

Whole-Genome Analysis of *Exserohilum rostratum* from an Outbreak of Fungal Meningitis and Other Infections

Anastasia P. Litvintseva,^a Steven Hurst,^a Lalitha Gade,^a Michael A. Frace,^b Remy Hilsabeck,^d James M. Schupp,^d John D. Gillece,^d Chandler Roe,^d David Smith,^d Paul Keim,^d Shawn R. Lockhart,^a Shankar Changayil,^c M. Ryan Weil,^c Duncan R. MacCannell,^c Mary E. Brandt,^a David M. Engelthaler^d

Mycotic Diseases Branch,^a Biotechnology Core Facility, Division of Scientific Resources,^b and Office of the Director, National Center for Emerging and Zoonotic Infectious Diseases,^c Centers for Disease Control and Prevention, Atlanta, Georgia, USA; Translational Genomics Research Institute, Flagstaff, Arizona, USA^d

Exserohilum rostratum was the cause of most cases of fungal meningitis and other infections associated with the injection of contaminated methylprednisolone acetate produced by the New England Compounding Center (NECC). Until this outbreak, very few human cases of *Exserohilum* infection had been reported, and very little was known about this dematiaceous fungus, which usually infects plants. Here, we report using whole-genome sequencing (WGS) for the detection of single nucleotide polymorphisms (SNPs) and phylogenetic analysis to investigate the molecular origin of the outbreak using 22 isolates of *E. rostratum* retrieved from 19 case patients with meningitis or epidural/spinal abscesses, 6 isolates from contaminated NECC vials, and 7 isolates unrelated to the outbreak. Our analysis indicates that all 28 isolates associated with the outbreak had nearly identical genomes of 33.8 Mb. A total of 8 SNPs were detected among the outbreak genomes, with no more than 2 SNPs separating any 2 of the 28 genomes. The outbreak genomes were separated from the next most closely related control strain by ~136,000 SNPs. We also observed significant genomic variability among strains unrelated to the outbreak, which may suggest the possibility of cryptic speciation in *E. rostratum*.

Beginning in September, 2012, the United States experienced one of the largest outbreaks of health care-associated infections ever reported. More than 13,000 people were exposed to three lots of contaminated methylprednisolone acetate (MPA) produced by a single compounding pharmacy, and >750 developed fungal infections following their MPA injections (1–3). Infections included meningitis, as well as localized epidural, paraspinous, and peripheral joint infections, complicated by arachnoiditis and abscess formation (4, 5). Although several fungal species were implicated in the outbreak, the vast majority of infections were caused by *Exserohilum rostratum*, a saprobic mold found in soil and plant debris worldwide (2, 6–10). Most species of *Exserohilum* are plant pathogens. *Exserohilum rostratum* and the species formerly known as *Exserohilum longirostratum* and *Exserohilum mcginnisii* are morphologically similar and are the only species known to cause human infections; however, their taxonomic status remains under investigation (11–13).

DNA fingerprinting has been used as an epidemiological tool to answer questions pertaining to strain relatedness (14, 15). However, no methodologies for DNA fingerprinting of *E. rostratum* existed at the time of the outbreak. While whole-genome sequencing (WGS) is a relatively new tool of molecular epidemiology, it has quickly been embraced as a mechanism for inferring phylogenetic relationships among organisms. This has translated to the recent use of whole-genome single nucleotide polymorphism (SNP) typing (whole-genome sequence typing [WGST]) as a mechanism for inferring relatedness of both bacteria (16–19) and fungi (20, 21) during outbreak investigations. One of the benefits of WGST is that prior knowledge of the genome is not necessary to conduct epidemiologic analyses. The relationships between organisms are inferred from the number of SNPs: typically the higher the number of SNPs in shared loci, the more distant the relationship between organisms. While this can be confounded by recombination, statistical bioinformatic analysis helps to discern the nature of such events.

Here, we report the WGST and initial analysis of the whole-genome sequence of *E. rostratum*. We use WGST to address two important epidemiologic questions that arose during this outbreak investigation. First, separate lots of MPA, independently compounded 42 days apart, were injected into patients and then subsequently found to be contaminated with *E. rostratum* (8). Were different lots of MPA contaminated with the same or different fungal strains? Second, the CDC case-patient population was classified into three groups: those with meningitis, those with localized epidural, paraspinous, or peripheral joint infections, and those with both infections (8). Were the differences in clinical manifestation due to different strains of *E. rostratum*? To answer the epidemiologic questions posed above, the genomes of 22 *E. rostratum* isolates from 19 case patients, 3 isolates from each of two contaminated lots of MPA, and 7 unrelated control *E. rostratum* isolates from the CDC collection were sequenced by WGS. These data offer molecular insight into the epidemiology of the outbreak and provide a foundation for subsequent comparative genomic studies to identify the factors that are responsible for the virulence of this pathogen in humans.

