

# Optical Whole-Genome Restriction Mapping as a Tool for Rapidly Distinguishing and Identifying Bacterial Contaminants in Clinical Samples

RICHARD R. CHAPLEAU<sup>1</sup>, JAMES C. BALDWIN<sup>2</sup>

## ABSTRACT

**Introduction:** Optical restriction genome mapping is a technology in which a genome is linearized on a surface and digested with specific restriction enzymes, giving an arrangement of the genome with gaps whose order and size are unique for a given organism. Current applications of this technology include assisting with the correct scaffolding and ordering of genomes in conjunction with whole-genome sequencing, observation of genetic drift and evolution using comparative genomics and epidemiological monitoring of the spread of infections. Here, we investigated the suitability of genome mapping for use in clinical labs as a potential diagnostic tool.

**Materials and Methods:** Using whole genome mapping, we investigated the basic performance of the technology for identifying

two bacteria of interest for food-safety (*Lactobacilli* spp. and Enterohemorrhagic *Escherichia coli*). We further evaluated the performance for identifying multiple organisms from both simple and complex mixtures.

**Results:** We were able to successfully generate optical restriction maps of four *Lactobacillus* species as well as a strain of Enterohemorrhagic *Escherichia coli* from within a mixed solution, each distinguished using a common compatible restriction enzyme. Finally, we demonstrated that optical restriction maps were successfully obtained and the correct organism identified within a clinical matrix.

**Conclusion:** With additional development, whole genome mapping may be a useful clinical tool for rapid *invitro* diagnostics.

**Keywords:** Assay development, Bacterial detection, Genome identification, Technical evaluation, Whole genome mapping

## INTRODUCTION

One of the primary goals of public health agencies is the early detection of infectious disease and emergent biological agents. Whole-genome mapping (WGM) is a recent technology capable of generating a visible signature specific to a given pathogen [1] with the possibility of being used as a clinical detector. By utilizing the entire pathogen's genome, a high degree of confidence in diagnostic value could potentially be obtained. This technology is currently used in basic research laboratories to aid in DNA sequence analysis, but its applicability in clinical situations has yet to be realized. WGM has been applied primarily for assembly of whole-genome sequencing [2-4] and in strain typing [1,5-8]. Recently, the use in strain typing has been advanced even further to include rapid assessment of genome instabilities in highly pathogenic *Staphylococcus aureus* [9].

Unlike many currently employed diagnostic technologies, this technology does not rely upon DNA amplification [10] and is thereby less prone to enzymatic errors or a prior knowledge of the suspected pathogens. This technology requires as input only purified, stable genomic DNA [11]. This input can be successfully obtained using a number of commercially available high molecular weight extraction kits. The genomic DNA is then gently added via pipette into charged, microfluidic channels which ensure a linear deposition of the DNA. This linear DNA is critical for correct restriction mapping, as individual fragments will be analysed by the instrument. The linearized DNA is then treated with restriction endonucleases, which remove short fragments of DNA, leaving larger fragments present and the order of these fragments remains intact. The final step requires the addition of a fluorescent dye and imaging using a digital camera. Data analysis is performed by overlapping fragment patterns to assemble full-length chromosomes, genomes and/or plasmids.

Because of the lack of a need for early knowledge, this instrument could be used to identify completely unknown infectious agents

within a patient sample. Furthermore, this technology could readily be multiplexed, or can be seen as tolerant of contaminating DNA, because it can assemble multiple optical maps in a single sample individually.

## AIM

The potential of WGM was assessed, within a clinical scope, by evaluating the impacts of mixed cultures and complex, clinical sample backgrounds.

## MATERIALS AND METHODS

The study was performed over the course of one year (2011-2012) at the US Air Force School of Aerospace Medicine in the Applied Technology and Genomics Division. Bacterial cultures were purchased from the American Type Culture Collection (Manassas, VA). Culture media and supplies were purchased from Sigma-Aldrich (St. Louis, MO), VWR (Radnor, PA), or Fisher (Waltham, MA), as appropriate. Chemicals and reagents for DNA extraction and map creation were purchased from OpGen (Gaithersburg, MD). All bacterial operations were performed within Class II Biosafety cabinets and DNA preparations were conducted on freshly cleaned and DNA AWAY (Fisher) decontaminated bench.

