nature research

Peer Review Information

Journal: Nature Methods
Manuscript Title: Massively parallel single-cell sequencing of diverse microbial populations
Corresponding author name(s): Ophelia Venturelli

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:

10th Apr 2022

Dear Dr Venturelli,

Thank you for your inquiry about submitting your manuscript, "Ultrahigh-throughput single-cell sequencing of genetic loci in microbial populations" to Nature Methods. The paper sounds like it may be of interest, and should fit the scope of the journal.

Of course, it is very difficult to judge a paper based only on the limited information available in a presubmission inquiry. Therefore I am sure you understand that we cannot promise to send your paper out for peer review and must read it in its entirety before deciding if this would be suitable.

Please keep in mind that the journal is aimed at a large, interdisciplinary audience and places a strong emphasis on the practical value of the work presented. We strongly encourage you to include data to validate method performance and demonstrate its advantages over existing tools.

You will find our Guide to Authors at http://www.nature.com/naturemethods to assist you in preparing your manuscript. However, it is not necessary at this stage to spend major effort adhering to our detailed formatting instructions.

Thank you for your interest in Nature Methods.

Sincerely,
Lei

Lei Tang, Ph.D.
Senior Editor
Nature Methods

Author Rebuttal to Initial comments

[There is no rebuttal letter at this stage.]

Decision Letter, first revision:

13th Jun 2022

Dear Dr Venturelli,

Your Article entitled "Ultrahigh-throughput single-cell sequencing of genetic loci in microbial populations" has now been seen by 3 reviewers, whose comments are attached. While they find your work of potential interest, they have raised serious concerns which in our view are sufficiently important that they preclude publication of the work in Nature Methods, at least in its present form.

As you will see, the reviewers raise technical concerns about DoTA-seq's throughput and robustness.

Should further experimental data allow you to fully address these criticisms we would be willing to look at a revised manuscript (unless, of course, something similar has by then been accepted at Nature Methods or appeared elsewhere). This includes submission or publication of a portion of this work somewhere else. We hope you understand that until we have read the revised paper in its entirety we cannot promise that it will be sent back for peer-review.

If you are interested in revising this manuscript for submission to Nature Methods in the future, please contact me to discuss your appeal before making any revisions. Otherwise, we hope that you find the reviewers' comments helpful when preparing your paper for submission elsewhere.

Best regards,
Lei

Lei Tang, Ph.D.
Senior Editor
Nature Methods

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:
In their manuscript, "Ultrahigh-throughput single-cell sequencing of genetic loci in microbial populations", Lan et al. develop a microfluidics-based PCR method for fusing barcodes with taxonomic or genetic markers. They perform a proof-of-concept experiment using two organisms, one containing a plasmid. They then go on to show that the method works on a 25-member synthetic community. This is a nice demonstration of the technology. The method builds on several previously published techniques, adding a barcoded primer so that frequencies of genes can be calculated, though the manuscript has several omissions that would go farther in validating the method (mainly showing the utility of the method on a gut microbiome community with a spiked-in bacteria). Additionally, while there is no doubt that microfluidic devices can be used in ultrahigh-throughput, the throughput of this particular method is questionable.

Major comments:

1. The claim that the method is ultrahigh-throughput (title, line 113) and generates "millions of droplets within minutes" (line 154) is true but describes the microfluidic device they use, rather than the method. The authors assay thousands of cells per replicate (around 10,000 at most in a single replicate), which is hardly high-throughput.
2. Similarly, the method does not appear to be high-throughput in terms of numbers of samples, if each requires 8 hours of hands-on time and is limited by the number of chips that can be run simultaneously.
3. The authors highlight the need for a rare and sensitive approach (lines 81 and 103), yet there are reasons to believe that DoTA-Seq would not be very sensitive in an actual diverse community based on the results from their 25-member synthetic community experiment. A real community would have hundreds of members at different frequencies. The variability in the representation of 25 members with supposedly equal distribution is also concerning. This variability will obviously affect the sensitivity of the assay, as some of the species surveyed had relatively low capture. Overall, there should be more discussion on how "quantitative" DoTA-Seq is.
4. The authors propose DoTA-Seq as a method for assaying complex bacterial communities; however, they do not actually apply the method to a real gut microbiome community. This is an essential experiment and would need validation (maybe with a spike-in experiment as well) to validate the method. The fact that they have not performed this experiment is mildly concerning. Is their microfluidic chip robust to having environmental matter in gut samples? In this experiment, the 16S rRNA data should be compared to the original 16S rRNA abundances to assess capture.
5. The authors must provide substantial data and information about the sequencing quality. We assume that given the multiplexed PCR and the single-cell reactions that much of the sequencing reaction generated junk reads (poor barcodes, messy amplicons, reads that did not align as expected). This data should be provided and discussed. Similarly, data should be provided on the actual numbers of doublets etc. that are generated, and an analysis of whether there are other doublets given that some of the 16S rRNA are not being captured as well. Doublets in a natural community will be important and should be accounted for.
6. How does the amplification efficiency map according to the number of amplicons in each droplet or the number of genes being assessed? Did they note any changes here? Or have any data to show that?

Minor comments:
1. It appears that the authors used a different barcoding method for the experiments performed in Figure 1 and their 25-member community. This should be made more explicit.

2. Line 124: Microfluidics is still relatively inaccessible, so "democratizing" is an overreach.
3. Line 138: There are important citations missing here: doi.org/10.1038/ismej.2015.124 and doi.org/10.3389/fmicb.2015.00195. Also, updating the references to include single-cell sequencing approaches: DOI: 10.1126/science.abm1483.
4. Figure 3b, line 380: The authors decide to use a cutoff of 10%, which only removes a single sample from the graph and doesn't seem necessary
5. Line 450-452: They discuss normalization methods in the discussion, but these should be discussed more thoroughly earlier in the paper.
6. Line 512: Curious about why ultramers were needed
7. Line 517: The file used to fabricate the microfluidic device should be included as supplemental information (CAD file).
8. Line 555-557: The lysozyme and proteinase K incubation be increased from 30 minutes to 1 hour in line with the recommended time for gram-positives in many commercial kits.

Reviewer #2:
Remarks to the Author:
This paper introduces DoTA-seq, a novel high-throughput approach for targeted single cell DNA sequencing of bacteria. DoTA-seq encloses single bacterial cells in droplets, co-localizing each cell with a unique barcode. PCR is performed in the droplet with primers that amplify the genomic loci of interest as well primers that amplify the barcode. The PCR results in molecular chimera between a copy of the barcode molecule and the gene amplicon. Subsequent sequencing of these chimeras makes it possible to group reads by cell of origin using the barcode. DoTa-seq is applied to a set of increasingly complex biological systems from a simple mixture between E. coli and B. subtilis to studying phase variation in B. fragilis and finally a synthetic consortium of 25 strains. The analysis of CPS gene promoter orientation in B. fragilis is particularly exciting case study for single cell DNA-sequencing.

DoTA-seq is a very elegant approach and the biological applications are exciting. This manuscript can be published with very minor changes. Great work!

1. Both cell encapsulation and barcode molecule addition are Poisson processes. For valuable samples with low cell counts this could become a concern if cells are lost because they are not in a barcoded droplet. This might be less of an issue for bacteria than it is for mammalian cells, but it would be interesting to discuss this potential limitation.

2. If I understand the approach correctly, a single DNA barcode molecule has to be included in a droplet and subsequently amplified in order to ensure that all gDNA amplicons in that droplet acquire the same barcode. It would be interesting to briefly discuss the optimal conditions for this PCR reaction, even if it is not an original contribution. For example, does barcode amplification have to be more efficient than template amplification in order to ensure that an excess of barcodes is available even if there are multiple genomic loci that are getting amplified? It seems quite surprising that a fusion PCR reaction starting with such low template concentration results in efficient formation of the desired barcoded amplicon.

3. "Thus, in any B. fragilis population, there are 128 possible CPS operon promoter states, where each combination of promoter ON and OFF states could yield a unique

CPS phenotype. " Are all of these combinations actually biologically possible? I am not an expert in this field but was under the impression that only one type of CPS is expressed per cell. The data shown here clearly show that this is not the case at least at the transcriptional level.

4. Primer dimers are mentioned in the discussion, but I still find it surprising that they do not generate major problems given the very low template abundance and complexity of the primer sets used. How are dimers removed before sequencing? More generally, it would be interesting to quantify sequencing efficiency. I.e. what fraction of the reads in a typical run correspond to barcoded genomic amplicons and what fraction are dimers or other side products?

Reviewer #3:
Remarks to the Author:
This work aims to introduce a droplet / bead microfluidics technique for encapsulation of bacteria culture, lysis, barcoding-PCR and subsequent sequencing. The work can be regarded as a simplified extension of previously published work by the first author (REF 12). The work is full with hidden methodological details that are either sealed or hidden somewhere in the context (see below). The analytical sensitivity of the technique is questionable, often lacking critical details (see below). The data analysis is described superficially making it impossible to properly peer-review the results. If I understood the technique correctly in the best case scenario, at the last step of the technique only 1% of droplets will have a cell and a barcode? If so, such technique becomes low-throughput and unproductive. The method, as nice as it sounds, does not provide sufficient confidence in sensitivity, robustness, whilst throughput seems to be rather limited. With all due respect, I cannot recommend this work for publication in Nature Methods.

I hope you will find my comments useful:

- It is known that encapsulated bacteria tend to adhere at the water-oil (or hydrogel-oil) interface (e.g., Song et al., Biomacromolecules 2019, 20, 4437−4446; Dorobantu et al., Appl. Environ. Microbiol., 2004, 70, 6333–6336) and will be lost during bead gelation. This should clearly explained, as this will be one of the limitations of the DoTA-Seq technique.

- The DotA technique involves multiple steps. Authors should have conducted proper analysis (using Sybr Green I dye) quantifying the occupancy of cells in water-in-oil droplets, after hydrogel formation, after lysis, after washes (given the throughput preferably with n > 1000).

- The technique argues about "high-throughput" but in fact the results are obtained on only very few cells n = 6 to 1014. (See also Figure 3, Figure S3, Figure S4, Figure S5).

- How do you prevent cells from clumping and premature lysis during the encapsulation process?

- Researchers working with microbial samples are well aware that bacteria samples tend to clump and clog the filters of the microfluidics device. Understandably, it is possible to show a selective images of input during the first minutes of encapsulation, but it would be useful to show how such filters, on a 20 µm device, look at the end of encapsulation?

- The "barcoding" step relies on a single-DNA molecule encapsulation and amplification by PCR, the

process which is prone to DNA loss.

- It's not clear what is the false positive rate and specificity of DNA amplification reaction given that 40 cycles of PCR is used. Under such conditions any ambient DNA will be amplified very efficiently.

- It is not clear how efficient is DNA assembly in droplets, which should depend on genetic background of microorganisms

-Assuming that occupancy of hydrogel beads by microbes is 0.1, and occupancy of DNA barcodes is ~0.1 it leads to many cells being non-amplified and lost in a process. If I understood the technique correctly in the best case scenario, at the last step of the technique only 1% of droplets will have a cell and a barcode? If so, such technique becomes low-throughput and unproductive.

- it is not clear how efficient lysis is for each species out of 25 strains. Lysis of some bacteria may need harsher conditions (e.g. 0.1M NaOH).

- How authors prevented bacteria clumping of previously frozen samples when conducting the experiment involving 25-members of synthetic community?

- Figure 1b. The authors show "lysed" cells, but in image they appear as intact or incompletely lysed cells. You may consider showing encapsulated cells at each step of a protocol (water droplet in oil, hydrogel bead, lysed cell)

- Figure 1e. On Y-axis please provide actual cell number being analyzed, and not the percentage. Alternatively, provide n for each bar.

- Figure 3. It would be useful to the unexperienced reader to understand why so few cells were sequenced? Why to use the ultrahigh-throughput technique to sequence few hundred cells?

- Why for some species recovery was only at 10-20 cells, while for other species it was 10-times higher, even though the species were mixed at even ratios? Do these results show the intrinsic bias of the technique?

Material and Methods

Provide cat not for Aquapel Glass Treatment. The company offers many choices.
What is cat no for the lysozyme?
Please make sure all reagents are provided with respective cat no.
What equipment was used to make droplets?
It is not clearly articulated how 25 µl of PCR primer mix is injected at 200 µl/hr rate, while beads are suspended in 1 ml and injected at 200 µl/hr. Does that imply that only 25 µl of beads are injected along with PCR primer mix. This needs to be clarified.
Another confusing statement that emulsion is being produced at 200 µl/hr + 900 µl/hr for oil can be collected into a PCR tube those volume is only 200 µL?
Please mention that TCEP is added to break the PAAG beads.
L631: Its not clearly explained how single-cell digital PCR sample was prepared. It seems that authors conducted PCR in the presence of ethanol, which does not make sense since even small traces of

ethanol will inhibit PCR.
L671: Please explain how many raw reads are typically retained and how many reads are thrown out after filters.

Author Rebuttal, first revision:

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:

In their manuscript, "Ultrahigh-throughput single-cell sequencing of genetic loci in microbial populations", Lan et al. develop a microfluidics-based PCR method for fusing barcodes with taxonomic or genetic markers. They perform a proof-of-concept experiment using two organisms, one containing a plasmid. They then go on to show that the method works on a 25-member synthetic community. This is a nice demonstration of the technology. The method builds on several previously published techniques, adding a barcoded primer so that frequencies of genes can be calculated, though the manuscript has several omissions that would go farther in validating the method (mainly showing the utility of the method on a gut microbiome community with a spiked-in bacteria). Additionally, while there is no doubt that microfluidic devices can be used in ultrahigh-throughput, the throughput of this particular method is questionable.

Response: We thank the reviewer for the comments. We understand that the reviewer has doubts about the throughput of the method. Sequencing even 10,000 cells at a time with DoTA-seq is an advance in single-cell bacterial sequencing especially single-cell genetic analysis is not widely used in the microbiology and microbiome research community. The throughput demonstrated the previous version of the manuscript (~3,000 - 10,000 cells per run) is not the upper limit of this method. The rate-limiting step in scaling up library prep (droplet making) was typically around ~6 minutes in our libraries thus far. To address this reviewers' concerns, we have conducted ~30 minutes of droplet making, sequencing up to **~37,000** cells in a full run of the Miseq **(Figure 4c)**, which better highlights the throughput capabilities of this method. Notably, this run did not require any significant changes to the workflow other than the run-time for the second microfluidics step (~30 mins instead of ~6 mins).

We would also like to note that the widely used commercial high-throughput single cell sequencing instruments such as the Mission Bio Tapestri and the 10X Genomics Chromium processes

~5,000 to ~10,000 cells at a time, using microfluidic devices, with longer workflows compared to DoTA-seq. Therefore, even this degree of throughput can be still very powerful for providing new biological insights into single-cell heterogeneity in diverse microbial populations.

Major comments:

Remark: 1. The claim that the method is ultrahigh-throughput (title, line 113) and generates "millions of droplets within minutes" (line 154) is true but describes the microfluidic device they use, rather than the method. The authors assay thousands of cells per replicate (around 10,000 at most in a single replicate), which is hardly high-throughput.

