labeled with FITC- or PE-conjugated mAbs directed against the following markers: CD45 (HLe1), CD34 (HPCA-2), CD38, CD15, CD33, CD71, CD2, CD3, CD4, CD7, CD8, CD19, CD20, CD16, CD56 (Becton-Dickinson), GPA, and CD71 (Pharmingen,). FITC- or PE-conjugated isotype-matched irrelevant mAbs were used as controls. BM, spleen, and PB cells from nontransplanted mice were used as negative control. Positive controls consisted of human BM or CB MCs. BM MCs were also cultured in semisolid media selective for human HPCs (*20*).
23. M. M. Verstegen *et al.*, *Blood* **91**, 1966 (1998).
24. M. Bhatia, D. Bonnet, B. Murdoch, O. I. Gan, J. E. Dick, *Nature Med.* **4**, 1038 (1998).
25. M. E. Garcia-Ojeda, S. Dejbakhsh-Jones, I. L. Weissman, S. Strober, *J. Exp. Med.* **187**, 1813 (1998).
26. For the fetal sheep xenografts (*3*, *27*, *28*), PB and BM MCs from chimeric fetuses or newborns, separated by Ficoll gradient, were evaluated for the presence of human cells by flow cytometry. BM MCs were also assayed for human HPCs in clonogenic culture by karyotyping of hematopoietic colonies. Human $CD34^+$ cells, isolated by MiniMACS column from BM MCs (*16*), were transplanted in secondary recipients.
27. C. I. Civin *et al.*, *Blood* **88**, 4102 (1996).
28. I. Kawashima *et al.*, *ibid.* **87**, 4136 (1996); D. R. Sutherland *et al.*, *Exp. Hematol.* **24**, 795 (1996); N. Uchida *et al.*, *Blood* **88**, 1297 (1996).
29. $CD34^-lin^-$ cells were purified as in (*24*). The $CD34^-lin^-KDR^+$ cell subfraction was analyzed and sorted as in (*16*).
30. R. J. Berenson *et al.*, *J. Clin. Invest.* **81**, 951 (1988); *Blood* **77**, 1717 (1991); W. I. Bensinger *et al.*, *ibid.* **88**, 4132 (1996); E. D. Zanjani, G. Almeida-Porada, A. W. Flake, *Int. J. Hematol.* **63**, 179 (1996).
31. U. Testa *et al.*, *Blood* **88**, 3391 (1996).
32. B. L. Ziegler *et al.*, *ibid.* **93**, 3355 (1999).
33. B. I. Terman *et al.*, *Oncogene* **6**, 1677 (1991).
34. P. E. Warburton, G. M. Greig, T. Haaf, H. F. Willard, *Genomics* **11**, 324 (1991)

3 February 1999; accepted 8 July 1999

Whole-Genome Shotgun Optical Mapping of *Deinococcus radiodurans*

Jieyi Lin,[1]* Rong Qi,[1] Christopher Aston,[1]† Junping Jing,[1]‡ Thomas S. Anantharaman,[2] Bud Mishra,[2] Owen White,[3] Michael J. Daly,[4] Kenneth W. Minton,[4] J. Craig Venter,[5] David C. Schwartz[1,2]§

A whole-genome restriction map of *Deinococcus radiodurans*, a radiation-resistant bacterium able to survive up to 15,000 grays of ionizing radiation, was constructed without using DNA libraries, the polymerase chain reaction, or electrophoresis. Very large, randomly sheared, genomic DNA fragments were used to construct maps from individual DNA molecules that were assembled into two circular overlapping maps (2.6 and 0.415 megabases), without gaps. A third smaller chromosome (176 kilobases) was identified and characterized. Aberrant nonlinear DNA structures that may define chromosome structure and organization, as well as intermediates in DNA repair, were directly visualized by optical mapping techniques after γ irradiation.

Detailed, structural knowledge of whole microbial genomes is of primary importance to many genomic studies, but this information has been difficult to obtain. Pulsed-field gel electrophoresis (PFGE) plus Southern (DNA) blot analysis (*1*) provides primary genome information but does not confidently size large circular genomes and frequently obscures the analysis of large episomal elements. Although an eight-enzyme restriction map of the *Escherichia coli* K12 genome was constructed in 1987 by Kohara *et al.* (*2*), this required a laborious approach involving partial digestion of 3400 phage clones followed by Southern blot analysis. Physical maps of the *Saccharomyces cerevisiae* genome were also prepared by painstaking restriction mapping of clones (*3*). Modern microbial genome analysis uses shotgun sequencing, followed by finishing efforts (*4, 5*). Whole-genome restriction maps may become an indispensable resource for large-scale genome sequencing projects. They facilitate sequence assembly by providing a scaffold for high-resolution alignment and verification of sequence assemblies (contigs), accurate genome sizing, and discernment of extrachromosomal elements (*6*).

