nature portfolio

Peer Review File

More than just disorder - metabolite diversity of Microcystis strains shows tight correspondence to genotype and may contribute to ecotype specificities

Corresponding Author: Dr Benjamin Marie

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)
In this manuscript, the authors use genomic and metabolomic approaches to characterize the chemical diversity of secondary metabolites produced by the cyanobacteria genus Microcystis. They specifically analyze 65 new strains isolated from Europe (most are from France). The authors perform some comparative genomic analyses of these strains in addition to the other publicly available Microcystis genomes from NCBI. From these analyses the authors find that 57 of the 65 European strains can be classified into 11 genotypes and 13 metabotypes, which correspond with each other, and that similar genotypes may also be considered specific toxico-ecotypes based on toxicological screening. They also find that strains isolated from Europe capture the expansive diversity of this genus, which is consistent with other culture collections from around the world.

Overall, I think there are some very strong points in the manuscript, but there are also additional analyses and considerations that should be addressed. This work significantly expands representation of sequenced Microcystis genomes from Europe, and the conservation between genotype and metabotype has not been this clearly articulated in previous studies. However, additional details are needed for metabolomic annotations, additional analyses could be completed linking metabolomic and genomic data, and some conclusions from the data should be rephrased.

1. Throughout the text, the authors use the term “metabotype” when referring to specific chemical profiles/phenotypes. In my opinion, the term “chemotype” is more appropriate, as the authors focus specifically on secondary metabolites, and do not address the broader metabolome beyond these natural products.
2. The authors should also state more clearly the level of confidence of the identification of each metabolite included in the results section and figures/tables. I recommend doing so by adopting the Schymanski scheme, which is largely used in MS for small and larger metabolites. It is important to be clear on the uncertainties of identification of all presented metabolites, and if in doubt, the confidence level can only be claimed to be level 3. For higher level confidence, manual annotations may need to be completed to assess completeness of predicted spectra (level 2) or reference standards should be used (level 1).
Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J.,
Identifying small molecules via high resolution mass spectrometry: communicating confidence.
Environ Sci Technol 2014, 48, (4), 2097-2098
3. I would have liked to have seen the authors use their data to perform in silico linking between genomic and metabolomic data. There are several tools that are available such as NPlinker and NPOmix that could be used, and the authors should have all the needed data to run this analysis. This would greatly increase the impact of this paper and provide greater understanding of how slight variation in BGC structure may influence congener type production – especially for poorly annotated/undescribed BGCs and chemical features.
4. Lines 192-195: The authors observe levels of sampled strain diversity varies considerably among lakes. This could be due to a multitude of factors, including time of year and nutrient availability. It would be helpful for the authors to include some of this information in table S1 (as is available) and include some details regarding the environmental differences of these lakes during the time of sampling – especially for those with very high and low diversity.
5. Lines 214-215/Figure 1 and Figure S6: I actually think that Figure S6 should replace Figure 1 in the manuscript. Figure S6 contains all the information provided in S6, but S6 also demonstrates the high conservation of genotypes when comparing the core and pangenome. This is the first time I’ve seen this comparison in literature, and I think it’s a point that should be

more prominently displayed in the main text.
6. Lines 274-283/Figure 6: While almost every genotype corresponds nicely to a conserved metabotype, this condition is not met for the genotype i4 which splits into metabotype 1 and 13. I find this extremely interesting and would like to the authors to include more commentary on this finding. Is this a result of different congener production, but largely the same relative abundances of reported metabolites? Are there variations in the composition of BGCs reported – either at the gene content or gene structure level (presence of point mutations, different domain structure, etc)? Having some insight/commentary as why strains that share the same genotype, and are cultured in the identical conditions, but produce different metabotypes would be very valuable.
7. Lines 309-313: The authors should clarify that BGCs are not constitutively expressed in natural microbial communities. Therefore, the presence of diverse secondary metabolites observed in culture do not necessarily reflect metabolite pools observed in nature. I agree that these traits ecotoxicological traits are conserved by genotype in culture, but in a natural system, there may be greater variability of what is produced by different strains.
8. Lines 335-338: I do not understand this sentence. The authors cite 2 papers that look at the specific roles of certain variables (N, P) and the production of certain metabolites – and how it is very challenging to identify specific abiotic variable that drive metabolite production. I'm not sure how potential health repercussions could impact studies attempting to make this link.
9. Lines 341-373 and 393-405: While well these sections provide nice summaries of current literature, I found it to be distracting in the results/discussion section. These sections were also not contextualized with the current findings presented in this paper, and it diverts my attention from the novel results being presented. I would suggest either shortening these sections, moving some of this text to the introduction, or finding ways to better integrate the new results presented in this paper with this text.
10. Lines 432-433: These results are in agreement with previous studies that the authors should consider citing including:
a. https://doi.org/10.3389/fmicb.2019.00791
b. https://doi.org/10.1111/1462-2920.12904
c. https://doi.org/10.1128/msystems.00334-24
11. Line 495-496: Could the authors please specify why 90% identity was set as the threshold? This number is lower than traditional cutoffs used (95% or higher).
12. Line 512: Could the authors please specify which R package was used in the text.
13. Line 535-538: I think this should be stated earlier in the methods, perhaps in the Strain collection and study design section. I think it's important to note that not all the strains are axenic before getting into the details of genomic/metabolomic analysis, and should also be added to Table S1 for reference.
14. Figure 3: I would argue that BGC annotations with less than 50% identity to known clusters should not be reported. There is too much divergence of these sequences at this point, and the genes present may be involved in a different (but similar) biosynthesis pathway. If the authors want to report the presence of these BGCs with greater confidence they could perform read mapping in the event of poor assembly, or provide other rationale for including these annotations.
15. Figure 3 & 4: This figure captions are lengthy and contain text that should be moved to the results section.

Reviewer #2

(Remarks to the Author)
This study pairs genomics with metabolomics to understand the genomic diversity of European (primarily French) strains of Microcystis. Based on a strict threshold, the authors propose a genotype classification based on >99% ANI as the most accurate operational taxonomic unit, rather than genospecies at >97% ANI. This genetic analysis was paired with characterization of corresponding metabotypes.
Major Items:
Overall, the study is an incremental advancement in our understanding of the metabolic capabilities of the genus Microcystis. While the study is worthy of publication, the potential impact of the findings is minimal.
The authors extrapolate the finds to all European strains of Microcystis, but the locations of isolation are primarily located in one country. Isolation sites should encompass lakes in numerous locations or the geographic extrapolation should be reduced.
A discussion of the implications of the interactions within the bloom holobiont with other microbes (bacteria, fungi, viruses) and the potential for evolutionary and ecological consequences reflected in Microcystis genome content and metabolite profiles is missing.
L376- The contrast to Aspergillus here is interesting, but a new comparison to a eukaryotic fungus when all previous context and comparisons have been to other cyanobacteria is discordant. Perhaps a longer discussion here is warranted, especially since we now know that fungi are ubiquitous during blooms and may be contributing to the holobiont metabolome profile.
Specific items:
L40- "corteges" is likely not the correct word to use here
L71- dominate – remove the "s"
L91- Mo or Mb?
L184- should be nine not neine?

Reviewer #3

(Remarks to the Author)
The manuscript by Marie and colleagues provides a comparative genomic, metabolomic and ecotoxicologic study of the ecologically relevant cyanobacterium Microcystis. It provides novel insights into the extensive diversity of Microcystis based on whole genome sequencing data sourced from various locations across western Europe. The manuscript challenges

postulated concepts such as that the genetic distance within Microcystis strains increases with geographical distance. It identifies various Microcystis genospecies encompassing multiple genotypes. A main finding of this study is that genotypes within a genospecies can show far greater metabolic diversity when compared to the metabolic landscape of genotypes of more distantly related genospecies. The data provided further enhances our understanding of environmental adaptations and evolution of bloom forming, toxic freshwater cyanobacteria and allows a glimpse at the rich and largely undiscovered chemical diversity of Microcystis. In general, the manuscript is well structured but legibility partly suffers from typos and misuse of terms. Data interpretation, conclusion and used methodology seem valid. A key problem is a lack of detail in the description of the experimental part that does not allow reproduction (detailed below). Some suggestions and corrections:
L 71 enables… to dominate
L 84-87 revise sentence
L 91 Do you mean Mb? If you mean Mb, the reference that is cited does not support the stated claim. In the cited research the authors claim Microcystis genome sizes range between 3.2 And 4.9 Mb and not 3.5 to 6 Mb.
L 97 …Microcystis exhibit complex genomes
L 99 …arsenals…
L 103 Revise language, “some of these several”
L 116 Environmental sampling “has” shown
L 121 Features
L 123 omit “strain culture”
L 128 remain
L 130 analysis
L 139 “This original approach clearly allows us to explore the functional aspect within the Microcystis tree of life.” functional aspect of what?
L 157 and following: the interchangeable terms genomospecies and genospecies are not used consistently
L 184 nine
L 198 use “.” for decimals
L 229-230 “Some RiPP-BGCs that present intra-genotype variations (microcylamide in i5, …., Fig. 3).” This sentence is incomplete
L238-239 “while subgroups within a genotype appeared much more diversified in terms of metabolites (e.g. Genotype I from i1 to i5)”. The “subgroups” are the genotypes and what is referred to as genotype here is the genospecies. While i1-i5 are genotypes, I represents the genospecies. In Fig S7 the color coding is hard to distinguish, especially the shades of blue for Genospecies I.
L 247 full stop missing after reference 18
L 264 “relative quantities that can be monitored for strains biomass cultured under lab conditions” ? revise sentence
L 303 “… to genotypes i2 and k2…” genotype i2 not presented in fig. 7, should be i1?
Page 12 The discussion focusses solely on diversity on the genomic level and metabolome when cultivated under lab conditions. Might be good to discuss the possibility of regulation of expression of BGCs, as this will allow for immediate environmental plasticity and responses on a metabolic level. Possibly to occupy and thrive in ecological niches. Differential regulation of BGCs could also depict another level of diversity within the same genotype.
L 354 remain
L 360 remove “)”
L 367 cyanobacteria-specific
L 389 novel novelties
L 422 patterns
L 432 a significant portion
L 470-471 rephrase, cultures are not performed, cultivation volume is missing, rpm and shaker geometry is missing
L 472 revise unit “µmol.m-2.s-1”
L 473 how long was the centrifugation step?
L 475) missing
L 480 inconsistent citation style
L 509 space missing
L 518 what is the reasoning for selecting a cutoff of 5.5 kb? Please elaborate
L 529 Rephrase, BGCs are not annotated as terpenes,
Why were BGCs that give rise to terpenes excluded?
L 535 and following. Lack of information necessary for reproducibility
L 548 triplicates
L 546 and following. Lack of information. You do not state the cultivation volume. If you remove 12 ml twice a week from the culture and you cultivate for 28 days (4 weeks) you end up removing 96 ml of culture if you do not replenish it. Thus, you not only monitor the variations of intracellular metabolite contents as a function of time but the drastic differences in culture volume will have wide ranging effects.
L 587 do you mean peak area?

