; Garbeva, P. The Prey’s Scent – Volatile Organic Compound Mediated Interactions between Soil Bacteria and Their Protist Predators. *ISME J* **2017**, *11* (3), 817–820. <https://doi.org/10.1038/ismej.2016.144>.
- (66) Hart, A. C.; Chao, M. Y. From Odors to Behaviors in *Caenorhabditis Elegans*. In *The Neurobiology of Olfaction*; Menini, A., Ed.; Frontiers in Neuroscience; CRC Press/Taylor & Francis: Boca Raton (FL), 2010.
- (67) Tran, A.; Tang, A.; O’Loughlin, C. T.; Balistreri, A.; Chang, E.; Coto Villa, D.; Li, J.; Varshney, A.; Jimenez, V.; Pyle, J.; Tsujimoto, B.; Wellbrook, C.; Vargas, C.; Duong, A.; Ali, N.; Matthews, S. Y.; Levinson, S.; Woldemariam, S.; Khuri, S.; Bremer, M.; Eggers, D. K.; L’Etoile, N.; Miller Conrad, L. C.; VanHoven, M. K. C. *Elegans* Avoids Toxin-Producing *Streptomyces* Using a Seven Transmembrane Domain Chemosensory Receptor. *eLife* **2017**, *6*, e23770. <https://doi.org/10.7554/eLife.23770>.
- (68) Cabreiro, F.; Gems, D. Worms Need Microbes Too: Microbiota, Health and Aging in *Caenorhabditis Elegans*. *EMBO Mol Med* **2013**, *5* (9), 1300–1310. <https://doi.org/10.1002/emmm.201100972>.
- (69) Wang, Z.; Shao, J.; Xu, Y.; Yan, B.; Li, R. Genetic Basis for Geosmin Production by the Water Bloom-Forming Cyanobacterium, *Anabaena Ucrainica*. *Water* **2014**, *7* (12), 175–187. <https://doi.org/10.3390/w7010175>.
- (70) Whitbourne, K. What Causes Petrichor, the Earthy Smell After Rain? <https://science.howstuffworks.com/nature/climate-weather/atmospheric/question479.htm>.
- (71) Kim, T. K. T Test as a Parametric Statistic. *Korean J Anesthesiol* **2015**, *68* (6), 540. <https://doi.org/10.4097/kjae.2015.68.6.540>.
- (72) Silhavy, T. J.; Kahne, D.; Walker, S. The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology* **2010**, *2* (5), a000414–a000414. <https://doi.org/10.1101/cshperspect.a000414>.

- (73) Moriyama, Y.; Watanabe, E.; Kobayashi, K.; Harano, H.; Inui, E.; Takeda, K. Secondary Structural Change of Bovine Serum Albumin in Thermal Denaturation up to 130 °C and Protective Effect of Sodium Dodecyl Sulfate on the Change. *J. Phys. Chem. B* **2008**, *112* (51), 16585–16589. <https://doi.org/10.1021/jp8067624>.
- (74) Runnels, L. W.; Scarlata, S. F. Theory and Application of Fluorescence Homotransfer to Melittin Oligomerization. *Biophysical Journal* **1995**, *69* (4), 1569–1583. [https://doi.org/10.1016/S0006-3495\(95\)80030-5](https://doi.org/10.1016/S0006-3495(95)80030-5).
- (75) Twining, S. S. Fluorescein Isothiocyanate-Labeled Casein Assay for Proteolytic Enzymes. *Analytical Biochemistry* **1984**, *143* (1), 30–34. [https://doi.org/10.1016/0003-2697\(84\)90553-0](https://doi.org/10.1016/0003-2697(84)90553-0).
- (76) Pierce Biotechnology. Pierce Fluorescent Protease Assay Kit. Thermo Fisher Scientific Inc. 2014.
- (77) Margie, O.; Palmer, C.; Chin-Sang, I. C. Elegans Chemotaxis Assay. *JoVE* **2013**, No. 74, 50069. <https://doi.org/10.3791/50069>.
- (78) Bargmann, C. I.; Hartwig, E.; Horvitz, H. R. Odorant-Selective Genes and Neurons Mediate Olfaction in *C. Elegans*. *Cell* **1993**, *74* (3), 515–527. [https://doi.org/10.1016/0092-8674\(93\)80053-H](https://doi.org/10.1016/0092-8674(93)80053-H).
- (79) Lans, H.; Rademakers, S.; Jansen, G. A Network of Stimulatory and Inhibitory Gα-Subunits Regulates Olfaction in *Caenorhabditis Elegans*. *Genetics* **2004**, *167* (4), 1677–1687. <https://doi.org/10.1534/genetics.103.024786>.
- (80) Milward, K.; Busch, K. E.; Murphy, R. J.; Bono, M. de; Olofsson, B. Neuronal and Molecular Substrates for Optimal Foraging in *Caenorhabditis Elegans*. *PNAS* **2011**, *108* (51), 20672–20677. <https://doi.org/10.1073/pnas.1106134109>.
- (81) Appleby, P. A. A Model of Chemotaxis and Associative Learning in *C. Elegans*. *Biol Cybern* **2012**, *106* (6), 373–387. <https://doi.org/10.1007/s00422-012-0504-8>.
- (82) Jiang, J.; He, X.; Cane, D. E. Biosynthesis of the Earthy Odorant Geosmin by a Bifunctional *Streptomyces Coelicolor* Enzyme. *Nature Chemical Biology* **2007**, *3* (11), 711–715. <https://doi.org/10.1038/nchembio.2007.29>.
- (83) Becher, P. G.; Vershut, V.; Bibb, M. J.; Bush, M. J.; Molnár, B. P.; Barane, E.; Al-Bassam, M. M.; Chandra, G.; Song, L.; Challis, G. L.; Buttner, M. J.; Flärdh, K. Developmentally Regulated Volatiles Geosmin and 2-Methylisoborneol Attract a Soil Arthropod to *Streptomyces* Bacteria Promoting Spore Dispersal. *Nature Microbiology* **2020**, *5* (6), 821–829. <https://doi.org/10.1038/s41564-020-0697-x>.
- (84) Mak, S.; Nodwell, J. R. Actinorhodin Is a Redox-Active Antibiotic with a Complex Mode of Action against Gram-Positive Cells. *Molecular Microbiology* **2017**, *106* (4), 597–613. <https://doi.org/10.1111/mmi.13837>.
- (85) Jüttner, F.; Watson, S. B. Biochemical and Ecological Control of Geosmin and 2-Methylisoborneol in Source Waters. *Appl. Environ. Microbiol.* **2007**, *73* (14), 4395–4406. <https://doi.org/10.1128/AEM.02250-06>.