Optical mapping is a system for the construction of ordered restriction maps from individual DNA molecules (*7, 8*) and has been used to prepare restriction maps of a number of clone types, including phage clones (*9*), yeast artificial chromosomes (*10*), bacterial artificial chromosomes (*6*), and, more recently, an entire electrophoretically separated chromosome (~1 Mb) from *Plasmodium falciparum* (*11*). An optical mapping approach for whole bacterial genome analysis is feasible because we can now mount and map extremely large, randomly sheared DNA molecules (0.4 to 2.4 Mb) that are digested with high cutting efficiency (70 to 90%). These parameters critically control the success rate of assembling the fragments and are well modeled by prior probabilistic (Bayesian) analysis (*12*). The contigs covering a whole genome were initially assembled manually, or later with the Gentig algorithm (*13*), which automatically computes contigs of genomic maps.

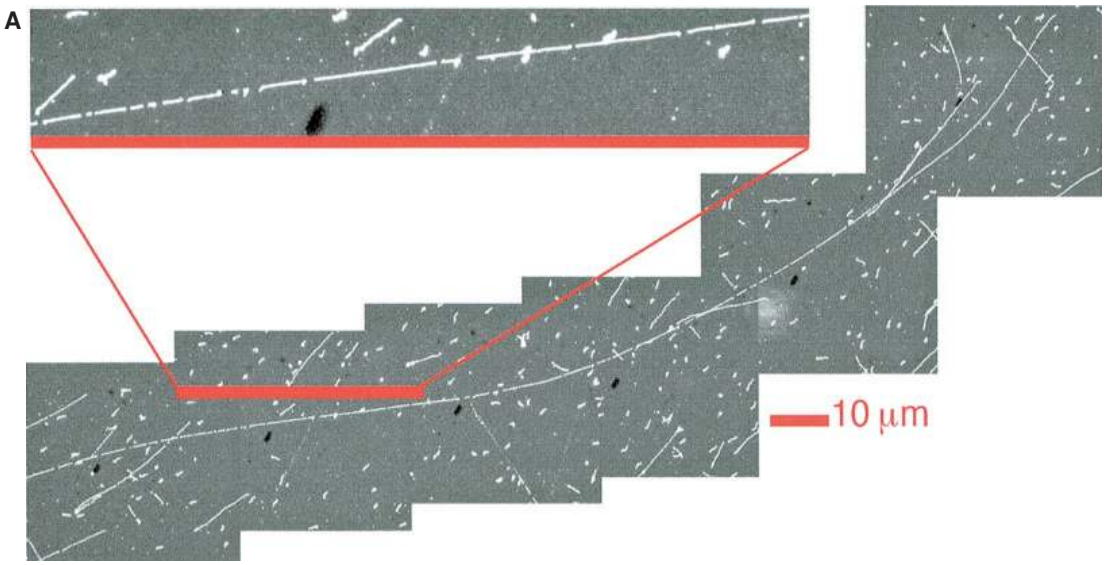
To efficiently collect such large molecules, we developed a semiautomated image acquisition system that collects successive images and correctly assembles them into one large superimage while maintaining proper pixel registration between images. A new image analysis system was developed [Visionade (*14*)] that enables markup of molecular images, allows for editing, and automatically calculates fragment masses and cutting efficiencies. A λ bacteriophage DNA sizing standard was used as follows. First, the total integrated fluorescence intensity of the standard was determined to be in the correct range. Second, the size of each genomic fragment in a particular image was calculated by dividing the fluorescence intensity of the DNA fragment by the average fluorescence intensity of the standards in the image and then multiplying this by the size of the standard. The images were not used as data if the cutting efficiency along the length of the molecule was less than 75%.

The development of Gentig enabled the rapid assembly of raw maps into a complete genome-wide map in minutes rather than months, with negligible false positives. Contigs of the *E. coli* genome were assembled with Gentig into a consensus map, which both reproduced the map constructed by hand and correlated with the map predicted by sequence. Gentig automatically generates contigs from optical mapping data by repeatedly combining the two islands that produce the greatest increase in probability density, excluding any contigs whose false positive overlap probability is unacceptable. The stan-

[1]W. M. Keck Laboratory for Biomolecular Imaging, Department of Chemistry, New York University, 31 Washington Place, New York, NY 10003, USA. [2]Courant Institute of Mathematical Sciences, Department of Computer Science, New York University, 251 Mercer Street, New York, NY 10012, USA. [3]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. [4]Uniformed Services University of the Health Sciences, Bethesda, MD 20814, USA. [5]Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA.

