Genomic DNA Amplification from a Single Bacterium

Arumugham Raghunathan,† Harley R. Ferguson, Jr., Carole J. Bornarth,†
Wanmin Song, Mark Driscoll,‡ and Roger S. Lasken*

Molecular Staging, Inc., New Haven, Connecticut

Genomic DNA was amplified about 5 billion-fold from single, flow-sorted bacterial cells by the multiple displacement amplification (MDA) reaction, using φ 29 DNA polymerase. A 662-bp segment of the 16S rRNA gene could be accurately sequenced from the amplified DNA. MDA methods enable new strategies for studying nonculturable microorganisms.

The multiple displacement amplification (MDA) reaction uses the φ 29 DNA polymerase and random primers to amplify DNA templates (1-3, 5-6). Amplification from small specimens has enabled novel research approaches (reviewed in reference 7), including genetic analysis of single blastomeres for use in preimplantation diagnosis of embryos (4). A method to amplify genomic DNA from nonculturable bacteria would allow direct analysis of virtually any microbe. We demonstrate here the use of MDA to achieve several-billion-fold amplification of genomic DNA from a single bacterium. MDA could be used for a wide range of approaches for discovery of new species, population and polymorphism analysis, diagnostics, and rapid detection of pathogens.

As a test case, *E. coli* cells (ATCC 10798; K-12 strain) were isolated (fluorescence-activated cell sorter Vantage flow cytometer [Becton Dickinson] using CellQuest and CytoCount softwares). To demonstrate proficiency in flow sorting, 180 putative cells were collected and vigorously vortexed in 10 μl phosphate-buffered saline to separate cells in the event that more than one cell was obtained, and the number of CFU was determined (Fig. 1). We have not excluded the possibility that dead cells or free DNA may be present along with the viable cell and could potentially contribute a template to amplification reaction mixtures. Next, single cells were collected in 27 μl Tris-EDTA, lysed with 3 μl 0.4 M KOH–10 mM EDTA at 65°C for 3 min, and neutralized with 3 μl Tris-HCl, pH 4, and whole-genome amplification was carried out with 100-μl reaction mixtures according to the instructions of the manufacturer (REPLI-g kit; QIAGEN Sciences Inc., Germantown, MD).

The amplified DNA was a suitable template for cycle sequencing (HHMI/Keck Biotechnology Resource Laboratory, Yale University) of the multicopy 16S rRNA gene directly from the amplified DNA (Fig. 2). Reaction mixtures contained the 5′-GCGT GGACTACCAGGGTATCTAATCC-3′ sequencing primer and 5 μl (3.5 μg) amplified DNA for 50 cycles. No miscalls resulted in the 662-bp read length (nucleotide databank at the NCBI site http://www.ncbi.nlm.nih.gov/BLAST/). Of 10 other single cells

amplified by MDA, 7 also supported accurate cycle sequencing of the 16S rRNA gene and 3 failed (data not shown). For a negative control, phosphate-buffered saline lacking *E. coli* cells was run through the flow cytometer, collected into four wells, and added to MDA reaction mixtures. No 16S rRNA sequence was obtained from these samples by cycle sequencing (data not shown). Therefore, addition of the *E. coli* cell to the MDA reaction is required and the obtained sequence is not derived from contaminating DNA in the flow cytometer or MDA reagents. The ability to sequence the 16S rRNA gene directly from DNA amplifications using a single cell should be a very useful method for researchers in the fields of ecology, evolution, and taxonomy and for discovery of new species. We were not able to sequence several single-copy genes directly from either amplified or unamplified genomic DNA (data not shown), presumably because the complexity of bacterial genomes makes specific sequencing and primer annealing difficult. Therefore, library construction or PCR amplification of intended sequencing targets from the MDA products may be needed for most sequencing. MDA was carried out from single cells, a 2-μl aliquot of the MDA reaction mixture was used in PCRs (Table 1), and 2.5-μl aliquots of the PCR mixtures (cleaned up with a QIAquick PCR purification kit [Qiagen Inc.]) were successfully used in cycle sequencing reactions (data not shown).

