The Journal of Infectious Diseases

SUPPLEMENT ARTICLE

IDSA Infectious Diseases Society of America

hivma hiv medicine association

OXFORD

Ebola Virus Epidemiology and Evolution in Nigeria

Onikepe A. Folarin,[1,2,a] Deborah Ehichioya,[1,2,a] Stephen F. Schaffner,[7,8,a] Sarah M. Winnicki,[7,8,a] Shirlee Wohl,[7,8,a] Philomena Eromon,[1,2] Kendra L. West,[7] Adrianne Gladden-Young,[7] Nicholas E. Oyejide,[1] Christian B. Matranga,[7] Awa Bineta Deme,[13] Ayorinde James,[3] Christopher Tomkins-Tinch,[7] Kenneth Onyewurunwa,[1,2] Jason T. Ladner,[10] Gustavo Palacios,[10] Iguosadolo Nosamiefan,[7] Kristian G. Andersen,[11] Sunday Omilabu,[4] Daniel J. Park,[2,7] Nathan L. Yozwiak,[2,7,8] Abdusallam Nasidi,[6] Robert F. Garry,[2,12] Oyewale Tomori,[1,5] Pardis C. Sabeti,[2,7,8,9] and Christian T. Happi[1,2]

[1]Department of Biological Sciences, [2]African Center of Excellence for Genomics of Infectious Diseases, Redeemer's University, Ede, Osun State, Departments of [3]Biochemistry, and [4]Medical Microbiology and Parasitology, College of Medicine, University of Lagos, [5]Nigerian Academy of Science, Akoka-Yaba, Lagos, and [6]Nigeria Centre for Disease Control, Abuja; [7]Broad Institute of Harvard and MIT, [8]FAS Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, [9]Howard Hughes Medical Institute, Chevy Chase, and [10]Center for Genome Sciences, US Army Medical Research Institute of Infectious Diseases, Frederick, Massachusetts; [11]The Scripps Research Institute, Scripps Translational Science Institute, La Jolla, California; [12]Department of Microbiology and Immunology, Tulane University, New Orleans, Louisiana; and [13]Department de Parasitologie et Mycologie, Université Cheikh Anta Diop de Dakar, Fann, Dakar, Senegal

Containment limited the 2014 Nigerian Ebola virus (EBOV) disease outbreak to 20 reported cases and 8 fatalities. We present here clinical data and contact information for at least 19 case patients, and full-length EBOV genome sequences for 12 of the 20. The detailed contact data permits nearly complete reconstruction of the transmission tree for the outbreak. The EBOV genomic data are consistent with that tree. It confirms that there was a single source for the Nigerian infections, shows that the Nigerian EBOV lineage nests within a lineage previously seen in Liberia but is genetically distinct from it, and supports the conclusion that transmission from Nigeria to elsewhere did not occur.

Keywords. Ebola; genomic; phylogeny; epidemiology; Nigeria; sequencing; outbreak.

The 2014 outbreak of Ebola virus (EBOV) disease (EVD) in Nigeria was one branch of the major West African epidemic that spanned 2013–2016. As of 13 March 2016, a total of 28 639 EVD cases and 11 316 deaths have been reported in 10 countries. The majority of EVD burden has occurred in Liberia, Sierra Leone, and Guinea, with exported cases responsible for additional transmissions in the United States, Mali, and Nigeria, and diagnosed cases with no transmissions in the United Kingdom, Italy, Senegal, and Spain [1].

The Nigeria EVD outbreak began on 20 July 2014, when a traveler from Liberia (the index case patient), who was infected with EBOV, arrived by commercial aircraft to Murtala Muhammed International Airport in Lagos. The traveler's movement was quickly restricted, patient samples were confirmed EBOV positive by independent polymerase chain reaction (PCR) tests within days, and intensive contact tracing was conducted. The Nigeria EVD outbreak ended on 20 October 2014, when the country was declared Ebola free by the World Health Organization. During that period, 20 individuals are reported to have been infected, of whom 8 died.

Despite emerging in the megacity of Lagos, the Nigeria EVD outbreak was well documented and well contained because of rapid detection of the index case and thorough contact tracing throughout the outbreak. Contact tracing provides a detailed understanding of viral spread, which is key to controlling any viral outbreak. Sequencing of patient samples can also be used to understand transmission routes and is especially important in cases where contact tracing is not available, or when contact tracing cannot completely resolve a transmission chain.

The EVD outbreak in Nigeria is unique because both genetic and contact tracing data are available. The complete transmission chain could be reconstructed with considerable confidence, and detailed clinical records were available for most patients. Viral sequencing data and sampling dates can be used to estimate general transmission patterns between patients and regions, and are used in this case to confirm and inform the transmission chain suggested by contact tracing. Comparing the 2 methods highlights the strengths of each, and the importance of both contact tracing and genomic sequencing during an outbreak.

We present here an account of the Nigeria 2014 EVD outbreak that includes clinical, epidemiological, and viral sequence data for most of the affected patients. We also describe sequencing results generated in Nigeria and in duplicate in the United States for the purposes of both outbreak investigation and validation of viral sequencing capabilities in new laboratories.

