

RESEARCH

Open Access



# Spatio-temporal spread and evolution of Lassa virus in West Africa

Xia Wang<sup>1,2†</sup>, Xianwei Ye<sup>2,3†</sup>, Ruihua Li<sup>1†</sup>, Xiaodong Zai<sup>1</sup>, Mingda Hu<sup>1</sup>, Shaoyan Wang<sup>1</sup>, Hongguang Ren<sup>1</sup>, Yuan Jin<sup>1\*</sup>, Junjie Xu<sup>1\*</sup> and Junjie Yue<sup>1\*</sup>

## Abstract

**Background** Lassa fever is a hemorrhagic disease caused by Lassa virus (LASV), which has been classified by the World Health Organization as one of the top infectious diseases requiring prioritized research. Previous studies have provided insights into the classification and geographic characteristics of LASV lineages. However, the factor of the distribution and evolution characteristics and phylodynamics of the virus was still limited.

**Methods** To enhance comprehensive understanding of LASV, we employed phylogenetic analysis, reassortment and recombination detection, and variation evaluation utilizing publicly available viral genome sequences.

**Results** The results showed the estimated the root of time of the most recent common ancestor (TMRCA) for large (L) segment was approximately 634 (95% HPD: [385879]), whereas the TMRCA for small (S) segment was around 1224 (95% HPD: [10301401]). LASV primarily spread from east to west in West Africa through two routes, and in route 2, the virus independently spread to surrounding countries through Liberia, resulting in a wider spread of LASV. From 1969 to 2018, the effective population size experienced two significant increased, indicating the enhanced genetic diversity of LASV. We also found the evolution rate of L segment was faster than S segment, further results showed zinc-binding protein had the fastest evolution rate. Reassortment events were detected in multiple lineages including sub-lineage IIg, while recombination events were observed within lineage V. Significant amino acid changes in the glycoprotein precursor of LASV were identified, demonstrating sequence diversity among lineages in LASV.

**Conclusion** This study comprehensively elucidated the transmission and evolution of LASV in West Africa, providing detailed insights into reassortment events, recombination events, and amino acid variations.

**Keywords** Lassa virus, Phylogeography, West Africa, Transmission, Variation

<sup>†</sup>Xia Wang, Xianwei Ye and Ruihua Li contributed equally to this work.

\*Correspondence:

Yuan Jin  
jin\_0220@sina.com  
Junjie Xu  
xujunjie@sina.com  
Junjie Yue  
yuejj@bmi.ac.cn

<sup>1</sup> Laboratory of Advanced Biotechnology & State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Biotechnology, Beijing 100071, China

<sup>2</sup> Medical College of Guizhou University, Guiyang 550025, China

<sup>3</sup> Guizhou Provincial People's Hospital, Guiyang 550002, China

## Introduction

Lassa fever (LF) is an acute viral hemorrhagic disease caused by Lassa virus (LASV) [1], with a case fatality rate of approximately 30% in hospitals in Nigeria [2]. LF is widespread in West Africa, particularly in Nigeria and the Mano River Union (MRU), which includes Guinea, Liberia, and Sierra Leone. In the past two decades, cases have also been reported in Mali, Côte d'Ivoire, Ghana, Benin, and Togo [3]. LASV is a single-stranded RNA virus with the genome consisting of two segments, the large (L) segment encoding RNA-dependent RNA polymerase (LP) and zinc-binding protein (ZP), and the small (S) segment encoding the glycoprotein precursor (GPC)



and nucleoprotein (NP) [4]. LASV belongs to the genus *Mammarenavirus* and the family *Arenaviridae*, with its reservoir is a rodent of the genus *Mastomys* known as “multimammate rat”. And LASV is transmitted to humans mainly through food or household items contaminated by infected rats’ urine and faeces.

Early studies identified at least seven lineages: lineages I to III were circulating in Nigeria, while lineage IV was found in MRU [5]. The strains isolated from Mali and Côte d’Ivoire were proposed as lineage V [6]. The new strain (*Hylomyscus pamfi*) isolated from Nigeria, as well as the strains from nosocomial outbreaks in Togo and Benin, may be considered lineages VI and VII [7–9]. Subsequent studies further refined the sub-lineages within each lineage [10, 11]. While the origin of LASV had been reported [12], comprehensive information regarding its spread in West Africa and the phylodynamic characteristics is still limited.

Different strains could exchange genetic information through intergenetic reassortment and recombination, leading to the generation of new genotypes and phenotypes, thereby driving viral evolution [13]. LASV belongs to the *Arenaviridae* family, where reassortment and recombination events had been detected, indicating their role in the evolutionary process within this viral family [14, 15]. Although some studies had mentioned the reassortment strains of LASV [11, 12], a global description of recombination events was lacking.

As the sole antigen on the viral surface, GPC was commonly used as an immunogen in vaccine development. The GPC virion forms a trimer, consisting of the receptor-binding subunit GP1 and the transmembrane fusion-mediating subunit GP2. It also contains a stable signal peptide (SSP), which remains as part of the complex in the virus particles. Additionally, GPC contained 11 N-glycosylation motifs, which combine with nearby amino acids to form glycans that affect the immune response [16]. While the study mentioned the variation range of GPC among lineages [17], the specific substitutions were not described in detail.

Based on the above information, we comprehensively reconstructed the spatio-temporal and phylodynamics of LASV in West Africa using as many sequences as possible from public databases, and detected possible reassortment and recombination events in sequences. Finally, we compared the variations of GPC among lineages to gain a global understanding of LASV.

## Materials and methods

### Data preparation and alignments

All sequences up to 2018 were downloaded from the National Center for Biotechnology Information (NCBI), resulting in 1322 initial genome sequences. After

removing sequences with lengths less than 95% and incomplete strain attribute information (collection time and location), a set of 500 genome sequences was finally obtained (Additional file 1). These nucleotide sequences were aligned as amino acids using MAFFT v7.490 [18]. Subsequently, the aligned amino acid sequences were converted back to nucleotide sequences to maintain codon alignment.

IQ-TREE v2.0.3 [19] was utilized to construct maximum likelihood (ML) trees for the genomic sequences. A bootstrap value of 1000 was employed to assess the reliability of the ML tree and obtain the best possible score.