Fig.1 2x black circles in figure legend and figure. Maybe change colour or shape.
L1066 ‘clade; 3, known BGC…’
L1072-1073 How does the size of the circles indicate percentage of identity? Small circles, filled circles, hollow circles? Also, this data was not retrieved experimentally but only from genome analysis?
L1068 ‘terpene’
L1077-1086 This is a discussion
L 1085 fell to be recovered? Use core or leader peptide instead of pro peptide
L1093-1097 This is also a discussion, should go to the main text

Figure 4
The dendrogram on top and left is the same? One might be enough then.
Figure 5
The figure is referred to once in the text and is not adressed. Hard to read and understand based on the description given. Is it important or can it go to the supplement?
Each circle represents 1 molecule? And size represents the quantity? What are the number in the circles?
Figure 7
“(%age)” [%] is sufficient

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Biology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have carefully and thoughtfully addressed most of my comments and suggestions. It appears that they have also taken careful time to review and revise the manuscript based on the suggestions of other reviewers as well. Most of my points have been sufficiently addressed, but there are still some questions or suggestions that need more careful review and consideration. While it may seem pedantic, I encourage the authors to revisit some of the points I previously raised, which were either insufficiently addressed or dismissed without adequate justification. Given the visibility and impact of this journal, it is essential that the manuscript reflects the highest standards of rigor and clarity, especially in its interpretation, to ensure it serves as a reliable resource for a broad and diverse readership. Below are my follow up comments for those points of interest.

Question 3:
I would have liked to have seen the authors use their data to perform in silico linking between genomic and metabolomic data. There are several tools that are available such as NPlinker and NPOmix that could be used, and the authors should have all the needed data to run this analysis. This would greatly increase the impact of this paper and provide greater understanding of how slight variation in BGC structure may influence congener type production – especially for poorly annotated/undescribed BGCs and chemical features.
Answer: We thank the reviewer for this remark as in silico screening of joint genomic and metabolomic dataset have been proposed to assist the discovery of novel BGC and related products based on co-occurrence detection within a large strain panel. However, in silico tools such as NPlinker and NPOmics (which by the way is already not supported anymore) have not demonstrated significant contribution. Indeed, preliminary description of novel BGC and potentially related metabolite family based on co-occurrence in certain strains present limited evidence, that would still require to be validated and supported by extended and in-depth direct structural and biosynthesis characterization. For this reason, we assume not to solely provide additional in silico screening in the present paper and consider that the already provided information (supplementary figure S11 and supplementary table S2), could already support valuably the discovery of novel metabolite family. However, according to the reviewer comments we have added to the discussion several sentences dedicated at the potential of the present dataset to further initiate the pointed characterization of novel metabolite families and related BGC: “Our dataset offers novel opportunities for discovering new BGCs associated within metabolites families. Future effort would now be specifically dedicated for charactering such novel molecules and potentially corresponding BGC. For example, the 2 stains from Chem06 (PMC 673 and 674) both present several uncharacterized RiPP BGCs (presenting faint similarities with PcpA, hglE and lanthipeptide class-v, respectively), together with a singular molecular cluster of 7 unknown metabolites ranging from around 1052-1092 Da, that aim at being further de novo characterized by direct structural analysis (supplementary figure S11; supplementary table S2a and 2G).”.

Response: While I can understand the perspective of the authors, I believe that additional analyses would enhance rigor, thereby making it a more suitable manuscript for publication. The authors have generated a very rich dataset, and it would not require too much effort to perform in silico linking, or co-occurance/similarity networking with other tools. The field is currently flooded with publications that list strain BGCs and produced metabolites, without making real efforts to try to link these data. While I agree that these tools are not sufficient to directly link an orphan BGC and metabolite, they provide valuable clues to other researchers and support hypothesis generation for future metabolite isolation and characterization. There are numerous publications to support these types approaches including tools such as NPlinker (previously mentioned), NPClassScore (DOI https://doi.org/10.1186/s40168-022-01444-3), and additional MetCalf scoring methods. Other publications have also validated the utility of these tools (DOI: https://doi.org/10.1186/s40168-022-01444-3, DOI : https://doi.org/10.3390/md19020103). It is becoming more common to at least attempt some hypothetical linking (if experimental validation is not completed). This would greatly improve the impact of this paper, and put it more in line in the direction the field is moving.

Question 4:
Lines 192-195: The authors observe levels of sampled strain diversity varies considerably among lakes. This could be due to a multitude of factors, including time of year and nutrient availability. It would be helpful for the authors to include some of this information in table S1 (as is available) and include some details regarding the environmental differences of these lakes during the time of sampling – especially for those with very high and low diversity.
Answer: We thank the reviewer for this comment and understand the interest he(she) put on trying to explain or to better understand ecological influence on the strain diversity. We are also concerned by the fact that a single sampling strategy is far from being enough to characterize the complexity of ecological influence of bloom genetic diversity and could constitute a quite finalistic deliverable. For this reason, and because we do not possess sufficiently informative environmental dataset concerning the sampled lake ecological features, we assume not to further investigate this question regarding the present dataset. However, we have now added basic information concerning the sampling origin of the strains within supplementary table S1.

Response : Thank you for including additional information in Table S1. My original comment was directed toward the new European strains that were described for the first time in the manuscript. I'm not at all suggesting that the authors should try to make assumptions of ecological complexity based on the isolation of single strains (which comes with it's own set of biases). However, I still think it is important to provide basic meta data where available, especially for the new strains discussed. Ideally I would like to see isolation year, collection date (or month if exact date is unknown), and a general trophic status for the sampled lake (oligotrophic, eutrophic, hypereutrophic, etc.). These parameters should have been collected during the time of sampling and are important details that should be included in the manuscript.

Additionally, the text in Figure 1 is pretty small for each individual strain and is hard to read. Would it be possible to add the genotype and chemotype to this table as well for easy reference ? I found it challenging to read even in the .tif file.

Question 6:
Lines 274-283/Figure 6: While almost every genotype corresponds nicely to a conserved metabotype, this condition is not met for the genotype i4 which splits into metabotype 1 and 13. I find this extremely interesting and would like to the authors to include more commentary on this finding. Is this a result of different congener production, but largely the same relative abundances of reported metabolites? Are there variations in the composition of BGCs reported – either at the gene content or gene structure level (presence of point mutations, different domain structure, etc)? Having some insight/commentary as why strains that share the same genotype, and are cultured in the identical conditions, but produce different metabotypes would be very valuable.
Answer: Thank you for this remark. This aspect has now been directly discussed, accordingly: "Although, strains from chem01 and chem13 belong to the same genotype and present globally similar metabolite composition (comprising microcystins, aeruginosins, cyanopeptolins and microcyclamides) they also present noticeable molecular distinction characterized by variations in variant contents, that appears congruent with the respective genotypic position of the strains".

Response : Thank you for addressing my initial comment. Now with the additional information reported in Table S1, it appears the genotype i4 is represented by strains of Microcystis that are both xenic and axenic. The presence of other microbial taxa can influence secondary metabolite production in terms variants produced, and regulation of biosynthetic pathways. The authors should include some additional text that discusses that axenic status can also influence differences in chemotypes of strains that share the same genotype.

Question 7:
Lines 309-313: The authors should clarify that BGCs are not constitutively expressed in natural microbial communities. Therefore, the presence of diverse secondary metabolites observed in culture do not necessarily reflect metabolite pools observed in nature. I agree that these traits ecotoxicological traits are conserved by genotype in culture, but in a natural system, there may be greater variability of what is produced by different strains.
Answer: We thank the rewiever for this judicious remark. We have now added to the discussion a sentence highlighting this limitation of experimental approach that could not fully represent to what could happen in naturae as BGC expression slightly vary regarding local environmental factors: "Also, several key abiotic of biotic factors (e.g. nutrient availability, temperature, light intensity, or allelopathy and predation) may influence BGC expression and slightly influence the composition of the Microcystis metabolite cocktails, potentially modulating the subsequent ecotoxicological threats that may be observed in both experimental and environmental context".

Response : Thank you once again for addressing my comment. However, I am concerned that the use of the word 'slightly' greatly downplays the significat impact that abiotic and biotic factors have on secondary metabolite production, especially in situ. This is well documented in the literature, mostly for the production of microcystin, but for other secondary metabolites as well. Work that immediately comes to mind that highlight this point includes DOI : 10.1007/s10452-016-9571-6.

Question 14:
Figure 3: I would argue that BGC annotations with less than 50% identity to known clusters should not be reported. There is too much divergence of these sequences at this point, and the genes present may be involved in a different (but similar) biosynthesis pathway. If the authors want to report the presence of these BGCs with greater confidence they could perform read mapping in the event of poor assembly, or provide other rationale for including these annotations.
Answer: We thank the reviewer for this comment. We believe he(she) is judicious to notice that similar score between

suggest that these BGC might produce similar molecules, when higher score may show more confidence in the proposed metabolite family. However, the consideration of a universal threshold for AntiSmash annotation is a complex concern that require global vista of the scoring principle (based on structural similarity to a reference sequence) regarding the respective architecture of the distinct BGC families. Indeed, when NRPS BGC of the same family present global high sequence similarity, when presenting little modifications in RIPP-BGC sequences, can strongly influence the structure of the product and its annotation. Thus, it would maybe appears adapted to consider distinct significance scores for the distinct BGC families. Although, lower score is lesser relevant in term of genuine annotation, it remains informative in term of sequence similarity that present homogenous pattern between strains of the same genotypes. For these reasons, we would rather keep this information accessible for the reader.

Response : I stand firmly with my initial comment regarding reporting BGC annotations with antismash below a 50% threshold. This is misleading, and has caused amplification in the literature of mis-reporting of biosynthetic gene clusters being present in strains or populations when they are not. Several biosynthetic genes such as those encoding NRPSs and PKSs have highly conserved regions that can sporadically, and incorrectly, map and hit to known BGCs when using antismash. Since the authors have high quality, assembled genomes, it is highly unlikely that these are artifacts of fragmented assembly. If the authors insist on reporting these data it should be done in a table or other figure where the % identity and alignment length are reported with the gene cluster. Alternatively, the authors could rename these clusters generically to reflect biosynthesis pathways for example RiPP_1, RiPP_2, etc.

Reviewer #3

(Remarks to the Author)
The manuscript by Marie et al was improved significantly, still some critical information is missing and several points made in the last round of review were not adressed.
Remove "cyanobacteria" from line 396
In the material and method section L525 the information for how much cultivation volume was used is still missing, thus no reproduction possible. Line 606, also here information about vessel volume is missing.
Line 535 remove .

Besides these minor points i suggest acceptance of the manuscript.