**Design of experimental samples:** Two user-blinded experiment sets were performed to test whether multiple bacteria could be uniquely identified within mixtures. In the first set of experiments, three unique organisms (*Bacillus subtilis* subsp. *globigii*, *Enterococcus faecalis*, and *B. anthracis* (Sterne)) were independently cultured and combined into one of three combinations: a single organism (*B. subtilis*), two organism at equal concentration (*B. subtilis* and *E. faecalis*), or all three at equal concentrations. The mixtures were assigned letters X-Z by a third-party responsible for bacterial cultures. The second set of experiments introduced new organisms

and higher complexity. In these experiments, six organisms were randomly selected by the culture specialist and provided in a blinded manner similar to the previous experiment, with the exception that one organism (*Pseudomonas aeruginosa*) was mixed at one-fifth the concentration of the other five organisms. In this manner, we were able to simultaneously evaluate the ability of the technology to detect and identify individual mixed organisms and to detect minor constituents when presented with overwhelming contaminant genomes.

Finally, to test the clinical applicability, *E. faecalis* was spiked into a commercial nasal wash sample. This organism was selected as it is not normally found in the nasal passages and therefore any observed *E. faecalis* came from the spike and not from background presence. This ensures that a target bacterium can be identified within clinically-relevant samples.

**Extraction of genomic DNA:** All bacteria were processed according to the manufacturer's directions for gram-positive bacterial DNA extraction using the HMW DNA Isolation Kit and the MapCard II Kit for Microbial Genomes (instructions provided with kits, catalog numbers 14310-020 and 14001-010, respectively). Preliminary tests indicated no detriment to gram-negative bacterial DNA using this method (unpublished results). Briefly, 100  $\mu$ L of bacterial culture or *E. faecalis* spiked nasal wash (25% from 3 McFarland) was spun at 5000g to pellet bacteria. The pellet was resuspended in 500  $\mu$ L Cell Wash Buffer and spun a second time. Spheroplasting was achieved using 0.5  $\mu$ L Ready-Lyse lysozyme and 3  $\mu$ L of Mutanolysin added to 100  $\mu$ L Spheroplasting Buffer and the reaction was carried out for 2 hour at 37°C. Spheroplasting was stopped and cells were lysed by adding 90  $\mu$ L Isolation Buffer and 10  $\mu$ L Proteinase K at 56°C for 30 min. Isolated DNA was diluted in Dilution Buffer and quality was checked on QCard surfaces prior to placing the optimal DNA dilution on a MapCard surface for analysis.

**Whole-genome restriction mapping:** Optical maps of isolated genomic DNA from bacterial samples were generated using the Argus System (OpGen) as per the manufacturer's directions using the Stain Kit DIL and the appropriate Enzyme Kit for the desired reaction. A minor modification to the MapCard protocol was found necessary to ensure high quality cards with no introduction of air bubbles: a stepwise application of the port seal as solutions were added to the MapCard. Antifade Solution was slowly pipetted into the top well of the MapCard, followed by slow addition of the appropriate Reaction Buffer for the enzyme chosen and the Enzyme itself. Finally, diluted JOJO was added to the MapCard, which was then placed in the MapCard Processor for automated restriction enzyme digestion. Following digestion, the MapCard was transferred to the Argus Optical Mapper for image acquisition and analysis. Organism identification was performed using the supplied software and the provided genome database. Database entries can be edited and uploaded using standard formatting with GenBank data files.

## RESULTS

The ability of the WGM to detect multiple organisms in a single sample was evaluated in two experiments. In the first study, three samples were prepared in a single-blind method. The organisms used in this study included two vegetative bacteria and a spore preparation. By preparing the samples as per the instrument manufacturer's Gram-positive isolation protocol, the instrument was able to successfully detect all vegetative bacteria in each of the three samples [Table/Fig-1].

In a second, single-blinded multi-organism study, six bacteria [Table/Fig-2] were combined into a single sample. Without any preselection of restriction enzymes caused by sample bias, we were successfully able to detect two of the bacteria, although neither was *P. aeruginosa* [Table/Fig-2], which had poor representation in all experiments. Limitations imposed by the instrumentation restricted

the work to only three restriction endonucleases. Here, the enzymes were chosen by comparing the vendor-provided enzyme kits against a vendor-provided database filtered for targets of potential food safety and public health interest so that the maximum number of potential targets could be identified. Using *in silico* estimates, these three enzymes (*Afl*III, *Nco*I, and *Nhe*I) were theoretically sufficient for distinguishing between *B. cereus* (*Nhe*I), *Escherichia coli* (*Afl*III, *Nco*I), *Listeria monocytogenes* (*Nco*I, *Nhe*I), and *S. aureus* (*Afl*III) from our mixed culture.