Response: We thank the reviewer for pointing out this potential source of confusion for readers. We intended to state that the process of encapsulating bacteria into hydrogels is very fast, and therefore, not the time-limiting step in the workflow. We have accordingly amended the manuscript to clarify this point. In addition, we have carried out new experiments to demonstrate that DoTA-seq is capable of sequencing **~37,000** cells per library **(Figure 4c)**, which is broadly considered high-throughput("Encyclopedia of Cell Biology" n.d.).

Remark: 2. Similarly, the method does not appear to be high-throughput in terms of numbers of samples, if each requires 8 hours of hands-on time and is limited by the number of chips that can be run simultaneously.

Response: Indeed, due to the use of microfluidic devices, this method in its current state is not high-throughput in terms of samples. However, multiple samples can be processed in parallel. Therefore, we respectfully disagree that each sample requires 8 hours of hands-on time. The only non-parallelizable step is the droplet making step, which is currently ~6 minutes per sample for ~10,000 cells. For comparison, the commercially available instruments Mission Bio Tapestri and the 10X Genomics Chromium are also not high-throughput in terms of samples but have found widespread use in single-cell sequencing research. We do not claim to be high-throughput in terms of samples and we have amended the manuscript to make this clear. In addition, we elaborated on this point in the discussion to suggest potential modifications to the workflow that could increase the throughput of the number of samples in the future (this is beyond the scope of our study).

Remark: 3. The authors highlight the need for a rare and sensitive approach (lines 81 and 103), yet there are reasons to believe that DoTA-Seq would not be very sensitive in an actual diverse community based on the results from their 25-member synthetic community experiment. A real community would have hundreds of members at different frequencies.

Response: We thank the reviewer for this comment. Yes, natural microbial communities can be extremely diverse, and the sensitivity of the method would depend on the number of cells sequenced, which is highly flexible depending on how much sequencing one is prepared to do. To address the reviewer's concern, we have conducted new experiments using highly complex fecal derived natural communities and profiled a large number of cells (~37,000) to demonstrate the potential applicability of DoTA-seq to complex natural communities **(Figure 4)**. In addition, we provided a discussion of the sensitivity of detection **(Supplemental Note 4**) and removed any mentions of high sensitivity or detection of rare events from the manuscript.

Remark: The variability in the representation of 25 members with supposedly equal distribution is also concerning. This variability will obviously affect the sensitivity of the assay, as some of the species surveyed had relatively low capture. Overall, there should be more discussion on how "quantitative" DoTA-Seq is.

Response: We thank the reviewer for requesting further clarification of this point. The relationship between OD600 and cell number varies across species (Venturelli et al. 2018). As such, we did not expect that each species in the 25-member synthetic community would have equal proportions based on DoTA-seq, which captures the number of cells of each species. We amended the manuscript to avoid this confusion (line 345-346).

To develop a better understanding of DoTA-seq's quantitative capabilities, we have conducted experiments using commercial microbial community standard to assess relative abundance measurements (**Supplemental Figure 2**). With the exception of two species, we observed a reasonable concordance between the relative abundance from DoTA-seq compared with measurements made by metagenomic shotgun sequencing of the sample, even though these two workflows are extremely different. We further conducted experiments with varying levels of two spike-in species in the context of a complex natural community. Our results show that the observed abundance of the spike-in cells decreased as their initial abundance was reduced, as expected **(Supplemental Figure 8)**. Overall, these new experiments demonstrate that DoTA-seq experiments can be used to reliably measure relative changes between samples. Finally, we added a discussion of the quantitative capabilities of DoTA-seq to the Discussion section.

Remark: 4. The authors propose DoTA-Seq as a method for assaying complex bacterial communities; however, they do not actually apply the method to a real gut microbiome community. This is an essential experiment and would need validation (maybe with a spike-in experiment as well) to validate the method. The fact that they have not performed this experiment is mildly concerning. Is their microfluidic chip robust to having environmental matter in gut samples? In this experiment, the 16S rRNA data should be compared to the original 16S rRNA abundances to assess capture.

Response: We thank the reviewer for this suggestion. We agree that applying DoTA-seq to a microbial community derived from a natural sample will significantly strengthen the paper and thus we have carried out the experiments the reviewer proposes. Specifically, we performed DoTA-seq on a microbial community derived from a human fecal sample with two spike-in species and performed the requested analyses **(Figure 4)**. Our new data shows that DoTA-seq provides comparable results to 16S profiling in terms of family level taxonomic composition. Our data also revealed the host associations of plasmids in this sample. To address unique challenges in data analysis for natural samples, we have developed a new analysis pipeline specifically for analyzing DoTA-seq data from natural samples (**Supplementary Figure 9**).

Remark: 5. The authors must provide substantial data and information about the sequencing quality. We assume that given the multiplexed PCR and the single-cell reactions that much of the sequencing reaction generated junk reads (poor barcodes, messy amplicons, reads that did not align as expected). This data should be provided and discussed.

Response: We appreciate the reviewers' concerns regarding PCR side-products. With proper quality control over multiplex primer sets, and significant investment in primer selection and PCR conditions, we have been able to minimize the generation of PCR side-products. Therefore, this question highlights a key success of our approach. With our size-selection method to remove primer dimers, ~70-90% of our reads have passed both the barcoding filter and also map to the expected target sequences. Reads that do not pass barcode filter are typically low-quality reads from over-loading the sequencer. Reads that do not map to the target sequence are typically short primer dimers. We have found that these unmapped reads do not represent a significant portion of the overall sequencing reads. We have provided a breakdown of the read yields throughout the filtering steps **(Supplemental Table 11).** We have additionally provided a discussion on primer design for DoTA-seq to maximize success (line 528-535), as well as an automated pipeline for generating DoTA-seq primer sets (see Methods: Generating DoTA-seq primers), which in our hands provides very successful primer set designs requiring only minor tweaks.

Remark: Similarly, data should be provided on the actual numbers of doublets etc. that are generated, and an analysis of whether there are other doublets given that some of the 16S rRNA are not being captured as well. Doublets in a natural community will be important and should be accounted for.

Response: Doublets (two or more cells encapsulated by a single barcode) are expected to be present in ~10% of barcoded droplets **(Supplementary Note 2)**. Often, we can screen them out when multiple different 16S amplicons are detected. There will, of course, be rare cases where the 16S sequence of one of multiple encapsulated species is not sampled randomly through sequencing. We mitigate these events by specifying that all droplets should contain at least 10 marker gene reads. If two cells are from the same species, they would not be screened out based on 16S sequence. In this case, the results would be minimally affected since we aggregate cells of the same species in the data analysis. We have provided an analysis of potential doublets, our observations from data, and ways we mitigate them in our experiments within the manuscript as requested **(Supplementary Note 2).**

Remark: 6. How does the amplification efficiency map according to the number of amplicons in each droplet or the number of genes being assessed? Did they note any changes here? Or have any data to show that?

Response: We are unsure what the reviewer meant by amplification efficiency, so we have attempted to address two possible interpretations of this comment. Our first interpretation is that the reviewer is wondering whether the increasing number of primers in multiplex amplification impacts the overall efficiency of the PCR reaction. To our understanding, the amplification efficiency of PCR is dependent on the thermodynamics of primer annealing to the target sequence. Many factors affect this primer annealing, including interactions with off-target sequences and primer dimer formation. Binding to off-target sequences is unlikely because a single bacterial genome is on the order of megabases and the probability of encountering an off-target sequence for a PCR primer is low. Primer dimers can become a challenge with increasing numbers of primer sets. To address this challenge, we have developed an automated pipeline for designing multiplex primer pools (see Methods: Generating DoTA-seq primers). Our current pipeline (available on Github) provides working primer-pools for ~10-plex targets almost every time. Occasionally, it is necessary to manually replace one or two primers that do not perform well in initial PCR tests. A discussion on primer design and PCR conditions has been added to the manuscript (lines 517-527).

Our second interpretation is that the reviewer is wondering whether increasing the number of targeted amplicons would result in more potential for certain targets to not be amplified relative to the other targets, or not sampled through sequencing. First, the relative amplification efficiencies will differ for each multiplex primer set, which determines the depth of sequencing required to obtain reads for all target amplicons. We can compensate for differences in amplification efficiency of different target loci by adjusting relative primer concentrations, as discussed in **Supplemental Note 5.**

With increasing number of amplicons targeted, the DNA yield for each individual amplicon may be expected to decrease in each droplet. However, this is not a problem as there is always more than enough DNA for sequencing. When detecting multiple amplicons per cell, higher depth of sequencing is needed to ensure coverage of every amplicon for each cell. Differences in amplification efficiency between the different targets can result in an imbalance in the amplicons present in each droplet. Therefore, the optimal depth of sequencing is best determined empirically. As a point of reference, in **Figure 1,** we sequenced ~15,000 cells, targeted 3 genes per cell, received 200-400 reads per cell, resulting in 92% +/- 2% (mean, stdev, 2 replicates) of cells containing at least 5 reads mapping to each target. In **Figure 2**, we targeted simultaneous amplification of 7 different loci for every cell. We sequenced ~16,000 – 20,000 cells at a depth of ~100 -200 reads per cell. After filtering for cells that contained at least 5 reads for each locus, we filtered the data to 50% +/- 5% (mean, stdev, 3 replicates) of the cells.

Minor comments:

1. It appears that the authors used a different barcoding method for the experiments performed in Figure 1 and their 25-member community. This should be made more explicit.

Response: We have used the same DoTA-seq protocol for experiments performed in Figure 1 and Figure 3, except with different primers (and primer concentrations) targeting different loci (Methods: DoTA-seq of mixture of E. coli and B. subtilis vs. Methods: DoTA-seq of 25 member synthetic community).

2. Line 124: Microfluidics is still relatively inaccessible, so “democratizing” is an overreach.

Response: Yes, we agree with the reviewer but we are hopeful that droplet microfluidics will become highly adopted in the future. To satisfy the reviewers, we have amended the manuscript accordingly.

3. Line 138: There are important citations missing here: doi.org/10.1038/ismej.2015.124 and doi.org/10.3389/fmicb.2015.00195. Also, updating the references to include single-cell sequencing approaches: DOI: 10.1126/science.abm1483.

Response: We thank the reviewer for pointing out these references and we have accordingly amended the manuscript.

4. Figure 3b, line 380: The authors decide to use a cutoff of 10%, which only removes a single sample from the graph and doesn't seem necessary

Response: We decided to use a 10% cutoff to be safe from any potential false positives, as we see some unexpected cross-talk at 1-2% in figure 3a.

5. Line 450-452: They discuss normalization methods in the discussion, but these should be discussed more thoroughly earlier in the paper.

Response: We have added a more thorough discussion of normalization methods earlier in the paper (lines 362-369).

6. Line 512: Curious about why ultramers were needed

Response: When we include the sequencing adaptors P5 onto the sequencing primers, the sequences get very long, and thus ultramers are needed to ensure higher synthesis yields. We have revised our Methods section to provide this explanation (lines 656-657).

7. Line 517: The file used to fabricate the microfluidic device should be included as supplemental information (CAD file).

Response: We have included the CAD file for photolithography masks in the supplemental materials.

8. Line 555-557: The lysozyme and proteinase K incubation be increased from 30 minutes to 1 hour in line with the recommended time for gram-positives in many commercial kits.

Response: We thank the reviewers for this suggestion. We have included this change as well as other improvements in lysis protocol into our standard protocol as well as new experiments.

Reviewer #2:

Remarks to the Author:

This paper introduces DoTA-seq, a novel high-throughput approach for targeted single cell DNA sequencing of bacteria. DoTA-seq encloses single bacterial cells in droplets, co-localizing each cell with a unique barcode. PCR is performed in the droplet with primers that amplify the genomic loci of interest as well primers that amplify the barcode. The PCR results in molecular chimera between a copy of the barcode molecule and the gene amplicon. Subsequent sequencing of these chimeras makes it possible to group reads by cell of origin using the barcode. DoTa-seq is applied to a set of increasingly complex biological systems from a simple mixture between E. coli and B. subtilis to studying phase variation in B. fragilis and finally a synthetic consortium of 25 strains. The analysis of CPS gene promoter orientation in B. fragilis is particularly exciting case study for single cell DNA-sequencing.

DoTA-seq is a very elegant approach and the biological applications are exciting. This manuscript can be published with very minor changes. Great work!

Response: We are very glad to hear that the reviewer finds DoTA-seq to be an elegant approach with exciting biological applications. We are currently in the process of using DoTA-seq for multiple applications including studying phase-variation and phage host range. We hope that many researchers will also find DoTA-seq to be useful to their research.

Remark: 1. Both cell encapsulation and barcode molecule addition are Poisson processes. For valuable samples with low cell counts this could become a concern if cells are lost because they are not in a

barcoded droplet. This might be less of an issue for bacteria than it is for mammalian cells, but it would be interesting to discuss this potential limitation.

Response: We thank the reviewer for pointing out this important consideration. Indeed, we considered that there is usually an excess of cells for bacterial populations, so the sample loss would not be a problem in most cases. Regardless, we believe it is possible to avoid the loss of cells due to the Poisson encapsulation process by ensuring that every droplet contains at least one barcode. This could be accomplished by using barcoded gels instead of single molecule barcodes, which pack closely to achieve one gel per drop(Abate et al. 2009). Alternatively, we could encapsulate multiple unique barcodes per droplet to ensure that every droplet contains barcodes. However, the statistical distribution of unique barcodes per droplet will have to be accounted for when assessing quantitative "cell counts" in the analysis step. We believe that these adaptations are possible and we have included a discussion of this in the manuscript (lines 145-152 and lines 500-509).

Remark: 2. If I understand the approach correctly, a single DNA barcode molecule has to be included in a droplet and subsequently amplified in order to ensure that all gDNA amplicons in that droplet acquire the same barcode. It would be interesting to briefly discuss the optimal conditions for this PCR reaction, even if it is not an original contribution. For example, does barcode amplification have to be more efficient than template amplification in order to ensure that an excess of barcodes is available even if there are multiple genomic loci that are getting amplified? It seems quite surprising that a fusion PCR reaction starting with such low template concentration results in efficient formation of the desired barcoded amplicon.

Response: The reviewer is correct in understanding that a single DNA barcode molecule is amplified to ensure that all gDNA amplicons acquire the same barcode. This is indeed an original contribution as we are not aware of any previous publications of this reaction. It does seem counter intuitive that fusion PCR could work effectively to tag multiple gDNA amplicons starting with a single barcode molecule. Our speculation is that the amplification of the barcode and gDNA amplicons happens in parallel for the first few cycles of PCR, such that there are multiple copies of both before the fusion reaction occurs. This is because the primers do not contain any complementary sequences, rather the complementary overlaps are generated through PCR products, so there is no complementary hybridization between the gDNA amplicons and barcode amplicons until there is sufficient concentration of both PCR products in the later PCR cycles. We have not discussed this in the manuscript since we currently have no experimental evidence and a detailed investigation of this mechanism is beyond the scope of this paper.

Remark: 3. "Thus, in any B. fragilis population, there are 128 possible CPS operon promoter states, where each combination of promoter ON and OFF states could yield a unique

CPS phenotype. " Are all of these combinations actually biologically possible? I am not an expert in this field but was under the impression that only one type of CPS is expressed per cell. The data shown here clearly show that this is not the case at least at the transcriptional level.

