

Functional capacities drive recruitment of bacteria into plant root microbiota

Corresponding Author: Professor Simona Radutoiu

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Plant microbiome plays crucial roles for plant fitness, however, mechanistic understanding about the principles governing root microbiome selection and assembly is still in its infancy. This study inoculate different plants with large amount of known isolates from previous studies. They analyzed the enriched functional genes overlap among root microbiome in different plants under different nutrient conditions. Thus successfully characterized both conserved “core” and species specific microbiome functional genes, as well as their distribution patterns among different taxa. The work could be interesting as a large scale microbiome reconstruction and functional description study. However, the findings in this work are still superficial; there is no molecular or SynCom based validation of their correlative analyses; there is no strong and clear key finding and their significance in guiding microbiome engineering is weak.

1. The manuscript try to study the principles of microbiome assembly in the rhizosphere. However, the current descriptive analyses lacks experimental validations. They have to integrate root metabolomics and essential validations into this study. To test whether root metabolites changes can be matched to the key functional changes in the assembled genomes. This limits the depth and mechanistic importance of this work.

2. The introduction is not focused on a clear question with a strong logic. The first paragraph of the introduction did not precisely introduce the key scientific question. This work did not address how “microbes are selected from the environment to form these niche-specific communities”, also not related to dysbiosis mechanisms or growth promotion. The manuscript could more carefully propose the key scientific question and novelty in the introduction.

3. The authors identified 266 functions enriched across all root microbiomes. Can they test what will happen if they chose strains defect in those key functions? The authors need to validate at least some of the predicted core functional genes required for rhizosphere colonization, similar to their cited study did <https://www.nature.com/articles/s41588-017-0012-9>.

4. What exactly is the key novel finding important for the field from this study? Can this study help design a SynCom with the potential to enhance plant growth? Can they design different SynComs with different colonization activities (through quantitative microbiome profiling) to guide rational design of SynComs? Can they manipulate SynCom composition based on functional studies?

5. A pretty interesting finding is that none of the 988 isolates encoded the full set of core functions, and most strains encoded fewer than 25# core functions required for rhizosphere fitness. However, it seems that most strains can colonize roots even better in mono-association compared to in natural soil system, due to less competition. Does that means in real world most strains does not need full set of core functions to colonize rhizosphere? Or in other words, the “full set functions” Identified here are not all needed for colonization?

6. The authors could analyze the Co-occurrence networks of major functional groups, to better illustrate the relationship and potential significance of different gene groups in rhizosphere ecosystem.

7. The authors only selected three very different plant species for comparing “core functional genes”. There are two potential problems here: 1) three plants are not sufficient for effectively characterize any “core” features for plant kingdom (since inoculating more plant species does not cost too much time as long as they prepared the strains already); 2) Only one genotype in each plant is not sufficient for characterizing species specific “core” stuff.

8. The authors successfully identified 76 to 146 isolates (averaging 119, RA >0.05%) were consistently associated with the roots across treatments. This could be interesting but lacks further investigation. Why those strains are “core” to different plants, are there any genomics features? What is the implication of significance for manipulating plant microbiome?

9. The authors should test whether knock-out of type VI secretion system negatively regulates the success of Rhizobiaceae colonization?

10. Line 495-506 raises concern about the novelty of this study. It seems there are a lot overlap with previous study, and there are fewer potential host plants in this study.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

This is a groundbreaking study. In it, Selton et al. perform a truly visionary experiment where they inoculate three different host plants with large root-derived bacterial isolate collections cultured from each separate host. In addition to inoculating each collection separately, they also assemble a super SC composed of all three AND include a nutrient treatment in a fully crossed manner. Replicate metagenomes of each treatment combination were comprehensively characterized using a battery of creative and informative analyses. The forward-looking design makes significant progress towards a mechanistic understanding of host plant-specificity and the shared and unique genetic components of competency across plant microbiome inhabitants and host plant species.

Importantly, the authors are very upfront about the limitations of their approach throughout the text, however, the scale and depth of the study are indeed unprecedented. Thus, this paper will make a great impact on the field.

My only major comment is that the results are very much interpreted through a lens of active recruitment and selection on the part of plant hosts. I agree that the host plants demonstrate differences in the taxonomic and functional composition of the isolates occurring in their roots across all SCs, but why must this be the result of 'active' recruitment? Furthermore, what does that even mean? It seems analogous to saying that alpine plants are actively recruited to montane habitats by virtue of their benefit to the ecosystem rather than a more simple explanation that they possess attributes that allow them to persist and reproduce in those environments to a greater extent than other species. Can a similar explanation not be used for the microbial taxa enriched in the microbiome of plants and particular plant species?

Minor comments

Line 142: How do we interpret this proportion of shared KOs among the three isolate collections? What might serve as a good reference is KO profiles generated from different soil metagenomes, ideally from the same source as the isolate collections, so that one can place this shared proportion among root bacterial endophytes in the context of soil bacterial KO variability.

Line 217: Would there be any value in attempting to incorporate an abundance weighting into the characterization of functional diversity? Just as the authors used Bray-Curtis dissimilarity, which is calculating differences in community composition based on taxon presence AND abundance, is there a way to capture the abundance of different KO in order to provide a more refined calculation of functional variation.

Line 274: I'm not entirely sure what this means. Does it mean that 70% of the maximum KO diversity harboured by Lj (across all SC?) is coming from individual members of SSC, LjSC, or HvSC? But doesn't the SSC contain the LjSC and HvSC?

Line 215 and 281: Perhaps contrast this to: <https://doi.org/10.1038/s41396-020-0695-2>

Line 299: What about actually dropping it out of the SC? I imagine such a dominant colonizer is likely having an effect on community assembly that will not be observed by in silico removal.

Line 305: Selective host recruitment implies that the host is actively participating in this process. However, inoculum dependent enrichment contradicts this statement.

Line 317: Can this be illustrated with a figure e.g. Bray-Curtis dissimilarity versus KO dissimilarity?

Figure 2b,c: Can the authors explain why the observed strains in the input does not match the number of strains included in the inoculum?

Figure 2f,g: I think the grey versus black outlines refer to the nutrient status but this may be worth including in the legend somewhere more clearly.

Figure 2h: I'm surprised that the nutrient treatment yielded such a weak effect on the taxonomic and functional root microbial composition? Fig S3 reveals quite a large nutrient effect on the biomass of *Hordeum* and *Lotus* but little on *Arabidopsis*. Could the authors speculate why a biologically relevant nutrient stress (at least for 2 of the 3 hosts) is having little effect on the identity of the enriched KOs? Are there particular environmental conditions where this might not be the case? I see the paragraph addressing this in the discussion (lines 585-608), but I'm trying to place this result in the context of disease states or other abiotic stress, where plants are often observed to recruit a unique microbiota.

Figure 3a: I think the authors need to provide more operational and biological explanation of these values. If I understand correctly, the authors are decomposing the KO functional variation. The inner ring represents the R2 value output from Adonis for the host species term for each inoculum. The outer ring represents the contribution of each host on to that effect represented by the species term, the mathematical formula underlying this provided in the supplemental text. If I have this wrong, please clarify

in text. If I have it right, great! I still need a bit of help interpreting it. When the authors report that Lotus has a contribution of 82.1% of the functional variation observed with the SSC inoculum, I interpret this as Lotus assembling a root microbiome that occupies a more unique multivariate functional space versus Hordeum or Arabidopsis when exposed to the same inoculum. Is this correct?

Line 325: I need more help interpreting this result and Figure 3b. Why is the intersection of the explanatory power of the Adonis terms when analyzing either taxonomic or functional composition important (i.e. the selection of Family from Figure 3b)?

Line 340: I'm not sure I'm convinced by the data to support this claim. There are several instances in Fig S14 where the most abundant isolate is not harbouring the highest proportion of KO diversity at the genus level. Is there a randomization test to perform here? I doubt r values would be significant for anything in Fig S13 or S14.

Lines 402-408: What an interesting result. Is this surprising to the authors given the high host-specificity reported elsewhere in the study? Does this suggest other genetic factors are at play during host selection, e.g. interactions with host-immunity? The following published article show a similar finding with a single bacterial strain: <https://doi.org/10.1038/s41467-025-61266-3>

Lines 484-493: Another fascinating result. Perhaps beyond the scope of the current paper but I'm very curious to know whether this represents a shared feature among all members of these families, including those not plant-associated? Additionally, this result yields very clear and distinct ecological hypotheses regarding the competition among members within these families occurring in the plant niche.

Line 502: A chi-squared test or Fishers exact test might help us interpret the significance of these different fractions.

Line 504: This is a remarkable result given that none of the genomes in Levy et al. are present in the current study, correct?

Line 566: I agree that there isn't much in the plant microbiome literature to support this claim, possibly <https://doi.org/10.1128/mbio.02892-21>

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

Gijs and co-workers presented a compelling story on the deconstructing and reconstructing the microbial communities thriving in association with plant roots, designated root microbiota. In analogy with the communities inhabiting the digestive tract of the vertebrates, the root microbiota impacts on the growth, development and health of their hosts. Since the 2010s, access to high-throughput sequencing technologies enabled scientists to gather insights into the molecular basis of plant-microbiota interactions. Where many of these studies fall short is in providing an understanding beyond taxonomy [i.e., what for?] and comparing host recruitment strategies [i.e. are all plants doing the same?]. This submission aims at filling these knowledge gaps and provide the scientific communities with a blueprint for follow-up engineering strategies.

Therefore, in my vision, this submission is novel and holds a potential for impact, both in basic science and translational applications.

Furthermore, the research approach is ingenious and will likely benefit scholars and professionals beyond the remit plant sciences.

However, there are a number of elements of this submission that are unclear to me:

1) the ecological links with the soil data are tenuous: while the authors provide evidence of strains recovery, what is missing is a like-with-like comparison of microbiota assembly in soil and under the conditions used by the authors. Results will inform on the "efficiency of translation" of the proposed work, or limitations between the two approaches - either way an informative result.

2) There is no closing model: do I interpret correctly that the majority of the 266 KOs are undefined (Fig S27)? In other words, you've got a number but we still do not know why plants want them? Resolving this knowledge gap will greatly strengthen the significance of your work.

3) Dataset S8 requires a clarification[see also comments in the main text and supplementary] as you have sample types sequenced not mentioned in the main text - what was their fate? Likewise, there is an average on 30% unmapped reads [including in the inoculum...] that challenges the whole concept of gnotobiotic system and, by extension, your discoveries.

4) Likewise, there are controls that are mentioned but not discussed in the main text. Those. e.g., unplanted control, mycorrhized specimens..., would be instrumental in strengthening your conclusions.

5) I'd soften some of the conclusions for which you do not have experimental support [other than in silico analysis]

I attached an annotated version of main text and supplementary with further feedback

#End

Davide Bulgarelli [signed for 'transparent review' requirements]

(Remarks on code availability)

Checked the codes on line, not reviewed though - at first glance everything is there.

However, instead of a unique R script I suggest creating different scripts for figures [or block of figures]. This will streamline data checking and reproducibility.

Decision Letter:

12th November 2025

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Professor Radutoiu,

Thank you for your patience while your manuscript "Functional capacities drive recruitment of bacteria into plant root microbiota" was under peer-review at Nature Microbiology. It has now been seen by 3 referees, whose expertise and comments you will find at the end of this email. Although they find your work of some potential interest, they have raised several concerns that will need to be addressed before we can consider publication of the work in Nature Microbiology.

