



HAL
open science

Optical mapping as a routine tool for bacterial genome sequence finishing

Phil Latreille, Stacie Norton, Barry S. Goldman, John Henkhaus, Nancy Miller, Brad Barbazuk, Helge B. Bode, Creg Darby, Zijin Du, Steve Forst, et al.

► **To cite this version:**

Phil Latreille, Stacie Norton, Barry S. Goldman, John Henkhaus, Nancy Miller, et al.. Optical mapping as a routine tool for bacterial genome sequence finishing. *BMC Genomics*, 2007, 8, pp.321. 10.1186/1471-2164-8-321 . hal-02666315

HAL Id: hal-02666315

<https://hal.inrae.fr/hal-02666315>

Submitted on 31 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Methodology article

Open Access

Optical mapping as a routine tool for bacterial genome sequence finishing

Phil Latreille¹, Stacie Norton¹, Barry S Goldman¹, John Henkhaus², Nancy Miller¹, Brad Barbazuk³, Helge B Bode⁴, Creg Darby⁵, Zijin Du¹, Steve Forst⁶, Sophie Gaudriault⁷, Brad Goodner⁸, Heidi Goodrich-Blair⁹ and Steven Slater*¹⁰

Address: ¹Monsanto Company, 800 North Lindbergh Boulevard St. Louis, MO 63167, USA, ²OpGen Technologies, Inc., 510 Charmany Drive, Suite 151, Madison, WI 53719, USA, ³Donald Danforth Plant Sciences Center, 975 North Warson Road St. Louis, MO 63132, USA, ⁴Institut für Pharmazeutische Biotechnologie, Universität des Saarlandes, 66123 Saarbrücken, Germany, ⁵University of California, San Francisco, Department of Cell and Tissue Biology, San Francisco, CA 94143, USA, ⁶University of Wisconsin, Milwaukee, Department of Biological Sciences, Milwaukee, WI 53211, USA, ⁷Institut National de la Recherche Agronomique-Université de Montpellier II, 34095 Montpellier, France, ⁸Hiram College, Department of Biology, Hiram, OH 44234, USA, ⁹University of Wisconsin, Department of Bacteriology, Madison, WI 53076, USA and ¹⁰Arizona State University, The Biodesign Institute and Department of Applied Biological Sciences, 7001 E. Williams Field Road, Mesa, AZ 85212, USA

Email: Phil Latreille - jplatre@monsanto.com; Stacie Norton - stgatt@monsanto.com; Barry S Goldman - barry.s.goldman@monsanto.com; John Henkhaus - jhenkhaus@opgen.com; Nancy Miller - nmmill@monsanto.com; Brad Barbazuk - bbarbazuk@danforthcenter.org; Helge B Bode - h.bode@mx.uni-saarland.de; Creg Darby - creg.darby@ucsf.edu; Zijin Du - zzdu@monsanto.com; Steve Forst - sforst@uwm.edu; Sophie Gaudriault - sgaudriault@univ-montp2.fr; Brad Goodner - goodnerbw@hiram.edu; Heidi Goodrich-Blair - hgblair@bact.wisc.edu; Steven Slater* - scslater@asu.edu

* Corresponding author

Published: 14 September 2007

Received: 19 December 2006

BMC Genomics 2007, 8:321 doi:10.1186/1471-2164-8-321

Accepted: 14 September 2007

This article is available from: <http://www.biomedcentral.com/1471-2164/8/321>

© 2007 Latreille et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: In sequencing the genomes of two *Xenorhabdus* species, we encountered a large number of sequence repeats and assembly anomalies that stalled finishing efforts. This included a stretch of about 12 Kb that is over 99.9% identical between the plasmid and chromosome of *X. nematophila*.

Results: Whole genome restriction maps of the sequenced strains were produced through optical mapping technology. These maps allowed rapid resolution of sequence assembly problems, permitted closing of the genome, and allowed correction of a large inversion in a genome assembly that we had considered finished.

Conclusion: Our experience suggests that routine use of optical mapping in bacterial genome sequence finishing is warranted. When combined with data produced through 454 sequencing, an optical map can rapidly and inexpensively generate an ordered and oriented set of contigs to produce a nearly complete genome sequence assembly.

