A Genomic Day in the Life of a Clinical Microbiology Laboratory

S. Wesley Long,^{a,b} Dawn Williams,^a Chandni Valson,^{a,b} Concepcion C. Cantu,^{a,b} Patricia Cernoch,^a James M. Musser,^{a,b} Randall J. Olsen^{a,b}

Department of Pathology and Genomic Medicine, The Methodist Hospital,^a and Center for Molecular and Translational Human Infectious Disease Research, The Methodist Hospital Research Institute,^b Houston, Texas, USA

Next-generation sequencing technology is available to many clinical laboratories; however, it is not yet widely used in routine microbiology practice. To demonstrate the feasibility of using whole-genome sequencing in a routine clinical microbiology workflow, we sequenced the genome of every organism isolated in our laboratory for 1 day.

Whole-genome sequencing of microbes is an important tool for infectious disease researchers (1, 2). It has been used to understand the virulence of organisms of uncertain provenance and investigate the molecular basis of host-pathogen interactions (3–6). Whole-genome sequencing has also been used to evaluate genetic relationships among strains recovered in potential outbreaks of *Mycobacterium tuberculosis*, *Staphylococcus aureus*, and other human pathogens (7–9). However, despite tremendous advances in benchtop next-generation sequencing instrumentation, laboratory automation, and bioinformatics tools, whole-genome sequencing of microbes has not yet been widely adopted by clinical microbiology laboratories to identify unknown organisms.

To test the feasibility of integrating whole-genome sequencing into the routine workflow of a clinical laboratory and identify potential impediments to this approach, we sequenced the genome of every organism recovered from cultures in our microbiology laboratory on a single day. In total, 130 samples were collected from 116 patient cultures (Table 1). A sample was defined as a loopful of colonial material (approximately 5 colonies) taken from solid agar. Whenever possible, each organism was sampled individually if multiple organisms were present on a plate. If organisms could not be sampled individually, then the mixture was sampled *in toto*. The study set included 107 samples collected from the aerobic bacteriology bench (Table 1, samples ADL101 to ADL124, ADL201 to ADL229, and ADL301 to ADL354), 9 samples collected from the anaerobic bacteriology bench (samples ADL401 to ADL409), and 14 samples collected from the acid-fast bacilli and mycology bench (samples ADL501 to ADL514).

Bacteria and yeast were collected from the agar plate using a sterile 1- μ l calibrated loop and transferred to a 2-ml tube containing 1 ml Tris-EDTA and 0.1-mm silica spheres (Lysing matrix B; MP Biomedicals, Santa Ana, CA). Fungal isolates were collected in tubes containing 2-mm zirconia beads and 1.6-mm aluminum oxide particles (Lysing matrix J). Cells were lysed using ballistic disintegration (FastPrep96 automated homogenizer; MP Biomedicals), and genomic DNA was extracted using standard methods (DNeasy 96 blood and tissue kit; Qiagen Inc., Valencia, CA). The quantity and quality of genomic DNA were confirmed (Qubit 2.0 fluorometer; Invitrogen, Life Technologies, Grand Island, NY), and dual-indexed sequencing libraries were prepared according to the manufacturer's instructions (NexteraXT DNA sample preparation kit; Illumina Inc., San Diego, CA). The first 96 libraries were sequenced using a 2- by 250-bp paired-end protocol (MiSeq personal sequencer; Illumina). Since yeast and fungi have much larger genomes than bacteria (~3- to 10-fold), these organ-

isms were grouped in a second run containing fewer libraries so that greater fold coverage could be generated. Using the automated on-instrument *de novo* assembly workflow (MiSeq Reporter version 2.0; Illumina) and default software settings, sequencing reads were assembled into contigs with VELVET (10). VELVET takes in the short reads, removes sequencing errors and low-quality sequences, and produces high-quality unique contigs. Each contig was then analyzed using BLAST (www.blast.ncbi.nlm.nih.gov) to identify high-homology matches within the nucleotide collection (NT) database. The NT database was selected because it consists of a broad range of DNA sequences from GenBank, EMBL, DDBJ, PDB, and RefSeq. It is partially nonredundant and updated daily. Since all divisions of the NT database are not curated, contigs were also analyzed with the RefSeq database for comparison. Organism identification by whole-genome sequence analysis was compared to the result generated by standard methods used in our laboratory, including matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry (Biotyper; Bruker Daltonics, Fremont, CA), biochemical phenotyping (BD Phoenix system; BD Diagnostics, Sparks, MD), or ribosomal DNA sequencing by a reference laboratory (ARUP Laboratory, Salt Lake City, UT).