In particular, Reviewer 1 is keen to see further experimental validation to understand the functional genomics of the identified community in root colonization. While we feel a metabolomic approach is NOT necessary here, we DO feel that this question can be addressed experimentally in the lab within a reasonable timeframe and would indeed further strengthen your study. Thus, we would like to see the functional characteristics described in your study experimentally supported in your revised manuscript. Reviewer 1 has also suggested SymCom validation, we feel this is beyond the scope of the current study and therefore it is NOT necessary for this manuscript. Aside from this, much of the feedback is regarding statistical approaches and additional analysis, clarification on your experimental approaches and further discussion regarding your findings and implications for these in the broader plant microbiology field, which are all addressable.

Should further experimental data allow you to address these criticisms, we would be happy to look at a revised manuscript.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

We strongly support public availability of data. Please place the data used in your paper into a public data repository, if one exists, or alternatively, present the data as Source Data or Supplementary Information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor. For some data types, deposition in a public repository is mandatory - more information on our data deposition policies and available repositories can be found at <https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data>.

Please include a data availability statement as a separate section after Methods but before references, under the heading "Data Availability". This section should inform readers about the availability of the data used to support the conclusions of your study. This information includes accession codes to public repositories (data banks for protein, DNA or RNA sequences, microarray, proteomics data etc...), references to source data published alongside the paper, unique identifiers such as URLs to data repository entries, or data set DOIs, and any other statement about data availability. At a minimum, you should include the following statement: "The data that support the findings of this study are available from the corresponding author upon request", mentioning any restrictions on availability. If DOIs are provided, we also strongly encourage including these in the Reference list (authors, title, publisher (repository name), identifier, year). For more guidance on how to write this section please see: <http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf>

If revising your manuscript:

* Include a "Response to referees" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

* If you have not done so already we suggest that you begin to revise your manuscript so that it conforms to our Article format instructions at <http://www.nature.com/nmicrobiol/info/final-submission>. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

When submitting the revised version of your manuscript, please pay close attention to our [href="https://www.nature.com/nature-portfolio/editorial-policies/image-integrity">Digital Image Integrity Guidelines](https://www.nature.com/nature-portfolio/editorial-policies/image-integrity) and to the following points below:

- that unprocessed scans are clearly labelled and match the gels and western blots presented in figures.
- that control panels for gels and western blots are appropriately described as loading on sample processing controls
- all images in the paper are checked for duplication of panels and for splicing of gel lanes.

EXTENDED DATA FIGURES

When re-submitting your manuscript, please ensure that any supplementary figures and tables that are crucial to the manuscript's conclusions are converted into Extended Data figures and tables to increase visibility of these data. Extended Data figures and tables are online-only (present in the online PDF and full-text HTML versions of the paper), peer-reviewed display items that provide essential background to the article but are not included in the main article due to space constraints. A maximum of ten Extended Data display items (figures and tables) is permitted.

Finally, please ensure that you retain unprocessed data and metadata files after publication, ideally archiving data in perpetuity, as these may be requested during the peer review and production process or after publication if any issues arise.

Please use the link below to submit a revised paper:

Link Redacted

Note: This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first.

Nature Microbiology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. This applies to primary research papers only. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit <http://www.springernature.com/orcid>.

If you wish to submit a suitably revised manuscript we would hope to receive it within 6 months. If you cannot send it within this time, please let us know. We will be happy to consider your revision, even if a similar study has been accepted for publication at Nature Microbiology or published elsewhere (up to a maximum of 6 months).

In the meantime we hope that you find our referees' comments helpful.

Yours sincerely,

Reviewer Expertise:

Referee #1: mechanisms of shaping/re-shaping rhizosphere microbiome, regulation of root colonization

Referee #2: plant-microbe interactions

Referee #3: molecular plant-microbe interactions, rhizosphere

Reviewer Comments:

Reviewer #1 (Remarks to the Author):

Plant microbiome plays crucial roles for plant fitness, however, mechanistic understanding about the principles governing root microbiome selection and assembly is still in its infancy. This study inoculate different plants with large amount of known isolates from previous studies. They analyzed the enriched functional genes overlap among root microbiome in different plants under different nutrient conditions. Thus successfully characterized both conserved "core" and species specific microbiome functional genes, as well as their distribution patterns among different taxa. The work could be interesting as a large scale microbiome reconstruction and functional description study. However, the findings in this work are still superficial; there is no molecular or SynCom based validation of their correlative analyses; there is no strong and clear key finding and their significance in guiding microbiome engineering is weak.

1.The manuscript try to study the principles of microbiome assembly in the rhizosphere. However, the current descriptive analyses lacks experimental validations. They have to integrate root metabolomics and essential validations into this study. To test whether root metabolites changes can be matched to the key functional changes in the assembled genomes. This limits the depth and mechanistic importance of this work.

2.The introduction is not focused on a clear question with a strong logic. The first paragraph of the introduction did not precisely

introduce the key scientific question. This work did not address how “microbes are selected from the environment to form these niche-specific communities”, also not related to dysbiosis mechanisms or growth promotion. The manuscript could more carefully propose the key scientific question and novelty in the introduction.

3. The authors identified 266 functions enriched across all root microbiomes. Can they test what will happen if they chose strains deficient in those key functions? The authors need to validate at least some of the predicted core functional genes required for rhizosphere colonization, similar to their cited study <https://www.nature.com/articles/s41588-017-0012-9>.

4. What exactly is the key novel finding important for the field from this study? Can this study help design a SynCom with the potential to enhance plant growth? Can they design different SynComs with different colonization activities (through quantitative microbiome profiling) to guide rational design of SynComs? Can they manipulate SynCom composition based on functional studies?

5. A pretty interesting finding is that none of the 988 isolates encoded the full set of core functions, and most strains encoded fewer than 25# core functions required for rhizosphere fitness. However, it seems that most strains can colonize roots even better in mono-association compared to in natural soil system, due to less competition. Does that mean in real world most strains do not need full set of core functions to colonize rhizosphere? Or in other words, the “full set functions” identified here are not all needed for colonization?

6. The authors could analyze the Co-occurrence networks of major functional groups, to better illustrate the relationship and potential significance of different gene groups in rhizosphere ecosystem.

7. The authors only selected three very different plant species for comparing “core functional genes”. There are two potential problems here: 1) three plants are not sufficient for effectively characterizing any “core” features for plant kingdom (since inoculating more plant species does not cost too much time as long as they prepared the strains already); 2) Only one genotype in each plant is not sufficient for characterizing species specific “core” stuff.

8. The authors successfully identified 76 to 146 isolates (averaging 119, RA >0.05%) were consistently associated with the roots across treatments. This could be interesting but lacks further investigation. Why those strains are “core” to different plants, are there any genomics features? What is the implication of significance for manipulating plant microbiome?

9. The authors should test whether knock-out of type VI secretion system negatively regulates the success of Rhizobiaceae colonization?

10. Line 495-506 raises concern about the novelty of this study. It seems there is a lot of overlap with previous study, and there are fewer potential host plants in this study.

Reviewer #2 (Remarks to the Author):

This is a groundbreaking study. In it, Selton et al. perform a truly visionary experiment where they inoculate three different host plants with large root-derived bacterial isolate collections cultured from each separate host. In addition to inoculating each collection separately, they also assemble a super SC composed of all three AND include a nutrient treatment in a fully crossed manner. Replicate metagenomes of each treatment combination were comprehensively characterized using a battery of creative and informative analyses. The forward-looking design makes significant progress towards a mechanistic understanding of host plant-specificity and the shared and unique genetic components of competency across plant microbiome inhabitants and host plant species.

Importantly, the authors are very upfront about the limitations of their approach throughout the text, however, the scale and depth of the study are indeed unprecedented. Thus, this paper will make a great impact on the field.

My only major comment is that the results are very much interpreted through a lens of active recruitment and selection on the part of plant hosts. I agree that the host plants demonstrate differences in the taxonomic and functional composition of the isolates occurring in their roots across all SCs, but why must this be the result of ‘active’ recruitment? Furthermore, what does that even mean? It seems analogous to saying that alpine plants are actively recruited to montane habitats by virtue of their benefit to the ecosystem rather than a more simple explanation that they possess attributes that allow them to persist and reproduce in those environments to a greater extent than other species. Can a similar explanation not be used for the microbial taxa enriched in the microbiome of plants and particular plant species?

Minor comments

Line 142: How do we interpret this proportion of shared KOs among the three isolate collections? What might serve as a good reference is KO profiles generated from different soil metagenomes, ideally from the same source as the isolate collections, so that one can place this shared proportion among root bacterial endophytes in the context of soil bacterial KO variability.

Line 217: Would there be any value in attempting to incorporate an abundance weighting into the characterization of functional diversity? Just as the authors used Bray-Curtis dissimilarity, which is calculating differences in community composition based on taxon presence AND abundance, is there a way to capture the abundance of different KO in order to provide a more refined calculation of functional variation.

Line 274: I'm not entirely sure what this means. Does it mean that 70% of the maximum KO diversity harboured by Lj (across all SC?) is coming from individual members of SSC, LjSC, or HvSC? But doesn't the SSC contain the LjSC and HvSC?

Line 215 and 281: Perhaps contrast this to: <https://doi.org/10.1038/s41396-020-0695-2>

Line 299: What about actually dropping it out of the SC? I imagine such a dominant colonizer is likely having an effect on community assembly that will not be observed by *in silico* removal.

Line 305: Selective host recruitment implies that the host is actively participating in this process. However, inoculum dependent enrichment contradicts this statement.

Line 317: Can this be illustrated with a figure e.g. Bray-Curtis dissimilarity versus KO dissimilarity?

Figure 2b,c: Can the authors explain why the observed strains in the input does not match the number of strains included in the inoculum?

Figure 2f,g: I think the grey versus black outlines refer to the nutrient status but this may be worth including in the legend somewhere more clearly.

Figure 2h: I'm surprised that the nutrient treatment yielded such a weak effect on the taxonomic and functional root microbial composition? Fig S3 reveals quite a large nutrient effect on the biomass of *Hordeum* and *Lotus* but little on *Arabidopsis*. Could the authors speculate why a biologically relevant nutrient stress (at least for 2 of the 3 hosts) is having little effect on the identity of the enriched KOs? Are there particular environmental conditions where this might not be the case? I see the paragraph addressing this in the discussion (lines 585-608), but I'm trying to place this result in the context of disease states or other abiotic stress, where plants are often observed to recruit a unique microbiota.

Figure 3a: I think the authors need to provide more operational and biological explanation of these values. If I understand correctly, the authors are decomposing the KO functional variation. The inner ring represents the R² value output from Adonis for the host species term for each inoculum. The outer ring represents the contribution of each host on to that effect represented by the species term, the mathematical formula underlying this provided in the supplemental text. If I have this wrong, please clarify in text. If I have it right, great! I still need a bit of help interpreting it. When the authors report that *Lotus* has a contribution of 82.1% of the functional variation observed with the SSC inoculum, I interpret this as *Lotus* assembling a root microbiome that occupies a more unique multivariate functional space versus *Hordeum* or *Arabidopsis* when exposed to the same inoculum. Is this correct?

Line 325: I need more help interpreting this result and Figure 3b. Why is the intersection of the explanatory power of the Adonis terms when analyzing either taxonomic or functional composition important (i.e. the selection of Family from Figure 3b)?

Line 340: I'm not sure I'm convinced by the data to support this claim. There are several instances in Fig S14 where the most abundant isolate is not harbouring the highest proportion of KO diversity at the genus level. Is there a randomization test to perform here? I doubt *r* values would be significant for anything in Fig S13 or S14.

Lines 402-408: What an interesting result. Is this surprising to the authors given the high host-specificity reported elsewhere in the study? Does this suggest other genetic factors are at play during host selection, e.g. interactions with host-immunity? The following published article show a similar finding with a single bacterial strain: <https://doi.org/10.1038/s41467-025-61266-3>

Lines 484-493: Another fascinating result. Perhaps beyond the scope of the current paper but I'm very curious to know whether this represents a shared feature among all members of these families, including those not plant-associated? Additionally, this result yields very clear and distinct ecological hypotheses regarding the competition among members within these families occurring in the plant niche.