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Biology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewers' comments:

Reviewer #1: secondary metabolite biosynthesis in cyanobacteria, metabolomics, genomics, Microcystis

(Remarks to the Author):

In this manuscript, the authors use genomic and metabolomic approaches to characterize the chemical diversity of secondary metabolites produced by the cyanobacteria genus Microcystis. They specifically analyze 65 new strains isolated from Europe (most are from France). The authors perform some comparative genomic analyses of these strains in addition to the other publicly available Microcystis genomes from NCBI. From these analyses the authors find that 57 of the 65 European strains can be classified into 11 genotypes and 13 metabotypes, which correspond with each other, and that similar genotypes may also be considered specific toxico-ecotypes based on toxicological screening. They also find that strains isolated from Europe capture the expansive diversity of this genus, which is consistent with other culture collections from around the world.

Overall, I think there are some very strong points in the manuscript, but there are also additional analyses and considerations that should be addressed. This work significantly expands representation of sequenced Microcystis genomes from Europe, and the conservation between genotype and metabotype has not been this clearly articulated in previous studies. However, additional details are needed for metabolomic annotations, additional analyses could be completed linking metabolomic and genomic data, and some conclusions from the data should be rephrased.

Question 1:
Throughout the text, the authors use the term “metabotype” when referring to specific chemical profiles/phenotypes. In my opinion, the term “chemotype” is more appropriate, as the authors focus specifically on secondary metabolites, and do not address the broader metabolome beyond these natural products.
Answer: Thank you for this remark. We have now replaced the term metabotype by chemotype in all the text, figures and supplementary materials, accordingly.

Question 2:
The authors should also state more clearly the level of confidence of the identification of each metabolite included in the results section and figures/tables. I recommend doing so by adopting the Schymanski scheme, which is largely used in MS for small and larger metabolites. It is important to be clear on the uncertainties of identification of all presented metabolites, and if in doubt, the confidence level can only be claimed to be level 3. For higher level confidence, manual annotations may need to be completed to assess completeness of predicted spectra (level 2) or reference standards should be used (level 1).
Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J., Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol 2014, 48, (4), 2097-2098
Answer: We thank the reviewer for this remark. This information has now been added in the M&M and in the supplementary table S2A, accordingly.

Question 3:
I would have liked to have seen the authors use their data to perform in silico linking between genomic and metabolomic data. There are several tools that are available such as NPlinker

and NPOmix that could be used, and the authors should have all the needed data to run this analysis. This would greatly increase the impact of this paper and provide greater understanding of how slight variation in BGC structure may influence congener type production – especially for poorly annotated/undescribed BGCs and chemical features.
Answer: We thank the reviewer for this remark as in silico screening of joint genomic and metabolomic dataset may assist the discovery of novel BGC and related products. However, any description of novel BGC and/or metabolite family might be supported by solid structural validation dataset, according to further abundant investigations that are beyond the scope of the present paper. For this reason, we assume not to solely provide in silico screening in the present paper and rather keep our purpose focused on the main message already provided by our data treatment.

Question 4:
Lines 192-195: The authors observe levels of sampled strain diversity varies considerably among lakes. This could be due to a multitude of factors, including time of year and nutrient availability. It would be helpful for the authors to include some of this information in table S1 (as is available) and include some details regarding the environmental differences of these lakes during the time of sampling – especially for those with very high and low diversity.
Answer: We thank the reviewer for this comment and understand the interest he(she) put on trying to explain or to better understand ecological influence on the strain diversity. We are also concerned by the fact that a single sampling strategy is far from being enough to characterize the complexity of ecological influence of bloom genetic diversity and could constitute a quite finalistic deliverable. For this reason, and because we do not possess sufficiently informative environmental dataset concerning the sampled lake ecological features, we assume not to further investigate this question regarding the present dataset. However, we have now added basic information concerning the sampling origin of the strains within supplementary table S1.

Question 5:
Lines 214-215/Figure 1 and Figure S6: I actually think that Figure S6 should replace Figure 1 in the manuscript. Figure S6 contains all the information provided in S6, but S6 also demonstrates the high conservation of genotypes when comparing the core and pangenome. This is the first time I've seen this comparison in literature, and I think it's a point that should be more prominently displayed in the main text.
Answer: We thank the reviewer for this comment and share his(her) enthusiasm on the interest of considering the high conservation of genome relationship regarding core and pan genomes, and have now inserted fig S6 as a regular figure.

Question 6:
Lines 274-283/Figure 6: While almost every genotype corresponds nicely to a conserved metabotype, this condition is not met for the genotype i4 which splits into metabotype 1 and 13. I find this extremely interesting and would like to the authors to include more commentary on this finding. Is this a result of different congener production, but largely the same relative abundances of reported metabolites? Are there variations in the composition of BGCs reported – either at the gene content or gene structure level (presence of point mutations, different domain structure, etc)? Having some insight/commentary as why strains that share the same genotype, and are cultured in the identical conditions, but produce different metabotypes would be very valuable.
Answer: Thank you for this remark. This aspect has now been directly discussed, accordingly: "Although, strains from chem01 and chem13 belong to the same genotype

and present globally similar metabolite composition (comprising microcystins, aeruginosins, cyanopeptolins and microcyclamides) they also present noticeable molecular distinction characterized by variations in variant contents, that appears congruent with the respective genotypic position of the strains".

Question 7:
Lines 309-313: The authors should clarify that BGCs are not constitutively expressed in natural microbial communities. Therefore, the presence of diverse secondary metabolites observed in culture do not necessarily reflect metabolite pools observed in nature. I agree that these traits ecotoxicological traits are conserved by genotype in culture, but in a natural system, there may be greater variability of what is produced by different strains.
Answer: We thank the rewiever for this judicious remark. We have now added to the discussion a sentence highlighting this limitation of experimental approach that could not fully represent to what could happen *in naturae* as BGC expression slightly vary regarding local environmental factors: "Also, several key abiotic of biotic factors (*e.g.* nutrient availability, temperature, light intensity, or allelopathy and predation) may influence BGC expression and slightly influence the composition of the *Microcystis* metabolite cocktails, potentially modulating the subsequent ecotoxicological threats that may be observed in both experimental and environmental context".

Question 8:
Lines 335-338: I do not understand this sentence. The authors cite 2 papers that look at the specific roles of certain variables (N, P) and the production of certain metabolites – and how it is very challenging to identify specific abiotic variable that drive metabolite production. I'm not sure how potential health repercussions could impact studies attempting to make this link.
Answer: This sentence has been reworded in order to better fit with the flow of the discussion and our general purpose, accordingly: "Interestingly, Jackrel and co-wokers[10] and latter Kuijpers and co-wokers[71] have correlated global genome reductions and expansions in gene supporting specific metabolic capabilities (*i.e.* N- and P-related genes) in relation with specific ecological features of their originating lakes (low- or high-nutrient lakes).".

Question 9:
Lines 341-373 and 393-405: While well these sections provide nice summaries of current literature, I found it to be distracting in the results/discussion section. These sections were also not contextualized with the current findings presented in this paper, and it diverts my attention from the novel results being presented. I would suggest either shortening these sections, moving some of this text to the introduction, or finding ways to better integrate the new results presented in this paper with this text.
Answer: These paragraphs have now been significantly shortened, accordingly.

Question 10:
Lines 432-433: These results are in agreement with previous studies that the authors should consider citing including:
a. https://doi.org/10.3389/fmicb.2019.00791
b. https://doi.org/10.1111/1462-2920.12904
c. https://doi.org/10.1128/msystems.00334-24
Answer: Citations have been added, as suggested.

Question 11:

Line 495-496: Could the authors please specify why 90% identity was set as the threshold? This number is lower than traditional cutoffs used (95% or higher).
Answer: Indeed, the identity threshold may depend on the taxa considered. In the present case, the investigation of a largely diversified *Microcystis* clade justify the use of a slightly lower threshold. However, although the use of a higher threshold decreases the number of orthologous gene considered it does not presently modify the global phylogenetic picture observed. With 90%-threshold

Question 12:
Line 512: Could the authors please specify which R package was used in the text.
Answer: This information has been specified.

Question 13:
Line 535-538: I think this should be stated earlier in the methods, perhaps in the Strain collection and study design section. I think it's important to note that not all the strains are axenic before getting into the details of genomic/metabolomic analysis, and should also be added to Table S1 for reference.
Answer: As the reference of the axenicity of the culture co concern the biomass production and not collection condition of the respective strains, we assume to keep this description in this specific position of the M&M. However, the axenicity statue of the strains in PMC and PCC collection have been added in table S1.

Question 14:
Figure 3: I would argue that BGC annotations with less than 50% identity to known clusters should not be reported. There is too much divergence of these sequences at this point, and the genes present may be involved in a different (but similar) biosynthesis pathway. If the authors want to report the presence of these BGCs with greater confidence they could perform read mapping in the event of poor assembly, or provide other rationale for including these annotations.
Answer: We thank the reviewer for this comment. We believe he(she) is judicious to notice that similar score between suggest that these BGC might produce similar molecules, when higher score may show more confidence in the proposed metabolite family. However, the consideration of a universal threshold for AntiSmash annotation is a complex concern that require global vista of the scoring principle (based on structural similarity to a reference sequence) regarding the respective architecture of the distinct BGC families. Indeed, when NRPS BGC of the same family present global high sequence similarity, when presenting little modifications in RIPP-BGC sequences, can strongly influence the structure of the product and its annotation. Thus, it would maybe appears adapted to consider distinct significance scores for the distinct BGC families. Although, lower score is lesser relevant in term of genuine annotation, it remains informative in term of sequence similarity that present homogenous pattern between strains of the same genotypes. For these reasons, we would rather keep this information accessible for the reader.

Question 15:
Figure 3 & 4: This figure captions are lengthy and contain text that should be moved to the results section.
Answer: Several sentences of Fig. 3 and 4 legends have now been moved to the main text of the MS, accordingly: " Overall, these results illustrate the large variation observed between the different European *Microcystis* strains (ranging from 3.8 to 5.9 Mb),

together with an overall homogeneity within the distinct genetic clades. The accessory metabolites of some genotypes are highly similar despite the diversity of the strain origins (*e.g.* genotype i5 containing isolated strains from north and south of France), while chemotypes within certain genotypes appeared much more diversified than between certain others (*e.g.* within I compared to between D and T; supplementary Fig. S7). It is also worth noticing that although the two neighboring strains belonging to the genospecies P (namely PMC 566.08 and PMC 568.08) did not constitute a genuine genotype (their ANI score being above 97%, but below 99% thresholds), the metabolomes of these two strains were grouped together within a single chemotype (see figure 4).”
“We observed that strain clusters discriminated at the genomic level considering either core- and pan-genome or BGC contents strictly correspond to characteristic chemotypes producing well determined metabolite categories. It is worth noticing that this chemotype distinction was remarkably identical to those retrieved from the merged data matrix obtained by combining both analyses performed on positive and negative MS mode (Supplementary Fig. S6) and that the chemotype belonging of strains was largely stable across successive cultures (Supplementary Fig. S7-8). Only minor variations in terms of molecular variants and/or relative metabolite quantities can be monitored regarding the biomass of strains cultured under laboratory conditions (Supplementary Fig. S8-10).”