MATERIALS AND METHODS

Culture of isolates and preparation of DNA. Isolates of *Exserohilum rostratum* ($n = 22$) were submitted by state public health laboratories from

Received 31 March 2014 Returned for modification 16 April 2014

Accepted 12 June 2014

Published ahead of print 20 June 2014

Editor: G. A. Land

Address correspondence to Anastasia P. Litvintseva, frq8@cdc.gov.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.00936-14

TABLE 1 Sequenced *E. rostratum* strains

Isolate origin	Lab no.	Source of specimen	Date specimen collected (mo/day/yr)	No. (lot ^c) of injection(s) received	Origin of specimen	Depth coverage (×)			
						GAll (CDC)	HiSeq (TGen)	MiSeq (TGen)	PacBio (CDC)
Patient 1_1 ^a	B9826	VA ^b	9/16/2012	1 (06)	CSF	55	205		
Patient 1_2	B9827	VA	9/16/2012		CSF	57	153		
Patient 2_1 ^a	B9846	MI	9/21/2012	1 (06)	CSF	74	ND		
Patient 2_2	B9847	MI	9/21/2012		CSF	57	ND		
Patient 3	B9823	TN	10/4/2012	2 (06)	CSF	118	ND		
Patient 4	B9824	MD	10/11/2012	Unk	CSF	64	ND		
Patient 5	B9860	MI	10/14/2012	6 (Unk)	CSF	67	ND		
Patient 6	B9831	IN	10/5/2012	1 (08)	CSF	48	130		
Patient 7	B9832	IN	10/5/2012	2 (1 08, 1 Unk)	CSF	71	191		
Patient 8	B9833	IN	10/5/2012	2 (1 05, 1 08)	CSF	45	125		
Patient 9	B9834	TN	10/4/2012	3 (2 06, 1 08)	Brain stem	21	153		
Patient 10	B9835	TN	10/2/2012	3 (2 06, 1 Unk)	CSF	37	137		
Patient 11	B9862	TN	10/18/2012	1 (08)	CSF	61	123		
Patient 12	B9839	VA	10/4/2012	2 (06)	CSF	63	148		
Patient 13_1 ^a	B9842	VA	10/4/2012	1 (06)	CSF	45	144		
Patient 13_2	B9843	VA	10/4/2012		CSF	77	129		
Patient 14	B9845	MI	10/5/2012	1 (06)	CSF	93	259		
Patient 15	B9877	MI	11/8/2012	1 (06)	Lumbar wound	N/D	114		
Patient 16	B9879	MI	11/8/2012	1 (06)	Epidural abscess	N/D	85		
Patient 17	B9892	MI	10/31/2012	1 (06)	Spinal abscess	N/D	103		
Patient 18	B9893	MI	11/6/2012	1 (06)	Spinal abscess	N/D	115		
Patient 19	B9908	MI	11/30/2012	3 (2 06, 1 Unk)	Spinal abscess	N/D	ND	32	91
Vial 1	B10011	Sample 744517 sub 3D	10/18/2012 ^d		NECC vial, MPA 06292012@26	59	143	28	131
Vial 2	B10016	Sample 756981 sub 7	10/18/2012		NECC vial, MPA 06292012@26	72	111		
Vial 3	B9935	Sample 693965 sub 1-1	10/12/2012		NECC vial, MPA 08102012@51	86	131		
Vial 4	B9970	Sample 693966 sub 3-I	10/12/2012		NECC vial, MPA 08102012@51	76	119		
Vial 5	B9960	Sample 760871 sub 4	10/12/2012		NECC vial, MPA 08102012@51	79	119		
Vial 6	B10014	Sample 756981 sub 12	10/18/2012		NECC vial, MPA 06292012@26	39	147		
<i>E. rostratum</i>	B6094	UTHSC 01-109	2001		Patient isolate, Texas	69	159		
<i>E. rostratum</i>	B6096	UTHSC 00-1435	2001		Patient isolate, Texas	62	106		
<i>E. rostratum</i>	B6171		2001		Environmental isolate, Alabama	62	109		
<i>E. rostratum</i>	B6177		2001		Environmental isolate, Alabama	88	135		
<i>E. rostratum</i>	B6207		2001		Environmental isolate, Alabama	202	102		
<i>E. rostratum</i> ^e	B6272	NCMH 1889	2002		Patient isolate, France (11)	50	122		
<i>E. rostratum</i> ^f	B6284	NCMH 2445, CBS 325.87	2002		Patient isolate, Arizona (11)	51	123		

^a Duplicate isolates from the same patient were obtained and analyzed.

^b State in which the patient received MPA injection(s).

^c 06 refers to MPA lot 06292012@26, 05 refers to MPA lot 05212012@68, 08 refers to MPA lot 08102012@51. Unk, lot injected is unknown.

^d Date listed for vials is the date when the isolate from that vial was received at the CDC.

^e Formerly *E. longirostratum*.

^f Formerly *E. mcginnisii*.

cultures of specimens collected from 19 case patients who were exposed to one of the implicated lots of preservative-free MPA produced by the New England Compounding Center (NECC) after May 21, 2012 (22) and met the CDC case definition as described previously (2) (Table 1). Patients were from Virginia (3), Michigan (8), Tennessee (4), Maryland (1), and Indiana (3). Specimens were cerebrospinal fluid (CSF) (16 specimens), brain stem tissue (1 specimen), spinal/epidural abscess tissue (4 specimens), and lumbar wound tissue (1 specimen). One isolate per patient

was submitted except in three cases where duplicate isolates from the same CSF culture were submitted to the CDC, and each isolate was sequenced separately.

Additional isolates ($n = 6$) (Table 1) for identification were sent to the CDC by the Food and Drug Administration (FDA). These isolates were recovered from MPA vials from lots 06292012@26 and 08102012@51 that were cultured by the FDA (2, 8). Since an *Exserohilum* isolate was never recovered from lot 05212012@68, this lot was not included in this analysis.

Seven unrelated *E. rostratum* isolates from the culture collection of the Centers for Disease Control (CDC) were used as controls. Some of these isolates were previously described (23). All isolates were identified to species by microscopic examination and DNA sequencing as previously described (8). The isolates were stored at -70°C until used. The isolates were cultured on Sabouraud dextrose agar, and DNA was extracted using a Qiagen DNeasy blood and tissue kit (Qiagen, Gaithersburg, MD), following the manufacturer's recommendations.

