nature portfolio

Peer Review File

Exploring and expanding the natural chemical space of bacterial diterpenes

Corresponding Author: Dr Jeffrey Rudolf

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Wei, Ning and co-workers present a comprehensive survey of bacterial terpene synthases (TSs), in which they identified putative terpene cyclases from the bacterial kingdom from the Uniprot database. They prioritized their hits and ordered synthetic genes for 334 TSs. They screened all the enzymes by heterologous expression in a diterpene producing E. coli platform and find that 125 show diterpene synthase activity. Of these enzymes the authors characterized 31 thoroughly and elucidated the produced structures by NMR, MS and VCD, elucidating their absolute configuration. They go through their characterized TS list and organize the enzymes by their products' representation in previous literature.
The chemical data are very thorough and clearly presented in the SI. Generally, these data are very useful to understand TS activities. As there are relatively few TSs characterized it is impossible to predict a TSs activity or even its preferred substrate by substrate alone, given the highly complex reaction they catalyze, the authors state this, too. Towards this aim, the collected data are a valuable contribution, and they were collected by a lot of hard work.

However, the organization of the article is very hard to follow. The premise of the article as stated in the introduction is to show terpene diversity of the bacterial phylum. Yet the only criterion for grouping is the chemical structure of the compounds and how they are represented in the literature as reported chemical entities (new, new to bacteria, new isomers, and – as a break from this scheme - a single tridomain enzyme). This makes the article repetitive and not easy to read, and the authors themselves state that this organization has drawbacks (several compounds fall into several categories). The methods for structure elucidation and isolation are listed in several sections but are the same. The biological background information (e.g., what phyla of bacteria had how many terpene cyclases per species, which group didn't have any, what is the occurrence of diterpene synthases per phylum etc.), are not comprehensively presented. The prioritization of the TSs from the initial batch (identified enzymes in uniprot) and the further prioritization for structure elucidation is not clearly presented. The authors should show data here.
By presenting and connecting the presented chemical data to the biological background data, which the authors should have because of their comprehensive analysis, they could provide a more natural structure to the article.
If the authors would organize the article by the phylogenetic origin of the discovered diterpene diversity it would be easier to highlight differences, similarities, abundance of terpene cyclases and terpene products (what bacterial phyla produce how many). This could be compared with previously characterized diterpene synthases (newly identified interesting groups of bacteria? Are the all-time classics (actinobacteria) really the all-time classics? Is there a surprising accumulation of diterpene synthases in certain groups? As diterpenes are not that volatile they are probably not used as volatile natural products: Are all diterpene cyclases characterized in this work in the proximity to oxidative enzymes?

Without this added context, the presentation of the results is a list of structures, some of which are new, most of which have been reported before, albeit some not in bacteria. As stated earlier the data from this study is very useful, but as it is, the article itself is not suited to be published in this journal with a general audience. I think it could be, however, by rewriting the presentation of the results and adding more background data to give some systematic biological background of the characterized enzymes and contextualize the presented data.
Therefore, I recommend to not accept this article for publication in its current state until it has been rewritten and the suggested additions and data have been introduced.

A few more detailed suggestions to this specific version of the article are listed below.

-The first paragraph of results belongs in the introduction.
-The second paragraph has the Figure-label font.
-lines 68-72: “we selected 334 TSs from 8 phyla, 17 classes, and 83 genera of bacteria for functional analysis.” – Please support this with data. As a suggestion: The authors could generate a schematic bacterial taxonomic tree and show which phyla have terpene cyclases/diterpene cyclases and which ones do not.
-lines 72-74: Fig 1 and Fig S1 does not clearly show the broad phylogenetic distribution of the chosen 334 sequences, since they make up the bulk of the sequences in the tree. For this statement, a more thorough phylogenetic analysis needs to be done including more sequences.
-line 79: Reference 14 is not superscript
-Figure 1: The figure is hard to read. For the bigger version in the SI: Please find a way to label the tips of the tree in a more unambiguous way. Sometimes several enzymes produce the same product, yet the name/ number is only noted once. Also, the names and numbers are just printed in the proximity of the tips, being ambiguous for the reader to interpret. What is the general bootstrap support? Please indicate values or thresholds.
-lines 85-87: TSs were chosen based on sequence diversity, genetic source, location in a unique BGC, and/or high initial yields.” Please support all of this with data. Show the genetic surroundings of the terpene cyclases, explicitly group different enzymes that make the same product (see general suggestions above).
-line 249: This should be relative configuration, not absolute configuration, as NMR is used as the grounds for this statement.
-294-297: “Tetraprenyl-β-curcumenes are often precursors for a 2nd cyclization reaction, but the BGC encoding StMpcS does not have a type II TS nearby suggesting different downstream tailoring modifications.” This is not too surprising since this is a different carbon scaffold.
-line 279: Non-identical NMR means that the relative configuration is not the same. If it is a different enantiomer cannot be seen by NMR.
-lines 350-354: Why were these enzymes chosen for in vitro analysis? Is there anything particular about their activities or the structures of their products?
-lines 379-381: From how the results are presented, this is not evident.