Response: We thank the reviewer for this question. It is unclear whether all of these promoter combinational states express unique CPS phenotypes, as there are additional layers of regulation downstream of the promoters (Chatzidaki-Livanis, Coyne, and Comstock 2009). In our single-cell sequencing experiments**,** we observed 117 different CPS promoter combination states at the single cell level. In addition, our analysis suggests that the unobserved promoter states were expected to be present at very low frequencies simply due to the slow rates of promoter inversion (Lan et al. 2022). In other words, we hypothesize that there are no mechanisms directly preventing any two CPS promoters to be simultaneously in the ON orientation. The data presented in this paper shows only the CPS promoter orientations on the genetic level. We do not know the CPS phenotypes expressed by these variants (mapping of genotype to phenotype). However, we are currently in the process of developing ways to expand DoTA-seq's capabilities to include phenotyping with barcoded-antibodies to address this question (beyond the scope of our paper).

Remark: 4. Primer dimers are mentioned in the discussion, but I still find it surprising that they do not generate major problems given the very low template abundance and complexity of the primer sets used. How are dimers removed before sequencing? More generally, it would be interesting to quantify sequencing efficiency. i.e. what fraction of the reads in a typical run correspond to barcoded genomic amplicons and what fraction are dimers or other side products?

Response: We thank the reviewer for this question. Primer dimers are removed through size selection using magnetic beads. Indeed, primer dimers can become a problem if they form large smears on gels. When validating primer sets for DoTA-seq, we look for primer sets that do not form smears on a gel in the absence of any template. We find that this simple test can identify primer sets that work for DoTA-seq since primer dimers can be efficiently removed by DNA size selection techniques. In a typical sequencing run of a cleaned library, ~90% of reads correspond to barcoded genomic amplicons, with the rest corresponding to primer dimers. We thank the reviewer for pointing out the need for clarification on this topic and we have included additional discussion and clarification regarding primer selection and primer dimer removal (lines 517-535), as well as a breakdown of sequencing read yields after filtering **(Supplemental Table 11).**

Reviewer #3:

Remarks to the Author:

This work aims to introduce a droplet / bead microfluidics technique for encapsulation of bacteria culture, lysis, barcoding-PCR and subsequent sequencing. The work can be regarded as a simplified extension of previously published work by the first author (REF 12). The work is full with hidden methodological details that are either sealed or hidden somewhere in the context (see below). The analytical sensitivity of the technique is questionable, often lacking critical details (see below). The data analysis is described superficially making it impossible to properly peer-review the results. If I understood the technique correctly in the best case scenario, at the last step of the technique only 1% of droplets will have a cell and a barcode? If so, such technique becomes low-throughput and unproductive. The method, as nice as it sounds, does not provide sufficient confidence in sensitivity, robustness, whilst throughput seems to be rather limited. With all due respect, I cannot recommend this work for publication in Nature Methods.

Response: We thank the reviewer for examining our manuscript and providing valuable feedback. We attempted to provide full methodological details via a detailed protocol on protocols.io as well as provided detailed breakdown of our data analysis pipeline to provide more clarity on the method (see protocols.io page, link below). While it is true that only 1% of droplets will contain a cell and a barcode, we are able to regularly sequence >10,000 cells per replicate (up to 37,000 cells in the revised manuscript), and have already used DoTA-seq to provide biological insights in important biological systems (Lan F et al. bioRxiv). We hope the reviewer will reconsider the usefulness and applicability of this method in context of our substantial new experimental data and analyses.

I hope you will find my comments useful:

Remark: - It is known that encapsulated bacteria tend to adhere at the water-oil (or hydrogel-oil) interface (e.g., Song et al., Biomacromolecules 2019, 20, 4437−4446; Dorobantu et al., Appl. Environ. Microbiol., 2004, 70, 6333–6336) and will be lost during bead gelation. This should clearly explained, as this will be one of the limitations of the DoTA-Seq technique.

Response: While encapsulated bacteria do sometimes adhere to the interface and is subsequently lost during gelation, we have not found this phenomenon to impact the efficacy of our method. We believe

this is because gelation happens rapidly after encapsulation in DoTA-seq. As such, single bacteria do not have time to adhere or divide, as opposed to the systems cited by the reviewer. Compared to multiple growing cells actively attaching to the surface, single bacteria are statistically less likely to be found at the interface. To assess for loss of cells, we SYBR stained cells for each step of the lysis protocol, counting >1000 droplets/gels for each condition **(Supplemental Figure 4).** The proportion of gels/droplets with SYBR positive cells were 4.5%-5.4% before and 3.9% after the lysis steps, showing that there is minimal cell loss throughout these steps.

Remark: - The DotA technique involves multiple steps. Authors should have conducted proper analysis (using Sybr Green I dye) quantifying the occupancy of cells in water-in-oil droplets, after hydrogel formation, after lysis, after washes (given the throughput preferably with n > 1000).

Response: We thank the reviewer for this suggestion and we have accordingly performed these experiments and provided the relevant data **(Supplemental Figure 4).** We stained cells with both CF555 (membrane/cell wall tracking dye) and SYBR Green I dye and used fluorescence microscopy to analyze the droplets and gels (n > 1000 for each condition) after each step of the DoTA-seq encapsulation and lysis workflows. Our results show that the proportion of cells in droplets remains similar after breaking droplets and washing, and after lysis and washing (**Supplementary Figure 4b**).

Remark: - The technique argues about “high-throughput” but in fact the results are obtained on only very few cells n = 6 to 1014. (See also Figure 3, Figure S3, Figure S4, Figure S5).

Response: We apologize for any confusion. The low throughput experiments in question are not DoTA-seq experiments but colony counting or digital PCR experiments used to independently validate the results of the DoTA-seq experiments (n = ~5000 - ~10,000). We have amended the manuscript to make this more clear. In addition, we have carried out new experiments with ~37,000 cells, which is broadly considered high-throughput(“Encyclopedia of Cell Biology” n.d.) **(Figure 4c).**

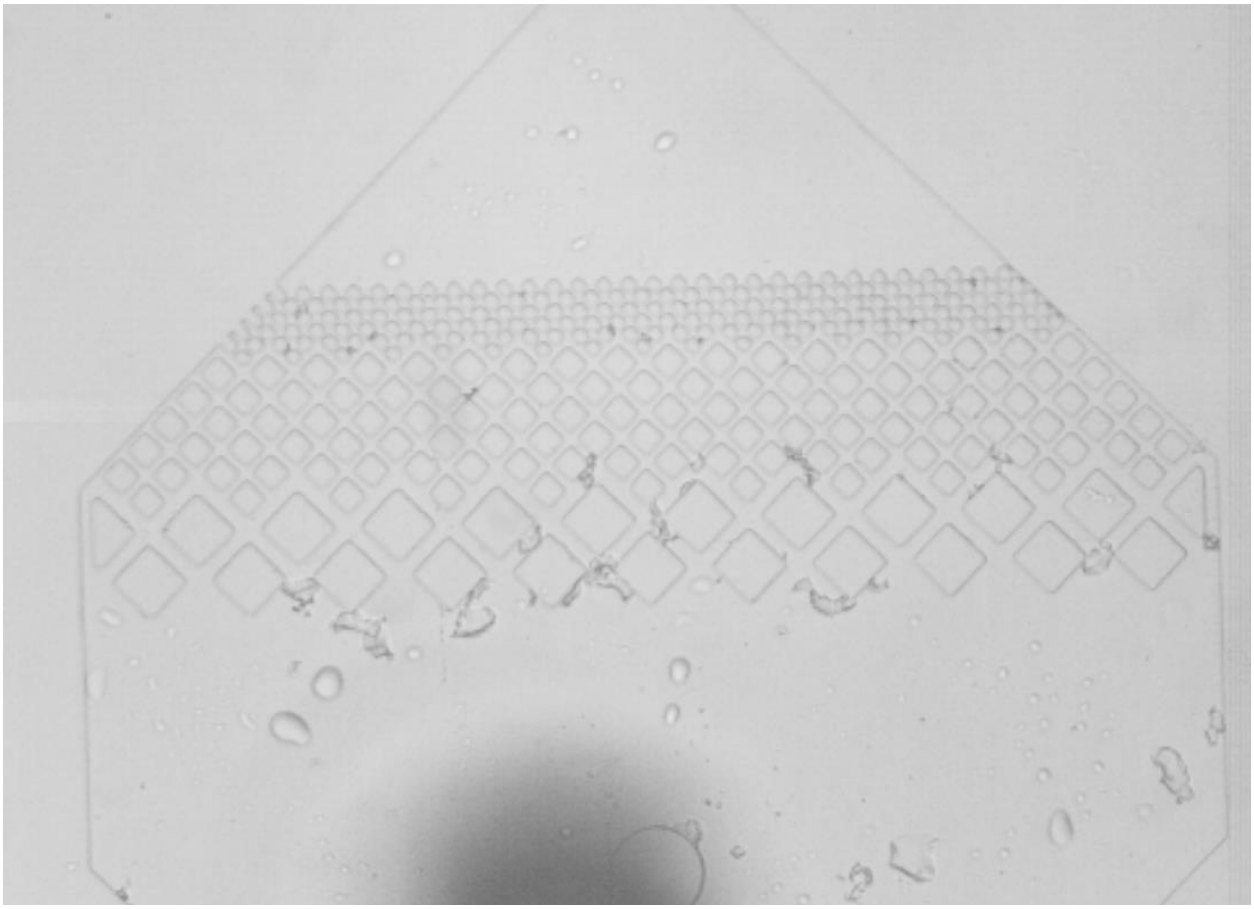
Remark: - How do you prevent cells from clumping and premature lysis during the encapsulation process?

Response: We have found that adding Tween-20 or pluronic-F68 at 0.1% w/v and vortexing cell suspensions has been sufficient to resolve cell clumps. The encapsulation process is performed on the

order of minutes and we have not found lysis to be a noticeable issue. We have provided microscopy images of the device filters after ~45 minutes of encapsulation to show that there is minimal clogging of the device or evidence of cell lysis **(Review figure 1: Cell-filter-post45min.png).**

Remark: - Researchers working with microbial samples are well aware that bacteria samples tend to clump and clog the filters of the microfluidics device. Understandably, it is possible to show a selective images of input during the first minutes of encapsulation, but it would be useful to show how such filters, on a 20 µm device, look at the end of encapsulation?

Response: We have so far not found significant clumping of bacteria and or clogged filters in our device during DoTA-seq workflows which encapsulate cells at very low concentrations (limiting dilution). We can see how this might be a problem when cells are introduced at high concentrations (multiple cells per droplet) but this is not the case in DoTA-seq. We have provided an image of our cell inlet filter as requested of the cell encapsulation device after encapsulating cells for 45 minutes **(Review figure 1: Cell-filter-post45min.png).**



Review figure 1. Cell filter of the microfluidic encapsulation device after 45 minutes of encapsulation of the Zymo microbial community standard cells, showing minimal clogging of the filters of the device.

Remark: - The "barcoding" step relies on a single-DNA molecule encapsulation and amplification by PCR, the process which is prone to DNA loss.

Response: We understand that the process of DNA amplification and barcoding may seem inefficient. However, we nearly always obtain high yielding libraries **(Supplementary Table 11).** We hope that the reviewer will reconsider their assessment of DoTA-seq with the newly provided data and analyses.

Remark: - It's not clear what is the false positive rate and specificity of DNA amplification reaction given that 40 cycles of PCR is used. Under such conditions any ambient DNA will be amplified very efficiently.

Response: We understand that typical PCR protocols do not utilize 40 cycles of PCR. However, it is necessary in droplet PCR, since the droplets that amplify first become saturated at earlier cycles and stop amplifying while the droplets that amplify later (or less efficiently) catch up in the later cycles, thus providing a more even library. For DoTA-seq, we make sure that the PCR primers do not produce large (>300bp) non-specific products, so that they can be removed through size-selection. We have developed an automated primer selection pipeline that does a good job of selecting primers that fit these requirements (See Methods: Generating DoTA-seq primers). In response to the reviewers' comments, we have provided further discussions of primer selection in the manuscript (lines 517-540). We have further provided a breakdown of the sequencing read yields in each analysis step **(Supplementary Table 11)**. Typically ~70-90% of reads map to the expected amplicons.

Remark: - It is not clear how efficient is DNA assembly in droplets, which should depend on genetic background of microorganisms

Response: We thank the reviewer for the question. We have provided additional experimental evidence that DoTA-seq can be used to sequence a wide range of microbial organisms **(Figure 3, 4, Supplementary Figure 3).** Overall, our results show that DoTA-seq can provide reliable single-cell sequencing data for a wide range of organisms.

Remark: -Assuming that occupancy of hydrogel beads by microbes is 0.1, and occupancy of DNA barcodes is ~0.1 it leads to many cells being non-amplified and lost in a process. If I understood ithe technique correctly in the best case scenario, at the last step of the technique only 1% of droplets will have a cell and a barcode? If so, such technique becomes low-throughput and unproductive.

Response: Indeed, only 1% of droplets are expected to contain a cell and a barcode. However, droplets are produced rapidly in the microfluidic device, and each of our typical libraries contains ~1 million droplets in a PCR tube, resulting in ~10,000 cells being barcoded. We demonstrated the capability of profiling up to 37,000 cells in the revised manuscript (**Figure 4c**). We have already used the technique to provide novel biological insights into the *B. fragilis* CPS system in a separate manuscript (Lan et al. 2022).

Remark: - it is not clear how efficient lysis is for each species out of 25 strains. Lysis of some bacteria may need harsher conditions (e.g. 0.1M NaOH).

Response: We thank the reviewer for pointing out the need to better assess the lysis efficiency. We have improved our lysis protocol by using a cocktail of lytic enzymes (see DoTA-seq V3 entry protocols.io:dx.doi.org/10.17504/protocols.io.n92ldzox7v5b/v1), and conducted experiments using a well-characterized and commercially available control microbial community to evaluate the lysis and DoTA-seq capture efficiency on a wide range of microbes **(Supplementary Figures 3 and 4)**.

Remark: - How authors prevented bacteria clumping of previously frozen samples when conducting the experiment involving 25-members of synthetic community?

Response: The reviewer is correct that cell clumping has been observed from frozen and fixed cell samples in our hands as well. However, when we included tween-20 or pluronic F68 at 0.1% into the resuspension and wash buffers of cells, we have found cell clumping to be largely resolved. We have now conducted DoTA-seq on frozen glycerol stocks of a wide range of samples without experiencing significant cell clumping issues. If there are more than one cell encapsulated, we remove barcodes that contain more than one 16S sequence, which should exclude data from bacterial clumps containing more than one species in our data analysis.

Remark: - Figure 1b. The authors show "lysed" cells, but in image they appear as intact or incompletely lysed cells. You may consider showing encapsulated cells at each step of a protocol (water droplet in oil, hydrogel bead, lysed cell)

Response: We thank the author for this suggestion and we have included representative images of encapsulated cells at each step of the workflow **(Supplementary Figure 4).**

Remark: - Figure 1e. On Y-axis please provide actual cell number being analyzed, and not the percentage. Alternatively, provide n for each bar.

Response: We thank the reviewer for this suggestion and have amended the figure accordingly.

Remark: - Figure 3. It would be useful to the unexperienced reader to understand why so few cells were sequenced? Why to use the ultrahigh-throughput technique to sequence few hundred cells?