MATERIALS AND METHODS

Management of Contacts and Cases of EVD

The index case patient presented to a private hospital in Lagos on 20 July 2014 with fever and body weakness, denied contact

[a]O. A. F., D. E., S. F. S., S. M. W., and S. W. contributed equally to this work.

Correspondence: C. T. Happi, Redeemer's University, Ede, Osun State, Nigeria (happic@run.edu.ng). S. F. Schaffner, Broad Institute of Harvard and MIT, Cambridge, MA 02142 (sfs@broadinstitute.org).

The Journal of Infectious Diseases® 2016;214(S3):S102–9

DOI: 10.1093/infdis/jiw190

with known EVD cases or funeral attendance, and was treated with antimalarial drugs and analgesics. Over the next 3 days, the patient's condition worsened (fever escalated, and vomiting and diarrhea persisted), and EVD was suspected. Filovirus PCR testing was conducted at Lagos University Teaching Hospital, and on 23 July the index case was reported as filovirus positive. Samples were then shared with Redeemer's University (RUN) for EBOV-specific PCR testing, which was confirmed on 25 July 2014. The index patient died on 25 July 2014 (see Case Supplement and Supplementary Data 1).

All persons who were exposed to the index patient and their contacts were traced, placed under surveillance, and monitored for clinical features of EVD. If contacts exhibited fever or other symptoms, they were admitted into the Ebola treatment center (ETC) as suspected case patients; blood samples were then collected and tested with reverse-transcription (RT) PCR for presence of EBOV at both Lagos University Teaching Hospital and RUN. Patients who tested positive with RT-PCR were moved to the confirmed ward of the ETC. This combination—history of contact with an EVD case patient, presentation with symptoms, and RT-PCR evidence of EBOV infection—defined a confirmed case. Each patient was counseled on the need for ≥4 L of oral rehydration solution daily. Treatment was started with antibiotics because of their immunosuppression and antimalarials because of the endemicity of malaria in Nigeria. Patients were also placed on a regimen of nutritional supplements and vitamins. The only analgesic administered was paracetamol. Injectables and invasive procedures were avoided unless patients were too ill or weak to take oral rehydration solution.

Infection prevention and control procedures and protocols were strictly adhered to in patient management. Before discharge, patients were confirmed negative for EVD by RT-PCR. When discharged, they were decontaminated before being allowed to leave the ETC and were not allowed to take clothing or other personal items. Replacement clothes, footwear, and basic personal effects were provided by family or the ETC, depending on individual circumstances.

Data Collection and Review

ETC case management, clinical data, and laboratory data of all confirmed EVD cases identified between 20 July and 30 September were reviewed by qualified medical professionals in the case management team. The following case data were compiled: sociodemographic (age, sex, occupation, and city of residence), clinical (respiratory rate, pulse rate, blood pressure, presenting symptoms, signs, syndromes, and outcome), laboratory (RT-PCR), and administrative data (date of symptoms onset, duration of symptoms, and length of stay).

Each patient's exposure history, presenting symptoms, history of presenting symptoms, course of illness, excerpts of clinical management, and illness outcome were abstracted from medical records or contact tracing interview notes (including suspect evacuation forms, case investigation forms, laboratory request and report forms, clinical notes and charts, and contact tracing interview notes) and summarized as case histories.

Sample Collection and Processing

Samples from patients with suspected EVD were shipped both to the virology laboratory at Lagos University Teaching Hospital for diagnostics and to the African Center of Excellence for Genomics of Infectious Diseases (ACEGID) at RUN for diagnostics and sequencing. Whole-blood samples shipped to RUN were inactivated with AVL buffer (Qiagen) or TRIzol LS reagent (Life Technologies) in a 4:1 ratio, both according to the manufacturer's protocol. Inactivated samples were stored in a −20°C freezer. AVL buffer and TRIzol LS reagent have been used extensively in virus inactivation including for EBOV [2–7]. Samples inactivated in AVL buffer were extracted using the QIAamp Viral RNA Mini Kit extraction protocol (Qiagen), according to the manufacturer's protocol. Samples inactivated in TRIzol reagent were extracted using chloroform modified with an AVL buffer inactivation and QIAamp Viral RNA Mini Kit extraction protocol. Following this modified protocol, 140 µL of chloroform was added to 1 mL of a TRIzol-inactivated sample. After vortex and centrifugation, 200 µL of the aqueous phase was transferred to a tube with 700 µL of AVL buffer without carrier RNA added. The sample was then processed according to the manufacturer's protocol for extraction, using the QIAamp Viral RNA Mini Kit. Extracted RNA samples were divided into aliquots for sequencing at both RUN and the Broad Institute of MIT and Harvard. Samples destined for the Broad Institute were shipped on dry ice and subsequently stored at −80°C.

Diagnostics Performed at RUN

EBOV-specific diagnostic tests were performed on the suspected EBOV samples at RUN with RT-PCR using the SuperScript III One-Step RT-PCR System with Platinum Taq High Fidelity DNA Polymerase (Life Technologies). The 25-µL assay mix included 5 µL of RNA, KGH primer set [2] at a 250 nmol/L final concentration (forward, GTC GTT CCA ACA ATC GAG CG; reverse, CGT CCC GTA GCT TTR GCC AT), 12.5 µL of ×2 Reaction Mix and 0.5 µL of SuperScript III RT/Platinum Taq High Fidelity Enzyme Mix. The cycling conditions were 60°C for 20 minutes and 94°C for 5 minutes, followed by 35 cycles of 94°C, 58°C, and 68°C for 15 seconds each, with a final extension at 68°C for 2 minutes. RT-PCR was performed on an Eppendorf Mastercycler thermocycler. The samples were analyzed on 1.5% agarose gel, and visual results were recorded.