Line 502: A chi-squared test or Fishers exact test might help us interpret the significance of these different fractions.

Line 504: This is a remarkable result given that none of the genomes in Levy et al. are present in the current study, correct?

Line 566: I agree that there isn't much in the plant microbiome literature to support this claim, possibly <https://doi.org/10.1128/mbio.02892-21>

Reviewer #3 (Remarks to the Author):

Gijs and co-workers presented a compelling story on the deconstructing and reconstructing the microbial communities thriving in association with plant roots, designated root microbiota. In analogy with the communities inhabiting the digestive tract of the vertebrates, the root microbiota impacts on the growth, development and health of their hosts. Since the 2010s, access to high-throughput sequencing technologies enabled scientists to gather insights into the molecular basis of plant-microbiota interactions. Where many of these studies fall short is in providing an understanding beyond taxonomy [i.e., what for?] and comparing host recruitment strategies [i.e. are all plants doing the same?]. This submission aims at filling these knowledge gaps and provide the scientific communities with a blueprint for follow-up engineering strategies. Therefore, in my vision, this submission is novel and holds a potential for impact, both in basic science and translational

applications.

Furthermore, the research approach is ingenious and will likely benefit scholars and professionals beyond the remit plant sciences.

However, there are a number of elements of this submission that are unclear to me:

1) the ecological links with the soil data are tenuous: while the authors provide evidence of strains recovery, what is missing is a like-with-like comparison of microbiota assembly in soil and under the conditions used by the authors. Results will inform on the "efficiency of translation" of the proposed work, or limitations between the two approaches - either way an informative result.

2) There is no closing model: do I interpret correctly that the majority of the 266 KOs are undefined (Fig S27)? In other words, you've got a number but we still do not know why plants want them? Resolving this knowledge gap will greatly strengthen the significance of your work.

3) Dataset S8 requires a clarification[see also comments in the main text and supplementary] as you have sample types sequenced not mentioned in the main text - what was their fate? Likewise, there is an average on 30% unmapped reads [including in the inoculum..] that challenges the whole concept of gnotobiotic system and, by extension, your discoveries.

4) Likewise, there are controls that are mentioned but not discussed in the main text. Those. e.g., unplanted control, mycorrhized specimens..., would be instrumental in strengthening your conclusions.

5) I'd soften some of the conclusions for which you do not have experimental support [other than in silico analysis]

I attached an annotated version of main text and supplementary with further feedback

#End

Davide Bulgarelli [signed for 'transparent review' requirements]

Reviewer #3 (Remarks on code availability):

Checked the codes on line, not reviewed though - at first glance everything is there.

However, instead of a unique R script I suggest creating different scripts for figures [or block of figures]. This will streamline data checking and reproducibility.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The manuscript conducted large-scale SynCom based microbiome re-construction and functional analysis. During revision, this work tested the community assembly rules related to SynCom composition and functions, which validates previous microbial ecological analyses. This version addressed my previous comments.

Minor checks:

*Line 162, which figure is related to the claim should be listed.

*Line 171, the title mentioned commensal interactions, while the results did not really do interaction analyses.

*Line 220, check whether this can be seen in Fig.2 bc.

(Remarks on code availability)

Due to limited time.

Reviewer #2

(Remarks to the Author)

No further comments. Very nice work.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

First of all let me apologise to the Authors and the Editorial Board for the delay in answering - due to unforeseen circumstances I had to rearrange my schedule and deadlines.

In this revised version, Selten and co-workers satisfactory addressed all my criticisms and suggestions.

You are "good-to-go" for what is my concern.

I'd like to take this occasion to praise the truly compelling experimentation performed by the teams. This will undoubtedly benefit the scientific community and beyond - keep up the good work!

Sincerely yours,

Davide Bulgarelli [singled for transparent peer-review requirements]

(Remarks on code availability)

Checked the code - not reviewed.

However, it seems complete and the authors restructured the repository according to my suggestion. I am confident this will streamline data reproducibility.

Decision Letter:

Our ref: NMICROBIOL-25093246A

18th June 2026

Dear Dr. Radutoiu,

Thank you for submitting your revised manuscript "Functional capacities drive recruitment of bacteria into plant root microbiota" (NMICROBIOL-25093246A). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Microbiology, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

If the current version of your manuscript is in a PDF format, please email us a copy of the file in an editable format (Microsoft Word or LaTeX)-- we can not proceed with PDFs at this stage.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in a couple of weeks. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Microbiology Please do not hesitate to contact me if you have any questions.

Sincerely,

Reviewer #1 (Remarks to the Author):

The manuscript conducted large-scale SynCom based microbiome re-construction and functional analysis. During revision, this work tested the community assembly rules related to SynCom composition and functions, which validates previous microbial ecological analyses. This version addressed my previous comments.

Minor checks:

*Line 162, which figure is related to the claim should be listed.

*Line 171, the title mentioned commensal interactions, while the results did not really do interaction analyses.

*Line 220, check whether this can be seen in Fig.2 bc.

Reviewer #1 (Remarks on code availability):

Due to limited time.

Reviewer #2 (Remarks to the Author):

No further comments. Very nice work.

Reviewer #3 (Remarks to the Author):

First of all let me apologise to the Authors and the Editorial Board for the delay in answering - due to unforeseen circumstances I had to rearrange my schedule and deadlines.

In this revised version, Selten and co-workers satisfactory addressed all my criticisms and suggestions.

You are "good-to-go" for what is my concern.

I'd like to take this occasion to praise the truly compelling experimentation performed by the teams. This will undoubtedly benefit the scientific community and beyond - keep up the good work!

Sincerely yours,

Davide Bulgarelli [singed for transparent peer-review requirements]

Reviewer #3 (Remarks on code availability):

Checked the code - not reviewed.

However, it seems complete and the authors restructured the repository according to my suggestion. I am confident this will streamline data reproducibility.

INTERNAL USE ONLY AIP modules required: ["Life Science"]

Version 2:

Decision Letter:

18th August 2026

Dear Professor Radutoiu,

I am pleased to accept your Article "Functional capacities drive recruitment of bacteria into plant root microbiota" for publication in Nature Microbiology. Thank you for having chosen to submit your work to us and many congratulations.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Microbiology style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

You may wish to make your media relations office aware of your accepted publication, in case they consider it appropriate to organize some internal or external publicity. Once your paper has been scheduled you will receive an email confirming the publication details. This is normally 3-4 working days in advance of publication. If you need additional notice of the date and time of publication, please let the production team know when you receive the proof of your article to ensure there is sufficient time to coordinate. Further information on our embargo policies can be found here:

<https://www.nature.com/authors/policies/embargo.html>

After the grant of rights is completed, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. You will not receive your proofs until the publishing agreement has been received through our system

Due to the importance of these deadlines, we ask you please us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see <https://www.nature.com/nmicrobiol/editorial-policies>). In particular your manuscript must not be published elsewhere.

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to <https://www.springernature.com/gp/open-science/plan-s-compliance> or the [Plan S principles](https://www.springernature.com/gp/open-science/us-federal-agency-compliance)) or the [NIH public access policy](https://www.springernature.com/gp/open-science/us-federal-agency-compliance)) then you

should select the gold OA route, and we will direct you to the compliant route where possible. Because authors warrant under our subscription licensing terms that they haven't committed to licensing any version of their article under a licence inconsistent with the terms of our agreement – including the applicable embargo period – publication under the subscription model isn't suitable for authors whose funders require no embargo.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

An online order form for reprints of your paper is available at <https://www.nature.com/reprints/author-reprints.html>. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Microbiology as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Microbiology logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

With kind regards,

P.S. Click on the following link if you would like to recommend Nature Microbiology to your librarian
<http://www.nature.com/subscriptions/recommend.html#forms>

** Visit the Springer Nature Editorial and Publishing website at http://editorial-jobs.springernature.com?utm_source=ejp_NMicro_email&utm_medium=ejp_NMicro_email&utm_campaign=ejp_NMicro for more information about our career opportunities. If you have any questions please click [here](mailto:editorial.publishing.jobs@springernature.com). **

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source. The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Editorial Comments

Thank you for your patience while your manuscript "Functional capacities drive recruitment of bacteria into plant root microbiota" was under peer-review at Nature Microbiology. It has now been seen by 3 referees, whose expertise and comments you will find at the end of this email. Although they find your work of some potential interest, they have raised several concerns that will need to be addressed before we can consider publication of the work in Nature Microbiology.

In particular, Reviewer 1 is keen to see further experimental validation to understand the functional genomics of the identified community in root colonization. While we feel a metabolomic approach is NOT necessary here, we DO feel that this question can be addressed experimentally in the lab within a reasonable timeframe and would indeed further strengthen your study. Thus, we would like to see the functional characteristics described in your study experimentally supported in your revised manuscript. Reviewer 1 has also suggested SynCom validation, we feel this is beyond the scope of the current study and therefore it is NOT necessary for this manuscript. Aside from this, much of the feedback is regarding statistical approaches and additional analysis, clarification on your experimental approaches and further discussion regarding your findings and implications for these in the broader plant microbiology field, which are all addressable.

Should further experimental data allow you to address these criticisms, we would be happy to look at a revised manuscript.

Response:

We thank the Editor for the encouraging response to our submission. We would also like to thank the three Reviewers for their comments and suggestions that have helped improving our submission. We have now carefully addressed the editorial suggestions and the comments from the three reviewers in the revised manuscript. We include here results from additional SynCom experiments as well as from additional analyses. Together these support and further inform conclusions derived from the original study. We have revised the main text to make our interpretations clearer and included additional explanations in the Supplementary Information documents. The suggested changes were numerous hence; to keep a good order of the alterations in the revised version and support the review of the revised manuscript, we have also provided an overview document with the changes compared to the original submission. The author list, contributions and acknowledgments have also been changed accordingly to illustrate the contribution to the resubmitted manuscript. We look forward to hearing from you.

Reviewer 1 (Remarks to the Author):

Plant microbiome plays crucial roles for plant fitness, however, mechanistic understanding about the principles governing root microbiome selection and assembly is still in its infancy. This study inoculate different plants with large amount of known isolates from previous studies. They analyzed the enriched functional genes overlap among root microbiome in different plants under different nutrient conditions. Thus successfully characterized both conserved "core" and species specific microbiome functional genes, as well as their distribution patterns among different taxa. The work could be interesting as a large scale microbiome re-construction and functional description study.

Response:

We thank Reviewer 1 for the careful evaluation of the manuscript and for highlighting points that helped us further sharpen the conceptual framing of the study. We have carefully revised the manuscript to address the points raised. In particular, we clarified the central questions and scope of

the work, refined the presentation of the key findings, and revised sections where the rationale or interpretation required greater precision. We also incorporated additional analyses and new SynCom experiments to further strengthen the main conclusions. Below, we provide detailed responses to the major conceptual and methodological points raised by the Reviewer.

However, the findings in this work are still superficial; there is no molecular or SynCom based validation of their correlative analyses; there is no strong and clear key finding and their significance in guiding microbiome engineering is weak.

Response:

We thank Reviewer 1 for the careful evaluation. Our conclusions stem from deep genomic and functional analyses of ~1,000 soil-derived isolates that are tested with large, highly diverse SynComs, revealing reproducible principles of root microbiota assembly. Reviewers 2 and 3 highlight the novelty and conceptual advance of this framework. In the revision, we include results from additional SynCom experiments designed to test and further develop the principles identified in the complex SynComs. These experiments replicate the original findings and strengthen the robustness of our conclusions. A detailed response is provided in the point-by-point comments and revised figures.