- (86) Martín-Sánchez, L.; Singh, K. S.; Avalos, M.; van Wezel, G. P.; Dickschat, J. S.; Garbeva, P. Phylogenomic Analyses and Distribution of Terpene Synthases among *Streptomyces*. *Beilstein Journal of Organic Chemistry* **2019**, *15*, 1181–1193. <https://doi.org/10.3762/bjoc.15.115>.
- (87) Churro, C.; Semedo-Aguiar, A. P.; Silva, A. D.; Pereira-Leal, J. B.; Leite, R. B. A Novel Cyanobacterial Geosmin Producer, Revising GeoA Distribution and Dispersion Patterns in Bacteria. *Scientific Reports* **2020**, *10* (1). <https://doi.org/10.1038/s41598-020-64774-y>.
- (88) Saporito, R. A.; Zuercher, R.; Roberts, M.; Gerow, K. G.; Donnelly, M. A. Experimental Evidence for Aposematism in the Dendrobatid Poison Frog *Oophaga Pumilio*. *Copeia* **2007**, *2007* (4), 1006–1011. [https://doi.org/10.1643/0045-8511\(2007\)7\[1006:EEFAIT\]2.0.CO;2](https://doi.org/10.1643/0045-8511(2007)7[1006:EEFAIT]2.0.CO;2).
- (89) Alatalo, R. V.; Mappes, J. Tracking the Evolution of Warning Signals. *Nature* **1996**, *382* (6593), 708–710.
- (90) Exnerová, A.; Štys, P.; Fučíková, E.; Veselá, S.; Svádová, K.; Prokopová, M.; Jarošík, V.; Fuchs, R.; Landová, E. Avoidance of Aposematic Prey in European Tits (Paridae): Learned or Innate? *Behavioral Ecology* **2007**, *18* (1), 148–156. <https://doi.org/10.1093/beheco/arl061>.
- (91) Jones, R. S.; Fenton, A.; Speed, M. P. “Parasite-Induced Aposematism” Protects Entomopathogenic Nematode Parasites against Invertebrate Enemies. *Behav Ecol* **2016**, *27* (2), 645–651. <https://doi.org/10.1093/beheco/arv202>.
- (92) Seto, H.; Watanabe, H.; Furihata, K. Simultaneous Operation of the Mevalonate and Non-Mevalonate Pathways in the Biosynthesis of Isopentenyl Diphosphate in *Streptomyces Aerioouifer*. *Tetrahedron letters* **1996**, *37* (44), 7979–7982.
- (93) Dairi, T. Studies on Biosynthetic Genes and Enzymes of Isoprenoids Produced by Actinomycetes. *The Journal of Antibiotics* **2005**, *58* (4), 227–243. <https://doi.org/10.1038/ja.2005.27>.
- (94) Alghanmi, H. A.; Alkam, F. M.; AL-Tae, M. M. Effect of Light and Temperature on New Cyanobacteria Producers for Geosmin and 2-Methylisoborneol. *Journal of Applied Phycology* **2018**, *30* (1), 319–328. <https://doi.org/10.1007/s10811-017-1233-0>.
- (95) Pérez, J.; Muñoz-Dorado, J.; Braña, A. F.; Shimkets, L. J.; Sevillano, L.; Santamaría, R. I. *Myxococcus Xanthus* Induces Actinorhodin Overproduction and Aerial Mycelium Formation by *Streptomyces Coelicolor*. *Microbial Biotechnology* **2011**, *4* (2), 175–183. <https://doi.org/10.1111/j.1751-7915.2010.00208.x>.
- (96) Stensmyr, M. C.; Dweck, H. K. M.; Farhan, A.; Ibba, I.; Strutz, A.; Mukunda, L.; Linz, J.; Grabe, V.; Steck, K.; Lavista-Llanos, S.; Wicher, D.; Sachse, S.; Knaden, M.; Becher, P. G.; Seki, Y.; Hansson, B. S. A Conserved Dedicated Olfactory Circuit for Detecting Harmful Microbes in *Drosophila*. *Cell* **2012**, *151* (6), 1345–1357. <https://doi.org/10.1016/j.cell.2012.09.046>.
- (97) Melo, N.; Wolff, G. H.; Costa-da-Silva, A. L.; Arribas, R.; Triana, M. F.; Gugger, M.; Riffell, J. A.; DeGennaro, M.; Stensmyr, M. C. Geosmin Attracts *Aedes Aegypti*

- Mosquitoes to Oviposition Sites. *Current Biology* **2020**, *30* (1), 127-134.e5.
<https://doi.org/10.1016/j.cub.2019.11.002>.
- (98) Huang, H.; Ren, L.; Li, H.; Schmidt, A.; Gershenzon, J.; Lu, Y.; Cheng, D. The Nesting Preference of an Invasive Ant Is Associated with the Cues Produced by Actinobacteria in Soil. *PLOS Pathogens* **2020**, *16* (9), e1008800.
<https://doi.org/10.1371/journal.ppat.1008800>.
- (99) Vazquez-Martinez, M. G.; Rodríguez, M. H.; Arredondo-Jiménez, J. I.; Méndez-Sánchez, J. D.; Bond-Compeán, J. G.; Gold-Morgan, M. Cyanobacteria Associated with *Anopheles Albimanus* (Diptera: Culicidae) Larval Habitats in Southern Mexico. *Journal of medical entomology* **2002**, *39* (6), 825–832.
- (100) Sherratt, T. N. The Evolution of Müllerian Mimicry. *Naturwissenschaften* **2008**, *95* (8), 681–695. <https://doi.org/10.1007/s00114-008-0403-y>.
- (101) McDaniel, L. D.; Young, E.; Delaney, J.; Ruhnau, F.; Ritchie, K. B.; Paul, J. H. High Frequency of Horizontal Gene Transfer in the Oceans. *Science* **2010**, *330* (6000), 50.
<https://doi.org/10.1126/science.1192243>.
- (102) Vos, M.; Hesselman, M. C.; Beek, T. A. te; Passel, M. W. J. van; Eyre-Walker, A. Rates of Lateral Gene Transfer in Prokaryotes: High but Why? *Trends in Microbiology* **2015**, *23* (10), 598–605. <https://doi.org/10.1016/j.tim.2015.07.006>.
- (103) Chouteau, M.; Arias, M.; Joron, M. Warning Signals Are under Positive Frequency-Dependent Selection in Nature. *Proceedings of the National Academy of Sciences* **2016**, *113* (8), 2164–2169. <https://doi.org/10.1073/pnas.1519216113>.
- (104) Tran, A.; Tang, A.; O’Loughlin, C. T.; Balistreri, A.; Chang, E.; Coto Villa, D.; Li, J.; Varshney, A.; Jimenez, V.; Pyle, J.; Tsujimoto, B.; Wellbrook, C.; Vargas, C.; Duong, A.; Ali, N.; Matthews, S. Y.; Levinson, S.; Woldemariam, S.; Khuri, S.; Bremer, M.; Eggers, D. K.; L’Etoile, N.; Miller Conrad, L. C.; VanHoven, M. K. C. *Elegans* Avoids Toxin-Producing *Streptomyces* Using a Seven Transmembrane Domain Chemosensory Receptor. *eLife* **2017**, *6*, e23770. <https://doi.org/10.7554/eLife.23770>.
- (105) Jousset, A.; Rochat, L.; Péchy-Tarr, M.; Keel, C.; Scheu, S.; Bonkowski, M. Predators Promote Defence of Rhizosphere Bacterial Populations by Selective Feeding on Non-Toxic Cheaters. *The ISME Journal* **2009**, *3* (6), 666–674.
<https://doi.org/10.1038/ismej.2009.26>.
- (106) Servedio, M. R. The Effects of Predator Learning, Forgetting, and Recognition Errors on the Evolution of Warning Coloration. *Evolution* **2000**, *54* (3), 751–763.
<https://doi.org/10.1111/j.0014-3820.2000.tb00077.x>.
- (107) De la Fuente, I. M.; Bringas, C.; Malaina, I.; Fedetz, M.; Carrasco-Pujante, J.; Morales, M.; Knafo, S.; Martínez, L.; Pérez-Samartín, A.; López, J. I.; Pérez-Yarza, G.; Boyano, M. D. Evidence of Conditioned Behavior in Amoebae. *Nature Communications* **2019**, *10* (1), 1–12. <https://doi.org/10.1038/s41467-019-11677-w>.
- (108) Ardiel, E. L.; Rankin, C. H. An Elegant Mind: Learning and Memory in *Caenorhabditis Elegans*. *Learn. Mem.* **2010**, *17* (4), 191–201. <https://doi.org/10.1101/lm.960510>.