DNA extraction and library preparation were completed in approximately 12 h. Each MiSeq sequencing run took approximately 39 h, and on-instrument *de novo* assembly of contigs took 2 to 4 h. In total, 11.08 GB of sequence data was generated (range, 0.1- to 28.4-fold coverage per organism; mean, 6.3-fold coverage per organism; Table 1). Organism identification by whole-genome sequencing (BLAST results of the assembled contigs) was concordant with conventional methods for 115 samples (88.5%; Table 1), including several mixed-organism samples that were collected and analyzed *in toto*. The remaining 15 samples (11.5%) had contig BLAST results that did not produce high-confidence organism identifications and were designated with “no reliable identification” (NRI). Four NRI results were likely due to insufficient sequence coverage (samples ADL110, ADL115, ADL119,

Received 8 December 2012 Returned for modification 17 December 2012

Accepted 18 January 2013

Published ahead of print 23 January 2013

Address correspondence to Randall J. Olsen, RJolsen@tmhs.org.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.03237-12

TABLE 1 Identification of organisms by whole-genome sequencing^a

Sample no. ^a	Source of specimen ^b	Organism(s) identified by conventional culture (MALDI-TOF, biochemical phenotype, and/or 16S)	Organism(s) identified by whole-genome sequencing of 500-bp sequence reads	Depth of coverage (fold)
ADL101	Blood	<i>Staphylococcus aureus</i>	<i>S. aureus</i>	8.4
ADL102	Blood	<i>S. aureus</i>	<i>S. aureus</i>	7.1
ADL103	Blood	<i>S. aureus</i>	<i>S. aureus</i>	3.2
ADL104	Blood	Coagulase-negative <i>Staphylococcus</i>	<i>Staphylococcus haemolyticus</i>	4.5
ADL105	SST	<i>S. aureus</i>	<i>S. aureus</i>	10.3
ADL106	SST	<i>Pseudomonas aeruginosa</i>	<i>P. aeruginosa</i>	1.2
ADL107	SST	<i>P. aeruginosa</i>	<i>P. aeruginosa</i>	1.3
ADL108	SST	Coagulase-negative <i>Staphylococcus</i>	<i>Staphylococcus epidermidis</i>	6.4
ADL109	SST	Coagulase-negative <i>Staphylococcus</i>	<i>S. epidermidis</i>	9.5
ADL110	Paracentesis fluid	<i>S. aureus</i>	NRI ^g	<0.1
ADL111	SST	<i>P. aeruginosa</i>	<i>P. aeruginosa</i>	1.5
ADL112	URT	<i>S. aureus</i>	<i>S. aureus</i>	4.8
ADL113	URT	<i>P. aeruginosa</i>	<i>P. aeruginosa</i>	4.8
ADL114	SST	Coagulase-negative <i>Staphylococcus</i>	<i>Staphylococcus lugdunensis</i>	28.4
ADL115	Sputum	Group B <i>Streptococcus</i> , <i>P. aeruginosa</i> , and <i>S. aureus</i>	NRI	NA ^o
ADL116	SST	<i>Candida albicans</i>	<i>C. albicans</i>	2.5
ADL117	SST	<i>S. aureus</i>	<i>S. aureus</i>	0.1
ADL118	SST	<i>Enterococcus faecalis</i> , group B <i>Streptococcus</i> , <i>P. aeruginosa</i> , and <i>S. aureus</i>	<i>E. faecalis</i> , <i>Streptococcus agalactiae</i> , <i>P. aeruginosa</i> , and <i>S. aureus</i>	NA ^o