1.(A) The manuscript try to study the principles of microbiome assembly in the rhizosphere. However, the current descriptive analyses lacks experimental validations.

Response:

To address this, we designed two independent SynCom studies that test central conclusions of the original work and provide additional experimental support for principles derived from the complex SynComs. We present these in more detail below and include these results in the revised manuscript.

Study 1.

Our study of four unprecedentedly large Syncoms showed that within the five most abundant families, isolates encoding a higher proportion of family-specific functions were more prevalent in root communities than other family members (Figures 3c, S13b, S14). To independently test this finding, we constructed seven new 200-member SynComs by random selection from the 447-member AtSC and inoculated them onto Arabidopsis. Across SynComs, we observed a consistent positive association between proportion of family specific functions and relative abundance of the isolate in root microbiota across multiple families. This was most notably for *Pseudomonadaceae* and *Caulobacteraceae* indicating that, within families, strains with broader KO repertoires reach higher abundance (Figure 3e & S15). These results independently confirm that intrafamilial functional diversity contributes to root colonization success and is now summarized in Figure 3e.

Study2

The original study identified that microbiota of all three hosts consistently enriched a core set of KOs that were distributed across multiple families. We designed Study 2 to i) test contributions of individual families to the recruitment of the 266 KOs and ii) determine if KOs encoded by a specific family have a significant impact on bacterial load. For this, we inoculated the 177-member from LjSC onto Lotus and generated a set of family drop-out SynComs, each lacking one of the major families and their corresponding functions. One Rhizobiaceae isolate remained detectable, so the Rhizobiaceae drop-out could not be interpreted.

We examined redundancy of key pathways delineated by the 266 KOs using a niche-replacement analysis (Fig. S31). This showed overall small changes, indicating that for many pathways derived from the 266 core KOs the remaining families already support these functions (19/40 negative; 17/40 positive; significance of limited loss of core KOs shown in Table S8), consistent with the robust, distributed architecture of core root microbiome pathways. This pattern underscores that these

functions are fundamental and maintained at the community level through redundancy, aligning with our finding that colonisation advantages arise cumulatively across taxa in complex communities. Oxidative phosphorylation in *Pseudomonas* displayed the highest positive niche-replacement score, indicating that *Pseudomonas* is the dominant supplier of these KOs in the intact community and that its loss triggers compensatory enrichment by other taxa, underscoring the community's functional robustness.

Next, we investigated the impact of family drop-out on bacterial load. We observed that this did not alter bacterial-to-plant read ratios (Fig. S32), showing that colonisation is maintained by redundant, community-level core functions that allow the microbiota to preserve colonisation capacity despite taxa perturbations.

Together, these results obtained from additional Syncom experiments, reinforce the findings of our manuscript that functional integrity, rather than taxonomic composition, is the key determinant of root microbiome performance, and taxonomic differences are only informative when they coincide with functional disruption.

The manuscript text reads now:

Lines 348 To further probe this principle, we assembled seven independent, near-randomly selected 200-member SynComs from the AtSC based on high functional diversity principles and inoculated these to *Arabidopsis* (Supplementary methods). We analysed the root-associated communities and observed that, within families, strains with broader KO repertoires consistently achieved higher relative abundance (Kendall correlation across SynComs), with particularly strong support for Pseudomonadaceae and Caulobacteraceae (Fig. 3e, S15, Table S6).

Lines 501 : Consistent with this distributed and redundant organisation, removal of individual bacterial families from Lotus synthetic communities retained a large, statistically significant fraction of the 266 core KOs, albeit with family- and pathway specific differences in contribution and compensation (Figure S31, Table S8). Moreover, the overall bacterial load was not significantly affected by the removal of individual families (Figure S32).

1(B) They have to integrate root metabolomics and essential validations into this study. To test whether root metabolites changes can be matched to the key functional changes in the assembled genomes. This limits the depth and mechanistic importance of this work.

Response:

We appreciate the suggestion which we as well evaluated very carefully during our studies. Our conclusion is that indeed, metabolite profiles can offer valuable correlative insights, and we have shown this previously for Lotus (Tao et al 2025). However, they maintain a descriptive nature and do not directly address the mechanistic questions we focus on in our manuscript namely, how genomic functional capacities shape recruitment and assembly principles in the root microbiome. Our study is designed around causal perturbations at the level of community composition by carefully designed SynComs, which directly test functional contributions. We therefore consider this a valuable direction for future studies but beyond the scope of the current framework.

2.The introduction is not focused on a clear question with a strong logic. The first paragraph of the introduction did not precisely introduce the key scientific question. This work did not address how “microbes are selected from the environment to form these niche-specific communities”, also not related to dysbiosis mechanisms or growth promotion. The manuscript could more carefully propose the key scientific question and novelty in the introduction.

Response:

We rephrased the opening paragraph to explicitly state that identifying principles underlying microbial recruitment into host-associated communities remain critical knowledge gaps and the aim of the study is also clearly stated in the last paragraph of the introduction. We also clarified that applications to dysbiosis treatment and understanding, and growth promotion and crop productivity are long-term motivations rather than objectives of the present study.

3. The authors identified 266 functions enriched across all root microbiomes. Can they test what will happen if they chose strains deficient in those key functions? The authors need to validate at least some of the predicted core functional genes required for rhizosphere colonization, similar to their cited study did <https://www.nature.com/articles/s41588-017-0012-9>.

Response:

We thank the reviewer for this relevant question and appreciate the suggestion of using genetics to confirm the mechanisms presented here. We fully agree that mutant analyses are a powerful approach for establishing isolate-level genetic dependencies. We have carefully evaluated this possibility. However, the 266 KOs identified in our study do not represent isolate-specific determinants of colonisation (see Figures S29a, S30). Instead, as confirmed by the results from Study 2, they constitute an emergent, community-level functional feature that predicts colonisation through their cumulative representation across many taxa and isolates.

Mechanistically, this is evident in two ways. First, isolates enriched for these KOs show increased competitiveness (Fig. 5f). Second, these KOs are redundantly distributed across taxa (Fig. S29a, S30), such that deleting any single KO in a single strain would not be expected to eliminate the community-level effect. Thus, while mutant strains are highly informative for dissecting single-strain phenotypes, they cannot test the emergent property described here, which depends on functional representation at the community scale rather than on isolate-specific gene requirements.

In direct response to the reviewer's question about testing the consequences of losing strains defective in these functions, we conducted a family drop-out experiment in Lotus. The rationale behind is motivated by the finding that functional variation in root communities is encapsulated at the family level. Hence, removing entire families allows us to perturb the collective representation of the enriched KOs. Consistent with our mechanistic interpretation, plants inoculated with the full LjSC or drop-outs do not display statistically significant differences.

We have clarified this mechanistic interpretation in the revised manuscript text, which now reads: Line 505 *"Together, these observations indicate that the predictive value of the core KOs for root colonisation arises from their cumulative representation across taxa and isolates defining an emergent, community level functional feature of root microbiomes, as opposed to the minimal gene set required for colonization in isolation."*

4. (A) What exactly is the key novel finding important for the field from this study?

Response:

Our study advances the plant root microbiota field on both mechanistic as well as ecological fronts. First, we show that isolates carrying a greater subset of familial functions exhibit higher root competitiveness (Fig. 3d, 5f). Second, we identify that ~3% of all analysed functions (266 KOs) provide a colonisation advantage that arises cumulatively at the community level, defining communal, redundantly distributed functions of root microbiome assembly. These insights are enabled by in-depth, complex experiments at unprecedented SynCom scale. As Reviewer 2 noted, our "forward-looking design... makes significant progress towards a mechanistic understanding of host plant-specificity and the shared and unique genetic components of competency across plant microbiome inhabitants and host plant species." This work bridges reductionist SynCom studies (which can miss community-emergent features and are prone to coincidental microbe-microbe interactions driving community assembly) and large comparative genomics (which lack validation in complex

communities), thereby revealing communal principles that also translate into isolate-level competitiveness.

From an ecological perspective, our results reveal a previously unrecognized principle of plant root-microbiota assembly that could only be uncovered using highly diverse, complex, and well-characterised SynComs: ecological performance and competitive hierarchy emerge from family-defined functional repertoires rather than taxonomic relatedness among genera. Moreover, our data reveal that host plants impose distinct ecological filters, exemplified by Lotus, which enriches for root-associated isolates with greater functional breadth than those selected by Arabidopsis or Barley.

4.(B) Can this study help design a SynCom with the potential to enhance plant growth ? (C) Can they design different SynComs with different colonization activities (through quantitative microbiome profiling) to guide rational design of SynComs? (D) Can they manipulate SynCom composition based on functional studies?

Response:

We appreciate the interest in the translational potential of our findings. While the design of plant growth-promoting SynComs is an important long-term goal for the field, our study was not aimed at engineering beneficial consortia but at uncovering the fundamental principles that govern root-microbiota naturally assembled across plant hosts when grown in native agricultural soils and complex microbial communities. Our data suggest that isolates or communities enriched for distinct functions may show enhanced resilience and competitiveness in the root environment, and that targeted community-level perturbations (e.g., family drop-outs) can modulate root microbial assembly and result in different colonization activities (Figure S32). Developing consortia with dedicated application scope lies beyond the scope of the present work, yet it provides a means to achieve this goal in the future.

5.A pretty interesting finding is that none of the 988 isolates encoded the full set of core functions, and most strains encoded fewer than 25# core functions required for rhizosphere fitness. However, it seems that most strains can colonize roots even better in mono-association compared to in natural soil system, due to less competition. Does that mean in real world most strains do not need full set of core functions to colonize rhizosphere? Or in other words, the “full set functions” Identified here are not all needed for colonization?

Response:

This comment raises a conceptual point central to interpreting our study. The 266 core functions we identify do not constitute a minimal, isolate-level gene set required for colonization in mono-association, but instead represent functions consistently enriched under competitive, community-level conditions characteristic of the rhizosphere. Mono-associations disregard the ecological pressures acting in the rhizosphere and therefore reveal only minimal host compatibility, let alone ecological fitness. The 266 core functions are distributed across taxa and act collectively within the assembled microbiota; their predictive value arises from cumulative representation across community members rather than from per-strain genetic requirements. The absence of a single strain encoding the full repertoire is consistent with ecological principles in which community stability emerges through functional complementarity and redundancy. Hence, the observation that individual isolates encode only subsets of these functions, while the community collectively covers the complete set, aligns with this framework.

This distinction is now clarified explicitly in the main text, which reads: *“Together, these observations indicate that the predictive value of the core KOs for root colonisation arises from their cumulative representation across taxa and isolates defining an emergent, community level functional feature of root microbiomes, as opposed to the minimal gene set required for colonization in isolation”*

6. The authors could analyze the Co-occurrence networks of major functional groups, to better illustrate the relationship and potential significance of different gene groups in rhizosphere ecosystem.

Response:

We share the interest in clarifying relationships among functional groups. To address this, we now include the Upset plot of the 266 KOs across the major families as a main figure (Figure 5g), rather than in the supplement (and added Figure S30b showing the distribution of the 266 KOs in the different families). This visualization directly illustrates the distribution of the 266 KOs, their redundancy, and patterns of family-level complementarity.

These KOs were identified as community-enriched features and show largely shared prevalence patterns across taxa. Such patterns can inflate correlation signals and complicate the interpretation of co-occurrence relationships. For this reason, KO co-occurrence network analysis may provide limited additional insight in this context. The Upset plot therefore provides a direct representation of how these functional groups are distributed and shared across the dominant families.

7. The authors only selected three very different plant species for comparing “core functional genes”. There are two potential problems here: 1) three plants are not sufficient for effectively characterize any “core” features for plant kingdom (since inoculating more plant species does not cost too much time as long as they prepared the strains already); 2) Only one genotype in each plant is not sufficient for characterizing species specific “core” stuff.