- (109) Musselman, H. N.; Neal-Beliveau, B.; Nass, R.; Engleman, E. A. Chemosensory Cue Conditioning with Stimulants in a *Caenorhabditis Elegans* Animal Model of Addiction. *Behavioral Neuroscience* **2012**, *126* (3), 445–456. <https://doi.org/10.1037/a0028303>.
- (110) Nishijima, S.; Maruyama, I. N. Appetitive Olfactory Learning and Long-Term Associative Memory in *Caenorhabditis Elegans*. *Front. Behav. Neurosci.* **2017**, *11*. <https://doi.org/10.3389/fnbeh.2017.00080>.
- (111) Philippe, H.; Douady, C. J. Horizontal Gene Transfer and Phylogenetics. *Current Opinion in Microbiology* **2003**, *6* (5), 498–505. <https://doi.org/10.1016/j.mib.2003.09.008>.
- (112) Lu, G.; Edwards, C. G.; Fellman, J. K.; Mattinson, D. S.; Navazio, J. Biosynthetic Origin of Geosmin in Red Beets (*Beta Vulgaris* L.). *J. Agric. Food Chem.* **2003**, *51* (4), 1026–1029. <https://doi.org/10.1021/jf020905r>.
- (113) Bucarey, S. A.; Penn, K.; Paul, L.; Fenical, W.; Jensen, P. R. Genetic Complementation of the Obligate Marine Actinobacterium *Salinispora Tropica* with the Large Mechanosensitive Channel Gene *MscL* Rescues Cells from Osmotic Downshock. *Appl. Environ. Microbiol.* **2012**, *78* (12), 4175–4182. <https://doi.org/10.1128/AEM.00577-12>.
- (114) Jensen, P. R.; Williams, P. G.; Oh, D.-C.; Zeigler, L.; Fenical, W. Species-Specific Secondary Metabolite Production in Marine Actinomycetes of the Genus *Salinispora*. *Appl. Environ. Microbiol.* **2007**, *73* (4), 1146–1152. <https://doi.org/10.1128/AEM.01891-06>.
- (115) Ziemert, N.; Lechner, A.; Wietz, M.; Millan-Aguinaga, N.; Chavarria, K. L.; Jensen, P. R. Diversity and Evolution of Secondary Metabolism in the Marine Actinomycete Genus *Salinispora*. *Proceedings of the National Academy of Sciences* **2014**, *111* (12), E1130–E1139. <https://doi.org/10.1073/pnas.1324161111>.
- (116) Manivasagan, P.; Kang, K.-H.; Sivakumar, K.; Li-Chan, E. C. Y.; Oh, H.-M.; Kim, S.-K. Marine Actinobacteria: An Important Source of Bioactive Natural Products. *Environmental Toxicology and Pharmacology* **2014**, *38* (1), 172–188. <https://doi.org/10.1016/j.etap.2014.05.014>.
- (117) Dufresne, A.; Ostrowski, M.; Scanlan, D. J.; Garczarek, L.; Mazard, S.; Palenik, B. P.; Paulsen, I. T.; Tandeau de Marsac, N.; Wincker, P.; Dossat, C.; Ferriera, S.; Johnson, J.; Post, A. F.; Hess, W. R.; Partensky, F. Unravelling the Genomic Mosaic of a Ubiquitous Genus of Marine Cyanobacteria. *Genome Biol* **2008**, *9* (5), R90. <https://doi.org/10.1186/gb-2008-9-5-r90>.
- (118) Amiri Moghaddam, J.; Poehlein, A.; Fisch, K.; Alanjary, M.; Daniel, R.; König, G. M.; Schäberle, T. F. Draft Genome Sequences of the Obligatory Marine Myxobacterial Strains *Enhygromyxa Salina* SWB005 and SWB007. *Genome Announc* **2018**, *6* (17), e00324-18, /ga/6/17/e00324-18.atom. <https://doi.org/10.1128/genomeA.00324-18>.
- (119) Albatineh, H.; Stevens, D. Marine Myxobacteria: A Few Good Halophiles. *Marine Drugs* **2018**, *16* (6), 209. <https://doi.org/10.3390/md16060209>.

- (120) Wang, Z.; Xu, Y.; Shao, J.; Wang, J.; Li, R. Genes Associated with 2-Methylisoborneol Biosynthesis in Cyanobacteria: Isolation, Characterization, and Expression in Response to Light. *PLoS ONE* **2011**, *6* (4), e18665. <https://doi.org/10.1371/journal.pone.0018665>.
- (121) Zhou, A.; Chen, Y. I.; Zane, G. M.; He, Z.; Hemme, C. L.; Joachimiak, M. P.; Baumohl, J. K.; He, Q.; Fields, M. W.; Arkin, A. P.; Wall, J. D.; Hazen, T. C.; Zhou, J. Functional Characterization of Crp/Fnr-Type Global Transcriptional Regulators in *Desulfovibrio Vulgaris* Hildenborough. *Appl. Environ. Microbiol.* **2012**, *78* (4), 1168–1177. <https://doi.org/10.1128/AEM.05666-11>.
- (122) Körner, H.; Sofia, H. J.; Zumft, W. G. Phylogeny of the Bacterial Superfamily of Crp-Fnr Transcription Regulators: Exploiting the Metabolic Spectrum by Controlling Alternative Gene Programs. *FEMS Microbiol Rev* **2003**, *27* (5), 559–592. [https://doi.org/10.1016/S0168-6445\(03\)00066-4](https://doi.org/10.1016/S0168-6445(03)00066-4).
- (123) Nakaya, A.; Katayama, T.; Itoh, M.; Hiranuka, K.; Kawashima, S.; Moriya, Y.; Okuda, S.; Tanaka, M.; Tokimatsu, T.; Yamanishi, Y.; Yoshizawa, A. C.; Kanehisa, M.; Goto, S. KEGG OC: A Large-Scale Automatic Construction of Taxonomy-Based Ortholog Clusters. *Nucleic Acids Research* **2012**, *41* (D1), D353–D357. <https://doi.org/10.1093/nar/gks1239>.
- (124) PDB ID: 3GIW. CRYSTAL STRUCTURE OF a DUF574 Family Protein (SAV_2177) FROM STREPTOMYCES AVERMITILIS MA-4680 AT 1.45 Å RESOLUTION. April 14, 2009.
- (125) Hudson, C. M.; Lau, B. Y.; Williams, K. P. Islander: A Database of Precisely Mapped Genomic Islands in tRNA and tmRNA Genes. *Nucleic Acids Research* **2015**, *43* (D1), D48–D53. <https://doi.org/10.1093/nar/gku1072>.
- (126) Langille, M. G.; Hsiao, W. W.; Brinkman, F. S. Evaluation of Genomic Island Predictors Using a Comparative Genomics Approach. *BMC Bioinformatics* **2008**, *9* (1), 329. <https://doi.org/10.1186/1471-2105-9-329>.
- (127) Hsiao, W.; Wan, I.; Jones, S. J.; Brinkman, F. S. L. IslandPath: Aiding Detection of Genomic Islands in Prokaryotes. *Bioinformatics* **2003**, *19* (3), 418–420. <https://doi.org/10.1093/bioinformatics/btg004>.
- (128) Waack, S.; Keller, O.; Asper, R.; Brodag, T.; Damm, C.; Fricke, W.; Surovcik, K.; Meinicke, P.; Merkl, R. Score-Based Prediction of Genomic Islands in Prokaryotic Genomes Using Hidden Markov Models. *BMC Bioinformatics* **2006**, *7* (1), 142. <https://doi.org/10.1186/1471-2105-7-142>.
- (129) Zhang, J.; Kumar, S.; Nei, M. Small-Sample Tests of Episodic Adaptive Evolution: A Case Study of Primate Lysozymes. *Molecular Biology and Evolution* **1997**, *14* (12), 1335–1338. <https://doi.org/10.1093/oxfordjournals.molbev.a025743>.
- (130) Zhang, J. Evolution by Gene Duplication: An Update. *Trends in Ecology & Evolution* **2003**, *18* (6), 292–298. [https://doi.org/10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8).
- (131) Stiernagle, T. Maintenance of *C. Elegans*. *WormBook* **2006**. <https://doi.org/10.1895/wormbook.1.101.1>.