Quantitative RT-PCR Performed at RUN and the Broad Institute

To assess sample quality, extracted RNA was quantified using quantitative RT-PCR (qRT-PCR) for both EBOV and human ribosomal RNA (18S). RNA selected for sequencing was quantified using the Power SYBR Green RNA-to-Ct 1-Step qRT-PCR assay (Life Technologies). The Kulesh assay protocol was adapted from a probe-based quantitative PCR (qPCR) assay to a

SYBR qPCR assay by omitting the probe [8]. The 10-µL assay mix included 3 µL of RNA, 0.3 µmol/L primer Kulesh forward (TCT GAC ATG GAT TAC CAC AAG ATC), 0.3 µmol/L Kulesh reverse (GGA TGA CTC TTT GCC GAA CAA TC), 5 µL of ×2 Power SYBR Green RT-PCR Mix and 0.08 µL of RT Enzyme Mix (Life Technologies). The cycling conditions were 48° C for 30 minutes and 95°C for 10 minutes, followed by 45 cycles of 95°C for 15 seconds and 60°C for 30 seconds with a melt curve of 95°C for 15 seconds, 55°C for 15 seconds, and 95°C for 15 seconds. qRT-PCR was performed on the LightCycler 96 (Roche) instrument at both RUN and the Broad Institute. Synthetic oligonucleotide amplicons were prepared as a standard to quantify the viral copy number in the qRT-PCR assays. These amplicons represent a portion of the EBOV segment within the L gene as a template for PCR. The amplicons were cleaned using AMPure XP beads (Beckman Coulter Genomics) and quantified by the TapeStation system (Ambion). Amplicon concentrations were converted to EBOV copies per microliter for quantification.

RNA Processing and Library Preparation

DNA was depleted from the RNA samples using TURBO DNase (Ambion), and host ribosomal RNA was then depleted from the samples using an RNase H selective depletion method described elsewhere [2, 9, 10]. Complementary DNA was then synthesized from the resulting depleted RNA, Nextera XT libraries were constructed, and Illumina sequencing was carried out according to methods described elsewhere [2, 11], with the modification that Nextera libraries were generated using 16–18 cycles of PCR. Samples were sequenced on the MiSeq platform at RUN, and on both the MiSeq and HiSeq 2500 platforms (Illumina) at the Broad Institute.

EBOV Genome Assembly and Analysis

Raw sequencing reads from all sequencing runs were processed together and assembled using the viral-ngs pipeline (version 1.0.0) [12, 13] with mostly default parameters. Reads from 2 flow cells were not included owing to suspected contamination. Two parameters were varied from defaults: the minimum length of assembly (expressed as a fraction of the reference genome length) and minimum fraction of unambiguous bases were both decreased to allow assembly of lower-quality samples; these parameters were 0.8 and 0.7, respectively.

Consensus variants were called using a custom pipeline and annotated using the program SnpEff (version 4.1) [14]. Multiple alignments were performed using MAFFT software (version 7.017) [15, 16] with default parameters. Within-host variants were identified as part of the viral-ngs pipeline with default minimum read and strand bias filters.

The maximum likelihood tree was produced using IQ-TREE software (version 1.3.13) [17], a TIM+I (a transitional model with a proportion of invariable sites) substitution model selected by ModelFinder (implemented in IQ-TREE), and 1000 bootstrap replicates. Liberian EBOV sequences included all genomes publicly available on GenBank as of 17 February 2016 (Supplementary Data 3). (Sequence assemblies are available from GenBank and reads available from the sequence read archive, accessible under BioProject PRJNA316870.)

Data Analysis

As noted in the Discussion, new single-nucleotide polymorphisms (SNPs) were observed to be clustered, with 6 SNPs appearing in 1 sample, 2 in another, and none in the remaining 9 samples. To determine whether this was unlikely given a uniform mutation rate per transmission, a P value was calculated as follows. From the transmission tree, the sequenced cases represent a minimum of 11 transmissions from the index case. Assume that new SNPs in a transmission occur in a Poisson process at an unknown rate, μ_s. For a given μ_s, we calculate the probability of seeing 4 new SNPs in ≥1 case and then integrate over all values of μ_s, weighting by the probability of observing 6 SNPs in 11 transmissions. That is,

$$p = \frac{\int p(S_t = 6|\mu_s)(1 - p(S_s < 4|\mu_s))^{N_t} \mathrm{d}\mu_s}{\int p(S_t = 6|\mu_s)\mathrm{d}\mu_s} \qquad (1)$$

where S_t is the total number of new SNPs, S_s is the number of new SNPs seen in a single case, and N_t is the number of transmissions. The first probability is the Poisson probability density function, $p(S_t|\mu_s) = ((\mu_s N_t)^{S_t} e^{-\mu_s N_t})/S_t!$, and the second is the cumulative distribution function, $e^{-\mu_s} \sum_{i=0}^{N_t-1} (\mu_s^i/i!)$.