Finally, *E. faecalis* was spiked into a commercially obtained nasal wash sample externally tested to be negative for the target bacterium (as well as many other pathogens). Although not normally associated with the nasal passages, this bacterium was chosen as an example because a whole-genome map was successfully obtained within the laboratory (99% DB coverage), it is a normal

Sample ID	Organism identified	DC	CC	Spiked organism
Sample X	<i>E. faecalis</i> , V583	98%	99%	<i>E. faecalis</i>
	<i>B. atrophaeus</i> , 1942	57%	90%	<i>B. atrophaeus</i> <i>B. anthracis</i> (Sterne), spore
Sample Y	<i>E. faecalis</i> , V583	99%	100%	<i>E. faecalis</i>
	<i>B. atrophaeus</i> , 1942	39%	87%	<i>B. atrophaeus</i>
Sample Z	<i>B. atrophaeus</i> , 1942	98%	99%	<i>B. atrophaeus</i>

**[Table/Fig- 1]:** Unrestricted database searching in a single-blind mixed sample study. *B. atrophaeus* is the identifier provided in the vendor-provided database for *B. subtilis* subsp. *globigii*.

Sample ID	Organism identified	DC	CC	Spiked organism
Sample A				<i>B. cereus</i>
	<i>E. coli</i> , O157:H7 str. Sakai	11%	48%	<i>E. coli</i> O157:H7
				<i>K. pneumonia</i>
	<i>L. monocytogenes</i> , EGD-e	24%	65%	<i>L. monocytogenes</i>
				<i>P. aeruginosa</i>
			<i>S. aureus</i>	

**[Table/Fig- 2]:** Unrestricted database searching in a single-blind, complex mixture study.

Spiked Organism	Hit ID	Organism Identified	DC	CC
<i>E. faecalis</i>	1	<i>E. faecalis</i> , V583	21%	56%
	2	<i>Mycoplasma arthritidis</i> , 158L3-1	7%	5%
	3	<i>Chlamydia trachomatis</i> , D/UW-3/CX	7%	8%

**[Table/Fig- 3]:** Unrestricted database searching in a spiked nasal wash sample using a single 1200 kilobase contig

Database Map Name	DC	CC	BC Factor
<i>V. cholerae</i> , MJ-1236 chromosome 2	88%	24%	2112
<i>V. cholerae</i> , M66-2 chromosome 2	84%	22%	1848
<i>V. cholerae</i> , O1 biovar eltor str. N16961 chromosome 2	81%	22%	1782
<i>V. cholerae</i> , M66-2 chromosome 1	77%	60%	4620
<i>V. cholerae</i> , MJ-1236 chromosome 1	76%	59%	4484
<i>V. cholerae</i> , O1 biovar El Tor str. N16961 chromosome 1	74%	57%	4218
<i>V. cholerae</i> , O395 chromosome 2	58%	44%	2552
<i>V. cholerae</i> , O395 chromosome 1	39%	10%	390
<i>Brucella abortus</i> , bv. 1 str. 9-941 chromosome 2	37%	11%	407
<i>Acidilobus saccharovorans</i> , 345-15 chromosome	31%	10%	310

**[Table/Fig-4]:** Demonstration of the BC Factor to mitigate potentially misleading detection calls based default sorting algorithms. Red highlighted hits indicate the shorter, reordered chromosome in *V. cholerae*, whereas green hits indicate the longer, native chromosome

component of the natural human GI biome [12], and at large concentrations it can become pathogenic, causing bacteremia and urinary tract infections [13]. When combined in a 50% V/V mixture with a nasal wash background, enough high molecular weight DNA was isolated to provide a contig size sufficiently large to search the *in silico* database and provide a positive identification [Table/Fig-3], albeit with a lesser coverage than from a pure bacterial culture.

## DISCUSSION

One major challenge to using this technology in the diagnostic laboratories is restriction enzyme selection. Design constraints of the current commercially available instruments only permit simultaneous measurement of three samples, which can be individual patient samples with a single restriction enzyme or a single patient with up to three enzymes. Without any a priori knowledge regarding the organism, the correct selection of enzymes and subsequent pathogen identification, is unlikely. For example, only 66% of the organisms seeded in our study could be potentially identified using the selected enzymes, and of these, only 50% actually were observed. Future optimization of the instrument or sample card design may permit multiple experiments in one chip; however, the original application of this technology has not been for diagnostic use. Although we show that WGM can be used to generate multiple maps within a single sample, clinical application may be challenging without redesign considering multiple contaminating pathogens. Furthermore, identifying new and emergent strains may be complicated when found in mixed samples containing related, benign species.