Response:

We appreciate the Reviewer’s point that broader sampling of plant species and genotypes would help refine the identification of conserved and host-specific functional features. The aim of this study, however, was to establish a mechanistic baseline using three phylogenetically distant model hosts. The core set of KOs was identified to be enriched by all three hosts, and we show it is largely overlapping with the KOs identified by Levy et al., 2018 in plant-associated bacterial genomes (Figure S29b) and is prevalent when comparing to metagenome datasets of plants grown in soil (Table S9). While expanding the analysis to additional species or genotypes would certainly provide further insights into host- or genotype-specific patterns, such experiments fall beyond the scope of the presented study.

8. The authors successfully identified 76 to 146 isolates (averaging 119, RA >0.05%) were consistently associated with the roots across treatments. This could be interesting but lacks further investigation. Why those strains are “core” to different plants, are there any genomics features? What is the implication of significance for manipulating plant microbiome?

Response:

We appreciate the interest in this observation. To clarify, the reported range of 76 to 146 isolates refers to the *average number of strains detected per plant-SynCom combination* (RA > 0.05%), rather than a single shared “core” set of strains present across all comparisons. Across all plant-SynCom combinations analysed in this study, these communities collectively comprise approximately 550 distinct isolates. The consistent recruitment of a similar number of strains per experiment is nevertheless notable and may suggest saturation of ecological niches available on the root. However, because the total pool of strains involved across all experiments is large and heterogeneous, a genome-level analysis of individual “core” strains would be difficult to interpret. For this reason, our analysis focuses instead on the consistently enriched KOs and evaluates their distribution among the strains that cumulatively account for 90% of the community’s relative abundance (Figure S30b). This approach allows us to extract more generalizable functional patterns from the dataset.

9. The authors should test whether knock-out of type VI secretion system negatively regulates the success of Rhizobiaceae colonization?

Response:

The suggestion to perform targeted knockouts of the type VI secretion system highlights an important direction for future mechanistic work. However, the impact of T6SS on colonisation is determined primarily by the specific effector proteins delivered through the system. Because different isolates encode distinct sets of effectors, addressing their mechanism at the community level would require systematic identification and functional testing of effector repertoires across contributing taxa. This represents a substantial undertaking and falls outside the scope of the present study. Moreover, the enriched functions we identify, including T6SS, act as part of a communal, redundantly distributed set of capacities at the community level, making single-strain knockouts insufficient for capturing their ecological contribution within complex microbiota.

10. Line 495-506 raises concern about the novelty of this study. It seems there are a lot overlap with previous study, and there are fewer potential host plants in this study.

Response:

The study by Levy et al. (2018) represents an important computational survey, but its resolution was inherently constrained by reliance on annotated genomes which identified thousands of KOs present in plant-associated bacteria. Such broad signatures are difficult to operationalize for functional or mechanistic interpretation. In contrast, our analyses center on microbes directly recruited by three phylogenetically distant plants grown in defined soils, with findings then tested in large, well-designed reconstitution experiments. The resulting datasets are analysed comprehensively to extract robust, generalisable principles of root microbiota assembly and community-level functional determinants of colonization. This experimental framework allows us to pinpoint, with far greater specificity, a core set of KOs that are consistently and actively recruited by these plants from complex communities. Importantly, our core KO set was identified through an entirely independent and agnostic experimental framework based on plant-recruited isolates and community-level reconstitution, yet it resolves to a functionally relevant subset of the thousands of plant-associated KOs reported by Levy et al., thereby distilling broad genomic signals into a tractable and biologically interpretable group. Altogether, our broad set of findings among which is the core set of KOs, cannot be inferred from comparative genomics alone.

Reviewer 2 (Remarks to the Author):

1. This is a groundbreaking study. In it, Selten et al. perform a truly visionary experiment where they inoculate three different host plants with large root-derived bacterial isolate collections cultured from each separate host. In addition to inoculating each collection separately, they also assemble a super SC composed of all three AND include a nutrient treatment in a fully crossed manner. Replicate metagenomes of each treatment combination were comprehensively characterized using a battery of creative and informative analyses. The forward-looking design makes significant progress towards a mechanistic understanding of host plant-specificity and the shared and unique genetic components of competency across plant microbiome inhabitants and host plant species.

Importantly, the authors are very upfront about the limitations of their approach throughout the text, however, the scale and depth of the study are indeed unprecedented. Thus, this paper will make a great impact on the field.

Response:

We thank Reviewer 2 for the appreciative and encouraging assessment of our work. We are pleased that the scale, design, and mechanistic insights of the study resonated as intended, and we appreciate the Reviewer's recognition of both the ambition and the transparency of our approach. We have carefully revised the manuscript to address the points raised. In particular, we clarified several aspects of the interpretation, improved wording and structure where appropriate, and incorporated additional analyses and figures to further strengthen the presentation of the results. Below, we provide detailed responses to the points raised by the Reviewer.

2. My only major comment is that the results are very much interpreted through a lens of active recruitment and selection on the part of plant hosts. I agree that the host plants demonstrate differences in the taxonomic and functional composition of the isolates occurring in their roots across all SCs, but why must this be the result of 'active' recruitment? Furthermore, what does that even mean? It seems analogous to saying that alpine plants are actively recruited to montane habitats by virtue of their benefit to the ecosystem rather than a more simple explanation that they possess attributes that allow them to persist and reproduce in those environments to a greater extent than other species. Can a similar explanation not be used for the microbial taxa enriched in the microbiome of plants and particular plant species?

Response:

We appreciate this thoughtful comment and agree that the sensed "active recruitment" in our manuscript text can unintentionally suggest intentionality, which is not what our data demonstrate. Our results are fully consistent with established ecological concepts of host-mediated filtering, in which plant physiological traits and root-associated environments shape the microbiome assembly. We have therefore revised the text to and when appropriate we used terms such as "host-driven filtering," "host-mediated structuring," or "host-imposed selection pressures," which more accurately capture the ecological processes supported by our experiments. These revisions preserve the core interpretation that hosts differ in the functional and taxonomic profiles they favour while avoiding language that implies intentional selection.

Minor comments:

3. Line 142: How do we interpret this proportion of shared KOs among the three isolate collections? What might serve as a good reference is KO profiles generated from different soil metagenomes, ideally from the same source as the isolate collections, so that one can place this shared proportion among root bacterial endophytes in the context of soil bacterial KO variability.

Response:

We appreciate this point and agree that contextualizing KO overlap with soil metagenomes is valuable. Although matched soils are available only for the Arabidopsis isolate collection (Stringlis et al., 2018), we addressed the reviewer's question by comparing this dataset with the published rhizosphere and soil metagenomes from Ofek-Lalzar et al., 2014 enabling us to compare multiple plant-soil systems. Using the data structures compiled in Fourie et al. (2024) and Lopez et al. (2023), we extracted all annotated KOs from Arabidopsis rhizosphere and Reijerscamp soil (Stringlis et al., 2018; Sánchez-Gil et al., BioProject 1PRJNA108915), and from cucumber and wheat rhizospheres and Maon soil (Ofek-Lalzar et al., 2014). A comparison of KO overlap between rhizosphere and soil communities revealed that the two soils share 6,713 of 11,161 KOs (60.1%), while the rhizospheres share 6,896 of 10,466 KOs (65.9%), a significant enrichment based on a binomial exact test (P -value $< 2.2e-16$). These results indicate that root metagenomes show greater functional overlap than soil metagenomes, providing a meaningful reference point for interpreting the shared KO proportions among our isolate collections. We included this analysis in the manuscript (Supplementary Table S9) and a sentence in the text that reads now: "To assess whether this pattern is recapitulated in published datasets (Table S9), we compared KO profiles across rhizosphere and soil metagenomes and found that rhizospheres share a larger fraction of KOs than their corresponding soils."

4. Line 217: Would there be any value in attempting to incorporate an abundance weighting into the characterization of functional diversity? Just as the authors used Bray-Curtis dissimilarity, which is calculating differences in community composition based on taxon presence AND abundance, is there a way to capture the abundance of different KO in order to provide a more refined calculation of functional variation.

Response:

We included in Figure S7 the information on the KO abundances. Supplementary Fig. S7c–d, depict the Shannon diversity reflecting both KO richness and their relative abundances. When interpreted together with the observed KO counts in Fig. 2d–e, these metrics provide the abundance-weighted view of functional diversity we believe the reviewer is seeking. We refer now to this supplementary figure in the text context mentioned by the Reviewer.

5. Line 215 and 281: Perhaps contrast this to: <https://doi.org/10.1038/s41396-020-0695-2>

Response:

We revisited Schneijderberg et al. (2020) with the reviewer's comment in mind and examined whether their results could be placed directly alongside ours. Because their work is based on natural OTU-level rhizosphere communities which display a much higher alpha diversity than our isolate-based synthetic systems, a quantitative comparison is not straightforward. That said, one qualitative pattern is strikingly parallel: Schneijderberg et al. report that Arabidopsis hosts a higher bacterial alpha diversity than *Lotus corniculatus*, which aligns well with the Arabidopsis-*Lotus japonicus* contrast we observe. Although their study did not include functional diversity analyses, their beta-diversity patterns similarly show pronounced host-specific community differences, providing a useful conceptual complement to our findings. We now include this reference when interpreting the results on L226-230 of the revised manuscript text.

6. Line 299: What about actually dropping it out of the SC? I imagine such a dominant colonizer is likely having an effect on community assembly that will not be observed by in silico removal.

Response:

We see the interest in testing how assembly changes if a dominant strain is removed. However, such a strain-specific perturbation shifts the focus toward isolate biology rather than the community-assembly principles. Our framework examines how functions and families interact within complex communities; notably, when this isolate is embedded in the original HvSC or the full SSC. The relative abundance of P2_G4 drops in the SSC, showing that competitive outcomes and niche occupation are determined by community-level interactions. Accordingly, we analyse the communities with and

without the dominant isolates to uncover patterns of the remaining members, whose contributions would otherwise be obscured by the strong competitive signatures of the dominating isolates.

7. Line 305: Selective host recruitment implies that the host is actively participating in this process. However, inoculum dependent enrichment contradicts this statement.

Response:

We appreciate this thoughtful comment and agree that the sensed “active recruitment” in our manuscript text can unintentionally suggest intentionality, which is not what our data demonstrate. As mentioned in Comment 2 we have adjusted the text accordingly.

8. Line 317: Can this be illustrated with a figure e.g. Bray-Curtis dissimilarity versus KO dissimilarity?

Response:

This is a valid point made by the Reviewer. Therefore, we refer to Figure 2i in addition to Figures 2h and S11. Figure 2i shows that functional dissimilarity among KOs is substantially lower and more constrained than the taxonomic dissimilarity observed at the genus level.

9. Figure 2b,c: Can the authors explain why the observed strains in the input does not match the number of strains included in the inoculum?

Response:

The discrepancy between the number of strains inoculated and those detected in the input likely reflects technical rather than biological factors. Equalising hundreds of isolates by OD₆₀₀ is inherently imprecise, leading to uneven starting abundances; and biases during DNA extraction, library preparation, and sequencing depth can cause low-abundance strains to fall below the detection threshold. We included this explanation in the Material and Methods section.

10. Figure 2f,g: I think the grey versus black outlines refer to the nutrient status but this may be worth including in the legend somewhere more clearly.

Response:

We thank Reviewer for raising this comment. We include this in the revised manuscript and figure legend.

11. Figure 2h: I’m surprised that the nutrient treatment yielded such a weak effect on the taxonomic and functional root microbial composition? Fig S3 reveals quite a large nutrient effect on the biomass of Hordeum and Lotus but little on Arabidopsis.