### Phylogeographic reconstruction

To reduce computational complexity during phylogeographic reconstruction, cd-hit v 4.8.1 [20] with a threshold of 0.98 was used to select representative strains. The goal was to ensure that there was at least one strain in each branch of the ML trees, as well as in each time, location, and host category. This resulted in 151 L segment sequences that met these criteria, and their corresponding S segment sequences were selected as well.

Phylogeographic inferences were performed using BEAST v1.10.4 [21] with the BEAGLE library [22] to improve computational efficiency. The substitution processes were modelled according to a HKY parametrization [23], and Bayesian skyline plot (BSP) coalescent model was specified as tree topology prior [24]. Relaxed clock models with rates drawn from an underlying log-normal distribution [25] were applied to both segments. Markov chain Monte Carlo chains were run for more than 300 and 200 million iterations for segments L and S, respectively. The chains were sampled every 20,000 generations, and other parameters used default values. Convergence and mixing properties were inspected using Tracer v1.7.2 [26] to ensure that estimated sampling size values associated with estimated parameters were all great than 200. Maximum clade credibility (MCC) trees were summarized using TreeAnnotator v1.10.4 [21] from 10,000 trees regularly sampled from each posterior distribution. The resulting trees were visualized using Figtree v1.10.4 [21], and the dynamic process was visualized using SPREAD3 [27].

### Reassortment and recombination detection

To detect recombination events in the genomic sequences, RDP, GENECONV, BOOTSCAN, CHIMAERA, and MAXCHI of RDP5 [28] were employed. Only recombination events longer than 300 bp and detected by at least three of the five methods with a *P*-value cutoff of 0.05 were considered significant [5]. SimPlot [29] was subsequently used to describe

inter-lineage recombination and identify breakpoints with their parent strains.

If a breakpoint was detected at the junction of two segments in the genomic sequences, it would be considered a reassortment event. Further ML trees were constructed to determine whether the topological structure of the strains in the two segments was consistent, and the tangram analysis was performed by “matplotlib” in Python.

#### Variation feature evaluation

GPC amino acid sequences were retrieved from NCBI. Sequences with a length of less than 95% of the full length were excluded, resulting in a final set of 621 GPC amino acid sequences. The Pinneo strain was selected as the reference sequence for comparison with other strains to assess variation features in GPC. The Shannon entropy of GPC was calculated using a Python script written by Joe R. J. Healey (<https://github.com/jrjhealey/bioinfo-tools/blob/master/Shannon.py>). Then, the LASV GPC structure file was downloaded from the RCSB Protein Data Bank.

## Results

### Reconstruction of the spatio-temporal dynamics of LASV propagation

The time-scaled phylogeographic MCC trees of 151 LASV sequences and the root state posterior probability are illustrated in Fig. 1. Both LASV segments originated in Nigeria, and further results showed that the estimated the time of the most recent common ancestor (TMRCA) for the root of the L segment was approximately 634 (95% HPD: [385879]), whereas the TMRCA of the S segment was around 1224 (95% HPD: [10301401]), other nodes are shown in Table S1. Compared to other study [12], we had more comprehensively considered representative strains from seven lineages to estimate the older root TMRCA of LASV. Overall, the results still showed that the root TMRCA of L segment was earlier than the S segment.

To further our understanding of LASV, we estimated the geographic diffusion characteristics of LASV in West Africa based on MCC trees. The result indicated that LASV had two main transmission routes in West Africa (Fig. 2). Route 1 originated from Nigeria, passed through Benin, and ultimately reached Togo. While route 2 also started from Nigeria, spread westward to Liberia, and then spreads through Liberia to Guinea, Sierra Leone, Mali, Côte d’Ivoire and Ghana. The transmission patterns of the two segments in LASV were similar, but the transmission time before each location of the S segment was later than that of the L segment. This may be related to the fact that the root TMRCA of the S segment is later than that of the L segment, which also suggests that they

may be facilitated by the transmission of two independent lineages [30, 31]. Additionally, we discovered that the strains from Côte d’Ivoire had distinct origins for the S segment and the L segment. The Côte d’Ivoire strains in the L segment were connected to the strains from Mali and Guinea, whereas in the S segment, they were solely linked to the strains from Mali. This suggested that during the ongoing transmission process, the imported L segment sequences in Côte d’Ivoire may have been a result of recombination involving strains from Mali and Guinea.

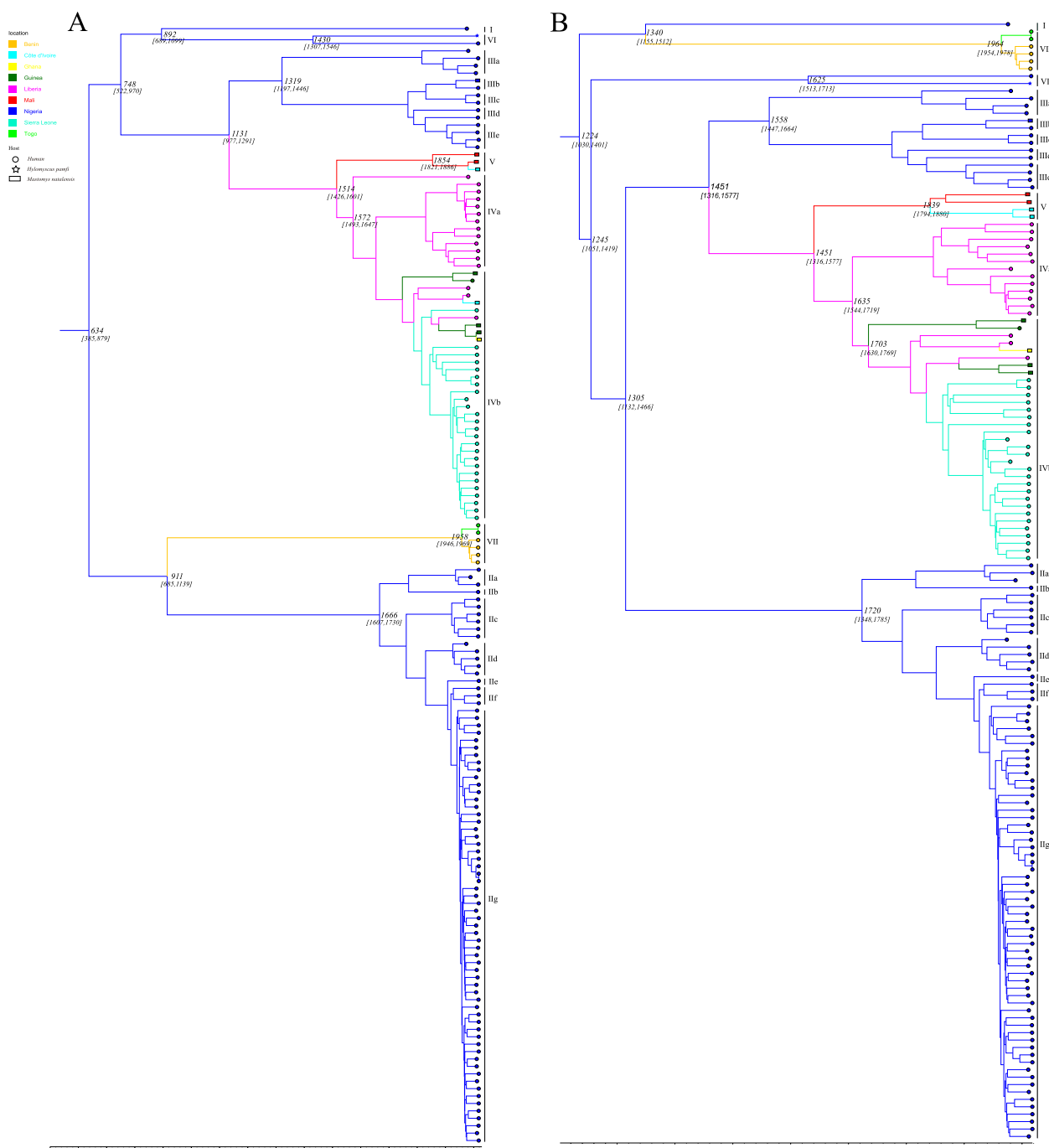
### Estimation of the phylogeographic features of LASV

