Peer-review

Dingiloglu, Defne. 2025. “The Untapped Potential of the Endophytic Actinomyces, Streptomyces Scabiei, in Producing Antibiotic and Antitumor Secondary Metabolites.” *Journal of High School Science* 9 (1): 18–30.

I have some comments that will need to be addressed.

1.Literature suggests approx 80 BGC in certain species of Streptomyces; see: https://doi.org/10.1038/s41598-020-58904-9, therefore, your number of BGC (approx 15) is significantly lesser. Why then did you choose to concentrate on this species/strain when there are species that produce many more BGCs ? Maybe a re-write of your hypothesis is in order to make it more specific? “The hypothesis of this study is that the plant pathogenic bacterium S. scabiei will exhibit a greater BGC diversity than previously studied Streptomyces species, and potentially synthesize new antibiotics.” You can also reference https://doi.org/10.1371/journal.pone.0289280 and https://doi.org/10.3389/fmicb.2024.1408479 (and references therein) which state that endophytes have a propensity for a greater # of gene cluster diversity.

2.Need bootstrap values for the phylogenetic tree in Figure 4. Generally, bootstrap values below 70 are not acceptable. Also, if 002155 is dis-similar from the other 12 strains as well (Figure 4), why does it not show a larger number of BGC in Figure 3b ? (it should either show very low or very high BGC counts.).

3.From point 2, since 002155 is dissimilar but shows similar BGC counts to the other strains (except 000738), this implies that absolute BGC counts do not necessarily mean that an exclusive study of 000738 will yield the most # of secondary metabolites (because none of the other strains may be subsets of 000738). For example, studying the other 11 strains simultaneously may yield many more compounds (approximately from 3x30=30 BGC) than studying 000738 alone (because none of them may be asumed to be subsets of 000738). This is also illustrated in https://doi.org/10.1038/s41598-020-58904-9. Hence, the premise that because one strain exhibits the most # of BGC; therefore it will yield the most # of secondary metabolites (antibiotics or antitumor agents) at the expense of the rest is false.

4.However, you can quote “Streptomyces strains with a greater number of BGCs might also contain a greater number of cryptic silent BGCs due to the presence of complex transcriptional regulatory systems that function to avoid coexpression of multiple BGCs that might be harmful to cell growth” from the reference https://doi.org/10.1128/msystems.00489-21 in support of your argument in point 3. However, please also present the limitation of point 3 in your manuscript.

5.Again from point 3, the # of useful secondary metabolites that can be extracted out of a strain depends on how many of those BGCs can be successfully uncrypticized (which, as you mention is a bottleneck in current research). Therefore, the # of BGC may not be the only metric by which extractable useful compounds are evaluated against.

6.In the disucssion section, you may want to include the following “…..discovered that small promoter insertion– deletion lesions (indels) differentiate weaker PTM producers from stronger ones……” from the reference: https://www.pnas.org/doi/pdf/10.1073/pnas.2103515118. Hence, future studies may also want to document these indels flanking the BGC so as to be able estimate whether a particular BGC will be productive (with regard to secondary metabolite production) or not.

7.The title needs to be changed. As written, it does not adequately capture the objective of the work. Suggest “The untapped potential of endophytic Sreptomyces scabiei in producing antibiotic and antitumor secondary metabolites” or something along those lines.
8.I can (partly) understand a BGC count of 41.643; however you have mentioned in parts of the manuscript of a “BGC diversity”. Can you please change this to “BGC count”? Also, how can the count be a non-integer? Similar question for RiPP counts?

9.you state “….Further analysis of antiSMASH results showed that S. scabiei has a greater number of polyketide synthase BGCs compared to other Streptomyces species……” The references I found do not support this result; for example, this reference contains approximately 37% type 1 PKS (Figure 4), https://doi.org/10.1038/s41598-020-58904-9, this reference contains about the same percentage , https://doi.org/10.1128/mSystems.00489-21, please describe how exactly this “further analysis of antiSMASH” was performed and why your results are not consistent with those reported in the literature.