Could the authors speculate why a biologically relevant nutrient stress (at least for 2 of the 3 hosts) is having little effect on the identity of the enriched KOs? Are there particular environmental conditions where this might not be the case? I see the paragraph addressing this in the discussion (lines 585-608), but I’m trying to place this result in the context of disease states or other abiotic stress, where plants are often observed to recruit a unique microbiota.

Response:

We agree with the Reviewer that the limited effect of nutrient availability on taxonomic and functional composition is initially surprising, especially given the biomass differences observed for Barley and Lotus. A closer look, however, indicates that nutrient regime does contribute to community variation: the PERMANOVA for nutrients is significant for both taxonomic and functional profiles (Fig. 2h), although the effect size is modest ($R^2 = 0.01-0.02$). These values are small relative to the much stronger structuring effects of inoculum identity and host species, but they remain detectable in comparison to the residual fraction. Similar impact was previously reported in maize (Zhang et al 2022) and soybean (Wang et al 2024). To improve clarity, we revised the text of Supplementary Results SR1 to better situate our findings within the broader literature on nutrient stress and host-mediated microbiome restructuring.

12. Figure 3a: I think the authors need to provide more operational and biological explanation of these values. If I understand correctly, the authors are decomposing the KO functional variation. The inner ring represents the R² value output from Adonis for the host species term for each inoculum. The outer ring represents the contribution of each host on to that effect represented by the species term, the mathematical formula underlying this provided in the supplemental text. If I have this wrong, please clarify in text. If I have it right, great! I still need a bit of help interpreting it. When the authors report that Lotus has a contribution of 82.1% of the functional variation observed with the SSC inoculum, I interpret this as Lotus assembling a root microbiome that occupies a more unique multivariate functional space versus Hordeum or Arabidopsis when exposed to the same inoculum. Is this correct?

Response:

We appreciate the Reviewer's careful reading of Fig. 3a, and we recognise that this figure condenses several layers of analysis and therefore rewards a careful read. We have tried to guide the reader as clearly as possible, and we believe the figure effectively illustrates how the two major sources of variation; the SynCom and host species, jointly contribute to the Host R². Your interpretation on this matter is correct. To assist readers, we revised the figure legend to highlight that the figure depicts the components of the Host R².

13. Line 325: I need more help interpreting this result and Figure 3b. Why is the intersection of the explanatory power of the Adonis terms when analyzing either taxonomic or functional composition important (i.e. the selection of Family from Figure 3b)?

Response:

Figure 3b (Figure 3c in the revised manuscript) contrasts the explanatory power based on Adonis R² of the same factor (host or inoculum) for taxonomic profiles (solid line, evaluated across ranks) and for functional profiles (dashed line, KO-based). The intersection of the two curves is the key: it marks the taxonomic rank at which taxonomic composition best matches functional composition, or the scale at which lineage identity most closely encapsulates functional differences. In our data, that convergence occurs at the family level; this is why Family is highlighted throughout as the most informative unit for interpreting host/inoculum effects on function. We added the following explanatory sentence to the revised manuscript in the figure legend: *The intersection of solid and dashed curves indicates the taxonomic rank at which taxonomic and functional variation coincide.*

14. Line 340: I'm not sure I'm convinced by the data to support this claim. There are several instances in Fig S14 where the most abundant isolate is not harbouring the highest proportion of KO diversity at the genus level. Is there a randomization test to perform here? I doubt r values would be significant for anything in Fig S13 or S14.

Response:

To address this point, we carried out a permutation test to assess whether the relationship between isolate abundance and KO diversity could arise by chance. We permuted family labels within each SynCom-family pair 1,000 times and recalculated the log-correlation statistics. During this analysis, we recognised that a log-linear transformation is the appropriate way to model the relationship between abundance and KO counts, and therefore re-computed the correlations using this approach. We compared the empirical p-values from Fig. 3c with the 0.025 confidence interval of the permuted distributions. In 11 out of 17 significant cases, the observed p-values were considerably lower than expected under randomization, indicating that these associations are unlikely to be artefacts. The results of the second analysis are included in the revised manuscript as Supplementary Table S5.

15. Lines 402-408: What an interesting result. Is this surprising to the authors given the high host-specificity reported elsewhere in the study? Does this suggest other genetic factors are at play during

host selection, e.g. interactions with host-immunity? The following published article show a similar finding with a single bacterial strain: <https://doi.org/10.1038/s41467-025-61266-3>

Response:

We thank Reviewer 2 for the appreciative comments on these results and for raising the question regarding the level of surprise. The pattern we observe is consistent with previous studies of root microbiota assembly, where host specificity has primarily been characterised at the taxonomic level. What our study adds is a functional dimension: even when plants recruit from highly complex communities, certain KOs are consistently enriched in a given host and across SynComs. This suggests that strains originating from other hosts can carry host-associated functions that become selectively favoured when they encounter another suitable plant, indicating that many soil-borne bacteria maintain a versatile functional repertoire that enables interactions with evolutionarily distant hosts. Regarding the mechanisms involved, a contribution of host immunity is plausible. One *Arabidopsis*-associated candidate, *ipdC*, could interact with indole-glucosinolate turnover in Brassicaceae. More broadly, our findings are in line with those reported by Russ et al. (2025) and support the view that host specificity can arise independently of community context, likely influenced by a combination of immune-related processes and host-derived metabolites. We incorporated these points into the revised manuscript text and Supplementary Results 2 and have integrated the literature suggested by the Reviewer.

16. Lines 484-493: Another fascinating result. Perhaps beyond the scope of the current paper but I'm very curious to know whether this represents a shared feature among all members of these families, including those not plant-associated? Additionally, this result yields very clear and distinct ecological hypotheses regarding the competition among members within these families occurring in the plant niche.

Response:

We appreciate Reviewer 2's insightful reflections on these results. They indeed raise compelling questions about whether the observed patterns represent a shared feature across all members of these families, including non-plant-associated lineages, and point toward clear ecological hypotheses regarding niche specialisation and intra-family competition within the plant environment. While highly interesting, these directions fall beyond the scope of the current study and would require dedicated follow-up work.

We also note that our observations concerning the *Burkholderiaceae*, which is the largest family represented in our dataset, suggest considerable functional and ecological heterogeneity. This may indicate that the current taxonomic boundaries of the family warrant re-evaluation and reopening the discussion on the need for subdivision into more coherent groups.

17. Line 502: A chi-squared test or Fishers exact test might help us interpret the significance of these different fractions.

Response:

We thank the Reviewer for suggesting this. We performed a binomial exact test, indicating this overlap in plant-associated KOs is indeed highly significant with $p = 2.5 \times 10^{-13}$. We have added this information to the manuscript text line 519: "*Importantly, plant-associated bacterial genomes were found to encode between 40-60% of them, whereas non-plant-associated and soil bacterial genomes only have 10-20%, supporting their signature of plant-adapted lifestyle in the bacterial genomes (Binomial exact test, $p=2.5.10^{-13}$).*"

18. Line 504: This is a remarkable result given that none of the genomes in Levy et al. are present in the current study, correct?

Response:

We appreciate the Reviewer's observation, and yes this is correct: none of the genomes analysed by Levy et al. are included in our study, as our isolates were not deposited in NCBI at the time of their analysis.

19. Line 566: I agree that there isn't much in the plant microbiome literature to support this claim, possibly <https://doi.org/10.1128/mbio.02892-21>

Response:

We thank Reviewer 2 for highlighting this paper. Indeed, this could be a comparable paper in the plant microbiota field, and we thank Reviewer 2 for the heads up on this. We included this note in the revised manuscript text, line 591.

Reviewer 3 (Remarks to the Author): Davide Bulgarelli

Gijs and co-workers presented a compelling story on the deconstructing and reconstructing the microbial communities thriving in association with plant roots, designated root microbiota. In analogy with the communities inhabiting the digestive tract of the vertebrates, the root microbiota impacts on the growth, development and health of their hosts. Since the 2010s, access to high-throughput sequencing technologies enabled scientists to gather insights into the molecular basis of plant-microbiota interactions. Where many of these studies fall short is in providing an understanding beyond taxonomy [i.e., what for?] and comparing host recruitment strategies [i.e. are all plants doing the same?]. This submission aims at filling these knowledge gaps and provide the scientific communities with a blueprint for follow-up engineering strategies.

Therefore, in my vision, this submission is novel and holds a potential for impact, both in basic science and translational applications.

Furthermore, the research approach is ingenious and will likely benefit scholars and professionals beyond the remit plant sciences.

Response:

We thank Davide Bulgarelli (Reviewer 3) for the thoughtful assessment of the work and the appreciative recognition of its conceptual contribution and impact.

We have carefully revised the manuscript to address the points raised. We have clarified terminology, improved phrasing and structure, and corrected or simplified statements to better explain the rationale and more accurately reflect the interpretations of the results.

To improve readability and accessibility, we have streamlined several sections of the manuscript, clarified definitions, and reduced potential ambiguities (e.g., regarding host effects, functional interpretation, and terminology). We have also incorporated additional analyses and figures where relevant (e.g., BUSCO completeness, within-soil comparisons) and improved the description of datasets and experimental design.

Below, we provide detailed responses to the main conceptual and methodological points raised by the Reviewer 3.

However, there are a number of elements of this submission that are unclear to me:

1) the ecological links with the soil data are tenuous: while the authors provide evidence of strains recovery, what is missing is a like-with-like comparison of microbiota assembly in soil and under the conditions used by the authors. Results will inform on the "efficiency of translation" of the proposed work, or limitations between the two approaches - either way an informative result.

Response:

We thank the Reviewer for this thoughtful suggestion. A direct like-for-like comparison would indeed be informative, but it presents substantial technical and methodological challenges knowing the experimental system used here to underpin host-dependent microbiota assembly. SynCom members cannot be quantitatively resolved against the high diversity of native soil communities.

To evaluate ecological relevance of our findings we analysed an independent soil-rhizosphere metagenome dataset from López et al. 2023 We found that 121 of the 266 KOs identified here are significantly enriched in natural rhizosphere communities. Importantly, this is a significant enrichment compared to all significant KOs detected in that dataset (hypergeometric test $p = 2.3 \times 10^{-6}$; Fisher's exact test $p = 3.6 \times 10^{-6}$). This analysis shows that functional signals identified in our complex SynCom studies are also preferentially associated with the rhizosphere in natural environments. This is included in the revised manuscript from line 524: *"We found that 121 of the 266 KOs identified in our experimental framework were significantly enriched relative to the total pool of rhizosphere-enriched KOs detected in the natural community dataset (hypergeometric test $p = 2.3 \times 10^{-6}$; Fisher's exact test $p = 3.6 \times 10^{-6}$) (Table S9)."*

Furthermore, we addressed the potential confounding between host lineage and soil origin by performing a within-soil comparison. Specifically, we compared Lotus-derived isolates used in this study with Arabidopsis-derived isolates from Bai et al. (2015), all derived from Cologne soil. This confirmed that the functional breadth of Lotus-enriched isolates when soil origin is held constant (Figure S24). Together, these analyses provide additional support that the patterns we observe can be extrapolated to soil conditions and inform translation of the acquired knowledge into practise.

2) There is no closing model: do I interpret correctly that the majority of the 266 KOs are undefined (Fig S27)? In other words, you've got a number but we still do not know why plants want them? Resolving this knowledge gap will greatly strengthen the significance of your work.

