

Rapid Microbial Identification through the Application of Cartridge-Based Optical Mapping



R. Moore, T. Wagner, R. Fraley, C. Cornetta, N. Xiao, A. Briska, T. Heiland, R. Ptashkin, T. Rockweiler, J. Henkhaus, and J. Ray
OpGen, Inc. 708 Quince Orchard Road, Gaithersburg, Maryland 20878

Background

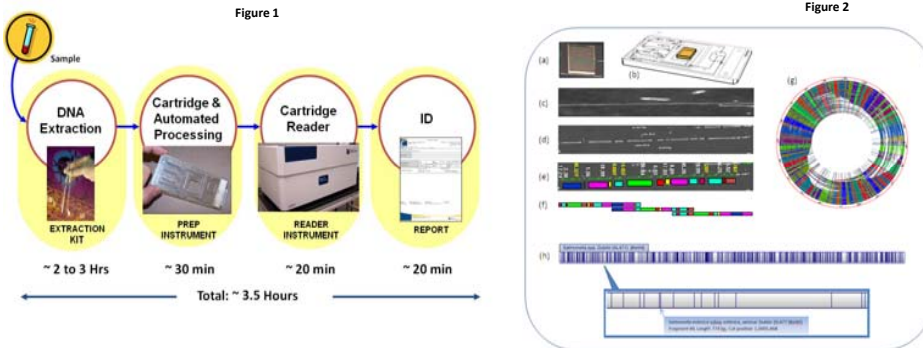
Optical Maps are ordered restriction maps of entire microbial genomes. For a typical 4 million base-pair bacterial chromosome, Optical Maps allow the sizing of 400-600 contiguous fragments, maintaining their relative order. Comparison of sequenced genomes with Optical Maps allows the precise sizing and mapping of chromosomal changes with a 1 to 2 kb limit of resolution across entire genomes. These variations in bacterial chromosomes include insertions, deletions, and large inversions (ref 1).

Optical Mapping technology can also be used for identification (ID) of microorganisms. In this application, restriction maps derived from single linear chromosomal fragments >200kb (molecules) are compared to a database of complete optical maps from known organisms. A molecule with high similarity to a database map is considered to be a 'hit'. A positive ID can be defined after a given number of hits to any given species or isolate in the databases. Since this approach uses information from single molecules, it can be applied to identification of microbes from both pure and mixed samples.

Here we demonstrate the use of a new cartridge-based Optical Mapping system for the identification of bacteria. This cartridge-based implementation of Optical Mapping provides for rapid and simple microorganism ID and may be suitable for clinical applications. The purpose of this study was to assess the ability of a cartridge-based Optical Mapping system to identify unknown bacteria.

Methods

Eight bacterial samples were prepared and 'blinded' by a third party and sent to OpGen, Inc. on slants. Reference ID for these samples was established by Vitek (bioMérieux). Organisms were grown overnight in liquid media and then suspended at a density of 2×10^7 to 1×10^8 /ml. High molecular DNA was purified from the unknown microorganisms using a modified Agencourt Genfind v2 protocol (Beckman Coulter, FL). Unknowns that tested positive in a Gram stain assay were pre-treated 1 hour at 37°C in lysozyme-containing SP Buffer (OpGen, Inc., Gaithersburg, Maryland) to create protoplasts. All the samples were lysed in Agencourt lysis buffer, bound to Agencourt Genfind beads using OpGen Binding Buffer, washed and then eluted in DNA Dilution Buffer (OpGen). In order to reduce shearing of DNA only wide-bore pipette tips were used and DNA-containing solutions were not vortexed. Single molecule restriction maps were generated using both traditional Optical Mapping techniques as previously described (Reference 1) as well as Optical Mapping performed in a cartridge format. An overview of the entire cartridge-based process is shown in Figure 1. Briefly, genomic DNA from each isolate was captured in parallel arrays of long (>200 kb) chromosomal fragments on positively charged glass using a PDMS microfluidic device (Figure 2(a)). A microfluidic pumped cartridge device (b) was used to control digestion of the immobilized chromosomal fragments (c) with a restriction endonuclease (NcoI), to reveal cut restriction sites as "gaps" in the DNA by fluorescence microscopy after staining with the intercalating DYE JoJo-1 (d). The contiguous immobilized restriction fragments were sized by image analysis software that measures the fluorescence associated with each fragment and converts the optical images to ideograms where restriction fragments are represented by colored rectangles (e). The pattern or map of restriction fragments from each molecule is compared to a database of complete restriction maps (g) generated from Optical Mapping or "in silico" from sequenced reference genomes (h). The ID database used in this study contained 588 maps from 79 bacterial species. This database was generated by OpGen, Inc. prior to the start of the study and did not contain maps made from the specific isolates used in the blinded study. To perform microbial ID, each single molecule restriction map is aligned against every optical map in the database, and each pairwise comparison is assigned a p-value according to probability of seeing an alignment at least that good by random chance. If the individual p-value passes a threshold of 0.001, that comparison is considered a hit. The final significance of the microbial ID is the combination of the individual p-values, combined according to Bailey and Grobsov (Reference 2).