10.You state “…..Nine out of fourteen of the strains encode for the antitumor BGC Concanamycin A (19). This supports the fact that although S. scabiei produces the same volume of antibiotic products as other Streptomyces species, it produces a wider variety of useful products, specifically with antitumor properties.” I don’t understand how you can conclude this based on the prior sentence. Please explain and discuss in the manuscript.
https://doi.org/10.1128/Spectrum.00571-21
https://www.jstage.jst.go.jp/article/jgam/58/5/58_363/_pdf
https://doi.org/10.1128/mSystems.00489-21
https://doi.org/10.1128/mbio.02729-20
https://doi.org/10.1101/2024.02.20.581055
https://www.pnas.org/doi/pdf/10.1073/pnas.2103515118
https://doi.org/10.3389/fmicb.2024.1408479
https://doi.org/10.1038/s42255-024-01024-9
https://doi.org/10.1186/s12934-024-02416-y
https://doi.org/10.1038/s41598-020-58904-9
https://doi.org/10.1186/s12934-024-02478-y
https://doi.org/10.1093/jimb/kuae014
https://doi.org/10.1371/journal.pone.0289280

---

1. Based on references https://doi.org/10.1111/1751-7915.14382, https://doi.org/10.3389/fmicb.2024.1408479, and references within these articles, the introduction was changed to include further information on the BGC production of endophytic actinomycin and why it is important to perform further analysis on this group. The reference section was updated with these new additions.
2. The hypothesis was specific based on *S. scabiei* having the potential for greater BGC diversity due to it being an endophytic actinomycin. The hypothesis was changed from “the plant pathogenic bacterium S. scabiei will exhibit a greater BGC diversity than previously studied Streptomyces species, and potentially synthesize new antibiotics” to “strain-level investigation of the endophytic actinomycin *S. scabiei* will reveal a greater diversity of BGCs in its genome and the presence of antitumor products”. This change was made in the introduction, and discussion sections.
3. The bootstrap value for the evolutionary tree was included in the last subsection of the results.

4. Comments specifying that Figures 3 and 4 both support the diversity of strain GCF_002155 were removed from the results section. In the discussion section, it was explained that due to the variance in the unique BGC count (Figure 3.B) and evolutionary tree (Figure 4) it is not possible to conclude that investigating a single strain based off its unique BGC count will reveal further BGC diversity.
5. Although unique BGC count alone cannot determine the biosynthetic potential of a strain, there may still be some benefit to using this metric. The reference https://doi.org/10.1128/msystems.00489-21 (given by reviewer) was used to explain that further study of GCF_000738695 (the strain with the highest unique BGC count) may reveal some hidden BGCs.
6. Analyzing cryptic BGCs within a bacterium are stated as a possible additional measure of a bacterium's biosynthetic potential in the discussion section.
7. The reference https://www.pnas.org/doi/pdf/10.1073/pnas.2103515118 (given by reviewer) was used to explain that analyzing indels surround a BGC can allow a greater understanding of the productivity level of a gene cluster. This was stated as a path for further research in the discussion section.
8. The title was changed based on recommendation by the reviewer.
9. All mentions of "BGC diversity" where changed to "BGC count". BGC count is not an integer because it is an average of the total BGCs found in all 14 strains divided by the number of strains (14). This calculation was explained in the first subsection of the results.
10. "Further analysis of antiSMASH results" was clarified in the second subsection of the results. The claim that *S. scabiei* produces more PKS BGCs that other bacteria was removed from this section and the discussion section.
11. The conclusion that *S. scabiei* "produces a wider variety of useful products" was removed from the second subsection of the results and the discussion section. The results were explained to support previous research and showing that due to the presence of PKS gene clusters and Concanamycin A, *S. scabiei* has the potential to produce metabolites with antitumor properties.
12. The abstract was updated according to all above stated changes.

You have not addressed all my comments satisfactorily, rather only inserted or deleted content based on my earlier comments. This has not improved the quality of the manuscript. A well thought out content with detailed assumptions (including those implicit) provide depth and erudition to the manuscript.

The premise of your hypothesis is incorrect and based on an implicit assumption for which you do not provide any references; which is not a deal breaker; **but needs to be identified in the discussion and results. You have not done so.** I do not see how your hypothesiss is "partially supported". Please replace this phrase with "**not supported"**. Please explicity define what "diversity" means (see point 3). Introduce the verbiage that you assume no or limited subsets of BGCs on different strains (is this really the case? references?) In the discussion section, add verbiage to imply that your premise of studying one strain based on BGC count as being the most diverse may not be true (explicitly quote and reproduce verbiage from point 3).
You stater "AntiSMASH analysis revealed that S. scabiei contained a total of 675 BGCs among all fourteen strains (average = 48.2) when run on Relaxed mode and 583 BGCs (average = 41.6) when run on Strict mode. **Make sure this is correct, Figure 3b only shows an average of approx 6 BGCs per strain.**
Please insert http links for all the software used in the manuscript. Please list exactly what webpage and what genome IDs were used.