ADL119	SST	Group A <i>Streptococcus</i> , <i>Moraxella</i> species, <i>S. aureus</i> , and diphtheroids	NRI	NA ^p
ADL120	SST	<i>S. aureus</i>	<i>S. aureus</i>	11.1
ADL121	SST	<i>S. aureus</i>	<i>S. aureus</i>	18.0
ADL122	Hardware	<i>K. pneumoniae</i>	<i>K. pneumoniae</i>	3.2
ADL123	SST	<i>E. faecalis</i>	<i>E. faecalis</i>	14.8
ADL124	SST	Coagulase-negative <i>Staphylococcus</i>	<i>S. epidermidis</i>	19.4
ADL201	Blood	<i>K. pneumoniae</i>	<i>K. pneumoniae</i>	6.6
ADL202	Blood	<i>K. pneumoniae</i>	<i>K. pneumoniae</i>	5.1
ADL203	Blood	Coagulase-negative <i>Staphylococcus</i>	<i>S. epidermidis</i>	18.3
ADL204	Blood	Coagulase-negative <i>Staphylococcus</i>	<i>S. haemolyticus</i>	1.0
ADL205	Blood	Coagulase-negative <i>Staphylococcus</i>	<i>S. epidermidis</i>	10.4
ADL206	Blood	<i>S. aureus</i>	<i>S. aureus</i>	18.7
ADL207	URT	<i>Serratia marcescens</i>	<i>Serratia</i> species ^h	NA ^p
ADL208	Sputum	<i>P. aeruginosa</i>	<i>P. aeruginosa</i>	1.7
ADL209	URT	<i>Acinetobacter</i> species	<i>A. baumannii</i>	6.4
ADL210	Blood	<i>S. aureus</i>	<i>S. aureus</i>	19.0
ADL211	URT	<i>Stenotrophomonas maltophilia</i>	<i>S. maltophilia</i>	0.9
ADL212	URT	<i>S. maltophilia</i>	<i>S. maltophilia</i>	0.6
ADL213	SST	<i>P. aeruginosa</i>	<i>P. aeruginosa</i>	4.1
ADL214	URT	<i>Enterobacter cloacae</i> complex	<i>E. cloacae</i> subsp. <i>cloacae</i>	0.7
ADL215	Blood	Coagulase-negative <i>Staphylococcus</i>	<i>Staphylococcus saprophyticus</i>	0.6
ADL216	Bone	<i>S. aureus</i>	<i>S. aureus</i>	8.0
ADL217	Sputum	<i>C. albicans</i> and mixed flora	<i>C. albicans</i> , <i>Streptococcus parasanguinis</i> , and <i>Lactobacillus</i> species	NA ^o
ADL218	SST	<i>Escherichia coli</i>	<i>E. coli</i>	3.5
ADL219	PEG site drainage	Mixed flora with more than 3 organisms	<i>K. pneumoniae</i> and <i>Serratia</i> species	NA ^o
ADL220	Blood	Diphtheroids	NRI	NA ^p
ADL221	SST	<i>E. cloacae</i> complex	<i>E. cloacae</i> subsp. <i>cloacae</i>	0.7
ADL222	SST	<i>E. cloacae</i> complex	<i>E. cloacae</i> subsp. <i>cloacae</i>	0.1
ADL223	SST	<i>E. coli</i>	<i>E. coli</i>	3.9
ADL224	Sputum	<i>S. maltophilia</i>	<i>S. maltophilia</i>	0.1
ADL225	SST	<i>S. aureus</i>	<i>S. aureus</i>	12.6
ADL226	SST	<i>E. coli</i>	<i>E. coli</i>	4.2
ADL227	Pleural fluid	<i>S. aureus</i>	<i>S. aureus</i>	16.3
ADL228	SST	Alpha-hemolytic <i>Streptococcus</i> species	<i>Streptococcus gordonii</i>	0.5
ADL229	SST	Coagulase-negative <i>Staphylococcus</i>	<i>S. epidermidis</i>	11.7
ADL301	Urine	Group B <i>Streptococcus</i> , <i>Lactobacillus</i> species, and mixed flora	<i>Lactobacillus gasseri</i> , <i>S. aureus</i> , and <i>Gardnerella vaginalis</i> ⁱ	NA ^o

(Continued on following page)

TABLE 1 (Continued)