Results

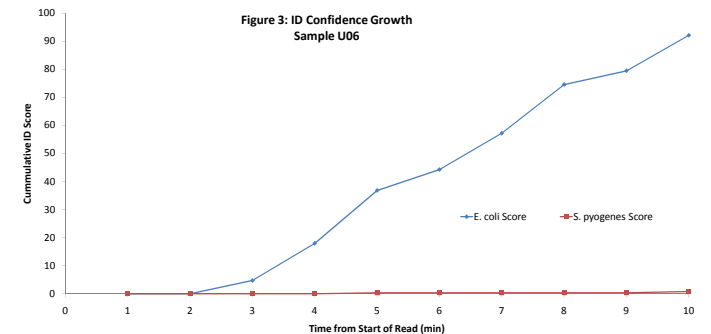
All 8 samples were correctly identified to the species level (Table 1). In this case, the Optical Mapping ID is the species of the database entry with the highest number of 'hits'. In practice, the unknown samples had hits across a number of isolates from the same species and a few hits from other species. An example full 'ID report' for unknown U06 is shown in Table 2. The database entry with the largest number of hits and highest confidence is *E. coli* isolate GLMC 102. This isolate was 'hit' 67 times with molecules that in total covered 2.56MB or just over half of the genome. Several other *E. coli* isolates also had significant numbers of hits. After *E. coli*, the species with the next highest confidence score is *S. pyogenes* with 1 hit, a confidence score of 3 and ~10% genome coverage. Given that the confidence score is closely related to the negative natural log of the p value for ID, this implies that we are $\sim e^{(197/3)}$ or $\sim 10^{28}$ times more confident in the *E. coli* ID than we are of the *S. pyogenes* ID. These extremely high confidence levels imply that reasonably confident ID could be made with far less data. In this study we collected data from a complete MapCard cartridge. This process takes ~1hr and typically generates several thousand single molecule restriction maps. In an effort to understand the amount of data required to make a reasonably confident ID we performed an "in silico" re-analysis of the data collected from sample U06 and computed the ID confidence score as a function of time and amount of data for *E. coli* and the next most confident species.

In our system, DNA flows down channels and is deposited in lanes. Many images are taken and stitched together within each lane to produce a single, long, thin image. This lane or channel image is then processed by the image processing software in near real time. This means that the images for channel 1 are stitched and processed at about the same time as channel 2 has completed image acquisition. Database analysis also happens in near real time. Molecules from channel 1 that are submitted for analysis after channel 2 completes image acquisition have been compared against the database at about the same time that channel 3 completes image acquisition. Each lane or channel of DNA is imaged in ~1 minute. The reanalysis of the first 10 minutes of data acquisition for sample U06 is shown in table 3 and figure 3. Note that scores shown in these figures are aggregate confidence scores for all isolates of the same species and are not pre-filtered for a p threshold of 0.001. Given that a confidence score of ~7 corresponds to a p value of ~0.001, it appears likely that a high confidence ID could be made from this system after several minutes of data collection.