*Present address: Genome Therapeutics Corp., 100 Beaver Street, Waltham, MA 02154, USA.
†Present address: CNS Disorders, Wyeth-Ayerst Research, Princeton, NJ 08543, USA.
‡Present address: Smith Kline Beecham Pharmaceuticals, King of Prussia, PA 19406, USA.
§To whom correspondence should be addressed. Present address: Departments of Chemistry and Genetics, UW Biotechnology Center, University of Wisconsin, 425 Henry Mall, Madison, WI 53706, USA. E-mail: schwad01@mcrcr.med.nyu.edu

dard deviation, digestion rate, false cut rate, and false match rate are given as parameters to Gentig and affect the false negative probability and hence the number and distribution of contigs and gaps. Gentig also facilitates assembly of whole-genome maps for much larger eukaryotic genomes such as *Plasmodium falciparum* (*15*). Here, Gentig assembled a consensus map derived from 50 molecules



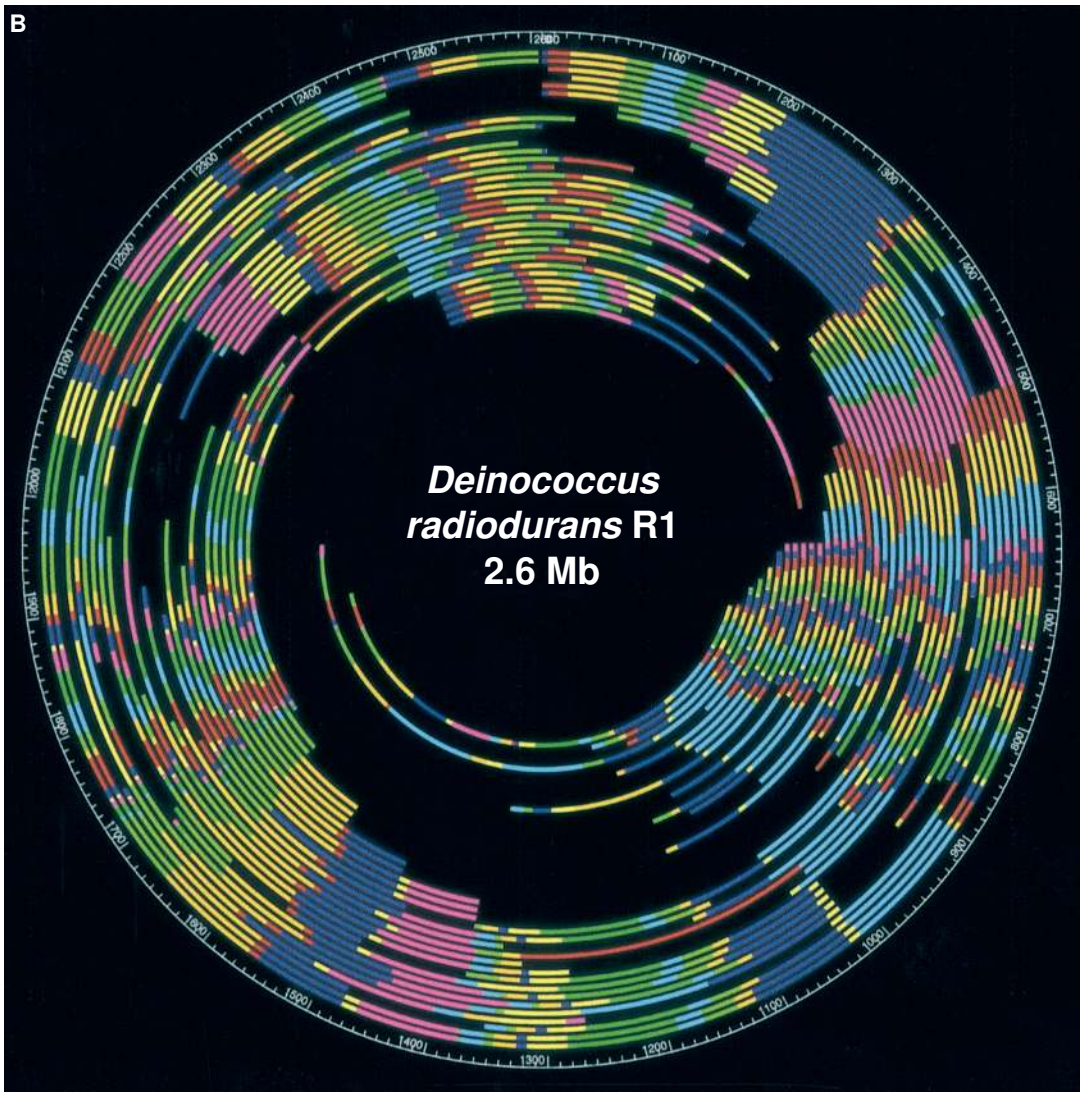


Fig. 1. Optical mapping of *D. radiodurans* genomic DNA with Nhe I. (**A**) Representative DNA molecule, 0.7 mm (2.4 Mb) in length, spanning six microscope fields, used to make an optical map. (**B**) Large chromosome restriction map generated by shotgun optical mapping. The outer circle shows the consensus map; the inner circles show the contig from which the consensus map was generated. The contig was assembled from 100 molecules by means of the Gentig algorithm. Nhe I fragment sizes (in kilobases) can be measured from the figure (*20*). Colors are arbitrarily assigned to homologous overlapping fragments.