- (132) Page, F. C. *A New Key to Freshwater and Soil Gymnamoebae: With Instructions for Culture*; Freshwater Biological Assoc.: Ambleside, Cumbria, 1988.
- (133) Caspers, H. F. C. Page: An Illustrated Key to Freshwater and Soil Amoebae with notes on Cultivation and Ecology. – With 64 fig., 155 pp. The Ferry House, Far Sawrey, Ambleside, Cumbria: Freshwater Biological Association, Scientific Publication No. 34. 1976. SBN 900 386 26 6, ISSN 0367-1857. £ 2.50. *Int. Revue ges. Hydrobiol. Hydrogr.* **1978**, 63 (2), 289–289. <https://doi.org/10.1002/iroh.19780630231>.
- (134) Weinstein, M. P.; Clinical and Laboratory Standards Institute. *Performance Standards for Antimicrobial Susceptibility Testing*; 2019.
- (135) Takeda, K.; Miura, M.; Takagi, T. Stepwise Formation of Complexes between Sodium Dodecyl Sulfate and Bovine Serum Albumin Detected by Measurements of Electric Conductivity, Binding Isotherm, and Circular Dichroism. *Journal of Colloid and Interface Science* **1981**, 82 (1), 38–44. [https://doi.org/10.1016/0021-9797\(81\)90121-1](https://doi.org/10.1016/0021-9797(81)90121-1).
- (136) Clinical and Laboratory Standards Institute; Weinstein, M. P. *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically*; 2018.
- (137) Forli, S. Epithelones: From Discovery to Clinical Trials. *Curr Top Med Chem* **2014**, 14 (20), 2312–2321. <https://doi.org/10.2174/1568026614666141130095855>.
- (138) Gray, K. C.; Palacios, D. S.; Dailey, I.; Endo, M. M.; Uno, B. E.; Wilcock, B. C.; Burke, M. D. Amphotericin Primarily Kills Yeast by Simply Binding Ergosterol. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, 109 (7), 2234–2239. <https://doi.org/10.1073/pnas.1117280109>.
- (139) *Imaris*; BitPlane: South Windsor, CT, USA.
- (140) *WormLab*; MBF Bioscience: Williston, VT USA.
- (141) Xiong, H.; Pears, C.; Woollard, A. An Enhanced C. Elegans Based Platform for Toxicity Assessment. *Sci Rep* **2017**, 7 (1), 9839. <https://doi.org/10.1038/s41598-017-10454-3>.
- (142) Letunic, I.; Bork, P. Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees. *Nucleic Acids Res* **2016**, 44 (W1), W242–W245. <https://doi.org/10.1093/nar/gkw290>.
- (143) Kim, P.-J.; Price, N. D. Genetic Co-Occurrence Network across Sequenced Microbes. *PLoS Comput Biol* **2011**, 7 (12), e1002340. <https://doi.org/10.1371/journal.pcbi.1002340>.

Appendix: Supporting Information

List of Supporting Figures and Tables

Figure A1: Effect of geosmin and 2-MIB on <i>C. elegans</i> feeding.	73
Figure A2: Analysis of RGP in various bacterial strains using Island viewer 4.....	78
Sample SPARQL Queries.....	81
Table A1: Effect of geosmin on <i>C. elegans</i> movement.....	74
Table A2: BLASTp readout of comparison of <i>M. xanthus</i> DK1622 GS to <i>Beta vulgaris</i> L., <i>Penicillium expansum</i> and <i>Aspergillus tubingensis</i>	79
Table A3: BLAST readout of CNBP comparison to various Crp-Fnr genes.	80

List of Additional Files and Videos

Additional file 1. Bacterial genome, GS gene and 16S rRNA selection. XLS sheet with the bacterial family and strain name, genbank accession number, base pair and amino acid sequence of GS and 16S rRNA of all strains used in the study.

Additional file 2. Isolation environment and classification of actinobacterial strains used in this study. DOC highlighting in yellow the Actinobacteria orders, in green the families and in cyan the genera that contain GS, in blue the strains that contain GS and are present in marine environments, and a delta (Δ) indicates the presence of CNBPs (legend at bottom of the page).

Additional file 3. Isolation environment and classification of myxobacterial strains used in this study. DOC highlighting in yellow the myxobacteria orders, in green the families and in cyan the genera that contain GS, in blue the strains that contain GS and are present in marine environments, and a delta (Δ) indicates the presence of CNBPs (legend at bottom of the page).

Additional file 4. Isolation environment and classification of cyanobacterial strains used in this study. DOC highlighting in yellow the cyanobacterial orders, in green the families and in cyan the genera that contain GS, in blue the strains that contain GS and are present in marine environments, and a delta (Δ) indicates the presence of CNBPs (legend at bottom of the page).

Additional file 5. Python scripts used for the orthologue-based approach for co-occurring genes determination (TXT file).

Additional file 6. Raw output of co-occurrence genes with GS using orthologue-based approach. XLS file of the co-occurrence of geosmin synthase and 2,110,338 orthologue clusters.

Additional file 7. Codon-based Fisher's Exact Test readout on 257 actinobacterial strains (XLS file).

Additional file 8. Codon-based Fisher's Exact Test readout on 25 myxobacterial strains (XLS file).

Additional file 9. Codon-based Fisher's Exact Test readout on 23 cyanobacterial strains (XLS file).

Additional file 10. Video of *C. elegans* N2 crawling on agar. (MP4 file)

Additional file 11. Video of *C. elegans* N2 crawling on agar that contains geosmin. (MP4 file)

Additional file 12. Video of *C. elegans* NL2105 crawling on agar. (MP4 file)

Additional file 13. MP4 video of *C. elegans* NL2105 crawling on agar that contains geosmin.

Additional file 14. Nematode distress behaviour exhibited when in *S. coelicolor* colonies. (MP4 file)

Additional file 15. Imaris Python analysis script. (TXT file).

Supplementary Figures and Tables

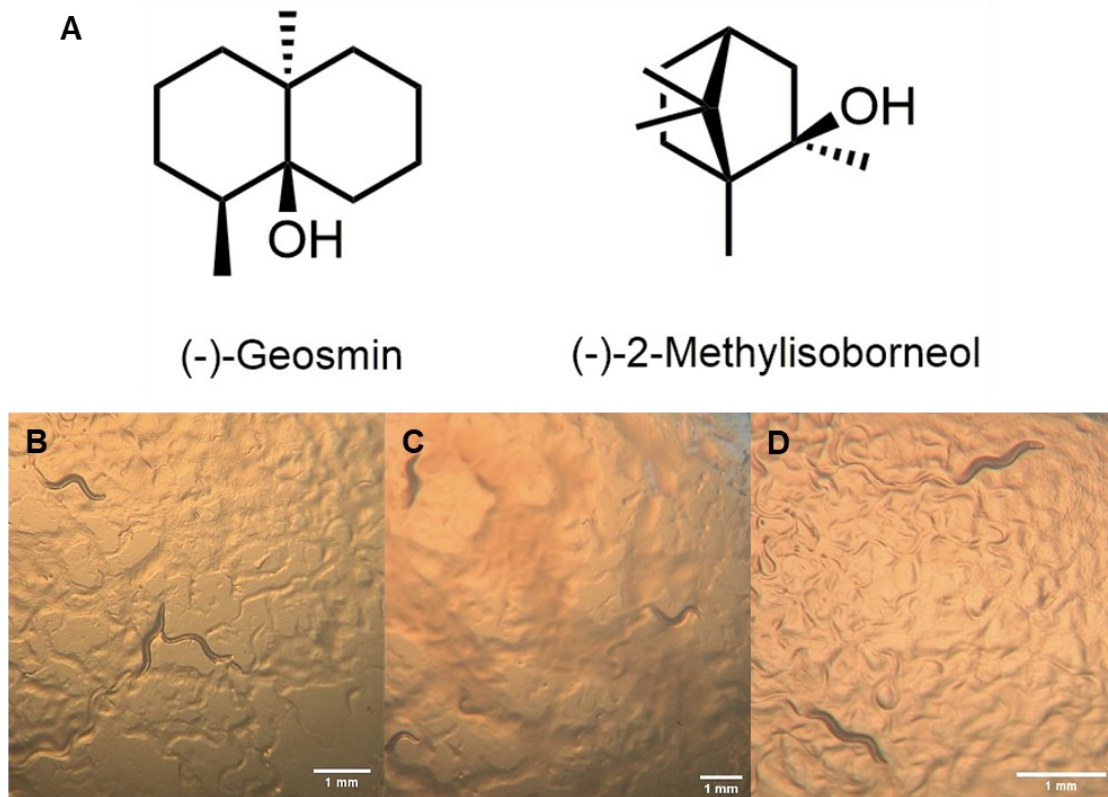


Figure A1: Effect of geosmin and 2-MIB on *C. elegans* feeding.