Each reference needs to have a live link. For references with more than 6 authors, the first 6 must be listed followed by an et al.

I have 'bolded' the part of the comment not addressed in the manuscript. Address these explicitly in the manuscript with the requisite reference(s).

1.**Literature suggests approx 80 BGC in certain species of Streptomyces; see: https://doi.org/10.1038/s41598-020-58904-9, therefore, your number of BGC (approx 15) is significantly lesser**. Why then did you choose to concentrate on this species/strain when there are species that produce many more BGCs ? Maybe a re-write of your hypothesis is in order to make it more specific? "The hypothesis of this study is that the plant pathogenic bacterium S. scabiei will exhibit a greater BGC diversity than previously studied Streptomyces species, and potentially synthesize new antibiotics." **You can also reference https://doi.org/10.1371/journal.pone.0289280 and https://doi.org/10.3389/fmicb.2024.1408479 (and references therein) which state that endophytes have a propensity for a greater # of gene cluster diversity.**

2.Need bootstrap values for the phylogenetic tree in Figure 4. Generally, bootstrap values below 70 are not acceptable. Also, if 002155 is dis-similar from the other 12 strains as well (Figure 4), why does it not show a larger number of BGC in Figure 3b ? (it should either show very low or very high BGC counts.).**I find it difficult to believe that all these tree branches were bootstrapped 100. Recheck.**

3.**From point 2, since 002155 is dissimilar but shows similar BGC counts to the other strains (except 000738), this implies that absolute BGC counts do not necessarily mean that an exclusive study of 000738 will yield the most # of secondary metabolites (because none of the other strains may be subsets of 000738). For example, studying the other 11 strains simultaneously may yield many more compounds (approximately from 3x30=30 BGC) than studying 000738 alone (because none of them may be asumed to be subsets of 000738). This is also illustrated in https://doi.org/10.1038/s41598-020-58904-9. Hence, the premise that because one strain exhibits the most # of BGC; therefore it will yield the most # of secondary metabolites (antibiotics or antitumor agents) at the expense of the rest is false.**

4.However, you can quote "Streptomyces strains with a greater number of BGCs might also contain a greater number of cryptic silent BGCs due to the presence of complex transcriptional regulatory systems that function to avoid coexpression of multiple BGCs that might be harmful to cell growth"**from the reference https://doi.org/10.1128/msystems.00489-21 in support of your argument in point 3. However, please also present the limitation of point 3 in your manuscript.**

5.Again from point 3, the # of useful secondary metabolites that can be extracted out of a strain depends on how many of those BGCs can be successfully uncrypticized (which, as you mention is a bottleneck in current research). Therefore, the # of BGC may not be the only metric by which extractable useful compounds are evaluated against.

6.In the disucssion section, you may want to include the following "…..discovered that small promoter insertion– deletion lesions (indels) differentiate weaker PTM producers from stronger ones……" from the reference: https://www.pnas.org/doi/pdf/10.1073/pnas.2103515118. Hence, future studies may also want to document these indels flanking the BGC so as to be able estimate whether a particular BGC will be productive (with regard to secondary metabolite production) or not.

7.The title needs to be changed. As written, it does not adequately capture the objective of the work. Suggest “The untapped potential of endophytic Sreptomyces scabiei in producing antibiotic and antitumor secondary metabolites” or something along those lines.

8.I can (partly) understand a BGC count of 41.643; however you have mentioned in parts of the manuscript of a “BGC diversity”. Can you please change this to “BGC count”? Also, how can the count be a non-integer? Similar question for RiPP counts?

9.you state “….Further analysis of antiSMASH results showed that S. scabiei has a greater number of polyketide synthase BGCs compared to other Streptomyces species……” The references I found do not support this result; for example, this reference contains approximately 37% type 1 PKS (Figure 4), https://doi.org/10.1038/s41598-020-58904-9, this reference contains about the same percentage , https://doi.org/10.1128/mSystems.00489-21, please describe how exactly this “further analysis of antiSMASH” was performed and why your results are not consistent with those reported in the literature.