To investigate the genetic diversity changes during the migration of LASV in West Africa, we employed BSP to reconstruct its demographic history. Figure 3A showed the BSP reconstruction of the L and S segments in which temporal changes of effective population size was plotted. For the L segment, the effective population size underwent a nearly stable period from 1970 to 1975. From approximately 1975, the first rapid increase in genetic diversity started, and around from 1985 to 2000, the effective population size stabilized again. Then, starting from 2000, the second rapid growth occurred and reached its maximum level around 2005, lasting for more than 10 years. Although there has been a decline since 2015, the population size during this period remained higher than during the first stage. Analysis of the BSP for S segment showed that except for slight differences in the first rapid increase period, the trends and time of other demographic patterns were similar.

Due to the observation of inconsistent root TMRCA between the two segments of LASV, we calculated the evolutionary meanRate of two segments and found the evolutionary meanRate of the L segment was faster than that of the S segment (Fig. 3B). Further analysis of the proteins they encoded revealed that ZP had the fastest evolution meanRate, with the rate almost 1.5 times that of other encoded proteins (Fig. 3C). This was consistent with the result described by O. K. Oloniniyi [32], which documented the fastest evolution rate of ZP within lineage II.

### Reassortment and recombination affect the lineage expansion of LASV

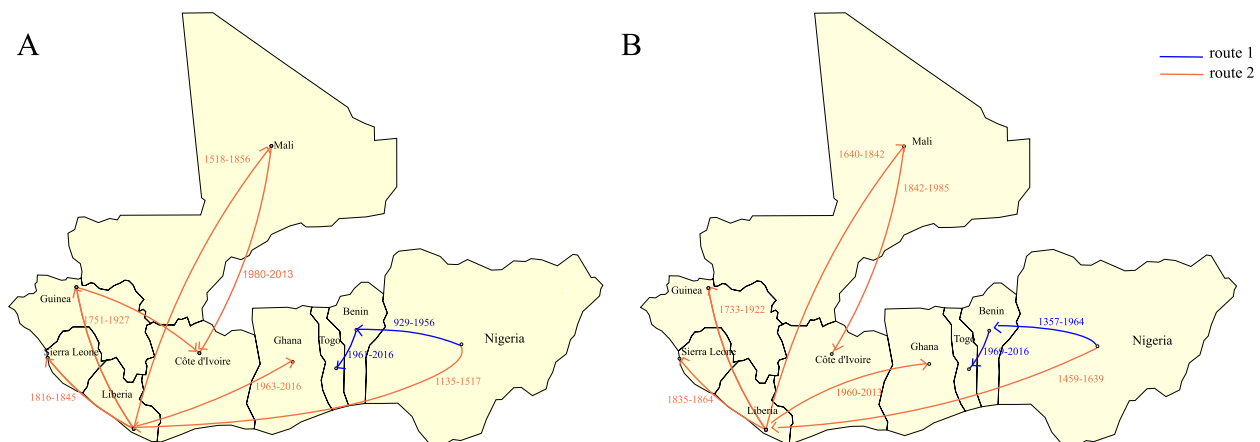
Previous reports had indicated that recombination events may lead to strains exhibiting different evolutionary branches in phylogenetic tree [15], and our MCC trees also revealed variations in the topological structures of LASV strains in the two segments. To investigate the underlying reasons for this, we conducted relevant detections, and the specific results were presented in Table 1.



**Fig. 1** BEAST analysis of complete or almost complete L and S segment nucleotide sequences. The resulting trees for the L segment (A) and S segment (B) were visualized using Figtree v1.10.4. The node ages, in years, are included on the major nodes, with the 95% confidence ranges displayed in parentheses below the median node ages. The isolates are grouped by their lineages, as represented by the bars to the right of the trees. The reverse axis represents the age, in years, from the most recent isolate. Colours indicate different geographical origins

In the whole genome, nine strains showed significant reassortment events. The first four strains all belong to lineage V, and their major and minor parents mainly came from lineage IV and lineage V. As for the

remaining strains, they all came from the sub-lineage IIg, and the detected results also showed that their parent strains also isolated from the same sub-lineage. And the strains LASV250 and LASV263 had been proven to be



**Fig. 2** Temporal dynamics of LASV geographic dispersal in West Africa. The plot simply shows the geographic diffusion model of the L segment (A) and S segment (B) of the LASV over time. The blue line represents route 1, and the orange line indicates route 2. The number displays the time of diffusion from country 1 to country 2, and they are all within the 95% confidence interval of TMRCA

reassortment strains [12], indicating that there was reassortment phenomena within the sub-lineage IIg. Figure 4 further demonstrated that in addition to these strains, there may also be some strains that had undergone reassortment, especially the LASV003 strain, which showed strong reassortment signals in two methods. Additionally, it was worth noting that the strain of lineage VII did not show any reassortment signals in detection by RDP5, but in Fig. 4, it clearly showed its distinct topological difference in the two segments, which may also be related to potential reassortment effects [33].