Referee #2: Mycrocystis, genomics

(Remarks to the Author):

This study pairs genomics with metabolomics to understand the genomic diversity of European (primarily French) strains of Microcystis. Based on a strict threshold, the authors propose a genotype classification based on >99% ANI as the most accurate operational taxonomic unit, rather than genospecies at >97% ANI. This genetic analysis was paired with characterization of corresponding metabotypes.

Major Items:
Overall, the study is an incremental advancement in our understanding of the metabolic capabilities of the genus Microcystis. While the study is worthy of publication, the potential impact of the findings is minimal.

Question 1:
The authors extrapolate the finds to all European strains of Microcystis, but the locations of isolation are primarily located in one country. Isolation sites should encompass lakes in numerous locations or the geographic extrapolation should be reduced.
Answer: The geographic extrapolation has now been reduced and further considered as strains “originating from France and surrounding countries”, accordingly.

Question 2:
A discussion of the implications of the interactions within the bloom holobiont with other microbes (bacteria, fungi, viruses) and the potential for evolutionary and ecological consequences reflected in Microcystis genome content and metabolite profiles is missing.
Answer: We thank the reviewer for this comment. A specific sentence on chemical interaction within the microbial community have been added, accordingly: “Indeed, several cyanobacterial metabolites have been proposed to play key ecological functions

(*e.g.* allelopathy, anti-grazing or antimicrobial) within the microbial communities and may contribute to the blooming capabilities of these micro-organisms."

Question 3:
L376- The contrast to Aspergillus here is interesting, but a new comparison to a eukaryotic fungus when all previous context and comparisons have been to other cyanobacteria is discordant. Perhaps a longer discussion here is warranted, especially since we now know that fungi are ubiquitous during blooms and may be contributing to the holobiont metabolome profile.
Answer: This paragraph has been reworded and the reference to Aspergillus better integrated to the rest of the discussion, accordingly: "The patterns of diversity that we observe here suggest that each distinct *Microcystis* clade generates and maintains biochemical diversity in particular ways. In contrast to certain other micro-organisms, where the chemical diversity increases linearly with taxonomic repertoire and genome size,[88] freshwater cyanobacteria demonstrate noteworthy examples of convergence in their accessory metabolite structural diversity, that do not appear systematically related to the genome size of the different taxa."

Specific items:
Question 4:
L40- "corteges" is likely not the correct word to use here
Answer: This word has been replaced by "cocktails"

Question 5:
L71- dominate – remove the "s"
Answer: This has been corrected. Thank you.

Question 6:
L91- Mo or Mb?
Answer: Mb, this has been corrected. Thank you.

Question 7:
L184- should be nine not neine?
Answer: This has been corrected. Thank you.

Referee #3 & #4 (co-reviewers): cyanobacteria metabolism

(Remarks to the Author):

The manuscript by Marie and colleagues provides a comparative genomic, metabolomic and ecotoxicologic study of the ecologically relevant cyanobacterium Microcystis. It provides novel insights into the extensive diversity of Microcystis based on whole genome sequencing data sourced from various locations across western Europe. The manuscript challenges postulated concepts such as that the genetic distance within Microcystis strains increases with geographical distance. It identifies various Microcystis genospecies encompassing multiple genotypes. A main finding of this study is that genotypes within a genospecies can show far greater metabolic diversity when compared to the metabolic landscape of genotypes of more distantly related genospecies. The data provided further enhances our understanding of environmental adaptations and evolution of bloom forming, toxic freshwater cyanobacteria and allows a glimpse at the rich and largely undiscovered chemical diversity of Microcystis.

In general, the manuscript is well structured but legibility partly suffers from typos and misuse of terms. Data interpretation, conclusion and used methodology seem valid. A key problem is a lack of detail in the description of the experimental part that does not allow reproduction (detailed below). Some suggestions and corrections:

Question 1:
L 71 enables… to dominate
Answer: This has been corrected. Thank you.

Question 2:
L 84-87 revise sentence
Answer: This sentence has been revised: "Since then, fully sequenced genomes of *Microcystis* strains have been progressively providing unprecedented taxonomic resolution and have described a distinct phylogenetic picture compare to those previously identified through morphology-based taxonomy."

Question 3:
L 91 Do you mean Mb? If you mean Mb, the reference that is cited does not support the stated claim. In the cited research the authors claim Microcystis genome sizes range between 3.2 And 4.9 Mb and not 3.5 to 6 Mb.
Answer: Mb, this has been corrected. The place of this citation has been moved, accordingly. Thank you.

Question 4:
L 97 …Microcystis exhibit complex genomes
Answer: This has been corrected. Thank you.

Question 5:
L 99 …arsenals…
Answer: This has been corrected.

Question 6:
L 103 Revise language, "some of these several"
Answer: This has been corrected.

Question 7:
L 116 Environmental sampling "has" shown
Answer: This has been corrected.

Question 8:
L 121 Features
Answer: This has been corrected.

Question 9:
L 123 omit "strain culture"
Answer: This has been modified.

Question 10:
L 128 remain
Answer: This has been corrected.

Question 11:
L 130 analysis
Answer: This has been corrected.

Question 12:
L 139 “This original approach clearly allows us to explore the functional aspect within the Microcystis tree of life.” functional aspect of what?
Answer: This has now been specified: “ecological functional aspect of this diversity”. Thank you.

Question 13:
L 157 and following: the interchangeable terms genomospecies and genospecies are not used consistently
Answer: As geno-species (genotypes presenting similar genetic markers, such as 16S sequence score) and genomo-species (genotypes presenting a global genome similarity, such as ANI score) have near but distinct meaning, we assume their proper use in the text.

Question 14:
L 184 nine
Answer: This has been corrected.

Question 15:
L 198 use “.” for decimals
Answer: This has been corrected.

Question 16:
L 229-230 “Some RiPP-BGCs that present intra-genotype variations (microcylamide in i5, …., Fig. 3).” This sentence is incomplete
Answer: This sentence has been reworded, accordingly: “While the BGC presence for NRPS/PKS and/or their corresponding metabolite detection are highly conserved within genotypes, but not into the genomo-species, some RiPP-BGCs present intra-genotype variations (microcylamide in i5, aeruginosamide in k2 and piricyclamide and microviridin in t1, Fig. 3).”

Question 17:
L238-239 “while subgroups within a genotype appeared much more diversified in terms of metabolites (e.g. Genotype I from i1 to i5)”. The “subgroups” are the genotypes and what is referred to as genotype here is the genospecies. While i1-i5 are genotypes, I represents the genospecies.
Answer: We understand that the formulation of this sentence maybe confusing and we have modified it, accordingly: “The accessory metabolites of some genotypes are highly similar despite the diversity of the strain origins (*e.g.* genotype i5 containing isolated strains from north and south of France), while metabotypes within certain genotypes appeared much more diversified than between certain others (*e.g.* within I compared to between D and T; supplementary Fig. S7).”.

Question 18:

L 247 full stop missing after reference 18
Answer: This has been corrected.

L 264 “relative quantities that can be monitored for strains biomass cultured under lab conditions” ? revise sentence
Answer: These sentences have been reworded: “We observed that strain clusters discriminated at the genomic level considering either core- and pan-genome or BGC contents strictly correspond to characteristic metabotypes producing well determined metabolite categories. Indeed, only minor variations in terms of molecular variants and/or relative metabolite quantities can be monitored regarding the biomass of strains cultured under laboratory conditions (Supplementary Fig. S8-10).”

Question 19:
L 303 “… to genotypes i2 and k2…” genotype i2 not presented in fig. 7, should be i1?
Page 12 The discussion focusses solely on diversity on the genomic level and metabolome when cultivated under lab conditions. Might be good to discuss the possibility of regulation of expression of BGCs, as this will allow for immediate environmental plasticity and responses on a metabolic level. Possibly to occupy and thrive in ecological niches. Differential regulation of BGCs could also depict another level of diversity within the same genotype.
Answer: Thank you. This has been corrected on Fig. 7. Also, we have added to the discussion a specific sentence discussing environmental metabolome regulations and potential ecotoxicological consequences.

Question 20:
L 354 remain
Answer: This has been corrected.

Question 21:
L 360 remove “)”
Answer: This has been corrected.

Question 22:
L 367 cyanobacteria-specific
Answer: This has been corrected.

Question 23:
L 389 novel novelties
Answer: This has been corrected.

Question 24:
L 422 patterns
Answer: This has been corrected.

Question 25:
L 432 a significant portion
Answer: This has been corrected.

Question 26:
L 470-471 rephrase, cultures are not performed, cultivation volume is missing, rpm and shaker geometry is missing

Answer: This sentence has been reworded: "To sequence the new *Microcystis* genomes from PMC, cultures were grown in Z8 medium[104] at 25°C in 250-mL Erlenmeyer flasks. The cultures were exposed to a photon flux density of 20 µmol.m^{-2}.s^{-1} and maintained on a 16:8h light:dark cycle."

Question 27:
L 472 revise unit "µmol.m-2.s-1"
Answer: This has been corrected.

Question 28:
L 473 how long was the centrifugation step?
Answer: This information has been added.

Question 29:
L 475) missing
Answer: This has been corrected.

Question 30:
L 480 inconsistent citation style
Answer: This has been corrected.

Question 31:
L 509 space missing
Answer: This has been corrected.

Question 32:
L 518 what is the reasoning for selecting a cutoff of 5.5 kb? Please elaborate
Answer: Answer: A citation have been added.

Question 33:
L 529 Rephrase, BGCs are not annotated as terpenes,
Why were BGCs that give rise to terpenes excluded?
Answer: This sentence has been corrected and specified.

Question 34:
L 535 and following. Lack of information necessary for reproducibility
Answer: This M&M section have been improved to ensure reproducibility, accordingly.

Question 35:
L 548 triplicates
Answer: This has been corrected.

Question 36:
L 546 and following. Lack of information. You do not state the cultivation volume. If you remove 12 ml twice a week from the culture and you cultivate for 28 days (4 weeks) you end up removing 96 ml of culture if you do not replenish it. Thus, you not only monitor the variations of intracellular metabolite contents as a function of time but the drastic differences in culture volume will have wide ranging effects.
Answer: We thank the reviewer for this interesting remark. Indeed, we placed a careful attention at the experimental design, and thus the culture and sample volume were

selected to ensure that the final volume (154 mL) was not lower than 50% of the initial volume (250 mL) to maintain culture volume as homogenous as possible along the experiment. This has now been more clearly specified within the M&M section.