-Methods:
Purification of terpenes: What solvent system was used for the purification of terpene alcohols?

Reviewer #2

(Remarks to the Author)
The manuscript by Wei and colleagues reports the discovery of more than 100 new bacterial diterpene synthases from a screening program starting with “334 diverse terpene synthases from 8 phyla, 17 classes, and 83 genera” (Abstract, line 3). The information content of this paper is incredible, not only to scientists working in the terpene synthase field but also to those with general interests in natural products chemistry and biology. Prior to this work, the magnitude and diversity of bacterial diterpenes was unknown. Importantly, this work describes the discovery of new diterpene skeletons – tetraisoquinene, salbirenol, and chitinotriene are described as representative examples. This work also demonstrates that bacteria can generate diterpenes generated in other organisms such as fungi that have not previously been observed in bacteria, e.g., peysonnosene and variediene. Furthermore, this work reports the discovery of a tridomain bifunctional terpene cyclase similar to enzymes previously known to exists only in fungi and plants, e.g., abietadiene synthase. Like plant abietadiene synthase, the newly discovered bacterial enzyme sandaracopimaradiene synthase catalyzes a class II cyclization of GGPP to generate copalyl diphosphate, which in turn undergoes class I cyclization to generate the eponymous product.

The chemical and biological diversity described in this manuscript is, in a word, staggering – this work represents a major advance for the fields of natural products chemistry and biology. I recommend immediate publication after consideration of the following minor points:

1. Line 57: while prenyltransferases such as farnesyl diphosphate synthase contain two metal-binding Asp-rich motifs, the second of these motifs has diverged to “NSE” or “DTE” in most cyclases. The only exception to this feature to this reviewer’s knowledge is delta-cadinene synthase, a cyclase that indeed contains two Asp-rich motifs. Perhaps a clarification of this sentence might be worthwhile, and especially so if any of the bacterial enzymes discovered contain two Asp-rich motifs.

2. Lines 14, 100, 102, 126, 131, 146, 382: the authors use the word “novel” to indicate previously unobserved, unique, new, and newly discovered. Some journals discourage the use of “novel”, at least in article titles, since it is a word that tends to be overused. The authors might consider using synonyms in place of some occurrences of the word “novel”.

Reviewer #3

(Remarks to the Author)

The manuscript reported a systematic screening of terpene synthases from bacteria, which has been known for producing

much less terpenoids compared to plants and fungi. Based on an engineered E. coli diterpene production system, they screened 334 diverse terpene synthases from 83 genera of bacteria and found 125 active for diterpene production. The isolation and structural elucidation of 28 bacterial diterpenes from 31 terpene synthases revealed three new diterpene skeletons previously unseen in nature, known diterpene skeletons from other organisms with unknown biosynthetic pathways, and new diterpene isomers. The structures of these diterpene products were elucidated based on MS, NMR, and VCD analyses. Besides using the engineered E. coli production system, the authors also carried out in vitro experiments for selected terpene synthases catalysis to confirm their products. These results greatly expand the discovery of diterpenes from bacteria and open the door for future investigation of new biosynthetic pathways. The manuscript is qualified for publication on Nature Communications, and I only have some minor questions/suggestions to be addressed:
1) For the relative configuration of compound 2, did the author really observe the NOE correlation between H-8 and H-19?
2) For compound 4, more NOE correlations need to be added at page 18 in SI.
3) For compound 15, the configuration of the double bond at C10/C11 should be elucidated by NOE.
4) For compound 16, how was the 6R configuration established?
5) Page 1, line 55, the sentence "…have conserved protein folds, they often…" needs to be adjusted.
6) Page 2, line 120, the C and D rings were referred, but no ring numbering was labeled in Figure 2.
7) Line 352, directly putting compound numbers after the enzyme names is a bit misunderstanding.
8) In SI, in all the DFT calculated and experimental VCD figures, carbon numbers should be put before the absolute configuration R or S.