(A) Structure of geosmin and 2-MIB.

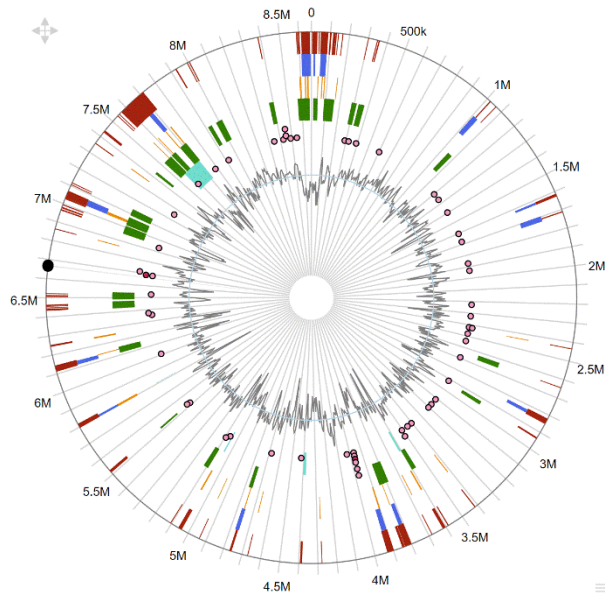
(B) Consumption of *E. coli* by *C. elegans* as they crawl into the bacterial lawn leaving tracks and reducing the bacterial population.

(C) Consumption of *E. coli* by *C. elegans* in the presence of geosmin (54 $\mu\text{g/mL}$) as they crawl into the bacterial lawn leaving tracks and reducing the bacterial population; as observed in (B).

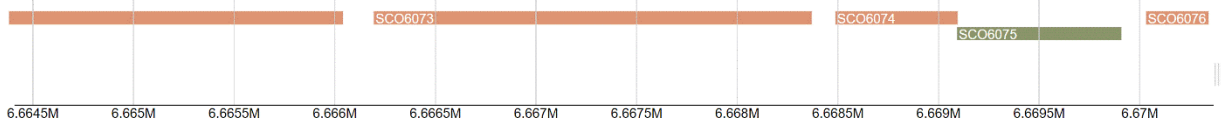
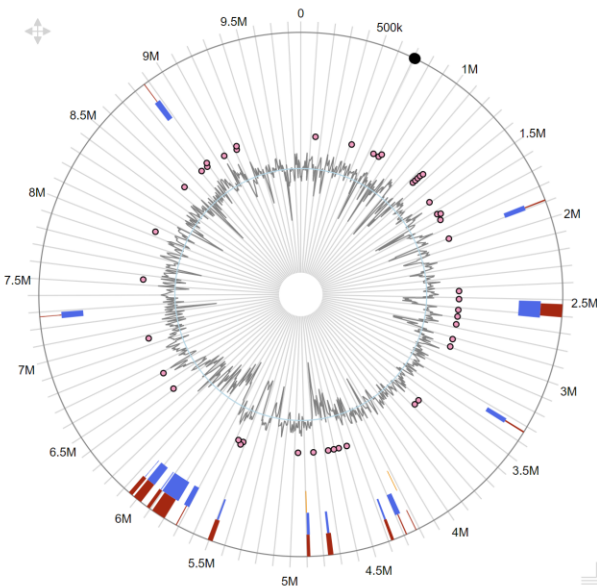
(D) Consumption of *E. coli* by *C. elegans* in the presence of 2-MIB (150 $\mu\text{g/mL}$) as they crawl into the bacterial lawn leaving tracks and reducing the bacterial population; as observed in (B).

Table A1: Effect of geosmin on *C. elegans* movement.

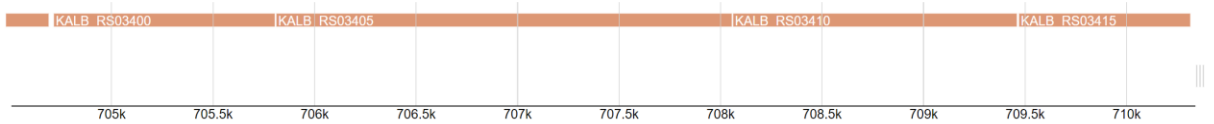
Analysis	<i>C. elegans</i>	Mutant Deficiency	Control		Geosmin		P-value
			Average	Standard Deviation	Average	Standard Deviation	
Track Line (%)	N2	Wild type	0.4495	0.2464	0.5547	0.3115	0.0498
	BR5514	ADF, ASE, AWC	0.4371	0.3032	0.5704	0.3091	0.0407
	CE1248	AWA	0.5447	0.3098	0.3973	0.2477	0.0098
	CX2065	AWB, partial AWC	0.5715	0.2969	0.7174	0.2617	0.0531
	CX2205	Olfactory system	0.5676	0.2667	0.5015	0.3330	0.2376
	CX5893	AWC	0.3797	0.2383	0.5124	0.3134	0.0233
	NL2105	Chemosensation	0.5417	0.3210	0.5919	0.3253	0.4034
	PR674	ASE	0.5617	0.2997	0.6174	0.2746	0.3523
Peristaltic Speed ($\mu\text{m/s}$) [Absolute peristaltic track length/time]	N2	Wild type	30.30027	8.036627	40.89631	5.74027	1.13E-06
	BR5514	ADF, ASE, AWC	28.60974	11.43549	28.62741	10.73547	0.8719
	CE1248	AWA	38.47773	5.18592	30.76179	5.137974	1.55E-05
	CX2065	AWB, partial AWC	31.69875	3.094122	45.48192	6.194929	5.57E-05
	CX2205	Olfactory system	41.43203	1.540853	35.47248	1.379966	0.00353
	CX5893	AWC	27.08556	9.214374	35.17561	5.317928	0.001547
	NL2105	Chemosensation	24.34918	3.870071	26.03611	5.0001	0.2812
	PR674	ASE	52.00835	3.823389	47.42004	6.401746	0.2098
Periodicity (μm)	N2	Wild type	19.096	4.489433	19.49362	0.499788	0.03744
	BR5514	ADF, ASE, AWC	18.12805	1.356552	17.20374	2.388581	0.2657
	CE1248	AWA	16.38904	0.431564	12.12216	1.405526	2.44E-18
	CX2065	AWB, partial AWC	16.08523	1.203601	18.68935	0.364582	1.94E-05
	CX2205	Olfactory system	17.03812	0.239728	15.15458	0.604553	2.91E-05
	CX5893	AWC	19.58694	4.840422	22.032	0.885555	0.09487
	NL2105	Chemosensation	16.4984	2.778458	15.42235	0.691968	0.6440
	PR674	ASE	22.65395	2.821125	21.9781	0.986451	0.1370

A**STREPTOMYCES COELICOLOR A3(2) CHROMOSOME, COMPLETE GENOME.**

Presence of RGP	Gene Name	Accnum	Product
No	SCO6072	NP_630181.1	hypothetical protein
No	SCO6073	NP_630182.1	cyclase
No	SCO6074	NP_630183.1	hypothetical protein
No	SCO6075	NP_630184.1	hypothetical protein
No	SCO6076	NP_630185.1	dipeptidase