Question 37:
L 587 do you mean peak area?
Answer: This has been corrected.

Question 38:
Fig.1 2x black circles in figure legend and figure. Maybe change colour or shape.
Answer: This has been modified. Thank you.

Question 39:
L1066 'clade; 3, known BGC…'
Answer: This has been corrected.

Question 40:
L1072-1073 How does the size of the circles indicate percentage of identity? Small circles, filled circles, hollow circles? Also, this data was not retrieved experimentally but only from genome analysis?
Answer: Legend has now been clarified, accordingly. Thank you.

Question 41:
L1068 'terpene'
Answer: This has been corrected.

Question 42:
L1077-1086 This is a discussion
Answer: This paragraph has now been transferred to the result & discussion section, accordingly.

Question 43:
L 1085 fell to be recovered? Use core or leader peptide instead of pro peptide
Answer: This sentence has now been reworded and clarified, accordingly: "These observations also highlight the global limitations of the automatic annotation of RiPP-BGC products – based on leader peptide and accessory enzyme recognition - as the high structural diversity generated by limited mutation events on pro-peptide fell to be established by LC-MSMS, while NRPS, NRPS/PKS and PKS products offer more automatic annotations according to the GNPS-based metabolomic pipeline."

Question 44:
L1093-1097 This is also a discussion, should go to the main text
Answer: This paragraph has now been transferred to the result & discussion section too.

Question 45:
Figure 4
The dendrogram on top and left is the same? One might be enough then.
Answer: The horizontal dendrogram has been removed, accordingly.

Question 46:

Figure 5
The figure is referred to once in the text and is not addressed. Hard to read and understand based on the description given. Is it important or can it go to the supplement?
Each circle represents 1 molecule? And size represents the quantity? What are the number in the circles?
Answer: Thank you for this remark. We assume to keep this figure in the core of the main text. Thus, we have now added a paragraph regarding Fig. 5 in the text, accordingly. The meaning of the node size and the number written inside have been now specified in the figure legend. Thank you.

Question 47:
Figure 7
"(%age)" [%] is sufficient
Answer: This has been corrected.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Biology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

General answer: we would like to thank all reviewers for the numerous constructive suggestions they made concerning our MS. Accordingly, we have now added a specific sentence in the acknowledgement paragraph.

Reviewers' comments:

Reviewer #1: secondary metabolite biosynthesis in cyanobacteria, metabolomics, genomics, Microcystis

(Remarks to the Author):

In this manuscript, the authors use genomic and metabolomic approaches to characterize the chemical diversity of secondary metabolites produced by the cyanobacteria genus Microcystis. They specifically analyze 65 new strains isolated from Europe (most are from France). The authors perform some comparative genomic analyses of these strains in addition to the other publicly available Microcystis genomes from NCBI. From these analyses the authors find that 57 of the 65 European strains can be classified into 11 genotypes and 13 metabotypes, which correspond with each other, and that similar genotypes may also be considered specific toxico-ecotypes based on toxicological screening. They also find that strains isolated from Europe capture the expansive diversity of this genus, which is consistent with other culture collections from around the world.

Overall, I think there are some very strong points in the manuscript, but there are also additional analyses and considerations that should be addressed. This work significantly expands representation of sequenced Microcystis genomes from Europe, and the conservation between genotype and metabotype has not been this clearly articulated in previous studies. However, additional details are needed for metabolomic annotations, additional analyses could be completed linking metabolomic and genomic data, and some conclusions from the data should be rephrased.

Question 1:
Throughout the text, the authors use the term “metabotype” when referring to specific chemical profiles/phenotypes. In my opinion, the term “chemotype” is more appropriate, as the authors focus specifically on secondary metabolites, and do not address the broader metabolome beyond these natural products.
Answer: Thank you for this remark. We have now replaced the term metabotype by chemotype in all the text, figures and supplementary materials, accordingly.

Question 2:
The authors should also state more clearly the level of confidence of the identification of each metabolite included in the results section and figures/tables. I recommend doing so by adopting the Schymanski scheme, which is largely used in MS for small and larger metabolites. It is important to be clear on the uncertainties of identification of all presented metabolites, and if in doubt, the confidence level can only be claimed to be level 3. For higher level confidence, manual annotations may need to be completed to assess completeness of predicted spectra (level 2) or reference standards should be used (level 1).
Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J., Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol 2014, 48, (4), 2097-2098
Answer: We thank the reviewer for this remark. This information has now been added in the M&M and in the supplementary table S2A, accordingly.

Question 3:
I would have liked to have seen the authors use their data to perform in silico linking between genomic and metabolomic data. There are several tools that are available such as NPlinker

and NPOmix that could be used, and the authors should have all the needed data to run this analysis. This would greatly increase the impact of this paper and provide greater understanding of how slight variation in BGC structure may influence congener type production – especially for poorly annotated/undescribed BGCs and chemical features.
Answer: We thank the reviewer for this remark as *in silico* screening of joint genomic and metabolomic dataset have been proposed to assist the discovery of novel BGC and related products based on co-occurrence detection within a large strain panel. However, *in silico* tools such as NPlinker and NPOmics (which by the way is already not supported anymore) have not demonstrated significant contribution. Indeed, preliminary description of novel BGC and potentially related metabolite family based on co-occurrence in certain strains present limited evidence, that would still require to be validated and supported by extended and in-depth direct structural and biosynthesis characterization. For this reason, we assume not to solely provide additional *in silico* screening in the present paper and consider that the already provided information (supplementary figure S11 and supplementary table S2), could already support valuably the discovery of novel metabolite family. However, according to the reviewer comments we have added to the discussion several sentences dedicated at the potential of the present dataset to further initiate the pointed characterization of novel metabolite families and related BGC: "Our dataset offers novel opportunities for discovering new BGCs associated within metabolites families. Future effort would now be specifically dedicated for charactering such novel molecules and potentially corresponding BGC. For example, the 2 stains from Chem06 (PMC 673 and 674) both present several uncharacterized RiPP BGCs (presenting faint similarities with PcpA, hglE and lanthipeptide class-v, respectively), together with a singular molecular cluster of 7 unknown metabolites ranging from around 1052-1092 Da, that aim at being further *de novo* characterized by direct structural analysis (supplementary figure S11; supplementary table S2a and 2G).".

Question 4:
Lines 192-195: The authors observe levels of sampled strain diversity varies considerably among lakes. This could be due to a multitude of factors, including time of year and nutrient availability. It would be helpful for the authors to include some of this information in table S1 (as is available) and include some details regarding the environmental differences of these lakes during the time of sampling – especially for those with very high and low diversity.
Answer: We thank the reviewer for this comment and understand the interest he(she) put on trying to explain or to better understand ecological influence on the strain diversity. We are also concerned by the fact that a single sampling strategy is far from being enough to characterize the complexity of ecological influence of bloom genetic diversity and could constitute a quite finalistic deliverable. For this reason, and because we do not possess sufficiently informative environmental dataset concerning the sampled lake ecological features, we assume not to further investigate this question regarding the present dataset. However, we have now added basic information concerning the sampling origin of the strains within supplementary table S1.

Question 5:
Lines 214-215/Figure 1 and Figure S6: I actually think that Figure S6 should replace Figure 1 in the manuscript. Figure S6 contains all the information provided in S6, but S6 also demonstrates the high conservation of genotypes when comparing the core and pangenome. This is the first time I've seen this comparison in literature, and I think it's a point that should be more prominently displayed in the main text.

Answer: We thank the reviewer for this comment and share his(her) enthusiasm on the interest of considering the high conservation of genome relationship regarding core and pan genomes, and have now inserted fig S6 as a regular figure.

Question 6:
Lines 274-283/Figure 6: While almost every genotype corresponds nicely to a conserved metabotype, this condition is not met for the genotype i4 which splits into metabotype 1 and 13. I find this extremely interesting and would like to the authors to include more commentary on this finding. Is this a result of different congener production, but largely the same relative abundances of reported metabolites? Are there variations in the composition of BGCs reported – either at the gene content or gene structure level (presence of point mutations, different domain structure, etc)? Having some insight/commentary as why strains that share the same genotype, and are cultured in the identical conditions, but produce different metabotypes would be very valuable.
Answer: Thank you for this remark. This aspect has now been directly discussed, accordingly: "Although, strains from chem01 and chem13 belong to the same genotype and present globally similar metabolite composition (comprising microcystins, aeruginosins, cyanopeptolins and microcyclamides) they also present noticeable molecular distinction characterized by variations in variant contents, that appears congruent with the respective genotypic position of the strains".

Question 7:
Lines 309-313: The authors should clarify that BGCs are not constitutively expressed in natural microbial communities. Therefore, the presence of diverse secondary metabolites observed in culture do not necessarily reflect metabolite pools observed in nature. I agree that these traits ecotoxicological traits are conserved by genotype in culture, but in a natural system, there may be greater variability of what is produced by different strains.
Answer: We thank the rewiever for this judicious remark. We have now added to the discussion a sentence highlighting this limitation of experimental approach that could not fully represent to what could happen *in naturae* as BGC expression slightly vary regarding local environmental factors: "Also, several key abiotic of biotic factors (*e.g.* nutrient availability, temperature, light intensity, or allelopathy and predation) may influence BGC expression and slightly influence the composition of the *Microcystis* metabolite cocktails, potentially modulating the subsequent ecotoxicological threats that may be observed in both experimental and environmental context".

Question 8:
Lines 335-338: I do not understand this sentence. The authors cite 2 papers that look at the specific roles of certain variables (N, P) and the production of certain metabolites – and how it is very challenging to identify specific abiotic variable that drive metabolite production. I'm not sure how potential health repercussions could impact studies attempting to make this link.
Answer: This sentence has been reworded in order to better fit with the flow of the discussion and our general purpose, accordingly: "Interestingly, Jackrel and co-wokers[10] and latter Kuijpers and co-wokers[71] have correlated global genome reductions and expansions in gene supporting specific metabolic capabilities (*i.e.* N- and P-related genes) in relation with specific ecological features of their originating lakes (low- or high-nutrient lakes).".

Question 9:

Lines 341-373 and 393-405: While well these sections provide nice summaries of current literature, I found it to be distracting in the results/discussion section. These sections were also not contextualized with the current findings presented in this paper, and it diverts my attention from the novel results being presented. I would suggest either shortening these sections, moving some of this text to the introduction, or finding ways to better integrate the new results presented in this paper with this text.
Answer: These paragraphs have now been significantly shortened, accordingly.

Question 10:
Lines 432-433: These results are in agreement with previous studies that the authors should consider citing including:
a. https://doi.org/10.3389/fmicb.2019.00791
b. https://doi.org/10.1111/1462-2920.12904
c. https://doi.org/10.1128/msystems.00334-24
Answer: Citations have been added, as suggested. Also, discussion sentences have been added, accordingly: " Interestingly, Yancey and co-workers[48] recently developed a similar approach to describe the chemical diversity of 21 xenic *Microcystis* strains isolated form Great Lake Erie (Ohio, USA). They observed that several genotypes were associated with the production of distinct metabolite cocktails, suggesting and highlighting the limited understanding of the determinism of chemical repertoire of *Microcystis* ".