Sample no. ^a	Source of specimen ^b	Organism(s) identified by conventional culture (MALDI-TOF, biochemical phenotype, and/or 16S)	Organism(s) identified by whole-genome sequencing of 500-bp sequence reads	Depth of coverage (fold)
ADL302	Urine	Alpha-hemolytic <i>Streptococcus</i> species	<i>S. aureus</i> , <i>Streptococcus</i> species, and <i>G. vaginalis</i> ⁱ	NA ^o
ADL303	Urine	<i>E. coli</i>	<i>E. coli</i> and <i>S. aureus</i>	4.2
ADL304	Urine	<i>E. coli</i>	<i>E. coli</i>	7.2
ADL305	Urine	<i>E. coli</i>	<i>E. coli</i>	5.6
ADL306	Urine	<i>E. coli</i>	<i>E. coli</i>	2.2
ADL307	Urine	<i>E. coli</i>	<i>E. coli</i>	2.3
ADL308	Urine	<i>E. coli</i>	<i>E. coli</i>	4.6
ADL309	Urine	<i>E. coli</i>	<i>E. coli</i>	4.1
ADL310	Urine	<i>E. coli</i>	<i>E. coli</i>	6.7
ADL311	Urine	<i>E. coli</i>	<i>E. coli</i>	3.8
ADL312	Urine	<i>E. coli</i>	<i>E. coli</i>	4.7
ADL313	Urine	<i>E. coli</i>	<i>E. coli</i>	4.6
ADL314	Urine	<i>E. coli</i>	<i>E. coli</i>	5.3
ADL315	Urine	<i>P. aeruginosa</i>	<i>P. aeruginosa</i>	2.9
ADL316	Urine	<i>E. coli</i>	<i>E. coli</i>	5.3
ADL317	Urine	<i>K. pneumoniae</i>	<i>K. pneumoniae</i>	5.3
ADL318	Urine	<i>P. aeruginosa</i>	<i>P. aeruginosa</i>	12.6
ADL319	Urine	<i>Proteus mirabilis</i>	<i>P. mirabilis</i>	6.4
ADL320	Urine	<i>P. aeruginosa</i>	<i>P. aeruginosa</i>	4.8
ADL321	Urine	<i>P. mirabilis</i>	<i>P. mirabilis</i>	9.3
ADL322	Urine	<i>K. pneumoniae</i>	<i>K. pneumoniae</i>	3.8
ADL323	Stool	No <i>Salmonella</i> , <i>Shigella</i> , or <i>Campylobacter</i> species identified ^{c,d}	<i>E. aerogenes</i>	4.1
ADL324	Stool	No <i>Salmonella</i> , <i>Shigella</i> , or <i>Campylobacter</i> species identified ^{c,d}	<i>E. coli</i>	3.4
ADL325	Stool	No <i>Salmonella</i> , <i>Shigella</i> , or <i>Campylobacter</i> species identified ^{c,d}	<i>P. fluorescens</i>	<0.1
ADL326	Stool	No <i>Salmonella</i> , <i>Shigella</i> , or <i>Campylobacter</i> species identified ^{c,d}	<i>Citrobacter koseri</i>	6.1
ADL327	Stool	No <i>Salmonella</i> , <i>Shigella</i> , or <i>Campylobacter</i> species identified ^{c,d}	<i>E. coli</i>	1.2
ADL328	Stool	No <i>Salmonella</i> , <i>Shigella</i> , or <i>Campylobacter</i> species identified ^{c,d}	<i>C. koseri</i>	5.8
