Assessing the Resolution of Methyltransferase-Mediated DNA Optical Mapping

L. d'Huys, Raffaele Vitale, E. Ruppeka-Rupeika, V. Goyvaerts, Cyril Ruckebusch, J. Hofkens

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Assessing the Resolution of Methyltransferase-Mediated DNA Optical Mapping

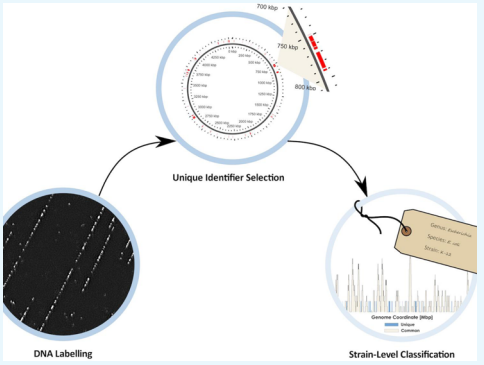
Laurens D'Huys,[‖] Raffaele Vitale,[‖] Elizabete Ruppeka-Rupeika, Vince Goyvaerts, Cyril Ruckebusch, and Johan Hofkens*

Cite This: *ACS Omega* 2021, 6, 21276–21283

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Interest in the human microbiome is growing and has been, for the past decade, leading to new insights into disease etiology and general human biology. Stimulated by these advances and in a parallel trend, new DNA sequencing platforms have been developed, radically expanding the possibilities in microbiome research. While DNA sequencing plays a pivotal role in this field, there are some technological hurdles that are yet to be overcome. Targeting of the 16S rRNA gene with amplicon sequencing, for instance, is frequently used for sample composition profiling due to its short sample-to-result time and low cost, which counterbalance its low resolution (genus to species level). On the other hand, more comprehensive methods, namely, whole-genome sequencing (WGS) and shallow shotgun sequencing, are capable of yielding single-gene- and functional-level resolution at a higher cost and much higher sample processing time. It goes without saying that the existing gap between these two types of approaches still calls for the development of a fast, robust, and low-cost analytical platform. In search of the latter, we investigated the taxonomic resolution of methyltransferase-mediated DNA optical mapping and found that strain-level identification can be achieved with both global and whole-genome analyses as well as using a unique identifier (UI) database. In addition, we demonstrated that UI selection in DNA optical mapping, unlike variable region selection in 16S amplicon sequencing, is not limited to any genomic location, explaining the increase in resolution. This latter aspect was highlighted by SCCmec typing in methicillin-resistant *Staphylococcus aureus* (MRSA) using a simulated data set. In conclusion, we propose DNA optical mapping as a method that has the potential to be highly complementary to current sequencing platforms.



INTRODUCTION

The relationship between the human microbiota and its host has been extensively studied in the past decade due to its strong association with human health and diseases.[1−6] This interest has recently promoted a fundamental paradigm shift in the way our biological complexity is investigated and explained. As an example, the healthy genomic ensemble of the gut microbiota is now considered to be complementary to that of the host, and research has been able to shed light on otherwise inaccessible routes for carbohydrate metabolism, vitamin synthesis, and even regulation of the immune system.[7,8] The presence of forthcoming metabolic products and signaling molecules in distant organs is evident of the enormous influence the gut microbiota can have on human health. Nevertheless, the exact nature and functioning of the interplay between these microbial communities and the host organism are still buried beneath their overwhelmingly large genomic landscape.[9,10]

In this regard, DNA sequencing technologies (also driven by rapid advances in both hardware and data analysis strategies) have played a pivotal role in shaping our current understanding of the human microbiome. When no functional information at the gene level is required, targeting of the 16S rRNA gene with amplicon sequencing has been a mainstay for sample composition profiling due to its short sample-to-result time and low cost compared to that of the far more expensive whole-genome sequencing (WGS).[11] However, there is a very large resolution gap between amplicon sequencing (genus to species level) and WGS (single-gene and functional levels), and the need to close this gap is evident from ongoing technological advances in the sequencing field. More recent third-generation sequencing platforms allow for sequencing of the full 16S gene due to the long read lengths achieved, pushing the resolution toward the strain level.[12,13] Unfortunately, the base-calling error rates of such platforms are still

Received: March 15, 2021
Accepted: June 25, 2021
Published: August 10, 2021

https://doi.org/10.1021/acsomega.1c01381
ACS Omega 2021, 6, 21276−21283

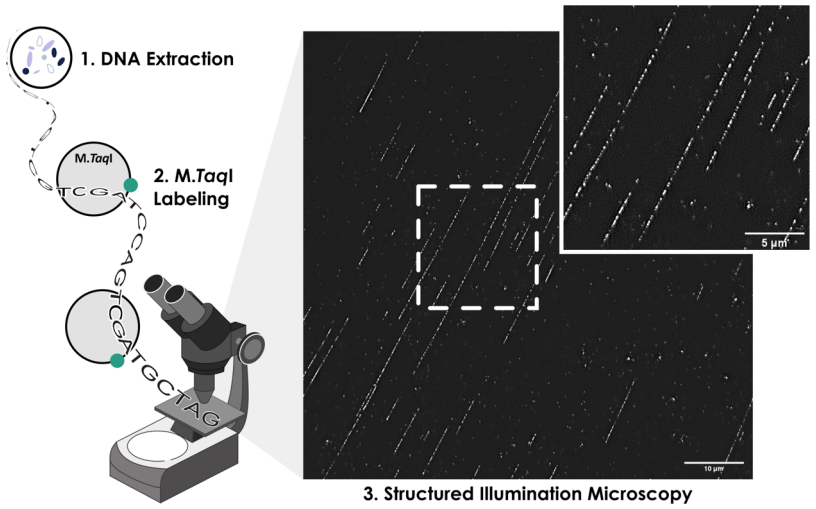


Figure 1. Schematic representation of the methyltransferase-mediated DNA optical mapping workflow. A characteristic SIM image (together with a representative zoom-in inset) extracted from the experimental *E. coli* data set is shown on the right side of the figure.

higher than desired, reducing the accuracy of strain identification.[14−16] On the WGS side of the sequencing spectrum, recent efforts have narrowed the resolution gap with shallow shotgun sequencing.[17] Taken together, these efforts indicate the need for a robust and low-cost platform for longitudinal taxonomic studies.