10.You state “…..Nine out of fourteen of the strains encode for the antitumor BGC Concanamycin A (19). This supports the fact that although S. scabiei produces the same volume of antibiotic products as other Streptomyces species, it produces a wider variety of useful products, specifically with antitumor properties.” I don’t understand how you can conclude this based on the prior sentence. Please explain and discuss in the manuscript.
https://doi.org/10.1128/Spectrum.00571-21
https://www.jstage.jst.go.jp/article/jgam/58/5/58_363/_pdf
https://doi.org/10.1128/mSystems.00489-21
https://doi.org/10.1128/mbio.02729-20
https://doi.org/10.1101/2024.02.20.581055
https://www.pnas.org/doi/pdf/10.1073/pnas.2103515118
https://doi.org/10.3389/fmicb.2024.1408479
https://doi.org/10.1038/s42255-024-01024-9
https://doi.org/10.1186/s12934-024-02416-y
https://doi.org/10.1038/s41598-020-58904-9
https://doi.org/10.1186/s12934-024-02478-y
https://doi.org/10.1093/jimb/kuae014
https://doi.org/10.1371/journal.pone.0289280
In summary, this manuscript needs in depth discussion, extensive revision and inclusion of all (or most) of the references suggested.

Thank you very much for your feedback, I greatly appreciate it. The comments helped to clarify and better articulate sections of the paper. Please let me know if there are other areas I can improve.

13. I have stated that my hypothesis is based on an “implicit inference” and explained the premise that this inference was made on in the discussion section. The conclusion of the study is changed to the hypothesis being “not supported”. I further explained that studying only GCF_000738695 “will yield the greatest number of bioactive compounds” in the discussion section. I have added in verbiage as you stated and also used horizontal gene transfer as an additional reason why it may be more beneficial to study multiple strains from the reference https://doi.org/10.1038/s41598-020-58904-9.

14. I realized that I was not clear about what “Strain-Specific BGC Count” is defined as. I have added in an explicit definition in the introduction as “Strain-specific BGCs are those that are only found in one strain of a bacterial species”. For example, the BGC foxicin A is only found in strain GCF_00148125.1 and is a “Strain-Specific BGC”. In Figure 3.B, the strain-specific BGC count represents this, not the total number of BGC that were found in a strain. I have added Figure 2A in order to make this more clear. Figure 2.A shows the total number of BGCs found in a strain when AntiSMASH was run on strict.
15. I have included the statistic that up to 80 BGCs have been found in certain *Streptomyces* strains from the reference https://doi.org/10.1038/s41598-020-58904-9 in the discussion section to explain why my hypothesis was not supported.
16. In the second subsection of Results I have stated that studies have found up to 37% PKS in a bacterial genome. I believe that the difference between this percentage and my 31% is due to the different species that were examined. For the reference https://doi.org/10.1128/mSystems.00489-21 that you pointed out it is stated that, “The total number of BGCs identified by antiSMASH across 11 different genomes was 576 (on average, 52 BGCs per genome) including 109 T1PKS, 96 NRPS, and 44 hybrid gene clusters”. This would mean that 18.9% of the genome is T1PKS. I found 583 BGCs (on strict) with 114 T1PKS gene clusters (also data from strict), which makes 19.6% of the genome T1PKS (the 31% stated in the manuscript is a total of T1PKS and T2PKS). This shows my results are consistent with those reported in literature.
17. To perform “further analysis of AntiSMASH results” I imported the data I found into Excel. I have explained this process in the second subsection of Methods.
18. I have defined “BGC diversity” in the introduction as “the number of strain-specific BGCs found in a strain”.
19. I double-checked the bootstrap values used on the phylogenetic tree and they are 100. The pangenome matrix shown in the Figure 4 is generated by Roary which infers the evolutionary tree. I checked the accuracy of this tree by building one using RAxML and then visualizing it using Interactive Tree of Life. The two trees are the same except for 4 strains at the bottom which are classified on different branches in Roary but on the same branch in RAxML. However, I don’t think this is significant enough of a difference as RAxML still reports different branch lengths. I have included attachments to the iTOL tree visualization and RAxML output summary file for the tree below.

20. I have included http links for all the software used after they are mentioned in the manuscript. I have also included a link to the webpage where the genomes were downloaded from and a table with genome ID information in the first subsection of Methods.
21. I have updated the reference section with the first 6 authors listed and doi links for each reference.

Thank you for addressing my comments. Accept.