DNA sequencing library preparation. The Illumina GAIIX and Pacific Biosciences (PacBio) libraries were prepared at the CDC. For WGS paired-end sequencing on the Illumina GAIIX DNA sequencer, genomic DNAs were prepared using standard library construction protocols and reagents from Illumina Inc. (San Diego, CA). Approximately 1 μg of DNA was sheared using a Covaris S2 sonicator (Covaris Inc., Woburn, MA) to a mean size of 350 bp. The DNA sequencing libraries were then prepared using Illumina TruSeq chemistry and size selected using double AMPure (Beckman Coulter) selection. Paired-end flow cells underwent cluster formation using an Illumina cBot, followed by $100 \times 100\text{-bp}$ cycle sequencing using SBS cycle sequencing V5 kits. The sequence data were processed using CASAVA (v1.8.2) into paired FASTQ paired-end reads for downstream assembly and analysis.

For Pacific Biosciences sequencing, fungal genomic DNAs were sheared to 10 kbp utilizing g-Tubes (Covaris Inc.). Sheared DNAs were then used to generate large SMRTbell libraries with a Pacific Biosciences (Menlo Park, CA) DNA template preparation kit using their standard 10-kb library protocol. Finished libraries were bound to a proprietary XL polymerase and sequenced on a PacBio RS sequencer for 120-min movies using C2 chemistry and V2 SMRT cells. Sequence reads were filtered and assembled *de novo* utilizing the PacBio hierarchical genome assembly process (HGAP1) or a modified Celera Assembler (PacBioToCA 2).

At the Translational Genomics Research Institute (TGen), additional samples were prepared for paired-end sequencing using the Kapa Biosystems (Woburn, MA) library preparation kit protocol with an 8-bp index modification. One to 2 μg double-stranded DNA (dsDNA) was sheared to an average size of 600 bp using SonicMan (Brooks Automation, Spokane, WA), which was then prepared using the Kapa Illumina paired-end library preparation protocol as described by the manufacturer. Modified oligonucleotides (Integrated DNA Technologies, Coralville, IA) with 8-bp indexing capability (24) were substituted at the appropriate step. Prior to sequencing, the libraries were quantified with quantitative PCR (qPCR) on an ABI 7900HT system (Life Technologies Corporation, Carlsbad, CA) using the Kapa Library quantification kit.

Parallel sequencing and data analysis. In order to ensure accurate sequencing and SNP detection, 25 of 35 samples were sequenced in two separate laboratories (CDC Biotechnology Core Facility and TGen, Flagstaff, AZ) using different sequencing platforms and analyzed using multiple bioinformatic methods.

Twenty-five of the CDC-prepared libraries were sequenced both at the CDC using the Illumina GAIIX platform and at TGen using the Illumina HiSeq 2000 platform, following the recommendations of the manufacturer. Five libraries were sequenced only at CDC. Four libraries were sequenced by TGen only using the Illumina HiSeq 2000 platform. Two libraries were sequenced at TGen using the MiSeq platform; one of these was also sequenced on the HiSeq platform. In addition, to improve assembly metrics, the two MiSeq strains were also sequenced at the CDC using the PacBio RSII platform (Table 1).

Read data from the 35 strains were deposited at the NCBI in the short-read archive under BioProject accession number PRJNA245241.

Genome assembly and SNP analysis. Bioinformatic analysis and clustering from the Illumina data at the CDC were conducted using a two-pass approach. *De novo* assemblies of all isolates were performed using Genomics Workbench 6.5 (CLC bio, Aarhus, Denmark) and assessed for quality and assembly metrics. For *de novo* assembly, the following parameters were used: minimum contig length of 300 and similarity and length fractions both 0.7. In the absence of an established reference sequence for

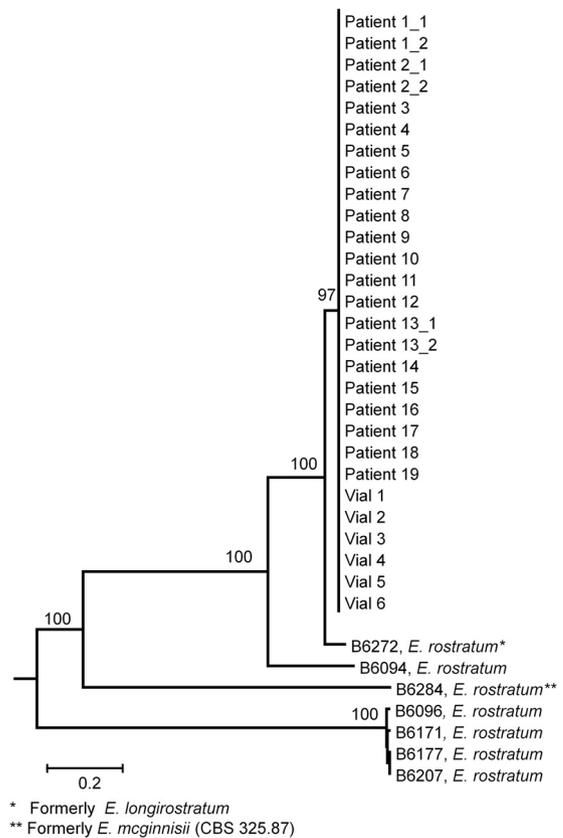


FIG 1 Genetic relationships among 28 outbreak-related and 7 control strains inferred using the neighbor-joining method based on 1,239,535 variable SNP positions. The optimal tree with the branch length sum of 3.75039049 is drawn to scale and rooted at midpoint, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Gaps and missing data are excluded. Numbers above the branches show bootstrap values based on 500 replicates. A phylogenetic tree with the identical topology was obtained using MP analysis (data not shown).