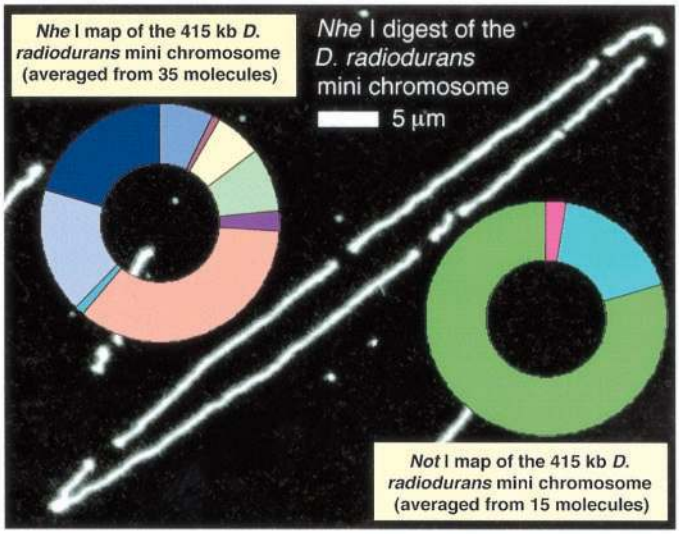


Fig. 2. Second (mini) chromosome of *D. radiodurans* identified by optical mapping. The image shows a representative second chromosome mapped with Nhe I (fragment sizes for the consensus map are 29.8, 4.6, 26.7, 35.4, 11.8, 145.0, 6.3, 69.8, and 85.5 kb). Second chromosomes were also mapped with Not I (fragment sizes for the consensus map are 11.5, 72.3, and 331.2 kb). Colors are arbitrary and correspond to fragment sizes, starting at the 12 o'clock position.

and found 147 out of the 150 cut sites predicted from the manually assembled map. The three cuts that Gentig failed to call were due to the lower depth of coverage applied in this benchmarking exercise. Whereas manually constructed maps of *E. coli* and *D. radiodurans* required several months to completely assemble, Gentig required only 20 to 30 min to generate contigs.

The radiation resistance of *D. radiodurans* may be a serendipitous result of its ability to survive periods of severe dehydration, which also fragments DNA. Efforts to understand the detailed mechanisms underlying this uniquely effective feat of DNA repair are now centered on the annotation results of a complete genome sequence. We thought an optical map would facilitate ongoing sequencing efforts, as a scaffold for sequence assembly, and might identify aberrant DNA structures associated with mechanisms of DNA repair. Before mapping the uncharacterized *D. radiodurans* genome, we benchmarked our genomic mapping system by mapping the sequenced (*4*) *E. coli* K12 substrain MG1665 (*16*). There was a <1% size difference in the optical versus sequence map [the same error rate was then seen when a Nhe I optical map of *D. radiodurans* was compared with preliminary sequence information (*17*)]. *Deinococcus radiodurans* R1 (American Type Culture Collection 13939) is a radiation-resistant organism with a comparably sized genome (*18*). A representative molecule is shown in Fig. 1A. The final map was assembled without gaps at an average depth of 35×, using 157 molecules, with an average fragment size of 29 kb (*19*). This depth of coverage gave the map very low errors (*20*). Contigs were also assembled with Gentig (Fig. 1B) (*21*). The genome size was calculated to be 2.61 Mb by manual assembly and 2.59 Mb by Gentig assembly. Gentig assembled a consensus map derived from 100 molecules. Manual editing can assign more molecules to the contig in a way that is not yet modeled in Gentig. To further confirm our map, we constructed a rare-cutter, Not I, map (macro) and aligned it, using Gentig, with the Nhe I (micro) map, which was analyzed by Southern blots (*22*).

Large extrachromosomal elements are notoriously difficult to characterize. For example, circular DNA molecules (25 kb) identified in *Entamoeba histolytica* (*23*) migrated as high molecular weight bands on pulsed-field gels, resembling chromosomes, and were only characterized after single-molecule analysis with electron microscopy. Optical mapping of the *D. radiodurans* genome also provided insight into the genomic structure of the organism. The genome is reported to be 3.17 Mb in size, as assessed by PFGE (*24*). We calculated the genome to be 2.6 Mb in size by optical means and also discovered the presence of a large episome, 415 kb in size, that was previously considered to be a part of a large single chromosome and which accounted for the missing portion (Fig. 2). We believe this episome to be a second (mini) chromosome, on the basis of its size and 1:1 stoichiometry with the large chromosome as analyzed by PFGE. This second chromosome has been shown from sequence assembly to contain genic regions critical to cell function (*25*), and thus is not likely to have resulted from a duplication event. In addition, we found no map homology between the second chromosome and any other genomic region. Interplasmidic and intrachromosomal recombination have been found to occur in *D. radiodurans* at high frequency after exposure to ionizing radiation (*26*). Dessication may induce similar changes. Consequently, the same type of chromosomal architecture may not exist in all *D. radiodurans* strains. We also saw images containing a third circular molecule, which was sized at 176 kb [SD = 10 kb, n = 8 (*25*)]. Nhe I cut sites were apparent, but an insufficient number of molecules were collected to enable us to make Nhe I maps. Other smaller chromosomes or large plasmids (*27*) may be present in the *D. radiodurans* genome, al-

