Whole-Genome Shotgun Optical Mapping of *Rhodospirillum rubrum*

Running Title: Optical Mapping of *Rhodospirillum rubrum*

Susan Reslewic[1,2.7], Shiguo Zhou[1,2,7], Mike Place[1,2,7], Yaoping Zhang[3], Adam Briska[4], Steve Goldstein[1,2,7], Chris Churas[1], Rod Runnheim[1], Dan Forrest[1], Alex Lim[1,2], Alla Lapidus[5], Cliff S. Han[6], Gary P. Roberts[3], David C. Schwartz[1,2,7*]

[1]The Laboratory for Molecular and Computational Genomics, University of Wisconsin-Madison, Madison, Wisconsin 53706 USA; [2]Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin, 53706 USA; [3]Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin 53706 USA; [4]OpGen, Inc., Madison, Wisconsin, 53719 USA; [5]Microbial Genomics DOE Joint Genome Institute 2800 Mitchell Drive, B400 Walnut Creek, California 94598 USA; [6]Los Alamos National Laboratory Center for Human Genome Studies, Los Alamos, New Mexico 87545 USA; [7]Laboratory of Genetics, University of Wisconsin-Madison, Madison, Wisconsin 53706 USA

*Corresponding author

Laboratory for Molecular and Computational Genomics, University of Wisconsin-Madison, UW-Biotechnology Center, 425 Henry Mall, Madison, Wisconsin, 53706, USA

Email: dcschwartz@facstaff.wisc.edu; Phone: 608-265-0546; Fax: 608-265-6743

ABSTRACT

Rhodospirillum rubrum is a phototrophic purple non-sulfur bacterium known for its unique and well-studied nitrogen fixation and carbon monoxide oxidation systems, and as a source of hydrogen and biodegradable plastics production. To better understand this organism and to facilitate assembly of its sequence, three whole-genome restriction maps (*Xba* I, *Nhe* I, and *Hind* III) of *R. rubrum* strain ATCC 11170 were created by optical mapping. Optical mapping is a system for creating whole-genome ordered restriction maps from randomly sheared genomic DNA molecules extracted directly from cells. During the sequence finishing process, all three optical maps confirmed a putative error in sequence assembly, while the *Hind* III map acted as a scaffold for high resolution alignment with sequence contigs spanning the whole genome. In addition to highlighting optical mapping's role in the assembly and validation of genome sequence, our work underscores the unique niche in resolution occupied by the optical mapping system. With a resolution ranging from 6.5 kb (previously published) to 45 kb (reported here), optical mapping advances a "molecular cytogenetics" approach to solving problems in genomic analysis.

INTRODUCTION

Rhodospirillum rubrum is one of two spiral-shaped *Rhodospirillum* species belonging to the alpha-Proteobacteria class of prototrophic purple non-sulfur bacteria (13, 14, 23). Found in aquatic environments like lakes, streams, and standing water, *R. rubrum* is known for its unique carbon and nitrogen metabolism (5, 13, 14, 27), and its potential to produce hydrogen and biodegradable plastics (6, 9, 12, 24, 25, 28). Specifically, *R. rubrum* possesses the very rare ability to oxidize carbon monoxide to carbon dioxide, and has been the subject of studies seeking to understand the mechanisms and regulation of this process (5). *R. rubrum* has proven to efficiently convert hydrogen to current in fuel cells (12), and produce novel forms of biodegradable thermoplastics when grown on assorted β-hydroxycarboxylic acids and n-alkanoic acids (6). More recently, Handrick *et al*. (12) reported on *R. rubrum's* activator role in the degradation of PHB, a polymer of interest for its thermoplasticity and breakdown to water and carbon dioxide. Finally, from a biochemical standpoint, Zhang and others (29)studied the mechanisms behind *R. rubrum's* posttranslational regulation of nitrogenase activity.

To supplement our understanding of *R. rubrum*'s unique biology, the organism was sequenced by the Department of Energy Joint Genome Institute (DOE JGI) and finished by the Los Alamos Finishing Group, a part of the DOE JGI at the Los Alamos National Laboratory. In order to aid in sequence assembly and validation, we created three whole-genome restriction maps of *R. rubrum* with resolution ranging from 11 to 45 kb. Physical maps are an excellent means by which to independently validate sequence, close sequence contig gaps, and resolve repeat-rich regions, which consistently confound sequence assembly methods (10, 15-19, 22). The automation and unique resolution of

optical mapping make this system ideal for addressing a wide range of problems in genomics, including the finishing and validation of microbial sequencing projects (4, 18, 30-32).

Optical mapping enables the construction of whole-genome restriction maps from ensembles of single DNA molecules that have been elongated and immobilized on positively charged glass surfaces and subsequently cut with a restriction enzyme. The resulting single DNA molecule restriction maps are stained with a fluorochrome and visualized by fluorescence microscopy. Because the order of restriction fragments is retained on the optical mapping surface, there is no need for sorting fragments by size. The mass of each restriction fragment is determined by integrated fluorescence intensity measurements. The single molecule restriction maps are assembled into contigs (2, 3, 17, 19), in a process similar to shotgun sequence assembly, that span entire microbial genomes.

Here we present three optical maps (*Xba* I, *Nhe* I, and *Hind* III) of the *R.rubrum* strain ATCC 11170 genome that aided in sequence assembly and validation. The high resolution *Hind* III map confirmed the correct placement of the sequence contigs generated during the finishing process, and the lower resolution *Nhe* I and *Xba* I maps confirmed the final sequence assembly. All three maps confirmed a sequence assembly error, the correction of which filled a gap in the sequence. With a resolution of 45 kb, the *Xba* I map is the lowest resolution optical map reported to date. With a documented resolution of 6.5 kb (30) to 45 kb, optical mapping fills a critical niche in the resolution capabilities of genomic analysis systems. Here we show the utility of optical mapping's

resolution range in microbial sequence assembly and validation, and comment briefly on the advantages of optical mapping for solving additional problems in genomics.

MATERIALS AND METHODS

DNA preparation. *R. rubrum* strain ATCC 11170 genomic DNA gel inserts (26) were prepared from a culture grown aerobically at 30°C in SMN medium as described previously (8) and stored in 0.5 M EDTA (pH 8.0). Prior to use, the DNA inserts were washed thoroughly overnight in TE (10 mM Tris, 1 mM EDTA; pH 8.0) to remove excess EDTA. After melting the agarose inserts at 72°C for 7 min, the agarose was digested at 42°C for 2 h in β-agarase solution (100 μL TE, 1 μL [1 unit] β-agarase [New England Biolabs]). The resulting concentrated DNA was diluted in TE to ensure minimal crowding of single DNA molecules on the optical mapping surfaces. Lambda DASH II bacteriophage DNA (Stratagene) was added to the genomic DNA dilution (10 pg/μL) as an internal standard for fragment sizing. The samples were mounted onto an optical mapping surface and inspected by fluorescence microscopy (details below) for molecular integrity and appropriate concentration.

Surface preparation. Glass cover slips (22 x 22 mm, Fisher's Finest; Fisher Scientific) were cleaned and derivatized as described previously (30). Surface properties were assayed by digesting lambda DASH II bacteriophage DNA with 40 units of *Xba* I, *Hind* III, and *Nhe* I diluted in 200 μL of digestion buffer with 0.2% Triton X-100 (SIGMA) at 37° C to determine optimal digestion times, which ranged from 30 – 120 minutes.

DNA mounting, overlay, digestion, and staining. DNA molecules were mounted on derivatized glass surfaces by capillary action using a microfluidic device (7). Following DNA elongation and deposition, a thin layer of acrylamide (3.3% containing 0.02% Triton X-100 [SIGMA]) was applied to the surface. After polymerization, the surfaces were washed with 400 μL of TE for 2 min, followed by washing with 200 μL of digestion buffer for another 2 min. The digestion was then performed by adding 200 μL of digestion buffer with enzyme (20 μL of NEB [New England Biolabs] Buffer 2, 176 μL high purity water, 2 μL 2% Triton X-100 [SIGMA] and 4 μL of NEB-*Hind* III or *Eco*RI (10 unit/μL) or 2 μL NEB-*Xba* I (20 units/μL)) to the surface and incubating in a humidified chamber at 37°C for 30 – 120 min. Following digestion, the surfaces were washed twice by adding 400 μL of TE, waiting 2-5 min, and the solution removed by aspiration. The surfaces were mounted onto a glass slide with 12 μL 0.2 μM YOYO-1 solution {containing 5 parts YOYO-1; 1,1'-[1,3-propanediylbis[(dimethyliminio) -3,1-propanediyl]]bis[4-[(3-methyl-2(3H)- benzoxazolylidene)-methyl]]-,tetraiodide [Molecular Probes] in 95 parts of β-mercaptoethanol in TE 20% v/v}. The edges of the glass surface were sealed to the glass slide with nail polish and incubated (4°C in the dark) for at least 20 min so the staining dye could diffuse before checking by fluorescence microscopy.