Then, we detected the occurrence of recombination within the segment. The results showed that no recombination signals were detected in the S segment, while in the L segment, four strains isolated from Côte d'Ivoire showed significant recombination signals (Table 1). The LASV\_3625 strain belonged to lineage V, and its major parent (BA366) came from lineage IV, while its minor parent (LASV\_3604) was from lineage V. Figure 5A showed before the recombination breakpoint (6661), the similarity between this strain and BA366 reached almost 100%, while after the breakpoint, the similarity with LASV3604 was higher. LASV\_3713 and LASV\_3715 was similar to the LASV\_3625, with both parents from lineage IV and lineage V, and the similarity before breakpoint remained around 100% with the major parent

(Figure S2A). The last strain (LASV\_3523) was different from the above strains, its major parent (Mad39) and minor parent (LASV\_3604) were isolated from lineage V, and before the breakpoint, the similarity with the major parent was around 90% (Figure S2B). The four detected strains all isolated from lineage V, and their respective parental results showed that some strains of lineage IV and lineage V recombined, leading to the formation of most strains in Côte d'Ivoire and further exacerbating the expansion of lineage V.

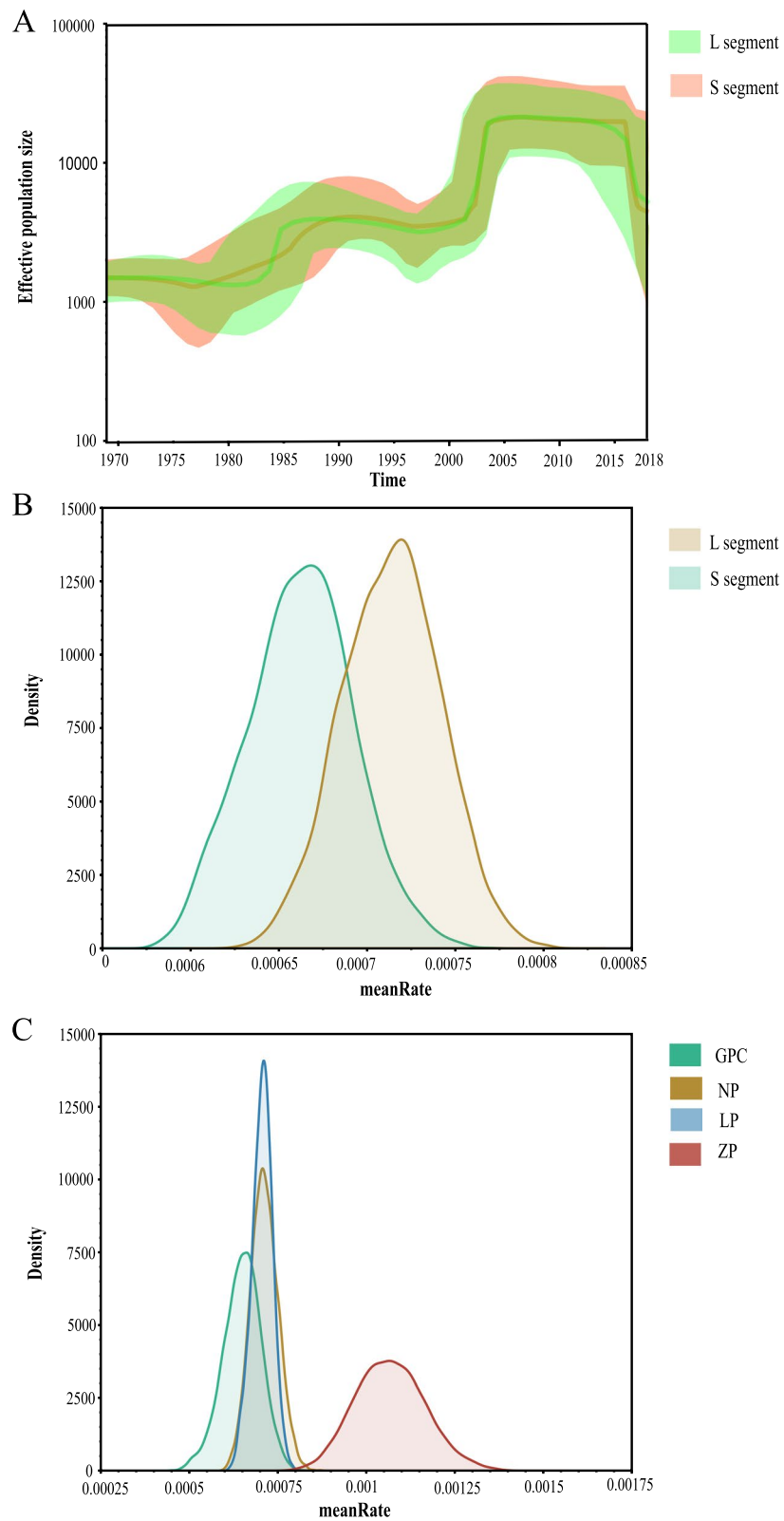
Previous study showed the intragenic recombination could play a role in driving the evolution of virus [34]. However, it was unclear whether intragenic recombination could contribute to gene mutations in LASV. Here, we conducted recombination detection on four encoding proteins, but no recombination signals were detected, which indicating that intragenic recombination did not contribute to the evolution of LASV.

#### GPC variations among lineages showed the diversity and antigen change of LASV

GPC was the main surface membrane protein of LASV, and it mediated the virus fusion [35]. Research had shown that inter-lineage variations in the GPC ranging from 4.9% to 11.0% [17]. To understand these changes, we compared the sequences of 621 GPC and identified 55

(See figure on next page.)