ADL329	Urine	<i>S. aureus</i>	<i>S. aureus</i>	12.9
ADL330	Urine	<i>S. aureus</i>	<i>S. aureus</i>	13.3
ADL331	Urine	<i>S. aureus</i>	<i>S. aureus</i>	16.1
ADL332	Urine	Coagulase-negative <i>Staphylococcus</i>	<i>S. epidermidis</i>	10.9
ADL333	Urine	<i>Enterococcus faecium</i>	<i>E. faecium</i>	8.8
ADL334	Urine	<i>E. faecium</i>	<i>E. faecium</i>	13.2
ADL335	Urine	<i>E. faecalis</i>	<i>E. faecalis</i>	15.1
ADL336	Urine	<i>E. faecalis</i>	<i>E. faecalis</i>	11.7
ADL337	Urine	<i>E. faecalis</i>	<i>E. faecalis</i>	12.6
ADL338	Urine	<i>E. faecalis</i>	<i>E. faecalis</i>	11.7
ADL339	Urine	<i>Candida tropicalis</i>	<i>C. tropicalis</i>	<0.1
ADL340	Urine	<i>Candida glabrata</i>	<i>C. glabrata</i>	0.7
ADL341	Urine	<i>C. glabrata</i>	<i>C. glabrata</i>	<0.1
ADL342	Urine	<i>C. glabrata</i>	<i>C. glabrata</i>	1.6
ADL343	Urine	<i>C. glabrata</i>	<i>C. glabrata</i>	0.2
ADL344	Urine	<i>C. glabrata</i>	<i>C. glabrata</i>	1.0
ADL345	Urine	<i>C. glabrata</i>	<i>C. glabrata</i>	1.6
ADL346	Urine	Group B <i>Streptococcus</i> , <i>Lactobacillus</i> species, and mixed flora	<i>Lactobacillus</i> species, <i>S. gordonii</i> , and <i>S. sanguinis</i>	NA ^o
ADL347	Urine	Coagulase-negative <i>Staphylococcus</i> and mixed flora	<i>S. epidermidis</i> , <i>S. agalactiae</i> , and <i>G. vaginalis</i> ⁱ	NA ^o
ADL348	Urine	Alpha-hemolytic <i>Streptococcus</i> species	<i>Streptococcus</i> species and <i>G. vaginalis</i> ⁱ	NA ^o
ADL349	Urine	Alpha-hemolytic <i>Streptococcus</i> species	<i>Lactobacillus</i> species and <i>G. vaginalis</i> ⁱ	NA ^o
ADL350	Urine	Group B <i>Streptococcus</i>	<i>S. agalactiae</i>	17.0
ADL351	Urine	Mixed flora	<i>L. gasseri</i>	<0.1
ADL352	Urine	Group B <i>Streptococcus</i>	<i>E. faecalis</i> and <i>S. agalactiae</i>	NA ^o
ADL353	Urine	Gram-negative rods (10 ² CFU) ^e and group B <i>Streptococcus</i>	<i>E. coli</i> and <i>S. agalactiae</i>	NA ^o
ADL354	Urine	Group B <i>Streptococcus</i>	<i>S. agalactiae</i>	7.1
ADL401	SST	<i>P. mirabilis</i>	<i>P. mirabilis</i>	4.4
ADL402	Blood	<i>Bacteroides fragilis</i>	<i>B. fragilis</i>	5.9
ADL403	Sinus	<i>Prevotella</i> species	<i>P. melaninogenica</i>	0.1
ADL404	SST	<i>B. fragilis</i>	<i>B. fragilis</i> and <i>P. aeruginosa</i>	NA ^o