Image Acquisition and Processing. The samples were imaged by fluorescence microscopy as previously described (18) using a 63x objective (Zeiss) and a high-resolution digital camera (Princeton Instruments). Single overlapping images, spanning the full length of the microfluidic channels, were collected, flattened, and superimposed by a fully automated image acquisition system, ChannelCollect (7) developed by our

laboratory. These flattened and overlapped superimages were then processed through the Pathfinder software, which identifies digested molecules to be made into single molecule maps. Features that are recognized as DNA molecules are denoted and created into an ordered restriction map for that molecule. Co-mounted Lambda DASH II molecules were used to estimate the digestion rate and to provide internal fluorescence standards for accurately sizing the DNA fragments (1, 19, 20). Each digested genomic DNA molecule selected by Pathfinder becomes a single molecule optical map.

Optical map assembly. The custom written software, "Gentig" (1-3, 17-19), overlapped the single-molecule restriction maps by aligning restriction sites based on fragment sizes. Via a greedy algorithm with limited backtracking, Gentig assembles the individual molecule restriction maps into a contig that spans the entire genome. Gentig avoids the great computational complexity that would occur in attempts to find the optical assembly, and instead finds an almost optimal scoring set of map contigs. Bayesian inference estimates the probability that two single-molecule restriction maps, while subject to various data errors stemming from sizing, missing restriction sites (missing cuts), and spurious restriction sites (false cuts), may have been derived from the proposed placement. A known statistical distribution of the error sources is required for the Bayesian approach, as is fine-tuning of parameters such as standard deviation, digestion rate, false cut, and false match probability. These parameters can be re-estimated from the data using a limited number of iterations of Bayesian probability density maximization. Once these parameters have been accurately estimated from the data, an efficient dynamic programming algorithm computes the best offset and alignment between a pair of maps.

DNA Sequencing. Two randomly sheared libraries of the *R. rubrum* strain ATCC 11170 genome were produced with inserts in the 3kb (plasmids) and 40kb (fosmids) size range. These libraries were sequenced to a total depth of approximately 11X and all reads quality assessed and trimmed for vector before being used for assembly.

For the 3 kb DNA shearing and plasmid sub-cloning, approximately 3-5 ug of isolated DNA was randomly sheared to 3-4 kb fragments (25 cycles at speed code 12) in a 100 μL volume using a HydroShear™ (GeneMachines, San Carlos, CA). The sheared DNA was immediately blunt end-repaired at room temperature for 40 min using 6 U of T4 DNA Polymerase (Roche), 30 U of DNA Polymerase I Klenow Fragment (NEB, Beverly, MA), 10 μL of 10 mM dNTP mix (Amersham), and 13 μL of 10x Klenow Buffer in a 130 μL total volume. After incubation the reaction was heat inactivated for 15 min at 70°C, cooled to 4°C for 10 min and then frozen at -20°C for storage. The end-repaired DNA was run on a 1% TAE agarose gel for ~ 30-40 min at 120 volts. Using ethidium bromide stain and UV illumination, 3-4 kb fragments were extracted from the agarose gel and purified using QIAquick™ Gel Extraction Kit (QIAGEN). Approximately 200-400 ng of purified fragment was blunt-end ligated for 40 min into the *Sma* I site of 100 ng of pUC 18 cloning vector (Roche) using the Fast-Link™ DNA Ligation Kit (Epicentre, Madison, WI). Following standard protocols, 1 μL of ligation product was electroporated into DH10B Electromax™ cells (Invitrogen, Carlsbad, CA) using the GENE PULSER® II electroporator (Bio-Rad, Hercules, CA). Transformed cells were transferred into 1000 μL of SOC and incubated at 37°C in a rotating wheel for 1 h. Cells (usually 20-50 μL) were spread on 22 x 22 cm LB agar plates containing 100

μg/mL of ampicillin, 120 μg/mL of IPTG, and 50 μg/mL of X-GAL. Colonies were grown for 16 h at 37°C. Individual white recombinant colonies were selected and picked into 384-well microtiter plates containing LB/glycerol (7.5%) medium containing 50 μg/mL of ampicillin using the Q-Bot™ multitasking robot (Genetix, Dorset, U.K.). To test the quality of the library (XYG), 24 colonies were directly PCR amplified with pUC m13 -28 and -40 primers using standard protocols. Libraries are considered to pass PCR QC if they have > 90% 3 kb inserts. For more details see research protocols at www.jgi.doe.gov.

For the plasmid amplification and sequencing steps, 2 μL aliquots of saturated *E. coli* cultures (DH10B) containing pUC18 vector with random 3-4 kb DNA inserts grown in LB/glycerol (7.5%) medium containing 50 μg/mL of ampicillin were added to 8 μL of a 10 mM Tris-HCl pH 8.2 + 0.1 mM EDTA denaturation buffer. The mixtures were heat lysed at 95°C for 5 min then placed at 4°C for 5 min. To these denatured products 10 μL of an RCA reaction mixture (TempliPhi™ DNA Sequencing Template Amplification Kit, Amersham Biosciences) were added. The amplification reactions were carried out at 30°C for 12-18 hr. The amplified products were heat inactivated at 65°C for 10 min then placed at 4°C until used as template for sequencing. Aliquots of the 20 μl of amplified plasmid RCA products were sequenced with standard M13 –28 or –40 primers. The reactions contained 1 μL of RCA product, 4 pmoles of primer, 5 μL of dH_2O, and 4 μL of DYEnamic™ ET terminator sequencing kit (Amersham Biosciences). Cycle sequencing conditions were 30 rounds of 95°C-25 sec, 50°C-10 sec, 60°C-2 min, and then held at 4°C. The reactions were then purified by a magnetic bead protocol (see research protocols, www.jgi.doe.gov) and run on a MegaBACE 4000 (Amersham Biosciences).

Alternatively, 1 μL of the RCA product was sequenced with 2 pmoles of standard M13 –28 or –40 primers, 1 μL 5x buffer, 0.8 μL H_2O, and 1 μL BigDye sequencing kit (Applied Biosystems) at 1 min denaturation and 25 cycles of 95°C-30 sec, 50°C-20 sec, 60°C-4 min, and finally held at 4°C. The reactions were then purified by a magnetic bead protocol (see research protocols, www.jgi.doe.gov) and run a ABI PRISM 3730 (Applied Biosystems) capillary DNA sequencer. Detailed protocols for fosmid library creation, fosmid DNA isolation and cleanup procedure can be found at http://www.jgi.doe.gov/Internal/protocols/prots_production.html.

In the sequence finishing process, all drafted reads were assembled together with SPS Phrap (SPSOFT, Albuquerque, NM). Repetitive regions of the genome were resolved with repFinisher (Cliff S. Han, unpublished). Autofinish (11) was used in the first cycle of finishing to select sequencing reactions. Remaining gaps and low quality regions closed with primer walking on subclones or by shattering PCR fragments covering the gaps.

Optical maps versus *in silico* maps. Alignments between the optical maps and *in silico* maps from the seven finishing-stage sequence contigs were created with the MapViewer software, a Perl/Tk application that provides an intuitive graphical interface for optical map analysis. In addition to creating and displaying alignments of optical maps, MapViewer allows the user to manipulate the relative positions and orientations as well as the scale of the optical maps to better understand these alignments. The map alignments are generated with a dynamic programming algorithm that finds the optimal alignment of two restriction maps according to a scoring model that incorporates fragment sizing errors, false and missing cuts, and missing small fragments. For a given

alignment, the score is proportional to the log of the length of the alignment, penalized by the differences between the two maps, such that longer, better matching alignments will have higher score.

Using Gentig, the *Xba* I, *Nhe* I, and *Hind* III maps were aligned separately with the *in silico Hind* III, *Xba* I, and *Nhe* I maps generated from the finished sequence. These initial alignments enabled determination of missing fragments, false cuts or missing cuts. The relative sizing error for each fragment in the optical maps was calculated from the formula [100 × (|optical map fragment size – corresponding *in silico* map fragment size|)/corresponding *in silico* map fragment size] and was plotted against the *in silico* map fragment sizes to show the relationship between fragment size and relative error.

RESULTS

Acquisition of optical map data and construction of optical maps. The three restriction maps presented here were created via whole-genome shotgun optical mapping (1-3, 17, 19). This mapping strategy has parallels to whole genome shotgun sequencing; large numbers of optical maps, analogous to "sequence reads", are assembled to cover any given locus. Depth of coverage minimizes mapping error and the overlapping cascades of optical maps create continuity of coverage across an entire genome. The use of genomic DNA as the source of single molecules for mapping eliminates the need for libraries, PCR, or separations and thus makes optical mapping advantageous for whole genome mapping and sequence assembly.

Our sample preparation methods create randomly sheared DNA molecules (200 – 3000 kb) which were elongated and immobilized onto the optical mapping surfaces,

digested with three restriction enzymes (*Hind* III, *Xba* I, and *Nhe* I) in separate experiments, and imaged by fluorescence microscopy using ChannelCollect (Materials and Methods). Maps derived from images of digested single molecules were assembled into genome-wide consensus maps using Gentig (1-3, 17).