**Fig. 3** Phylogenetic features of LASV in West Africa. **A** Effective population size (genetic diversity) of the spatial temporal reconstruction of LASV. The thick solid line indicates the median value, and the dark area is the 95% HPD of the genetic diversity estimates. The L segment and S segment are represented in green and light red, respectively. **B-C** meanRate estimation of LASV. **B** Grass green represents the S segment (5.366E-4), and the other segment is light yellow (5.9831E-4); **C** GPC (4.9877E-4), NP (5.4523E-4), LP (5.9939E-4) and ZP (9.857E-4) are expressed in light green, light yellow, light blue and light red, respectively. The horizontal axis indicates the estimated mean rate, and the vertical axis represents the corresponding population density



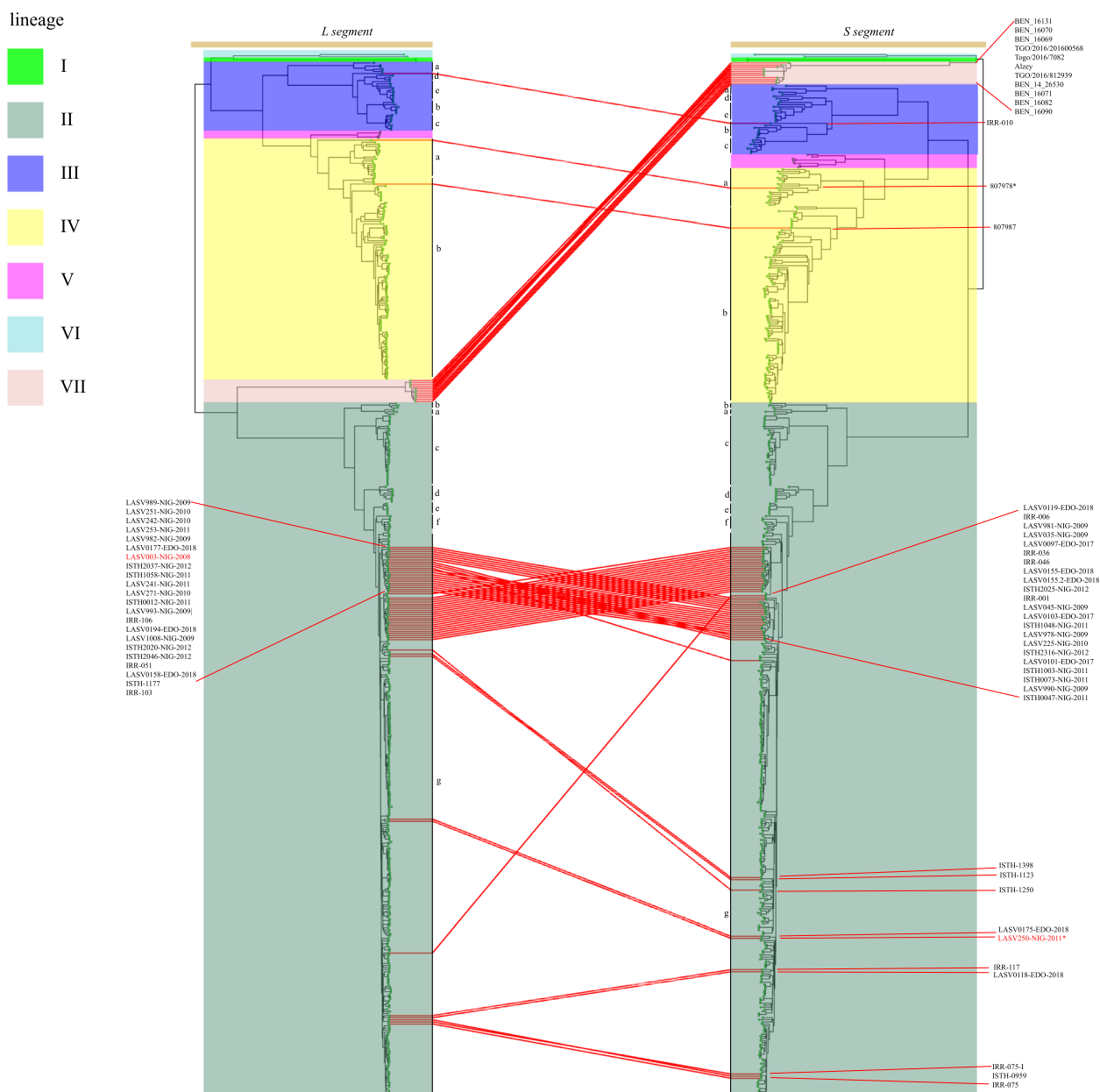
**Fig. 3** (See legend on previous page.)



**Table 1** Detected reassortment and recombination strains by RDP5

Type	Strain	Major parent	Minor parent	P value		GENECONV	Bootscan	Maxchi	Chimaera
				RDP					
reassortment	LASV_3625_V	BA366_IVb	LASV_3604_V	8.116E-213		2.37E-167	8.95E-201	2.74E-58	2.86E-54
	LASV_3523_V	Mad39_IVb	LASV_3604_V	3.27E-190		1.05E-139	NS	8.71E-59	4.96E-53
	LASV_3713_V	G1529_IVb	Mad63_IVb	3.32E-160		2.09E-105	5.48E-170	2.45E-54	1.30E-22
	LASV_3715_V	G2431_IVb	Mad69_IVb	6.42E-156		4.59E-108	1.63E-179	8.40E-56	7.49E-23
	LASV250_Ilg [12]	IRR-006_Ilg	LASV1015_Ilg	2.21E-50		5.84E-46	2.51E-53	1.68E-21	5.89E-19
	IRR-016_Ilg	LASV0070_Ilg	ISTH1096_Ilg	2.49E-36		5.24E-04	1.87E-23	3.48E-24	2.76E-19
	LASV0159_Ilg	ISTH-0664_Ilg	LASV0104_Ilg	2.80E-07		4.06E-02	6.81E-05	4.37E-06	1.08E-06
	LASV003_Ilg	LASV241_Ilg	LASV242_Ilg	1.19E-06		NS	NS	1.52E-05	6.33E-06
	LASV263_Ilg [12]	LASV975_Ilg	IRR-017_Ilg	7.64E-05		NS	NS	1.72E-04	1.44E-02
	LASV_3625_V	BA366_IVb	LASV_3604_V	9.25E-71		8.66E-57	4.44E-65	1.94E-12	3.69E-11
	LASV_3713_V	Mad63_IVb	Bamba-RT114_V	2.93E-57		1.44E-26	1.96E-54	8.56E-12	3.60E-09
	LASV_3715_V	Mad69_IVb	Bamba-RT114_V	5.20E-54		6.03E-21	6.17E-43	2.46E-12	5.78E-09
	LASV_3523_V	Mad39_V	LASV_3604_V	1.23E-50		1.14E-28	1.15E-48	2.51E-09	3.27E-07

NS No significant, P value was recorded for this recombination event using this method



**Fig. 4** Tanglegram between phylogenetic trees of LASV segments. Phylogenetic trees of the L segment and S segment were constructed using ML. Each colour represented one lineage, and the green dots at the end of the tree were strains. And those strains with different topological structures in the two segments were connected by red lines

amino acids with significant changes (Fig. 6, Additional file 2).