Response:

We see the value of providing an overview of the pathways delineated by these KOs and therefore include Figures 5f-g and Supplementary Figure S28, an extended version of Figure 5f, summarising the information provided in the manuscript text. In short, the KOs map to carbohydrate degradation (plant-derived sugars), broad nutrient transport (ABC systems), respiratory energy metabolism with flexible electron donors (including sulphur compounds), xenobiotic/aromatic compound degradation, and the biosynthesis of cofactors, exopolysaccharides, and secondary metabolites. A subset of these KOs is not resolved at the pathway level (i.e., annotated at the KO level but lacking specific functional placement in current databases). We have intentionally avoided speculating on host “needs” for these, as the functions are communal and emergent rather than attributable to a single lineage or host-specific dependency and represent a shared metabolic backbone.

3) Dataset S8 requires a clarification[see also comments in the main text and supplementary] as you have sample types sequenced not mentioned in the main text - what was their fate? Likewise, there is an average on 30% unmapped reads [including in the inoculum..] that challenges the whole concept of gnotobiotic system and, by extension, your discoveries. Likewise, there are controls that are mentioned but not discussed in the main text. Those. e.g., unplanted control, mycorrhized specimens..., would be instrumental in strengthening your conclusions.

Response:

We appreciate the Reviewer’s careful examination of the complete dataset and the opportunity to clarify these points.

First, regarding the concern about the proportion of unmapped reads and its implications for the gnotobiotic system, we thank the Reviewer for drawing attention to this issue. Upon re-examination of Data S8, we identified an error in the original export of values to Excel. After correction, the proportion of unmapped reads in root-associated samples averages ~3.8%, which is well within the expected range for metagenomic analyses of defined synthetic communities and supports the validity of the gnotobiotic system used in this study. The corrected values are provided in the updated Data S8.

With respect to the sample types included in Dataset S8 but not discussed in detail in the main text, the reconstitution experiments were explicitly designed to identify host-driven principles of root colonisation by monitoring bacterial dependency on host-derived metabolites. For this purpose, the rhizoplane and endosphere provide the most informative and mechanistically interpretable readout, as they capture bacteria in direct physical contact with the root and exposed to host-derived signals. In contrast, under the inert substrate conditions used in the reconstitution experiments, the rhizosphere cannot be reproducibly delineated. The substrate adheres poorly to roots compared to soil and does not support robust bacterial amplification, resulting in high variability and limited compartment-specific insight, as reflected in the low number of successfully sequenced rhizosphere samples in the first replicate experiment. Rarefaction curves for rhizosphere and unplanted control samples are provided here (Figure A and B) to allow full evaluation of these data.

Accordingly, to maintain a clear and focused scientific narrative, we limited the main body of our analyses to root-associated communities, where the data were consistently recovered and directly informative with respect to host-mediated assembly processes (Figure S5). Nevertheless, we agree with the Reviewer that the additional control samples may have value, and therefore all datasets are retained in the Supplementary Information and deposited in full under BioProject PRJNA1131994.

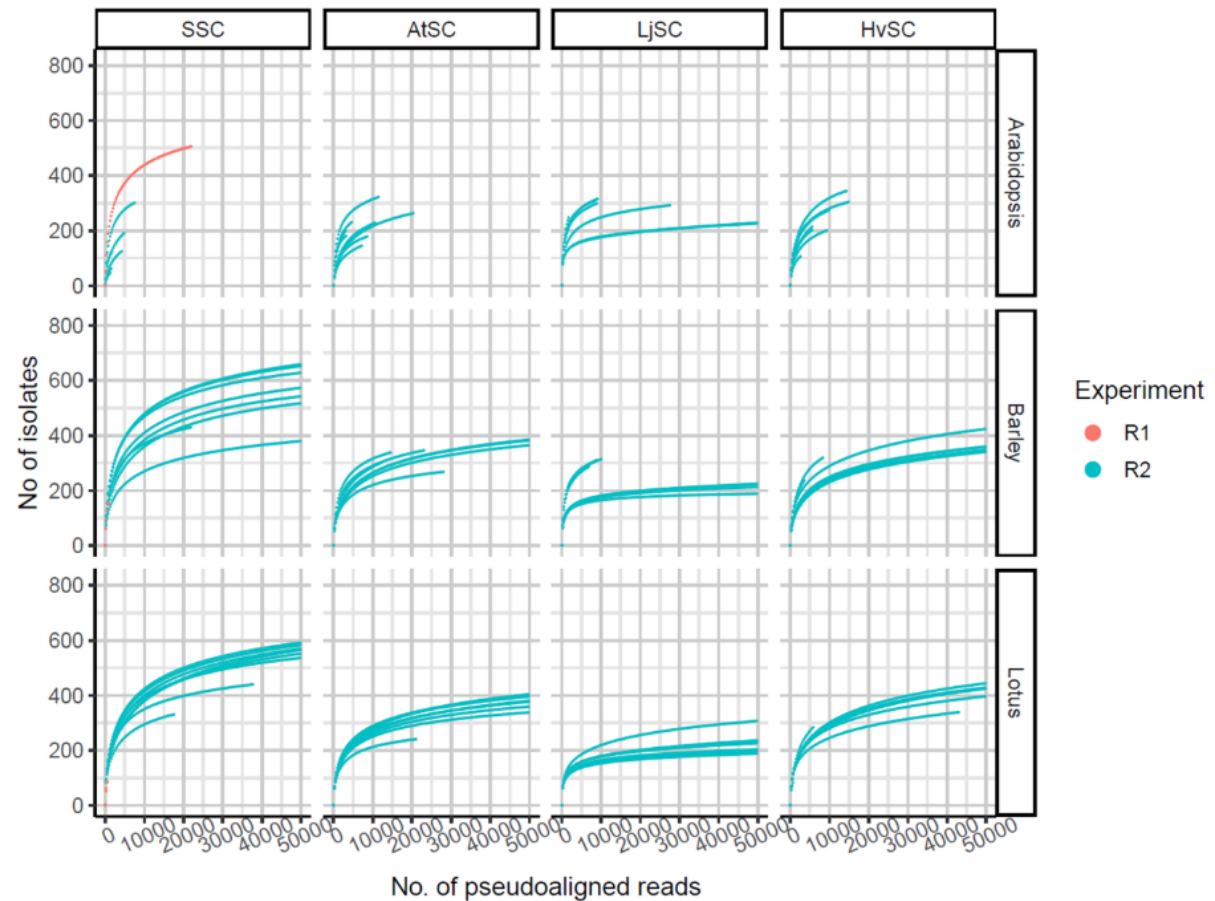


Figure A. Rarefaction curves of rhizosphere samples in the first (R1) and second (R2) replicate of the reconstitution experiment.

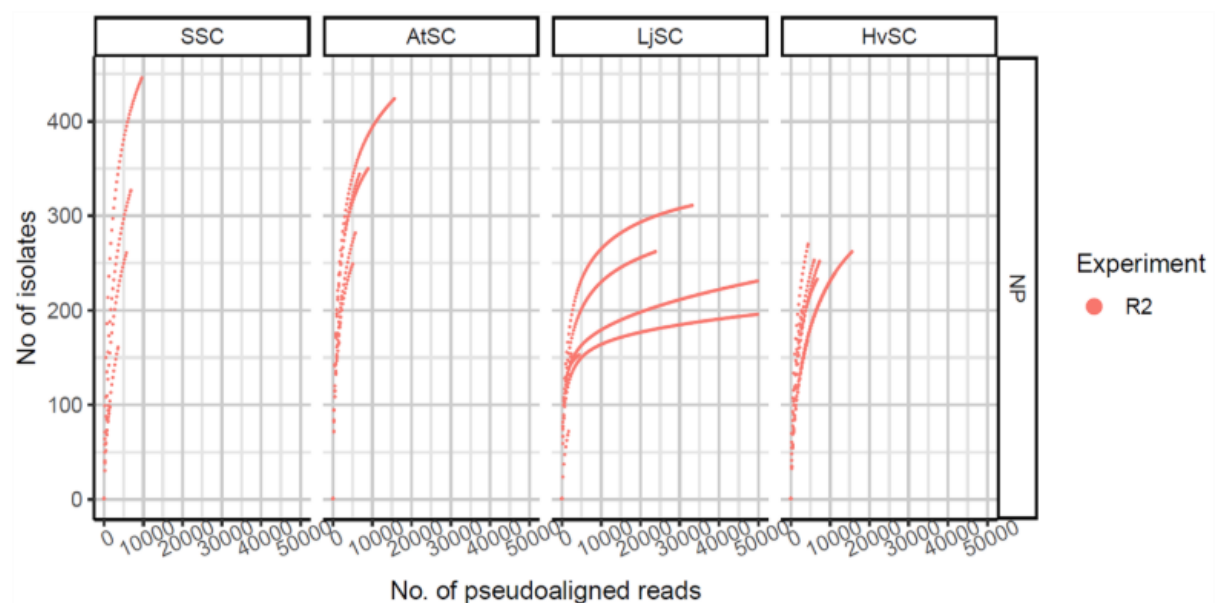


Figure B. Rarefaction curves of unplanted samples (SynCom inoculated substrate) in the second (R2) replicate of the reconstitution experiment.

4) I'd soften some of the conclusions for which you do not have experimental support [other than in silico analysis]

Response:

We appreciate the Reviewer's observation. Where appropriate we have adjusted the conclusions supported only by in silico data so that they reflect the appropriate degree of inference and remain fully aligned with the analytical evidence.

Remarks on code availability:

Checked the codes on line, not reviewed though - at first glance everything is there.

However, instead of a unique R script I suggest creating different scripts for figures [or block of figures]. This will streamline data checking and reproducibility.

Response:

We revised the entire R script and formatted as indicated.

Extracted comments from Davide Bulgarelli's PDF.

We thank Prof. Davide Bulgarelli for a detailed evaluation of the manuscript. We provide below responses to the comments provided in the pdf file that raise conceptual and methodological points. They are organised by page in the original manuscript. The remaining comments related to style or wording were carefully evaluated during revision of the manuscript text or were addressed in the other comments.

Page 2

3....or cellulose degradation...

Response:

We do not find KOs associated with cellulose degradation per se, but we provide an overview of the KOs predicted functions in the text, and these represent metabolic activities required for adaptation to plant-root niche.

Page 3

8. What do you mean by that?

Response:

By "nutritional feedback loops" we refer to the reciprocal exchange of nutrients: microbes increase the availability of mineral nutrients to the plant, and the plant supplies carbon-rich exudates that fuel microbial growth. This two-way nutrient flow is what we mean by a feedback loop.

Page 4

16. Though the readership will lean towards the microbial side of the story, I'd spend a few words motivating your host selection. i.e., why those three and not e.g. coconut or spruce...

Response:

These three hosts were chosen on account of being phylogenetically distinct, represent important model plants for crops (brassicacae, cereals, and legumes), are studied extensively in literature, and have differential capacities to engage in plant-microbe relationships. We provide additional details to a similar comment from Reviewer 1 (see comment 7 of Reviewer 1)

24. The origin of the collection confounds with the host lineage - how can you exclude that, instead of a host effect, you are simply describing a soil effect? I mean if all three hosts were grown in let's say Cologne, they could have all displayed a "Swiss army knife approach"....

Response:

We thank Reviewer 3 for highlighting the potential confounding between host lineage and soil origin. To address this, we performed a within-soil comparison using isolates derived from the same Cologne soil: Lotus-derived isolates used in this study versus Arabidopsis-derived isolates from Bai et al. (2015). This analysis shows that the "Swiss-army knife" functional breadth is more pronounced in the Lotus-derived isolates relative to the Arabidopsis-derived ones when soil is held constant. We have added this result and figure to Supplementary Figure 24.

Page 5

25. What do you mean by that?

Response:

We mean all isolates from all collections, which is, in essence, all isolates present in the SuperSynCom (SSC = AtSC+HvSC+LjSC). The text was revised.

26) have you run a **BUSCO** completeness of your isolates? In other words, are your data describing a functional overlap or a limitation of the sequencing/annotation approaches?