E. rostratum, an initial k-mer-based maximum likelihood tree was generated using kSNP (v2), with default parameters and a k-mer size of 25 bp (25). This first-pass clustering was used to assess the overall diversity of sample and comparator isolates and to determine suitable reference sequences for high-quality, alignment-mapped SNP analyses. The isolate from patient 1_1 (B9826) was selected as an internal CDC reference sequence (Fig. 1), based on the assembly quality and the epidemiologic context of the isolate (this strain was isolated from one of the first confirmed cases) (Table 1). Quality-checked and trimmed reads from all other isolates were mapped to the concatenated reference sequence using Genomics Workbench, and SNPs were called from the resulting alignments with a quality-based variant caller, using Phred quality thresholds of Q20 (99%) at the central position and Q15 for five flanking bases (26). The specific parameters for CLC bio reference mapping were: similarity, 0.7; length fraction, 0.7; insertion cost, 3; deletion cost, 3; mismatch cost, 2; global alignment, no; override paired distance, yes; and maximum distance, 400. Putative SNP calls for all strains were imported into a relational database for postprocessing analyses. During postprocessing analysis, a locus was added, if one among all the samples had a SNP with 100% frequency and at least $25\times$ coverage. For other samples, a minimum of 70% unbiased read-level consensus at $40\times$ coverage was required. If a sample did not meet either of these two criteria at that particular locus, then that locus was not considered a SNP for that sample. A matrix of final SNP alleles based on 1,239,535 variable positions for each isolate was exported for WGST analysis.

TABLE 2 Summary statistics for reference assemblies

Strain	Origin	No. of contigs	Total length (Mb)	N50	GC content (%)
B6177	Control	1,121	35.5	110,956	50.8
Patient 19	Outbreak	256	33.8	256,382	51.1

TGen utilized a reference-based method in which all sequences were compared against an assembled genome of a single isolate. For preliminary phylogenetic analyses, since no *Exserohilum* reference genomes were previously published, TGen assembled B9908 (patient 19) from 250-bp paired-end reads that were obtained on the Illumina MiSeq sequencer. Specifically, FLASH was used to match overlapping MiSeq pairs using the default parameters of the program and the following setting: min-overlap = 25 (27). MIRA was used to generate an assembly from the overlapping and unmerged pairs using the following settings: job = genome, *de novo*, accurate (28). Any contigs that were <500 bp in length and/or were assembled with less than two-thirds of the average coverage were removed from the assembly.

For the final WGST analysis, TGen utilized the reference generated by the CDC from B9826 (patient 1_1) for the kSNP analyses to ensure consistency. The paired reads corresponding to all of the isolates were aligned to the reference genome using the commercially available aligner, Novoalign (Novocraft). Thereafter, the alignment file was processed using the Genome Analysis Toolkit to identify all callable loci. Any position that was not present at a minimum of 10 reads and/or was at a proportion of <90% of all bases at that position was removed prior to downstream analysis. The positions passing these filters were assigned as references or variants. Duplicated regions were identified by a self-comparison of the reference genome using MUMmer (v3.22) (29), and calls within these repetitive regions were removed. A matrix was assembled for positions in which all isolates at that loci contained a confident call and at least one of the positions was a variant. This final matrix was used for inferring phylogenetic relationships using the maximum parsimony (MP) and neighbor-joining (NJ) methods in molecular evolutionary genetic analysis (MEGA) (30). Gaps and missing data were excluded. Bootstrap values for the NJ phylogenetic tree were obtained using 500 replications.

Discrepant SNP analysis. Because SNP calling took place at two separate institutions using different methodologies, discrepant SNPs were independently verified by Sanger DNA sequencing. The targeted oligonucleotide primers were designed upstream and downstream of each targeted SNP so that the final amplicon size was approximately 500 nucleotides with the targeted SNP roughly in the middle, and resultant amplicons were sequenced using Sanger methodology (20).

Ploidy analysis. The alignment sequence read depths of the coverage of the reference genome Illumina sequence data were compared across the assembled contigs to detect possible aneuploidy. Genome ploidy was examined by comparing the variant allele frequencies across the assembled genome (31).

RESULTS

The average sequencing depth obtained for the outbreak strains was 135×, which provided 99.9% coverage across the genomes (Table 1). The preferred assembly of the outbreak strains using MiSeq data produced 256 contigs with N50 (32) of 256,382 bp and the draft genome size of 33.8 Mb (Table 2). Since only HiSeq and GAI data were available for the control strains, on average 91% of the genomes of the control strains were sequenced at a depth of ≥10×, resulting in 1,124 contigs with N50 of 110,956 bp and the draft genome size of 35.5 Mb (Table 2). The comparative depth of coverage and the variant allele frequency across the assembled contigs provided no evidence of aneuploidy (fold coverage varia-

tion across contigs) or polyploidy (variant allele frequencies other than >0.9 or near zero), indicating that the outbreak genome is likely haploid.

Sequencing and SNP analysis of reads from both the Illumina and PacBio instruments produced nearly identical results. Within the set of isolates associated with the outbreak, only seven non-parsimonious concordant SNPs were identified. An additional SNP identified by the Illumina but not the PacBio sequencing was subsequently confirmed by the Sanger sequencing, for a total of eight SNPs detected among the 28 outbreak strains, from 33.8 million bases in the genome. Conversely, the comparison between the outbreak and unrelated control strains identified 1.2 million SNPs, and the genome of the closest unrelated control B6272 was differentiated from the genomes of the outbreak strains by 136,000 SNPs (Fig. 1).