The resolution of the optical maps affects the contig rate and average molecule size in the contig (Table 1). For the *Xba* I map, a total of 405 digested molecules were imaged and processed. Of this total, 204 were included into the final map contig, giving a contig rate of 50.37%. This low contig rate can be explained in part by the low resolution of the optical map. A map with an average fragment size of 44.73 kb requires very large genomic DNA molecules for contig assembly, due to the number of fragments in a single molecule map required for confidence in merging that map into a map contig. The average size of molecules in the *Xba* I contig was about 900 kb, whereas the average size of collected DNA molecules was 637.69 kb. In addition, the digestion rate was calculated at 76.01%, which is lower than our target digest rate of 80% or higher. While still acceptable, a digest rate of 76.01% will reduce the density of apparent restriction sites, or markers, thus increasing difficulty in creating a contig. However, even with the low rate of contig formation, the total mass of molecules in the contig was 184.23 Mb, which corresponds to about 42x coverage. Figure 1 shows the finished *Xba* I map, which resembles a classical macrorestriction map in terms of resolution. Notably, the contig was circularized without gaps, and a typical restriction fragment was calculated from the average of about 30 fragments. The sizing error per single fragment, or precision, was calculated from the set of restriction fragments used to determine each consensus fragment. The average standard deviation of fragment size about the mean was 4.31 kb.

The size of the *R. rubrum Xba* I circular contig was 4323.08 kb, and was calculated by summing the restriction fragments in the *Xba* I consensus map. This map is the lowest resolution optical map created to date.

With an average fragment size of 31.53 kb, the *Nhe* I map has a resolution in between those of the *Xba* I and *Hind* III maps. A total of 409 molecules were collected and processed to form the *Nhe* I map. Of the total, 345, or 84.35% of the molecules went into the circular contig. The average molecule size of the collected molecules was 635.72 kb, only 2 kb smaller than average size of molecules collected for the *Xba* I map. However, the average size of molecules in the contig was 731.30 kb, almost 200 kb smaller than the average size of the molecules in the *Xba* I contig. As the average fragment size of the *Nhe* I map is smaller, a molecule of a given size will have more restriction fragments, and smaller molecules can be included into the contig. Thus, in comparison to the *Xba* I map, the increased *Nhe* I contig rate is due in part to the smaller average fragment size. In addition, the digestion rate was calculated to be 87%, which means that the patterns were more informative and accurate in this map. The mass of the molecules in the contig was 252.30 Mb; this represents approximately 60x coverage. The contig circularized without gaps, and about 42 fragments were used to calculate a given fragment mass in the consensus map. Based on the final *Nhe* I consensus map, the average standard deviation of the fragment size about the mean was 2.83 kb. Summing the masses of all of the restriction fragments in the consensus map gave a total genome size of 4223.13 kb.

Finally, 932 molecules were collected for the production of the *Hind* III map. With an average fragment size of 10.95 kb, this map is the highest resolution of the set of

maps presented here. Of the 932 maps, 623, or 66.84% of maps went in to the final contig. The smaller average fragment size loosened the requirements for molecule size in the contig; the average size of molecules in the contig was 405.49 kb. The total mass represented by the molecules in the contig was 252.68 Mb, which corresponds to about 57x coverage, based on the *Hind* III contig size of 4456.35 kb. Again, the size of the *Hind* III contig was calculated by summing the masses of the restriction fragments in the consensus map. The digestion rate of the molecules in the contig was calculated to be 78.9%. An average of about 31 fragments was used to calculate the mass of each fragment in the consensus map. For each fragment in the consensus map, the standard deviation about the mean was 1.16 kb.

In all of the maps, the high coverage ensured accurate calling of restriction sites, fragment sizing, and sizing of the entire circularized genome map. Below, the accuracy of the optical maps as compared to the sequence is assessed. The accuracy of an optical map as its own entity is estimated by Gentig's (see Materials and Methods) ability to assess a set of hypothetical maps against the optical map data set and, using error models, report a false-positive probability. The false circularization probability for the *Xba* I, *Nhe* I, and *Hind* III maps were 0.00738, 0.00329, and 0.00440, respectively. Since the *Xba* I map had the lowest coverage, contig rate, and digestion rate, it is not surprising that the false circularization probability for this map is the highest. However, for all the maps, the false probability values were below 0.05, which is considered to be the upper limit for confident map circularization (30). The restriction patterns generated by the *Xba* I, *Nhe* I, and *Hind* III maps appeared random; no particular restriction patterns or structural features were observed.

Use of optical maps in sequence assembly. All of the optical maps were made in order to guide and validate the *R. rubrum* genome sequence assembly process. Near the end of the finishing effort, nine sequence contigs were generated ranging in size from 2.311 kb to 1465.886 kb. Alignment of the optical maps against the *in silico* maps of the sequence contigs gave three independent indications of sequence contig assembly and order. Two of the sequence contigs did not aligned against the optical maps. They were contig 83, the plasmid sequence contig, and contig 82c, which, with a size of 2.311 kb, was too small to align with the optical maps. Six of the seven remaining contigs aligned to both the *Xba* I and *Nhe* I maps. Only the *Hind* III map, with its higher resolution, was able to align all seven sequence contigs, including contig 83 with its small size of 80.404 kb (Figure 2A). All three comparisons of optical map to sequence supported a problematic assembly of sequence contig 90. Alignment of the *Nhe* I map to the sequence contigs best illustrates this (Figure 2B). Contig 87 and the right most eight fragments of contig 90 align to the region in red in the *Nhe* I map. Inspection shows a cleaner alignment of the region with contig 87. Removing the leftmost eight fragments from contig 90 and inverting the orientation produced a solid alignment with the gap in the *Nhe* I map that was between contig 90 and contig 85c (Figure 2C). Our realization of this problematic assembly confirmed the Los Alamos finishing group's suspicions of an erroneous assembly in this region (Cliff S. Han, personal communication). Elsewhere in the genome, we found good agreement between sequence contigs and optical maps.

The finished sequence contained a 4.353 Mb circular chromosome (contig 94) and a 54.412 kb plasmid (contig 93). Minor differences have been found between the optical maps and finished sequence and are described below.

Assessment of Optical Mapping Errors. Comparisons between sequence and optical mapping data were made in order to evaluate the errors and accuracy in the *Xba* I, *Nhe* I, and *Hind* III maps (Figure 3). The relative sizing error was calculated by the alignment of optical maps with the *in silico* maps made from the finished sequence (Figure 3A-F). The error bars in Figures 3A-C reflect the standard deviation about the means of the restriction fragment sizes used in calculating the consensus map fragments shown in Figures 5 and 6. In general, a high degree of correspondence was evident between the optical map and *in silico* map fragment sizes. The regression values for the trendlines are 0.9985, 0.9995, and 0.9947 for the *Xba* I, *Nhe* I, and *Hind* III maps respectively. Figures 3 D-F are scatter plots showing the relationship between absolute relative fragment sizing error (optical map versus *in silico* map) and restriction fragment size. For the *Xba* I map, the average relative sizing error [100 × (|optical map fragment size – corresponding *in silico* map fragment size|)/corresponding *in silico* map fragment size] was 6.20% for fragments smaller than 5 kb and 2.87% for fragments larger than 5 kb. For fragments smaller than 5 kb in the *Nhe* I map, the average relative sizing error was 5.87%, and was 3.11% for fragments larger than 5 kb. Finally, for the *Hind* III map, the average relative sizing error was 16.70% for fragments smaller than 5 kb and 8.05% for fragments larger than 5 kb. The positive skew in the relative error plots for small fragments in Figure 3 are discussed in greater detail in the following section.

Figure 4 shows the cumulative distribution of fragment sizes for the three optical maps. Only the *Xba* I map has fragments greater than 135 kb, thus this value was chosen as an endpoint in the figure to facilitate visual comparison of the three maps' fragment size distributions. Each bar represents the cumulative percentage of consensus map

fragments in 5 kb intervals. For each of the maps, the distribution is roughly exponential as expected. One key difference between the *Hind* III and the lower resolution *Xba* I and *Nhe* I maps is the proportion of fragments smaller than 5 kb and 10 kb. In the *Hind* III map, about 35% of all fragments in the consensus map are smaller than 5 kb. 72% of fragments are smaller than 10 kb, and 100% of fragments are under 40 kb (the largest fragment is 37.82 kb). These numbers are in stark contrast to those for the *Nhe* I and *Xba* I maps. In the *Nhe* I map, only 11.6 % of fragments are smaller than 5 kb, and 31.9% of fragments are smaller than 10 kb. Similarly, in the *Xba* I map, only 13.3% of fragments are smaller than 5 kb and 25.3% of fragments smaller than 10 kb. For the *Xba* I map, there were only 2 additional fragments greater than 135 kb: a 242.20 kb fragment and a 256.30 kb fragment (not shown). The increased average relative sizing error for small fragments (Figure 3F) seen in the *Hind* III map may be due to the high proportion of fragments 2 kb or smaller, many in tandem with each other, in this high resolution map.