RESULTS

Clinical Data

Available metadata on the Nigerian patients with EVD are summarized in Supplementary Data 1, and symptoms and outcome for all 20 are summarized in Supplementary Tables 1 and 2. Their median age was 33 years (range, 26–62 years), and 55% were female. Most (65%) were <40 years of age, and most (65%) were health workers. At presentation, the most common symptoms were fever (85%), fatigue (70%), and diarrhea (65%). The pulse rate and blood pressure were within normal range in 50% of the patients, but the respiratory rate was elevated in 90% of those with available data. The common clinical syndromes documented were gastroenteritis (45%), hemorrhage (30%), and encephalopathy (15%). Of 20 patients, 12 (60%) survived, with 1 having postillness mental health complication requiring follow-up. The mean (standard deviation) duration from onset of symptoms to presentation at the ETC was 3 (2) days among survivors, compared with 5 (2) days for nonsurvivors. The mean duration from symptom onset to death or discharge from the ETC was 15 (5) days for survivors and 11 (2) days for nonsurvivors.

Sequencing Data

We prepared 16 samples from 13 of the 20 patients with confirmed EVD and discharge samples for 3 of them. This includes case 9, which could not be confidently matched to a sample

Table 1. Sample Coverage

Sample No.	Case No.	Coverage, %[a]	× Coverage[b]	GenBank Accession No.
E001	Index	99.8	1364	KX013101
E020	2	99.5	158	KX013092
E021	3	99.8	520	KX013099
E023	4	99.7	525	KX013097
E024	5	99.8	4864	KX013091
E027	6	99.5	159	KX013098
E029	7	99.7	474	KX013093
E033	8	82.4	6	KX013090
E030	9	99.8	292	KX013094
E039	10	90.4	8	KX013100
E076	11	99.1	25	KX013096
E130	13	99.1	14	KX013095

[a] Percentage of bases with ≥1× coverage.

[b] Median depth of coverage.

(suspected match to E030). We prepared an additional 16 samples from suspected cases in which the sample could not be clearly associated with a particular case because of incomplete records. Dates and qRT-PCR results for each of these samples are reported in Supplementary Data 1. Because these data include retested and discharge samples, as well as incomplete information collated many months after the outbreak, we were not able to confirm that there were exactly 20 EVD cases in Nigeria. After inactivation and extraction at RUN, we divided RNA from each sample into 2 aliquots for independent library preparation and sequencing at RUN and the Broad Institute. Extracted RNA samples contained an average of 3.97×10^6 18S copies/mL (range, 3.28×10^4 to 2.31×10^7 copies/mL) as determined by qRT-PCR.

We prepared Nextera libraries for all 32 samples. Using the Kulesh qRT-PCR assay, we detected EBOV RNA in 18 of these samples, including 2 discharge samples and 3 samples unassociated with a particular case. After library construction, we used Kulesh qPCR to detect the presence of any EBOV copies in the libraries. Based on the results, we sequenced 23 samples using a combination of the MiSeq and HiSeq 2500 platforms (Illumina). We were able to generate assembled EBOV genomes from 12 of these samples, all from confirmed EVD cases with associated case histories. We combined the MiSeq and HiSeq sequencing data from RUN and the Broad Institute for analysis. The median sequencing coverage was 225.5× (range, 6–4864×) (Table 1). Although we recorded combined sequencing data, the MiSeq data from RUN separately confirmed EBOV reads in 6 of the 12 samples with assembled EBOV genomes.

Consensus and Within-Host Variants

We identified 17 consensus-level variants (9 synonymous, 5 nonsynonymous, 3 noncoding, all relative to the earliest EBOV sequence from the West African outbreak (accession No. KJ660346.2) in EBOV genomes from the 12 sequencing-positive Nigerian samples (Table 2). Variants characteristic of the LB5 (Liberia sublineage 5) [18] were shared by all Nigeria EBOV genomes. The Nigerian EBOV genomes also shared 3 variants not common in Liberia, at positions 4037, 17 016, and 18 754 (Table 2). These variants were present in all

Table 2. Consensus SNPs Seen in Nigeria[a]

Position	Reference Allele	Alternative Allele	Type	Gene	Substitution	Lineage	Count
800	C	T	Missense	NP	R111C	SL2	12
1849	T	C	Silent	NP	D460D	SL1	12
2895	C	T	Noncoding	. . .	. . .	. . .	1 (E020)
3336	A	G	Missense	VP35	N70D	. . .	1 (E020)
3920	G	A	Silent	VP35	Q264Q	. . .	1 (E020)
4037[b]	T	C	Silent	VP35	I303I	. . .	12
6056	A	C	Silent	GP	I6I	LB5	12
6283	C	T	Missense	GP	A82V	SL1	11[c]
7551	T	C	Missense	GP	V505A	. . .	1 (E030)
8928	A	C	Silent	VP30	P140P	SL2	12
10 503	A	G	Silent	VP24	G53G	. . .	1 (E030)
11 201	A	G	Noncoding	. . .	. . .	. . .	1 (E020)
15 963	G	A	Silent	L	K1461K	SL2	12
16 514	G	A	Missense	L	S1645N	LB5	12
17 016[b]	C	T	Silent	L	S1812S	. . .	12
17 142	T	C	Silent	L	F1854F	SL2	12
18 754[b]	A	T	Noncoding	. . .	. . .	. . .	10[d]

Abbreviation: SNPs, single-nucleotide polymorphisms.