Reviewer #4

(Remarks to the Author)
Research on terpene compounds and their terpene synthases and genes was developed in plants, where terpene compounds are generally understood as plant-specific metabolites, and until the beginning of the 21st century they were treated as extremely rare metabolites from bacteria. However, in 2015, Cane and Ikeda's breakthrough research method, i.e. statistical analysis of the amino acid sequence near the active centre of the terpene synthase using the hidden Markov model instead of sequence-homology search based on BLAST analysis of existing terpene synthases, led to the discovery of an extremely large number of terpene synthases and their genes. This led to the discovery of terpene compounds with novel skeletons. This manuscript is extremely difficult to find originality in view of the ground-breaking research results published in 2015. From 2015 to the present, advances in DNA sequencing and computer analysis have led to an explosion of genome sequences data on organisms, which is now hundreds of times greater than the NCBI database in 2015. It is therefore easy to infer that the 262 novel terpene synthases discovered by Cane and Ikeda in 2015 have now increased to nearly 10 000. The author has also produced diterpenes in genetically engineered E. coli, achieved by co-expression with prenyl diphosphate synthase (in this case geranylgeranyl diphosphate synthase), a resource of substrates for terpene synthases. This method is exactly the same as co-expression with the geranyl-, farnesyl- or geranylgeranyl diphosphate synthase genes by Cane and Ikeda. The only difference is that they use Streptomyces as a host, which is basically the same as the method of Cane and Ikeda. In general, bacterial terpene synthases often form inclusions in the E. coli expression system, even if the codon is optimised. This is less likely to occur when the gene is expressed in strains that are closely related to the strain carrying the gene concerned. This is essentially the same conclusion as their 2015 report. The present manuscript can therefore only be understood as the result of an extension of this brilliant milestone-like work of Cane and Ikeda. Furthermore, Nat. Commun. is generally understood to be a prestigious journal that takes up results based on originality. The results described in this paper are commendable for the recent discovery of a number of diterpene synthases, but considering their originality, they are most appropriate for submission to other leading chemistry journals rather than Nat. Commun.

There are no comments on the text, but the following points only need to be amended.
p. 5, l 428 and 434
'lysogeny broth' was the name given by Dr. Bertani in 1951 when he was working on bacteriophages in Luria's laboratory but was later used by many researchers in the study of E. coli genetics and now it has become commonplace in genetics and other textbooks and laboratory books as the 'Luria-Bertani broth'.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors addressed all my concerns and suggestions either with the requested revisions or provided understandable counterarguments. This revised version of the manuscript and specifically the added analysis of bioinformatic data and TS statistics gives a better overarching picture of where the knowledge about bacterial type 1 TSs stands and makes the objectives and outcomes of this studies much clearer at a big-picture level.
Hence, the manuscript is greatly improved, and this version is in my opinion suitable for publication in this journal.

Reviewer #2

(Remarks to the Author)
The manuscript has been revised to the satisfaction of this reviewer. Immediate publication is recommended.

Reviewer #3

(Remarks to the Author)
The authors have addressed all the concerns from this reveiwer and there is no more comments.

RE: Response to Reviewers

We would like to thank each of the reviewers for their time and effort to review this manuscript and appreciate each of the comments and suggestions. We considered every point very carefully, addressed their questions and suggestions, performed additional bioinformatics analyses, and made appropriate revisions throughout the manuscript and SI. The changes are highlighted in yellow in the "marked" version of the manuscript; they are additionally outlined point-by-point on the pages below. We believe these revisions have improved the manuscript and appreciate the reviewers' thoughtful reviews.

Reviewer 1:

"Wei, Ning and co-workers present a comprehensive survey of bacterial terpene synthases (TSs), in which they identified putative terpene cyclases from the bacterial kingdom from the Uniprot database. They prioritized their hits and ordered synthetic genes for 334 TSs. They screened all the enzymes by heterologous expression in a diterpene producing E. coli platform and find that 125 show diterpene synthase activity. Of these enzymes the authors characterized 31 thoroughly and elucidated the produced structures by NMR, MS and VCD, elucidating their absolute configuration. They go through their characterized TS list and organize the enzymes by their products' representation in previous literature. The chemical data are very thorough and clearly presented in the SI. Generally, these data are very useful to understand TS activities. As there are relatively few TSs characterized it is impossible to predict a TSs activity or even its preferred substrate by substrate alone, given the highly complex reaction they catalyze, the authors state this, too. Towards this aim, the collected data are a valuable contribution, and they were collected by a lot of hard work."