Comparing Optical Maps to the Sequence. An assessment of the previously described errors in the context of optical map to sequence alignment is necessary for distinguishing random errors from those that may consistently point to discrepancies between optical maps and sequence. Figure 5 shows the *Xba* I, *Nhe* I, and *Hind* III alignments of the consensus optical map to the corresponding *in silico* maps. These circular, more global alignments are linearized and expanded in Figure 6 in order to show the exact locations of discrepancies between the sequence and the optical maps.

The alignment of the *Xba* I map with the *in silico* map showed that there were no false cuts (apparent in the optical map but not in the *in silico* map) and 12 missing cuts (apparent in the *in silico* map but not the optical map) out of a total of 100 *Xba* I cuts in

the *in silico* map. Optical maps normally do not report restriction fragments smaller than 500 bp, and, due to the resolution of optical mapping, reporting of fragments smaller than 1 kb is incomplete (21)**.** The *Xba* I map had no missing fragments over 500 bp. Out of 100 fragments, the *in silico* map showed two fragments smaller than 500 bp, and two fragments between 500 bp and 1 kb.

In comparison to the *in silico* map, the *Nhe* I map showed no false cuts and one missing cut out of a total of 145 cuts in the *in silico* map. There were four missing fragments, over 500 bp, in the *Nhe* I map. The *in silico* map had no fragments smaller than 500 bp, and three fragments smaller than 1 kb, out of a total of 145 fragments.

Finally, the *Hind* III map showed no false cuts and five missing cuts in comparison to the 684 cuts in the *in silico* map. Of the 684 fragments, 664 were greater than 500 bp. Of these fragments, 125 were missing in the *Hind* III optical map. Fifty-eight of the missing fragment loci, corresponded to *in silico* fragments > 500 bp and $\leq$1 kb, 59 to fragments >1 kb and $\leq$2 kb, and the remaining eight to fragments >2 kb and <3 kb.

Comparing the locations of the missing cuts and missing fragments revealed no consistent errors among the three optical maps. Thus, errors appear to be random and not associated with any major discrepancy between the sequence and the optical maps.

DISCUSSION

The goal of whole-genome optical mapping of *R. rubrum* strain ATCC 11170 was to aid in sequence assembly and finished sequence validation. The enzymes *Xba* I, *Nhe* I, and *Hind* III were selected because of their different cutting frequencies. The advantages

of both low and high resolution optical maps in sequence assembly are demonstrated here. The high resolution *Nhe* I map was able to align and order the seven sequence contigs (not including the ~2 kb contig) generated at the end of the finishing effort without gaps. While the error in contig 90 was evident in the *Hind* III optical map to sequence alignment, but the lower resolution *Xba* I and *Nhe* I maps best displayed this error and how it could be corrected. All three maps were used to validate the final sequence 4.353 Mb sequence contig generated by the Los Alamos finishing group.

The finished *R. rubrum* strain ATCC 11170 sequence size of 4.353 Mb is closest the estimate of 4.323 Mb provided by the *Xba* I map. The overall sizing error for the *Xba* I map is 0.7%, which is smaller than the error associated with other whole-genome physical maps generated by pulsed-field gel electrophoresis (32). The sizing errors for the *Nhe* I map and *Hind* III map were 3% and 2% respectively. Yet, the alignment of the *Nhe* I and *Hind* III optical maps against the *in silico* maps showed no apparent overall size discrepancies, and thus this error most likely stems from the summation and increased error associated with small fragments. As the number of fragments summed to calculate genome size increases, so does the error associated with this calculation. As such, the low resolution *Xba* I map should, and does, give the most accurate estimate of genome size.

The high number of missing fragments in the *Hind* III map and increased sizing error of small fragments illustrate the challenges optical mapping faces for scoring of small fragments. Of the 125 missing fragments in the *Hind* III optical map, 116 corresponded to *in silico* map fragments less than or equal to 2 kb. This corresponds to a small fragment loss rate of 75%, as the *in silico* map contained 154 fragments less than or

equal to 2 kb. By contrast, the fragment loss rate for fragments >2 and ≤3 kb was 11.6% (8 out of 69 fragments were missing), and zero for fragments greater than 3 kb. A key element of the optical mapping system is the elongation and immobilization of single DNA molecules onto glass surfaces. The immobilization via electrostatic interactions between the negatively charged DNA and positively charged glass surface must be subtle enough to enable biochemical reactions, such as a restriction digest, yet strong enough to retain the resulting fragments. The loss of fragments 2 kb and smaller reflects the difficulty in retaining small fragments in their exact position on the surface after a restriction digest, but also, in identifying and correctly sizing the fragments during image acquisition and subsequent processing. The error models in the optical map assembly software (Gentig) take into account the likelihood of losing small fragments and enable alignment against the sequence, as seen here in the *Hind* III map, despite the significant small fragment loss. An increased positive sizing error is seen in both the *Hind* III map and *Nhe* I map for small fragments. One possible explanation is the likelihood of overestimating the size of small fragments when they are scored. In other words, when a small fragment is marked, it is unlikely that the fragment would be undersized, and thus errors in this size range do not balance each other as well as they do for the larger. New efforts in DNA mounting and small fragment sizing with the Pathfinder software are currently underway in order to improve our retention and scoring of small fragments.

With an average fragment size of 44.73 kb, the *Xba* I map represents the lowest resolution optical map created in our lab. There are significant advantages of a low resolution map. First, a low resolution map requires very large single molecule for assembly into a whole-genome contig. As average fragment sizes increases, so does the

molecule size required for achieving a unique pattern of restriction fragments for accurate map assembly. Here, the average size of molecules in the *Xba* I map approached 1 Mb. This scale approaches the lower resolution limit of more global cytogenetic methods that reveal chromosomal insertions, deletions, rearrangements, etc. With a documented resolution between 6.5 kb (32) and 45 kb (reported here), optical mapping's niche falls between low resolution, global methods, such as comparative genomic hybridization (CGH) and very high resolution genotyping systems. This "molecular cytogenetics" approach has enormous potential for aiding in large genome (such as mammalian) sequencing projects as well as for identifying genomic variation in the form of insertions, deletions, and repetitive elements, a difficult and often evasive task. With the ability to qualify conclusions drawn from low resolution cytogenetic techniques and contextualize the information gleaned from high resolution genotyping tools, the optical mapping system can be particularly powerful when used in conjunction with other methods. We are currently pursuing these directions with the optical mapping system, as well as working on improvements for larger molecules and improved small fragment retention for the goal of widening the range of optical mapping's resolution.

Here we have shown three optical maps that have aided in sequence assembly and validation of *R. rubrum*. In addition, we have widened the resolution range of the optical mapping system and contextualized this contribution to genomic analysis. Continual improvements and new applications of the optical mapping system are underway in our lab. For example, in a recent comparative genomics study, optical mapping revealed novel genomic insertions and rearrangements in *Shigella flexneri*, in addition to genomic differences between sequenced strains of *Escherichia coli* and *Yersinia pestis* that were

aligned as maps (31). Optical mapping's role in sequencing projects has expanded to larger, more complex genomes, such as the ~34 Mb genome of the diatom *Thalassiosira pseudonana* (4). Optical mapping projects will continue to encompass increasingly challenging questions, with the goal of providing new insights on genome structure and organization that will potentiate the capabilities of higher and lower resolution genomic analysis systems.

ACKNOWLEDGMENTS

This work was supported by DOE grant DE-FC02-01ER63175 and NIH grants 2 R01 HG000225-10, NIH 5 T32 GM08349, and NIH GM65891. We thank all members of the University of Wisconsin-Madison Laboratory for Molecular and Computational Genomics.

REFERENCES

1. **Anantharaman, T. S., B. Mishra, and D. C. Schwartz.** 1997. Genomics via optical mapping. II. Ordered restriction maps. J. Comput. Biol. **4:**91 - 118.
2. **Anantharaman, T. S., B. Mishra, and D. C. Schwartz.** 1999. Genomics via optical mapping. III. Contiging genomic DNA and variations, p. 18 - 27, The Seventh International Conference on Intelligent Systems for Molecular Biology, vol. 7.
3. **Anantharaman, T. S., B. Mishra, and D. C. Schwartz.** 1998. Genomics via optical mapping. III. Contiging genomic DNA and variations. Courant Technical Report 760. Courant Institute, New York University, New York, NY.
4. **Armbrust, E. V., J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kroger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, and D.**