The WGST analysis showed that most outbreak isolates had identical genomes, with no more than two SNPs separating any 2 genomes. Specifically, 19 genomes were identical, representing strains from CSF and abscess tissue specimens from patients in Indiana, Michigan, Tennessee, and Virginia as well as the 06292012@26 and 08102012@51 MPA vials (patients 1_2, 2 to 6, 9, 11 to 13, 15, 16, 18, and 19 and vials 1, 2, and 5); eight strains had single nonparsimonious SNPs (patients 1_1, 7, 8, 10, 14, and 17 and vials 4 and 6), and one strain from vial 3 had two nonparsimonious SNPs. There was a single SNP difference between the genomes of two duplicate strains from patient 1, while genomes of two other sets of duplicate strains were identical. Conversely, 20,043 SNPs were identified when we compared the two most closely related control strains, B6177 and B6171, collected from an outbreak in Alabama (Fig. 1). No differences in the genomes were found when we compared the two lots of MPA, the geographic location of the patient, or the syndrome of the patient.

DISCUSSION

To investigate the relatedness of *E. rostratum* strains associated with the outbreak, we sequenced the genomes of 22 isolates from 19 case patients, 6 isolates from contaminated vials of MPA, and 7 isolates of *E. rostratum* unrelated to this outbreak. To ensure the accuracy of SNP detection, isolates were sequenced at two different laboratories using four different platforms and multiple comparative tools. The resulting data were analyzed separately and in combination, resulting in > 100× sequencing depth for most of the strains and 99.9% coverage for the 33.8-Mb genome. Only eight individual nonparsimonious SNPs were identified among the 28 strains associated with the outbreak, and no more than two SNPs differentiated any pair of strains isolated from the case patients and NECC vials. Conversely, approximately 1.2 million SNPs were observed among the comparative isolate genomes not linked to the outbreak. The most closely related control *E. rostratum* isolate, B6272 (formerly *E. longirostratum*), differed from the outbreak cluster by approximately 136,000 SNPs, suggesting divergence between the outbreak and other available *E. rostratum* strains. Overall, genetic diversity among the control strains of *E. rostratum* was comparable to diversities observed in other groups of fungi with sequenced genomes. For example, approximately 300,000 fixed SNPs were observed between *Coccidioides immitis* and *Coccidioides posadasii* genomes (33), approximately 135,000 SNPs were identified among wild-type isolates of *Neurospora crassa* from geographically different populations (34), and ap-

proximately 28,000 SNPs were identified among strains of *Apophysomyces trapeziformis* (21).

Overall, our data demonstrate that the outbreak strains are highly clonal and strongly suggest their origination from a single common source. Furthermore, isolates from the two different lots of MPA were also indistinguishable from each other and from the isolates from the case patients, suggesting that the source of the contamination was persistent. However, the extent of diversity in the natural population of *E. rostratum* in New England requires further evaluation.

The most significant limitation of this study is that we did not have any environmental or unrelated patient isolates of *E. rostratum* from the same geographic area and/or from a time frame similar to that of the MPA-associated isolates. Only a limited number of *E. rostratum* strains were available in culture collections, and to our knowledge, all of them were isolated from specimens from the southern United States or other countries. The population genetics of *E. rostratum* is completely unknown at this time, and the number of SNPs which can be expected between any two strains in the United States was previously unknown. Our preliminary results based on sequencing of seven isolates not related to the outbreak suggest that diversity is readily identifiable in genome analysis of background isolates. Specifically, the two most closely related strains, B6177 and B6171, both recovered from an outbreak in Alabama (23), are separated from each other by 20,043 SNPs, which is considerably higher than the 0 to 2 SNPs separating any two of the outbreak strains. However, more data are needed to assess the genetic diversity among other strains of *E. rostratum* in the environment.

Our results may also suggest the possibility of cryptic speciation in *E. rostratum*; however, more research is necessary to investigate the genetic relationships among different clades. Phylogenetic analysis using WGS indicates three genetically distinct groups, which are separated by hundreds of thousands of SNPs, and may possibly represent distinct species. Group I includes all of the outbreak strains as well as B6272 and B6094, group II includes a single isolate B6284, and group III includes four isolates from Texas and Alabama (Fig. 1). Notably, all control isolates were morphologically indistinguishable from one another and had identical internal transcribed spacer 1 (ITS1) and ITS2 ribosomal DNA (rDNA) sequences (data not shown), suggesting that rDNA sequences and morphology are not sufficient for differentiating the distinct *Exserohilum* subpopulations identified here. Along this line, we note that B6272, the isolate separated from the outbreak cluster by ~136,000 SNPs, was originally identified as *Exserohilum longirostratum* (11), and we previously reported that the outbreak isolates had conidia with 6 to 12 septa, somewhat resembling the conidia of *E. longirostratum* (8). This species was later placed into conspecificity with *E. rostratum* on the basis of multi-locus sequence analysis (35). A further phylogenetic study of these organisms at the whole-genome level as well as detailed morphological and physiological analyses of different genetic groups will be needed to determine potential species boundaries. The genome of another closely related species, *Exserohilum turcicum*, has recently been released (36), but the genome size difference alone (43 Mb versus 33.8 Mb) shows just how much diversity lies within this group of organisms and indicates that more research is needed to understand the phylogeny of *Exserohilum*.