We thank reviewer 1 for their overall positive comments.

Overall suggestion:

"*However, the organization of the article is very hard to follow. The premise of the article as stated in the introduction is to show terpene diversity of the bacterial phylum. Yet the only criterion for grouping is the chemical structure of the compounds and how they are represented in the literature as reported chemical entities (new, new to bacteria, new isomers, and – as a break from this scheme - a single tridomain enzyme). This makes the article repetitive and not easy to read, and the authors themselves state that this organization has drawbacks (several compounds fall into several categories). The methods for structure elucidation and isolation are listed in several sections but are the same. The biological background information (e.g., what phyla of bacteria had how many terpene cyclases per species, which group didn't have any, what is the occurrence of diterpene synthases per phylum etc.), are not comprehensively presented. The prioritization of the TSs from the initial batch (identified enzymes in uniprot) and the further prioritization for structure elucidation is not clearly presented. The authors should show data here. By presenting and connecting the presented chemical data to the biological background data, which the authors should have because of their comprehensive analysis, they could provide a more natural structure to the article.*

If the authors would organize the article by the phylogenetic origin of the discovered diterpene diversity it would be easier to highlight differences, similarities, abundance of terpene cyclases and terpene products (what bacterial phyla produce how many). This could be compared with previously characterized diterpene synthases (newly identified interesting groups of bacteria? Are the all-time classics (actinobacteria) really

the all-time classics? Is there a surprising accumulation of diterpene synthases in certain groups? As diterpenes are not that volatile they are probably not used as volatile natural products: Are all diterpene cyclases characterized in this work in the proximity to oxidative enzymes?

Without this added context, the presentation of the results is a list of structures, some of which are new, most of which have been reported before, albeit some not in bacteria. As stated earlier the data from this study is very useful, but as it is, the article itself is not suited to be published in this journal with a general audience. I think it could be, however, by rewriting the presentation of the results and adding more background data to give some systematic biological background of the characterized enzymes and contextualize the presented data.

Therefore, I recommend to not accept this article for publication in its current state until it has been rewritten and the suggested additions and data have been introduced."

Thank you for the thoughtful comments, constructive feedback, and suggestions. We address each of the main points below; see further below for responses to the more specific comments and suggestions.

Organization: We agree that providing an ideal organization for this article is a challenge. After serious consideration, we prefer to not change the overall organization of the article for the following reasons:

(1) In a study of this size, there will be many conclusions, some of which are more important/impactful than others. We've outlined (in the abstract and organization of the main text) what we think the majority of the community will find most interesting, new diterpene skeletons and skeletons never seen before in bacteria, and put them together at the front of the manuscript. Placing the new skeletons together, in the text and in Fig. 2, also highlights that they are from three different phyla and thus emphasizes that new terpenes are present in underrepresented phyla (i.e., non-actinomycetota).

(2) Given the phylogenetic distribution of TSs in our library (see new revisions regarding biological data and comments above), separating the categories into phyla will not be effective. Some phyla will be overpopulated while others will only have one representative. While this would help convey the lopsided distribution of TSs in bacteria, we feel that point is now made clear with the revisions and new Fig. 1 (thank you for the suggestion).

(3) Writing sections by phylogenetic origin will also create redundancies as several diterpenes, or related isomers, are produced by TSs from different phyla. This would create additional repetitive descriptions for the same or very similar compounds.

We did, however, revise the organization of the introduction and include more details in the genome mining section that we feel improves the manuscript. We also removed the "tridomain, bifunctional" subheading to make it clear that this TS and product fit into the "diterpenes from other organisms" heading.

Repetitive text and not easy to read: We understand this criticism. The nature of the study, the identification and characterization of 31 TSs via diterpene isolation and structural elucidation, creates this problem. To fully appreciate the diversity of diterpenes, TSs, bacteria, and their relation, if any, to what is known in the literature (either within bacteria or not), we feel it is necessary to highlight similarities/differences/knowledge gaps with what is in literature for each compound/enzyme, and to properly cite those papers. In regard to the repetitiveness of the structural elucidation methods, the majority of the data is spectroscopic (NMR, MS, VCD) and therefore must either be described (as is standard in the field) similarly or completely removed. However, we do not feel that removing this already brief description of the data (for the new compounds, full structural elucidation is already in the SI) is beneficial to supporting the diterpene structures presented.