Additionally, the database searching in the supplied software contains four potential search methods: unrestricted (whole database) or restricted (user-selected organisms) and with or without plasmids (useful for identifying toxin-producing strains). Once a search has been selected and queued, the database search results are obtained in approximately 0.5 – 1 hour. A list of identified organisms can be accessed by double-clicking the results icon. From a potential clinical-use perspective, this low-intensity interaction is highly desirable; however, at this point, the software has reached its limit of ease-of-use, as further data interpretation requires complex user input and manipulations of data.

By default, the included software prioritizes identification based upon “Database Coverage” (DC), which is a percentage measure of the amount of the genome contained in the included database that aligned with the contig searched. This is a useful method of organizing if the primary goal of the user is to create a whole-genome restriction map for sequencing or scaffolding applications; however, in the context of pathogen identification, this method fails to account for the contig contribution. A more useful method of sorting the identification table for detection purposes would take into account the “Contig Coverage” (CC), which is a percentage measure of the amount of the contig aligned with the database genome. We propose a multiplication operation involving the DC and the CC, termed the Bacterial Coverage (BC) Factor ( $BC = (DC\% * 100) * (CC\% * 100)$ ). Using *Vibrio cholerae* as an example, the initial sorting method identified the shorter chromosome from three strains as the primary organisms [Table/Fig-4]. In contrast, using the BC Factor method, the longer chromosomes of these same strains were called as the top three hits. In the former case, the DC values were greater than 80% but the CC values were less than 25%, whereas for the latter, the DC values ranged between 74% and 77% and the CC values were about 60%. Clearly, including the CC contribution could result in a more reliable diagnostic value.

When considering incorporating this technology into a clinical environment, even in a lab-developed test capacity, the relevant infrastructure requirements must be considered. A rudimentary form of this technology could be developed using electrostatic glass slides, custom fabricated microfluidic coverslips and a basic

fluorescent microscope with at least 60x objectives and a digital camera [11]. Alternatively, a fully-equipped system inclusive of all required instruments and analytical capabilities as demonstrated herein is commercially available. The instrument cost is in-line with other standard clinical lab instruments and the per sample costs from the manufacturer would be dependent upon number of restriction enzymes chosen. Each of the commercial product's chips could be used to run up to three patient samples in parallel or a single patient with three unique enzyme combinations. The entire experiment, from sample receipt through data analysis, can be performed in a single 8-hour shift. The data reported from the software we used are objective, single-line indications of which organisms are present, limiting any subjective data interpretation to a single step in the procedure where the stained images are viewed and a decision to process the chip must be made. This step may be made objective with a few simple image analysis techniques, although this is not currently standard practice.

With some modifications to the standard procedure for data analysis, it may be possible to employ this technology as a lab developed test in a clinical setting. Here we showed that this technique could be used to identify bacteria within nasal samples and we expect it to perform equally well with other clinical samples since the first step is a purification of the high molecular weight DNA. Depending upon the concentration of organism in the sample, it may be possible to perform restriction mapping-based diagnostics without bacterial culturing; however it is more likely that some culturing would be required. Therefore, only minimal time would be saved using this method versus standard microbe identification techniques. Instead, this technique could prove useful in strain typing for determination of pathogenesis. By interrogating the DNA directly, pathogenic strains will be readily apparent to the map alignment software. Such rapid strain typing could be envisioned to be useful in monitoring nosocomial outbreaks in neonatal and intensive care wards, or even as an initial screen for antibiotic resistant strains such as MRSA.

## CONCLUSION

We have shown optical restriction genome mapping as capable of identifying pure, clinically relevant organisms from single-blinded samples in culture media, in clinical matrices such as nasal wash, and identifying complex mixtures of unknown bacteria. Furthermore, we present a few simple modifications to the data analysis steps with the potential to turn this technology into a valuable device for clinical use.

## ACKNOWLEDGEMENTS

The authors would like to thank Manuel Caballero, David Compton, Roel Escobar, Sonal Naik, and Kenney Reynolds for assistance with bacterial culture. The authors are grateful to Dr. Clarise Starr for critical comments on the manuscript. Funding for this work was arranged through the Defense Health Program to JCB. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the United States Air Force, the United States Department of Defense, or the United States Government. The use of trade names is for clarity and should not be viewed as an endorsement by the United States Government.