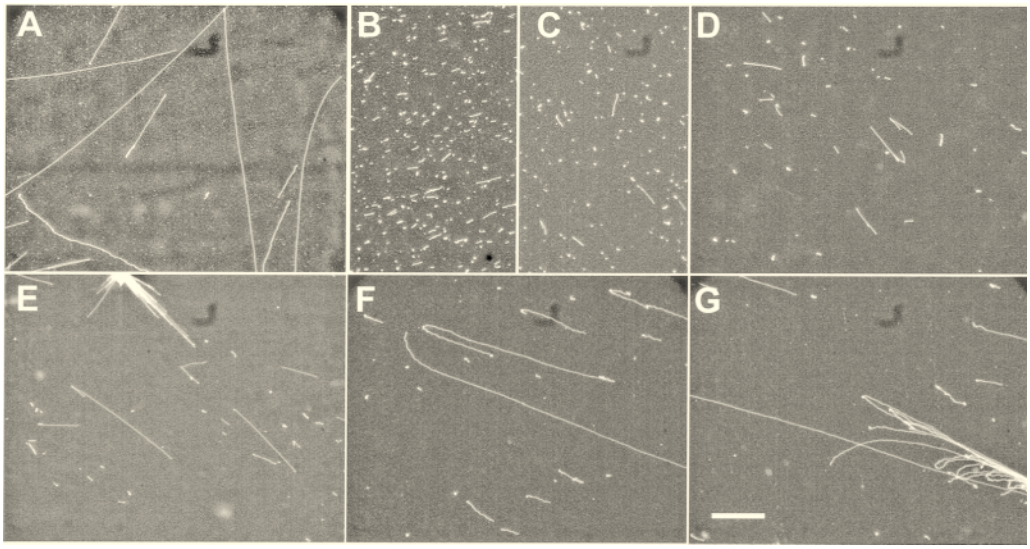


Fig. 3. DNA repair after irradiation, visualized by optical mapping. DNA samples were extracted at different time points from *D. radiodurans* R1 (wild type) cells. The irradiation and recovery conditions (time interval after irradiation) for each image were: (**A**) 0 Gy, 0 hours; (**B**) 17.5 Gy, 0 hours; (**C**) 17.5 Gy, 2 hours; (**D**) 17.5 Gy, 4 hours; (**E**) 17.5 Gy, 6 hours; (**F**) 17.5 Gy, 16 hours; (**G**) 17.5 Gy, 24 hours. Scale bar, 20 μm. Histogram analysis (*34*) showed consistent elongation of DNA molecules into full-length genomic molecules. The recA-deficient strain (rec 30) showed no evidence of repair, and the observable number of molecules decreased with time as a result of cell death (*35*).

though we did not map any such molecules.

Whereas the genomes of most organisms are irreversibly shattered by the effects of high levels of ionizing radiation, *D. radiodurans* is able to efficiently reconstruct an intact genome through means that are not yet fully understood. Dissecting its ability to deal with severe DNA damage may uncover new general mechanisms of DNA repair. The postulated DNA repair system of *D. radiodurans* is in part facilitated by multiple chromosome copies held in register. According to this model, double-stranded breaks are repaired by recA-dependent recombination, with the undamaged DNA duplex acting as the template (*28, 29*). One model for holding chromosome copies together in sequence alignment is by the presence of persistent Holliday junctions (*30, 31*). We saw no evidence of DNA molecules containing Holliday junctions, which might have been visualized as bundles of DNA molecules or molecules with a χ-shape. We also saw no evidence of back-to-back repeated regions containing Nhe I sites. These would have been noticeable as multiple ordered fragments of the same size. We did, however, see large repeated DNA molecules, which shared the Not I digestion pattern of the second chromosome (*32*). The pattern (10, 70, 340 kb) was repeated up to three times, and 10 such molecules were observed. We do not know the origin of such molecules, but we speculate that these may be observable replication intermediates.

Images of DNA extracted from cells after γ irradiation of 17.5 Gy (Fig. 3) were examined from both wild-type and recA-deficient [rec 30 (*30*)] strains. We have previously shown that DNA fragments flanked by identical 4-kb sequences are efficiently circularized after γ irradiation in both wild-type and recA strains (*31*). *Deinococcus radiodurans* contains hundreds of short repeats (~150 base pairs) and numerous long repeats, providing ample substrate for circularization of large portions of the genome. Such circles could be expected to protect vulnerable DNA ends produced by double-strand breaks from exonucleolytic degradation. Furthermore, homologous recombination among these circles could progressively generate larger circles, ultimately restoring the circular chromosomes without rearrangements. Optical mapping did not show the presence of circles of any dimension, although it should be noted that the resolution of light microscopy would not enable us to characterize circles smaller than ~30 kb.

However, the imaging results graphically showed the repair of the genome in wild-type cells, but not in recA cells, by unbranched linear elongation of extracted molecules from small kilobase fragments to essentially intact chromosomes. Although this is consistent with previous conclusions based on gel electrophoresis and sucrose gradient technologies, this finding eliminates circular intermediates as a generalized mechanism of DNA repair. What protects the ends of these molecules from degradation remains unknown. After irradiation we estimated that the average fragment size was ~15 kb (Fig. 3B), corresponding to about 200 double-strand breaks per genome.