This study confirmed the utility of WGST for the epidemiologic investigation of fungal infections caused by species with no

previously established typing system. This is especially important for outbreaks of fungal infections where rapid laboratory methods to establish the relatedness of different strains may be the critical factor in determining whether a cluster of rare mold infections represent a true outbreak or a just an unusual coincidence. In the past 18 months, the CDC has investigated several potential outbreaks of rare molds that did not have established DNA typing methods, including an outbreak of *Bipolaris hawaiiensis* associated with compounded triamcinolone for intraocular administration (37), *Curvularia (Bipolaris) spicifera* in cardiac surgical site infections (CDC unpublished data), and *Apophysomyces trapeziformis* soft tissue infections occurring after a natural disaster (21). The determination of whether these isolates were related relied on painstaking epidemiologic studies, as typing methods were not initially available.

Importantly, our data demonstrate that in mold outbreaks involving contaminated medications, a high degree of clonality may be expected. This is in contrast to outbreaks due to construction or local factors, where a lower degree of relatedness may occur. These findings will be important for investigators during future outbreaks involving molds and highlight the importance of WGS in these settings.

Similarly, WGS might also be a powerful tool in the investigation of potentially donor-derived transplant-related fungal infections. We earlier used WGS to investigate the etiology of three *Coccidioides immitis* isolates recovered from organ recipients after transplant, when transmission from a common organ donor was suspected but no isolate was recovered from the donor (20). Because of the ubiquitous nature of many fungi in the environment, it is often difficult, even when a donor and a recipient have the same culture-positive fungal infection, to determine with certainty that it was transmitted via organ transplant if no typing system exists. Knowing if an infection is donor-derived is time sensitive and has important implications for both the testing and treatment of organ recipients as well as surveillance to understand how often these infections occur. As medical science continues to advance, the population of persons susceptible to fungal infections due to critical illness or immune defects will continue to expand. With this expansion, we will likely see more clusters and outbreaks of emerging fungal infections, many agents of which lack existing typing methods, making it important that available laboratory tools such as WGS be utilized to their full extent in order to detect outbreaks quickly and to identify outbreak sources when possible.

This study demonstrates the usefulness of WGS for epidemiologic outbreak investigations and for investigations of fungal species with no previously established typing system. The importance of a culture collection of isolates for use as phylogenetic controls during an investigation is apparent. One of the hallmarks of WGS is that the data can be archived and stored in publicly accessible databases. As more fungal genomes become available in the public domain, they will serve as reference genomes during other investigations, enabling faster and more accurate genome assemblies and providing controls for population genetic analysis. Here, the data confirmed that isolates from the case patients could not be distinguished from isolates from the MPA vials, but were unrelated to *E. rostratum* isolates from an archival collection.

ACKNOWLEDGMENTS

We thank all the participants in the Multistate Fungal Infection Outbreak Response Team, especially state and local health department officials, clinicians, and health care professionals, who provided clinical samples and data for this investigation; all members of the CDC Mycotic Diseases Branch and the CDC emergency response volunteers for their help with receiving, processing, and cataloging clinical samples, especially Christina Scheel, Cau Pham, Mark Lindsley, Nina Grossman, Joyce Peterson, Carol Bolden, Ngoc Le, Shirley McClinton, Randy Kuykendall, Steve Hurst, Kizee Etienne, Noelle Bessette, Grant Williams, Paris Paul, Sarah Yi, Laura Dickmeyer and Jessica Halpin; Michael Palmieri of the FDA for providing viral isolates; Deanna Sutton, Michael Rinaldi, and Michael McGinnis for providing control isolates for this study; and Rachel Smith, Benjamin Park, and Tom Chiller for critically reading the manuscript.

The use of product names in this article does not imply their endorsement by the US Department of Health and Human Services. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the CDC.