S. Rokhsar. 2004. The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. Science **306:**79-86.

5. **Bonam, D., L. Lehman, G. P. Roberts, and P. W. Ludden.** 1989. Regulation of carbon monoxide dehydrogenase and hydrogenase in *Rhodospirillum rubrum*: effects of CO and oxygen on synthesis and activity. J. Bacteriol. **171:**3102 - 3107.
6. **Brandl, H., E. J. Knee, Jr., R. C. Fuller, R. A. Gross, and R. W. Lenz.** 1989. Ability of the phototrophic bacterium *Rhodospirillum rubrum* to produce various poly (β-hydroxyalkanoates): potential sources for biodegradable polyesters. Int. J. Biol. Macromol. **11:**49 - 55.
7. **Dimalanta, E. T., A. Lim, R. Runnheim, C. Lamers, C. Churas, D. K. Forrest, J. J. de Pablo, M. D. Graham, S. N. Coppersmith, S. Goldstein, and D. C. Schwartz.** 2004. A microfluidic system for large DNA molecule arrays. Anal. Chem. **76:**5293-5301.
8. **Fitzmaurice, W. P., L. L. Saari, R. G. Lowery, P. W. Ludden, and G. P. Roberts.** 1989. Genes coding for the reversible ADP-ribosylation system of dinitrogenase reductase from *Rhodosphirillum rubrum*. Mol. Gen. Genet. **218:**340 - 347.
9. **Fuller, R. C.** 1995. Polyesters and photosynthetic bacteria: from lipid cellular inclusions to microbial thermoplastics, p. 991 - 1003. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (ed.), Anoxygenic Photosynthetic Bacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.
10. **Giacalone, J., S. Delobette, V. Gibaja, L. Ni, Y. Skiadas, H. Zhao, T. Ananatharaman, B. Mishra, L. G. Brown, R. Saxena, D. C. Page, and D. C. Schwartz.** 2000. Optical mapping of BAC clones from the human Y chromosome *DAZ* locus. Genome Res **10:**1421 - 1429.
11. **Gordon, D., C. Desmarais, and P. Green.** 2001. Automated finishing with Autofinish. Genome Res. **11:**614 - 625.
12. **Handrick, R., S. Reinhardt, D. Schultheiss, T. Reichart, D. Schuler, V. Jendrossek, and D. Jendrossek.** 2004. Unraveling the function of the *Rhodospirillum rubrum* activator of polyhydroxybutyrate (PHB) degradation: the activator is a PHB-granule-protein (Phasin). J. Bacteriol. **186:**2466 - 2575.
13. **Imhoff, J. F.** 2000. The anoxygenic phototrophic purple bacteria, p. 631 - 637. *In* D. Boone and R. W. Castenholz (ed.), Bergey's Manual of Systematic Bacteriology, 2nd ed., vol. 1. Springer-Verlag, New York.
14. **Imhoff, J. F.** 2001, posting date. The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community. Springer-Verlag. [Online.]
15. **Jing, J., R. J, J. Huang, X. Hu, V. Clarke, J. Edington, D. Housman, T. S. Anantharaman, E. J. Huff, B. Mishra, B. Porter, A. Shenker, E. Wolfson, C. Hiort, R. Kantor, C. Aston, and D. C. Schwartz.** 1998. Automated high resolution optical mapping using arrayed, fluid-fixed DNA molecules. Proc. Natl. Acad. Sci. USA **95:**8046 - 8051.
16. **Jing, J., Z. Lai, C. Aston, J. Lin, D. J. Carucci, M. J. Gardner, B. Mishra, T. S. Anantharaman, H. Tettelin, L. M. Cummings, S. L. Hoffman, J. C. Venter, and D. C. Schwartz.** 1999. Optical Mapping of *Plasmodium falciparum* Chromosome 2. Genome Res **9:**175 - 181.

17. **Lai, Z., J. J., C. Aston, V. Clarke, J. Apodaca, E. T. Dimalanta, D. J. Carucci, M. J. Gardner, B. Mishra, T. S. Anantharaman, S. Paxia, S. L. Hoffman, J. C. V. J. Huff, and D. C. Schwartz.** 1999. A shotgun optical map of the entire *Plasmodium falciparum* genome. Nat. Genet. **23:**309 - 313.
18. **Lim, A., E. T. Dimalanta, K. D. Potamousis, G. Yen, J. Apodoca, C.Tao, J. Lin, R. Qi, J. Skiadas, A. Ramanathan, N. T. Perna, G. Plunkett, V. Burland, B. Mau, J. Hackett, F. R. Blattner, T. S. Anantharaman, B.Mishra, and D. C. Schwartz.** 2001. Shotgun optical maps of the whole *Escherichia coli* O157:H7 genome. Genome Res. **11:**1584 - 1593.
19. **Lin, J., R. Qi, C. Aston, J. Jing, T. S. Anantharaman, B. Mishra, O. White, M. J. Daly, K. W. Minton, J. C. Venter, and D.C.Schwartz.** 1999. Whole-genome shotgun optical mapping of *Deinococcus radiodurans*. Science **285:**1558 - 1562.
20. **Marra, M. A., T. A. Kucaba, N. L. Dietrich, E. D. Green, B. Brownstein, R. K. Wilson, K. M. McDonald, L. W. Hillier, J. D. McPherson, and R. H. Waterston.** 1997. High throughput fingerprint analysis of large-insert clones. Genome Res. **7:**1072 - 1084.
21. **Meng, X., K. Benson, K. Chada, J. E. Huff, and D. C. Schwartz.** 1995. Optical mapping of lambda bacteriophage clones using restriction endonucleases. Nat. Genet. **9:**432 - 438.
22. **Meyers, B. C., S. Scalabrin, and M. Morgante.** 2004. Mapping and sequencing complex genomes: let's get physical. Nat. Rev. Genet. **5:**578 - 589.
23. **Pfennig, N., H. Lünsdorf, J. Süling, and J. F. Imhoff.** 1997. *Rhodospira trueperi* gen. nov., spec. nov., a new phototrophic Proteobacterium of the alpha group. Arch. Microbiol. **168:**39 - 45.
24. **Sasaki, S., and I. Karube.** 1999. The development of microfabricated biocatalytic fuel cells. TIBTECH **17:**50 - 52.
25. **Schick, H. J.** 1971. Interrelationship of nitrogen fixation, hydrogen evolution and photoreduction in *Rhodospirillum rubrum*. Arch. Microbiol. **75:**102 - 109.
26. **Schwartz, D. C., and C. R. Cantor.** 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell **37:**67 - 75.
27. **Tabita, F. R.** 1995. The biochemistry and metabolic regulation of carbon metabolism and CO_2 fixation in purple bacteria, p. 885 - 914. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (ed.), Anoxygenic Photosynthetic Bacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.
28. **von Felten, P., H. Zürrer, and R. Bachofen.** 1985. Production of molecular hydrogen with immobilized cells of *Rhodospirillum rubrum*. Appl. Microbiol. Biotechnol. **23:**15 - 20.
29. **Zhang, Y., R. H. Burris, P. W. Ludden, and G. P. Roberts.** 1995. Comparison studies of dinitrogenase reductase ADP-ribosyl transferase/dinitrogenase reductase activating glycohydrolase regulatory systems in *Rhodospirillum rubrum* and *Azospirillum brasilense*. J. Bacteriol. **177:**2354 - 2359.
30. **Zhou, S., W. Deng, T. S. Anantharaman, A. Lim, E. T. Dimalanta, J. Wang, T. Wu, T. Chunhong, R. Creighton, A. Kile, E. Kvikstad, M. Bechner, G. Yen, A. Garic-Stankovic, J. Severin, D. Forrest, R. Runnheim, C. Churas, C. Lamers, N. T. Perna, V. Burland, F. R. Blattner, B. Mishra, and D. C.**

Schwartz. 2002. A whole-genome shotgun optical map of *Yersinia pestis* strain KIM. Appl. Environ. Microbiol. **68:**6321 - 6331.

31. **Zhou, S., A. Kile, M. Bechner, M. Place, E. Kvikstad, W. Deng, J. Wei, J. Severin, R. Runnheim, C. Churas, D. Forrest, E. T. Dimalanta, C. Lamers, V. Burland, F. R. Blattner, and D.C.Schwartz.** 2004. A single molecule approach to bacterial genomic comparisons via optical mapping. J. Bacteriol. **in press**.
32. **Zhou, S., E. Kvikstad, A. Kile, J. Severin, D. Forrest, R. Runnheim, C. Churas, J. W. Hickman, C. Mackenzie, M. Choudhary, T. Donahue, S. Kaplan, and D. C. Schwartz.** 2003. Whole-genome shotgun optical mapping of *Rhodabacter sphaeroides* strain 2.4.1 and its use for whole-genome shotgun sequence assembly. Genome Res. **13:**2142 - 2151.

FIGURE LEGENDS

Figure 1. The whole-genome circular *Xba* I map of *R. rubrum*. The outermost circle represents the consensus map created by Gentig from the single molecule maps shown here as arcs. The single molecule maps were made from single DNA molecules digested with *Xba* I. Fragments that created each consensus map fragment are indicated by a common color; the color ordering is for contrast and therefore is random.

Figure 2. Use of optical maps in confirming assembly and order of sequence contigs. *In silico* maps of the sequence contigs were made and aligned against the whole genome optical maps located in the center of each diagram. **A**) The high resolution *Hind* III map enabled alignment of the small (~80 kb) sequence contig (contig 83) between contigs 85c and 86c. Aside from the redundant alignment of contigs 87 and 90, no gaps remained in alignment between the optical map and sequence. **B**) The *Nhe* I map gave the clearest representation of the redundant alignment, shown in red in the optical map, of contigs 87 and 90. The alignment of contig 87 had a better match to the optical map than do the right most eight fragments in contig 90 (shown in the red square). **C**) Removing the rightmost 8 fragments from contig 90 to make contig 90a and 90b, and reversing the orientation of 90b permitted alignment to the gap between contig 97 and contig 85c. Alignments were made with MapViewer Software (OpGen, Inc.)