Question 11:
Line 495-496: Could the authors please specify why 90% identity was set as the threshold? This number is lower than traditional cutoffs used (95% or higher).
Answer: Indeed, the identity threshold may depend on the taxa considered. In the present case, the investigation of a largely diversified *Microcystis* clade justify the use of a slightly lower threshold. However, although the use of a higher threshold decreases the number of orthologous gene considered it does not presently modify the global phylogenetic picture observed. With 90%-threshold

Question 12:
Line 512: Could the authors please specify which R package was used in the text.
Answer: This information has been specified.

Question 13:
Line 535-538: I think this should be stated earlier in the methods, perhaps in the Strain collection and study design section. I think it's important to note that not all the strains are axenic before getting into the details of genomic/metabolomic analysis, and should also be added to Table S1 for reference.
Answer: As the reference of the axenicity of the culture co concern the biomass production and not collection condition of the respective strains, we assume to keep this description in this specific position of the M&M. However, the axenicity statue of the strains in PMC and PCC collection have been added in table S1.

Question 14:
Figure 3: I would argue that BGC annotations with less than 50% identity to known clusters should not be reported. There is too much divergence of these sequences at this point, and the genes present may be involved in a different (but similar) biosynthesis pathway. If the authors want to report the presence of these BGCs with greater confidence they could perform read

mapping in the event of poor assembly, or provide other rationale for including these annotations.

Answer: We thank the reviewer for this comment. We believe he(she) is judicious to notice that similar score between suggest that these BGC might produce similar molecules, when higher score may show more confidence in the proposed metabolite family. However, the consideration of a universal threshold for AntiSmash annotation is a complex concern that require global vista of the scoring principle (based on structural similarity to a reference sequence) regarding the respective architecture of the distinct BGC families. Indeed, when NRPS BGC of the same family present global high sequence similarity, when presenting little modifications in RIPP-BGC sequences, can strongly influence the structure of the product and its annotation. Thus, it would maybe appears adapted to consider distinct significance scores for the distinct BGC families. Although, lower score is lesser relevant in term of genuine annotation, it remains informative in term of sequence similarity that present homogenous pattern between strains of the same genotypes. For these reasons, we would rather keep this information accessible for the reader.

Question 15:

Figure 3 & 4: This figure captions are lengthy and contain text that should be moved to the results section.

Answer: Several sentences of Fig. 3 and 4 legends have now been moved to the main text of the MS, accordingly: "Overall, these results illustrate the large variation observed between the different European *Microcystis* strains (ranging from 3.8 to 5.9 Mb), together with an overall homogeneity within the distinct genetic clades. The accessory metabolites of some genotypes are highly similar despite the diversity of the strain origins (*e.g.* genotype i5 containing isolated strains from north and south of France), while chemotypes within certain genotypes appeared much more diversified than between certain others (*e.g.* within I compared to between D and T; supplementary Fig. S7). It is also worth noticing that although the two neighboring strains belonging to the genospecies P (namely PMC 566.08 and PMC 568.08) did not constitute a genuine genotype (their ANI score being above 97%, but below 99% thresholds), the metabolomes of these two strains were grouped together within a single chemotype (see figure 4)."

"We observed that strain clusters discriminated at the genomic level considering either core- and pan-genome or BGC contents strictly correspond to characteristic chemotypes producing well determined metabolite categories. It is worth noticing that this chemotype distinction was remarkably identical to those retrieved from the merged data matrix obtained by combining both analyses performed on positive and negative MS mode (Supplementary Fig. S6) and that the chemotype belonging of strains was largely stable across successive cultures (Supplementary Fig. S7-8). Only minor variations in terms of molecular variants and/or relative metabolite quantities can be monitored regarding the biomass of strains cultured under laboratory conditions (Supplementary Fig. S8-10)."

Referee #2: Mycrocystis, genomics

(Remarks to the Author):

This study pairs genomics with metabolomics to understand the genomic diversity of European (primarily French) strains of Microcystis. Based on a strict threshold, the authors

propose a genotype classification based on >99% ANI as the most accurate operational taxonomic unit, rather than genospecies at >97% ANI. This genetic analysis was paired with characterization of corresponding metabotypes.

Major Items:
Overall, the study is an incremental advancement in our understanding of the metabolic capabilities of the genus Microcystis. While the study is worthy of publication, the potential impact of the findings is minimal.

Question 1:
The authors extrapolate the finds to all European strains of Microcystis, but the locations of isolation are primarily located in one country. Isolation sites should encompass lakes in numerous locations or the geographic extrapolation should be reduced.
Answer: The geographic extrapolation has now been reduced and further considered as "Western European strains", "originating from France and surrounding countries", in all the manuscript, accordingly.
Line 153-156: "Our sequencing efforts have revealed the extensive diversity among the investigation of 65 strains originating from France and surrounding countries, which are largely distributed across the phylogenomic tree (Fig. 1)."
Line 177-178: "*Microcystis* diversity from France and surrounding countries and local genotype richness"
Line 169-171: "To further refine the phylogenetic delineation of the 65 Western European strains, we grouped those sharing an average identity of >99% together (Supplementary Figure S5A-B). "
Line 179-181: "Our investigation of various Western European *Microcystis* strains (mostly from France) shows that most strains belong to genospecies already observed in other geographic areas."
Line 250-253: "Overall, these results illustrate the large variation observed between the different European *Microcystis* strains from France and surrounding countries (ranging from 3.8 to 5.9 Mb), together with an overall homogeneity within the distinct genetic clades."
Line 279-283: "Our extended LC-MS-based metabolomic analyses clearly indicate that Western European *Microcystis* strains (*i.e.*, from France and surrounding countries) belonging to different genetic clades produce specific and distinct metabolite corteges, representing characteristic chemotypes even when cultured under standard laboratory conditions (Fig. 5)."
Line 334-336: "Our analysis of the diversity of accessory metabolites of Western European *Microcystis* strains corroborates the possibility for environmental filtering and ecologically relevant genomic variation between strains."
Line 342-344: "We further investigated 12 selected *Microcystis* strains belonging to six distinct genotypes (d2, i2, i3, i5, and k2 from Western European and a pair of strains (PCC 9804 from Australia and NIES 98 from Japan)"
Line 577-578: "The 65 Western European *Microcystis* strains (63 from France, 1 from Scotland, 1 from Netherland), …"
Legends of figures: "Western European"

Question 2:
A discussion of the implications of the interactions within the bloom holobiont with other microbes (bacteria, fungi, viruses) and the potential for evolutionary and ecological consequences reflected in Microcystis genome content and metabolite profiles is missing.

Answer: We thank the reviewer for this comment. A specific sentence on chemical interaction within the microbial community have been added, accordingly: "Indeed, several cyanobacterial metabolites have been proposed to play key ecological functions (*e.g.* allelopathy, anti-grazing or antimicrobial) within the microbial communities and may contribute to the blooming capabilities of these micro-organisms."

Question 3:
L376- The contrast to Aspergillus here is interesting, but a new comparison to a eukaryotic fungus when all previous context and comparisons have been to other cyanobacteria is discordant. Perhaps a longer discussion here is warranted, especially since we now know that fungi are ubiquitous during blooms and may be contributing to the holobiont metabolome profile.
Answer: This paragraph has been reworded and the reference to Aspergillus better integrated to the rest of the discussion, accordingly: "The patterns of diversity that we observe here suggest that each distinct *Microcystis* clade generates and maintains biochemical diversity in particular ways. In contrast to certain other micro-organisms, where the chemical diversity increases linearly with taxonomic repertoire and genome size,[88] freshwater cyanobacteria demonstrate noteworthy examples of convergence in their accessory metabolite structural diversity, that do not appear systematically related to the genome size of the different taxa."

Specific items:
Question 4:
L40- "corteges" is likely not the correct word to use here
Answer: This word has been replaced by "cocktails"

Question 5:
L71- dominate – remove the "s"
Answer: This has been corrected. Thank you.

Question 6:
L91- Mo or Mb?
Answer: Mb, this has been corrected. Thank you.

Question 7:
L184- should be nine not neine?
Answer: This has been corrected. Thank you.

Referee #3 & #4 (co-reviewers): cyanobacteria metabolism

(Remarks to the Author):

The manuscript by Marie and colleagues provides a comparative genomic, metabolomic and ecotoxicologic study of the ecologically relevant cyanobacterium Microcystis. It provides novel insights into the extensive diversity of Microcystis based on whole genome sequencing data sourced from various locations across western Europe. The manuscript challenges postulated concepts such as that the genetic distance within Microcystis strains increases with geographical distance. It identifies various Microcystis genospecies encompassing multiple genotypes. A main finding of this study is that genotypes within a genospecies can show far greater metabolic diversity when compared to the metabolic landscape of genotypes of more

distantly related genospecies. The data provided further enhances our understanding of environmental adaptations and evolution of bloom forming, toxic freshwater cyanobacteria and allows a glimpse at the rich and largely undiscovered chemical diversity of Microcystis. In general, the manuscript is well structured but legibility partly suffers from typos and misuse of terms. Data interpretation, conclusion and used methodology seem valid. A key problem is a lack of detail in the description of the experimental part that does not allow reproduction (detailed below). Some suggestions and corrections:

Question 1:
L 71 enables… to dominate
Answer: This has been corrected. Thank you.

Question 2:
L 84-87 revise sentence
Answer: This sentence has been revised: "Since then, fully sequenced genomes of *Microcystis* strains have been progressively providing unprecedented taxonomic resolution and have described a distinct phylogenetic picture compare to those previously identified through morphology-based taxonomy."

Question 3:
L 91 Do you mean Mb? If you mean Mb, the reference that is cited does not support the stated claim. In the cited research the authors claim Microcystis genome sizes range between 3.2 And 4.9 Mb and not 3.5 to 6 Mb.
Answer: Mb, this has been corrected. The place of this citation has been moved, accordingly. Thank you.

Question 4:
L 97 …Microcystis exhibit complex genomes
Answer: This has been corrected. Thank you.

Question 5:
L 99 …arsenals…
Answer: This has been corrected.

Question 6:
L 103 Revise language, "some of these several"
Answer: This has been corrected.

Question 7:
L 116 Environmental sampling "has" shown
Answer: This has been corrected.

Question 8:
L 121 Features
Answer: This has been corrected.

Question 9:
L 123 omit "strain culture"
Answer: This has been modified.

Question 10:
L 128 remain
Answer: This has been corrected.

Question 11:
L 130 analysis
Answer: This has been corrected.

