Genome sequence analysis of Ebola virus in clinical samples from three British healthcare workers, August 2014 to March 2015

A Bell[1,2], K Lewandowski (kuiama.lewandowski@phe.gov.uk)[1,2], R Myers[3], D Wooldridge[3], E Aarons[1], A Simpson[1], R Vipond[1,4], M Jacobs[5], S Gharbia[3,4], M Zambon[3]

1. Public Health England, Porton Down, Salisbury, United Kingdom
2. These authors contributed equally to the work and are joint first authors
3. Public Health England, Colindale, London, United Kingdom
4. NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Liverpool, United Kingdom
5. Department of Infection, Royal Free London NHS Foundation Trust, London, United Kingdom

We determined complete viral genome sequences from three British healthcare workers infected with Ebola virus (EBOV) in Sierra Leone, directly from clinical samples. These sequences closely resemble those previously observed in the current Ebola virus disease outbreak in West Africa, with glycoprotein and polymerase genes showing the most sequence variation. Our data indicate that current PCR diagnostic assays remain suitable for detection of EBOV in this epidemic and provide confidence for their continued use in diagnosis.

Monitoring of the evolution of the viral genome during the ongoing outbreak of Ebola virus disease (EVD) in West Africa is crucial for the early detection of mutants that may evade sequence-based diagnostics and for monitoring efficacy of therapeutic options. We present here our analysis of Ebola virus (EBOV) sequences obtained from blood samples from three British healthcare workers (HCWs) who were infected with EBOV in Sierra Leone.

Assessing sequence variation in Ebola virus

Between August 2014 and March 2015, three HCWs (Cases 1, 2 and 3) from the United Kingdom (UK) were infected with EBOV (Ebola virus/H.sapiens-wt/GBR/2014/Makona-UK1, Ebola virus/H.sapiens-wt/GBR/2014/Makona-UK2 and Ebola virus/H.sapiens-wt/GBR/2014/Makona-UK3, respectively; hereafter referred to as UK1, UK2 and UK3) in Sierra Leone.

Two were repatriated from Sierra Leone and the third became symptomatic upon return to the UK. All were transferred to the specialist isolation ward at the Royal Free Hospital in London, where they subsequently recovered. Informed consent was sought and received from each of the patients for viral whole genome sequencing and publication of the findings.

Viral genomes from pre-intervention whole blood and EDTA plasma samples were sequenced and analysed to provide a baseline for any subsequent transmission of EBOV in the UK and to identify and monitor mutations that may affect the sensitivity of treatment and diagnostics (Table 1).

Sequence analysis

RNA was extracted from patient samples using the EZ1 RNA Universal Tissue Kit (QIAgen). Confirmation of EVD diagnosis in all three patients was performed using PCR assays targeting the NP gene [1]. Samples for sequencing were treated with DNase I (Life Technologies) and purified using an RNA Clean and Concentrator kit (Zymo). Single primer isothermal linear amplification (SPIA) cDNA was prepared from total RNA following the Ovation RNA-seq V2 (NuGens) protocol [2], with the exception that RNA was denatured for 5 min at 85 °C before first-strand synthesis. Samples were purified using a MinElute column (QIAgen). Following amplification, paired-end libraries were prepared for Illuminia MiSeq sequencing following the Nextra XT protocol using 1.5 ng of SPIA cDNA. Reads were trimmed to a minimum of Q30. Genomes were mapped to KM233113.1 using BWA 0.7.5 and consensus called with Quasibam 1.0 using a local instance of The Galaxy Project [3-5]. Consensus sequences were produced at a minimum depth of five reads and single nucleotide polymorphisms (SNPs) at a minimum depth of 20. Ambiguous bases were included when present in 20% of reads.

Full viral genome sequences were obtained from samples from all three infected HCWs patient samples and were submitted to GenBank (accession numbers are listed in Table 1). Sequence analysis showed that