Figure 3. Comparisons of the *Xba* I, *Nhe* I, and *Hind* III optical maps to sequence data. **(A-C)** Plots of optical map fragment sizes versus the *in silico* map fragment sizes from

the finished sequence for *Xba* I **(A)**, *Nhe* I **(B)**, and *Hind* III **(C)**. The error bars represent the standard deviation of optical map fragment size about the mean. **(D-F)** Plots of the absolute relative fragment sizing error versus *in silico* map fragment size for *Xba* I **(D)**, *Nhe* I **(E)**, and *Hind* III **(F)**.

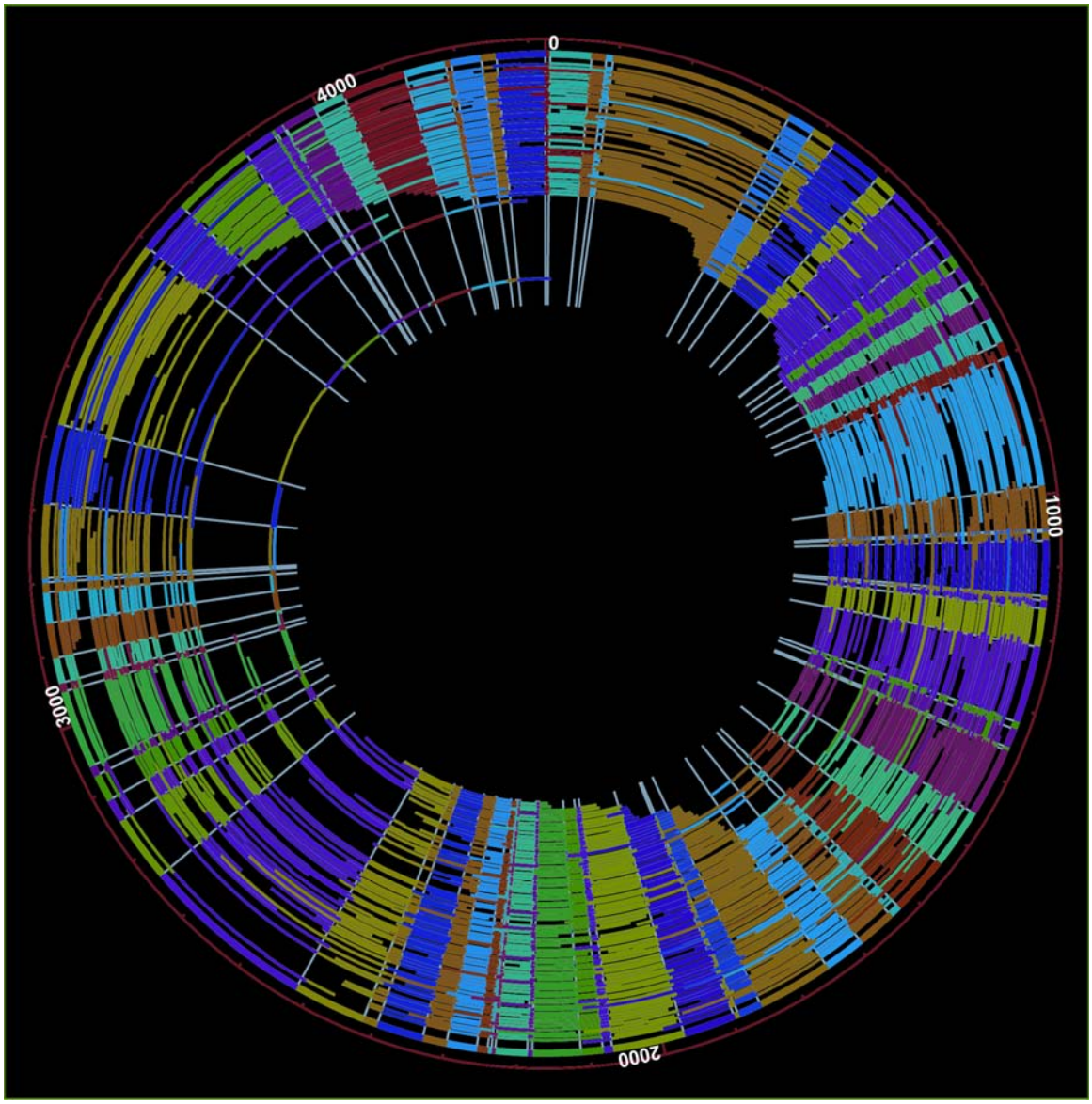
Figure 4. Cumulative distribution of optical map fragment sizes for the three optical maps. For each of the three whole-genome *R. rubrum* optical maps, the percentage of fragments within the size range from zero to 135 kb is plotted. Each bar represents the cumulative percentage of consensus map fragments in 5-kb intervals. From the figure, the drastic differences in map fragment sizes, particular for fragments less than 10 kb are evident.

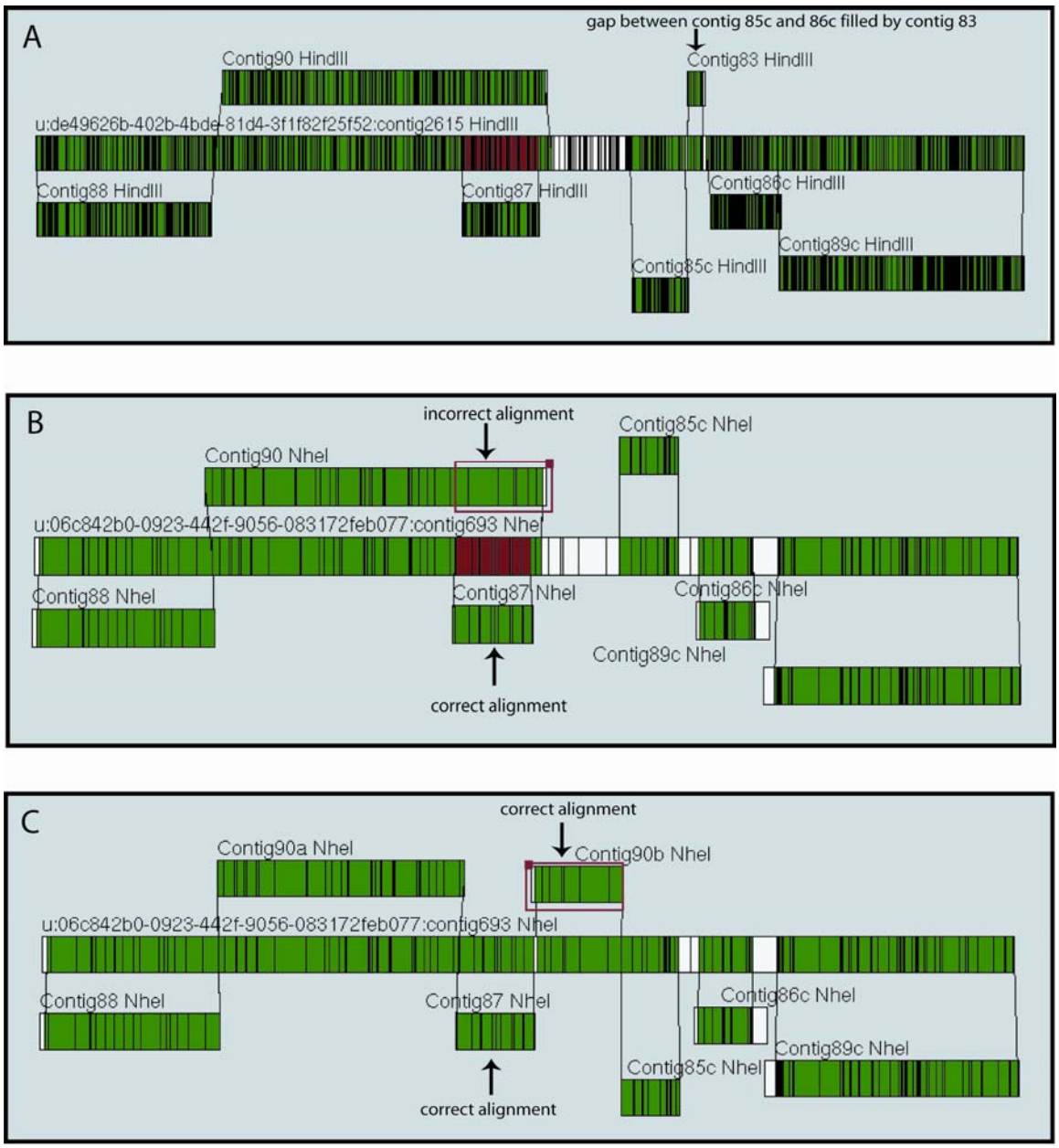
Figure 5. Alignments of consensus optical maps with the *in silico* maps. **A**) The lowest resolution *Xba* I map alignment **B**) the medium resolution *Nhe* I map alignment and **C**) the highest resolution *Hind* III map alignment.