Response: The total number of cells sequenced is divided among each species. Thus, thousands of cells in a population of 25 species become subpopulations of hundreds of cells for each species. We realize that the figure and the throughput of our method requires clarification, and we have accordingly amended the manuscript (line 348). We have additionally conducted experiments to demonstrate the high-throughput capabilities of the method **(Figure 4)**.

Remark: - Why for some species recovery was only at 10-20 cells, while for other species it was 10-times higher, even though the species were mixed at even ratios? Do these results show the intrinsic bias of the technique?

Response: The species abundances in the 25-member human gut community was not expected to be equal. We apologize for this confusion. We attempted to mix the species at even ratios based on their OD600. However, some species did not grow sufficiently to allow for that, and thus were present at lower than intended ratios. Additionally, OD600 is also not always an accurate indicator of cell numbers. To better address potential biases on species capture by DoTA-seq, we sequenced a well characterized commercially available microbial community **(Supplementary Figure 3).** Our results indicate general concordance (relative abundances within ~2 fold difference) with the expected distribution of gram-positive cells in the sample determined by metagenomic sequencing. Unfortunately, the gram-negative cells in this commercial standard were lysed by the storage method, precluding them from single-cell analysis. However, we've successfully sequenced gram-negative cells in our other experiments (Figure 1 and 2).

Material and Methods

Remark: Provide cat not for Aquapel Glass Treatment. The company offers many choices.

Response: Thank you for pointing this out. We had assumed that there was only one hydrophobic treatment option from Aquapel. We have now provided the catalogue number for clarity.

Remark: What is cat no for the lysozyme?

Response: We thank the reviewer for pointing out this oversight. We have provided this catologue number.

Remark: Please make sure all reagents are provided with respective cat no.

Response: We have amended the manuscript to include catalog numbers where possible and applicable in the detailed supplementary protocol (see link below).

Remark: What equipment was used to make droplets?

Response: We have made a comprehensive protocol on droplet making linked in our protocol.io entry. In this protocol, we also point readers towards additional references in video format for droplet making (Demaree et al. 2018).

Remark: It is not clearly articulated how 25 µl of PCR primer mix is injected at 200 µl/hr rate, while beads are suspended in 1 ml and injected at 200 µl/hr. Does that imply that only 25 µl of beads are injected along with PCR primer mix. This needs to be clarified.

Response: We thank the reviewer for pointing out this potential point of confusion. Indeed, as the reviewer assumed, equal volumes of the gel and PCR mix are injected into the device, even though a significantly larger volume of gels are prepared. We used 25µL of PCR mix in most of our experiments to save on reagents, while the high throughput capabilities of our devices allow us to make massive amounts of gels which we typically use in excess. We have clarified this procedure protocol.io entry (link below).

Remark: Another confusing statement that emulsion is being produced at 200 µl/hr + 900 µl/hr for oil can be collected into a PCR tube those volume is only 200 µL?

Response: Since we only collect the droplets for 6 minutes, we end up with ~50uL of emulsion which fits nicely into a 200µL PCR tube. We have revised the method section be more clear (line 681) and we have added a detailed protocol on protocol.io to address these points of confusion.

Remark: Please mention that TCEP is added to break the PAAG beads.

Response: We thank the reviewer for this suggestion, and we have amended the manuscript accordingly.

Remark: L631: Its not clearly explained how single-cell digital PCR sample was prepared. It seems that authors conducted PCR in the presence of ethanol, which does not make sense since even small traces of ethanol will inhibit PCR.

Response: The cells were fixed in ethanol then subsequently washed in Pre-injection buffer prior to adding to the PCR mix. We have clarified these points in the manuscript (line 808-809).

Remark: L671: Please explain how many raw reads are typically retained and how many reads are thrown out after filters.

Response: We have added a detailed breakdown of the sequencing statistics for our libraries including filtering and mapping of reads **(Supplementary Table 11).** A flow chart of our sequence analysis pipeline is now available **(Supplementary Figure 9).** Details for the scripts used are available in the README file on the DoTA-seq paper github: https://github.com/lanfreem/DoTA-seq-Paper

Protocols.IO link:

DOI: dx.doi.org/10.17504/protocols.io.n92ldzox7v5b/v1 (Private link for reviewers: https://www.protocols.io/private/2EEAAD7C224611ED9F1D0A58A9FEAC02 to be removed before publication.)

References

Abate, Adam R, Chia-Hung Chen, Jeremy J Agresti, and David A Weitz. 2009. “Beating {Poisson} Encapsulation Statistics Using Close-Packed Ordering.” *Lab on a Chip* 9 (18): 2628–31. https://doi.org/10.1039/B909386A.
Chatzidaki-Livanis, Maria, Michael J. Coyne, and Laurie E. Comstock. 2009. “A Family of Transcriptional Antitermination Factors Necessary for Synthesis of the Capsular Polysaccharides of Bacteroides Fragilis.” *Journal of Bacteriology* 191 (23): 7288–95. https://doi.org/10.1128/JB.00500-09.
Demaree, B., D. Weisgerber, F. Lan, and A.R. Abate. 2018. “An Ultrahigh-Throughput Microfluidic Platform for Single-Cell Genome Sequencing.” *Journal of Visualized Experiments* 2018 (135). https://doi.org/10.3791/57598.
“Encyclopedia of Cell Biology.” n.d. ScienceDirect. Accessed November 19, 2022. http://www.sciencedirect.com:5070/referencework/9780123947963/encyclopedia-of-cell-biology.
Lan, Freeman, Jason Saba, Yili Qian, Tyler D. Ross, Robert Landick, and Ophelia S. Venturelli. 2022. “Single-Cell Bacterial Sequencing of Invertible Loci Reveals Promoter Inversion Rates as a Strong Determinant of Population Heterogeneity.” bioRxiv. https://doi.org/10.1101/2022.10.31.514637.
Venturelli, Ophelia S, Alex V Carr, Garth Fisher, Ryan H Hsu, Rebecca Lau, Benjamin P Bowen, Susan Hromada, Trent Northen, and Adam P Arkin. 2018. “Deciphering Microbial Interactions in Synthetic Human Gut Microbiome Communities.” *Molecular Systems Biology* 14 (6). https://doi.org/10.15252/msb.20178157.

Decision Letter, second revision:

15th Jul 2022

Dear Ophelia,

Thank you for your letter asking us to reconsider our decision on your Article, "Ultrahigh-throughput single-cell sequencing of genetic loci in microbial populations". After careful consideration we have decided that we are willing to consider a revised version of your manuscript that include data to address the major concerns on the throughput and droplet quality of DoTA-seq.

When revising your paper:

* include a point-by-point response to our referees and to any editorial suggestions

* please underline/highlight any additions to the text or areas with other significant changes to facilitate review of the revised manuscript

* address the points listed described below to conform to our open science requirements

* ensure it complies with our general format requirements as set out in our guide to authors at www.nature.com/naturemethods

* resubmit all the necessary files electronically by using the link below to access your home page

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised paper within 8 weeks. If you cannot send it within this time, please let us know. In this event, we will still be happy to reconsider your paper at a later date so long as nothing similar has been accepted for publication at Nature Methods or published elsewhere.

OPEN SCIENCE REQUIREMENTS

REPORTING SUMMARY AND EDITORIAL POLICY CHECKLISTS
When revising your manuscript, please submit reporting summary and editorial policy checklists.

Reporting summary: https://www.nature.com/documents/nr-reporting-summary.zip
Editorial policy checklist: https://www.nature.com/documents/nr-editorial-policy-checklist.zip

If your paper includes custom software, we also ask you to complete a supplemental reporting summary.

Software supplement: https://www.nature.com/documents/nr-software-policy.pdf

Please submit these with your revised manuscript. They will be available to reviewers to aid in their evaluation if the paper is re-reviewed. If you have any questions about the checklist, please see http://www.nature.com/authors/policies/availability.html or contact me.

Please note that these forms are dynamic 'smart pdfs' and must therefore be downloaded and completed in Adobe Reader. We will then flatten them for ease of use by the reviewers. If you would like to reference the guidance text as you complete the template, please access these flattened versions at http://www.nature.com/authors/policies/availability.html.

IMAGE INTEGRITY
When submitting the revised version of your manuscript, please pay close attention to our href="https://www.nature.com/nature-research/editorial-policies/image-integrity">Digital Image Integrity Guidelines. 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.

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.

DATA AVAILABILITY
Please include a "Data availability" subsection in the Online Methods. This section should inform readers about the availability of the data used to support the conclusions of your study, including accession codes to public repositories, references to source data that may be 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", describing which data is available upon request and mentioning any restrictions on availability. If DOIs are provided, please include 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

CODE AVAILABILITY
Please include a "Code Availability" subsection in the Online Methods which details how your custom code is made available. Only in rare cases (where code is not central to the main conclusions of the paper) is the statement "available upon request" allowed (and reasons should be specified).

We request that you deposit code in a DOI-minting repository such as Zenodo, Gigantum or Code Ocean and cite the DOI in the Reference list. We also request that you use code versioning and provide a license.

For more information on our code sharing policy and requirements, please see: https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-computer-code

SUPPLEMENTARY PROTOCOL
To help facilitate reproducibility and uptake of your method, we ask you to prepare a step-by-step Supplementary Protocol for the method described in this paper. We encourage authors to share their step-by-step experimental protocols on a protocol sharing platform of their choice and report the protocol DOI in the reference list. Nature Research's Protocol Exchange is a free-to-use and open resource for protocols; protocols deposited in Protocol Exchange are citable and can be linked from the published article. More details can found at www.nature.com/protocolexchange/about.

ORCID
Nature Methods 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 please visit www.springernature.com/orcid.

We look forward to hearing from you soon.

Best regards,
Lei

Lei Tang, Ph.D.
Senior Editor
Nature Methods

Author Rebuttal, second revision:

[There is no rebuttal letter at this stage.]

Decision Letter, third revision:

1st Feb 2023

Dear Professor Venturelli,

Your Article entitled "Massively parallel single-cell sequencing of genetic loci in diverse microbial populations" has now been seen by 3 reviewers, whose comments are attached. Reviewers #1 and #3 have raised serious concerns which in our view are sufficiently important that they preclude publication of the work in Nature Methods, at least in its present form.

As you will see, the reviewers raise concerns about different aspects of the study/method. After discussion with the editorial team, we are fine with the novelty and throughput of the method (though clarification and toning down are needed). That said, other technical points (such as those about sensitivity and bias of the method) need to be fully addressed for us to consider a revision.

Should further experimental data and major revisions allow you to fully address these criticisms we would be willing to look at a revised manuscript (unless, of course, something similar has by then been accepted at Nature Methods or appeared elsewhere). This includes submission or publication of a

portion of this work somewhere else. We hope you understand that until we have read the revised paper in its entirety we cannot promise that it will be sent back for peer-review.

If you are interested in revising this manuscript for submission to Nature Methods in the future, please contact me to discuss your appeal before making any revisions. Otherwise, we hope that you find the reviewers' comments helpful when preparing your paper for submission elsewhere. Thank you for the opportunity to consider this paper.

Sincerely,

Lin Tang, PhD
Senior Editor
Nature Methods

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
The two major flaws from the previous version were the overstatement of the throughput of the method and the lack of validation on actual samples. Despite the effort from the authors, the added justification and added experiments are not convincing of the method. (If anything, they go further in raising questions).

The throughput of the method is still low (the comparison with 10X is not a fair comparison given that 10X is fully profiling cells). 10,000 cells on a method that cannot be easily performed for many samples is not high-throughput. 10,000 cells (at best) is also not sufficient for quantitative measurements of gene abundances in communities, especially of low abundant organisms, and most especially, with the low sensitivity rates they have. They claim it is highly parallelized and it is not at all in their rendition. With real samples, there is likely to be clumping etc. that further limit the potential to parallelize this.

Rather than use a real environmental sample, the authors use a commercial product (which seems to be a mock community as well) and has some problems with storage conditions. This does not allow them to answer questions about lysis in a mixed community. Most importantly, the improvement the authors suggest is including a barcode to identify specific cells. Yet, in their experiments on the commercial product shows that the sensitivity of their method goes down (60% average for the essential gene which should be 100%). Therefore, this negates any ability for the barcode to be quantitative! The authors suggest the false positives is the limit of detection, but it is this barcoding problem that is central to their improvements that does not seem to be working well.

would not recommend this method to be used in my lab.

Reviewer #2:
Remarks to the Author:
This is an excellent paper that should be published in Nature Methods as soon as possible. The authors

have done a great job addressing the issues raised by all reviewers.

Reviewer #3:
Remarks to the Author:
I appreciate the efforts that authors put in attempting to address the reviewers' concerns. The authors have shown that they can sequence target DNA region in bacteria, in up to 37k cells. The earlier results (provided in the original submission) also included results from 25 species synthetic community. In updated version the authors also performed DoTA-seq on one stool sample and included two spike-in species analyses. The new results show that DoTA-seq can be used to profile 16S DNA of single bacteria cells but at the cost of reduced accuracy, reduced resolution of cell composition and with biases in cell capture. The authors have supplemented their original draft with Figure 4, Supplementary Figure 2, Supplementary Note 4, etc. It is now more clearly explained the limitations and advantages of the technique. It is evident that the DoTA-Seq levareges the throughput of droplet microfluidics technology to isolate single-bacteria species, which are then gelled into beads, and genes of interest are amplified during indexing (barcoding) PCR reaction. While DoTA-seq may appear as an elegant approach schematically, yet careful analysis of the updated manuscript revealed analytical drawbacks that have significantly dampened my enthusiasm.

First, I would like to comment on the rebuke letter. Regarding authors' claim that „single bacteria are statistically less likely to be found at the interface". This cannot be true. Droplets have high surface to volume ratio, meaning that statistically the cells will have higher probability to be closer to the interface than reside in the core of the droplet. Being near the interface will cause genetic material loss during gelation and lysis. Although the work related to bacterial loss in hydrogel beads is known to the users, the public reports on this issue are scarce. However, the authors could refer to these two papers for consideration: Figure 1a in 10.1002/adhm.201200341 and Figure S6 in 10.1039/D0LC00660B. In this context, Supplementary Figure 4 must have some errors in the measurements as the number of encapsulated cells should not increase after gelation (pre-lysis gels).

The new experiments (Figure 4 and Figure S2) revealed that there are significant biases towards certain species of bacteria during microfluidic operations and workflow processing. For example there was a strong bias towards gram positive cells that can survive sample preparation and encapsulation procedures. Gram negative cells that are more sensitive are poorly captured and become underepresented. When working with a real-world sample authors attempt to put a blame on freezing (i.e. according to the authors freezing the cells causes their lysis prematurely), but then the authors should have validated the technique with a real-world sample without cryo-preservation. In another context authors explain the discrepancy in the results due to the use of DNA/RNA shield reagent. In yet another scenario, authors attempted to profile synthetic community of 25 laboratory-cultivated strains; the results again showed biases, but this time authors attributed the errors to OD600 measurements. When working with reference standard (reference community ZymoBIOMICS), again the DoTA-Seq revealed biases but according to the authors freezing/thawing was the cause, even though it is known that gram negative bacteria cells can withstand multiple cycles of freezing. It appears that DoTA-Seq is operational only on a very limited number of samples making it poorly suitable for an accurate and quantitative research that is needed for microbiology field. I could not identify a convincing indication that the DoTA-Seq technique can provide unbiased profiling of heterogenious bacteria populations.