Biological background information: Thank you very much for the suggestion. We have now included the following pieces of biological data along with corresponding text:

(1) A schematic bacterial taxonomic tree (as suggested in comment #3 below) of 43 phyla within bacteria including the total number of putative TSs from the IPR034686 subfamily (Fig. 1A).
(2) A sunburst plot of putative TSs from the IPR034686 revealing the phylogenetic distribution amongst the 13 phyla that encode these TSs (Fig. 1B). We also included additional data and pie charts (Fig. S1) breaking down the TSs present in each phylum based on their Orders and

provide statistics regarding TS per genome. The caveat here is that we used UniProt for putative TSs and NCBI for genomes and UniProt does not contain sequence information for all genome sequences in NCBI, and therefore the values are not exact TS / strain. This is made clear in the legend of Fig. S1.

(3) A pie chart showing the representation and active hits of our TS library (Fig. 1C). We've also included a small table in Fig. 1C that breaks down the previously known di-TSs from these phyla and those from this study, highlighting that we found the first myxobacterial and cyanobacterial TSs and essentially doubled the number of known di-TSs in actinobacteria and bacteroides.

(4) The BGCs of all 31 characterized TSs is now included in Fig. S4 (as suggested in comment #7 below).

Overall, this improves the understanding of the distribution of TSs in bacteria and our contributions to expanding the scope of what is known.

Prioritization data: Respectfully, we feel that further explanation of our initial genome mining does not add anything of value in this study and is not feasible to show. The phylogenetic tree graphically shows sequence diversity, in a similar way to that of a sequence similarity network, multiple sequence alignment, or sequence identity matrix table. Showing the thousands of BGCs that we analyzed will also not be informative, and likely no one will look at them, particularly when readers can simply look them up online using interactive user interfaces.

The prioritization of active TSs was simply as described: "sequence and phylogenetic diversity, location in a unique BGC, and/or high initial yields". The phylogenetic tree and % IDs detailed throughout highlight the sequence diversity; the BGCs encoding all 31 TSs are now included in Fig. S4; and isolation yields are already provided in the SI. Fig. 1C also now contributes to the distribution of TSs selected for isolation.

Detailed suggestions:

1. "*The first paragraph of results belongs in the introduction."*

Revised as suggested.

2. "*The second paragraph has the Figure-label font.*"

Revised as suggested.

3. "*lines 68-72: 'we selected 334 TSs from 8 phyla, 17 classes, and 83 genera of bacteria for functional analysis.' – Please support this with data. As a suggestion: The authors could generate a schematic bacterial taxonomic tree and show which phyla have terpene cyclases/diterpene cyclases and which ones do not.*

Revised as suggested. Please see response to general comment above.

4. "*lines 72-74: Fig 1 and Fig S1 does not clearly show the broad phylogenetic distribution of the chosen 334 sequences, since they make up the bulk of the sequences in the tree. For this statement, a more thorough phylogenetic analysis needs to be done including more sequences.*"

We have previously made phylogenetic trees including more uncharacterized, putative TSs, but they are even harder to read, label (see comment 6 below), and type I TSs are notorious for not being phylogenetic "harmony" with taxonomic trees (see doi: 10.3762/bjoc.15.115 as an example) and di-TSs are so disparate in sequences that they do not clade well together (sesqui-TSs are a little better for that but not helpful here). Therefore, adding more sequences will not solve the phylogenetic distribution question raised here. In addition, as mentioned in the text, TSs with very distinct sequences will sometimes make the same diterpene and therefore will not clade together in a phylogenetic tree. What the tree does show is that the TSs in our library (black, green, and red) are diverse, well-distributed, and it is not obvious based on

sequence alone if TSs are di-TSs or something else. We revised the figure according to comment 6 below, but did not include additional non-characterized TSs from the database for the reasons listed above.

5. "*line 79: Reference 14 is not superscript.*"

Thank you for catching this. Revised as suggested.

6. "*Figure 1: The figure is hard to read. For the bigger version in the SI: Please find a way to label the tips of the tree in a more unambiguous way. Sometimes several enzymes produce the same product, yet the name/ number is only noted once. Also, the names and numbers are just printed in the proximity of the tips, being ambiguous for the reader to interpret. What is the general bootstrap support? Please indicate values or thresholds.*"

Revised as suggested by including arrows for the labels. For bootstrap support, we removed the values for clarity as the values would either be too small to read or too crowded with each other and overlapping with the tree branches. We'll also note that given the diversity of TS sequences from bacteria (now Fig. 1D), general bootstrap values for this tree and ones like it are typically low for nodes near the root. We are not making any evolutionary conclusions based on this tree but are simply showing sequence diversity of the library vs characterized TSs.