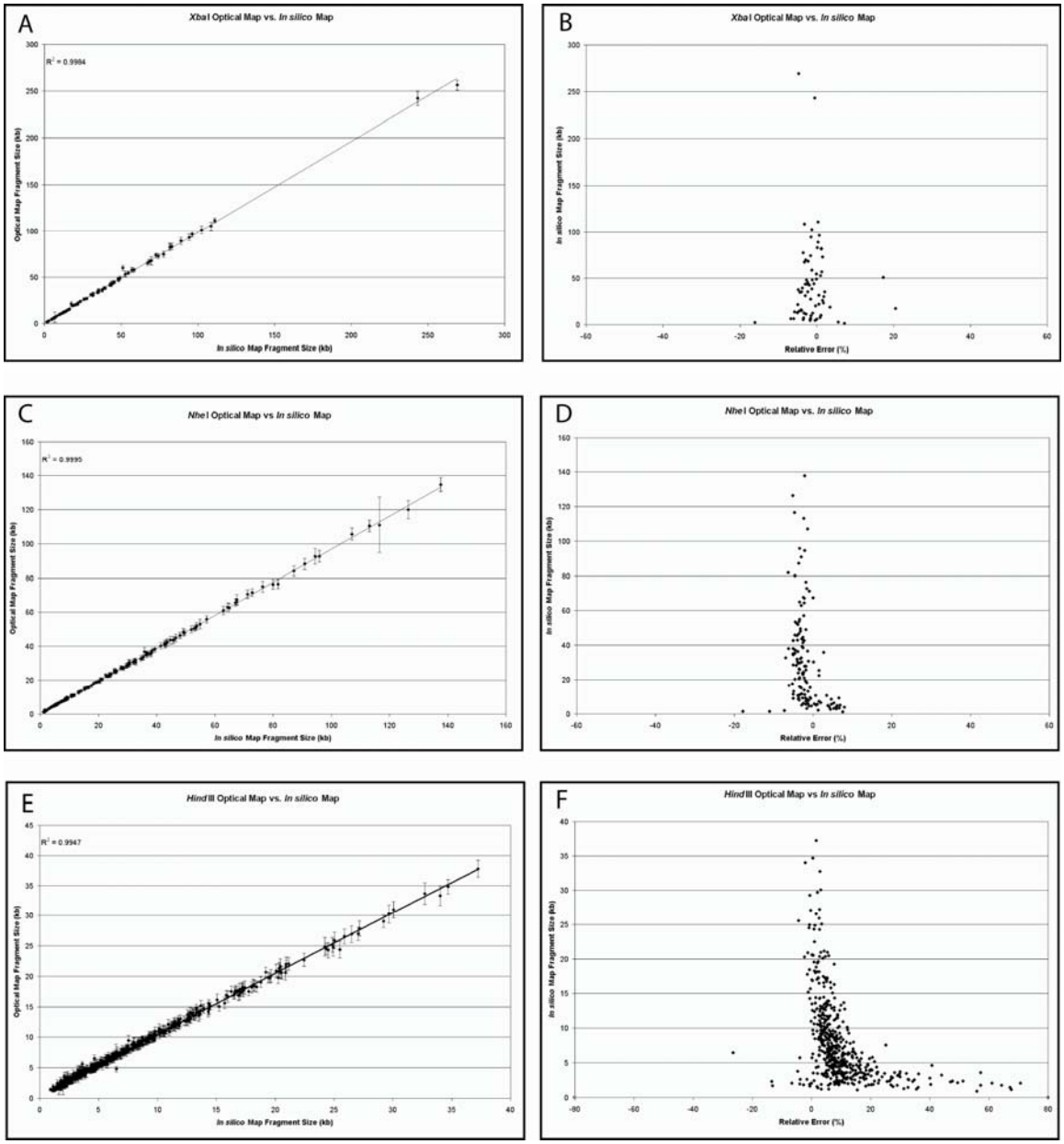
Figure 6. A linear view of the map alignments in Figure 6 for viewing errors in optical maps as compared to sequence. Solid black arrows represent the locations of missing cuts in the consensus optical maps. The *Xba* I and *Nhe* I consensus optical maps extend to the left of the origin of the *in silico* map because both optical maps had a missing cut at that position.

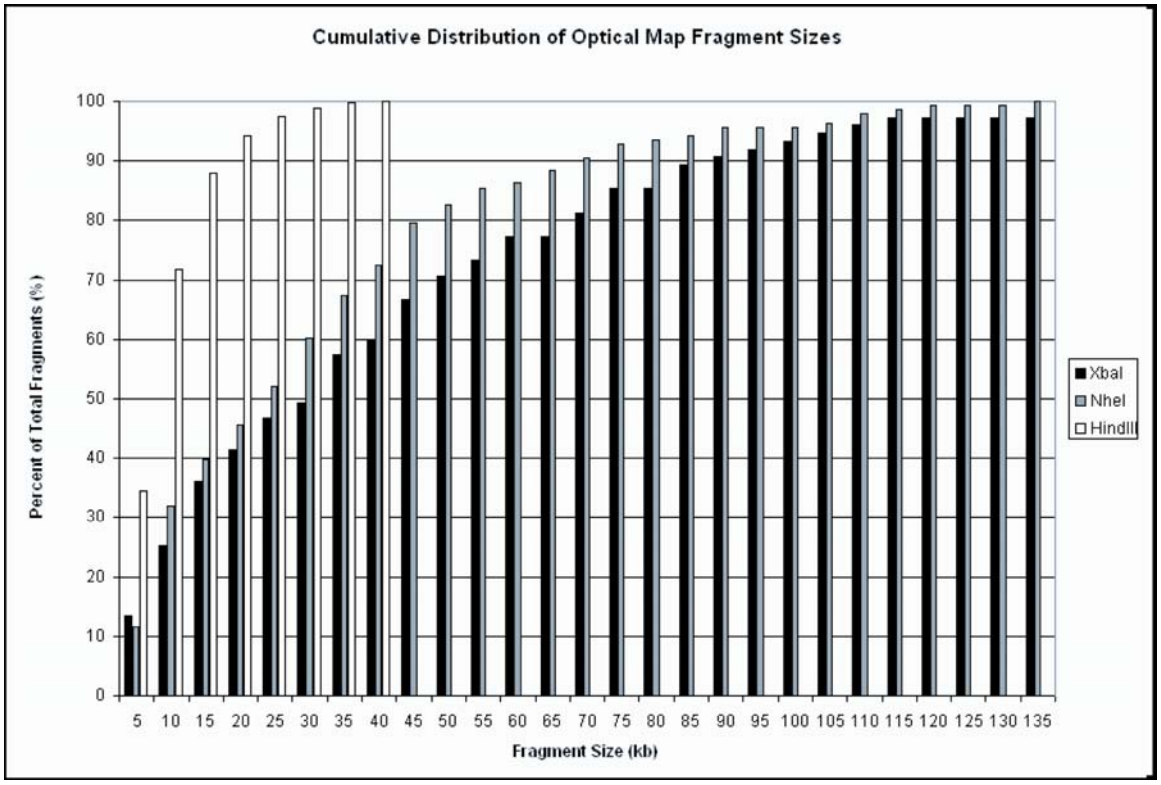
Table 1. *R. rubrum* Optical Mapping Data