Background

Xenorhabdus species are symbiotic bacteria associated with insectivorous nematodes of the genus *Steinernema* (for

review see [1]) They reside in a specialized segment of the nematode gut [2,3], and provide insecticidal proteins [4,5] and small molecules [6-10] that help to kill the

insect larvae that are the prey of the nematode. Both organisms reproduce in the dead larvae, the *Xenorhabdus* colonize the young nematodes, and the cycle repeats [11]. *Xenorhabdus* are closely related to the enteric gamma proteobacteria such as *Escherichia coli* [12], and are an emerging model for both mutualism and pathogenicity in invertebrate hosts. To better understand the genetic basis of these relationships, we are sequencing the genomes of two *Xenorhabdus* species: *X. nematophila* ATCC 19061 and an *X. bovienii* strain from Monsanto's collection.

In the course of this work, we found that the *X. nematophila* genome contained large numbers of highly repetitive DNA regions, and efforts to finish the genome stalled. We sought a means to produce whole-genome maps for comparison with the genomic DNA sequence, and identified optical mapping as a useful means to align and orient the genome sections *in silico*. In addition, we produced an optical map of a second genome that we had considered finished, and identified a large sequence inversion that would have otherwise been unnoticed.

Results

A whole-genome restriction map permits finishing of the *X. nematophila* genome sequence

Eight-fold genome sequence coverage of *X. nematophila* ATCC19061 (Goodrich-Blair et al, in preparation) was generated, with 26,976 reads from a 2–4 kb insert library and 41,376 reads from a 4–8 kb insert library. This yielded an initial assembly consisting of 100 contiguous sequences (contigs) greater than 2 kb, 14 contigs greater than 100 kb, and 2 contigs greater than 200 kb length. Our initial research had shown the presence of a 150 Kb plasmid in addition to the circular chromosome (Goodrich-Blair and Goodner, unpublished).

It became rapidly clear that multiple areas of repeated sequence were causing problems. In fact, the final *X. nematophila* sequence assembly shows a nearly identical 12 Kb region found on both the plasmid and chromosome, many transposons (including over 30 copies of a single transposon) scattered throughout the genome, and seven rRNA regions. Using the paired clone-end sequences and syntenic comparison to the related species *Photorhabdus luminescens* [13], resolution of misassemblies and gap closure was attempted by walking across individual clones and amplifying potentially adjacent regions using the polymerase chain reaction (PCR). The resulting assembly contained over 50 contigs, but most lacked linkage information from gap-spanning paired ends. Multiplex PCR resolved some gaps, but provided no indication about whether the amplified product was actually the correct size, whether a particular gap was resistant to amplification, or whether a reaction failed because the primers were not properly paired to cross a gap. After four months

of concerted effort, the assembly still contained 36 contigs which collectively contained several hundred copies of transposons plus seven ribosomal RNA coding regions. Given this complexity, optical mapping was attempted to provide a structural scaffold for aligning and orienting the contigs.

Optical mapping permits assembly of whole-genome restriction endonuclease maps by digesting immobilized DNA molecules and determining the size and order of fragments [14–22]. In collaboration with OpGen Technologies (Madison, WI), optical maps of *X. nematophila* ATCC19061 were produced using *AflIII* and *EagI* restriction enzymes. Through repeated overlapping of restriction maps from individual molecules (over 50-fold coverage), OpGen's assembler program reconstructed the ordered restriction map of the genome [23].

Each restriction map produced by optical mapping was aligned with the restriction map predicted from the *X. nematophila* genome sequence. The map permitted alignment and orientation of all 36 contigs, and identification of misassemblies, allowing production of PCR products to cover all remaining gaps in the sequence (Figure 1 panel A). Once the optical map was available, PCR, sequencing, and validation of the final assembly were accomplished in approximately one month. The map also detected several regions of misassembled sequence, including a plasmid that was integrated into the chromosomal sequence among the assembled contigs (Figure 1 panel B). The plasmid shares a highly conserved stretch of sequence with the chromosome (only 37 bp differences over approximately 12.5 kb), and this duplication led to the *in silico* misassembly. The final sequenced genome aligned directly to the restriction map generated by optical mapping (Figure 1 panel C).