(Continued on following page)

TABLE 1 (Continued)

Sample no. ^a	Source of specimen ^b	Organism(s) identified by conventional culture (MALDI-TOF, biochemical phenotype, and/or 16S)	Organism(s) identified by whole-genome sequencing of 500-bp sequence reads	Depth of coverage (fold)
ADL405	Kidney abscess	Alpha-hemolytic <i>Streptococcus</i> species and <i>Prevotella buccae</i>	<i>S. lugdunensis</i> ^j	NA ^o
ADL406	Abdomen abscess	<i>Veillonella</i> species	NRI ^l	NA ^p
ADL407	Kidney transplant	<i>Propionibacterium acnes</i>	<i>P. acnes</i> and <i>C. albicans</i>	NA ^o
ADL408	Pericardial fluid	<i>P. acnes</i>	<i>P. acnes</i> , <i>E. faecalis</i> , and <i>S. agalactiae</i>	NA ^o
ADL409	SST	<i>P. acnes</i>	<i>P. acnes</i>	9.4
ADL501	Bronchial washing	<i>Penicillium</i> species	<i>Penicillium marneffe</i>	NA ^p
ADL502	Bronchial washing	Sterile mycelia	NRI ^l	NA ^p
ADL503	Ear drainage	<i>Mycobacterium chelonae-abscessus</i> complex ^f and sterile mycelia	<i>Mycobacterium abscessus</i> and NRI ^l	<0.1
ADL504	BAL fluid	Organism failed to thrive for identification	NRI ^l	NA ^p
ADL505	Lung tissue	<i>Ramichloridium</i> species	NRI ^{l,m}	NA ^p
ADL506	Bronchial washing	<i>Ramichloridium</i> species	NRI ^{l,m}	NA ^p
ADL507	Sinus contents	<i>Bipolaris</i> species	<i>Bipolaris species</i>	NA ^p
ADL508	BAL fluid	Sterile mycelia	NRI ^l	NA ^p
ADL509	Toenail	<i>Trichophyton rubrum</i>	<i>Trichophyton rubrum</i>	NA ^p
ADL510	BAL fluid	Sterile mycelia	NRI ^l	NA ^p
ADL511	BAL fluid	Sterile mycelia	NRI ^l	NA ^p
ADL512	Sputum	<i>Mycobacterium avium-intracellulare</i> ^f and <i>Nocardia transvalensis</i>	<i>Mycobacterium avium</i> , <i>S. parasanguinis</i> , and NRI ⁿ	0.1
ADL513	Bronchial washing	<i>Acremonium</i> species	NRI ^l	NA ^p
ADL514	Bronchial washing	<i>Mycobacterium abscesses</i>	<i>Mycobacterium abscesses</i>	6.3

^a ADL100, ADL200, and ADL300 strains were collected from the three aerobic bacteriology benches; ADL400 strains were collected from the anaerobic bacteriology bench; and ADL500 strains were collected from the AFB/fungus bench.

^b BAL, bronchial alveolar lavage; SST, skin and soft tissue; URT, upper respiratory tract.

^c The standard operating procedure in our laboratory is to identify all lactose-fermenting Gram-negative bacilli recovered from stool, but only *Salmonella*, *Shigella*, and *Campylobacter* species are reported.

^d The organism identified by MALDI-TOF was concordant with the result by whole-genome sequencing.

^e The standard operating procedure is to identify organisms present at a concentration of $\geq 10^4$ CFU/ml urine.

^f This organism was identified by culture more than 10 days after sampling for the whole-genome sequencing analysis.

^g NRI, no reliable identification; contig BLAST results insufficient to identify an organism.

^h *S. marcescens* was not in the database.

ⁱ *G. vaginalis* was not identified by routine urine culture.

^j *P. buccae* was not in the database.

^k *Veillonella* species were not in the database.

^l Multiple contig BLAST results with low homology to various nonpathogenic fungi.

^m *Ramichloridium* species were not in the database.

ⁿ Only 1 contig BLAST result with homology to *Nocardia* species.

^o NA, not applicable; cannot calculate depth of coverage for samples containing multiple organisms.

^p NA, not applicable; cannot calculate depth of coverage without an organism identification and/or a reference genome in the database.

^q Sequence data were deposited in the NCBI Short Read Archive (SRA) in the Bioproject PRJNA184888.

and ADL220). The samples included clinically relevant organisms such as group A *Streptococcus*, *Staphylococcus aureus*, and *Pseudomonas aeruginosa*. One NRI result occurred with a mixed sample (sample ADL512) containing *Mycobacterium avium-intracellulare*, *Streptococcus parasanguis*, and *Nocardia transvalensis*. Although the first two organisms from sample ADL512 were readily identified by the contig BLAST, only one contig matching *Nocardia* species was assembled. Failure of whole-genome sequencing to identify this organism may be due to inefficient lysis of the

Nocardia cell wall or dilution of its genomic DNA in the complex sample mixture. Further optimization of the sample processing and library preparation protocols may be needed to address these limitations. Ten NRI results were obtained for organisms that did not have a complete reference genome in the NT or RefSeq database. That is, the lack of a comprehensive reference genome database of human pathogens precluded identification of these organisms using next-generation sequencing.