The complete restriction mapping of whole genomes may catalyze the development of new modes of genome analysis, unhindered by the need to generate and map large numbers of clones. Because ensembles of single molecules are analyzed, small amounts of starting material are required, enabling mapping of microorganisms that are problematic to culture. Shotgun mapping obviates the need for library construction with its associated cloning artifacts, and also enables mapping of organisms with DNA, which is difficult to clone. Perhaps most important, optical mapping renders new biological insight by readily providing a picture of the basic organization of an entire genome, revealing the number of chromosomes and the presence of extrachromosomal elements as well as providing a means of directly examining dynamic processes such as DNA repair. If maps can be rapidly generated, microbial populations can be analyzed at the whole-genome level to reveal genotypic differences that can be linked to phenotype.

References and Notes

1. D. C. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984); C. L. Smith, J. G. Econome, A. Schutt, S. Klco, C. R. Cantor, *Science* **236**, 1448 (1987).
2. Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987).
3. A. J. Link and M. V. Olson, *Genetics* **127**, 681 (1991); L. Riles *et al.*, *ibid.* **134**, 81 (1993).
4. F. R. Blattner *et al.*, *Science* **277**, 1453 (1997).
5. R. D. Fleischmann *et al.*, *ibid.* **269**, 496 (1995).
6. W. Cai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3390 (1998).
7. C. Aston, C. Hiort, D. C. Schwartz, *Methods Enzymol.* **303**, 55 (1999); D. C. Schwartz *et al.*, *Science* **262**, 110 (1993); W. Bautsch *et al.*, in *Genome Mapping, A Practical Approach*, P. H. Dear, Ed. (Oxford Univ. Press, New York, 1997), pp. 281–313; D. C. Schwartz and A. Samad, *Curr. Opin. Biotechnol.* **8**, 70 (1997); A. Samad *et al.*, *Nature* **378**, 516 (1995); W. Wang, J. Lin, D. C. Schwartz, *Biophys. J.* **75**, 513 (1998); J. Reed, E. Singer, G. Kresbach, D. C. Schwartz, *Anal. Biochem.* **259**, 80 (1998).
8. T. S. Ananthraman, B. Mishra, D. C. Schwartz, *J. Comp. Biol.* **4**, 91 (1997).
9. X. Meng, K. Benson, K. Chada, E. Huff, D. C. Schwartz, *Nature Genet.* **9**, 432 (1995).
10. W. Cai, H. Aburatani, D. Housman, Y. Wang, D. C. Schwartz, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5164 (1995).
11. J. Jing *et al.*, *Genome Res.* **9**, 175 (1999).
12. T. S. Anantharaman, B. Mishra, D. C. Schwartz, *Tech. Rep. 760* (Courant Institute, New York University, New York, 1998); T. S. Anantharaman, B. Mishra, D. C. Schwartz, in *Seventh International Conference on Intelligent Systems for Molecular Biology (ISMB99),* in press.
13. Gentig (*12*) uses an approximation algorithm for finding an almost optimal scoring set of contigs, while constraining the false positive error rate below a negligible value. Under a simple overlap rule, dubbed Type D, that determines when two genomic DNA molecules can be deemed to have a common subfragment, a conservative estimate of the false probability can be given as