Optical mapping identifies an assembly error in the *X. bovienii* sequence

In addition to *X. nematophila*, we had previously sequenced and assembled the genome of the related organism *X. bovienii* using traditional finishing technologies. Although the *X. bovienii* genome does not contain as many repeats as that of *X. nematophila*, the *X. nematophila* project had shown the value of non-sequence-based methodologies in validating sequence assemblies. After generating an optical map for *X. bovienii* (NCBI designation *Xenorhabdus bovienii* SS-2004) using *AflIII*, a large inversion was detected in the sequence assembly, permitting a simple re-orientation of the data and correction of the genome sequence (Figure 1 panel D). It is doubtful that this assembly inversion would have been detected without the optical map.

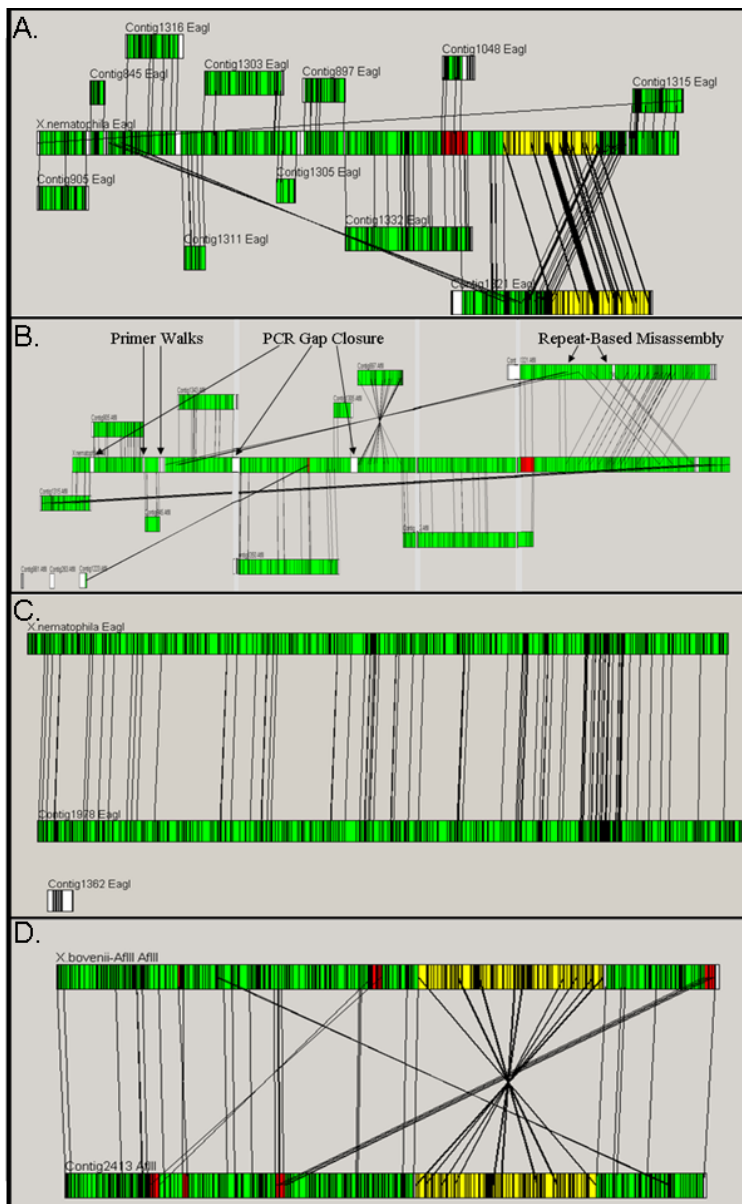


Figure 1
Alignments between the whole-genome optical maps and the *in silico* genome sequence assemblies at various stages of the project. Green regions indicate perfect alignment, white regions indicate no alignment, red regions indicate sequence that is present on at least two contigs, and yellow regions indicate inversions. Lines between maps indicate the position of identical sequences on the two maps, and can be used to visually identify misassemblies and inversions. **Panel A:** An early comparison of an optical map derived from *EagI* digestion of the *X. nematophila* genome to the assembled contigs generated by traditional sequencing technologies. All contigs could be ordered for gap closure. In addition, the optical map indicated an overlooked misassembly. **Panel B:** The finishing strategy, including gap closure and misassembly resolution, was simplified using the optical map as an assembly model. The *X. nematophila* optical map derived from an *AfIII* digestion of the chromosome is presented as a single contig in the center. The sequenced genome contains nine contigs that have a corresponding match to the optical map. The *X. nematophila* plasmid is 158 Kb and is too small to be identified using the current optical map technology. Nonetheless, small sections of the plasmid can be identified as regions that do not have corresponding optical map locations (white in figure). **Panel C:** Comparison of the final assembly of the *X. nematophila* genome (bottom) to the optical map (top) for the *EagI* digest. The non-aligned contig represents the plasmid, which was generated by traditional sequencing technologies. **Panel D:** Comparison of the finished sequence of *Xenorhabdus bovienii* to the *EagI* optical map revealed a large inverted region of the genome. The red regions indicate regions of repeats within the genome that cannot be resolved by optical mapping. These regions were resolved using traditional sequencing methods. The sequenced genome was easily re-oriented to correct the assembly.

Discussion