We learned several important lessons from this whole-genome

identification exercise. First, although insufficient sequencing coverage limited the evaluation of 4 of the 130 samples, we discovered that low read counts were not always problematic. For example, the strain with the fewest number of reads (sample ADL341) was easily identified as *Candida glabrata*, despite a very low depth of coverage across its genome (less than 0.1-fold average coverage). Second, we found that whole-genome sequencing was able to deconvolute mixed cultures. Several samples were collected from primary isolation plates containing a mixture of organisms that needed to be subcultured prior to identification by conventional methods. Whole-genome sequencing identified each organism in these mixed samples, confirming its power to quickly resolve complex bacterial populations. Importantly, whole-genome sequencing identified *Mycobacterium* species in two samples 10 days before they were identified by the conventional method (additional incubation until visible colonies could be collected for ribosomal DNA sequencing performed by a reference laboratory; samples ADL503 and ADL512). These two *Mycobacterium*-positive samples were included in the whole-genome sequencing study set because of another organism present on the agar plate at the time of collection (a suspected *Nocardia* species on sample ADL503 and an unknown fungus on sample ADL512). Although the comparatively slower-growing *Mycobacterium* species were not visible at the time of sample collection, a sufficient number of organisms were clearly collected while sampling the other visible organisms to successfully perform genomic analyses. Similarly, genomic DNA from *Gardnerella vaginalis*, a commensal organism of the genital tract and potential urogenital pathogen, was detected in five urine cultures from which this organism was not identified by the conventional method (samples ADL301, ADL302, ADL347, ADL348, and ADL349). There are two plausible explanations for this finding. First, *Gardnerella vaginalis* is a facultative anaerobe that grows slowly under the conditions used for a standard urine culture. In our laboratory, urine cultures are read at 18 to 24 h postinoculation, which may not have allowed sufficient incubation time for this slow-growing organism to form visible colonies in the mixed sample. Alternatively, if present at a concentration less than 10^4 CFU/ml urine, an organism is not routinely worked up in our laboratory. Under either scenario, whole-genome sequencing identified a potential pathogen that was not detected by the routine method. Taken together, these data demonstrate that whole-genome sequencing can accurately characterize mixed cultures and rapidly identify pathogens that are slow growing or otherwise difficult to cultivate.

To have the greatest impact on patient care, clinically actionable results must be generated in a timely fashion. One strategy to reduce the turnaround time of whole-genome sequencing is to use shorter read lengths. Compared to the 500-bp protocol described above, a 50-bp sequencing run can be completed in approximately 5 h. To determine if a 50-bp read length protocol would result in similar organism identifications, we trimmed the 500-bp reads to 50 bp *in silico* and repeated the bioinformatic analysis. Of the 115 samples with one or more organisms identified using the 500-bp read length data, 95 (82.6%) had the same organism(s) identified using the simulated 50-bp read length data (Table 1). Notably, each of the organisms present in 11 mixed samples could not be identified with the 50-bp read length data (ADL217, ADL301, ADL302, ADL346, ADL347, ADL348, ADL352, ADL404, ADL405, ADL407, and ADL408). Similarly, 5 organisms could be identified only to the genus (ADL117, ADL204, ADL228,

ADL512, and ADL514), and 4 samples had insufficient coverage to generate contigs or a reliable identification (ADL222, ADL339, ADL341, and ADL351). The *in silico* data were verified by sequencing the first 96 libraries using a 1- by 50-bp single-end protocol (MiSeq personal sequencer). Taken together, these findings confirm that clinically relevant organisms can be identified using the faster-running 50-bp protocol but suggest that the number of samples multiplexed per run should be reduced accordingly.

An unexpected finding of our study was the notable absence of several organisms in the reference genome database. Nine fungi recovered from patient cultures generated NRI results (Table 1) despite abundant genomic sequence data and contig assembly. Similarly, no complete reference genome was available for *Serratia marcescens*, *Prevotella buccae*, or *Veillonella* species, and only a few genomes have been deposited for *Nocardia* species and *Enterococcus faecium*. BLAST results using the NT and RefSeq databases were similar for bacteria and yeast; however, the RefSeq database contained fewer sequences from fungi and thus failed to provide a reliable identification of these strains.