Quantitative PCR by the TaqMan assay, as described previously (5), was used to investigate the quality and coverage of amplified DNA. The average locus representation values (5) of four loci, Exo, GlyS, HolA, and PcnB (Table 1), were about 135%, 85%, and 72%, starting from 100, 50, and 10 cells, respectively (Fig. 3A). A value of 100% indicates that the locus is present at the same copy number in the amplified DNA as in the starting template. The individual locus representation values tended to drop for amplifications from fewer cells. Nevertheless, of 84 different MDA reaction mixtures containing one *E. coli* cell, 10 different loci (Table 1) were detected about 50% of the time (Table 2). The efficiency of MDA was probably considerably higher than 50%, since not all of the tubes successfully receive a cell from flow cytometry (Fig. 1). No locus had an average representation higher that 54% or lower than 15% for the 84 MDA reactions (Fig. 3B), with an average across all 10 loci of about 30%. Based on these initial data from 10 randomly chosen *E. coli* loci, it can be estimated that about 30% of the amplified DNA is actually *E. coli* sequence, with the rest thought to be artifact synthesis, such as primer dimers or sequence derived from trace contaminating DNA (5). Ar-

* Corresponding author. Present address: Allegheny-Singer Research Institute, 320 East North Ave., Pittsburgh, PA 15212-4772. Phone: (412) 359-6214. Fax: (412) 359-8169. E-mail: rlasken2000 @Yahoo.com.
† Present address: QIAGEN Inc., 27220 Turnberry Lane, Valencia, CA 91355.
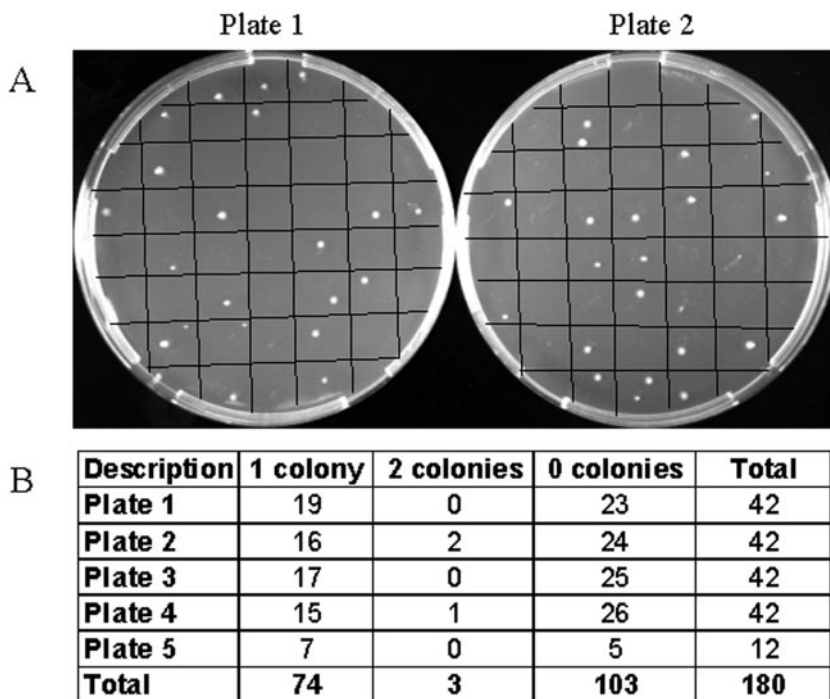‡ Present address: 454 Corporation, New Haven, Conn.

FIG. 1. Verification of single-cell isolation by flow sorting. (A) Two examples of the five plates (42 squares per plate grid). (B) Colonies growing from each well were counted after an overnight incubation.

TABLE 1. Probe and primer sequences for PCR and TaqMan analysis of MDA reactions

| Locus	Probe sequence (5′-3′)	Primer sequence (5′-3′)[a]
Exo	CGCATTCATGCCGTTCAGGG	L, CGTGGCTGTTCATTTGCTTA R, TCTGGTAAACGCTGATGACG
OmpA	CCCAGGCTCAGCATGCCGTT	L, TGACCGAAACGGTAGGAAAC R, TGAGTACGCGATCACTCCTG
GlyS	GACGCAGCGCAAACGGGTCT	L, ATACAGACGCACCGCTTCTT R, TGAATGAGCAGTATCAGCCG
TopA	TGCCACCTGGTCGAGGCTGT	L, TCGATCATTTCGACCATTCA R, CCCTCTTCCGGATCTTTTTC
HolA	CAAACAACGCACGCAGTGGC	L, CTGCCATACCCGATGCTTAT R, CAGCGAACCGGTTATTTTGT
CadA	GGCCACCGCTCATACCGCAT	L, GTGAGTGGACTGGGTTTCGT R, CAGCACGCTACCATTGCTAA
ExuR	CCTCAAGGCGCTGATTCGCA	L, AAAATGTGGACCCAGCGTAG R, ATAGATAGCGGTCGGCATTG
PcnB	ACAGGCGTGCCGGTGGGATA	L, TAGCCCGCTTGTAGCAGTTT R, GCGTTATCCGTCTGATTGGT
EutC	ATCCTGCTGGTGGGCGAACG	L, ATACGTTTCGCCAAATCCAC R, CGCGTGAAGATTGAAGATCA
Nth	AGGGCTTCAAGCGCAGCACG	L, CGCCTTTTGAATTGCTGATT R, ACGTTGGCTGTTTTACGACC