Heatmaps showing nucleotide and amino acid variation between three Ebola virus isolates from three British healthcare workers infected in Sierra Leone, August 2014–March 2015
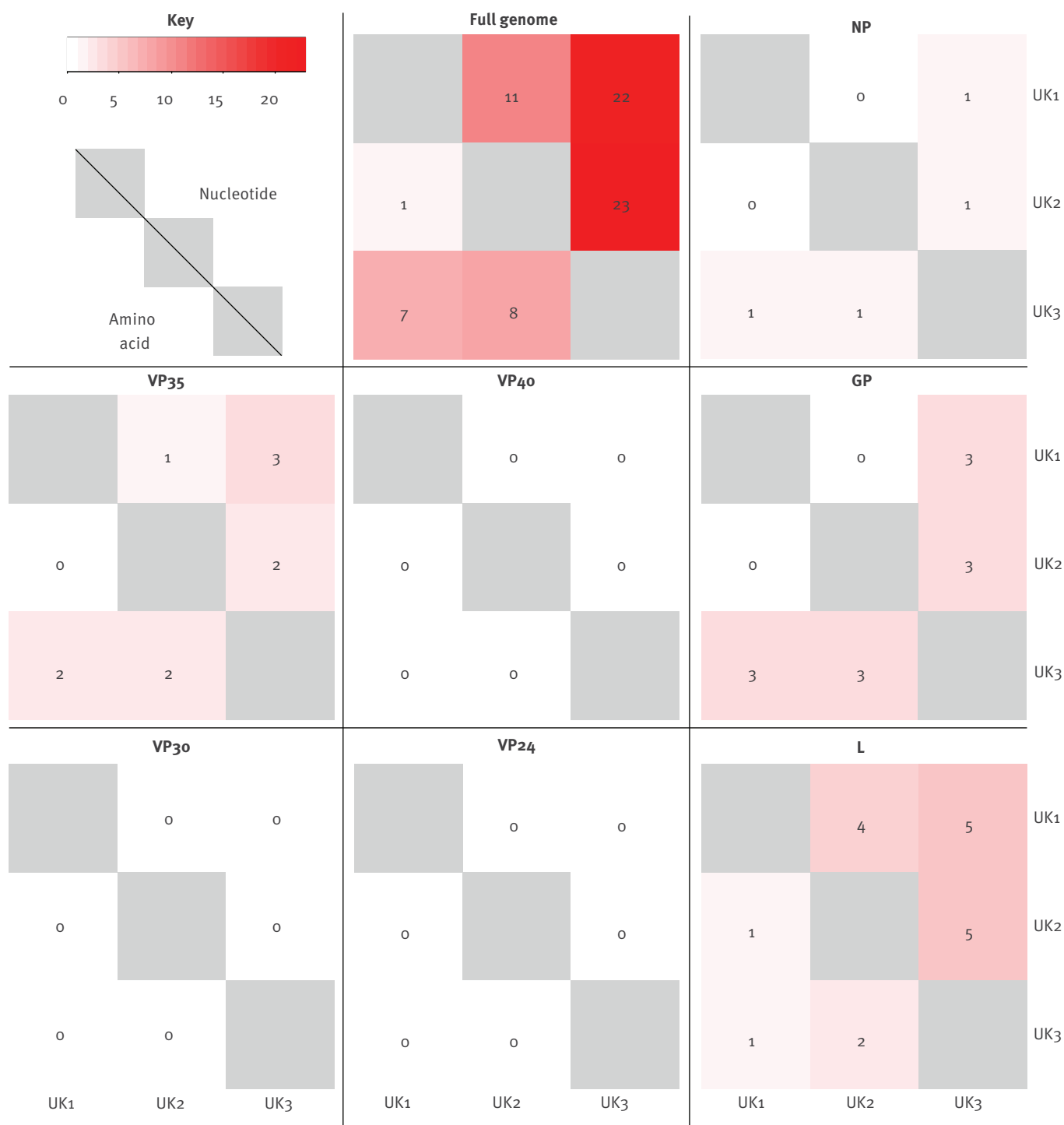
GP: glycoprotein; L: viral polymerase; NP: nucleoprotein; VP: virion protein.

Heatmaps showing the number of nucleotide and amino acid substitutions between Ebola virus isolates from three British healthcare workers infected in Sierra Leone, August 2014, December 2014 and March 2015, across the whole genome and in individual genes. Each heatmap is split in two diagonally, with the top right showing nucleotide changes and the bottom left amino acid differences between the three isolates. The number of differences are displayed in each box.

across the length of the EBOV genome, UK3 showed the most nucleotide variation (22 and 23 SNPs), but no insertions or deletions, compared with UK1 and UK2, respectively (Figure 1). These gave rise to seven and eight amino acid changes, respectively.

No nucleotide changes within the open reading frames (ORFs) for the virion protein (VP) 40, VP30 and VP24 genes were observed. Within the coding region for the nucleoprotein (NP) gene, no SNPs were seen between UK1 and UK2, although UK3 showed one non-synonymous SNP (P to S at position 1,957). One synonymous

Table 1

Sample details from three British healthcare workers with Ebola virus disease infected in Sierra Leone, August 2014–March 2015

Case	Medical centre worked at in Sierra Leone	Date sampled	Sample type	Genome length (bp)	Isolate name	GenBank accession number
1	Kenema Government Hospital	26 Aug 2014	Whole blood	18,920	Makona-UK1	KP120616
2	Kerry Town treatment centre	29 Dec 2014	EDTA plasma	18,929	Makona-UK2	KP658432
3	Kerry Town treatment centre	12 Mar 2015	EDTA plasma	18,921	Makona-UK3	KR025228

bp: base pairs.