DNA optical mapping has recently been proposed as an inexpensive and rapid complementary approach to existing sequencing modalities. Fundamentally, the approach relies on site-specific fluorescent labeling of genomic material, generating long (>20 kbp) barcodes or maps that are visualized using fluorescence microscopy. Previous works have already demonstrated the promising potential of DNA optical mapping for phage genome identification as well as its flexibility for alternative labeling and color schemes.[18−20] However, while recent work showed the bacterial identification potential of enzyme-free optical mapping using long (>100 kbp) maps,[21] there is still a lack of a thorough assessment of the taxonomic resolution for a methyltransferase-mediated approach. In this work, such an assessment is carried out for map sizes averaging at only 30−35 kbp using a simple yet well-documented genomic database. The results indicate that DNA optical mapping is capable of revealing genetic differences at the strain level in this database, therefore, spotting the unique genomic locations along bacterial DNA. These regions are referred to as unique identifiers (UIs), and their use as a standalone database for DNA optical mapping is additionally evaluated here. Furthermore, with a simulated data set, we demonstrated that the UI principle could also be applied in typing gene clusters, such as the SCCmec cassette in the methicillin-resistant *Staphylococcus aureus* (MRSA) genome. Together, these outcomes indicate that the UIs of DNA optical mapping are not limited to a single target within the organism's genome, as is the case for 16S sequencing, where only the variable regions within the 16S gene cluster serve as UIs.

In summary, given the potential of DNA optical mapping to neatly complement current sequencing platforms at the taxonomic resolution scale, this paper aims to position this low-cost analytical platform in the current state of the art through a comprehensive validation of its performance in identification.

MATERIALS AND METHODS

Experimental Data. *DNA Extraction.* An *Escherichia coli* K12 (MG1655) culture was grown overnight at 37 °C on an agar plate. Next, a single colony was inoculated in 5 mL of autoclaved LB medium (Invitrogen) with a sterile pipette and grown overnight at 37 °C with continuous shaking. DNA extraction was performed using the Qiagen Puregene Yeast/Bact Kit B according to the manufacturer's protocol. The concentration and purity of the DNA extracts were measured using a Biodrop µlite UV−vis spectrophotometer.

DNA Labeling and Combing. Fluorescent labeling of DNA extracts was performed at a final DNA concentration of 50 ng/µL using 45 µM rhodamine B-functionalized AdoMet analogue (referred to as compound 5a in earlier work[20]) and 0.18 mg/mL M.*TaqI* methyltransferase enzyme. CutSmart buffer (NEB) was added to the reaction mixture, which was then incubated at 60 °C for 1 h with gentle shaking. To quench the enzymatic reaction, 2 µL of proteinase k (800 units/mL, NEB) was added to the mixture and allowed to react for 1 h at 50 °C with gentle shaking. For purification, the sample was embedded in a 2% agarose plug. This was done by briefly melting UltraPure Low Melting Point Agarose (Thermo Fisher) in 1× tris-acetate-EDTA (TAE) buffer at 70 °C, adding it to the sample and allowing the agarose to set at 4 °C at the bottom of an Eppendorf tube. Next, this plug was washed every 30 min with 500 µL of 1× TAE buffer at room temperature. Washing consists of removing the buffer present and adding a new volume, pipetting in a manner that causes the plug to come loose from the bottom of the Eppendorf tube and float freely in solution, enabling maximum contact with the surrounding wash solution. After four washing steps, the agarose plug was washed on ice twice with 2 V of 1× β-Agarase I Buffer (NEB) for 30 min. Next, the plug was melted at 65 °C for 10 min and treated with β-agarase (NEB) at 42 °C for 1 h with gentle shaking. Purified labeled DNA was obtained after twofold dialysis for 45 min using 0.1-µm-diameter Millipore dialysis membranes (MF-Millipore membrane, Merck) floating on 1× TAE buffer. Finally, the labeled DNA was combed on Zeonex-coated coverslips following the standardized protocol that was reported previously.[18,22]

Imaging. A Zeiss structured illumination microscopy (SIM) Elyra microscope with a Zeiss Plan-APOCHROMAT 63× oil

https://doi.org/10.1021/acsomega.1c01381
ACS Omega 2021, 6, 21276−21283

immersion objective (numerical aperture 1.4) and an electron-multiplying charge-coupled device (EMCCD) camera (exposure time 300 ms/frame, EM gain setting 35) was used for imaging. An additional 1.6× image magnification was applied. The field of view per image was 75 × 75 μm^2. The camera pixel size projected in the sample was 80 nm/pixel. A power of ∼3 mW over the field of view was provided by a 561 nm excitation laser. The emission was filtered using a 570−620 nm bandpass filter. For each field of view, 25 frames were recorded for five SIM modulation angles and five phases/angle. The illumination patterns for SIM were created by a grating with a period of 34 μm. A drop of Milli-Q water was placed on top of the sample before imaging. SIM reconstruction was performed using the Zeiss Zen software package. DNA fragments were segmented manually on the SIM images using ImageJ.[23] The general workflow for DNA optical mapping is schematically depicted in Figure 1.