Response:

To address this, we complemented the CheckM analysis already included in the manuscript with an additional assessment using BUSCO. This analysis shows that the vast majority of genomes (887 strains) exceed 90% completeness, indicating that the observed KO overlap is unlikely to be driven by sequencing or assembly limitations. We have added the full BUSCO results to the Supplementary Information (Supplementary Data S3). We note, however, that annotation biases remain possible, as KEGG analyses inevitably exclude unannotated or poorly characterized functions, which may contribute additional diversity not captured here. This point is made in the manuscript text: *"This overlap supports the hypothesis that plant hosts recruit microbes based on functional traits, while keeping in mind the inherent limitations of KEGG orthology relying on existing biological information as well as functional constraints of soil microbial communities."*

27. I am confused: you just told me that 80% of their functional potential is shared, I would call them the opposite...

Response:

To clarify the point raised: the functional overlap applies across the three collections, but each culture collection itself contains high internal taxonomic and functional diversity

Page 8

42. I'll flip the coin - can this underpin the broad geographic adaptation of barley?

Response:

That is one possible explanation, although Arabidopsis also has a broad native range. As discussed in Supplementary Results 1, an alternative hypothesis relates to domestication: barley's long history of selection for specific agronomic traits may have reduced some host-driven selective pressures, increasing the relative influence of microbe-microbe interactions on its root microbiota.

Page 9

43. this is indeed an intriguing observation! You have an isolate that goes from 75% to less than 10% RA in an inoculum-dependent fashion [Fig. S9] while the overall number of KOs [Fig. 2] does not change. Could it be evidence of the onset of a dysbiosis? After all rhizobacter has been reported as a plant pathogenic bacterium. Have you considered a leave-one-out validation to infer the impact of P2_G4 on plant's performance/community dynamics?

Response:

We appreciate the thoughts around P2_G4. We do not interpret the shift in P2_G4 abundance as dysbiosis. As discussed in the literature, including Hooks et al. (2017), the term dysbiosis is often used inconsistently and should be applied with caution, particularly when changes in community composition are not accompanied by evidence of host dysfunction or a clearly abnormal host-associated state. In our experimental system, we do not observe host impairment (see figures S2 & S3- plant phenotypes), and we therefore interpret the marked reduction of P2_G4 from ~75% to <10% in the larger SSC as context-dependent community assembly, reflecting competitive interactions that depend on inoculum composition rather than dysbiosis. This indicates that P2_G4 abundance is largely dependent on microbe-microbe interactions. We have clarified this point in the manuscript text (see lines 305-314). A “leave-one-out” strategy for a dominant isolate would test isolate biology rather than the community-assembly principles we investigate here. Instead, our analyses compare communities with and without dominant isolates *in silico*, revealing how higher-order interactions shape assembly and behaviour of isolates which otherwise would be overshadowed in the analysis by the dominant isolates.

46. I am confused: figure S12 reports a R2 of .92 when correlating taxonomies with function - I argue that taxonomic diversity mirrors functional diversity [though the latter may not be predicted straightway from taxonomies]

Response:

Figure S12 shows that in simulated communities, taxonomic and functional diversity are tightly coupled ($R^2=0.92$). By contrast, the empirical inocula systematically deviate from this line: despite taxonomic differences, root microbiomes converge functionally. This indicates host-driven selection on functions that cut across lineages, i.e., functional convergence beyond genus identity. Accordingly, genus composition is not a sufficient proxy for functional assembly in our system.

Page 10

48. how did you conclude that? I mean your experimentation is *in-silico*: so these KOs may provide...

Response:

We take on Reviewer’s comment and revised the text with caution, as these KOs remain candidate explanations rather than demonstrated mechanisms. We identify four KOs (three in Lotus, one in Barley- see Figure 4b) that are strongly associated with Rhizobiaceae and contribute significantly to the KO composition of the respective root microbiomes. When we subset the data to Rhizobiaceae only, these KOs show significantly higher abundance in the hosts where they are enriched. This enrichment disappears when analysing all non-Rhizobiaceae isolates, indicating that these KOs are specifically linked to Rhizobiaceae and may confer a competitive advantage within this family. The text in the revised manuscript reads: *Importantly, these KOs are associated with greater competitive performance when encoded by Rhizobiaceae than when present in other families, consistent with a family-specific effect (Fig. 4c).*

Page 12

55. Very interesting observation; However, I'd argue that bacteria may not be the only compatible symbiont. Looking at Fig. S3, I'd argue that nodule-derived nitrogen isn't the rate-limiting factor for plant growth: have you looked at mycorrhization in your specimens [e.g., checking on sequencing data for evidence of extra-radical mycelium...] - as lotus is certainly more prone to this type of interaction than At and Barley

Response:

The highlighted text refers specifically to compatible bacterial symbionts, such as those present in the HvSC. Arbuscular mycorrhizal fungi (AMF) were included as a common fungal component across all conditions and were detected in all root microbiota samples. Plants were grown for three weeks in this system, a time frame that, based on established knowledge in the field and our previous work,

corresponds to an early stage of AMF root colonisation, prior to substantial metabolite exchange or support of extraradical hyphal growth. This differs from the root nodule symbiosis, which is fully established at this time point. Nevertheless, AMF abundance was quantified and confirmed across conditions. As expected, given the biology of the interaction, AMF levels were reduced under high nutrient conditions. Figure C presented here illustrates this.

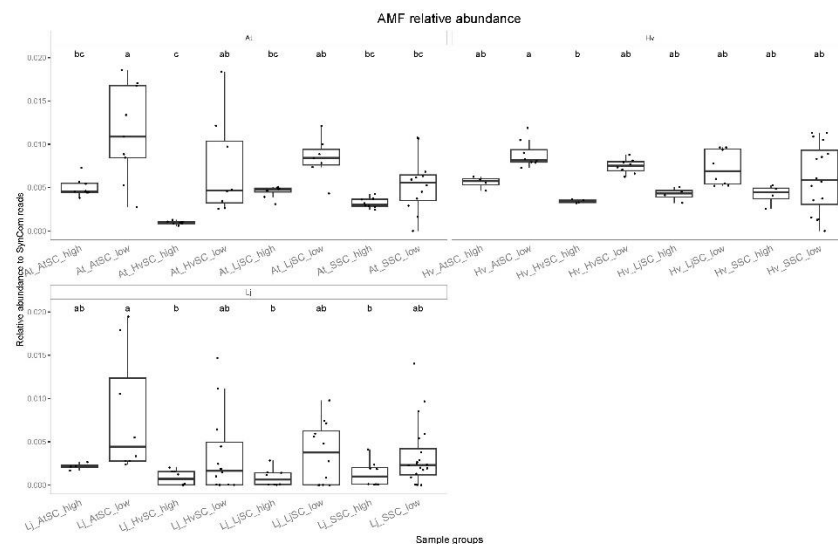


Figure C. AMF reads abundance in the root samples across hosts, conditions and SynComs.

56. are you sure about this claim? According to Fig. S9 HvSC communities are comprised by up to 50%/75% by just one strain- I'd argue that is a very selective environment...

Response:

The sentence in question pertains to host-driven functional selectivity across SynComs, showing that Lotus recruits functions more consistently than Arabidopsis and Barley. See also Comment 43.

Page 14

59. it does not seem to be the case when looking at Fig S12 or?

Response:

We thank Reviewer 3 for highlighting the need for this important clarification of the statement here and on the page 15. The text here on page 14 refers to the general principle emerging from our data: hosts select microbiomes more strongly by functional properties than by taxonomic identity. Fig. S12 shows that, when different SynComs are inoculated, root communities converge more at the functional level than at the taxonomic level, and the same relationship is observed for differences between hosts. Thus, the statement is directly supported by the finding that functional composition is more consistently structured than taxonomic composition across both inoculum and host variables.

Page 15

60. again - isn't this a contradiction of the statement function over taxonomy?

Response:

Our data (Fig. 2g-h) show that hosts preferentially recruit isolates from complex communities based on their functional attributes. Notably, this functional signal is structured at the family level (Fig. 3c) indicating that the observed clustering reflects a phylogenetically, non-random distribution of these capacities. This does not imply that families themselves constitute the unit of selection, but that the relevant traits display evolutionary coherence across related lineages, and their enrichment through host filtering generates a statistically detectable pattern at the family rank. We revised the text that reads now: *"The resulting patterns reveal guiding principles for the assembly of bacterial functions in root microbiota (Figs. 2–4). Across hosts and inocula, functional variation within root communities is*

structured at the family level (Fig. 3c), consistent with a phylogenetically non-random distribution of the relevant capacities and their enrichment under host filtering.”

61. you haven't tested that though - this is inferred in silico, right?

Response:

Our conclusions regarding the differential recruitment of strains emerge from detailed computational analysis of root communities derived from two independent experiments using the SSC. As shown in Figure S6, the input inocula of the SSC experiments contain a balanced representation of AtSC-, HvSC-, and LjSC-derived strains. In contrast, the root microbiome outputs of the SSC consistently display a marked enrichment of LjSC-derived strains, whereas AtSC and HvSC lineages remain comparatively underrepresented. This pattern is further reflected in Figure 2f, where root microbiomes inoculated with LjSC and those inoculated with the full SSC exhibit high taxonomic similarity, underscoring the reproducibility of this recruitment outcome across experiments.

Page 15

64. you haven't sequenced unplanted controls though - at least in part it could be a technical artifact

Response:

See Response to Main Comment 3

65. this could have been tested by mapping metagenomic reads obtained from soil-grown plants against your genomes

Response:

We agree that linking our findings to natural soil systems is important, and we addressed this in the revised manuscript by comparing our results with an independent soil-rhizosphere metagenomic dataset from López et al. This analysis allowed us to assess the extent to which the functional signals identified in our analyses of the Culture collections and in our SynCom reconstitution experiments are also observed under soil-grown conditions. This is outlined in our response to Comment 1.

Editorial Comments:

Thank you for submitting your revised manuscript "Functional capacities drive recruitment of bacteria into plant root microbiota" (NMICROBIOL-25093246A). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Microbiology, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

Response:

We thank the Editor for the encouraging response to our submission. We would also like to thank the three Reviewers for their suggestions and appreciation of our revision. We have now carefully addressed the editorial suggestions and the comments from reviewer #1 in the revised manuscript. We have revised the main text to support the editorial limitations on word count. We have also included a method section into the main text. This is a compressed version of the extended methods section included in supplementary information. The analyses included in our manuscript are ample and would therefore appreciate keeping the very detailed method section in the supplementary information as well. As suggested, we included 10 Extended Data supporting the main figures and the manuscript text. The supplementary information contains an overview of all the Supplementary Material which includes: Methods, results, Figures, Tables and information on the Source Data.

On a special note, we have used BioRender to produce part of the illustrations. Due to holiday time, we will be able to provide the license proof for BioRender only after August 12th. We hope this will be acceptable to move forward with the submission.

Lastly, we acknowledge the importance of transparent peer review and would like to participate in the proposed transparent peer review option by Nature Microbiology.

We look forward to the Editorial response.

Kins regards,

Simona and Ronnie

Below is the point-by-point response to the minor comments from reviewer #1

Minor checks:

*Line 162, which figure is related to the claim should be listed.

Response. The reference is to Table S2 as mentioned in the text.

*Line 171, the title mentioned commensal interactions, while the results did not really do interaction analyses.

Response: The revised manuscript was reduced according to the editorial recommendations. This title is no longer present as a result of this reduction.

*Line 220, check whether this can be seen in Fig.2 bc.

Response: We agree that the exact numbers must be inferred from the axis scale in Fig. 2b,c. We therefore report the precise values (76-146 isolates) in the text to ensure clarity.