## REFERENCES

- [1] Riley MC, Kirkup BC Jr, Johnson JD, Lesho EP, Ockenhouse CF. Rapid whole genome optical mapping of *Plasmodium falciparum*. *Malar J*. 2011;10:252.
- [2] Brown CT, Fishwick LK, Chokshi BM, Cuff MA, Jackson JM 4<sup>th</sup>, Oglesby T, et al. Whole-genome sequencing and phenotypic analysis of *Bacillus subtilis* mutants following evolution under conditions of relaxed selection for sporulation. *Appl Environ Microbiol*. 2011;77:6867–67.
- [3] Onmus-Leone F, Hang J, Clifford RJ, Yang Y, Riley MC, Kuschner RA, et al. Enhanced de novo assembly of high throughput pyrosequencing data using whole genome mapping. *PLoS One*. 2013;8:e61762.

- [4] Raeside C, Gaffé J, Deatherage DE, Tenailon O, Briska AM, Ptashkin RN, et al. Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. *M Bio*. 2014;5:e01377-14.
- [5] Kotewicz ML, Mammel MK, LeClerc JE, Cebula TA. Optical mapping and 454 sequencing of *Escherichia coli* O157: H7 isolates linked to the US 2006 spinach-associated outbreak. *Microbiology*. 2008;154:3518-28.
- [6] Johnson PD, Ballard SA, Grabsch EA, Stinear TP, Seemann T, Young HL, et al. A sustained hospital outbreak of vancomycin-resistant *Enterococcus faecium* bacteremia due to emergence of van BE. *faecium* sequence type 203. *J Infect Dis*. 2010;202:1278-86.
- [7] Petersen RF, Litrup E, Larsson JT, Torpdahl M, Sorensen G, Müller L, Nielsen EM. Molecular characterization of *Salmonella Typhimurium* highly successful outbreak strains. *Foodborne Pathog Dis*. 2011;8:655-61.
- [8] Jackson SA, Kotewicz ML, Patel IR, Lacher DW, Gangiredla J, Elkins, CA. Rapid genomic-scale analysis of *Escherichia coli* O104:H4 by using high-resolution alternative methods to next-generation sequencing. *Appl Environ Microbiol*. 2012;78:1601-5.
- [9] Sabirova JS, Xavier BB, Ieven M, Goossens H, Malhotra-Kumar S. Whole genome mapping as a fast-track tool to assess genomic stability of sequenced *Staphylococcus aureus* strains. *BMC. Res Notes*. 2014;7:704.
- [10] Datta S, Chatterjee S, Veer V. Recent advances in molecular diagnostics of hepatitis B virus. *World J Gastroenterol*. 2014;20:14615-25.
- [11] Ananiev GE, Goldstein S, Runnheim R, Forrest DK, Zhou S, Potamouis K, et al. Optical mapping discerns genome wide DNA methylation profiles. *BMC MolBiol*. 2008;9:68.
- [12] Wang X, Yang Y, Huycke MM. Commensal bacteria drive endogenous transformation and tumour stem cell marker expression through a bystander effect. *Gut*. 2015;64:459-68.
- [13] Gupta S, Kapur S, Padmavathi D. Comparative prevalence of antimicrobial resistance in community-acquired urinary tract infection cases from representative States of northern and southern India. *J ClinDiagn Res*.2014;8:DC09-12.

#### PARTICULARS OF CONTRIBUTORS:

1. Applied Technology and Genomics Center, United States Air Force School of Aerospace Medicine, 711<sup>th</sup> Human Performance Wing, Air Force Research Laboratory, Wright-Patterson AFB OH.
2. Applied Technology and Genomics Center, United States Air Force School of Aerospace Medicine, 711<sup>th</sup> Human Performance Wing, Air Force Research Laboratory, Wright-Patterson AFB OH.

#### NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:

Dr. Richard R. Chapleau,  
2510 Fifth St., Bldg 840, Area B, Wright-Patterson AFB OH 45433  
E-mail: richard.chapleau.1@us.af.mil

**FINANCIAL OR OTHER COMPETING INTERESTS:** As declared above.

Date of Submission: **Mar 13, 2015**

Date of Peer Review: **May 11, 2015**

Date of Acceptance: **Jun 09, 2015**

Date of Publishing: **Aug 01, 2015**