Question 12:
L 139 “This original approach clearly allows us to explore the functional aspect within the Microcystis tree of life.” functional aspect of what?
Answer: This has now been specified: “ecological functional aspect of this diversity”. Thank you.

Question 13:
L 157 and following: the interchangeable terms genomospecies and genospecies are not used consistently
Answer: As geno-species (genotypes presenting similar genetic markers, such as 16S sequence score) and genomo-species (genotypes presenting a global genome similarity, such as ANI score) have near but distinct meaning, we assume their proper use in the text.

Question 14:
L 184 nine
Answer: This has been corrected.

Question 15:
L 198 use “.” for decimals
Answer: This has been corrected.

Question 16:
L 229-230 “Some RiPP-BGCs that present intra-genotype variations (microcylamide in i5, …., Fig. 3).” This sentence is incomplete
Answer: This sentence has been reworded, accordingly: “While the BGC presence for NRPS/PKS and/or their corresponding metabolite detection are highly conserved within genotypes, but not into the genomo-species, some RiPP-BGCs present intra-genotype variations (microcylamide in i5, aeruginosamide in k2 and piricyclamide and microviridin in t1, Fig. 3).”

Question 17:
L238-239 “while subgroups within a genotype appeared much more diversified in terms of metabolites (e.g. Genotype I from i1 to i5)”. The “subgroups” are the genotypes and what is referred to as genotype here is the genospecies. While i1-i5 are genotypes, I represents the genospecies.
Answer: We understand that the formulation of this sentence maybe confusing and we have modified it, accordingly: “The accessory metabolites of some genotypes are highly similar despite the diversity of the strain origins (*e.g.* genotype i5 containing isolated strains from north and south of France), while metabotypes within certain genotypes appeared much more diversified than between certain others (*e.g.* within I compared to between D and T; supplementary Fig. S7).”.

Question 18:
L 247 full stop missing after reference 18
Answer: This has been corrected.

L 264 “relative quantities that can be monitored for strains biomass cultured under lab conditions” ? revise sentence
Answer: These sentences have been reworded: “We observed that strain clusters discriminated at the genomic level considering either core- and pan-genome or BGC contents strictly correspond to characteristic metabotypes producing well determined metabolite categories. Indeed, only minor variations in terms of molecular variants and/or relative metabolite quantities can be monitored regarding the biomass of strains cultured under laboratory conditions (Supplementary Fig. S8-10).”

Question 19:
L 303 “… to genotypes i2 and k2…” genotype i2 not presented in fig. 7, should be i1?
Page 12 The discussion focusses solely on diversity on the genomic level and metabolome when cultivated under lab conditions. Might be good to discuss the possibility of regulation of expression of BGCs, as this will allow for immediate environmental plasticity and responses on a metabolic level. Possibly to occupy and thrive in ecological niches. Differential regulation of BGCs could also depict another level of diversity within the same genotype.
Answer: Thank you. This has been corrected on Fig. 7. Also, we have added to the discussion a specific sentence discussing environmental metabolome regulations and potential ecotoxicological consequences.

Question 20:
L 354 remain
Answer: This has been corrected.

Question 21:
L 360 remove “)”
Answer: This has been corrected.

Question 22:
L 367 cyanobacteria-specific
Answer: This has been corrected.

Question 23:
L 389 novel novelties
Answer: This has been corrected.

Question 24:
L 422 patterns
Answer: This has been corrected.

Question 25:
L 432 a significant portion
Answer: This has been corrected.

Question 26:

L 470-471 rephrase, cultures are not performed, cultivation volume is missing, rpm and shaker geometry is missing

Answer: This sentence has been reworded: "To sequence the new *Microcystis* genomes from PMC, cultures were grown in Z8 medium[104] at 25°C in 250-mL Erlenmeyer flasks. The cultures were exposed to a photon flux density of 20 µmol.m^{-2}.s^{-1} and maintained on a 16:8h light:dark cycle."

Question 27:
L 472 revise unit "µmol.m-2.s-1"

Answer: This has been corrected.

Question 28:
L 473 how long was the centrifugation step?

Answer: This information has been added.

Question 29:
L 475) missing

Answer: This has been corrected.

Question 30:
L 480 inconsistent citation style

Answer: This has been corrected.

Question 31:
L 509 space missing

Answer: This has been corrected.

Question 32:
L 518 what is the reasoning for selecting a cutoff of 5.5 kb? Please elaborate

Answer: Answer: A citation have been added.

Question 33:
L 529 Rephrase, BGCs are not annotated as terpenes,
Why were BGCs that give rise to terpenes excluded?

Answer: This sentence has been corrected and specified.

Question 34:
L 535 and following. Lack of information necessary for reproducibility

Answer: This M&M section have been improved to ensure reproducibility, accordingly.

Question 35:
L 548 triplicates

Answer: This has been corrected.

Question 36:

L 546 and following. Lack of information. You do not state the cultivation volume. If you remove 12 ml twice a week from the culture and you cultivate for 28 days (4 weeks) you end up removing 96 ml of culture if you do not replenish it. Thus, you not only monitor the variations of intracellular metabolite contents as a function of time but the drastic differences in culture volume will have wide ranging effects.

Answer: We thank the reviewer for this interesting remark. Indeed, we placed a careful attention at the experimental design, and thus the culture and sample volume were selected to ensure that the final volume (154 mL) was not lower than 50% of the initial volume (250 mL) to maintain culture volume as homogenous as possible along the experiment. This has now been more clearly specified within the M&M section: "Simultaneously, a specific experiment was conducted over a single 28-day period to monitor variations in intracellular metabolite content. Experimental conditions (250 mL of BG-11 medium) were carried out in triplicates, with a 16h:8h photoperiod (12 µmol.m^{-2}.s^{-1}; 25°C) and constant homogenization maintained by magnetic agitators. Two weekly samples of 12 mL were taken throughout the 28 days (final volume above 150 mL). Each sample was divided into *i*) the extraction of extra-cellular metabolites (10 mL) and *ii*) the biomass monitoring using optical density (2 mL) at 750 nm performed by *in vivo* spectrophotometry (Cary 500; Varian, U.S.A.)."

Question 37:
L 587 do you mean peak area?
Answer: This has been corrected.

Question 38:
Fig.1 2x black circles in figure legend and figure. Maybe change colour or shape.
Answer: This has been modified. Thank you.

Question 39:
L1066 'clade; 3, known BGC…'
Answer: This has been corrected.

Question 40:
L1072-1073 How does the size of the circles indicate percentage of identity? Small circles, filled circles, hollow circles? Also, this data was not retrieved experimentally but only from genome analysis?
Answer: Legend has now been clarified, accordingly. Thank you.

Question 41:
L1068 'terpene'
Answer: This has been corrected.

Question 42:
L1077-1086 This is a discussion
Answer: This paragraph has now been transferred to the result & discussion section, accordingly.

Question 43:
L 1085 fell to be recovered? Use core or leader peptide instead of pro peptide
Answer: This sentence has now been reworded and clarified, accordingly: "These observations also highlight the global limitations of the automatic annotation of RiPP-BGC products – based on leader peptide and accessory enzyme recognition - as the high structural diversity generated by limited mutation events on pro-peptide fell to be established by LC-MSMS, while NRPS, NRPS/PKS and PKS products offer more automatic annotations according to the GNPS-based metabolomic pipeline."

Question 44:
L1093-1097 This is also a discussion, should go to the main text
Answer: This paragraph has now been transferred to the result & discussion section too.

Question 45:
Figure 4
The dendrogram on top and left is the same? One might be enough then.
Answer: The horizontal dendrogram has been removed, accordingly.

Question 46:
Figure 5
The figure is referred to once in the text and is not addressed. Hard to read and understand based on the description given. Is it important or can it go to the supplement?
Each circle represents 1 molecule? And size represents the quantity? What are the number in the circles?
Answer: Thank you for this remark. We assume to keep this figure in the core of the main text. Thus, we have now added a paragraph regarding Fig. 5 in the text, accordingly. The meaning of the node size and the number written inside have been now specified in the figure legend. Thank you.

Question 47:
Figure 7
"(%age)" [%] is sufficient
Answer: This has been corrected.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Biology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

General answer: we would like to thank all reviewers for the numerous constructive suggestions they made concerning our MS. Accordingly, we have now added a specific sentence in the acknowledgement paragraph.

REVIEWERS' COMMENTS:

Dear Dr Marie,

Your manuscript entitled "More than just disorder - metabolite diversity of Microcystis strains shows tight correspondence to genotype and may contribute to ecotype specificities" has now been seen again by our referees, and I apologize for the delay in the processing of your manuscript. In light of the referees' advice, I am delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Biology.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. Please thoroughly address the remaining concerns of Reviewers #1 and #3/#4 with the exception of Reviewer #1's request to add the in silico linking analysis (related to their Question 3), as we have decided we will not require you to add such analysis to the manuscript.

Please include a marked-up version of the manuscript showing all changes made in response to these reviewers' comments as well as a point-by-point response to these referees' comments.
We hope to hear from you within two weeks. If you expect the process to take longer than one month, please let us know.

Congratulations on an excellent paper!

Best regards,

David Favero, PhD
Senior Editor
Communications Biology

Reviewer #1 (Remarks to the Author):

The authors have carefully and thoughtfully addressed most of my comments and suggestions. It appears that they have also taken careful time to review and revise the manuscript based on the suggestions of other reviewers as well. Most of my points have been sufficiently addressed, but there are still some questions or suggestions that need more careful review and consideration. While it may seem pedantic, I encourage the authors to revisit some of the points I previously raised, which were either insufficiently addressed or dismissed without adequate justification. Given the visibility and impact of this journal, it is essential that the manuscript reflects the highest standards of rigor and clarity, especially in its interpretation, to ensure it serves as a reliable resource for a broad and diverse readership. Below are my follow up comments for those points of interest.

Question 3:
I would have liked to have seen the authors use their data to perform in silico linking between genomic and metabolomic data. There are several tools that are available such as NPlinker and NPOmix that could be used, and the authors should have all the needed data to run this analysis. This would greatly increase the impact of this paper and provide greater understanding of how slight variation in BGC structure may influence congener type production – especially for poorly annotated/undescribed BGCs and chemical features.