REFERENCES

- Centers for Disease Control and Prevention. 2012. Multistate outbreak of fungal infection associated with injection of methylprednisolone acetate solution from a single compounding pharmacy—United States, 2012. *MMWR Morb. Mortal. Wkly. Rep.* 61:839–842.
- Smith RM, Schaefer MK, Kainer MA, Wise M, Finks J, Duwve J, Fontaine E, Chu A, Carothers B, Reilly A, Fiedler J, Wiese AD, Feaster C, Gibson L, Griese S, Purfield A, Cleveland AA, Benedict K, Harris JR, Brandt ME, Blau D, Jernigan J, Weber JT, Park BJ, Multistate Fungal Infection Outbreak Response T. 2013. Fungal infections associated with contaminated methylprednisolone injections. *N. Engl. J. Med.* 369:1598–1609. <http://dx.doi.org/10.1056/NEJMoa1213978>.
- Kainer MA, Reagan DR, Nguyen DB, Wiese AD, Wise ME, Ward J, Park BJ, Kanago ML, Baumblatt J, Schaefer MK, Berger BE, Marder EP, Min JY, Dunn JR, Smith RM, Dreyzehner J, Jones TF, Tennessee Fungal Meningitis Investigation T. 2012. Fungal infections associated with contaminated methylprednisolone in Tennessee. *N. Engl. J. Med.* 367:2194–2203. <http://dx.doi.org/10.1056/NEJMoa1212972>.
- Centers for Disease Control and Prevention. 2013. Spinal and paraspinal infections associated with contaminated methylprednisolone acetate injections—Michigan, 2012–2013. *MMWR Morb. Mortal. Wkly. Rep.* 62:377–381.
- Chiller TM, Roy M, Nguyen D, Guh A, Malani AN, Latham R, Peglow S, Kerkering T, Kaufman D, McFadden J, Collins J, Kainer M, Duwve J, Trump D, Blackmore C, Tan C, Cleveland AA, MacCannell T, Muehlenbachs A, Zaki SR, Brandt ME, Jernigan JA, Multistate Fungal Infection Clinical Investigation T. 2013. Clinical findings for fungal infections caused by methylprednisolone injections. *N. Engl. J. Med.* 369:1610–1619. <http://dx.doi.org/10.1056/NEJMoa1304879>.
- Pettit AC, Kropski JA, Castilho JL, Schmitz JE, Rauch CA, Mobley BC, Wang XJ, Spires SS, Pugh ME. 2012. The index case for the fungal meningitis outbreak in the United States. *N. Engl. J. Med.* 367:2119–2125. <http://dx.doi.org/10.1056/NEJMoa1212292>.
- Gade L, Scheel CM, Pham CD, Lindsley MD, Iqbal N, Cleveland AA, Whitney AM, Lockhart SR, Brandt ME, Litvintseva AP. 2013. Detection of fungal DNA in human body fluids and tissues during a multistate outbreak of fungal meningitis and other infections. *Eukaryot. Cell* 12:677–683. <http://dx.doi.org/10.1128/EC.00046-13>.
- Lockhart SR, Pham CD, Gade L, Iqbal N, Scheel CM, Cleveland AA, Whitney AM, Noble-Wang J, Chiller TM, Park BJ, Litvintseva AP, Brandt ME. 2013. Preliminary laboratory report of fungal infections associated with contaminated methylprednisolone injections. *J. Clin. Microbiol.* 51:2654–2661. <http://dx.doi.org/10.1128/JCM.01000-13>.
- Litvintseva AP, Lindsley MD, Gade L, Smith R, Chiller T, Lyons JL, Thakur KT, Zhang SX, Grgurich DE, Kerkering TM, Brandt ME, Park BJ. 2014. Utility of (1–3)-beta-D-glucan testing for diagnostics and monitoring response to treatment during the multistate outbreak of fungal meningitis and other infections. *Clin. Infect. Dis.* 58:622–630. <http://dx.doi.org/10.1093/cid/cit808>.
- Ritter JM, Muehlenbachs A, Blau DM, Paddock CD, Shieh WJ, Drew CP, Batten BC, Bartlett JH, Metcalfe MG, Pham CD, Lockhart SR, Patel M, Liu L, Jones TL, Greer PW, Montague JL, White E, Rollin DC, Seales C, Stewart D, Deming MV, Brandt ME, Zaki SR, *Exserohilum* Infections Working Group. 2013. *Exserohilum* infections associated with contaminated steroid injections: a clinicopathologic review of 40 cases. *Am. J. Pathol.* 183:881–892. <http://dx.doi.org/10.1016/j.ajpath.2013.05.007>.
- McGinnis MR, Rinaldi MG, Winn RE. 1986. Emerging agents of phaeo-phycomycosis: pathogenic species of *Bipolaris* and *Exserohilum*. *J. Clin. Microbiol.* 24:250–259.
- Adler A, Yaniv I, Samra Z, Yacobovich J, Fisher S, Avrahami G, Levy I. 2006. *Exserohilum*: an emerging human pathogen. *Eur. J. Clin. Microbiol. Infect. Dis.* 25:247–253. <http://dx.doi.org/10.1007/s10096-006-0093-3>.
- Gauthier GM, Keller NP. 2013. Crossover fungal pathogens: the biology and pathogenesis of fungi capable of crossing kingdoms to infect plants and humans. *Fungal Genet. Biol.* 61:146–157. <http://dx.doi.org/10.1016/j.fgb.2013.08.016>.
- Balajee SA, Sigler L, Brandt ME. 2007. DNA and the classical way: identification of medically important molds in the 21st century. *Med. Mycol.* 45:475–490. <http://dx.doi.org/10.1080/13693780701449425>.
- McEwen J, Taylor J, Carter DA, Xu J, Felipes M, Vilgalys R, Mitchell TG, Kasuga T, White T, Bui T, Soares CM. 2000. Molecular typing of pathogenic fungi. *Med. Mycol.* 38(Suppl):189–197. <http://dx.doi.org/10.1080/mmy.38.s1.189.197>.
- Lewis T, Loman NJ, Bingle L, Jumaa P, Weinstock GM, Mortiboy D, Pallen MJ. 2010. High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. *J. Hosp. Infect.* 75:37–41. <http://dx.doi.org/10.1016/j.jhin.2010.01.012>.
- Reuter S, Ellington MJ, Cartwright EJ, Koser CU, Torok ME, Gouliouris T, Harris SR, Brown NM, Holden MT, Quail M, Parkhill J, Smith GP, Bentley SD, Peacock SJ. 2013. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. *JAMA Intern. Med.* 173:1397–1404. <http://dx.doi.org/10.1001/jamainternmed.2013.7734>.
- Harris SR, Cartwright EJ, Torok ME, Holden MT, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J, Peacock SJ. 2013. Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect. Dis.* 13:130–136. [http://dx.doi.org/10.1016/S1473-3099\(12\)70268-2](http://dx.doi.org/10.1016/S1473-3099(12)70268-2).
- Hendriksen RS, Price LB, Schupp JM, Gillette JD, Kaas RS, Engelthaler DM, Bortolaia V, Pearson T, Waters AE, Upadhyay BP, Shrestha SD, Adhikari S, Shakya G, Keim PS, Aarestrup EM. 2011. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *mBio.* 2:e00157-11. <http://dx.doi.org/10.1128/mBio.00157-11>.
- Engelthaler DM, Chiller T, Schupp JA, Colvin J, Beckstrom-Sternberg SM, Driebe EM, Moses T, Tembe W, Sinari S, Beckstrom-Sternberg JS, Christoforides A, Pearson JV, Carpten J, Keim P, Peterson A, Terashita D, Balajee SA. 2011. Next-generation sequencing of *Coccidioides immitis* isolated during cluster investigation. *Emerg. Infect. Dis.* 17:227–232. <http://dx.doi.org/10.3201/eid1702.100620>.
- Etienne KA, Gillette J, Hilsabeck R, Schupp JM, Colman R, Lockhart SR, Gade L, Thompson EH, Sutton DA, Neblett-Fanfair R, Park BJ, Tura-belidze G, Keim P, Brandt ME, Deak E, Engelthaler DM. 2012. Whole genome sequence typing to investigate the *Apophysomyces* outbreak following a tornado in Joplin, Missouri, 2011. *PLoS One* 7:e49989. <http://dx.doi.org/10.1371/journal.pone.0049989>.
- Centers for Disease Control and Prevention. 2012. Update: additional contamination identified in medical products from New England Compounding Center, Health Alert Network, vol HAN-0337. CDC, Atlanta, GA.
- Kainer MA, Keshavarz H, Jensen BJ, Arduino MJ, Brandt ME, Padhye AA, Jarvis WR, Archibald LK. 2005. Saline-filled breast implant contamination with *Curvularia* species among women who underwent cosmetic breast augmentation. *J. Infect. Dis.* 192:170–177. <http://dx.doi.org/10.1086/430613>.
- Kozarewa I, Turner DJ. 2011. 96-Plex molecular barcoding for the Illumina Genome Analyzer. *Methods Mol. Biol.* 733:279–298. http://dx.doi.org/10.1007/978-1-61779-089-8_20.
- Gardner SN, Hall BG. 2013. When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLoS One* 8:e81760. <http://dx.doi.org/10.1371/journal.pone.0081760>.
- Richterich P. 1998. Estimation of errors in “raw” DNA sequences: a val-