Data Analysis. To assign optical maps to the set of database genomes (or UIs), a custom MATLAB-based analysis code was used. Briefly, after image segmentation, an intensity trace was extracted for each DNA molecule. Cross correlation was then used to estimate the similarity of these experimental intensity traces with theoretical traces derived from the genome sequences of several target microbial species (barcodes, i.e., logical series of zeros and ones codifying the sites at which the full genomes should produce a fluorescence signal based on the specific enzymatic reaction performed during the experimental stage). For each species and each trace, the matching score obtained (i.e., the maximum of the resulting cross-correlation function) was contrasted against a *null* model. The latter was generated by (i) permuting the corresponding theoretical barcode, (ii) computing the cross-correlation function between the original experimental trace and the reshuffled barcode, (iii) retaining the maximum value of this new cross-correlation function, and (iv) repeating this procedure a sufficiently high number of times. An empirical *p*-value (p_1) could therefore be calculated for every pair of measured trace/target species: a match was deemed statistically significant if the *p*-value was found to be smaller than the preset threshold.

In the second step, if a single experimental trace was assigned to several species, the corresponding matching scores were compared to reduce potential identification ambiguities due to imperfect labeling. Here, an approach inspired by bootstrapping was adopted. The subregion of the theoretical barcode corresponding to the position of best significant matching (which returned the largest matching score at the global database level) was resampled by artificially removing two randomly selected fluorescent sites, and its correlation with the measured intensity trace was then estimated again. This procedure was repeated a sufficiently high number of times (i.e., the number of resampling steps as specified in the Results section) to define a *null* distribution that was used to test all of the other significant matching scores through a second *p*-value metric (p_2). In this case, if p_2 was found to be lower than the preset threshold, the multiple matches with the corresponding species were discarded, and the experimental trace was assigned to only the species featuring the largest significant matching score at the global database level (whose barcode was resampled). For a more detailed reading of the full analysis procedure, we refer to earlier work.[18]

For all UI analyses (experimental and simulated), 1000 permutation and 1000 resampling steps were performed, and for both steps, a significance threshold of 0.001 was imposed. For all other analyses (experimental and simulated), 300 permutation and 300 resampling steps were performed, and a threshold of 0.01 was set for both.

Simulated Data. A custom simulation toolbox, described previously,[18] was used to generate the simulated SIM data sets. For all *E. coli*-related analyses, 1000 MG1655 molecules with an average size of 35 kbp were simulated. For the MRSA-related simulation, 10 000 maps of 35 kbp were simulated. For the long-map analysis, 500 maps of 65 kbp were generated. All simulated maps exhibit an average overstretch factor of 1.73 (based on earlier work[18]).

RESULTS AND DISCUSSION

Assigning K12 MG1655 Maps to a Complete Database. In this section, the objective was to obtain a general indication of the taxonomic resolution of enzyme-mediated DNA optical mapping by matching all 443 experimental and 1000 simulated DNA maps, obtained from the K12 MG1655 (NC_000913) ground truth genome, to the full genome of the ground truth genome and seven additional *E. coli* genomes: K12 W3110 (NC_007779), K12 DH10B (NC_010473), BW2952 (NC_012759), BL21(DE3) (NC_012971), REL606 (NC_012967), HS (NC_009800), and ATCC8739 (NC_010468). The completely unrelated genome of *Limosilactobacillus reuteri* DSM 20016 (NC_009513) was used as a negative control. Whole-genome information was retrieved from NCBI.[24] While ideally each DNA map would only be significantly assigned to the single ground truth genome, one can naturally expect a single map to be assigned to multiple database entries when there is a high degree of genomic similarity (with the extremities being genomic regions that are conserved within multiple database entries). As demonstrated in previous works, in such cases, a comparison between all significant matching scores for a single DNA map is effectively accounted for by the resampling step of the analysis.[18] As such, the specificity of map assignments is maximized. This situation, regarding genomic similarity, is highly applicable for the database that was selected for this matching analysis, as can be observed from the phylogenetic tree. While few differences in the matching sensitivity are apparent for genomes within *E. coli* strain A (the first four genomes, which include the ground truth genome K12 MG1655) due to the high degree of sequence conservation within this clade of substrains, such a strain is clearly distinguishable from the genomes belonging to strain B (BL21(DE3) and REL6060) and strain C (HS and ATCC873). Indeed, the simulated data are clearly indicative of strain-level resolution (Figure 2A), and this claim is confirmed in the experimental data set (Figure 2B). No matches were found for *L. reuteri* after resampling. A *p*-value threshold of 0.01 was imposed for both p_1 and p_2.