$$\frac{4}{p_c^4}\exp(-\beta n/2)\sum_{i=k}^{\infty}\frac{(\beta n/2)^i}{i!}$$

where p_c is the digestion rate, β is the relative sizing error, n is the expected number of restriction fragments per genomic DNA molecule, and k is the integer parameter directly related to overlap threshold ratio θ.
14. E. Lee, B. Porter, E. Huff, unpublished data.
15. Z. Lai, unpublished data.
16. F. Blattner, unpublished data.
17. O. White *et al.*, data not shown.
18. *Deinococcus radiodurans* has a number of powerful DNA repair systems, including a fast-acting dimer excision system as well as an unusual double-stranded DNA break repair system (*31*) [V. Mattimore and J. R. Battista, *J. Bacteriol.* **178**, 633 (1996); K. W. Minton, *Mutat. Res.* **363**, 1 (1996)]. It carries multiple (4 to 10) copies of homologous chromosomes and plasmids, which are postulated to exist in pairs that are aligned relative to one another (*28*) [M. T. Hansen, *J. Bacteriol.* **134**, 71 (1978); M. J. Daly and K. W. Minton, *ibid.* **177**, 5505 (1995)].
19. J. Lin, data not shown.
20. The systematic error associated with the mapmaking for the second chromosome, for which complete sequence information is available from O. White, was <0.2%. The contig was assembled from 100 molecules by means of the Gentig algorithm. The sizes (in kilobases) of the 91 Nhe I fragments were 5.3, 18.9, 47.2, 23, 28, 18.4, 30.8, 40.8, 132.5, 3.8, 16.4, 29.9, 39, 13, 61.8, 34.4, 50.9, 45.6, 11.5, 12.5, 17.8, 22.9, 19.5, 3.6, 4.2, 3, 28.8, 6.4, 27.7, 9.6, 15, 26.9, 6.2, 15.6, 20.2, 7.5, 118.1, 11.3, 92, 15.7, 30, 23, 94.2, 34.1, 8.8, 18.9, 8.1, 20.1, 1.9, 87.5, 13.5, 79, 9.8, 93.2, 58.8, 3.1, 10.5, 5, 2.9, 8.6, 29.1, 6.4, 16.1, 26.3, 28.7, 17.7, 4.4, 18.4, 8.4, 17.1, 42.5, 30, 64.1, 47, 17.3, 27.7, 46.1, 5.1, 65.6, 46.2, 18.2, 14.4, 37.9, 35.7, 33.3, 23.2, 2.6, 28.8, 13.4, 37.5, and 50.5.
21. A direct calculation for *D. radiodurans* indicates that about 110 genomic DNA molecules, each of expected length 500 kb, digested with a six-cutter enzyme, with a digestion rate of 50% and a relative sizing error of 10%, would overlap these molecules into one single island with ~0.04% probability that a declared overlap in the computed contig is false.
22. J. Lin and R. Qi, data not shown. We mapped a series of probes developed from *D. radiodurans* sequences to the composite map by Southern blot analysis (J. Lin *et al.*, data not shown). Genomic DNA fragments were prepared by Not I digestion in solution (tube), surface-mounted, then optically mapped with Nhe I and overlapped with Gentig. The sizes (in kilobases) of the nine Not I fragments were 381, 354, 468, 239, 400, 180, 228, 253, and 103. Any given sequence or probe was localized to a particular large Not I restriction fragment and to a Nhe I restriction fragment by Southern blotting. Probes were generated by the polymerase chain reaction from genomic DNA templates with primers designed from preliminary shotgun sequence information: recA, (katA, orf1, plsx), DNA polymerase, and ribonuclease P. Results confirmed optical maps and preliminary sequence assemblies.
23. C. Lioutas, C. Schmetz, E. Tannich, *Exp. Parasitol.* **80**, 349 (1995).
24. M. J. Daly, unpublished data.
25. O. White *et al.*, unpublished data.
26. M. J. Daly and K. W. Minton, *Gene* **187**, 225 (1997); *J. Bacteriol.* **177**, 5495 (1995).
27. M. Kikuchi, S. Kitayama, S. H. Sjarief, H. Watanabe, *Radiat. Res.* **139**, 123 (1994).
28. M. J. Daly and K. W. Minton, *Science* **270**, 1318 (1995).
29. K. W. Minton and M. J. Daly, *Bioessays* **17**, 457 (1995).
30. DNA was prepared from γ ray–irradiated cells (D_{37}, the dose yielding a 37% survival rate for plateau- phase wild type R1), using gel insert methodology. Agarose

inserts were melted at 72°C, diluted 500× in TE buffer [1 mM EDTA and 10 mM tris (pH 8.0)], and mounted for optical mapping. More than 10,000 irradiated DNA molecules were studied for the presence of circles or other aberrant structures.
31. M. J. Daly and K. W. Minton, *J. Bacteriol.* **178**, 4461 (1996).
32. J. Lin, data not shown.
33. B. B. Bederson *et al.*, in *Human Factors in Web Development*, L. Ratner, E. Grose, C. Forsythe, Eds. (Erlbaum, Mahwah, NJ, 1998), p. 255.
34. J. Lin, data not shown.
35. _____, data not shown.
36. We thank T. Burland for Geneplot software, S. Paxia for Pad software (*33*), and F. Blattner for prepublication *E. coli* sequence. Partially funded by the U.S. Department of Energy and NIH, by Department of Energy Office of Biological and Environmental Research grants DE-FG02-98ER62583, DE-FG02-97ER62492, DE-FG07-97ER20293, and DE-FG02-98ER62554, and by U.S. Public Health Service grants GM39933 and HG0025-08.

11 February 1999; accepted 2 August 1999

Historical Genetics: The Parentage of Chardonnay, Gamay, and Other Wine Grapes of Northeastern France

John Bowers,[1]* Jean-Michel Boursiquot,[2] Patrice This,[3] Kieu Chu,[1] Henrik Johansson,[1] Carole Meredith[1]†

The origins of the classic European wine grapes (*Vitis vinifera*) have been the subject of much speculation. In a search for parental relationships, microsatellite loci were analyzed in more than 300 grape cultivars. Sixteen wine grapes that have long been grown in northeastern France, including 'Chardonnay', 'Gamay noir', 'Aligoté', and 'Melon', have microsatellite genotypes consistent with their being the progeny of a single pair of parents, 'Pinot' and 'Gouais blanc', both of which were widespread in this region in the Middle Ages. Parentage analysis at 32 microsatellite loci provides statistical support for these relationships.