- idation study. *Genome Res.* 8:251–259. <http://dx.doi.org/10.1101/gr.8.3.251>.
27. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <http://dx.doi.org/10.1093/bioinformatics/btr507>.
 28. Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Proceedings of the German Conference on Bioinformatics, GBC Braunschweig and University of Bielefeld, Bielefeld, Germany. <http://www.bioinfo.de/isb/gcb99/talks/chevreux/main.html>.
 29. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. <http://dx.doi.org/10.1186/gb-2004-5-2-r12>.
 30. Kumar S, Tamura K, Nei M. 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput. Appl. Biosci.* 10:189–191.
 31. Voelz K, Ma H, Phadke S, Byrnes EJ, Zhu P, Mueller O, Farrer RA, Henk DA, Lewit Y, Hsueh Y-P. 2013. Transmission of hypervirulence traits via sexual reproduction within and between lineages of the human fungal pathogen *Cryptococcus gattii*. *PLoS Genet.* 9:e1003771. <http://dx.doi.org/10.1371/journal.pgen.1003771>.
 32. Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95:315–327. <http://dx.doi.org/10.1016/j.ygeno.2010.03.001>.
 33. Neafsey DE, Barker BM, Sharpton TJ, Stajich JE, Park DJ, Whiston E, Hung CY, McMahan C, White J, Sykes S, Heiman D, Young S, Zeng Q, Abouelleil A, Aftuck L, Bessette D, Brown A, FitzGerald M, Lui A, Macdonald JP, Priest M, Orbach MJ, Galgiani JN, Kirkland TN, Cole GT, Birren BW, Henn MR, Taylor JW, Rounsley SD. 2010. Population genomic sequencing of *Coccidioides* fungi reveals recent hybridization and transposon control. *Genome Res.* 20:938–946. <http://dx.doi.org/10.1101/gr.103911.109>.
 34. Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW. 2011. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proc. Natl. Acad. Sci. U. S. A.* 108:2831–2836. <http://dx.doi.org/10.1073/pnas.1014971108>.
 35. Da Cunha KC, Sutton DA, Gene J, Capilla J, Cano J, Guarro J. 2012. Molecular identification and in vitro response to antifungal drugs of clinical isolates of *Exserohilum*. *Antimicrob. Agents Chemother.* 56:4951–4954. <http://dx.doi.org/10.1128/AAC.00488-12>.
 36. Condon BJ, Leng Y, Wu D, Bushley KE, Ohm RA, Otilar R, Martin J, Schackwitz W, Grimwood J, MohdZainudin N, Xue C, Wang R, Manning VA, Dhillon B, Tu ZJ, Steffenson BJ, Salamov A, Sun H, Lowry S, LaButti K, Han J, Copeland A, Lindquist E, Barry K, Schmutz J, Baker SE, Ciuffetti LM, Grigoriev IV, Zhong S, Turgeon BG. 2013. Comparative genome structure, secondary metabolite, and effector coding capacity across *Cochliobolus* pathogens. *PLoS Genet.* 9:e1003233. <http://dx.doi.org/10.1371/journal.pgen.1003233>.
 37. Mikosz CA, Smith RM, Kim M, Tyson C, Lee EH, Adams E, Straif-Bourgeois S, Sowadsky R, Arroyo S, Grant-Greene Y. 2014. Fungal endophthalmitis associated with compounded products. *Emerg. Infect. Dis.* 20:248. <http://dx.doi.org/10.3201/eid2002.131257>.