7. "*lines 85-87: TSs were chosen based on sequence diversity, genetic source, location in a unique BGC, and/or high initial yields." Please support all of this with data. Show the genetic surroundings of the terpene cyclases, explicitly group different enzymes that make the same product (see general suggestions above).*"

Please see response to general comment above. The BGCs are now shown in Fig. S4. These BGCs also answer the question above regarding "if all TSs are in proximity to oxidative enzymes": many are, but not all.

8. "*line 249: This should be relative configuration, not absolute configuration, as NMR is used as the grounds for this statement.*"

Line 249 refers to odyverdiene B2 (**12**) and its absolute configuration was determined by a combination of NMR and VCD as stated. No change was made.

9. "*294-297: "Tetraprenyl-β-curcumenes are often precursors for a 2nd cyclization reaction, but the BGC encoding StMpcS does not have a type II TS nearby suggesting different downstream tailoring modifications." This is not too surprising since this is a different carbon scaffold.*"

Thank you for the comment. We agree and were not trying to emphasize that it was surprising, only that there are similar skeletons known but that this diterpene and its presumed downstream product, if there is one, are different. We included a reference to the new BGC figure, but made no other changes.

10. "*line 279: Non-identical NMR means that the relative configuration is not the same. If it is a different enantiomer cannot be seen by NMR.*"

Thank you for the comment. The purpose to state that the NMR was not identical was to emphasize that these compounds are not the same and that we needed to determine the absolute configurations of both **16** and **17**, which we did. We clarified this in the text with the revision: "but its NMR was not identical indicating that **16** and **17** are not the same or enantiomers."

11. "*lines 350-354: Why were these enzymes chosen for in vitro analysis? Is there anything particular about their activities or the structures of their products?*"

We included in vitro experiments with six TSs to confirm that the products from the *E. coli* system and in vitro reactions were the same, as we and others sometimes see differences in product profiles between

systems or using different conditions. We selected six TSs as representative examples and all six worked. No change was made.

12. "*lines 379-381: From how the results are presented, this is not evident.*"

We thank the reviewer for the comment and hope that the added biological relevance helps to emphasize this point. We believe that this point is evident because this single study nearly doubled the number of characterized diterpene synthases from bacteria, nearly 40% of screened TSs showed di-TS activity on our first attempt, 3 out of 31 produced novel skeletons (a 10% hit rate for new chemical skeletons is high), and 7 out of 31 produced skeletons never seen before in bacteria. We also revised the conclusion to re-highlight that we characterized the first type I di-TSs in myxobacteria and cyanobacteria, supporting our claim regarding the size of the bacterial terpenome.

13. "*-Methods: Purification of terpenes: What solvent system was used for the purification of terpene alcohols?*"

All terpenes were isolated using the same solvent systems: isocratic hexanes from silica chromatography followed by preparative HPLC using a C18 column. We revised the methods to include our linear gradient solvent system for prep LC.

Reviewer 2:

"The manuscript by Wei and colleagues reports the discovery of more than 100 new bacterial diterpene synthases from a screening program starting with "334 diverse terpene synthases from 8 phyla, 17 classes, and 83 genera" (Abstract, line 3). The information content of this paper is incredible, not only to scientists working in the terpene synthase field but also to those with general interests in natural products chemistry and biology. Prior to this work, the magnitude and diversity of bacterial diterpenes was unknown. Importantly, this work describes the discovery of new diterpene skeletons – tetraisoquinene, salbirenol, and chitinotriene are described as representative examples. This work also demonstrates that bacteria can generate diterpenes generated in other organisms such as fungi that have not previously been observed in bacteria, e.g., peysonnosene and variediene. Furthermore, this work reports the discovery of a tridomain bifunctional terpene cyclase similar to enzymes previously known to exists only in fungi and plants, e.g., abietadiene synthase. Like plant abietadiene synthase, the newly discovered bacterial enzyme sandaracopimaradiene synthase catalyzes a class II cyclization of GGPP to generate copalyl diphosphate, which in turn undergoes class I cyclization to generate the eponymous product.

The chemical and biological diversity described in this manuscript is, in a word, staggering – this work represents a major advance for the fields of natural products chemistry and biology. I recommend immediate publication after consideration of the following minor points:"

We thank reviewer 2 for their extremely positive comments.