[a] All variants and positions are relative to the KJ660346.2 Guinea genome from early in the outbreak. The lineage column includes previously published clade-defining SNPs ancestral to the Nigeria lineage.

[b] These 3 SNPs are novel to Nigeria (except 18 754, which is shared by 2 Ebola virus genomes from Liberia) and are shared by all Nigerian samples.

[c] No coverage in sample E033.

[d] No coverage in sample E033 or E039.

Nigerian samples sequenced, including the index case (we note that 2 samples did not have coverage at position 18 574). Two of these variants were unique to Nigeria, and 1 variant, at position 18 754, was also seen in 2 EBOV genomes from Liberia (accession Nos. KT725314 and KT725261), suggesting a close relationship of the Nigeria clade to those samples. Two Nigerian samples had unique additional consensus variants.

We also identified 31 intrahost single-nucleotide variants (iSNVs) in 5 of the 12 EBOV genomes from Nigeria (5 synonymous, 5 nonsynonymous, 5 noncoding SNPs, and 16 insertions/deletions) (Supplementary Data 2). We sequenced each of the 5 samples with iSNVs at least twice from replicate libraries, and iSNV calls were concordant between libraries. Eight of these iSNVs were shared by ≥2 samples, and 2 iSNVs (positions 7551 and 10 503), both found in sample E027, were also consensus variants in sample E030. The presence and number of iSNVs found correlated roughly with sample coverage; only samples with >100× coverage had >1 iSNV call that passed our basic filters.

Phylogenetic Tree

To better understand the evolutionary relationship between the EVD outbreak in Nigeria and the West African outbreak as a whole, we created a maximum likelihood tree (Figure 1). The tree confirms that the EVD outbreak in Nigeria was due to a single introduction from Liberia, as suggested by contact tracing. More specifically, the EBOV genomes from Nigeria are descendants of the LB5 clade in Liberia [18]. No EBOV sequences yet sampled outside Nigeria descend from the Nigerian EBOV isolates [2, 19–23], indicating containment of EVD cases in Nigeria within the larger outbreak, as also suggested by contact tracing.

Reconstructed Transmission Tree

Given the phylogenetic tree of the sampled viruses, along with their dates, it is possible to infer at least the outlines of the chain of transmission from one patient to another (Figure 2*A*). Ten Nigerian EBOV have identical consensus sequences, suggesting that these sequences are closely connected by direct transmissions. Date information identifies sample E001, the index case, as the earliest-sampled case in Nigeria (collection date: 22 July 2014). Of the other 9 identical genomes, 7 have collection dates from 4 August 2014–8 August 2014.

The close proximity of the sample collection dates to each other suggests that each of the corresponding case patients was infected by the index case patient (ie, it is unlikely that an individual presenting