The substitution of isoleucine with valine occurred in I5, I31, I38, I43 and I44 of SSP, where I5 located in the N-terminal myristoylation motif and I31, I43, and I44 located in two helices. The degree of variation of these amino acids varied among lineages. Taking I5 as an example, it underwent mutations in lineage II-VII, and the variation rates were all above 50%. Mutational

analysis of GP1 showed that only S90N\T substitutions happened in N-glycosylation region, while asparagine replaced serine as the main amino acid in the lineage II-VII. It was worth noting that in lineage VI, threonine specifically appeared and was associated with strains isolated from humans. For GP2, the substitution of amino acids primarily appeared in N-terminal fusion peptide, heptad repeat and so on. The changes of GPC among lineages demonstrated the sequence diversity of LASV, and these





**Fig. 5** Similarity between recombined strains and their parents. The strains LASV\_3625 (A) and LASV\_3523 (B) show their own similarity with their parents. The major parent and minor parent are represented by the green line and blue line, respectively. The red coordinate points represent the breakpoint position and the similarity

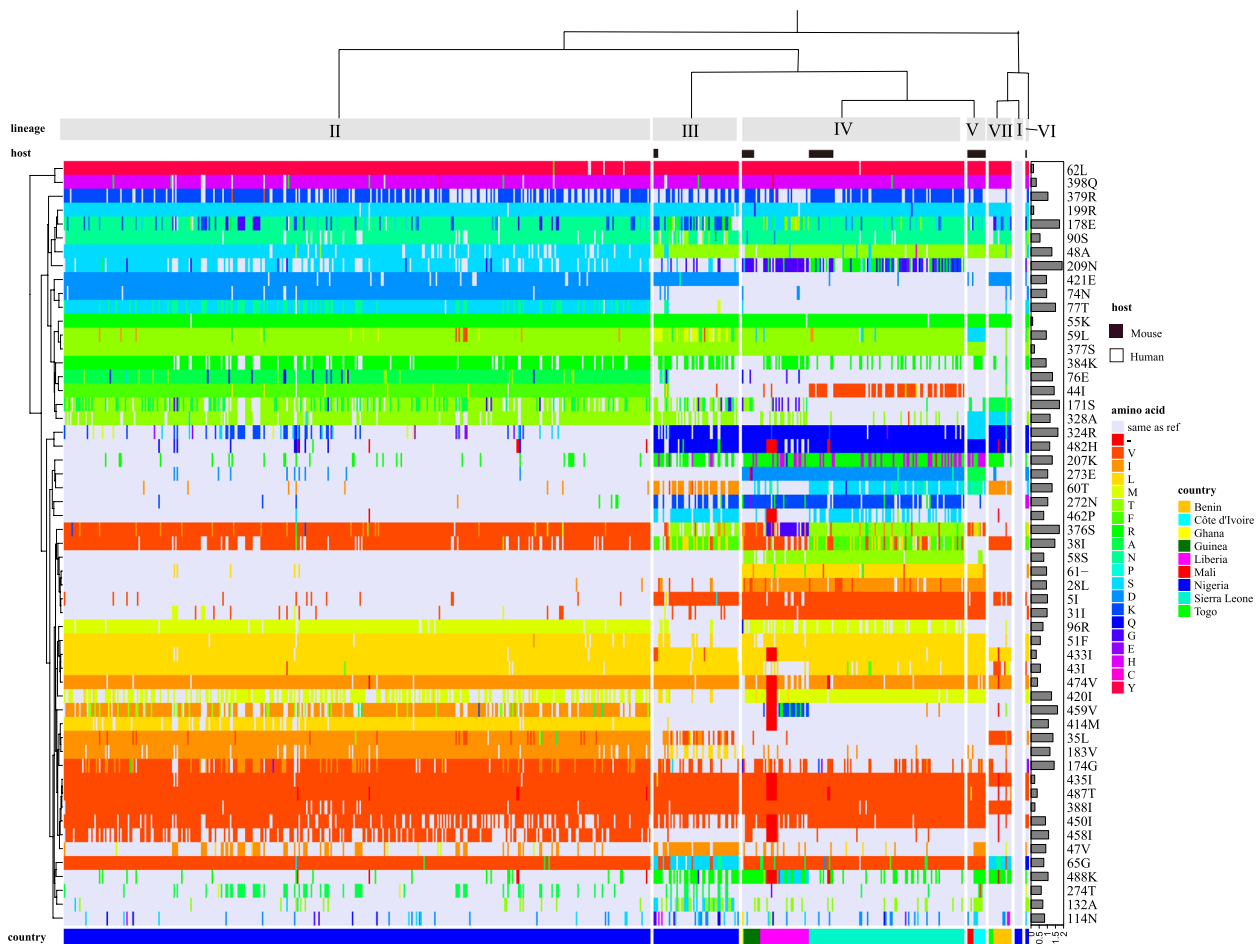
changes mainly occurred in functional regions related to membrane fusion, which was highly likely to affect the regulation of related functions [36, 37].

We also found amino acid substitutions in the known epitopes (Table 2). E1 (epitope 1) was reported to be the best predicted LBL epitope in GPC [38], but in our results, more than half of the amino acids in E1 showed changes among lineages. Spatial structure analysis showed E1 and E2 were located on the GPC trimer's surface, while E3 was present in heptad repeat 1 and the membrane proximal external region in GP2 (Fig. 7). E2 and E4 were located on the surface of GPC, and they both contained I112 and N114, indicating that they had a chance to bind with antibodies. E5 and E6 located in the membrane fusion region [39, 40] and antibody binding region [41], respectively. When changes occurred, the related functions were highly likely to be affected. These variations in amino acid mainly happened in lineage II, which led to the development and application of broad-spectrum vaccines.