Minor points:

1. "*Line 57: while prenyltransferases such as farnesyl diphosphate synthase contain two metal-binding Asp-rich motifs, the second of these motifs has diverged to "NSE" or "DTE" in most cyclases. The only exception to this feature to this reviewer's knowledge is delta-cadinene synthase, a cyclase that indeed contains two Asp-rich motifs. Perhaps a clarification of this sentence might be worthwhile, and especially so if any of the bacterial enzymes discovered contain two Asp-rich motifs."*

We apologize for the oversight here and revised it to clarify the two metal-binding motifs are the DDxxD motif and NSE/DTE triad.

2. "*Lines 14, 100, 102, 126, 131, 146, 382: the authors use the word "novel" to indicate previously unobserved, unique, new, and newly discovered. Some journals discourage the use of "novel", at least in*

article titles, since it is a word that tends to be overused. The authors might consider using synonyms in place of some occurrences of the word "novel".

Thank you for the comment. We revised "novel" to "new" or "previously unobserved in nature" throughout.

Reviewer 3:

"The manuscript reported a systematic screening of terpene synthases from bacteria, which has been known for producing much less terpenoids compared to plants and fungi. Based on an engineered E. coli diterpene production system, they screened 334 diverse terpene synthases from 83 genera of bacteria and found 125 active for diterpene production. The isolation and structural elucidation of 28 bacterial diterpenes from 31 terpene synthases revealed three new diterpene skeletons previously unseen in nature, known diterpene skeletons from other organisms with unknown biosynthetic pathways, and new diterpene isomers. The structures of these diterpene products were elucidated based on MS, NMR, and VCD analyses. Besides using the engineered E. coli production system, the authors also carried out in vitro experiments for selected terpene synthases catalysis to confirm their products. These results greatly expand the discovery of diterpenes from bacteria and open the door for future investigation of new biosynthetic pathways. The manuscript is qualified for publication on Nature Communications, and I only have some minor questions/suggestions to be addressed:"

We thank reviewer 3 for their positive comments.

Minor points:

1. "*For the relative configuration of compound 2, did the author really observe the NOE correlation between H-8 and H-19?"*

Thank you for taking the time to carefully review the NOE data and for catching this. Upon re-inspection of the NOESY data for **2**, we agree that there is not a confident correlation (very weak intensity) between H-8 and Me-19, and this makes sense based on those two groups being on a cyclopentane. This was revised in Table S6.

2. "*For compound 4, more NOE correlations need to be added at page 18 in SI.*"

Thank you for the careful review the NOE data. In the original submission, we only included the single NOE correlation for **4** because there was only one correlation that we could confidently assign (due to significant overlap in proton signals) that was a "key 2D NOESY correlation". We have revised Table S8 to include additional confident NOE correlations, even though they were not instrumental in our assignment of the relative configuration. Given the structure, NMR and VCD data, and literature of the peysonnosides, we are confident in the assigned configuration even without more NOE correlations.

3. "*For compound 15, the configuration of the double bond at C10/C11 should be elucidated by NOE.*"

Thank you for the careful review the NOE data. We included a correlation between H-10 and Me-15 in Table S19.

4. "*For compound 16, how was the 6R configuration established?*"

All 8 possible relative configurations (which gives spectra for all 16 possible isomers) were calculated at 6-31G(d) / B3PW91. There were two configurations, (2*S*,3*S*,6*R*,11*R*) and (2*S*,3*R*,6*R*,11*R*), which matched well. Therefore, we calculated these two configurations at cc-pVTZ / B3PW91 to discriminate between them. Both visual and CompareVoA results show that (2*S*,3*S*,6*R*,11*R*) is the better fit and the correct assignment.

5. "*Page 1, line 55, the sentence "…have conserved protein folds, they often…" needs to be adjusted.*"

We are not sure what the reviewer suggests to be adjusted here, but included “which are” to improve readability.

6. “*Page 2, line 120, the C and D rings were referred, but no ring numbering was labeled in Figure 2.*”

Thank you for pointing this out. The figure was revised as suggested. Upon confirming the previously labeled rings, we misassigned the C and D rings; it is now correctly labeled.

7. “*Line 352, directly putting compound numbers after the enzyme names is a bit misunderstanding.*”

Thank you for the comment. The compound numbers were deleted.

8. “*In SI, in all the DFT calculated and experimental VCD figures, carbon numbers should be put before the absolute configuration R or S.*”

Thank you for the suggestion. We revised the VCD figures and supplementary VCD data file to include carbon numbers and structures (in the figures) for all compounds.