SNP was seen between UK1 and UK2 in the VP35 ORF, while UK3 showed two non-synonymous SNPs to UK1 and UK2 (S to R at position 3,371 and E to G at position 3,380).

The GP gene showed no SNPs between UK1 and UK2, and three non-synonymous SNPs from UK3 to UK1 and UK2 (R to K at position 6,932, R to S at 7,265 and L to E at 7,352). The most SNPs within an ORF were found to be in the viral polymerase (L) gene, with UK1 and UK2 showing four nucleotide changes, and UK3 showing five changes in respect to UK1 and UK2. These SNPs total less than one third of SNPs found, for a gene that comprises 36% of the total genome. These data suggest that the L gene is conserved, with only two non-synonymous SNPs. One amino acid change is seen from UK2 to UK1 and UK3 (A to T at 17,848) and one amino acid change from UK3 to UK1 and UK2 (T to A at 16,894) (combined, UK3 differs in one position from UK1 and two positions from UK2).

A phylogenetic tree based on sequences from the three UK samples and all available published sequences was generated using a heuristic maximum likelihood algorithm (Figure 2). Analysis shows that the three UK sequences fall within one large Sierra Leonean clade, with UK2 and UK3 in a different subclade from UK1. UK3 appears to share a common ancestor with the group that UK2 sits within. Sequences from Mali and Liberia form a distinct outgroup from the Sierra Leonean clade.

Discussion

The ongoing EVD outbreak in West Africa is the largest known, with over 25,000 recorded cases up until April 2015 [6]. In response to the outbreak, a large number of international civilian and military aid teams have been deployed alongside local workers at multiple treatment and diagnosis centres in Guinea, Sierra Leone and Liberia. Over 860 HCWs are known to have been infected [6]. Monitoring of the evolution of the viral genome during outbreaks is crucial for the early detection of mutations that may have an impact on disease virulence or transmissibility or affect the sensitivity of sequence-based viral genome detection assays in widespread use. The high viral loads seen in individuals infected with Ebola virus shortly after symptom onset favours the development of whole genome sequencing using next generation sequencing. More

than 450 EBOV genome sequences derived using whole genome sequencing have been reported from samples isolated in Guinea, Sierra Leone, Mali and Liberia [7-9]. Analysis of 78 genomes isolated from samples from patients in Sierra Leone between May and June 2014 suggested an observed evolutionary rate double that seen in previous EVD outbreaks [10]. The importance of tracking sequence variation in relation to molecular detection strategies was highlighted in that analysis. More recent analysis, however, identified an observed evolutionary rate equivalent to that of past outbreaks [11].

In our study presented here, sequence analysis of the NP gene, the target for widely used diagnostic detection assays [1], identified no SNPs within the regions where diagnostic primers bind. The GP gene product is the viral receptor, and the target of neutralising antibodies. Synonymous SNPs are present in locations where primers and probe bind for real-time detection methodologies based on the GP gene [1] (Table 2).

The observation of SNPs within the primer/probe binding sites of the GP gene is consistent with other sequences obtained from this outbreak in West Africa (data not shown). These SNPs are not expected to affect primer binding, although this is yet to be formally determined, but this reinforces the necessity of regular review of diagnostic detection strategies

Table 2

Ebola virus real-time PCR assay primers and probes designed by Trombley et al. [1]

Gene target	Primer/ probe name	Sequence[a]	Base change
NP	F565	TCTGACATGGATTACCACAAGATC	None
	R640	GGATGACTCTTTGCCGAACAATC	None
	p597S	6FAM-AGGTCTGTCCGTTCAA-MGBNFQ	None
GP	F2000	TTT TCA ATCCTCAACCGTAAGGC	None
	R2079	CAG TCC GGT CCC A**G**A ATG TG	G to A
	p2058A	6FAM-CAT GTG CC**G** CCC CAT CGC TGC-TAMRA	G to A

GP: glycoprotein; NP: nucleoprotein.

[a] Single nucleotide polymorphisms in Ebola virus sequences from three patients from the United Kingdom infected in Sierra Leone are shown in bold.