## Discussion

Spatial temporal dynamics analysis showed LASV had two transmission routes in West Africa, and the reasons for these seem to be influenced by natural reservoir

spillovers and certain human activities. The natural host of LASV was mainly the *Mastomys natalensis*, which was semi-commensal habit with humans and had the characteristic of seasonal reproduction. A previous study showed that *Mastomys natalensis* reached its breeding peak at the end of the rainy season in West Africa [42]. When the rodent density increased, the ratio of their activity frequency to the crowd concentration area in the dry season also increased accordingly. This seasonal activity increased the probability of human infection, causing concentrated outbreaks of LASV in the dry season [2]. Moreover, the Triangular Trade route in the sixteenth century approximated the special and temporal traits of transmission route 2 and the armed conflicts in recent years also played an important role in the spatial distribution of LASV [43]. These human activities facilitated the incidental transport of rodents over geographical barriers, which may have been one of the reasons for the widespread spread of LASV in West Africa. Michael R once proposed that Liberia was the entry point of LASV into the MRU [11]. In route 2, the outbreak of the epidemic in West African countries was more closely linked to Liberia. Nigeria, as the source epicentre of LASV, spread the virus outward to Liberia, and then independently spread from Liberia to surrounding



**Fig. 6** GPC variation characteristics of LASV. The amino acid variation of GPC in different lineages is shown. Different colours on the clustering tree represent different amino acids, and the “lavender” represents the same amino acid as the reference strain. The colours outside the cluster tree represent different countries, hosts and lineages accordingly. The bar graph on the right shows Shannon entropy at these sites

countries, resulting in a wider spread of LASV in West Africa. This implied Liberia may be a secondary epicentre for the outbreak of LASV on the West Africa.

Not all the family of *Arenaviridae* contained a bi-segmented genome, and recently, the new virus with tri-segmented genomes had been detected [44], indicating the generation of new segments may have been through long-term evolution [45]. This may also be one of the reasons for the inconsistent root TMRCA of the two segments of LASV. From 1969 to 2018, several heavy rainfall events occurred in West Africa, especially in Nigeria in 1988. The increase in rainfall affected the host’s habits [43], thereby affecting the spread of LASV. The increase in genetic diversity during this period may be related to the cumulative effect of rainfall [46]. The virus lacking ZP was found in family of *Arenaviridae*, indicating that arenaviruses may have evolved a new mechanism that could replace ZP [44]. The high evolution rate of ZP in LASV suggested that LASV may also be

undergoing similar evolution. And other study had shown that the high evolution rate may be related to its short length [47]. Thus, when considering ZP as a drug target [48], it was necessary to consider the potential risks.

Two reported strains [12] indicated that there had been reassorted within the sub-lineage IIg of LASV, and more reassortment signals further suggested that reassortment had been still ongoing. In the whole genome sequences obtained in Côte d’Ivoire, four potential reassortment strains were identified. Further investigations revealed that the occurrence of reassortment signal of LASV\_3625 may be attributed to recombination events within L segment. For the L segment, four strains isolated from Côte d’Ivoire also showed significant recombination signals. The major and minor parents of these strains come from the lineage IV and V, respectively, indicating that these strains were produced by of recombination of two lineages. The strains of the Côte d’Ivoire branch were highly

**Table 2** Amino acid variations with known epitopes

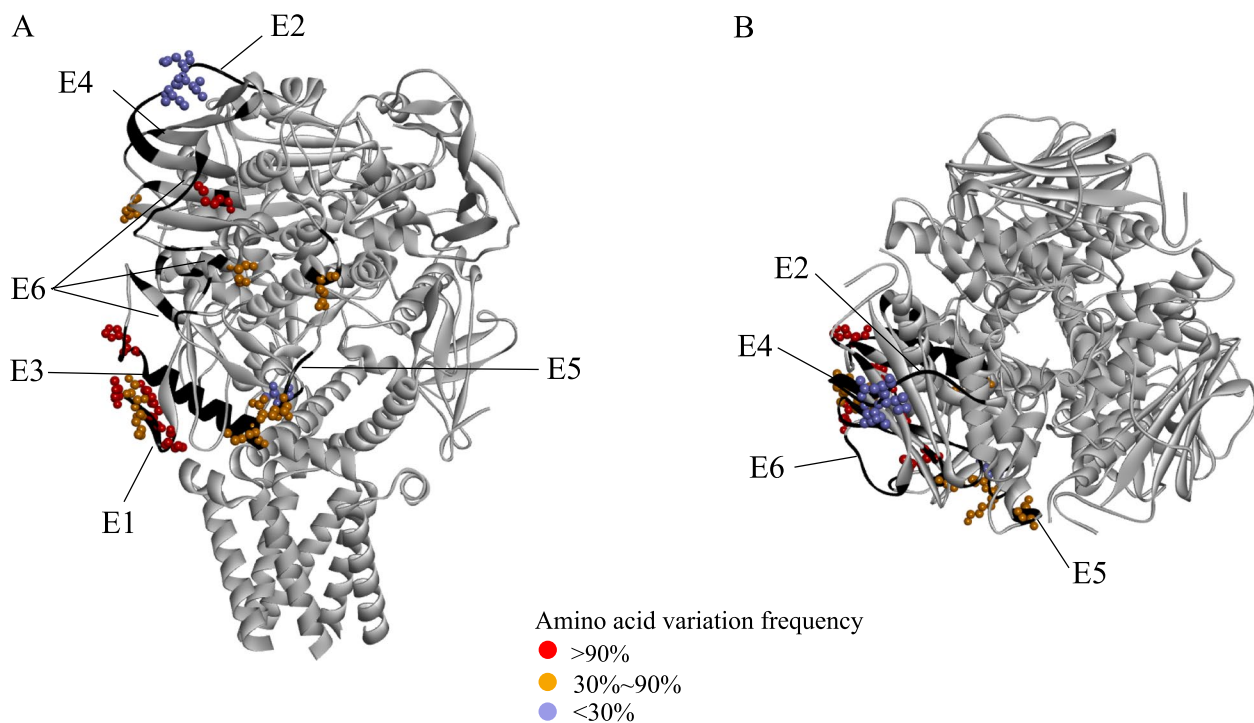
Epitope	Position and sequence	Mutation
E1	54-65 GKSCSLT-LKGG	55K: II: R (100%); III: R (100%); IV: R (99%); V: R (100%); V: R (100%); VI: R (100%); VII: R (100%)
		58S: IV: T (93%); V: T (25%)
		59 L: II: T (95%); III: M (41%); IV: T (95%); V: S (100%); VI: T (50%), A (50%)
		60T: III: I (64%); IV: S (58%); V: N (75%); VI: S (50%); VII: V (94%)
		61-: IV: L (98%); V: L (100%); VI: L (50%)
		62 L: II: Y (99%); III: Y (100%); IV: Y (99%); V: Y (100%); VI: Y (100%); VII: Y (100%)
		65G: II: V (99%); III: S (73%); IV: V (93%); V: V (100%); VI: S (100%); VII: S (73%)
		112I: V: L (50%)
		113I: V: L (33.3%); VII: L (20%)
		114 N: V: S (25%); VI: S (50%)
E2	110-119 TSIHNHKFCN	
E3	401-415 DDIEQQADNMITEML	414 M: II: L (90%)
E4	GP1: S91, H92, L107, T108, N109, T110, S111, I112, N114, K161, Y217, Y219	112I: V: L (50%)
		114 N: V: S (25%); VI: S (50%)
E5	GP2: S269, E270, G271, N272, E273, T274, P275, R324, R325, K327, R356, C360, L387, S389, N390, N395, E396, T397, Q398, F399, S400, D401, D402, E404, Q405, D408	272 N: III: K (50%); IV: K (84%); VI: K (100%)
		273E: IV: D (97%); V: A (100%)
		274T: III: A (23%); IV: A (25%); V: A (50%)
		324R: III: Q (79%); IV: Q (92%); V: S (100%); VI: Q (100%); VII: Q (93%)
		398Q: II: H (99%); III: H (98%); IV: H (98%); V: H (100%); VI: H (100%); VII: H (100%)
E6	GP1: N74, E76, K88, R96, T226, W227, E228, D229, H230, C231, Q232, R235, P236; GP2: I286, E287	74 N: II: D (98%); VI: D (50%)
		76E: II: A (93%)
		96R: II: M (96%); IV: M (78%)