Remarkably, while the genomic differences between, for example, K12 MG1655 and BL21(DE3) accounted for only approximately 7−10% of the total MG1655 genome, the difference in the number of assigned maps was more pronounced (373 MG1655 maps versus 145 BL21(DE3) maps after resampling in Figure 2B). The reason can be found in the unique genomic regions being scattered within the genome (see also next paragraph) and the long-range nature of the experimental DNA maps (30 kbp on average). With unique regions scattered along the genome coordinate, there is a higher probability for any map to contain such unique features, thus leading to an assignment that is unique to the ground

https://doi.org/10.1021/acsomega.1c01381
ACS Omega 2021, 6, 21276−21283

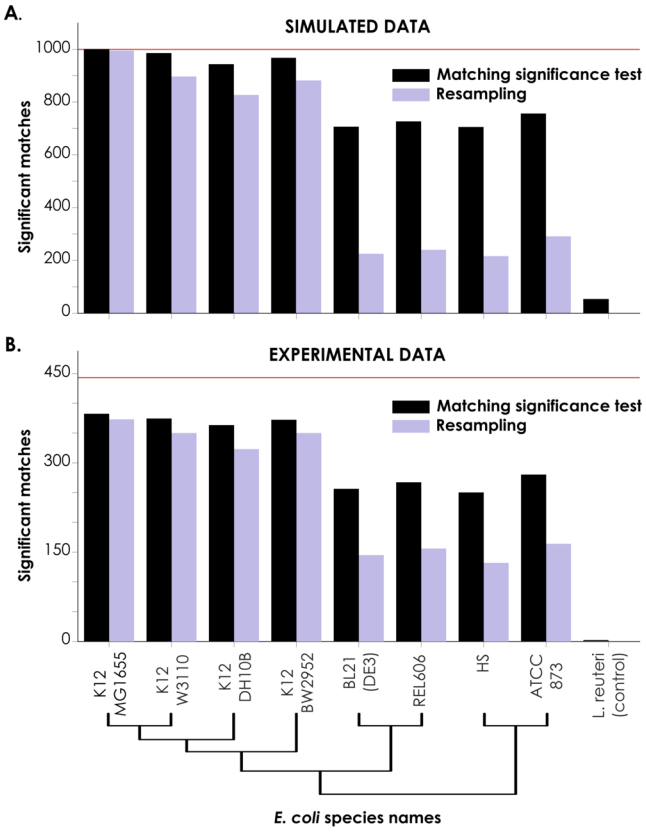


Figure 2. Global map assignments. (A) 1000 simulated experimental *E. coli* K12 MG1655 maps of 35 kbp and (B) 443 experimental *E. coli* K12 MG1655 maps (30 kbp average). For both data sets, black bars represent the results after the matching significance test (p_1 < 0.01 for significance, 300 permutations) and blue bars represent the results after resampling (p_2 > 0.01 for significance, 300 resampling steps). The red horizontal line indicates the total molecule number.

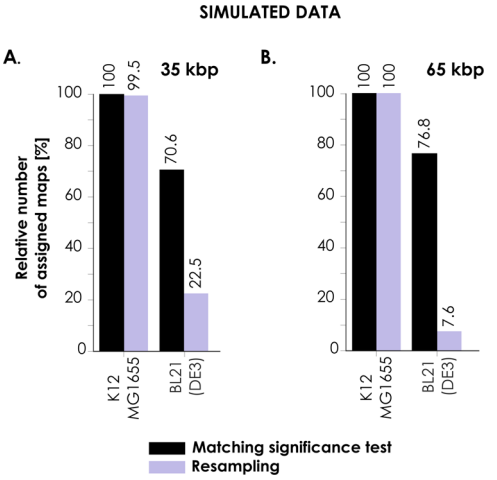


Figure 3. Effect of increased map size on global map assignments. (A) 1000 simulated *E. coli* K12 MG1655 maps of 35 kbp and (B) 500 simulated *E. coli* K12 MG1655 maps of 65 kbp. For both data sets, black bars represent the results after the matching significance test (p_1 < 0.01 for significance, 300 permutations) and blue bars represent the results after resampling (p_2 > 0.01 for significance, 300 resampling steps).

truth strain. This also implies that the difference in assigned maps between two strains would be amplified even more if the DNA map size is increased since the chance of capturing only conserved sequences in such a map (and consequentially ending up with a conserved assignment after resampling) is decreased. In other words, while the map assignment sensitivity already reached 100% at 35 kbp, the taxonomic resolving power for each map (specificity) increased upon increasing the map size. This was confirmed using a simulated data set consisting of 500 maps of 65 kbp (Figure 3). The same matching parameters were applied as indicated previously.

Projecting Significant Maps Reveals Unique Regions and Structural Variations. When comparing the ground truth MG1655 genome to any other database entry, the differences in matching sensitivity observed are expected to arise from genomic differences between the two. In other words, isolating unique maps (i.e., maps uniquely assigned to the ground truth species and not to a comparison species) should reveal those sequence regions unique to the ground truth. Conversely, DNA maps assigned to both genomes should be located in genomic regions conserved to both entries. To reveal the locations of such unique and conserved matches, strain BL21(DE3) was selected from the database and compared to the ground truth strain K12 MG1655. The unique regions are depicted as red vertical lines, with widths corresponding to their actual size as obtained by GView interactive Genome Viewer.[25] The blue area shows the matched map density (map coverage) at each location in the genome and is plotted to superimpose the red lines in the case of position overlap (Figure 4). Indeed, for the blue coverage plot related to the conserved matches (maps with p_2 > 0.01 for both strains), almost no overlap with the red unique regions was found in either simulated or experimental data sets (Figure 4A,B, respectively). This was also quantified by calculating the average coverage of unique regions as the base pair content of experimental maps overlapping with unique regions divided by the total unique base pair content, yielding 0.07 and 0.04 for the simulated and experimental data sets, respectively. In contrast, the blue coverage plot related to the unique maps (p_1 < 0.01 for K12 MG1655 and p_1 > 0.01 for BL21(DE3)) displays almost complete overlap with the red regions, as also evidenced by the average coverage of unique regions being 6.22 and 1.57 for the simulated and experimental data sets, respectively (Figure 4C,D, respectively). Together, these results corroborate the strain-level resolution.