The *Xenorhabdus* genomes analyzed in this project contain many highly repetitive regions, and these became a major obstacle in our attempts to assemble the genome sequences. Genome finishing traditionally relies on cosmid libraries or overlapping restriction maps of BACs to build larger meta-contigs. With the *X. nematophila* genome the traditional approach failed, and we used a genome-scale restriction map generated by optical mapping. This permitted rapid and accurate closing of *X. nematophila*, and provided savings of labor, reagents and time. Finishing the *X. nematophila* genome sequence would have otherwise required production of a fine-scale genetic-physical map at much greater cost in time and materials. Optical mapping also identified an inversion in the *X. bovienii* genome sequence assembly that we had considered finished.

High throughput processes like DNA sequencing normally require trade-offs among cost, speed, and data quality. Sequencing costs are being reduced, and speed increased, by novel methods such as the pyrosequencing technology of 454 Life Sciences [24,25]. However, 454 technology produces shorter sequences (100 to 250 bases per reaction) than traditional Sanger sequencing using ABI instrumentation (800–1000 bases per reaction). These shorter 454-derived sequences mean that sequence contigs are also, on average, shorter than those produced using ABI instruments. However, the lower quality of sequence assemblies from 454 data is compensated by speed and cost considerations. Excluding the cost of purchasing the instrumentation and labor, a typical 5 Mb bacterial genome takes approximately 2 days and costs about \$6,000 in consumables using 454. The same genome sequence produced by ABI instrumentation would cost approximately 10-fold more and take several weeks. In our experience, a typical 5 Mb assembly using 454 data would contain about 80–90 contigs, with an average length around 60–70 Kb. A similar genome assembled using data from ABI 3730 instruments would contain about 50 contigs with an average length >100 kb. Both strategies would typically add about 4,000 end-paired sequences from cosmids or phosmids to help scaffold the genome, at a cost of about another \$4,000.

The current cost for an optical map with a single enzyme is approximately \$7,000, and adding a second enzyme costs around another \$3,000 (in our experience, only one enzyme is typically required). The optical mapping system can accurately quantify fragments down to about 4 kb in size, and a contig of 40 kb has an approximately 80% probability of being placed within a whole genome optical map (OpGen, unpublished data). When all of these data are combined, a 454 shotgun sequence plus cosmid end sequences and an optical map, can produce an assem-

bled and oriented set of contigs containing about 95% of the genome for under \$20,000 with very limited input by a human finisher. This is about one-fifth the cost of a project produced through traditional means, provides very high quality data, and puts production of finished bacterial genomes within the reach of even small labs. We are currently working on a genome produced in this manner that will be primarily closed using undergraduate researchers supported by some bioinformatics infrastructure.