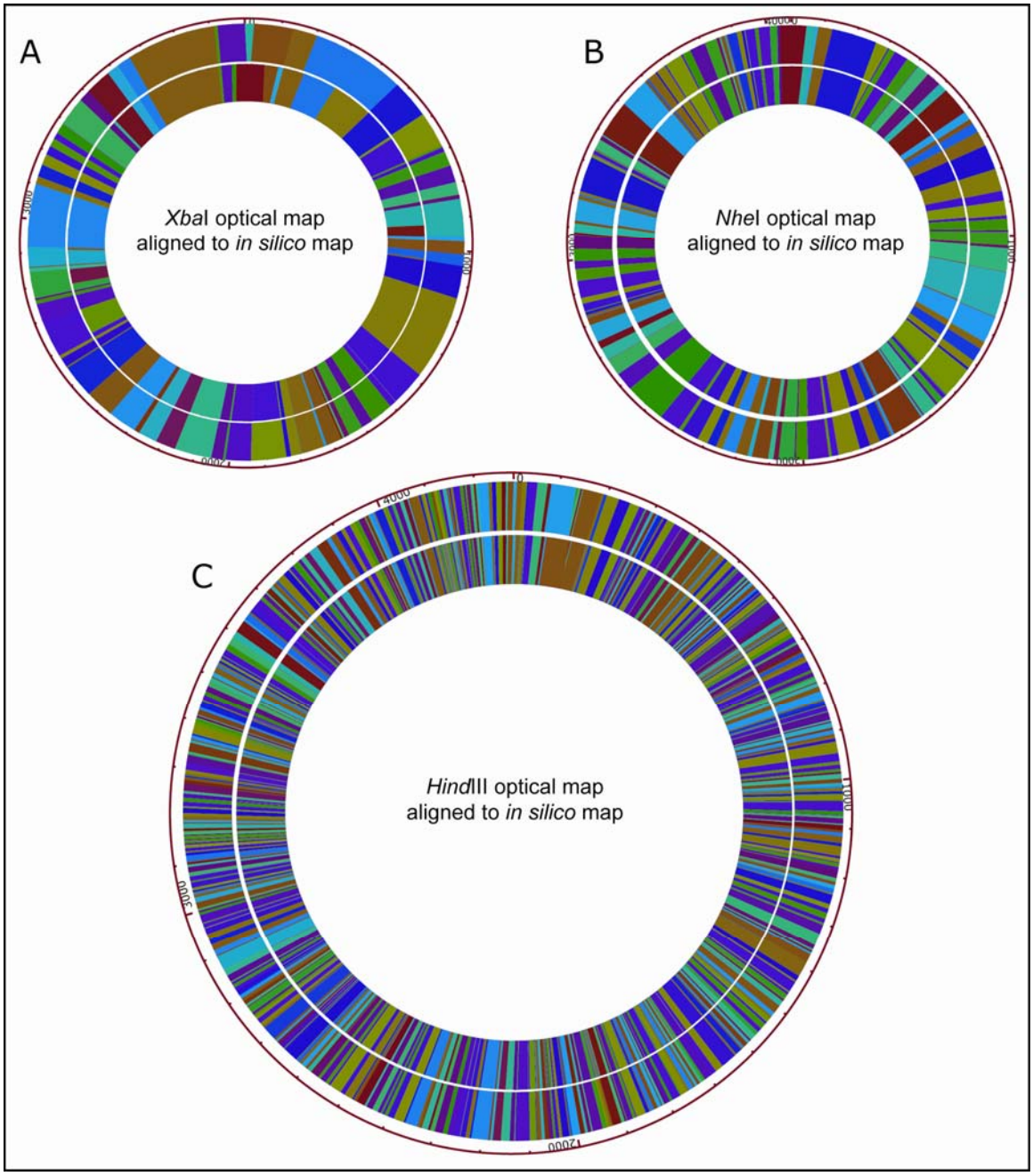
	Enzyme		
Data Collection	***Xba* I**	***Nhe* I**	***Hind* III**
Number of Molecules	405	409	932
Average Molecule Size (kb)	637.69	635.72	411.76
Total Mass (Mb)	258.265	260.009	383.758
Molecules in the Contigs			
Number of Molecules	204	345	623
Average Molecule Size (kb)	903.07	731.30	405.49
Total Mass (Mb)	184.23	252.30	252.68
Contig Rate (%)	50.37	84.35	66.84
Circular Contig Size (kb)	4323.08	4223.13	4456.35
Average Fragment Size (kb)	44.73	31.53	10.95
Average Fragment Size Standard Deviation (kb)	4.31	2.83	1.16
Circularization False Probability	0.00738	0.00329	0.00440

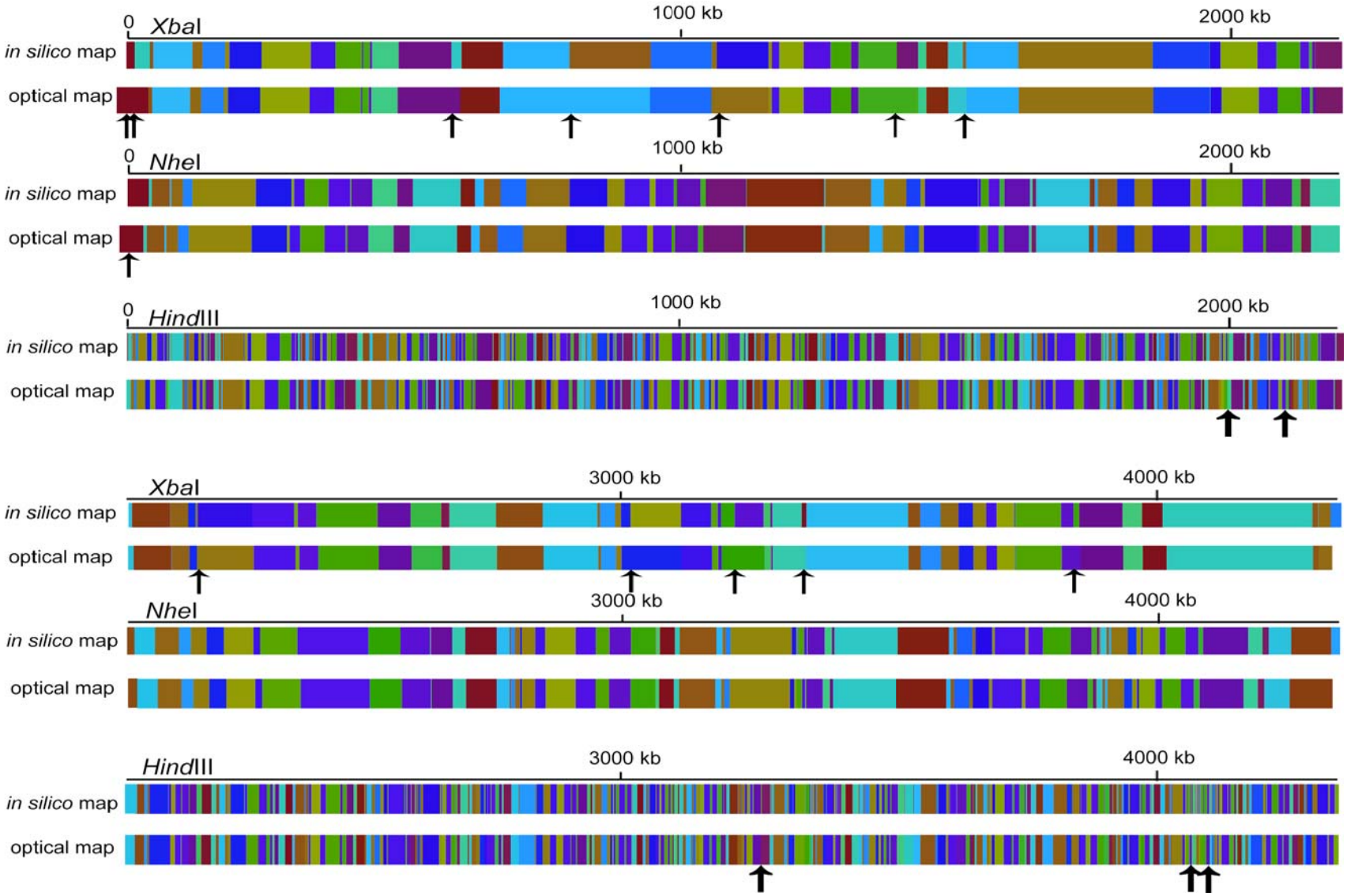

0
1000
2000
3000
4000


A
gap between contig 85c and 86c filled by contig 83
Contig90 HindIII
Contig83 HindIII
u:de49626b-402b-4bde-81d4-3f1f82f25f52:contig2615 HindIII
Contig88 HindIII
Contig87 HindIII
Contig86c HindIII
Contig89c HindIII
Contig85c HindIII
B
incorrect alignment
Contig85c NheI
Contig90 NheI
u:06c842b0-0923-442f-9056-083172feb077:contig693 NheI
Contig88 NheI
Contig87 NheI
Contig86c NheI
Contig89c NheI
correct alignment
C
correct alignment
Contig90a NheI
Contig90b NheI
u:06c842b0-0923-442f-9056-083172feb077:contig693 NheI
Contig88 NheI
Contig87 NheI
Contig86c NheI
Contig89c NheI
Contig85c NheI
correct alignment


A
XbaI Optical Map vs. In silico Map
R² = 0.9984
Optical Map Fragment Size (kb)
In silico Map Fragment Size (kb)
B
XbaI Optical Map vs. In silico Map
In silico Map Fragment Size (kb)
Relative Error (%)
C
NheI Optical Map vs In silico Map
R² = 0.9995
Optical Map Fragment Size (kb)
In silico Map Fragment Size (kb)
D
NheI Optical Map vs. In silico Map
In silico Map Fragment Size (kb)
Relative Error (%)
E
HindIII Optical Map vs. In silico Map
R² = 0.9947
Optical Map Fragment Size (kb)
In silico Map Fragment Size (kb)
F
HindIII Optical Map vs In silico Map
In silico Map Fragment Size (kb)
Relative Error (%)


Cumulative Distribution of Optical Map Fragment Sizes
Percent of Total Fragments (%)
0
10
20
30
40
50
60
70
80
90
100
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135
Fragment Size (kb)
XbaI
NheI
HindIII


A
XbaI optical map
aligned to in silico map
B
NheI optical map
aligned to in silico map
C
HindIII optical map
aligned to in silico map


0
XbaI
1000 kb
2000 kb
in silico map
optical map
0
NheI
1000 kb
2000 kb
in silico map
optical map
0
HindIII
1000 kb
2000 kb
in silico map
optical map
XbaI
3000 kb
4000 kb
in silico map
optical map
NheI
3000 kb
4000 kb
in silico map
optical map
HindIII
3000 kb
4000 kb
in silico map
optical map