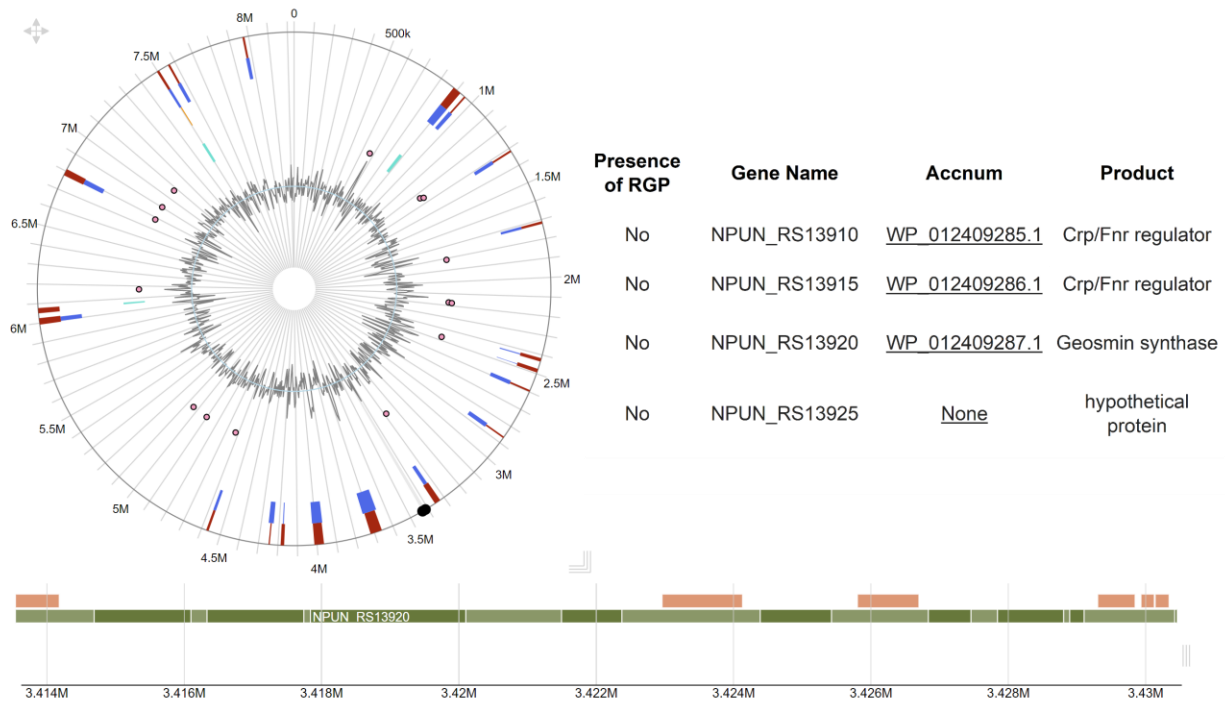
**B****KUTZNERIA ALBIDA DSM 43870, COMPLETE GENOME.**

Presence of RGP	Gene Name	Accnum	Product
No	KALB_RS03395	WP_042220182.1	Crp/Fnr regulator
No	KALB_RS03400	WP_025354321.1	hypothetical protein
No	KALB_RS03405	WP_025354322.1	hypothetical protein
No	KALB_RS03410	WP_025354323.1	Crp/Fnr regulator



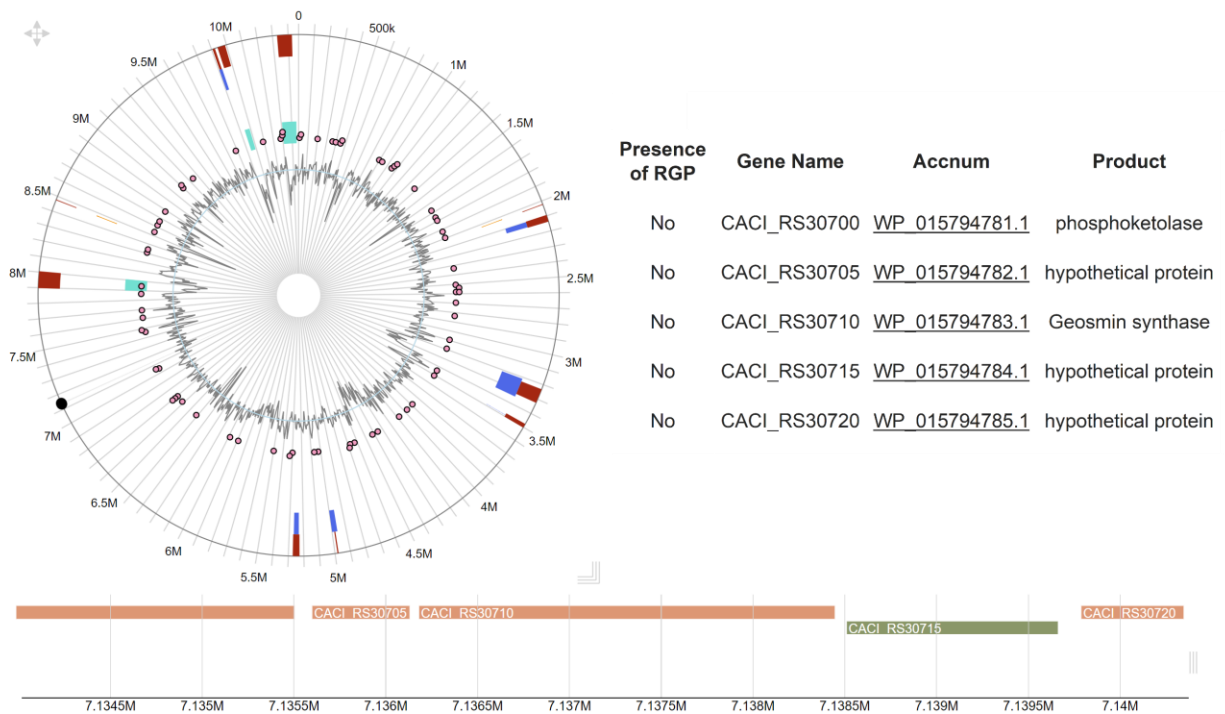
C

NOSTOC PUNCTIFORME PCC 73102, COMPLETE GENOME.