Answer: We thank the reviewer for this remark as in silico screening of joint genomic and metabolomic dataset have been proposed to assist the discovery of novel BGC and related products based on co-occurrence detection within a large strain panel. However, in silico tools such as NPlinker and NPOmics (which by the way is already not supported anymore) have not demonstrated significant contribution. Indeed, preliminary description of novel BGC and potentially related metabolite family based on co-occurrence in certain strains present limited evidence, that would still require to be validated and supported by extended and in-depth direct structural and biosynthesis characterization. For this reason, we assume not to solely provide additional in silico screening in the present paper and consider that the already provided information (supplementary figure S11 and supplementary table S2), could already support valuably the discovery of novel metabolite family. However, according to the reviewer comments we have added to the discussion several sentences dedicated at the potential of the present dataset to further initiate the pointed characterization of novel metabolite families and related BGC: "Our dataset offers novel opportunities for discovering new BGCs associated within metabolites families. Future effort would now be specifically dedicated for charactering such novel molecules and potentially corresponding BGC. For example, the 2 stains from Chem06 (PMC 673 and 674) both present several uncharacterized RiPP BGCs (presenting faint similarities with PcpA, hglE and lanthipeptide class-v, respectively), together with a singular molecular cluster of 7 unknown metabolites ranging from around 1052-1092 Da, that aim at being further de novo characterized by direct structural analysis (supplementary figure S11; supplementary table S2a and 2G).".

Response: While I can understand the perspective of the authors, I believe that additional analyses would enhance rigor, thereby making it a more suitable manuscript for publication. The authors have generated a very rich dataset, and it would not require too much effort to perform in silico linking, or co-occurance/similarity networking with other tools. The field is currently flooded with publications that list strain BGCs and produced metabolites, without making real efforts to try to link these data. While I agree that these tools are not sufficient to directly link an orphan BGC and metabolite, they provide valuable clues to other researchers and support hypothesis generation for future metabolite isolation and characterization. There are numerous publications to support these types approaches including tools such as NPlinker (previously mentioned), NPClassScore (DOI https://doi.org/10.1186/s40168-022-01444-3), and additional MetCalf scoring methods. Other publications have also validated the utility of these tools (DOI: https://doi.org/10.1186/s40168-022-01444-3, DOI : https://doi.org/10.3390/md19020103). It is becoming more common to at least attempt some hypothetical linking (if experimental validation is not completed). This would greatly improve the impact of this paper, and put it more in line in the direction the field is moving.

Answer: According to our previous answer and to the editor appreciation, we assume not to add supplementary *in silico* analysis to the MS.

Question 4:

Lines 192-195: The authors observe levels of sampled strain diversity varies considerably among lakes. This could be due to a multitude of factors, including time of year and nutrient availability. It would be helpful for the authors to include some of this information in table S1 (as is available) and include some details regarding the environmental differences of these lakes during the time of sampling – especially for those with very high and low diversity.

Answer: We thank the reviewer for this comment and understand the interest he(she) put on trying to explain or to better understand ecological influence on the strain diversity. We are also concerned by the fact that a single sampling strategy is far from being enough to characterize the complexity of ecological influence of bloom genetic diversity and could

constitute a quite finalistic deliverable. For this reason, and because we do not possess sufficiently informative environmental dataset concerning the sampled lake ecological features, we assume not to further investigate this question regarding the present dataset. However, we have now added basic information concerning the sampling origin of the strains within supplementary table S1.

Response: Thank you for including additional information in Table S1. My original comment was directed toward the new European strains that were described for the first time in the manuscript. I'm not at all suggesting that the authors should try to make assumptions of ecological complexity based on the isolation of single strains (which comes with it's own set of biases). However, I still think it is important to provide basic meta data where available, especially for the new strains discussed. Ideally, I would like to see isolation year, collection date (or month if exact date is unknown), and a general trophic status for the sampled lake (oligotrophic, eutrophic, hypereutrophic, etc.). These parameters should have been collected during the time of sampling and are important details that should be included in the manuscript.

Answer: When available, the isolation year, sampling date and trophic status of the lake have now been added to supplementary table S1, accordingly.

Additionally, the text in Figure 1 is pretty small for each individual strain and is hard to read. Would it be possible to add the genotype and chemotype to this table as well for easy reference ? I found it challenging to read even in the .tif file.

Answer: We are sorry that the figure definition on the pdf file does not allow a full view of each individual strain name. We have now provided fig. 1 in PDF format with an increased resolution and hope that publication format would make its readable. However, in order to make this information easy to find, we have also added this genotype/chemotype information within table S1.

Question 6:

Lines 274-283/Figure 6: While almost every genotype corresponds nicely to a conserved metabotype, this condition is not met for the genotype i4 which splits into metabotype 1 and 13. I find this extremely interesting and would like to the authors to include more commentary on this finding. Is this a result of different congener production, but largely the same relative abundances of reported metabolites? Are there variations in the composition of BGCs reported – either at the gene content or gene structure level (presence of point mutations, different domain structure, etc)? Having some insight/commentary as why strains that share the same genotype, and are cultured in the identical conditions, but produce different metabotypes would be very valuable.

Answer: Thank you for this remark. This aspect has now been directly discussed, accordingly: "Although, strains from chem01 and chem13 belong to the same genotype and present globally similar metabolite composition (comprising microcystins, aeruginosins, cyanopeptolins and microcyclamides) they also present noticeable molecular distinction characterized by variations in variant contents, that appears congruent with the respective genotypic position of the strains".

Response: Thank you for addressing my initial comment. Now with the additional information reported in Table S1, it appears the genotype i4 is represented by strains of Microcystis that are both xenic and axenic. The presence of other microbial taxa can influence secondary metabolite production in terms variants produced, and regulation of biosynthetic pathways. The authors should include some additional text that discusses that axenic status can also influence differences in chemotypes of strains that share the same genotype.

Answer: We thank the reviewer for this comment. Accordingly, we have now added a paragraph on the potential influence of secondary metabolite production regarding xenic or axenic statues of the strains: "Additionally, the xenic or axenic statue of each individual strain seems presently not to blur the overall chemotype signal of the cultures that remains linked to their respective *Microcystis* genotypes. Regarding the large size difference between heterotroph bacteria of xenic cultures and the cyanobacterial cells (above 1/500-1/5000 in terms of cellular biovolume) and the high secondary production rate of cyanobacteria cultures samples under exponential growth phase (also maximizing the heterotroph/cyanobacteria ratio), the potential xenic influence on *Microcystis* strain chemotype remained presently limited."

Question 7:

Lines 309-313: The authors should clarify that BGCs are not constitutively expressed in natural microbial communities. Therefore, the presence of diverse secondary metabolites observed in culture do not necessarily reflect metabolite pools observed in nature. I agree that these traits ecotoxicological traits are conserved by genotype in culture, but in a natural system, there may be greater variability of what is produced by different strains.

Answer: We thank the rewiever for this judicious remark. We have now added to the discussion a sentence highlighting this limitation of experimental approach that could not fully represent to what could happen in naturae as BGC expression slightly vary regarding local environmental factors: "Also, several key abiotic of biotic factors (e.g. nutrient availability, temperature, light intensity, or allelopathy and predation) may influence BGC expression and slightly influence the composition of the Microcystis metabolite cocktails, potentially modulating the subsequent ecotoxicological threats that may be observed in both experimental and environmental context".

Response: Thank you once again for addressing my comment. However, I am concerned that the use of the word 'slightly' greatly downplays the significant impact that abiotic and biotic factors have on secondary metabolite production, especially in situ. This is well documented in the literature, mostly for the production of microcystin, but for other secondary metabolites as well. Work that immediately comes to mind that highlight this point includes DOI : 10.1007/s10452-016-9571-6.

Answer: Thank you for this remark. We have now modified this sentence accordingly: "Also, several key abiotic of biotic factors (*e.g.* nutrient availability, temperature, light intensity, or allelopathy and predation) may influence BGC expression and influence the composition of the *Microcystis* metabolite cocktails in terms of relative and absolute quantities or variant and chemical family ratio, potentially modulating the subsequent ecotoxicological threats that may be observed in both experimental and environmental context."

Question 14:

Figure 3: I would argue that BGC annotations with less than 50% identity to known clusters should not be reported. There is too much divergence of these sequences at this point, and the genes present may be involved in a different (but similar) biosynthesis pathway. If the authors want to report the presence of these BGCs with greater confidence they could perform read mapping in the event of poor assembly, or provide other rationale for including these annotations.

Answer: We thank the reviewer for this comment. We believe he(she) is judicious to notice that similar score between suggest that these BGC might produce similar molecules, when higher score may show more confidence in the proposed metabolite family. However, the

consideration of a universal threshold for AntiSmash annotation is a complex concern that require global vista of the scoring principle (based on structural similarity to a reference sequence) regarding the respective architecture of the distinct BGC families. Indeed, when NRPS BGC of the same family present global high sequence similarity, when presenting little modifications in RIPP-BGC sequences, can strongly influence the structure of the product and its annotation. Thus, it would maybe appears adapted to consider distinct significance scores for the distinct BGC families. Although, lower score is lesser relevant in term of genuine annotation, it remains informative in term of sequence similarity that present homogenous pattern between strains of the same genotypes. For these reasons, we would rather keep this information accessible for the reader.

Response: I stand firmly with my initial comment regarding reporting BGC annotations with antismash below a 50% threshold. This is misleading, and has caused amplification in the literature of mis-reporting of biosynthetic gene clusters being present in strains or populations when they are not. Several biosynthetic genes such as those encoding NRPSs and PKSs have highly conserved regions that can sporadically, and incorrectly, map and hit to known BGCs when using antismash. Since the authors have high quality, assembled genomes, it is highly unlikely that these are artifacts of fragmented assembly. If the authors insist on reporting these data it should be done in a table or other figure where the % identity and alignment length are reported with the gene cluster. Alternatively, the authors could rename these clusters generically to reflect biosynthesis pathways for example RiPP_1, RiPP_2, etc.

Answer: We are also concerned about the significance of our work to provide reliable dataset and avoid misleading the reader in anyway. However, we still believe that the fact that for example regarding the genotype t1/chemotype 05, the relatively low Antismash annotation score observed for aeruginosin or nostophycin BGC (score between 28 and 18%, respectively, depending on the completeness of the assembled sequence; supp. table S2P) might correspond to the several spumigin variants - clearly observed in all corresponding strains (supp. fig. S12) - constitutes a reliable information that aim at been provided by our MS. In this case, the closest reference sequence of spumigin BGC (lacking in MiBig database) remain those of aeruginosin BGC leading to relatively low but still informative scoring.

According to the reviewer suggestion, we have now added a supplementary table S2P reporting the % identity score provided by AntiSmash, together with the parallel observation of their potential product by LC-MS/MS for all reported BGCs.

Reviewer #3 (Remarks to the Author):

The manuscript by Marie et al was improved significantly, still some critical information is missing and several points made in the last round of review were not addressed.

Remove "cyanobacteria" from line 396

Answer: Thank you, this has been corrected.

In the material and method section L525 the information for how much cultivation volume was used is still missing, thus no reproduction possible.

Answer: Culture volume (150 mL) has been specified, accordingly.

Line 606, also here information about vessel volume is missing.

Answer: The vessel volume (500-mL Erlenmeyers) has been specified, accordingly.

Line 535 remove “.”
Answer: Thank you, this has been corrected.

Besides these minor points i suggest acceptance of the manuscript.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Biology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts;
Answer: Thank you.