Concerning fact is that method relies on the use of non-ionic detergents to encapsulate cells. The non-ionic detergents can lyse bacteria cells. Indeed, as one could expect, the authors found that gram negative bacteria were severelly underepresented in their results (Figure S2). As I mentioned above these issues will rise serious concerns to microbiology community because it will always remain questionable which species have been lost (or retained) along the way.

While I agree that the method is a high-throughput in the first step (i.e. cell isolation), yet the bead reinjection and encapsulation with DNA molecules has rather limited throughput. The number of samples that can be handled is limited as well. Therefore, the entire story around „throughput" is not really convincing.

Authors should cite the work relevant to single-bacteria genome sequencing:
https://doi.org/10.1016/j.cell.2022.06.040
10.1126/science.abm1483
10.1038/s41564-020-0729-6

I could not find experimental evidence proving that perceived combinatorial states (phase variation) of promoters are a real result and not some kind PCR artifact. The authors will agree that it is well known in the microbiology field that PCR generates chimeras and recombination events, especially if the template contains polyndromic sequences. This has been know for decades. See for example:
https://doi.org/10.1111/j.1574-6941.1998.tb00500.x
https://doi.org/10.2144/000113219
https://doi.org/10.1093/nar/19.8.1949
https://doi.org/10.1093/oxfordjournals.molbev.a025797
It is also known that when analyzing mock community data, the results strongly deviate from the expected composition with the increasing number of PCR cycles. Considering that the DoTA-Seq technique relies on 40-cycles of PCR, false promoter states or even false 16S species are likely to appear due to accumulating PCR artifacts. Highly relevant and equally concerning is relevant report by Sze et al., titled "The Impact of DNA Polymerase and Number of Rounds of Amplification in PCR on 16S rRNA Gene Sequence Data", which have shown that DNA chimera artifacts reach >30% after only 35-cycles of PCR using the same PCR enzyme that authors employed in this work (i.e. Q5 from NEB).

In the introduction part authors build a case for simple-to-use bacteria sequencing, yet the technique that is being presented is not that simple and will not be easily adopted by microbiology community as it relies on multi-step microfluidics and expertise that requires special training and skills unavailable in the mol. biology laboratories.

There seems to be an over-statement here: "The highlighted applications are only possible using DoTA-seq and represent just a few of many possible future applications of this workflow." In fact the technique reported by the same authors Lan et al., 2017 should allow to achieve the same if not better results than DoTA-Seq given the use of the right off-shelf reagents.

The Reviewer #1 pointed out the concerning discrepancy between the initial distribution of the cells and experimental results obtained. I agree with reviewer and I believe that results indicate the biases and limited analytical resolution of the DoTA-Seq. In rebuke, the authors attribute this discrepancy to the errors of OD600 measurements stating that "The relationship between OD600 and cell number varies across species (Venturelli et al. 2018)." While I agree that bacteria shape and size may affect OD600 measurements yet the differences introduced by phenotypic changes of the cells do not affect

the OD600 measurements that much as is observed in the results. Moreover, the cited work confirms that OD600 can be used as reference. Venturelli et al states that their model relies on OD600 measurements directly contradicting authors claims. "Our model was trained on absolute abundance estimated from OD600 measurements and used to predict absolute abundance based on OD600" In the same work authors normalized species abundance to OD600; "Absolute species abundance was normalized to the monospecies maximum OD600 value". In short, the explanation that OD600 measurements in this work introduced errors that cause cell capture deviation is unconvincing.

Regarding Supplementary Figure 2.
The authors used ZymoBIOMICS Microbial Community standard to evaluate DoTA-Seq analytical capabilities. Initially I was very encouraged to see this effort. Unfortunately, the results by DoTA-Seq showed large discrepancy between the standard and obtained result. The vendor provides informative data sheet showing that the ZymoBIOMICS community standard is carefully calibrated population of cells comprising "three easy-to-lyse bacteria, five tough- to-lyse bacteria, and two tough-to-lyse yeasts."
DoTA-Seq could not accurately determine the composition of such community.

Regarding Supplementary Figure 8.
The methodological details of spike-in experiment are not revealed (e.g. were the spike-in cells exposed to non-ionic detergents to limit their clumping?). At which step the spike-in was added in feces sample (I can only assume it was done after all purification steps in order to limit the cell loss). Also, if the technique cannot discover species below 3% abundancy the whole story about throughput becomes of little value. The advantage of throughput is clear for discovery of rare cells; thus 0.1-1% spike-in experiments are missing. Also, the correct experiment would be to take the bacteria community of known composition and number of cells (e.g. reference community ZymoBIOMICS), prepare along along with spike-in cells and measure their abundance after completing the workflow. In this context the claim that "…DoTA-seq experiments can be used to reliably measure relative changes between samples" remains unsubstantiated.

Regarding Figure 4. It needs to be clarified whether replicates were technical or not. More specifically, where it was the same purified samples encapsulated once on the first microfluidics device and then split into three replicates, or whether it was three encapsulations of the same purified sample using the first microfluidics device, or whether it was the one feces sample, split into three tubes, encapsulated individually and processed through 2nd microfluidics chip separately. If this was the latter then that's great because I only the latter scenario reflects the most relevant accuracy of the technique. If results present one of the former two scenarios, then the technique introduces way too much technical variation. This should be clearly explained in the text (material and methods or in supplementary note).
Also, irrespectively where this work will be published Y axis should had been in linear scale, not log. As you may know the log scale often confuses the readers by giving a visually false impression as if there is a small variability between data points, which in fact variability might very large. Also, if I understood the symbols correctly some species (e.g. Propionibacteria, Peptostreptococcus, Clostridia) are not captured at all in some replicate experiments. It might be worth mentioning it.

I also share concerns by 1st reviewer regarding amplification efficiency of single DNA barcodes and the error rate. Authors use a primer with random sequence N15 as their barcode, but it is not explained what measures they take to identify and remove the spurious barcodes that will be

generated during PCR? Given that barcode amplification involves 40-cycles of PCR the probability of generating spurious barcodes will increase exponentially.

Supplementary Note 3 should be removed as it does disservice by promoting certain companies but not mentioning others providing the same services.

Regarding Supplementary Note 4 and statements in the article that are related to this note. It seems that authors provide a theoretical model that does not reflect DoTA-Seq real-world analytical features, because i) the model does not take into consideration biased cell encapsulation (as shown in the manuscript there is a bias towards gram positive cells) and ii) theoretical model overestimates the detection efficiency due to negligence of co-encapsulation events (beads + DNA).

DISCUSSION

Line 489: How can one call a technique "robust" if it can only process lab grown cultures, while introducing biases towards gram positive cells on real-world samples? Also, the claim that technique is "widely applicable " is arguable and should be removed.

Line 491: "Encapsulation into hydrogels enables multi-step procedure for efficient lysis of diverse bacteria and simplification of microfluidics steps." Are there advantages of DoTA-Seq over capsule-based approaches, which seemed to require only one microfluidic chip (10.1039/D0LC00660B; 10.1073/pnas.2109430119)?

Line 493: Misleading statement. My understanding is that it may take 30 min to encapsulate cells on the first microfluidics chip, but the technique is not finished during the first encapsulation. The process continues with a second chip and subsequent sample processing. The hands-on time is not 30 min but hours.

Line 498-510: I would recommend removing this section as this would represent sophisticated engineering solution that will be difficult to implement in the non-microfluidics labs. Also, having to reinject to hydrogel bead solutions might reduce the throughput and accuracy of the technique.

Line 514: Again misleading. DoTA-Seq relies on one droplet (bead) making device and one bead reinjecting device, the latter being more difficult to use for non-experts, will require tweaking of the flow rates, might clog due to bead polydispersity, dust etc. It is misleading to hear that DoTA-Seq "requires only droplet making steps".

Minor comments.

I assume the authors did analysis correctly but just to be sure the authors should indicate whether or not the results presented in Figure 2d and 2e were generated using identical parameters, thresholds and filtering settings; and if not, then present results that are obtain when using identical parameters for data analysis.

The scale bar in panel a shows bacteria containing bead larger than 20 μm, but in panel C the beads are smaller than 20 μm.

Microfluidics device fabrication be shorten to one sentence by citing most relevant reference, as this seems to be a standard approach for manufacturing.

Line 681: you should explain how many productive droplets (cell + barcode) do you obtain in one PCR tube; once "the PCR mix runs out".

Line 687: To my understanding droplet coalescence would create false-positive events. How to ensure that all coalesced droplets are removed?

Line 387: End-point PCR-driven detection is not a direct or quantitative method of measure. The FISH probes or qPCR would be. The sentence should be corrected: "DoTA-seq directly and quantitatively measures genes of interest within single cells"

Line 426: It seems it was a bad choice to use DNA/RNA shield reagent if it lyses the cells. The correct experiment should be done under conditions where gram negative cells in synthetic standard are not lysed. Zymo research provide relevant guidelines "HOW TO ESTABLISH AN ACCURATE MICROBIOMIC WORKFLOW IN YOUR LAB". I don't see a logical reason for using reagents that lyse the cells prematurely.

Line 433: In my opinion statement that "The relative abundances of species were consistent across three replicates (Fig 4b)." is not really supported by the data presented in Figure 4; where Y-axis is presented in log scale giving distorted visual impression that abundance is consistent.

In conclusion, with all due respect, the revised version has not changed my opinion. Rather the opposite; it convinced me that this technique will hardly find a broad use among microbiology community.

Although we cannot publish your paper, it may be appropriate for another journal in the Nature Portfolio. If you wish to explore the journals and transfer your manuscript please use our [REDACTED]manuscript transfer portal. You will not have to re-supply manuscript metadata and files, but please note that this link can only be used once and remains active until used. For more information, please see our manuscript transfer FAQ page.

Note that any decision to opt in to In Review at the original journal is not sent to the receiving journal on transfer. You can opt in to In Review at receiving journals that support this service by choosing to modify your manuscript on transfer. In Review is available for primary research manuscript types only.

Author Rebuttal, third revision:

Reviewer #1:

Remarks to the Author:

The two major flaws from the previous version were the overstatement of the throughput of the method and the lack of validation on actual samples. Despite the effort from the authors, the added justification and added experiments are not convincing of the method. (If anything, they go further in raising questions).

Response: We have provided validation on numerous samples including environmental samples (including a dedicated main figure to this) in the previous iteration. In order to provide maximum clarity to our method, in this iteration, we have provided even more data and removed some previous data that was potentially confusing to the reviewer. Briefly, it seemed that the reviewer was confused about the Zymo mock community and the Zymo Fecal community (which were two different communities). In our revised manuscript, we have carried out new experiments. Specifically, we replaced the data from the Zymo mock community with data from a new in-house mock community (**Fig. 1f**). This new community was cryopreserved and single-cell counted so as to provide the best true estimates of the community composition for the purposes of validating DoTA-seq. We further included a mouse fecal sample to represent yet another complex environmental sample for demonstration of DoTA-seq. At this point, we have now gone above and beyond to provide numerous examples of DoTA-seq on different samples. We respectfully request that the reviewer consider all of the data that we have shown in the manuscript that supports the utility and accuracy of DoTA-seq.

The throughput of the method is still low (the comparison with 10X is not a fair comparison given that 10X is fully profiling cells). 10,000 cells on a method that cannot be easily performed for many samples is not high-throughput. 10,000 cells (at best) is also not sufficient for quantitative measurements of gene abundances in communities, especially of low abundant organisms, and most especially, with the low sensitivity rates they have. They claim it is highly parallelized and it is not at all in their rendition. With real samples, there is likely to be clumping etc. that further limit the potential to parallelize this.

Response: We recognize that the reviewer believes that high parallelization and even higher throughput is desirable in a workflow. We want to emphasize that we have demonstrated throughputs of >38,000 cells per run in this manuscript. We have amended the manuscript to give precise and detailed numbers regarding the throughput. In addition, we have provided the total number of cells sequenced for all data in this paper in the respective figures (n = x cells) in Supplementary Table 2. We have not claimed that DoTA-seq is highly parallelized. Nevertheless, DoTA-seq has already found use in multiple projects in both our and collaborators' labs, as well as resulting in a new publication that was recently accepted (Lan, F and Saba J et al. Sci Adv 2023; PMID 37540747). Due to the relatively simple droplet making nature of this

workflow, a microfluidics-free implementation will become possible in the future based on particle-templated emulsification (PMID 30033717). We believe that DoTA-seq could adapted to become parallelized using this method but this is beyond the scope of our paper.

Rather than use a real environmental sample, the authors use a commercial product (which seems to be a mock community as well) and has some problems with storage conditions. This does not allow them to answer questions about lysis in a mixed community.

Response: The ZymoBIOMICS sample is a real environmental sample composed of human fecal matter directly preserved in DNA/RNA shield, which prematurely lysed the gram-negatives. As such, we were unable to assess lysis efficiency on gram negative bacteria. Gram negative bacteria can be easily lysed and thus this is not a detriment to DoTA-seq. In this revision, we have carried out substantial experimental and computational analyses using an in-house generated mock community (**Fig. 1f**), as well as a mouse fecal sample (**Fig. 3b,c**) to show that we are able to sequence complex microbiome environmental samples. In total, we have sequenced 4 additional microbial samples of varying origins to demonstrate DoTA-seq's ability for single cell sequencing. Therefore, we believe that we have presented strong evidence that DoTA-seq can be used to analyze a wide range of microbial communities.

Most importantly, the improvement the authors suggest is including a barcode to identify specific cells. Yet, in their experiments on the commercial product shows that the sensitivity of their method goes down (60% average for the essential gene which should be 100%). Therefore, this negates any ability for the barcode to be quantitative! The authors suggest the false positives is the limit of detection, but it is this barcoding problem that is central to their improvements that does not seem to be working well.

Response: We thank the reviewer for this question. As stated in the manuscript, the detection rate could be improved by optimizing primer capture efficiency. We did not optimize the primer capture efficiency for this particular experiment. We have carried out new experiments that demonstrate ~70-90% detection for essential genes (**Supplemental Figure 2**). Variation in prime capture efficiency will indeed affect the quantitative readout of DoTA-seq. However, as explained in the manuscript (lines 183-191), the capture efficiency does not need to be 100% to reliably decipher relative changes between conditions (**Fig 2**). Furthermore, we demonstrate that using DoTA-seq to analyze sequence variation rather than loci presence/absence can provide quantitative information (**Fig. 4**).

would not recommend this method to be used in my lab.

Response: We have already used DoTA-seq to gain key biological insights into microbial populations (DOI: 10.1126/sciadv.adg5476 and PMID 37540747), and currently have additional ongoing projects in 3 different collaborating labs utilizing DoTA-seq to study phage-bacterial interactions, horizontal gene transfer in a human gut pathogen and phase variation in human gut bacteria.

Reviewer #2:

Remarks to the Author:

This is an excellent paper that should be published in Nature Methods as soon as possible. The authors have done a great job addressing the issues raised by all reviewers.

Response: We are highly encouraged by the reviewer's response and thank the reviewer for the approval of our manuscript for publication.