Conclusion

Even on these relatively small genomes, the whole-genome maps were very valuable. In the *X. nematophila* project, we had the advantages of long sequence reads and clone end-pairing data, yet still were unable to assemble contigs because of the presence of numerous highly repetitive sequences. The optical map allowed rapid closure of one genome and identified an assembly error in a fully-assembled genome sequence that gave no prior indication of having errors.

As shotgun sequencing costs come down, the optical map becomes a significant portion of the budget for a new bacterial genome sequence. However, for genomes that contain particularly large numbers of repetitive sequences, require finishing, or simply require ordered and oriented contigs from shotgun sequence, an optical map can increase the speed and decrease the overall cost of the project. We also expect that mapping costs will come down as optical mapping becomes more routinely used by sequencing centers, and as resolution of fragment size moves toward the 1–2 kb range. We now routinely confirm the *in silico* assemblies of bacterial genomes using a whole-genome restriction map, and believe this is a relatively low cost method to speed finishing and ensure accuracy of finished bacterial genome sequences.

Methods

Genomic library construction, DNA sequencing, and finishing

The genomic DNA was sonicated at scale of 8.5 for two seconds, repeated 3 times (Misonnex Inc. Sonicator XL2020). The ends were repaired using T4 DNA polymerase and T4 kinase (NEB) and fractionated on a 1% agarose gel. Fractions representing size ranges 2–4 KB and 4–8 KB were excised from the gel and purified using a Qiagen Gel Quick extraction column (Qiagen, Cat No 28704). DNA samples from the isolated fractions were checked for size on an agarose gel and then ligated into pUC18.

Clones were plated and colonies picked on a Q-Bot (Genetix), to achieve 80% of sequence from the 2–4 KB library and 20% of sequence from the 4–8 KB library. Each template was sequenced using the Big Dye termina-

tor protocol (Applied Biosystems) and analyzed on ABI 3700 and ABI 3730 sequencers. Both the forward (M13 -40) and reverse (M13 -21) primer were used on each template, yielding two related sequences per subclone. Data were assembled using phred/phrap (ver. 0.990319; [26,27]), and finished in Consed and Autofinish (v.13.0; [28-30]) using a variety of directed primer walks on subclones, and using PCR/walking to close any gaps. The sequence assemblies were confirmed by OpGen using optical mapping, as described below and previously [14-22]. These alignments were viewed using OpGen's MapViewer software (Figure 1; see below).

Optical map construction

Optical maps were prepared at OpGen Technologies, Inc. (Madison, WI) according to methods described previously [22,23]. Briefly, high molecular weight DNA was prepared by first embedding bacterial cells harvested at stationary phase in low melting temperature agarose plugs, followed by treatment with bacterial lysing solutions. The genomic DNA was recovered after thoroughly rinsing the plugs in TE followed by melting the plugs at 42 C and subsequent treatment with β -agarase. The high molecular weight DNA was then immobilized as individual molecules onto Optical Chips, digested with EagI or AflIII restriction enzymes (New England Biolabs), fluorescently stained with YOYO-1 (Invitrogen) and positioned onto an automated fluorescent microscope system for image capture and fragment size measurement, resulting in high resolution single-molecule restriction maps. Collections of single molecule maps were then assembled to produce whole genome, ordered restriction maps.

Sequence-to-map comparison

Comparisons between Optical maps and sequence contigs were performed as described previously [22]. Sequence FASTA files were converted to *in silico* restriction maps via the MapViewer software (OpGen Technologies, Inc.) for direct comparison to the Optical maps. Comparisons were accomplished by aligning the sequence with the Optical maps according to their restriction fragment pattern. Alignments were generated with a dynamic programming algorithm which finds the optimal location, or placement, of a sequence contig by first performing a global alignment of the sequence contig against the Optical map. Local alignment analysis were also performed where segments of the sequence contigs were compared to the Optical map.

Competing interests

JH is employed by OpGen Technologies, Inc., the commercial provider of optical mapping technology.