New technologies, such as MALDI-TOF, are enhancing clinical microbiology workflow, decreasing patient length of stay, and reducing hospital costs (11). We believe that next-generation whole-genome sequencing also has the ability to enhance patient care. Herein, we confirmed that whole-genome sequencing can be integrated into the routine workflow of a clinical microbiology laboratory to identify unknown organisms. With increased fold coverage and additional bioinformatic analysis, these data could also assess virulence gene content and inform molecular epidemiology. As sequencing costs and technical requirements continue to fall, whole-genome sequencing will become an increasingly tractable option for diagnostic laboratories. A key obstacle to its routine use will be the construction of a high-quality reference genome database. Existing collections such as those comprising the NT database have gaps in the number and breadth of deposited whole-genome sequences of human pathogens, particularly medically important fungi. Also, the reference genome databases must be curated, ensuring that submissions contain high-quality sequence data that are accurately annotated. These deficiencies must be addressed before whole-genome sequencing can reach its full potential in the clinical microbiology laboratory. One limitation of genome-based studies is the lack of antibiotic susceptibility testing. For some organisms, the presence or absence of a particular resistance gene may hold predictive value (i.e., the staphylococcal cassette chromosome *mec* [SCC*mec*] that conveys beta-lactam resistance). However, the genetic mechanism underlying resistance phenotypes to many antimicrobial agents is complex or incompletely understood. Finally, run times and reagent costs for large-scale highly multiplexed strategies must be further reduced to optimize this technology for routine patient care. In the meantime, microbial genomics will continue to be applied in niche areas such as molecular epidemiology and outbreak investigations.

REFERENCES

- Olsen RJ, Long SW, Musser JM. 2012. Bacterial genomics in infectious disease and the clinical pathology laboratory. *Arch. Pathol. Lab. Med.* 136:1414–1422.
- Relman DA. 2011. Microbial genomics and infectious diseases. *N. Engl. J. Med.* 365:347–357.
- Beres SB, Carroll RK, Shea PR, Sitkiewicz I, Martinez-Gutierrez JC, Low DE, McGeer A, Willey BM, Green K, Tyrrell GJ, Goldman TD,

- Feldgarden M, Birren BW, Fofanov Y, Boos J, Wheaton WD, Honisch C, Musser JM. 2010. Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proc. Natl. Acad. Sci. U. S. A.* 107:4371–4376.
4. D'Souza CA, Kronstad JW, Taylor G, Warren R, Yuen M, Hu G, Jung WH, Sham A, Kidd SE, Tangen K, Lee N, Zeilmaker T, Sawkins J, McVicker G, Shah S, Gnerre S, Griggs A, Zeng Q, Bartlett K, Li W, Wang X, Heitman J, Stajich JE, Fraser JA, Meyer W, Carter D, Schein J, Krzywinski M, Kwon-Chung KJ, Varma A, Wang J, Brunham R, Fyfe M, Ouellette BF, Siddiqui A, Marra M, Jones S, Holt R, Birren BW, Galagan JE, Cuomo CA. 2011. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. *mBio* 2:e00342–00310.
 5. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio DA, Miller SI, Ramsey BW, Speert DP, Moskowitz SM, Burns JL, Kaul R, Olson MV. 2006. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc. Natl. Acad. Sci. U. S. A.* 103:8487–8492.
 6. Wright AM, Beres SB, Consamus EN, Long SW, Flores AR, Barrios R, Richter GS, Oh SY, Garufi G, Maier H, Drews AL, Stockbauer KE, Cernoch P, Schneewind O, Olsen RJ, Musser JM. 2011. Rapidly progressive, fatal, inhalation anthrax-like infection in a human: case report, pathogen genome sequencing, pathology, and coordinated response. *Arch. Pathol. Lab. Med.* 135:1447–1459.
 7. Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA. 2011. Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc. Natl. Acad. Sci. U. S. A.* 108:20142–20147.
 8. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodtkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJ, Brinkman FS, Brunham RC, Tang P. 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* 364:730–739.
 9. Koser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ. 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N. Engl. J. Med.* 366:2267–2275.
 10. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
 11. Perez KK, Olsen RJ, Musick WL, Cernoch PL, Davis JR, Land GA, Peterson LE, Musser JM. 6 December 2012. Integrating rapid pathogen identification and antimicrobial stewardship significantly decreases hospital costs. *Arch. Pathol. Lab. Med.* [Epub ahead of print.] doi:10.5858/arpa.2012-0651-OA.