Phylogenetic subtree of 233 near full-length Ebola virus genomes from the West African outbreak that started in 2014
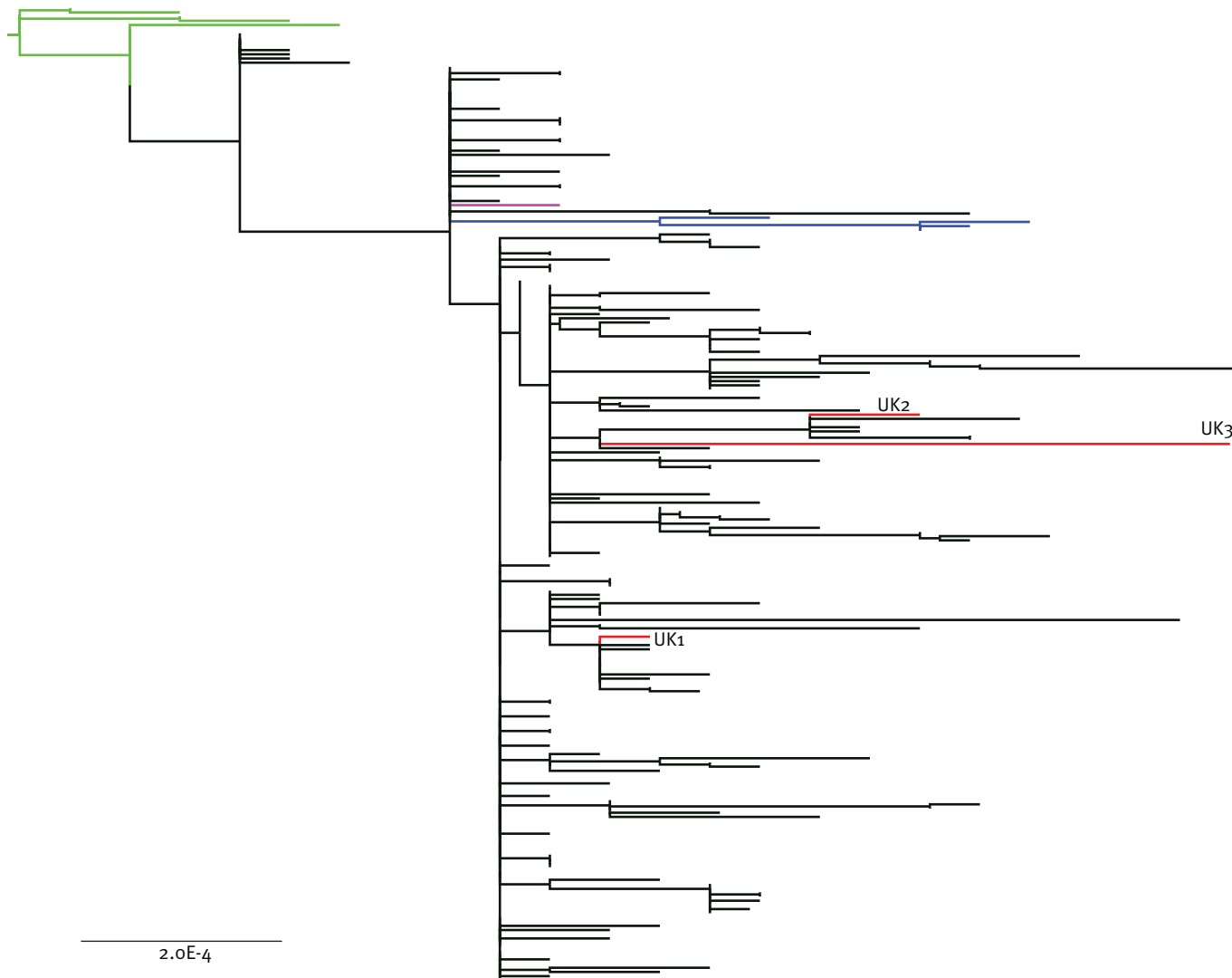
The subtree is taken from a larger tree containing 258 sequences, which includes sequences from earlier outbreaks of Ebola virus disease.

Ebola virus genomes from patients outside the United Kingdom (UK) were obtained from GenBank (n = 255). The tree was generated using a heuristic maximum likelihood algorithm (FastTree – version 2.1.8) and the HKY model of nucleotide substitution.

The position of sequences from three patients repatriated to the UK from Sierra Leone are shown in red and labelled (UK 1–3). Sequences from patients in Guinea are shown in green, those from Mali are shown in blue and those from Liberia in purple. The remaining sequences (black) are from patients in Sierra Leone.

against available sequence information. A recent analysis of sequences from nine EBOVs from Mali and other available sequences also indicated no effect of SNPs on PCR-based detection assays [12,13].