In addition to revealing unique genomic regions, a sequence inversion was found by comparing the ground truth strain K12 MG1655 with substrain K12 W3110. The location of this inversion was predicted by BLAST (NCBI) and is shown as a negative slope in the dot matrix representation.[26] Such an inversion can also be translated into map assignments. Indeed, a sequence inversion should give rise to map assignments that are conserved for the two genomes (p_2 > 0.01 for both substrains) but have opposite map matching directionality (forward or reverse) for both genomes. Matching directionality is a parameter that is registered during analysis and can easily be called for every map. When plotting only the assigned maps that fulfill these two requirements, the resulting coverage plots, depicted in blue for simulated and red for experimental data, perfectly coincide with the theoretical location of the sequence inversion (Figure 5).

Unique Identifier Analysis. As evident from the previous results, maps obtained from DNA optical mapping carry enough information to target strain-level genomic differences. As a result, to resolve the various strains present in a sample, clusters of adjacent UIs (instead of full genomes) could

https://doi.org/10.1021/acsomega.1c01381
ACS Omega 2021, 6, 21276−21283

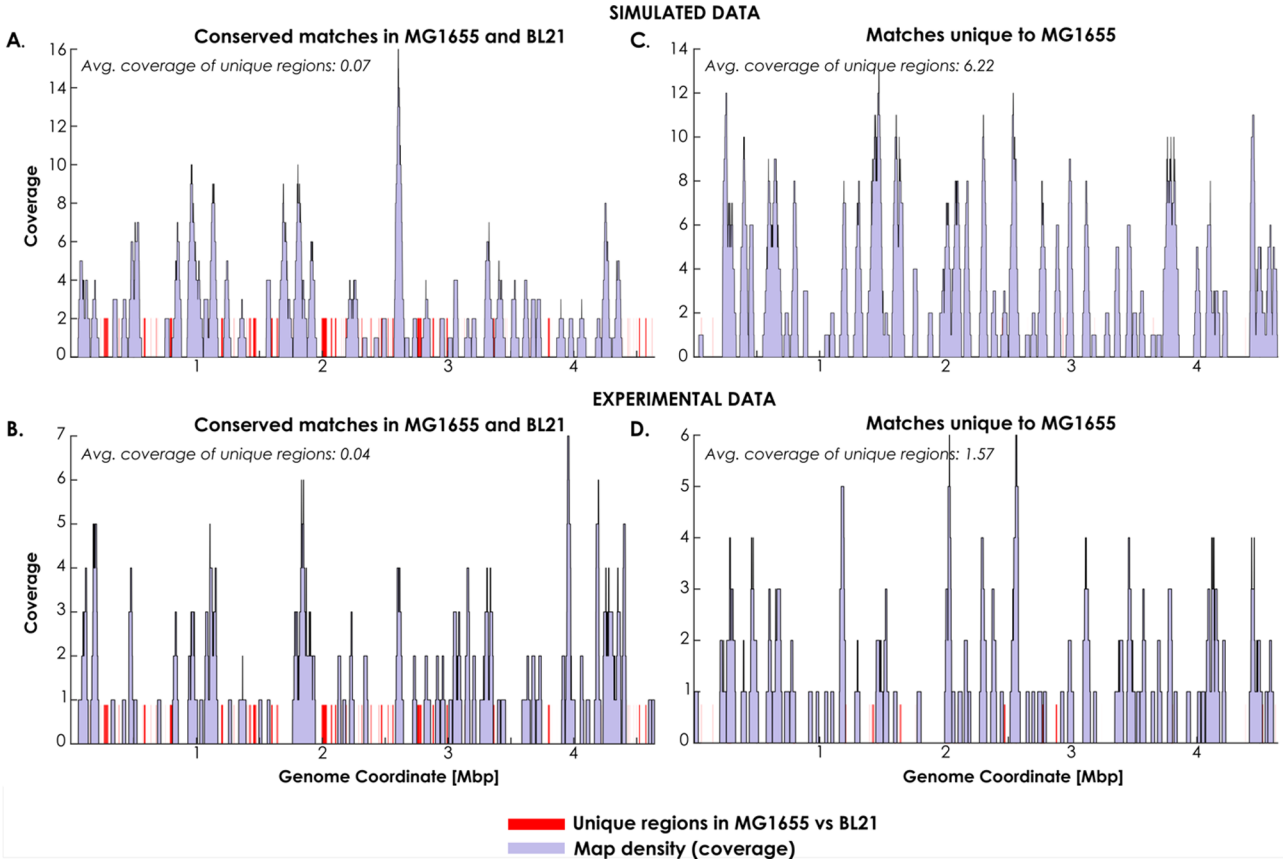


Figure 4. Plotting unique and conserved maps for the ground truth strain K12 MG1655 and comparison strain BL21(DE3). In all panels, vertical red lines represent unique regions in K12 MG1655 compared to BL21(DE3). Average (avg.) coverage of unique regions was calculated as the base pair content of experimental maps overlapping with unique regions divided by the total unique base pair content. (A) From the simulated data set, only the conserved maps (maps with $p_2 > 0.01$ for both strains, 300 resampling steps) were plotted and represented as a (blue) coverage map along the genome coordinate. (B) Same plot for the experimental data set. (C) From the simulated data set, only the unique maps (maps with $p_1 < 0.01$ for K12 MG1655, 300 permutations and $p_1 > 0.01$ for BL21(DE3), 300 permutations) were plotted and represented as a (blue) coverage map along the genome coordinate. (D) Same plot for the experimental data set.