Authors' contributions

PL performed quality control analysis on the sequence and prepared an early draft of the manuscript. JH produced the optical maps. SN, ZD and NM performed the genome finishing work. SS assisted with data analysis and was the primary writer of the manuscript. BB ran automated annotation on the genomes and provided an HTML interface for analysis. HGB, SF and BG performed genetic and molecular analysis of the strains to confirm their identity prior to sequencing and optical mapping. HGB, SF, BG, HB, CD and SG assisted with data analysis. BSG conceived and coordinated the project and helped to write the manuscript.

Acknowledgements

This work was funded by USDA Grant 2004-35600-14181, by NSF Grant 0603491, and by the Monsanto Company.

References

- Forst S, Dowds B, Boemare N, Stackebrandt E: **Xenorhabdus and Photorhabdus spp.: bugs that kill bugs.** *Annu Rev Microbiol* 1997, **51**:47-72.
- Bird AF, Akhurst RJ: **The nature of the intestinal vesicle in nematodes of the family Steinernematidae.** *Int J Parasitol* 1983, **13**:599-606.
- Poinar GO: **The presence of Achromobacter nematophilus in the infective stage of a Neoaeplectana sp. (Steinernematidae: Nematoda).** *Nematologica* 1966, **12**:105-108.
- Morgan JA, Sergeant M, Ellis D, Ousley M, Jarrett P: **Sequence analysis of insecticidal genes from Xenorhabdus nematophilus PMFI296.** *Appl Environ Microbiol* 2001, **67**(5):2062-2069.
- Sergeant M, Jarrett P, Ousley M, Morgan JA: **Interactions of insecticidal toxin gene products from Xenorhabdus nematophilus PMFI296.** *Appl Environ Microbiol* 2003, **69**(6):3344-3349.
- Hwang SY, Paik S, Park SH, Kim HS, Lee IS, Kim SP, Baek WK, Suh MH, Kwon TK, Park JW, Park JB, Lee JJ, Suh SI: **N-phenethyl-2-phenylacetamide isolated from Xenorhabdus nematophilus induces apoptosis through caspase activation and calpain-mediated Bax cleavage in U937 cells.** *Int J Oncol* 2003, **22**(1):151-157.
- Kim Y, Ji D, Cho S, Park Y: **Two groups of entomopathogenic bacteria, Photorhabdus and Xenorhabdus, share an inhibitory action against phospholipase A2 to induce host immunodepression.** *J Invertebr Pathol* 2005, **89**(3):258-264.
- Paik S, Park YH, Suh SI, Kim HS, Lee IS, Park MK, Lee CS, Park SH: **Unusual cytotoxic phenethylamines from Xenorhabdus nematophila.** *Bull Korean Chem Soc* 2001, **22**:372-374.
- Park Y, Kim Y: **Xenorhabdus nematophilus inhibits p-bromophenacyl bromide (BPB)-sensitive PLA2 of Spodoptera exigua.** *Arch Insect Biochem Physiol* 2003, **54**(3):134-142.
- Park Y, Kim Y, Putnam SM, Stanley DW: **The bacterium Xenorhabdus nematophilus depresses nodulation reactions to infection by inhibiting eicosanoid biosynthesis in tobacco hornworms, Manduca sexta.** *Arch Insect Biochem Physiol* 2003, **52**(2):71-80.
- Forst S, Clarke D: **Bacteria-nematode symbioses.** In *Entomopathogenic Nematology* Edited by: Gaugler R. Wallingford: CABI Publishing; 2002:57-77.
- Boemare N: **Biology, taxonomy and systematics of Photorhabdus and Xenorhabdus.** In *Entomopathogenic Nematology* Edited by: Gaugler R. Wallingford: CABI Publishing; 2002.
- Duchaud E, Rusniok C, Frangeul L, Buchrieser C, Givaudan A, Taourit S, Bocs S, Boursaux-Eude C, Chandler M, Charles JF, Dassa E, Derosé R, Derezelle S, Freyssinet G, Gaudriault S, Medigue C, Lanois A, Powell K, Siguier P, Vincent R, Wingate V, Zouine M, Glaser P, Boemare N, Danchin A, Kunst F: **The genome sequence of the entomopathogenic bacterium Photorhabdus luminescens.** *Nat Biotechnol* 2003, **21**(11):1307-1313.
- Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang YK: **Ordered restriction maps of Saccharomyces cerevisiae chro-**