[a] L, left; R, right.
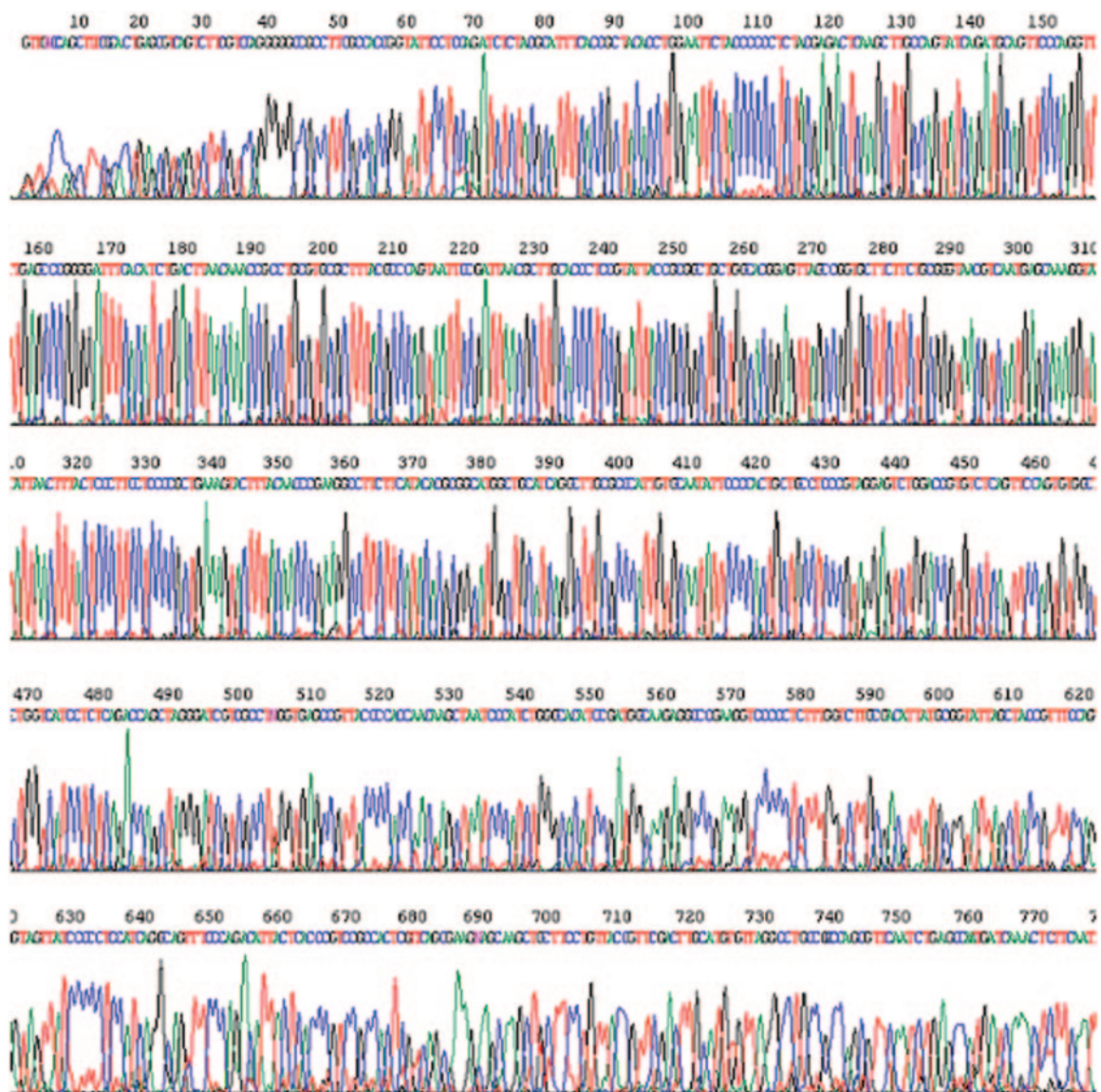
FIG. 2. DNA sequencing of the 16S rRNA gene by direct primer annealing to DNA amplified from a single cell. Five microliters of DNA amplified from a single cell was sequenced on an Applied Biosystems 3730 capillary DNA analyzer.
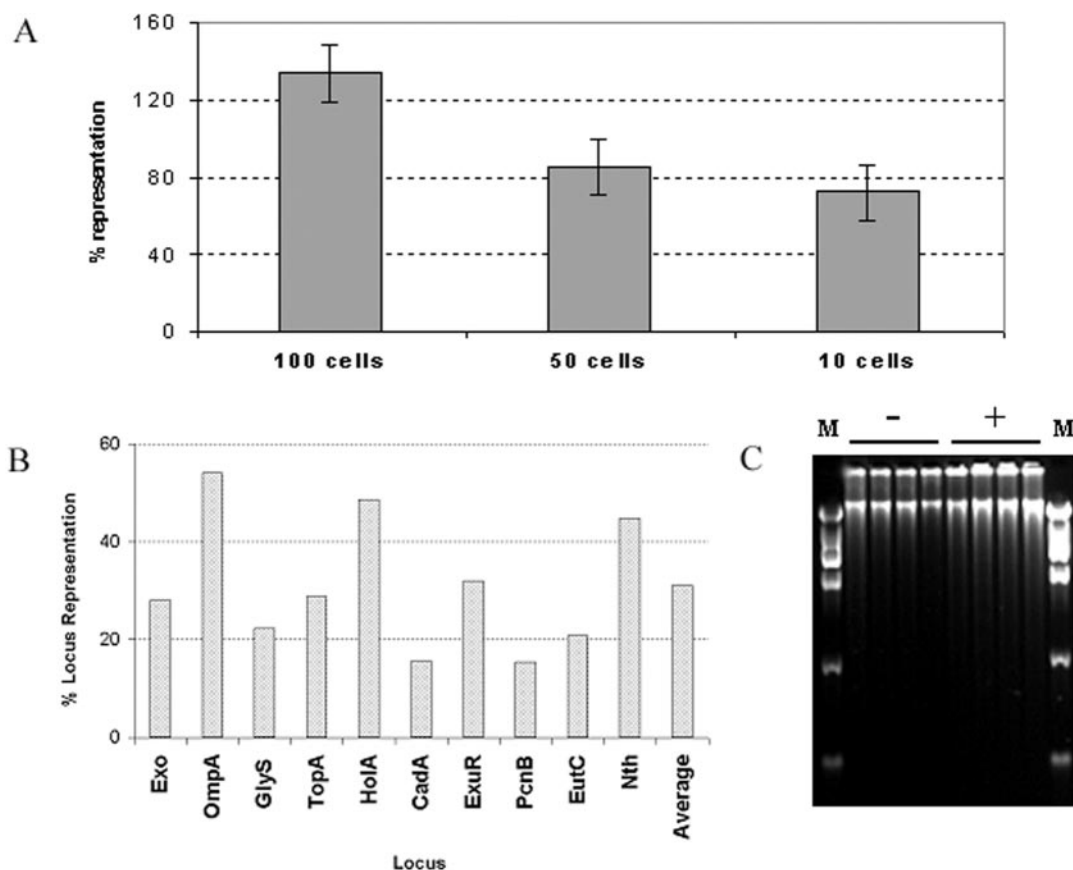
FIG. 3. Analysis of amplified DNA by quantitative PCR. (A) Each bar is the average of results for four different loci and 10 replicate MDA reactions for approximately 100, 50, and 10 flow-sorted *E. coli* cells. Error bars are 1 standard deviation. (B) Single cells were sorted into 84 microtiter plate wells, lysed, and subjected to whole-genome amplification by MDA. Average locus representation ($n = 84$) is plotted on the *y* axis for 10 individual loci in the *E. coli* genome. The average of all 10 loci is represented in the last bar. (C) Amplified DNA (negative controls and single-cell samples) resolved by agarose gel electrophoresis. Lanes M, molecular size markers.