Reviewer #3:

Remarks to the Author:

I appreciate the efforts that authors put in attempting to address the reviewers' concerns. The authors have shown that they can sequence target DNA region in bacteria, in up to 37k cells. The earlier results (provided in the original submission) also included results from 25 species synthetic community. In updated version the authors also performed DoTA-seq on one stool sample and included two spike-in species analyses. The new results show that DoTA-seq can be used to profile 16S DNA of single bacteria cells but at the cost of reduced accuracy, reduced resolution of cell composition and with biases in cell capture. The authors have supplemented their original draft with Figure 4, Supplementary Figure 2, Supplementary Note 4, etc. It is now more clearly explained the limitations and advantages of the technique. It is evident that the DoTA-Seq levareges the throughput of droplet microfluidics technology to isolate single-bacteria species, which are then gelled into beads, and genes of interest are amplified during indexing (barcoding) PCR reaction. While DoTA-seq may appear as an elegant approach schematically, yet careful analysis of the updated manuscript revealed analytical drawbacks that have significantly dampened my enthusiasm.

Response: We appreciate the reviewer's re-review of our manuscript. We understand that the reviewer has concerns about potential biases in cell capture of DoTA-seq as a method for relative abundance profiling of microbiomes. However, we believe that the reviewer's concerns are largely due to misinterpretations of our manuscript and primary applications of DoTA-seq. Firstly, we emphasize that the purpose of DoTA-seq is not to accurately profile microbial composition (shotgun sequencing and 16S profiling are more suitable for this purpose). Instead, the purpose of DoTA-seq is to provide single-cell genetic sequencing of microbial populations to reveal quantitative insights about the genetic structure of the populations. For example, in **Fig. 4**, we demonstrate the ability to profile phase variation in bacterial populations. In **Fig. 2**, we demonstrate the ability to track antibiotic resistance genes within a complex microbial community. In **Fig. 3**, we demonstrate the ability to track mobile genetic elements in a fecal derived community. These are the novel aspects of DoTA-seq that make it useful beyond relative abundance profiling, which can be done with other available methods.

First, I would like to comment on the rebuke letter. Regarding authors' claim that „single bacteria are statistically less likely to be found at the interface". This cannot be true. Droplets have high surface to volume ratio, meaning that statistically the cells will have higher probability to be closer to the interface than reside in the core of the droplet. Being near the interface will cause genetic material loss during gelation and lysis. Although the work related to bacterial loss in hydrogel beads is known to the users, the public reports on this issue are scarce. However, the authors could refer to these two papers for consideration: Figure 1a in 10.1002/adhm.201200341 and Figure S6 in 10.1039/D0LC00660B. In this context, Supplementary Figure 4 must have some errors in the measurements as the number of encapsulated cells should not increase after gelation (pre-lysis gels).

Response: With respect to the core-shell particles and references provided, we believe that core-shell particles can be very useful for single cell analysis, especially in the context of growing cells inside droplets, and we hope to explore them in the future. We agree that droplets have a higher surface to volume ratio than bulk reaction containers. However, there is still plenty of volume that is far away from the interface for a single bacterial cell to occupy. Our original statement was meant to convey that single bacteria (as compared to cell clusters or cells grown in droplets which occupy a lot more volume, which are the subject of the references provided) are statistically less likely to be found at the interface.

We agree with the reviewer that there is measurement error in Supplementary Figure 4, as our droplet counting methods are not perfect. Occasionally, some droplets do not get detected or get detected multiple times by the droplet counting algorithm. Therefore, the resulting numbers are approximations. However, we meant to convey that our data supported no major loss of encapsulated cells before and after removal of oil, suggesting that cell loss is kept at a minimum. The results of our

control experiments demonstrate that genetic material loss through the hydrogel does not present a substantial problem. We have updated the supplementary figure caption to clarify this point.

The new experiments (Figure 4 and Figure S2) revealed that there are significant biases towards certain species of bacteria during microfluidic operations and workflow processing. For example there was a strong bias towards gram positive cells that can survive sample preparation and encapsulation procedures. Gram negative cells that are more sensitive are poorly captured and become underepresented.

Response: The bias towards gram positive cells in this figure are the result of the sample preservation method of the commercial human fecal microbiome sample rather than the DoTA-seq workflow itself. The workflow for DoTA-seq does not have a strong bias towards gram positive cells. Corroborating this notion, we have carried out new experiments to sequence both a mock community (**Fig. 1f**) and a mouse fecal sample showing no such biases in the representation of gram negative or gram positive species (**Fig. 3**).

When working with a real-world sample authors attempt to put a blame on freezing (i.e. according to the authors freezing the cells causes their lysis prematurely), but then the authors should have validated the technique with a real-world sample without cryo-preservation. In another context authors explain the discrepancy in the results due to the use of DNA/RNA shield reagent.

Response: We have carried out a new experiment to demonstrate the ability of DoTA-seq to capture diverse gram positive and negative bacteria with DoTA-seq with a cryopreserved real-world sample in this revision (**Fig. 3**).

In yet another scenario, authors attempted to profile synthetic community of 25 laboratory-cultivated strains; the results again showed biases, but this time authors attributed the errors to OD600 measurements. When working with reference standard (reference community ZymoBIOMICS), again the DoTA-Seq revealed biases but according to the authors freezing/thawing was the cause, even though it is known that gram negative bacteria cells can withstand multiple cycles of freezing. It appears that DoTA-Seq is operational only on a very limited number of samples making it poorly suitable for an accurate and quantitative research that is needed for microbiology field. I could not identify a convincing indication that the DoTA-Seq technique can provide unbiased profiling of heterogenious bacteria populations.

Response: We have validated DoTA-seq on an in-house mock community (**Fig. 1f**) and the results support that DoTA-seq does not have substantial biases. The previous under-representation was due to the sample storage method of Zymobiomics, not freeze thawing. Due to this issue, we have removed the Zymobiomics mock community from the manuscript. We did not expect the 25-member community experiment to be quantitative since we did not count the individual cells of each species in the sample. We would like to emphasize that the purpose of DoTA-seq is not profiling microbial composition (shotgun sequencing and 16S profiling are more suitable for this purpose). Instead, the purpose of DoTA-seq is to provide single-cell genetic sequencing of microbial populations to reveal insights about the genetic structure of microbiomes. To clarify, we have removed the bar graph of the number of cells for the 25-member community to focus the reader on the gene prevalence rather than relative abundance. This information can now be found in Supplementary Figure 2a.

Concerning fact is that method relies on the use of non-ionic detergents to encapsulate cells. The non-ionic detergents can lyse bacteria cells. Indeed, as one could expect, the authors found that gram negative bacteria were severelly underepresented in their results (Figure S2). As I mentioned above these issues will rise serious concerns to microbiology community because it will always remain questionable which species have been lost (or retained) along the way.

Response: We use very small amounts of non-ionic detergents (0.1% tween-20, 0.1% pluronic) that are frequently used in cell washing protocols, including mammalian cells. We understand that our previous results have generated doubt about potential premature lysis of cells (which is due to DNA/RNA shield and not due to DoTA-seq). To provide further evidence, we carried out a new experiment to sequence an in-house mock community that was preserved by cryopreservation rather than DNA/RNA shield. Our results showed no substantial bias towards gram negative or gram positive bacteria, with all bacterial abundance estimates within approximately 2-fold of the theoretical expectations (Fig. 1f).

While I agree that the method is a high-throughput in the first step (i.e. cell isolation), yet the bead reinjection and encapsulation with DNA molecules has rather limited throughput. The number of samples that can be handled is limited as well. Therefore, the entire story around „throughput“ is not really convincing.

Response: The second encapsulation step uses a combined aqueous flow rate of 400 uL/hr generating droplets at ~33 pL volumes. This corresponds to a droplet making rate of ~3 kHz, which can produce ~1

million droplets in 5 minutes. Considering the typical encapsulation rate of ~1 barcode per 10 drops, and ~1 cell per 10 gels, this yields ~10,000 cells sequenced within ~5 minutes, which is what we typically see in our libraries (we usually run the droplet maker for 6 minutes per library). We believe it is reasonable to consider ~10,000 cells per 5 minutes of droplet making to be high-throughput, considering the 10x Genomics platforms take much longer (even with 8 channels running simultaneously).

Authors should cite the work relevant to single-bacteria genome sequencing:

https://doi.org/10.1016/j.cell.2022.06.040

10.1126/science.abm1483

10.1038/s41564-020-0729-6

Response: We thank the reviewer for updating us on these references and have included them in the manuscript. The last reference was in relation to RNA sequencing, which while important, is not directly relevant to this manuscript.

I could not find experimental evidence proving that perceived combinatorial states (phase variation) of promoters are a real result and not some kind PCR artifact. The authors will agree that it is well known in the microbiology field that PCR generates chimeras and recombination events, especially if the template contains polyndromic sequences. This has been know for decades. See for example: https://doi.org/10.1111/j.1574-6941.1998.tb00500.x

https://doi.org/10.2144/000113219

https://doi.org/10.1093/nar/19.8.1949

https://doi.org/10.1093/oxfordjournals.molbev.a025797

It is also known that when analyzing mock community data, the results strongly deviate from the expected composition with the increasing number of PCR cycles. Considering that the DoTA-Seq technique relies on 40-cycles of PCR, false promoter states or even false 16S species are likely to appear due to accumulating PCR artifacts. Highly relevant and equally concerning is relevant report by Sze et al., titled “The Impact of DNA Polymerase and Number of Rounds of Amplification in PCR on 16S rRNA Gene Sequence Data”, which have shown that DNA chimera artifacts reach >30% after only 35-cycles of PCR using the same PCR enzyme that authors employed in this work (i.e. Q5 from NEB).

Response: We appreciate the reviewer's concern about potential chimeras in effecting our readout of the phase-variation states. Chimeras are indeed a problem in PCR, but they typically affect PCR amplification of complex pools of highly homologous amplicons (e.g. 16S from a metagenomic sample or different alleles of the same gene as indicated by the papers that the reviewer has presented). In droplet PCR, the template is one cell, and therefore PCR chimeras cannot generate new 16S species as a bulk PCR with a complex community would. Therefore, PCR chimeras for 16S are not expected to be a problem for DoTA-seq.

With regards to the phase variation demonstration of DoTA-seq, the 7 different amplicons do not have a high percent identity, and thus should not generate chimeras. For example, if all promoters are in the ON orientation, the maximum percent identity is ~63% between pairs of amplicons . However, we frequently observe promoters in the OFF state, which further reduces any homology between amplicons (for most cells, only 2-3 promoters are oriented in the ON state for more details see our paper DOI: 10.1126/sciadv.adg54). In addition, we use unambiguously mapped reads to the expected template to determine the inversion state of each promoter for each cell. Therefore, even if any chimeras arose, we would have excluded any potential chimeras in our analysis pipelines.

Our studies on *B. fragilis* phase variation using DoTA-seq have been peer-reviewed by experts in phase-variation (manuscript published DOI: 10.1126/sciadv.adg54) and are consistent with controls and previous literature. In this study, we demonstrated that our control strains (recombinase *mpi* deletion and phase locked variants) yielded combinatorial promoter states that were expected based on an independent validation. In sum, chimeras do not present a problem in DoTA-seq.

In the introduction part authors build a case for simple-to-use bacteria sequencing, yet the technique that is being presented is not that simple and will not be easily adopted by microbiology community as it relies on multi-step microfluidics and expertise that requires special training and skills unavailable in the mol. biology laboratories.

Response: The microfluidics employed in this method consists solely of droplet making, which is becoming more widely used in biology labs, not to mention the abundance of commercial off-the-shelf solutions available. We have already transferred the technology to two collaborating labs with no droplet expertise and they have successfully performed DoTA-seq runs on their own. Future developments using particle templated emulsions, which enables droplet making without any microfluidics will enable this technique to be used completely microfluidics-free (https://doi.org/10.1021/acs.analchem.8b01759; PMID 30033717).

There seems to be an over-statement here: “The highlighted applications are only possible using DoTA-seq and represent just a few of many possible future applications of this workflow.” In fact the technique reported by the same authors Lan et al., 2017 should allow to achieve the same if not better results than DoTA-Seq given the use of the right off-shelf reagents.

Response: SiC-seq (the 2017 technique) randomly captures genomic DNA, and thus would have a lower chance of capturing specific loci of interest. In this case, DoTA-seq is both easier to use (less complex microfluidic workflows) and has a higher capture efficiency of loci of interest. These two factors are central to the demonstrated applications. For example, the best coverage of SiC-seq per cell is approximately 10% of the genome. Iit would be difficult to find rare phase variants in the *B. fragilis* populations if only 10% of the cells sequenced contained any information about single CPS promoters (let alone all 7 simultaneously sequenced within the same cell). As another example, to track changes in prevalence in the antibiotic resistance genes would require multiple samples with multiple replicates to be sequenced, which is difficult using a complex workflow like SiC-seq (and other whole-genome single cell sequencing pipelines). Thus, DoTA-seq is indeed the only currently available workflow practically capable of the demonstrated applications in this manuscript. To clarify, we’ve amended the manuscript as follows: “These highlighted applications are efficiently performed using DoTA-seq and represent just a few of many possible future applications of this workflow.”

The Reviewer #1 pointed out the concerning discrepancy between the initial distribution of the cells and experimental results obtained. I agree with reviewer and I believe that results indicate the biases and limited analytical resolution of the DoTA-Seq. In rebuke, the authors attribute this discrepancy to the errors of OD600 measurements stating that “The relationship between OD600 and cell number varies across species (Venturelli et al. 2018).” While I agree that bacteria shape and size may affect OD600 measurements yet the differences introduced by phenotypic changes of the cells do not affect the OD600 measurements that much as is observed in the results. Moreover, the cited work confirms that OD600 can be used as reference. Venturelli et al states that their model relies on OD600 measurements directly contradicting authors claims. “Our model was trained on absolute abundance estimated from OD600 measurements and used to predict absolute abundance based on OD600” In the same work authors normalized species abundance to OD600; “Absolute species abundance was normalized to the monospecies maximum OD600 value”. In short, the explanation that OD600 measurements in this work introduced errors that cause cell capture deviation is unconvincing.

Response: All methods for measuring absolute or relative abundance, such as CFU counting, OD600, metagenomic sequencing, 16S sequencing, or single cell sequencing have unavoidable biases. In the Venturelli et al 2018 paper, OD600 was used in all measurements as well as model training and

predictions, which indicates that any potential biases in composition are consistently propagated in both the model and the data. Any method of absolute abundance quantification would also have biases. In the case of this paper, DoTA-seq provides information about the number of single cells, which is not the same as OD600. Both methods (OD600 or DoTA-seq) will be internally consistent (if we are looking at changes in relative abundance over different conditions, as is usually the case). We have emphasized this point throughout the manuscript.

In addition, to better show internal consistency, we have conducted new experiments with mock communities of hemocytometer counted cells, which match closely with single-cell DoTA-seq data as expected **(Fig. 1f).** We have also removed the cell count data for the 25-member community to highlight the major advantage of DoTA-seq which is to reveal gene prevalence rather than community composition. This data can now be found in Supplementary Table 2a. We have further carried out new experiments to demonstrate that DoTA-seq is accurate in quantifying community composition in **Fig. 1f**.

Regarding Supplementary Figure 2.