- mosomes constructed by optical mapping. *Science* 1993, **262**(5130):110-114.
15. Cai W, Aburatani H, Stanton VP Jr, Housman DE, Wang YK, Schwartz DC: **Ordered restriction endonuclease maps of yeast artificial chromosomes created by optical mapping on surfaces.** *Proc Natl Acad Sci USA* 1995, **92**(11):5164-5168.
 16. Cai W, Jing J, Irvin B, Ohler L, Rose E, Shizuya H, Kim UJ, Simon M, Anantharaman T, Mishra B, Schwartz DC: **High-resolution restriction maps of bacterial artificial chromosomes constructed by optical mapping.** *Proc Natl Acad Sci USA* 1998, **95**(7):3390-3395.
 17. Anantharaman T, Mishra B, Schwartz D: **Genomics via optical mapping. III: Contigging genomic DNA.** *Proc Int Conf Intell Syst Mol Biol* 1999:18-27.
 18. Jing J, Reed J, Huang J, Hu X, Clarke V, Edington J, Housman D, Anantharaman TS, Huff EJ, Mishra B, Porter B, Shenker A, Wolfson E, Hiort C, Kantor R, Aston C, Schwartz DC: **Automated high resolution optical mapping using arrayed, fluid-fixed DNA molecules.** *Proc Natl Acad Sci USA* 1998, **95**(14):8046-8051.
 19. Lai Z, Jing J, Aston C, Clarke V, Apodaca J, Dimalanta ET, Carucci DJ, Gardner MJ, Mishra B, Anantharaman TS, Paxia S, Hoffman SL, Craig Venter J, Huff EJ, Schwartz DC: **A shotgun optical map of the entire *Plasmodium falciparum* genome.** *Nat Genet* 1999, **23**(3):309-313.
 20. Lin J, Qi R, Aston C, Jing J, Anantharaman TS, Mishra B, White O, Daly MJ, Minton KW, Venter JC, Schwartz DC: **Whole-genome shotgun optical mapping of *Deinococcus radiodurans*.** *Science* 1999, **285**(5433):1558-1562.
 21. Lim A, Dimalanta ET, Potamouisis KD, Yen G, Apodaca J, Tao C, Lin J, Qi R, Skiadas J, Ramanathan A, Perna NT, Plunkett G 3rd, Burland V, Mau B, Hackett J, Blattner FR, Anantharaman TS, Mishra B, Schwartz DC: **Shotgun optical maps of the whole *Escherichia coli* O157:H7 genome.** *Genome Res* 2001, **11**(9):1584-1593.
 22. Chen Q, Savarino SJ, Venkatesan MM: **Subtractive hybridization and optical mapping of the enterotoxigenic *Escherichia coli* H10407 chromosome: isolation of unique sequences and demonstration of significant similarity to the chromosome of *E. coli* K-12.** *Microbiology* 2006, **152**(Pt 4):1041-1054.
 23. Reslewic S, Zhou S, Place M, Zhang Y, Briska A, Goldstein S, Churas C, Runnheim R, Forrest D, Lim A, Lapidus A, Han CS, Roberts GP, Schwartz DC: **Whole-genome shotgun optical mapping of *Rhodospirillum rubrum*.** *Appl Environ Microbiol* 2005, **71**(9):5511-5522.
 24. Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC Jr, Rohwer F: **Using pyrosequencing to shed light on deep mine microbial ecology.** *BMC Genomics* 2006, **7**:57.
 25. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, et al.: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**(7057):376-380.
 26. Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities.** *Genome Res* 1998, **8**(3):186-194.
 27. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8**(3):175-185.
 28. Gordon D: **Viewing and Editing Assembled Sequences Using Consed.** In *Current Protocols in Bioinformatics* Edited by: Baxevanis AD, Davison DB. New York: John Wiley & Co; 2004:11.12.11-11.12.43.
 29. Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Res* 1998, **8**(3):195-202.
 30. Gordon D, Desmarais C, Green P: **Automated finishing with autofinish.** *Genome Res* 2001, **11**(4):614-625.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