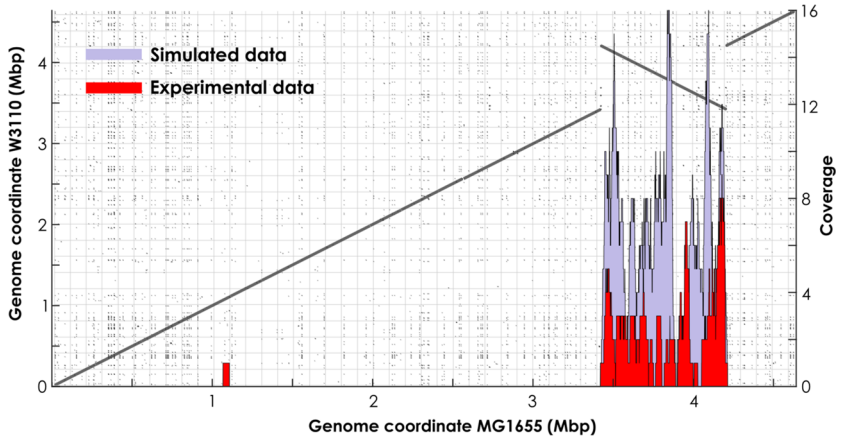


Figure 5. Plotting genome inversions with sequencing and mapping data. The dot matrix plot (in gray) was obtained with Nucleotide BLAST (NCBI) for the genomes of K12 MG1655 and K12 W3110. The negative slope represents an inversion location as obtained directly from sequencing data. For both simulated and experimental data sets, conserved maps (maps with $p_2 > 0.01$ for both substrains, 300 resampling steps) and opposite matching directionality (forward or reversed) were plotted along the genome coordinate representing the inversion location as obtained by mapping data.

potentially be used as a reference database. To test this hypothesis, the 443 experimental maps and 1000 simulated maps from *E. coli* K12 MG1655 presented above were rematched to a database containing the largest UI cluster for MG1655 compared to BL21(DE3) with a size of 41 kbp (bases 1995000 to 2036000) (Figure 6A) and the largest UI cluster for BL21(DE3) compared to MG1655 with a size of 41.5 kbp (bases 750000 to 791500) (Figure 6B). UI clusters

https://doi.org/10.1021/acsomega.1c01381
ACS Omega 2021, 6, 21276−21283

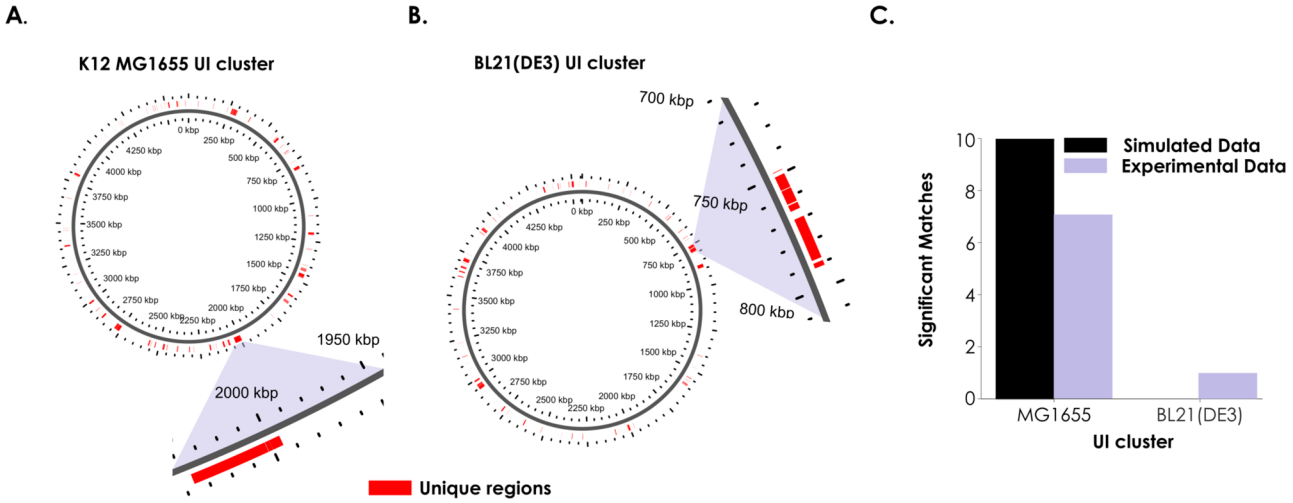


Figure 6. Unique identifier analysis. UI clusters are defined as clusters of unique sequences with a total size that is similar to the average map size. Gaps consisting of conserved sequences are allowed if these gaps are significantly smaller than the average map (max gap size in this case was set at 5 kbp). (A) Largest UI cluster found in the MG1655 genome (41 kbp). (B) Largest UI cluster for BL21(DE3) (41.5 kbp). (C) Matching results for all MG1655 maps (443 for the experimental data set and 1000 for the simulated data set) compared to a database consisting of both UI clusters (p_1 < 0.001, 1000 permutations and p_2 > 0.001, 1000 resampling steps).

are defined as clusters of unique sequences with a total size that is similar to the average map size. Gaps consisting of conserved sequences are allowed if these gaps are significantly smaller than the average map (max gap size in this case was set at 5 kbp). Maps were considered significant when p_2 > 0.001, which was a lower threshold to account for the smaller target sequence compared to all previously shown whole-genome-scale analyses. The results clearly indicate that the sole strain present in this sample is MG1655 (Figure 6C), with ten significant maps for the simulated data set and seven for the experimental data set. Note that such a low number of assigned maps is perfectly consistent with expectations since the database consists of less than 1% of each species' full genome. In addition, the single false-positive assignment out of the 443 experimental maps is consistent with the potential false positives at a confidence level of 0.1%. In conclusion, this UI approach not only allows for the estimation of relative abundances but also can increase the analysis speed due to the reduced database size, despite the increased number of permutation steps. In addition, the resampling step can ideally be skipped since the database itself is changed from being highly similar in the global analysis (for which the effect of resampling is very clearly demonstrated in Figures 2 and 3) to completely unrelated in UI analysis.