The wines of northeastern France, notably those of the Burgundy and Champagne regions, have been highly regarded for centuries. Like most of the world's finest wines, they are made entirely from old cultivars of *Vitis vinifera* L. The cultivars most strongly associated with this part of France are 'Pinot noir' and 'Chardonnay', which are used both for Champagne (*1*) and also for the best red and white wines, respectively, of the Côte d'Or in the heart of Burgundy. These two grapes are now grown in many of the world's wine regions. In the southern part of Burgundy, the red wines of Beaujolais are made primarily from 'Gamay noir'. Several other cultivars, including 'Aligoté', 'Melon', and 'Sacy', are used in wines carrying regional Burgundy appellations.

Grapevines are propagated vegetatively, so that the individual vines of a cultivar are genetically identical to each other (except for somatic mutations) and to the single original seedling from which the cultivar originated. While some cultivars may have originated in the regions with which they are now associated, others are thought to have been introduced by traders or conquerors, most notably the Romans. Although a few varieties have been produced by controlled crosses since the mid-1800s, most *V. vinifera* cultivars in existence today are centuries old and are thought to have arisen by several mechanisms—domestication of wild vines, spontaneous crosses between wild vines and cultivars, and spontaneous crosses between cultivars (*2*). Such a spontaneous cross between two cultivated varieties gave rise to 'Cabernet Sauvignon', the most important cultivar of Bordeaux and arguably the most highly regarded red wine grape in the world today (*3*).

Any wild vines that were parents of today's important wine grapes cannot be identified, because they no longer exist. Parents that are themselves cultivars, however, may still exist in collections if not in cultivation. In a search for the parents of some important French wine grapes, we analyzed 322 cultivars of *V. vinifera* (*4*), including most extant old French cultivars.

Samples of 51 cultivars were obtained from the vineyards at the University of California at Davis and the rest were taken from the variety collection maintained by Institut National de la Recherche Agronomique at Domaine de Vassal, near Montpellier, France. After an initial screening of all cultivars at 17 microsatellite loci (*5*), we compared microsatellite alleles within all possible sets of three to identify pairs of cultivars that could have contributed the alleles of the third cultivar (*6*). A subset of cultivars was then further analyzed at 15 additional loci (*7*).

On the basis of 32 loci, 16 cultivars had microsatellite genotypes consistent with their being the progeny of a single pair of parents—'Pinot' and 'Gouais blanc' (Fig. 1) (*8*–*10*). For each of the 16 putative 'Pinot' × 'Gouais blanc' progeny, we calculated parentage indices to compare the probability of the observed progeny alleles if it had the putative parents to the probability of those alleles, if it had two random parents, or if the parents were close relatives of the putative parents. We show the detailed parentage indices for 'Chardonnay' in Table 1 and summaries for all 16 progeny cultivars in Table 2. The likelihood ratios show that the putative parents are 10^{12} to 10^{15} times

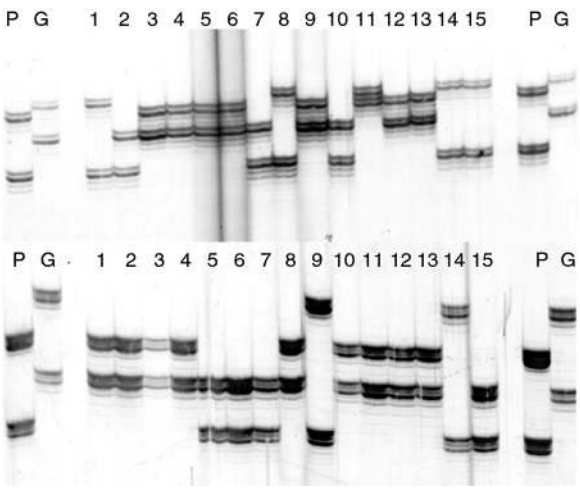


Fig. 1. Inheritance of parental microsatellite alleles by progeny cultivars for locus VVMD5 (top) and VVMD28 (bottom). Microsatellites were amplified from genomic DNA, electrophoresed in polyacrylamide gels, and visualized by silver staining. Lanes represent (from left to right) (P) Pinot noir, (G) Gouais blanc, (1) Aligoté, (2) Aubin vert, (3) Auxerrois, (4) Bachet noir, (5) Beaunoir, (6) Chardonnay, (7) Franc noir de la Haute Saône, (8) Gamay blanc Gloriod, (9) Gamay noir, (10) Knipperlé, (11) Melon, (12) Peurion, (13) Romorantin, (14) Roublot, (15) Sacy, (P) Pinot noir, and (G) Gouais blanc.

[1]Department of Viticulture and Enology, University of California, Davis, CA 95616, USA. [2]Unité de Formation et de Recherche de Viticulture, Ecole Nationale Supérieure Agronomique, 34060 Montpellier, France. [3]Unité de Recherches de Génétique et d'Amélioration des Plantes-Viticulture, Institut National de la Recherche Agronomique, 34060 Montpellier, France.

*Present address: Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA.
†To whom correspondence should be addressed. E-mail: cpmeredith@ucdavis.edu