Sample #	Species ID from Optical Mapping	Species ID from Vitek	Hits	Significance Value
U01	<i>S.aureus</i>	<i>S.aureus</i>	84	126.9
U02	<i>S.aureus</i>	<i>S.aureus</i>	92	142.5
U03	<i>S.epidermidis</i>	<i>S.epidermidis</i>	225	433.5
U04	<i>S.epidermidis</i>	<i>S.epidermidis</i>	24	17.40
U05	<i>E.coli</i>	<i>E.coli</i>	51	66.6
U06	<i>E.coli</i>	<i>E.coli</i>	67	197.0
U07	<i>K.pneumoniae</i>	<i>K.pneumoniae</i>	373	687.50
U08	<i>K.Pneumoniae</i>	<i>K.Pneumoniae</i>	423	802.50

Time (min)	Channels Imaged	Channel Images Processed	Molecules Submitted for ID	Molecules Analyzed	<i>E. coli</i> Score	<i>S. pyogenes</i> Score
1	1	0	0	0	0	0
2	2	1	137	0	0	0
3	3	2	274	137	5	0
4	4	3	411	274	18	0
5	5	4	548	411	37	0
6	6	5	685	548	44	0
7	7	6	822	685	57	0
8	8	7	959	822	75	0
9	9	8	1096	959	79	0
10	10	9	1233	1096	92	1

Hits	Confidence Score	% Genome	Mb	Name
67	197	51.8	2.56	<i>Escherichia coli</i> _(GLMC_102),
14	37	13.4	0.68	<i>Escherichia coli</i> _(SMS-3-5,
5	12	5	0.23	<i>Escherichia coli</i> _(HS,
2	6	4.2	0.21	<i>Escherichia coli</i> _(536,
2	5	10.1	0.49	<i>Escherichia coli</i> _(CV1129347),
1	4	3.6	0.17	<i>Escherichia coli</i> _(CB-6),
1	4	3.5	0.17	<i>Escherichia coli</i> _(GLMC_504),
1	3	10.1	0.19	<i>Streptococcus pyogenes</i> _(MGAS9429,
1	3	8.4	0.19	<i>Neisseria meningitidis</i> _(FAM18,
1	3	5.6	0.26	<i>Escherichia coli</i> _(Scarab1),
1	3	3.9	0.18	<i>Escherichia coli</i> _(GLMC_21),
1	3	3.7	0.17	<i>Escherichia coli</i> _(DHSAlpha),
1	3	3.6	0.17	<i>Escherichia coli</i> _(K12,
				<i>Escherichia coli</i> _(str_K-
1	3	3.6	0.17	<i>12_substr_DH10B,</i>
1	3	3.6	0.17	<i>Escherichia coli</i> _(W3110,
1	3	3.5	0.18	<i>Escherichia coli</i> _(GLMC_3),
1	3	3.5	0.17	<i>Escherichia coli</i> _(SE11,
1	3	3.5	0.17	<i>Escherichia coli</i> _(CB-3),
1	3	3.4	0.17	<i>Escherichia coli</i> _(CV1131494),
1	3	3.3	0.17	<i>Escherichia coli</i> _(UTI89,
1	3	3.3	0.16	<i>Escherichia coli</i> _(CV1118322),
1	3	3.2	0.16	<i>Escherichia coli</i> _(CV1118940),
1	3	2.7	0.13	<i>Salmonella enterica</i> _(subsp._enterica
1	3	2.6	0.14	<i>Escherichia coli</i> _(CFT073,
1	3	2.3	0.13	<i>Escherichia coli</i> _(O157:H7_EDL933,



Conclusions

This study demonstrates that Optical Mapping can produce extremely high confidence species level ID in less than 4 hours from positive cultures. The high confidence of the IDs is likely related to the fact that molecules compared against the database have an average size of ~225kb so even 1 'hit' covers restriction sites that span ~5% of a typical 5MB organism. In a clinical setting, it is likely that a user would use a universal lysis and extraction protocol that might require as long as 3 hours. Card preparation takes approximately 30 minutes. This study demonstrates that data collection and data analysis could likely take place in less than 30 minutes.

In addition to delivering rapid ID results from positive cultures, Optical Mapping is a de novo, single molecule technology. Universal reagents can be used for most samples and organisms and that isolated colonies are not required. Future work will focus on improving sample prep methods to enable use of Optical Mapping for ID directly from some clinical samples.

References

- Cai W, Jing J, Irvin B, Ohler L, Rose E, Shizuya H, Kim UJ, Simon M, Anantharaman T, Mishra B, Schwartz DC (1998): High-resolution restriction maps of bacterial artificial chromosomes constructed by optical mapping. *PNAS* 95:3390-3395.
- Bailey and Grobsov, "Combining evidence using p-values: application to sequence homology searches." *Bioinformatics*, 14, p. 48-54, 1998.