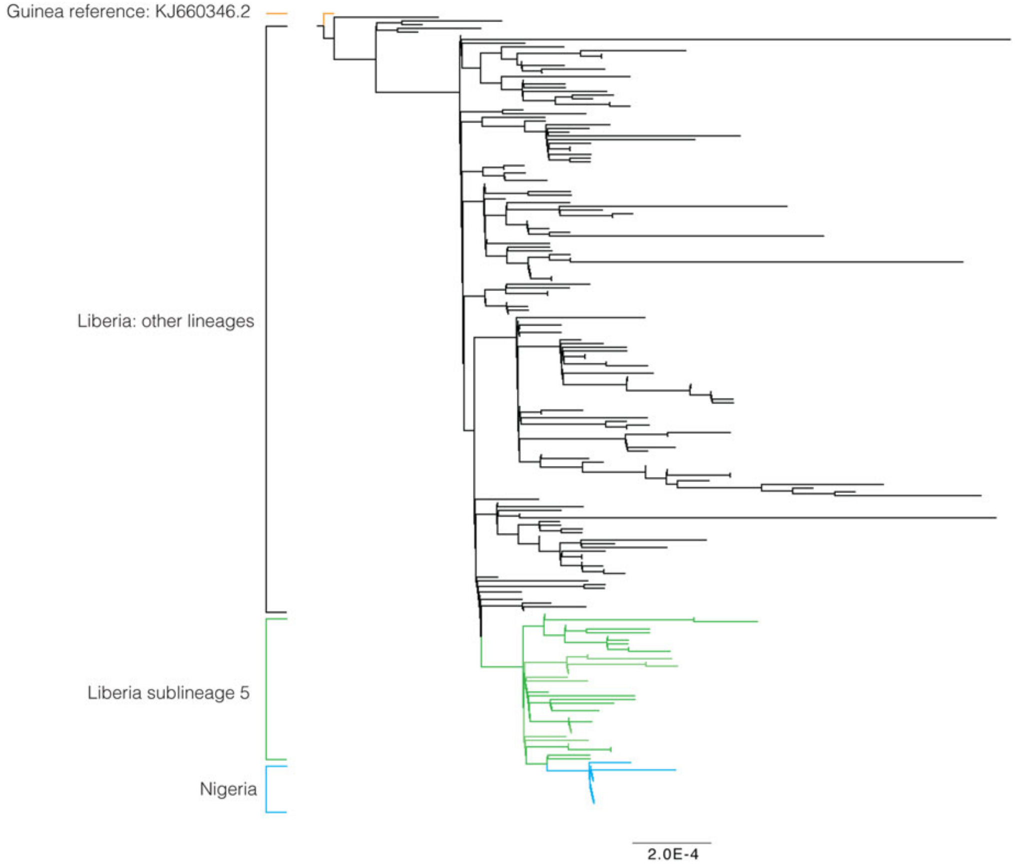

Figure 1. Maximum likelihood tree. Phylogenetic analysis confirms a single introduction of Ebola virus into Nigeria from Liberia and places all Nigerian sequences as descendents of Liberia sublineage 5. Two Liberia sublineage genomes (accession Nos. KT725314 and KT725261) cluster closely with Nigerian samples owing to a shared variant at position 18 754. (Scale bar indicates nucleotide substitutions per site.)

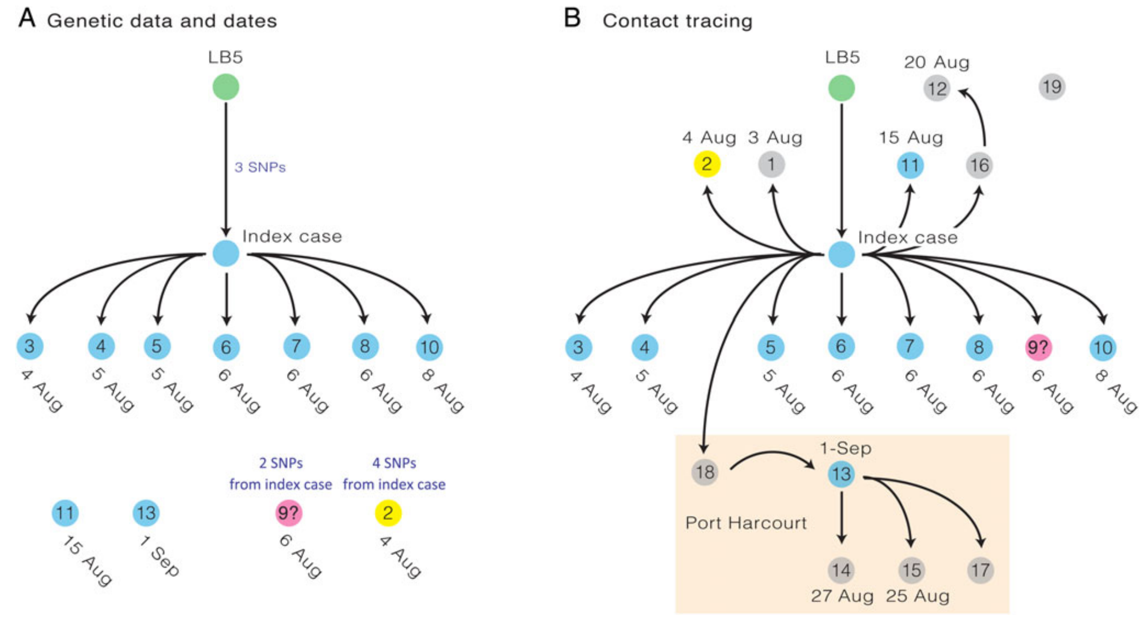

Figure 2. Transmission tree. *A*, Transmission reconstructed from of Ebola virus genome sequence and sample dates only. Arrows indicate likely transmission; cases not connected to arrows cannot be placed within the transmission tree given the available data. LB5, Liberia sublineage 5 reference. *B*, Transmission reconstructed from contact tracing only. Contact tracing provides more precise information, but is not always available. Samples were collected in Lagos, Nigeria, unless otherwise identified. Each case is labeled with its sample collection date; cases not connected to sequenced samples are labeled with date of hospitalization. Samples are colored by consensus sequence (ie, samples with identical viral genomes are similarly colored). Cases in gray are those for which genetic data are not available. Abbreviation: SNPs, single-nucleotide polymorphisms.

symptoms on 8 August 2014 would have been infected <4 days previously) [24]. The remaining 2 cases with viral genomes identical to the index case are dated 15 August and 1 September, and these patients therefore may have been infected by 1 of the earlier case patients. The presence of additional SNPs in the viral genomes corresponding to cases 2 and 9 make it difficult to place these samples within the transmission chain. However, case 6 has an iSNV at each of the 2 case 9 SNP positions (position 7551, 21% minor allele frequency; position 10 503, 16% minor allele frequency) (Supplementary Data 2), suggesting that these 2 cases are closely linked.