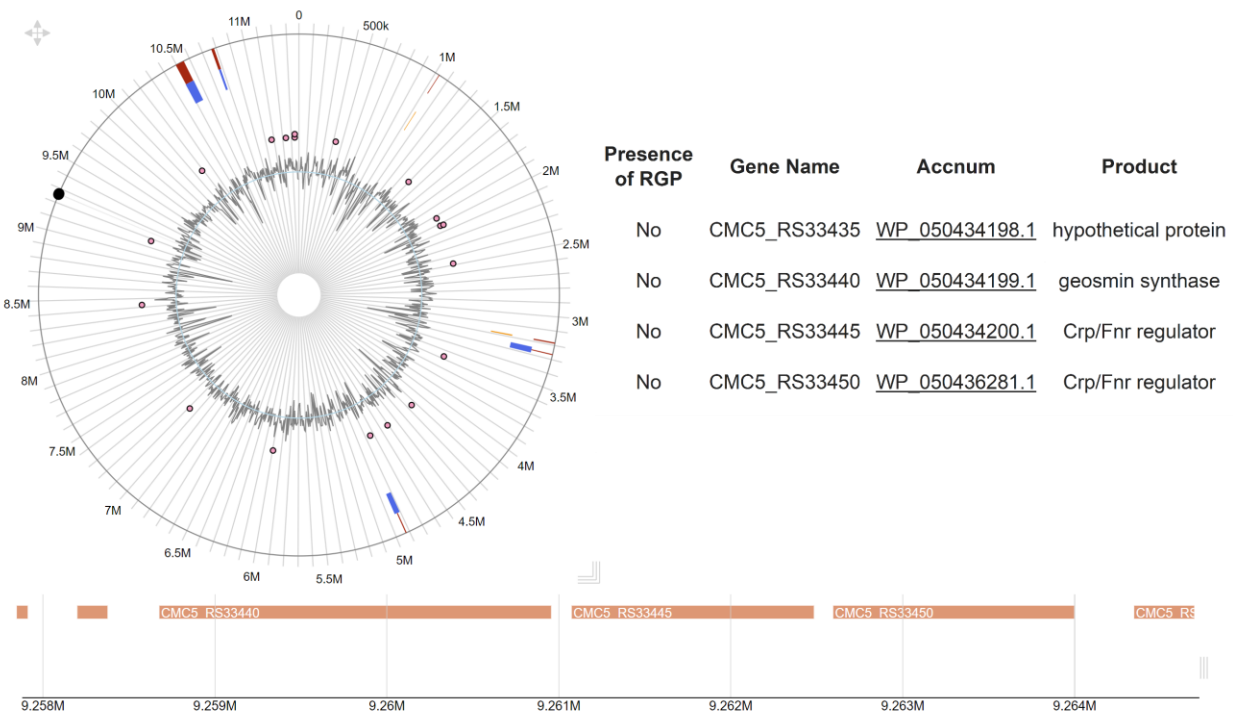


D

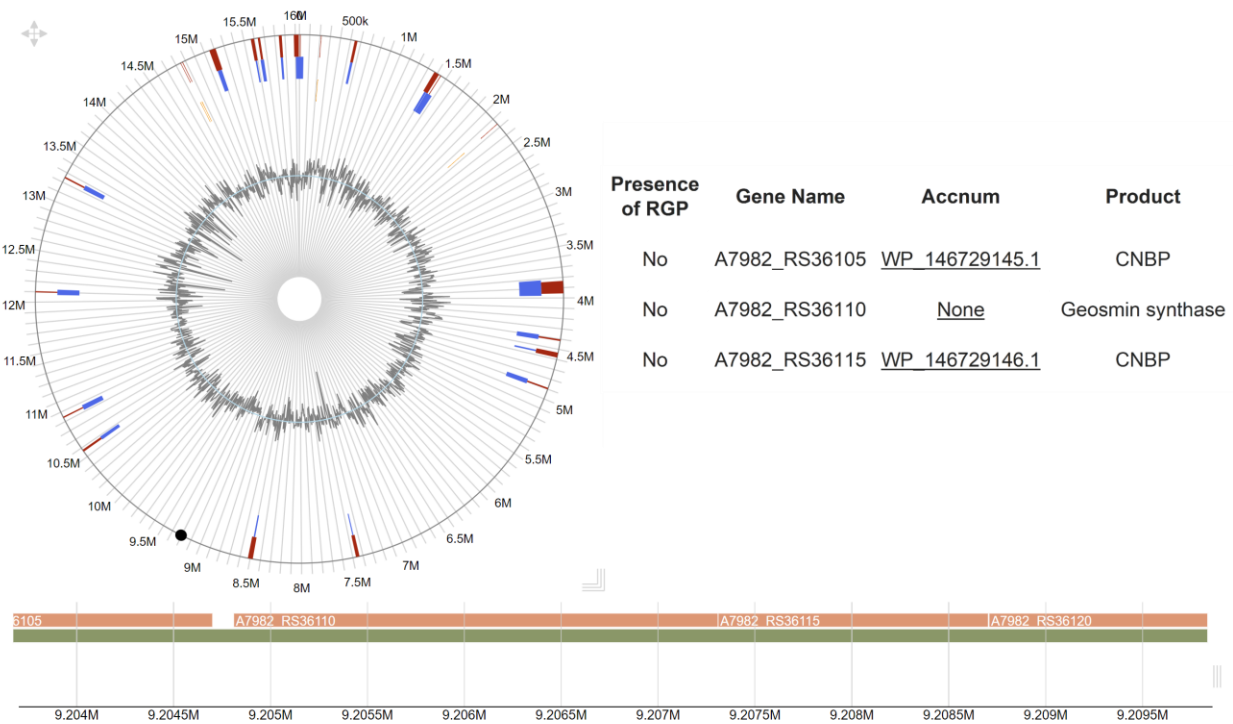
CATENULISPORIA ACIDIPHILA DSM 44928, COMPLETE GENOME.



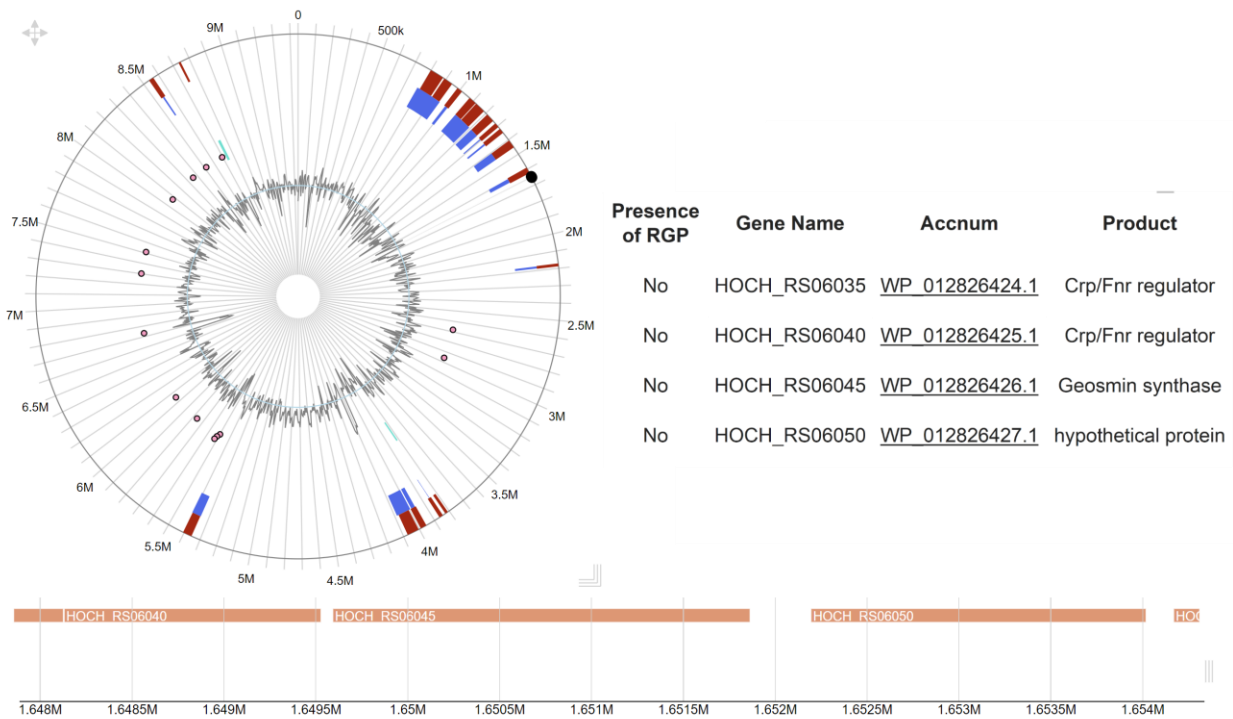
E **CHONDROMYCES CROCATUS STRAIN Cm c5, COMPLETE GENOME.**



F **MINICYSTIS ROSEA STRAIN DSM 24000 CHROMOSOME, COMPLETE GENOME.**



G HALIANGIUM OCHRACEUM DSM 14365, COMPLETE GENOME.



H STACKEBRANDTIA NASSAUENSIS DSM 44728, COMPLETE GENOME.