tifact synthesis occurs even in the absence of added cells (Fig. 3C) and is currently unavoidable. About 80 μg of amplified DNA was produced in a 100-μl MDA reaction mixture (PicoGreen analysis; Molecular Probes, Invitrogen, CA). Even with an estimate that only 30% of the DNA may be specific, a yield of 24 μg of *E. coli* sequence corresponding to a 5×10^9-fold amplification from the 5 fg of DNA present in a single *E. coli* cell is calculated. While the presence of artifact sequence can be problematic for some applications, it is not expected to be an obstacle for PCR-based methods where specific primers can distinguish the amplified bacterial sequence.

For the 84 replicate MDA reactions, most locus representation values (Table 2) ranged from 0.1% to 1,211%. Stochastic effects of starting with a single genomic copy probably result in this amplification bias, which overrepresents some sequences and underrepresents others. Nevertheless, extremely high levels of amplification occurred for all regions tested. Even a 0.1% efficiency corresponds to a 10^6- to 10^7-fold DNA amplification from the single cell and will enable many kinds of specific primer/probe applications.

Amplification of a second species, *Myxococcus xanthus* (a gift of Dan Santi, Kosan Bioscience), also yielded amplified sequence based on TaqMan PCR for the C-methyltransferase

gene (M12-ta locus; representative MDA reactions) (Fig. 4). Five of 44 MDA reactions carried out on single flow-sorted cells were successful and had a minimum of 13% and a maximum of 74% representation of the locus (data not shown).

While these results demonstrate valuable uses for MDA in single-cell analysis, several limitations of the method remain to be resolved. Some microbes are expected to be more difficult than *E. coli* to lyse. The considerable amplification bias for
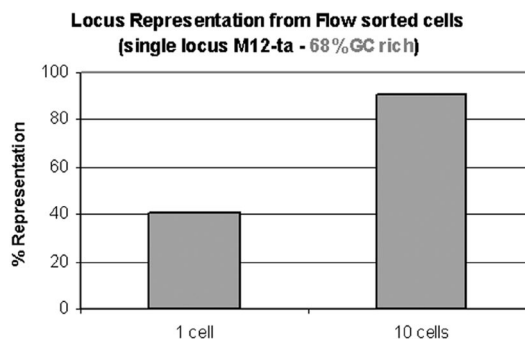

FIG. 4. Whole-genome amplification from 1 or 10 *M. xanthus* cells. Results of a quantitative assay for a representative locus are shown.

TABLE 2. Quantitative PCR analysis by TaqMan assay for representation of 10 loci in 84 different MDA reactions