In the limited Nigeria EVD outbreak, it was also possible to reconstruct a nearly complete transmission chain based on contact tracing alone (Figure 2*B*). Such a reconstruction is feasible in this case because (1) EBOV spreads primarily through direct contact, (2) there were few cases (multiple exposures were uncommon), and (3) intensive efforts were made to trace and monitor all suspected contacts. The contact tracing information resulted in a transmission tree similar to that suggested by genetic data, with the index case responsible for a majority of transmissions. This data also revealed that 1 individual (case patient 18) traveled from Lagos to Port Harcourt while infected with EBOV, where he acted as the index patient in a small secondary outbreak containing 4 additional EVD cases.

DISCUSSION

The 2014 Nigeria outbreak is unusual for an EVD outbreak in the detailed information available about its development: we have both a good reconstruction of the transmission chain of 20 patients, and viral genomic data from most cases in the chain. The completeness of the record reflects the public health situation: Nigeria was prepared for the arrival of EBOV and was able to implement thorough contact tracing promptly after the index case was diagnosed, while the number of cases was still small. That effort was critical in containing the outbreak, but it is also very helpful in reconstructing its details afterward. Combined with sequence data, the transmission chain helps us interpret the changes occurring in the virus, because it generally lets us pinpoint where in the chain each new mutation actually occurred.

Viewed by itself, sequence data can serve to provide a broad picture of an outbreak, and that is true of this EVD outbreak. This capability is obviously useful when contact tracing is absent or incomplete, as is usually the case with epidemics. In the 2014 Nigeria outbreak, sequencing alone makes it clear that the entire outbreak stemmed from a single introduction of EBOV into the country. It also places the Nigerian outbreak in its larger context, identifying a particular branch of the Liberian LB5 lineage of EBOV as the source and showing that the Nigerian lineage did not spread into other countries.

Identifying individual links in the transmission chain is usually beyond the resolution of sequence data, however, and requires contact tracing in the field. The resolution of genomic data is limited because new variants arise less often than new cases, meaning that many cases will be genetically indistinguishable. This can be seen in our data in Figure 2A, in which multiple successive links in the chain share identical genomes. In addition, when mutations do occur, >1 can arise in a single patient, making genetic distance an imperfect guide to the number of transmission links that have occurred. Thus, most of the cases infected directly by the index patient in Nigeria had identical genomes, but 1 case (case 4) differed by 4 mutations, even though it too resulted from a single transmission. Contact tracing (Figure 2B)—when it is available—does not suffer from such limitations.

Within-host variants (iSNVs) that are shared between patients can provide a more detailed picture of transmission routes, but our data point out some important caveats about their usefulness. First, detection of iSNVs requires deep sequencing of good-quality samples, and that is not always possible: deep enough sequencing could be achieved for only two-thirds of our sequenced samples. Second, even when iSNV data are available, it may not all be meaningful. Some of the iSNVs we observed have previously been documented in unrelated data sets from Sierra Leone and Liberia [2, 13, 18]; these included all 8 of the shared iSNVs. Most of our iSNVs, including most shared iSNVs, were low-frequency frameshift insertions or deletions. Because they can disrupt protein structure, they are unlikely to be transmitted. More likely, these recurrent iSNVs represent either recurring mutations in highly mutable regions of the EBOV genome or sequencing errors, especially because many of them occur in homopolymer regions. In either case, their value for determining transmission chains is uncertain. More research is necessary to fully make use of within-host genomic data in understanding transmission, including better sequencing coverage for all samples and improved methods to identify false-positives.

One aspect of our genomic data that is slightly surprising is the distribution of new variants, which is not at all uniform. Our sequenced samples include the results of 11 transmissions from the index case. Nine of these produced no new consensus SNPS, 1 produced 4 new SNPs, and 1 produced 2 (Figure 2A). This clustering of mutations in certain samples suggests the possibility that the mutation rate was not uniform across all of the cases. This is no more than a possibility, though, because the clustering is not statistically significant ($P = .07$).

Also puzzling is a pair of variants that were seen twice, once as consensus SNPs (in case 9) and once as iSNVs (in case 6). Based on sample dates and contact data, both of these patients were infected by the index patient, so presumably they inherited these variants from that patient. We do not, however, find them in the sample from the index patients, either as consensus SNPs or as iSNVs, despite high sequencing depth. Nor do they appear as consensus SNPs in the other cases derived from the index case, or as iSNVs in the one other case that was deeply sequenced and was sampled around the same time as samples 6 and 9. The explanation may simply be that the variants were present in the index patient but at too low a frequency for us to detect. It is also possible that their frequency changed in the index patient between the time he was sampled and transmission to the other cases, or that they differed across tissues within the patient. Better understanding of the dynamics of within-host evolution and transmission, and of our power to detect iSNVs, would help clarify this issue.

The genomic data were invaluable in revealing what was happening to the virus during the outbreak, but it would have been even more informative had samples been of uniformly high quality. Many samples did not produce whole-genome assemblies because of poor sample quality, and a third of those that did could not be used to detect iSNVs. This highlights the importance of rapid sequencing in clinical settings during outbreaks, with well-established sample collection and processing protocols. Although at the time of the outbreak sequencing was not yet ready on site, sequencing capability is now becoming increasingly available throughout many regions. With high-throughput deep sequencing now being routinely performed by ACEGID at RUN, high-resolution pathogen information can now be generated to elucidate outbreak dynamics and response, both in Nigeria and throughout West Africa.