Reviewer 4:

“Research on terpene compounds and their terpene synthases and genes was developed in plants, where terpene compounds are generally understood as plant-specific metabolites, and until the beginning of the 21st century they were treated as extremely rare metabolites from bacteria. However, in 2015, Cane and Ikeda's breakthrough research method, i.e. statistical analysis of the amino acid sequence near the active centre of the terpene synthase using the hidden Markov model instead of sequence-homology search based on BLAST analysis of existing terpene synthases, led to the discovery of an extremely large number of terpene synthases and their genes. This led to the discovery of terpene compounds with novel skeletons. This manuscript is extremely difficult to find originality in view of the ground-breaking research results published in 2015. From 2015 to the present, advances in DNA sequencing and computer analysis have led to an explosion of genome sequences data on organisms, which is now hundreds of times greater than the NCBI database in 2015. It is therefore easy to infer that the 262 novel terpene synthases discovered by Cane and Ikeda in 2015 have now increased to nearly 10 000. The author has also produced diterpenes in genetically engineered E. coli, achieved by co-expression with prenyl diphosphate synthase (in this case geranylgeranyl diphosphate synthase), a resource of substrates for terpene synthases. This method is exactly the same as co-expression with the geranyl-, farnesyl- or geranylgeranyl diphosphate synthase genes by Cane and Ikeda. The only difference is that they use Streptomyces as a host, which is basically the same as the method of Cane and Ikeda. In general, bacterial terpene synthases often form inclusions in the E. coli expression system, even if the codon is optimised. This is less likely to occur when the gene is expressed in strains that are closely related to the strain carrying the gene concerned. This is essentially the same conclusion as their 2015 report. The present manuscript can therefore only be understood as the result of an extension of this brilliant milestone-like work of Cane and Ikeda. Furthermore, Nat. Commun. is generally understood to be a prestigious journal that takes up results based on originality. The results described in this paper are commendable for the recent discovery of a number of diterpene synthases, but considering their originality, they are most appropriate for submission to other leading chemistry journals rather than Nat. Commun.”

We thank and respect reviewer 4 for their review and perspective on terpene synthases in bacteria. We agree that Cane and Ikeda’s seminal work with 2015, along with their previous insight in 2012, were ground-breaking and an inspiration for the field of bacterial terpenoids. Both papers are mentioned and cited in our introduction (refs. 9 and 10).

This manuscript, however, expands what we know about the chemical space and enzymological possibility of terpene biosynthesis in bacteria. This study nearly doubles the number of characterized diterpene synthases from bacteria, includes 3 novel skeletons and 7 additional skeletons never seen before in bacteria, and identifies diterpenes that are produced by distinct enzymes from different bacterial phyla. In

addition, we found the first type I di-TSs from myxobacteria and cyanobacteria. Finally, we identified a tridomain, bifunctional diterpene synthase in bacteria, which was previously only seen in plants and fungi.

Given *Nature Communications'* Aims & Scope to "represent important advances of significance to specialists within each field," we believe this study advances what we know about bacterial diterpenes and will become a launching point for follow-up studies in our lab and around the world, including enzyme mechanisms, natural product discovery, and enzyme engineering. In our lab alone, it has already led to several new collaborations in computational chemistry, structural biology, chemical synthesis, and machine learning, highlighting interest not only in the field of terpenoid natural products and enzymology, but in related chemistry fields.

Minor points:

1. "*There are no comments on the text, but the following points only need to be amended.*
p. 5, l 428 and 434: 'lysogeny broth' was the name given by Dr. Bertani in 1951 when he was working on bacteriophages in Luria's laboratory but was later used by many researchers in the study of E. coli genetics and now it has become commonplace in genetics and other textbooks and laboratory books as the 'Luria-Bertani broth'."

Thank you for the comment. However, no change was made. We prefer to use the acronym as Dr. Bertani intended. According to Dr. Bertani in https://doi.org/10.1128/jb.186.3.595-600.2004, "The acronym [LB] has been variously interpreted, perhaps flatteringly, but incorrectly, as Luria broth, Lennox broth, or Luria-Bertani medium. For the historical record, the abbreviation LB was intended to stand for "lysogeny broth."

Additional revisions:

During review, we realized that one of our TSs, StüTS, had been previously characterized as the sesterterpene synthase sesterviolene synthase (StSS). This is consistent with our result as in our diterpene synthase, only the acyclic β-springene was produced, which led us to hypothesize that was not a true di-TS. This is now stated with the reference cited in the main text, and all StüTS names have been revised to StSS for consistency in the literature.

We included the accession numbers for the raw data NMR deposition to np-mrd.org.

We also made a few minor typo/grammatical revisions to the main text.

Sincerely,

Jeff Rudolf