MDA no.	% Representation of indicated locus									
	Exo	OmpA	GlyS	TopA	HolA	CadA	ExuR	PcnB	EutC	Nth
1	1.4	0.0	0.1	0.0	14.5	0.3	0.0	0.2	0.0	2.3
2	0.3	0.0	6.5	1.7	0.6	3.6	7.6	0.0	0.9	0.4
3	139.0	95.3	0.0	0.0	4.0	0.5	0.0	0.0	10.7	0.5
4	0.0	52.4	0.3	0.2	0.4	0.0	0.0	0.0	0.7	24.7
5	62.1	102.1	0.0	0.3	0.0	20.8	15.4	0.0	0.1	0.1
6	130.2	0.6	0.0	0.1	4.6	0.0	0.2	12.2	2.1	1.1
7	45.4	0.0	0.0	0.0	0.0	2.1	0.1	0.0	0.0	0.0
8	0.0	11.2	28.6	0.0	2.8	0.0	0.1	0.0	0.2	0.1
9	0.7	77.1	20.7	230.5	50.0	8.8	12.3	1.0	12.6	74.2
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	55.3	214.8	29.4	56.8	17.4	20.5	8.5	35.7	52.2	58.6
12	119.7	90.0	31.9	55.5	15.4	22.2	23.4	14.3	1.6	65.1
13	41.5	260.9	17.9	4.2	0.1	0.0	5.5	0.0	39.0	13.8
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	2.4	0.9	0.0	9.4	737.0	0.0	140.7	1.2	19.6	0.0
17	100.6	0.0	0.0	0.4	0.0	2.8	0.2	0.0	0.0	0.0
18	93.9	6.3	0.0	0.0	0.0	169.0	6.0	0.4	0.0	0.0
19	0.1	0.0	0.0	0.0	0.4	0.0	0.1	0.0	0.0	0.0
20	4.5	0.0	54.6	59.1	1.3	67.2	127.7	2.4	31.5	2.6
21	226.4	1.0	0.0	67.1	100.0	0.0	28.7	14.2	74.6	0.6
22	8.4	178.7	0.9	146.4	293.5	0.6	1.8	8.7	9.6	2.3
23	0.0	0.5	1.6	214.0	1.9	26.5	0.1	14.9	0.4	0.0
24	14.8	3.0	136.0	12.7	11.7	18.9	2.7	0.9	0.4	47.5
25	7.7	63.2	167.8	141.2	14.2	0.4	17.3	5.2	0.0	44.6
26	1.3	7.2	0.0	0.0	0.0	26.9	14.9	257.3	0.0	1.8
27	60.9	23.5	24.7	39.7	16.6	49.4	24.1	39.8	16.2	4.2
28	0.0	0.0	53.7	0.0	0.0	11.4	0.2	0.0	1.3	0.0
29	84.5	24.1	33.3	1.9	302.0	0.6	0.0	1.6	0.9	0.0
30	19.5	0.0	0.0	0.6	0.0	26.8	76.7	11.3	0.0	266.1
31	49.2	1.2	58.6	14.3	71.2	18.8	135.4	6.4	104.5	10.8
32	0.0	8.7	143.2	0.0	0.0	5.6	4.7	0.0	6.0	0.1
33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
34	21.1	0.0	195.9	0.4	20.1	0.0	84.4	6.1	0.1	5.1
35	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
36	48.2	82.7	0.9	109.1	50.9	1.9	0.3	2.0	4.4	265.6
37	0.0	0.0	0.0	0.1	0.0	0.0	569.3	0.0	0.0	1.6
38	32.3	56.2	6.9	4.4	569.4	27.6	150.5	0.6	4.1	0.8
39	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
40	16.4	79.6	86.5	25.8	39.5	49.2	0.0	16.9	30.6	20.8
41	29.4	22.3	43.9	12.8	4.1	0.7	6.8	0.4	0.1	0.7
42	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.6	128.5
43	98.1	25.3	0.0	0.1	25.0	30.3	7.2	21.1	0.6	10.4
44	0.1	22.4	82.4	88.6	1.6	0.7	0.0	0.7	55.2	177.0
45	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
46	0.0	0.0	37.9	0.1	138.5	0.0	445.9	0.0	0.1	0.0
47	0.0	24.0	0.0	39.2	5.8	0.2	0.5	0.0	0.5	1.0
48	1.6	0.3	28.7	1.1	305.2	87.3	0.0	11.1	89.3	38.8
49	11.3	0.2	50.1	40.9	14.0	61.4	3.8	93.4	0.0	260.8
50	0.2	17.7	8.5	14.6	0.1	0.0	21.9	15.4	4.1	0.3
51	90.1	0.1	0.0	0.0	0.0	207.2	0.0	0.0	0.0	0.0
52	16.0	6.0	7.5	0.5	107.9	11.2	25.6	3.3	88.2	272.7
53	33.2	47.6	20.3	104.9	91.1	0.3	14.7	32.4	7.4	252.9
54	10.4	24.8	9.7	5.9	0.0	0.9	112.5	0.1	0.6	59.2
55	0.0	0.0	13.6	0.1	0.0	0.0	0.0	126.8	0.0	0.0
56	15.8	512.6	7.4	20.1	0.0	19.2	299.0	0.0	1.0	946.5
57	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
58	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
59	0.0	0.0	0.4	25.2	0.0	0.0	0.2	0.0	7.1	0.0
60	25.2	97.1	12.0	86.2	184.3	4.1	51.1	21.9	1.6	0.8
61	67.6	5.3	0.0	0.0	276.9	23.0	0.0	0.5	11.4	86.1
62	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
64	3.2	0.0	83.9	9.4	1.2	137.0	3.4	14.2	2.1	19.9