Application of Unique Identifier Analysis. Similarly, one clinical example in which DNA optical mapping analysis based on UIs could be employed is the case of MRSA. Here, the UI, compared to methicillin-sensitive *S. aureus* (MSSA), is a gene cluster carrying antibiotic resistance and is known as the staphylococcal cassette chromosome mec (SCCmec). This cassette is classified into 11 different types, varying in size between 21 and 67 kbp, with types I to III being the most common ones in healthcare-acquired MRSA (HA-MRSA).[27,28] A simulated data set of 10 000 maps for two MRSA strains (strain N315, accession NC_002745 carrying a type II cassette and strain T0131, accession CP002643 carrying a type III cassette) was used to show that the correct type of SCCmec gene cluster could be identified for both species (Figure 7). Analogous to the previous UI analysis, this time, the database consists of type I-III SCCmec cassettes, while the simulated

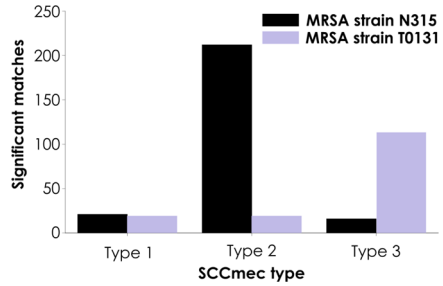


Figure 7. Unique identifier analysis applied to the MRSA genome. Matching results for 10 000 simulated maps drawn from the full genomes of two MRSA strains and compared to a database containing three types of SCCmec gene clusters (p_1 < 0.001, 1000 permutations and p_2 > 0.001, 1000 resampling steps).

DNA optical maps are drawn from the whole genome of the selected MRSA strain. Maps were considered significant when p_2 was > 0.001.

CONCLUSIONS

In this article, a thorough assessment of the taxonomic resolution of DNA optical mapping was carried out. This assessment clearly indicates that, although DNA optical mapping does not provide base-by-base reads as those typically resulting from DNA sequencing approaches, sequence variations at the strain level can be detected without being restricted to a single locus in the genome. Nonetheless, it should be noticed that, in the case of complex (highly mixed) samples, sufficient coverage of the genomes of all of the microbial species under study is required to perform UI analysis. Luckily, the throughput of DNA optical mapping is easily scalable toward such scenarios: in fact, rather than acquiring individual scattered images across very extended fields of view, entire specimens can nowadays be scanned through novel strategies for modular SIM microscopy.[29] This combined with the universal detection capability of DNA optical mapping and its fast and low-cost nature, can constitute the key point behind its possible extension to more challenging real-world case studies, ideally in the biomedical and clinical

https://doi.org/10.1021/acsomega.1c01381
ACS Omega 2021, 6, 21276−21283

fields. In addition, DNA optical mapping shows the potential to bridge the aforementioned taxonomic resolution gap and, as such, complement the current state-of-the-art sequencing platforms.

AUTHOR INFORMATION

Corresponding Author

Johan Hofkens − *Molecular Imaging and Photonics Unit, Department of Chemistry, KU Leuven, 3001 Leuven, Belgium; Max Planck Institute for Polymer Research, 55128 Mainz, Germany;* orcid.org/0000-0002-9101-0567; Email: johan.hofkens@kuleuven.be

Authors

Laurens D'Huys − *Molecular Imaging and Photonics Unit, Department of Chemistry, KU Leuven, 3001 Leuven, Belgium;* orcid.org/0000-0001-6325-9720

Raffaele Vitale − *Molecular Imaging and Photonics Unit, Department of Chemistry, KU Leuven, 3001 Leuven, Belgium; Dynamics, Nanoscopy and Chemometrics (DYNACHEM) Group, U. Lille, CNRS, LASIRE, Laboratoire Avancé de Spectroscopie pour les Interactions, la Réactivité et l'Environnement, F-59000 Lille, France;* orcid.org/0000-0002-7497-1673

Elizabete Ruppeka-Rupeika − *Molecular Imaging and Photonics Unit, Department of Chemistry, KU Leuven, 3001 Leuven, Belgium*

Vince Goyvaerts − *Molecular Imaging and Photonics Unit, Department of Chemistry, KU Leuven, 3001 Leuven, Belgium*

Cyril Ruckebusch − *Dynamics, Nanoscopy and Chemometrics (DYNACHEM) Group, U. Lille, CNRS, LASIRE, Laboratoire Avancé de Spectroscopie pour les Interactions, la Réactivité et l'Environnement, F-59000 Lille, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c01381

Author Contributions

‖L.D. and R.V. are first authors.

Notes

The authors declare the following competing financial interest(s): Johan Hofkens is a co-founder of the spin-off Chrometra which develops methyltransferase-directed modification of DNA-labelling kits and of the spin-off PerseusBiomics which is commercializing the optical mapping technology.