Data handling could similarly benefit from good protocols established in advance. In the case of the data presented here, clinical and contact data were separated from sequence data, and the correspondence between them had to be established post hoc, a process that was both laborious and uncertain. In an outbreak setting, keeping track of different kinds of data is not the highest priority, but valuable information can be lost as a result. Having a system for collecting and maintaining both clinical and laboratory data established in advance would be very helpful.

Supplementary Data

Supplementary materials are available at http://jid.oxfordjournals.org. Consisting of data provided by the author to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the author, so questions or comments should be addressed to the author.

Notes

Acknowledgments. We would like to thank Mike Lin and Yifei Men at DNAnexus for their engineering work to assist with analysis of data generated by the African Center of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer's University (RUN). We also thank the RUN management for the support provided to ACEGID staff during the 2014 EVD outbreak in Nigeria, members of the Emergency Operations Center in Nigeria during the outbreak, and the federal government of Nigeria.

Disclaimer. The content of this publication does not necessarily reflect the views, policies, or official positions of the US Army.

Financial support. This work was supported by grants from Illumina Corporation and the United States Agency for International Development (grant OAA-G-15-00001), the National Institutes of Health (NIH) (grant 5U01HG007480-03), the World Bank (ACE 019), the Bill and Melinda Gates Foundation (grant OPP1123407), and the Howard Hughes Medical Institute (to P. C. S.). Sequencing and analysis work at the Broad Institute was supported by federal funds from the National Institute of Allergy and Infectious Diseases, NIH, Department of Health and Human Services (under grant U19AI110818).

Potential conflicts of interest. R. F. G. is the founder of Zalgen Labs. All other authors report no potential conflicts. All authors have submitted the

ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. World Health Organization. Ebola situation report. http://apps.who.int/ebola/current-situation/ebola-situation-report-16-march-2016. Accessed 16 March 2016.
2. Gire SK, Goba A, Andersen KG, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science **2014**; 345:1369–72.
3. Günther S, Asper M, Röser C, et al. Application of real-time PCR for testing antiviral compounds against Lassa virus, SARS coronavirus and Ebola virus in vitro. Antiviral Res **2004**; 63:209–15.
4. Grard G, Biek R, Tamfum J-JM, et al. Emergence of divergent Zaire Ebola virus strains in Democratic Republic of the Congo in 2007 and 2008. J Infect Dis **2011**; 204(suppl 3):S776–84.
5. Kobinger GP, Leung A, Neufeld J, et al. Replication, pathogenicity, shedding, and transmission of *Zaire ebolavirus* in pigs. J Infect Dis **2011**; 204:200–8.
6. Hoenen T, Jung S, Herwig A, Groseth A, Becker S. Both matrix proteins of Ebola virus contribute to the regulation of viral genome replication and transcription. Virology **2010**; 403:56–66.
7. Blow JA, Mores CN, Dyer J, Dohm DJ. Viral nucleic acid stabilization by RNA extraction reagent. J Virol Methods **2008**; 150:41–4.
8. Trombley AR, Wachter L, Garrison J, et al. Comprehensive panel of real-time TaqMan polymerase chain reaction assays for detection and absolute quantification of filoviruses, arenaviruses, and New World hantaviruses. Am J Trop Med Hyg **2010**; 82:954–60.
9. Matranga CB, Andersen KG, Winnicki S, et al. Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. Genome Biol **2014**; 15:519.
10. Morlan JD, Qu K, Sinicropy DV. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. PLoS One **2012**; 7:e42882.
11. Adiconis X, Borges-Rivera D, Satija R, et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods **2013**; 10:623–9.
12. Park DJ, Jungreis I, Tomkins-Tinch C, Lin M, Andersen K. viral-ngs. http://dx.doi.org/10.5281/zenodo.17560. (Published 12 May 2015.)
13. Park DJ, Dudas G, Wohl S, et al. Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. Cell **2015**; 161:1516–26.
14. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly **2012**; 6:80.
15. Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **2002**; 30:3059–66.
16. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **2013**; 30:772–80.
17. Nguyen LT, Schmidt HA, Haeseler von A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **2015**; 32:268–74.
18. Ladner JT, Wiley MR, Mate S, et al. Evolution and spread of Ebola virus in Liberia, 2014–2015. Cell Host Microbe **2015**; 18:659–69.
19. Baize S, Pannetier D, Oestereich L, et al. Emergence of Zaire Ebola virus disease in Guinea—preliminary report. N Engl J Med **2014**; 371:1418–25.
20. Tong Y-G, Shi WF, Liu D, et al. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. Nature **2015**; 524:93–6.
21. Carroll MW, Matthews DA, Hiscox JA, et al. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. Nature **2015**; 524:97–U201.
22. Simon-Loriere E, Faye O, Faye O, et al. Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. Nature **2015**; 524:102–U210.
23. Kugelman JR, Wiley MR, Mate S, et al. Monitoring of Ebola virus Makona evolution through establishment of advanced genomic capability in Liberia. Emerg Infect Dis **2015**; 21:1135–43.
24. Chowell G, Nishiura H. Transmission dynamics and control of Ebola virus disease (EVD): a review. BMC Med **2014**; 12:196.