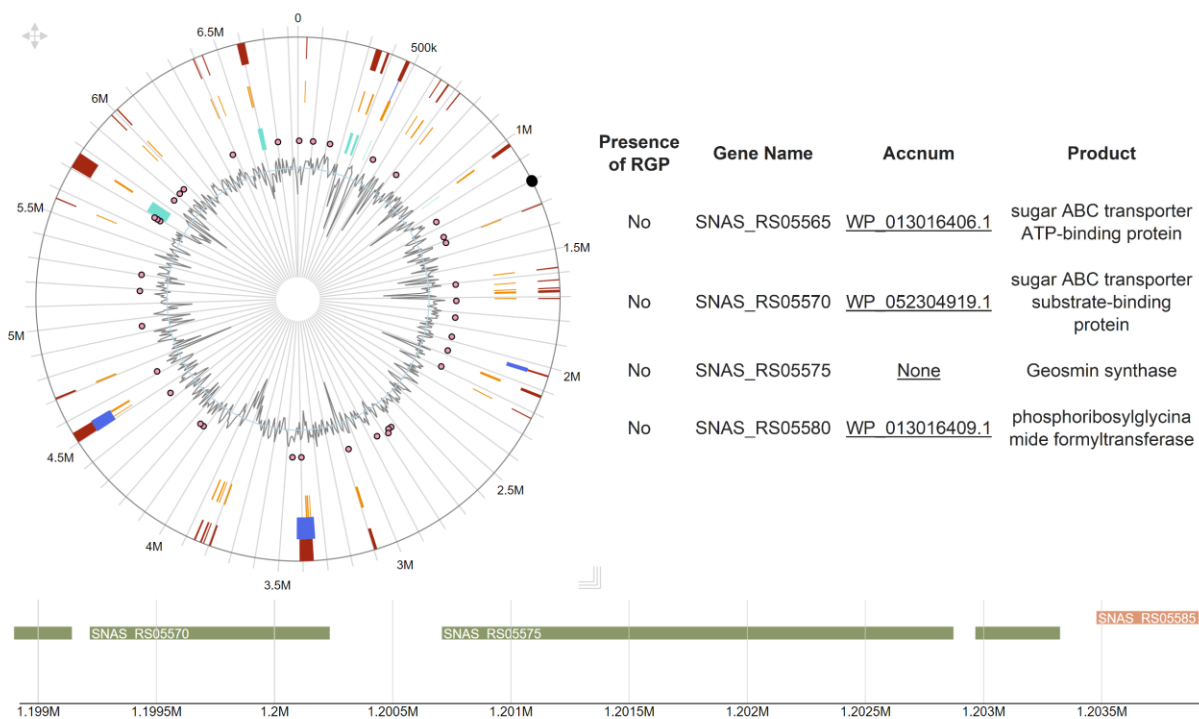


Figure A2: Analysis of RGP in various bacterial strains using Island viewer 4. Full circular genome with RGP indications (upper left), information window of selected region containing GS gene as a cyclase (upper right) and close-up of linear DNA region containing geosmin synthase as SCO6073 (lower center) adjacent to two hypothetical proteins. Recent indications of RGP are

absent around GS gene indicating HGT that occurred in the distant past.

(A) Analysis of RGP in *S. coelicolor* A3(2).

(B) Analysis of RGP in *Kutzneria albida* DSM 43870. KALB_RS03405 is the gene for GS.

(C) Analysis of RGP in *Nostoc punctiforme* PCC 73102.

(D) Analysis of RGP in *Catenulispora acidiphila* DSM 44928.

(E) Analysis of RGP in *Chondromyces crocatus* Cm c5.

(F) Analysis of RGP in *Minicystis rosea* DSM 24000.

(G) Analysis of RGP in *Haliangium ochraceum* DSM 14365.

(H) Analysis of RGP in *Stackebrandtia nassauensis* DSM 44728.

Table A2: BLASTp readout of comparison of *M. xanthus* DK1622 GS to *Beta vulgaris* L., *Penicillium expansum* and *Aspergillus tubingensis*.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
pentalene synthase [<i>Aspergillus tubingensis</i>]	<i>Aspergillus tubingensis</i>	88.6	88.6	37%	3.00E-18	25.17	365	XP_035361508.1
hypothetical protein ASPTUDRAFT_50022 [<i>Aspergillus tubingensis</i> CBS 134.48]	<i>Aspergillus tubingensis</i> CBS 134.48	78.6	78.6	36%	5.00E-15	24.45	331	OJI89048.1
hypothetical protein ASPTUDRAFT_61636 [<i>Aspergillus tubingensis</i> CBS 134.48]	<i>Aspergillus tubingensis</i> CBS 134.48	65.1	122	76%	1.00E-10	25.14	344	OJI88257.1
terpene synthase metal binding domain protein [<i>Aspergillus tubingensis</i>]	<i>Aspergillus tubingensis</i>	61.6	116	76%	2.00E-09	25.29	336	XP_035361915.1
Terpenoid synthase [<i>Penicillium expansum</i>]	<i>Penicillium expansum</i>	53.9	107	74%	4.00E-07	23.53	331	KGO48192.1
Terpenoid synthase [<i>Penicillium expansum</i>]	<i>Penicillium expansum</i>	53.9	105	73%	5.00E-07	23.53	331	XP_016595385.1
Terpenoid synthase [<i>Penicillium expansum</i>]	<i>Penicillium expansum</i>	47	47	24%	6.00E-05	20.97	291	XP_016595124.1
hypothetical protein ASPTUDRAFT_126705 [<i>Aspergillus tubingensis</i> CBS 134.48]	<i>Aspergillus tubingensis</i> CBS 134.48	42	83.9	20%	0.003	22.5	367	OJI81974.1
terpenoid synthase [<i>Aspergillus tubingensis</i>]	<i>Aspergillus tubingensis</i>	40	40	14%	0.011	25.66	391	XP_035358724.1

No significant alignments were observed with *Beta vulgaris* L. (red beets).

Table A3: BLAST readout of CNBP comparison to various Crp-Fnr genes.

Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Agrobacterium tumefaciens	40.8	40.8	19%	0.009	30	233	NTE84063.1
Agrobacterium tumefaciens	40.4	40.4	14%	0.01	36.36	233	WP_099086884.1
Agrobacterium	40.4	40.4	14%	0.01	36.36	217	WP_153316095.1
Agrobacterium tumefaciens	40.4	40.4	14%	0.01	36.36	217	WP_153463722.1
Agrobacterium tumefaciens	40.4	40.4	14%	0.01	36.36	233	WP_060724269.1
Agrobacterium	40.4	40.4	14%	0.01	36.36	233	WP_003504633.1
Rhizobium/Agrobacterium group	40.4	40.4	14%	0.01	36.36	233	WP_020809046.1
Agrobacterium fabrum	39.7	39.7	16%	0.021	32	233	WP_080814717.1
Rhizobium/Agrobacterium group	39.7	39.7	16%	0.022	32	233	WP_034499545.1
Agrobacterium fabrum	39.3	39.3	15%	0.022	32.43	217	WP_178118884.1
Agrobacterium tumefaciens	39.7	39.7	16%	0.022	32	233	AMD57539.1
Agrobacterium	39.3	39.3	16%	0.029	32	233	WP_078053290.1
Agrobacterium	38.9	38.9	16%	0.032	32	233	WP_003522052.1
Agrobacterium	38.9	38.9	16%	0.034	32	233	WP_020811077.1
Agrobacterium tumefaciens	38.9	38.9	14%	0.034	34.85	233	WP_065655244.1

Sample SPARQL Queries.

Full taxon database readout

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX orth: <http://purl.org/net/orth#>
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dct: <http://purl.org/dc/terms/>
```

```
PREFIX void: <http://rdfs.org/ns/void#>
SELECT ?org_name
WHERE {
    ?gene obo:RO_0002162 ?org_name
}
GROUP BY ?org_name
```

Orthologue clusters of organism clade "Myxococcus"

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX orth: <http://purl.org/net/orth#>
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX void: <http://rdfs.org/ns/void#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
SELECT DISTINCT ?orgs ?ocnum
WHERE {
    ?org_group rdfs:label "Myxococcus" ;
        void:subset* ?orgs .
```

```
?orgs rdfs:seeAlso ?org_name .
?gene obo:RO_0002162 ?org_name .
?oc orth:hasHomologous* ?gene ;
    rdfs:label ?ocnum
```

```
FILTER (REGEX (?ocnum, "^OC"))
}
GROUP BY ?orgs ?ocnum
```