Cases 2 and 3 from whom UK2 and UK3 were obtained, respectively, worked at the same treatment centre before infection and this is reflected in the close nature of the isolates' phylogeny. The patient from whom UK1 was obtained worked elsewhere: the UK1 sequence more closely resembles those reported by Gire et al. [10], who sampled from the same location.

During the intensive and widespread EVD epidemic in West Africa, the evolution of EBOV in Sierra Leone has been driven through person-to-person transmission in community settings, with a high number of HCW infections. HCW infections are less likely, because of rapid ascertainment through strict infection control and health monitoring, to lead to further transmission events. Currently, widely and increasingly used diagnostic detection strategies based on the NP gene have remained suitable for use. Molecular detection strategies based on the GP gene require close attention to ensure that SNPs occurring in this gene, perhaps as a result of host selective pressure, are evaluated for their impact on detection strategies. Viral sequences from any further cases of EVD in UK nationals or those imported into the UK will continue to be sequenced and analysed to ensure continued effectiveness of EVD diagnosis and monitoring of viral genome evolution.

Acknowledgments

Conflict of interest

None declared.

Authors' contributions

Andrew Bell - planned experiments, sample preparation, sequence analysis, wrote the manuscript. Kuiama Lewandowski - planned experiments, sample preparation, sequence analysis, wrote the manuscript. Richard Myers - sequence and phylogenetic analysis. David Wooldridge - sequence library preparation. Emma Aarons -clinical input. Andrew Simpson - clinical input. Richard Vipond - scientific management. Michael Jacobs - lead clinician. Saheer Gharbia - conceived study and scientific management of samples. Maria Zambon - conceived and coordinated the study of sequence comparison of the three UK clinical cases, clinical input, manuscript preparation.

Andrew Bell and Kuiama Lewandowski contributed equally and are joint first authors.

Authors' correction

M Jacobs was inadvertently left out of the author list. This was corrected on 22 May 2015 at the request of the authors.

References

1. Trombley AR, Wachter L, Garrison J, Buckley-Beason VA, Jahrling J, Hensley LE, 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(5):954-60. http://dx.doi.org/10.4269/ajtmh.2010.09-0636 PMID:20439981

2. Malboeuf CM, Yang X, Charlebois P, Qu J, Berlin AM, Casali M, et al. Complete viral RNA genome sequencing of ultra-low copy samples by sequence-independent amplification. Nucleic Acids Res. 2013;41(1):e13. http://dx.doi.org/10.1093/nar/gks794 PMID:22962364

3. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 2005;15(10):1451-5. http://dx.doi.org/10.1101/gr.4086505 PMID:16169926

4. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al., editors. Current protocols in molecular biology. Hoboken, NJ: John Wiley & Sons, Inc.; 2001.

5. Goecks J, Nekrutenko A, Taylor J, Galaxy Team T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010;11(8):R86. http://dx.doi.org/10.1186/gb-2010-11-8-r86 PMID:20738864

6. World Health Organization (WHO). Ebola situation report - 8 April 2015. Geneva: WHO. [Accessed 9 Apr 2015]. Available from: http://apps.who.int/ebola/current-situation/ebola-situation-report-8-april-2015

7. Brister JR, Bao Y, Zhdanov SA, Ostapchuck Y, Chetvernin V, Kiryutin B, et al. Virus Variation Resource--recent updates and future directions. Nucleic Acids Res. 2014;42(Database issue):D660-5. http://dx.doi.org/10.1093/nar/gkt1268 PMID:24304891

8. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2013;41(Database issue):D36-42. http://dx.doi.org/10.1093/nar/gks1195 PMID:23193287

9. Kugelman JR, Wiley MR, Mate S, Ladner JT, Beitzel B, Fakoli L, et al. Monitoring of Ebola virus Makona evolution through establishment of advanced genomic capability in Liberia. Emerg Infect Dis J. 2015 Jul. [Accessed 24 Apr 2015]. http://dx.doi.org/10.3201/eid2107.150522

10. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014;345(6202):1369-72.

11. Tong YG, Shi WF, Di Liu, Qian J, Liang L, Bo XC, et al. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. Nature. 2015. http://dx.doi.org/10.1038/nature14490 PMID:25970247

12. Hoenen T, Safronetz D, Groseth A, Wollenberg KR, Koita OA, Diarra B, et al. Virology. Mutation rate and genotype variation of Ebola virus from Mali case sequences. Science. 2015;348(6230):117-9.

13. Vogel G. Infectious Diseases. A reassuring snapshot of Ebola. Science. 2015;347(6229):1407. http://dx.doi.org/10.1126/science.347.6229.1407 PMID:25814564