ACKNOWLEDGMENTS

J.H. acknowledges financial support from the Horizon 2020 Framework Programme of the European Union ADgut (grant number 686271), the European Union Research Council through ERC-2017-PoC Metamapper (grant number 768826), FWO (Fonds voor Wetenschappelijk Onderzoek, grant number G0C1821N), the Flemish Government through long-term structural funding Methusalem (CASAS2, Meth/15/04), and the Max Planck Institute through the MPI fellowship. C.R. acknowledges support from LAI U.Lille-KU Leuven (HPFM 2020-23). L.D. acknowledges financial support from FWO (Fonds voor Wetenschappelijk Onderzoek, grant number 11D3718N) and Prof. Dr. Susana Rocha for the fruitful discussions. The Zeiss SIM Elyra microscope was acquired through a CLME grant from Minister Lieten to the VIB BioImaging Core.

REFERENCES

(1) Shreiner, A. B.; Kao, J. Y.; Young, V. B. The gut microbiome in health and in disease. *Curr. Opin. Gastroenterol.* **2015**, *31*, 69−75.
(2) Cho, I.; Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* **2012**, *13*, 260−270.
(3) Sun, J.-Y.; Yin, T.-L.; Zhou, J.; Xu, J.; Lu, X.-J. Gut microbiome and cancer immunotherapy. *J. Cell. Physiol.* **2020**, *235*, 4082−4088.
(4) Mager, L. F.; et al. Microbiome-derived inosine modulates response to checkpoint inhibitor immunotherapy. *Science* **2020**, *369*, 1481−1489.
(5) Gopalakrishnan, V.; et al. Gut microbiome modulates response to anti−PD-1 immunotherapy in melanoma patients. *Science* **2018**, *359*, 97−103.
(6) Kowalski, K.; Mulak, A. Brain-Gut-Microbiota Axis in Alzheimer's Disease. *J. Neurogastroenterol. Motil.* **2019**, *25*, 48−60.
(7) Knight, R.; et al. The Microbiome and Human Biology. *Annu. Rev. Genomics Hum. Genet.* **2017**, *18*, 65−86.
(8) Ruan, W.; Engevik, M. A.; Spinler, J. K.; Versalovic, J. Healthy Human Gastrointestinal Microbiome: Composition and Function After a Decade of Exploration. *Dig. Dis. Sci.* **2020**, *65*, 695−705.
(9) Cani, P. D. Human gut microbiome: hopes, threats and promises. *Gut* **2018**, *67*, 1716−1725.
(10) Knox, N. C.; Forbes, J. D.; Van Domselaar, G.; Bernstein, C. N. The Gut Microbiome as a Target for IBD Treatment: Are We There Yet? *Curr. Treat. Options Gastroenterol.* **2019**, *17*, 115−126.
(11) Schwarze, K.; et al. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genet. Med.* **2020**, *22*, 85−94.
(12) Johnson, J. S.; et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **2019**, *10*, No. 5029.
(13) Earl, J. P.; et al. Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes. *Microbiome* **2018**, *6*, No. 190.
(14) van Dijk, E. L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. The Third Revolution in Sequencing Technology. *Trends Genet.* **2018**, *34*, 666−681.
(15) Petersen, L. M.; Martin, I. W.; Moschetti, W. E.; Kershaw, C. M.; Tsongalis, G. J. Third-Generation Sequencing in the Clinical Laboratory: Exploring the Advantages and Challenges of Nanopore Sequencing. *J. Clin. Microbiol.* **2020**, *58*, No. e01315-19.
(16) Callahan, B. J.; et al. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res.* **2019**, *47*, No. e103.
(17) Hillmann, B.; et al. Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems* **2018**, *3*, No. e00069-18.
(18) Bouwens, A.; et al. Identifying microbial species by single-molecule DNA optical mapping and resampling statistics. *NAR Genom. Bioinform.* **2020**, *2*, No. lqz007.
(19) Wand, N. O.; et al. DNA barcodes for rapid, whole genome, single-molecule analyses. *Nucleic Acids Res.* **2019**, *47*, e68.
(20) Goyvaerts, V.; et al. Fluorescent SAM analogues for methyltransferase based DNA labeling. *Chem. Commun.* **2020**, *56*, 3317−3320.
(21) Müller, V.; et al. Cultivation-Free Typing of Bacteria Using Optical DNA Mapping. *ACS Infect. Dis.* **2020**, *6*, 1076−1084.
(22) Deen, J.; et al. Combing of Genomic DNA from Droplets Containing Picograms of Material. *ACS Nano* **2015**, *9*, 809−816.
(23) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671−675.
(24) NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46*, D8 D13. DOI: 10.1093/nar/gkx1095.
(25) Petkau, A.; Stuart-Edwards, M.; Stothard, P.; Van Domselaar, G. Interactive microbial genome visualization with GView. *Bioinformatics* **2010**, *26*, 3125−3126.
(26) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.

(27) Kateete, D. P.; et al. CA-MRSA and HA-MRSA coexist in community and hospital settings in Uganda. *Antimicrob. Resist. Infect. Control* **2019**, *8*, No. 94.
(28) Reichmann, N. T.; Pinho, M. G. Role of SCC mec type in resistance to the synergistic activity of oxacillin and cefoxitin in MRSA. *Sci. Rep.* **2017**, *7*, No. 6154.
(29) Van den Eynde, R.; et al. Self-contained and modular structured illumination microscope. *Biomed. Opt. Express* **2021**, *12*, 4414−4422.

https://doi.org/10.1021/acsomega.1c01381
ACS Omega 2021, 6, 21276−21283