The authors used ZymoBIOMICS Microbial Community standard to evaluate DoTA-Seq analytical capabilities. Initially I was very encouraged to see this effort. Unfortunately, the results by DoTA-Seq showed large discrepancy between the standard and obtained result. The vendor provides informative data sheet showing that the ZymoBIOMICS community standard is carefully calibrated population of cells comprising “three easy-to-lyse bacteria, five tough- to-lyse bacteria, and two tough-to-lyse yeasts.”

DoTA-Seq could not accurately determine the composition of such community.

Response: The discrepancies were due to the premature lysis of gram negative bacteria in the storage buffer as we explained in our responses above and the manuscript. We have conducted additional experiments using in-house generated mock communities (which were cryopreserved) to show that DoTA-seq was capable of determining the composition of a community **(Fig. 1f)**. As explained in our previous responses, relative abundance quantification is not the primary function of DoTA-seq (shotgun or 16S sequencing is more well-suited for this purpose). We respectfully ask that the reviewer please consider DoTA-seq on the merits of its usefulness as a single-cell sequencing technique.

Regarding Supplementary Figure 8.

The methodological details of spike-in experiment are not revealed (e.g. were the spike-in cells exposed to non-ionic detergents to limit their clumping?). At which step the spike-in was added in feces sample (I can only assume it was done after all purification steps in order to limit the cell loss).

Response: In the interest of streamlining the paper and avoiding further reader confusion, we have decided to remove the spike-in analysis from the manuscript, as we feel that our new experiments in Fig. 1f and Fig. 3 sufficiently and clearly demonstrate DoTA-seq's ability to detect rare species within microbial communities. Regardless, our explanation of the methodological details are described below for the reviewer's information:

The spike-in procedures were described as in the methods section as follows:

"The cell pellet containing supernatant was transferred to a new tube and washed twice with PBS + 1% Pluronic F68 and resuspended in 100 µL of PBS + 1% Pluronic F68. Next, 100 µL of overnight cultures of *Pseudomonas putida KT2440* and *Staphylococcus epidermidis ATCC 12228* grown in LB and BHI broth (Sigma # 53286) respectively were washed twice in PBS + 1% Pluronic F68 then resuspended in 100 µL of PBS + 1% Pluronic F68. All cells were stained with 10X SYBRGreen and counted on a hemocytometer to estimate cell concentrations. Fecal community and spike-in cells were combined at predetermined ratios and prepared for sequencing with the fecal sample DoTA-seq primer mix"

In summary, the spike-ins were prepared in non-ionic detergents and spiked in after isolation of the fecal cells from the storage buffer. The spike-ins were used to show that different spike-in levels corresponded to the expected DoTA-seq read-outs, indicating the ability to do relative quantification between different samples.

Also, if the technique cannot discover species below 3% abundancy the whole story about throughput becomes of little value. The advantage of throughput is clear for discovery of rare cells; thus 0.1-1% spike-in experiments are missing.

Response: We have demonstrated that DoTA-seq can accurately species below 3% as there are species below 1% in every sample sequenced (Figures 1-4). We further show quantification within 2-fold of theoretical expectations for populations below 1% abundance (Fig. 1f).

Also, the correct experiment would be to take the bacteria community of known composition and number of cells (e.g. reference community ZymoBIOMICS), prepare along along with spike-in cells and measure their abundance after completing the workflow.

Response: Spike-ins are typically done with complex communities as a way to control for consistency of the prep and for absolute quantification of cells. Our interpretation of this suggestion is to spike in a known amount of cells into a defined community. However, this would be equivalent to mixing known quantities of cells since we would already know the abundance of every species in a defined community.

In this context the claim that "…DoTA-seq experiments can be used to reliably measure relative changes between samples" remains unsubstantiated.

Response: We have shown that DoTA-seq can measure relative changes in gene prevalence in Fig. 2e which was corroborated by independent colony picking and single-cell digital PCR experiments (supplemental fig 7). There is further evidence of DOTA-seq for measuring relative changes in subpopulation abundance in our studies of *B fragilis* populations (DOI: 10.1126/sciadv.adg5476, PMID 37540747). Measurements of relative changes in taxonomic composition can be done more easily with 16S or metagenomic sequencing, and is not the primary use of DoTA-seq. We hope the reviewer can appreciate the single-cell sequencing aspects of DoTA-seq that can go far beyond taxonomic characterization.

Regarding Figure 4. It needs to be clarified whether replicates were technical or not. More specifically, where it was the same purified samples encapsulated once on the first microfluidics device and then split into three replicates, or whether it was three encapsulations of the same purified sample using the first microfluidics device, or whether it was the one feces sample, split into three tubes, encapsulated individually and processed through 2nd microfluidics chip separately. If this was the latter then that's great because I only the latter scenario reflects the most relevant accuracy of the technique. If results present one of the former two scenarios, then the technique introduces way too much technical variation. This should be clearly explained in the text (material and methods or in supplementary note).

Response: We have reworked the aforementioned figure for clarity and to reduce confusion and therefore, the spike-ins are no longer mentioned. However, to answer the question, in that experiment (Fig. 4 of previous iteration of manuscript, now Fig. 3), technical replicates were performed over 2 different days (the spike ins were done on one day, the replicate containing >30,000 cells was done on a different day). Each day begins with cell extraction from the fecal suspension in the preservation buffer, which gets split into 3 tubes for parallel processing. Technical replicates within the same day were encapsulated one at a time on different microfluidics chips.

Also, irrespectively where this work will be published Y axis should had been in linear scale, not log. As you may know the log scale often confuses the readers by giving a visually false impression as if there is a small variability between data points, which in fact variability might very large.

Response: Abundances are shown on log scale because the variation in species abundances in natural samples are extremely large. If shown on a linear scale, the entire graph would be overwhelmed by a few large bars and a host of tiny bars. This is a common way to visualize species abundances (e.g., Fig. 3 in PMID 20203603, doi: 10.1038/nature0882) We believe log scale is actually a more appropriate way of displaying the data since the variability displayed is in proportion to fold change (which is more relevant across species at different natural abundances) rather than absolute change. In Fig. 1f, we plotted the data on a linear scale because the dynamic range was not very large compared to Fig. 3b. In Fig. 3b, we would not be able to visualize the majority of taxa detected by DoTA-seq without plotting on log scale.

Also, if I understood the symbols correctly some species (e.g. Propionibacteria, Peptostreptococcus, Clostridia) are not captured at all in some replicate experiments. It might be worth mentioning it.

Response: Some rare species (~0.1% abundance) are not captured in all replicate experiments likely due to stochasticity of sampling near the limit of detection for the number of cells sequenced (Supplementary Figure 4). This is mentioned in lines 234-250 as follows: "The technical variation was small and negatively correlated (Spearman Rho = –0.57, P < 6.5e–10) to the number of sequenced cells for each species (Supplementary Fig. 4). This implies that increasing the total number of sequenced cells can reduce the technical variability observed in DoTA-seq, particularly in low abundance populations."

I also share concerns by 1st reviewer regarding amplification efficiency of single DNA barcodes and the error rate. Authors use a primer with random sequence N15 as their barcode, but it is not explained what measures they take to identify and remove the spurious barcodes that will be generated during PCR? Given that barcode amplification involves 40-cycles of PCR the probability of generating spurious barcodes will increase exponentially.

Response: Spurious barcodes are indeed generated by PCR, but due to the random nature of these errors, they tend to appear as barcode groups of fewer reads. We can clearly separate them from the real

barcodes based on the number of times they appear in the sequencing reads (the barcodes to the left of the vertical line in Supplementary Note 1).

Supplementary Note 3 should be removed as it does disservice by promoting certain companies but not mentioning others providing the same services.

Response: We have removed the mention of specific companies from the manuscript and replaced it with a more general statement that these devices are commercially available.

Regarding Supplementary Note 4 and statements in the article that are related to this note. It seems that authors provide a theoretical model that does not reflect DoTA-Seq real-world analytical features,

because i) the model does not take into consideration biased cell encapsulation (as shown in the manuscript there is a bias towards gram positive cells) and ii) theoretical model overestimates the detection efficiency due to negligence of co-encapsulation events (beads + DNA).

Response: This calculation in Supplementary Note 4 is used to estimate the number of cells that need to be sequenced in order to detect features of interest based on their estimated rarity. It assumes 100% capture efficiency for each target. With a lower capture efficiency, we would have to scale up the estimated number of cells that need to be sequenced in proportion to the inverse of the capture efficiency (i.e. 80% capture efficiency means the estimated number of cells that need to be sequenced would be multiplied by 1/0.8 = 1.25).

There will always be deviations from theory in real life experimentation, but modeling based on a set of assumptions remains useful as a first approximation before any experiments are performed. All models make certain assumptions to answer specific questions and can be useful. i) As explained earlier, there is no bias towards gram positive cells. ii) the theoretical model is meant to provide a starting point for which to design experiments. Co-encapsulation events (we assume the reviewer means co-encapsulation of multiple cells which causes a barcode to be filtered out) are rare according to the Poisson distribution and will not cause the model to deviate significantly from reality.

DISCUSSION

Line 489: How can one call a technique “robust” if it can only process lab grown cultures, while introducing biases towards gram positive cells on real-world samples? Also, the claim that technique is “widely applicable “ is arguable and should be removed.

Response: We have demonstrated numerous applications of DoTA-seq (Figures 2-4) and are using it for multiple projects both within our lab and with collaborating labs. As we show repeatedly (figures 4c,d), it is capable of processing natural samples as long as a cell suspension can be extracted from them.

Line 491: “Encapsulation into hydrogels enables multi-step procedure for efficient lysis of diverse bacteria and simplification of microfluidics steps.” Are there advantages of DoTA-Seq over capsule-based approaches, which seemed to require only one microfluidic chip (10.1039/D0LC00660B; 10.1073/pnas.2109430119)?

Response: Gels are simpler and easier to generate at high throughput than capsules. This is because capsules are generated using aqueous two-phase systems, which contain polymers that increase viscosity, making droplet generation more difficult. As far as DoTA-seq is concerned, capsule based approaches should also work. However, there is no comparison as there are no capsule based ultrahigh-throughput single cell sequencing methods published to date. The papers mentioned by the reviewer do not perform high-throughput single cell sequencing using the capsules. Due to its simplicity, we find our gel based approaches to be easy to implement and perfectly suited for the purposes of DoTA-seq.

Line 493: Misleading statement. My understanding is that it may take 30 min to encapsulate cells on the first microfluidics chip, but the technique is not finished during the first encapsulation. The process continues with a second chip and subsequent sample processing. The hands-on time is not 30 min but hours.

Response: We never stated 30 min hands on time. We have only stated that we encapsulated 37,000 cells with 30 min of the bottleneck droplet making step (**Fig. 1c**), which is not a misleading statement, as the droplet making ste is the bottleneck in our workflow. Regardless, this statement has been removed to prevent any potential confusion.

Line 498-510: I would recommend removing this section as this would represent sophisticated engineering solution that will be difficult to implement in the non-microfluidics labs. Also, having to reinject to hydrogel bead solutions might reduce the throughput and accuracy of the technique.

Response: This comment was in response to a section describing the use of hydrogel bead barcodes to improve Dota-seq**.** Our response is that reinjection of hydrogel beads is already implemented in DoTA-seq (cell genomes are encapsulated in hydrogel beads which are reinjected), and thus additional reinjection of hydrogel bead barcodes does not require sophisticated solutions to implement. We have plans to implement these workflows in future iterations of this method, including methods to do so without using any microfluidics devices (lines 509-514).

Line 514: Again misleading. DoTA-Seq relies on one droplet (bead) making device and one bead reinjecting device, the latter being more difficult to use for non-experts, will require tweaking of the flow rates, might clog due to bead polydispersity, dust etc. It is misleading to hear that DoTA-Seq "requires only droplet making steps".

Response: We have amended the manuscript to state droplet makers and gel bead reinjectors. Gel Bead reinjection is basically a droplet maker that has a channel for gel beads and adds only marginal complexity to the device (we load gels into the syringe instead of liquid). You do not have to tweak the flow rates, but can just use the flow rates provided in the protocol. The beads generated in our DoTA-seq workflow are not polydisperse. Dust clogs are a problem in all microfluidic workflows and simply require careful handling of reagents. The fact that multiple non-microfluidics researchers have successfully adopted DoTA-seq into their research is evidence that bead reinjection is not difficult for non-experts.

Minor comments.

I assume the authors did analysis correctly but just to be sure the authors should indicate whether or not the results presented in Figure 2d and 2e were generated using identical parameters, thresholds and filtering settings; and if not, then present results that are obtain when using identical parameters for data analysis.

Response: Yes, the results in figure 2d and 2e are performed using the exact same script with identical parameters.

The scale bar in panel a shows bacteria containing bead larger than 20 µm, but in panel C the beads are smaller than 20 µm.

Response: A close (zoomed in) inspection shows that the error bars are consistent with the gel sizes in both panels. The perceived discrepancy might be an optical illusion. We've reduced the size of the vertical ends of the scale bars to make it easier to compare.

Microfluidics device fabrication be shorten to one sentence by citing most relevant reference, as this seems to be a standard approach for manufacturing.

Response: We appreciate when authors provide brief summaries of methods used to allow easier interpretation by readers.

Line 681: you should explain how many productive droplets (cell + barcode) do you obtain in one PCR tube; once "the PCR mix runs out".

Response: We have added the requested details (line 629).

Line 687: To my understanding droplet coalescence would create false-positive events. How to ensure that all coalesced droplets are removed?

Response: Coalesced droplets generate unique signatures in the sequencing results and thus can be removed which is discussed in Supplementary Note 2.

Line 387: End-point PCR-driven detection is not a direct or quantitative method of measure. The FISH probes or qPCR would be. The sentence should be corrected: “DoTA-seq directly and quantitatively measures genes of interest within single cells”

Response: We agree that end-point PCR detection is not a quantitative method of quantifying genes. However, we do not believe we ever made this statement and neither were we able to find the text that the reviewer is referring to. As a result, we could not make the requested corrections.

Line 426: It seems it was a bad choice to use DNA/RNA shield reagent if it lyses the cells. The correct experiment should be done under conditions where gram negative cells in synthetic standard are not lysed. Zymo research provide relevant guidelines “HOW TO ESTABLISH AN ACCURATE MICROBIOMIC WORKFLOW IN YOUR LAB”. I don’t see a logical reason for using reagents that lyse the cells prematurely.

Response: We were originally told by the Zymo sales reps that it did not lyse the gram negative cells prematurely. Unfortunately, that was untrue and we found out only after doing single cell sequencing. However, we originally believed that it would not detract from the central message of our manuscript. We have carried out new experiments using an in-house generated mock communities to avoid this problem (**Fig. 1f**).

Line 433: In my opinion statement that “The relative abundances of species were consistent across three replicates (Fig 4b).” is not really supported by the data presented in Figure 4; where Y-axis is presented in log scale giving distorted visual impression that abundance is consistent.

Response: We have modified the statement to emphasize the differences in the relative abundances. However, we will keep the figure in log axis because due to the large distribution of relative abundances, a linear scale will be dominated only by the very abundant taxa. The main point of the figure is to show that the expected taxa based on 16S sequencing are also detected by DoTA-seq. This way of representing microbial community abundance is not unprecedented. For example, see Fig. 3 in PMID 20203603.

In conclusion, with all due respect, the revised version has not changed my opinion. Rather the opposite; it convinced me that this technique will hardly find a broad use among microbiology community.