TABLE 2—*Continued*

MDA no.	% Representation of indicated locus									
	Exo	OmpA	GlyS	TopA	HolA	CadA	ExuR	PcnB	EutC	Nth
65	5.2	219.8	49.0	18.2	82.7	27.7	8.7	173.2	144.1	0.0
66	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
67	3.0	6.9	0.6	6.0	0.0	0.0	0.9	0.8	736.4	1.1
68	73.3	304.5	12.8	41.9	21.4	0.0	28.9	0.8	16.8	65.7
69	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
70	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
71	0.6	164.8	2.8	0.0	41.6	2.2	9.9	2.6	0.6	0.0
72	0.1	1,211.1	68.9	3.6	103.0	2.7	35.5	111.2	0.0	0.1
73	0.1	67.5	50.9	53.5	24.2	2.3	0.4	18.6	0.0	236.5
74	0.0	11.3	0.0	68.1	52.5	19.7	0.0	11.0	0.0	187.7
75	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
76	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
77	175.6	6.2	24.2	26.1	233.7	11.7	37.9	7.5	22.1	23.9
78	209.0	0.0	158.9	0.0	0.0	0.0	0.0	0.0	4.4	0.0
79	14.5	8.4	0.0	90.0	0.0	0.0	22.0	3.0	0.4	82.8
80	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
81	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
82	0.1	141.5	0.1	463.4	0.0	74.1	0.4	0.0	0.0	0.1
83	0.0	152.8	0.0	0.0	0.0	8.0	0.2	0.0	0.0	0.0
84	86.3	0.0	0.5	4.0	29.4	0.0	98.6	175.3	117.2	0.4
Avg	28.1	54.1	22.3	28.8	48.6	15.6	32.1	15.5	20.8	44.9

single cells will result in the loss of some sequences and will also require more effort in preparation and interpretation of DNA libraries constructed from the amplified DNA. We have also not yet determined the degree to which MDA may result in sequence rearrangements and chimeric sequences. In previous work, a DNA library was constructed using DNA amplified by MDA, starting with about 1,000 cells of the bacterium *Xylella fastidiosa*, and sequenced to a depth of approximately sevenfold (3). It may be necessary to sequence to an even greater depth for single cells due to the amplification bias. Finally, it is important to recognize that MDA tends to amplify all of the DNA present in a sample, including contaminating DNAs, multicopy sequences, and plasmids. In general, the goal of whole-genome amplification is to conserve all template sequences with minimal amplification bias. The ability to amplify genomic DNA from individual bacteria could enable many new research strategies for studying nonculturable species, population diversity, and genetic heterogeneity within species.

REFERENCES

1. **Dean, F. B., J. R. Nelson, T. L. Giesler, and R. S. Lasken.** 2001. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res. **11:**1095–1099.
2. **Dean, F. B., S. Hosono, L. Fang, X. Wu, A. F. Faruqi, P. Bray-Ward, Z. Sun, Q. Zong, Y. Du, J. Du, M. Driscoll, W. Song, S. F. Kingsmore, M. Egholm, and R. S. Lasken.** 2002. Comprehensive human genome amplification using multiple displacement amplification. Proc. Natl. Acad. Sci. USA **99:**5261–5266.
3. **Detter, J. C., J. M. Jett, S. M. Lucas, E. Dalin, A. R. Arellano, M. Wang, J. R. Nelson, J. Chapman, Y. Lou, D. Rokhsar, T. L. Hawkins, and P. M. Richardson.** 2002. Isothermal strand-displacement amplification applications for high-throughput genomics. Genomics **80:**691–698.
4. **Handyside, A. H., M. D. Robinson, R. J. Simpson, M. Omar, Marie-Anne Shaw, J. G. Grudzinskas, and A. Rutherford.** 2004. Isothermal whole genome amplification from single and small numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease. Mol. Hum. Reprod. Epub 20 August 2004.
5. **Hosono, S., A. F. Faruqi, F. B. Dean, Y. Du, Z. Sun, X. Wu, J. Du, S. F. Kingsmore, M. Egholm, and R. S. Lasken.** 2003. Unbiased whole-genome amplification directly from clinical samples. Genome Res. **13:**954–964.
6. **Lage, J. M., J. H. Leamon, T. Pejovic, S. Hamann, M. Lacey, D. Dillon, R. Segraves, B. Vossbrinck, A. Gonzalez, D. Pinkel, D. G. Albertson, J. Costa, and P. M. Lizardi.** 2003. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. Genome Res. **13:**294–307.
7. **Lasken, R. S., and M. Egholm.** 2003. Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. Trends Biotechnol. **21:**531–535.