Response: We respect the reviewer's opinion, but we believe it is based on a mis-understanding of the method. We have thus attempted to make the paper easier to understand by restructuring and substantially re-writing the paper in addition to providing new data and analyses. DoTA-seq is currently in use by both our lab and collaborator labs by personnel with no microfluidics experience for microbial research. It has already resulted in a separate publication examining population dynamics in phase variation (DOI: 10.1126/sciadv.adg5476). Evidently, we believe that it is a extremely useful tool for the microbial research community both by itself and as a foundational platform for developing more microbial single-cell analysis tools in the future.

Decision Letter, fourth revision:

21st Mar 2023

Dear Professor Venturelli,

Thank you for your letter asking us to reconsider our decision on your Article, "Massively parallel single-cell sequencing of genetic loci in diverse microbial populations". We very much appreciate your patience when the editorial team evaluate and discuss this appeal. After careful consideration we have decided that we are willing to consider a revised version of your manuscript that addresses the reviewers' concerns.

When revising your paper:

* include a point-by-point response to our referees and to any editorial suggestions

* please underline/highlight any additions to the text or areas with other significant changes to facilitate review of the revised manuscript

* address the points listed described below to conform to our open science requirements

* ensure it complies with our general format requirements as set out in our guide to authors at www.nature.com/naturemethods

* resubmit all the necessary files electronically by using the link below to access your home page

[REDACTED]
Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised paper within 6 months. If you cannot send it within this time, please let us know. In this event, we will still be happy to reconsider your paper at a later date so long as nothing similar has been accepted for publication at Nature Methods or published elsewhere.

OPEN SCIENCE REQUIREMENTS

REPORTING SUMMARY AND EDITORIAL POLICY CHECKLISTS
When revising your manuscript, please submit reporting summary and editorial policy checklists.

Reporting summary: https://www.nature.com/documents/nr-reporting-summary.zip
Editorial policy checklist: https://www.nature.com/documents/nr-editorial-policy-checklist.zip

If your paper includes custom software, we also ask you to complete a supplemental reporting summary.

Software supplement: https://www.nature.com/documents/nr-software-policy.pdf

Please submit these with your revised manuscript. They will be available to reviewers to aid in their evaluation if the paper is re-reviewed. If you have any questions about the checklist, please see http://www.nature.com/authors/policies/availability.html or contact me.

Please note that these forms are dynamic 'smart pdfs' and must therefore be downloaded and completed in Adobe Reader. We will then flatten them for ease of use by the reviewers. If you would like to reference the guidance text as you complete the template, please access these flattened versions at http://www.nature.com/authors/policies/availability.html.

IMAGE INTEGRITY
When submitting the revised version of your manuscript, please pay close attention to our href="https://www.nature.com/nature-research/editorial-policies/image-integrity">Digital Image Integrity Guidelines. 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.

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.

DATA AVAILABILITY
Please include a "Data availability" subsection in the Online Methods. This section should inform readers about the availability of the data used to support the conclusions of your study, including accession codes to public repositories, references to source data that may be 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", describing which data is available upon request and mentioning any restrictions on availability. If DOIs

are provided, please include 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

MATERIALS AVAILABILITY
As a condition of publication in Nature Methods, authors are required to make unique materials promptly available to others without undue qualifications.

Authors reporting new chemical compounds must provide chemical structure, synthesis and characterization details. Authors reporting mutant strains and cell lines are strongly encouraged to use established public repositories.

More details about our materials availability policy can be found at https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#availability-of-materials

SUPPLEMENTARY PROTOCOL
To help facilitate reproducibility and uptake of your method, we ask you to prepare a step-by-step Supplementary Protocol for the method described in this paper. We encourage authors to share their step-by-step experimental protocols on a protocol sharing platform of their choice and report the protocol DOI in the reference list. Nature Portfolio 's Protocol Exchange is a free-to-use and open resource for protocols; protocols deposited in Protocol Exchange are citable and can be linked from the published article. More details can found at www.nature.com/protocolexchange/about.

ORCID
Nature Methods 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 please visit www.springernature.com/orcid.

We look forward to hearing from you soon. Thank you again for the opportunity to consider this paper.

Sincerely,

Lin Tang, PhD
Senior Editor
Nature Methods

Author Rebuttal, fourth revision:

[There is no rebuttal letter at this stage.]

Decision Letter, fifth revision:

Our ref: NMETH-A48893E

16th Oct 2023

Dear Dr. Venturelli,

Thank you for your letter explaining the potential overlap between your submission and your recent publication in Sci Adv. Given the fact that these two papers have small portion of overlapping contents and the submission to Nature Methods has a focus on DoTA-seq, we'll be happy in principle to publish it in Nature Methods, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

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

TRANSPARENT PEER REVIEW
Nature Methods offers a transparent peer review option for new original research manuscripts submitted from 17th February 2021. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. **Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't.** Failure to state your preference will result in delays in accepting your manuscript for publication.

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

ORCID
IMPORTANT: Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, please let your co-authors know that if they wish to have their ORCID added to the paper they must follow the procedure described in the following link prior to acceptance:
https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

Thank you again for your interest in Nature Methods. Please do not hesitate to contact me if you have any questions. We will be in touch again soon.

Sincerely,
Lei

Lei Tang, Ph.D.
Senior Editor
Nature Methods

Reviewer #1 (Remarks to the Author):

The authors have added substantially more information to the manuscript showing its limit of detection in complex communities. Thank you for doing these additional experiments.

I'm not sure why the authors removed the spike in experiment or didn't include a new spike in experiment since that is good for assessing ground truth, which is more difficult in a mixed community.

The figures in Sup. Fig 6 and Note 3 are hard to understand-- is note 3 log or linear scale? Some numbers on Sup Fig 6 would be helpful. It seems like if you don't have the 16S, the cell is non informative, so shouldn't the statistics be according to the number of total cells with 16S genes? Some better accounting of the cells that are actually able to be analyzed at the end would be helpful. The number of cells analyzed for example in Fig 2a (not just percentages) would be helpful.

Reviewer #2 (Remarks to the Author):

I remain very excited about this work and believe that it needs to be published in Nature Methods.

I have carefully read through the very detailed comments of the other reviewers as well as the excellent replies by the authors. It is clear that the authors have gone above and beyond in terms of additional experimental and analysis workflows and have more than addressed all the issues raised.

Reviewer #3 (Remarks to the Author):

The fact that authors have already published the same method in Science Advances journal, changes situation completely. Perhaps I am wrong, but it does appear that the authors submitted two papers with with the same technique (DoT-Seq); one to Nature Methods and one to Science Advances. I read Science Advances article, and it led to me to conclude that the manuscript submitted to Nature Methods has nothing outstanding to offer. It is conceptually the same method as the one already published. While one can argue that in the work submitted to Nature Methods authors put more efforts in single-cell *targeted* gene sequencing , yet the “meat” of both articles is the same technique, and the study of promoter inversions in bacteria.

Please understand me correctly, but I could find a good reason to publish two salami-slices in high impact journals.

In short, with all due respect, given the recent Fan et al., article published in Science Advances, I concluded that the manuscript submitted to Nature Methods does not bring enough methodological and technological novelty to the field, and therefore in my opinion it is not suitable for a high-impact journal such as Nature Methods.

Author Rebuttal, fourth revision:

Reviewer #1:

Remarks to the Author:

The authors have added substantially more information to the manuscript showing its limit of detection in complex communities. Thank you for doing these additional experiments.

s

I'm not sure why the authors removed the spike in experiment or didn't include a new spike in experiment since that is good for assessing ground truth, which is more difficult in a mixed community.

Response: We removed the analysis of the previous spike-in experiments as part of an effort for overall streamlining of the manuscript for ease-of-reading and clarity of the message. We prioritized experiments that were central to the main message of the paper while being clear and easy to interpret. We believe that the mixed community experiment demonstrates that DoTA-seq can reliably profile diverse bacteria at the single-cell level within multi-species communities.

The figures in Sup. Fig 6 and Note 3 are hard to understand-- is note 3 log or linear scale? Some numbers on Sup Fig 6 would be helpful. It seems like if you don't have the 16S, the cell is non informative, so shouldn't the statistics be according to the number of total cells with 16S genes? Some better accounting of the cells that are actually able to be analyzed at the end would be helpful. The number of cells analyzed for example in Fig 2a (not just percentages) would be helpful.

Response: The scales in sup fig 6 (now extended figure 3) and note 3 are in log scale on the y axis although it looks very similar plotted on a linear scale. Numbers on the axes were not included because we have

mostly been using these figures as a qualitative check of the relative target amplification balance. The number on the Y axis depends on the number of cells sequenced and the number on the X axis ranges from 0-1, representing the fraction of reads mapping to a specific gene (the 16S species marker gene in case of sup fig 6 - now extended data fig 3).

Indeed, if you don't have 16S gene, then the cell would not be informative and thus not counted in the final cell count (all cell counts given in all figures are cells with 16S gene information, except in the *B. fragilis* figure now extended figure 9). We thank the reviewer for pointing out this oversight and have amended the figures and captions to better explain the axes as well as the purpose of the figure.

Reviewer #2:

Remarks to the Author:

I remain very excited about this work and believe that it needs to be published in Nature Methods.

I have carefully read through the very detailed comments of the other reviewers as well as the excellent replies by the authors. It is clear that the authors have gone above and beyond in terms of additional experimental and analysis workflows and have more than addressed all the issues raised.

Response: We thank the reviewer for the valuable feedback and we are glad that the reviewer sees that we have gone above and beyond to address all issues raised to improve the manuscript.

Reviewer #3:

Remarks to the Author:

The fact that authors have already published the same method in Science Advances journal, changes situation completely. Perhaps I am wrong, but it does appear that the authors submitted two papers with with the same technique (DoT-Seq); one to Nature Methods and one to Science Advances. I read Science Advances article, and it led to me to conclude that the manuscript submitted to Nature Methods has nothing outstanding to offer. It is conceptually the same method as the one already published. While one can argue that in the work submitted to Nature Methods authors put more efforts in single-cell *targeted* gene sequencing , yet the "meat" of both articles is the same technique, and the study of promoter inversions in bacteria.

Please understand me correctly, but I could find a good reason to publish two salami-slices in high impact journals.

In short, with all due respect, given the recent Fan et al., article published in Science Advances, I concluded that the manuscript submitted to Nature Methods does not bring enough methodological and technological novelty to the field, and therefore in my opinion it is not suitable for a high-impact journal such as Nature Methods.

Response: We assume the reviewer is talking about the article we recently published (Lan et al., Science Advances) which used computational modeling and single-cell sequencing to understand the combinatorial phase variation dynamics of *B. fragilis.* There is minimal overlap between the Science Advances paper, which focuses on a biological question and this manuscript which describes a very useful technique and provides many examples of its use.

The method that was published in Lan F. et al. Sci Adv focuses on a **specific biological question about the regulation of phase variation in *Bacteroides fragilis* (*B. fragilis*) in substantial depth using computational modeling and experiments**. Therefore, Lan F. et al. Sci Adv is a biology focused paper and not a methods paper. By contrast, NMETH-A48893E has multiple case studies using different synthetic and natural microbiomes containing diverse gram negative and gram positive bacteria (Figures 1-3) and the methods of DoTA-seq used in these case studies is notably different (see below for details). In these microbial community and microbiome case studies, we demonstrate the ability to link the 16S gene to a wide range of different target genes including antibiotic resistance and mobile element genes using natural and synthetic communities.

In sum, all other methods in NMETH-A48893E (Figures 1-3 and corresponding SI figures) have not been published previously and do not directly overlap with the Lan F. et al. Sci Adv.

We provide more specific details below:

1. Lan F. et al. Sci Adv focuses on the biological phenomenon of regulation of combinatorial phase variation in a single gram negative bacterial population. NMETH-A48893E is focused on a general method of multiplexed single-cell targeted sequencing in diverse microbiomes.

2. Lan F. et al. Sci Adv does not mention the broad range of applications of targeted single cell sequencing. It is not possible to use the methods described in Lan F. et al. Sci Adv to achieve broad application of DoTA-seq for diverse microbiomes.
3. The method used in the Lan F. et al. Sci Adv paper is not the full DoTA-seq workflow as described in the DoTA-seq paper. It is a single cell workflow specifically for *B. fragilis*. It would not work for any of the other use cases demonstrated in the Nature Methods paper. The specific methodological differences are as follows:
    - NMETH-A48893E uses gels and enzyme cocktails to lyse diverse microbes, and composes of two microfluidic workflows. The Lan F. et al. Sci Adv method uses single cell sequencing technique in a one-pot reaction that only works on easy to lyse, single species populations of gram negative bacteria. The primer design is also substantially different in NMETH-A48893E than in Lan F. et al. Sci Adv.
4. In NMETH-A48893E, we provide substantial discussion on characterizing various aspects of broadly applicable method of DoTA-seq, which enables readers to apply our method to a large range of applications.

In summary, our manuscript provides a focused discussion and validation on a truly generalizable and flexible single-cell sequencing workflow (DoTA-seq) while Lan F. et al. Sci Adv describes the biology of combinatorial phase variation in *B. fragilis*. In our opinion, the contents of this manuscript are intended for a broad audience interested in single cell sequencing of any microbes and have not been overlapped in any significant way with the Lan F. et al. Sci Adv, and therefore, warrants publication in Nature Methods. Finally, to clarify, the data in NMETH-A48893E is distinct from the data in Lan F. et al. Sci Adv.

Final Decision Letter:

17th Dec 2023

Dear Dr Venturelli,

I am pleased to inform you that your Article, "Massively parallel single-cell sequencing of diverse microbial populations", has now been accepted for publication in Nature Methods. The received and accepted dates will be 24th Apr 2022 and 17th Dec 2023. This note is intended to let you know what to expect from us over the next month or so, and to let you know where to address any further questions.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see https://www.nature.com/natsustain/info/gta). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our website).

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Methods style. 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.

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.

Please note that Nature Methods is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

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 Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including self-archiving policies. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

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

Your paper will now be copyedited to ensure that it conforms to Nature Methods style. Once proofs are generated, they will be sent to you electronically and you will be asked to send a corrected version within 24 hours. It is extremely important that you let us know now whether you will be difficult to contact over the next month. If this is the case, we ask that you send us the contact information (email,

phone and fax) of someone who will be able to check the proofs and deal with any last-minute problems.

If, when you receive your proof, you cannot meet the deadline, please inform us at rjsproduction@springernature.com immediately.

Once your manuscript is typeset and you have completed the appropriate grant of rights, 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.

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

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

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.

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.

Nature Portfolio journals encourage authors to share their step-by-step experimental protocols on a protocol sharing platform of their choice. Nature Portfolio 's Protocol Exchange is a free-to-use and open resource for protocols; protocols deposited in Protocol Exchange are

citable and can be linked from the published article. More details can found at www.nature.com/protocolexchange/about.

Please note that you and any of your coauthors will be able to order reprints and single copies of the issue containing your article through Nature Portfolio's reprint website, which is located at http://www.nature.com/reprints/author-reprints.html. If there are any questions about reprints please send an email to author-reprints@nature.com and someone will assist you.

Please feel free to contact me if you have questions about any of these points.

Best regards,
Lei

Lei Tang, Ph.D.
Senior Editor
Nature Methods