The epitope was the Josiah strain, but in our results, the amino acid at the mutation site was artificially replaced by the amino acid of the Pinneo strain to unify the mutation results. The red highlighted amino acid indicates the mutated amino acid

likely to be affected by recombination, thereby exacerbating the expansion of lineage V.

The diversity of LASV could be reflected through GPC, which served as a surface antigen and often was used as a candidate immunogen in vaccine development [35].

Multiple residues substitutions also emerged in GPC. At E178, each lineage contained at least two mutated amino acids, and the continuous substitutions of amino acids was likely related to their positive selection or high Shannon entropy of GPC [49]. Analysis of known epitopes



**Fig. 7** Global amino acid variation in epitopes mapped onto the protein crystal structure of the GPC trimer. Epitopes were plotted in the 7puy model, and each epitope is represented in black and marked in the figure. The variation sites on epitopes used different colours to represent the variation rate. See the description in the figure for the specific change rate

showed that they were also undergoing amino acid substitutions. These changes in E2 and E4 sequences were mainly concentrated in lineages V to VII, while they were relatively conserved in the I-IV lineages. This special variation among lineages may be beneficial for them as candidate epitopes for specific vaccines in lineages I to IV.

## Conclusions

The results of this study investigated the temporal and spatial dynamics of LASV transmission in West Africa and its associated factors. We have highlighted the crucial role of Liberia as a secondary epicenter and estimated relevant evolutionary features such as TMRCA, effective population size, and evolution rate. Relevant results suggested that multiple lineages contained reassortment events, especially the sub-lineage IIg, while recombination played a role in expanding lineage V. Notably, significant amino acid variations in GPC demonstrated the sequences diversity of LASV, and this diversity change may affect the normal expression of epitopes.

## Abbreviations

LF	Lassa fever
LASV	Lassa virus
MRU	Mano River Union
L	Large
LP	RNA-dependent RNA polymerase

ZP	Zinc-binding protein
S	Small
GPC	Glycoprotein precursor
NP	Nucleoprotein
SSP	Stable signal peptide
NCBI	National Center for Biotechnology Information
ML	Maximum likelihood
MCC	Maximum clade credibility
LBL	Linear B-cell lymphocyte
TMRCA	The time of the most recent common ancestor
E1	Epitope 1

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12879-024-09200-8>.

- Supplementary Material 1.**
- Supplementary Material 2.**
- Supplementary Material 3.**
- Supplementary Material 4.**
- Supplementary Material 5.**

## Acknowledgements

I am grateful to all the collectors and authors of LASV sequences, as well as the researchers who participated in this study and for all the constructive suggestions they have provided.

## Authors' contributions

XW and SYW downloaded the data, and XW and MDH analysis the data. RHL support the funding. XDZ, XWY, HGR, YJ, JJX and JJY supervised. XW wrote

the main manuscript text, and YJ and JJY reviewed and edited. All authors reviewed the manuscript.

#### Funding

This work was supported by the National Natural Science Foundation of China [grant number: 82202520].

#### Availability of data and materials

All the relevant datasets relating to this study are available upon reasonable request to the co-authors.

#### Declarations

##### Ethics approval and consent to participate

This study did not involve human or animal subjects, and thus, no ethical approval was required. The study protocol adhered to the guidelines established by the journal.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 21 September 2023 Accepted: 6 March 2024

Published online: 14 March 2024

#### References

- Okokhere P, Colubri A, Azubike C, et al. Clinical and laboratory predictors of Lassa fever outcome in a dedicated treatment facility in Nigeria: a retrospective, observational cohort study. *Lancet Infect Dis*. 2018;18(6):684–95.
- Asogun DA, Adomeh DI, Ehimuan J, et al. Molecular diagnostics for lassa fever at Irrua specialist teaching hospital, Nigeria: lessons learnt from two years of laboratory operation. *PLoS Negl Trop Dis*. 2012;6(9):e1839.
- Bell-Kareem AR, Smither AR. Epidemiology of Lassa Fever. *Curr Top Microbiol Immunol*. 2023;440:87–109.
- Auperin DD, Romanowski V, Galinski M, et al. Sequencing studies of pichinde arenavirus S RNA indicate a novel coding strategy, an ambisense viral S RNA. *J Virol*. 1984;52(3):897–904.
- Forni D, Sironi M. Population structure of Lassa Mammarenavirus in West Africa. *Viruses* 2020;12(4):437.
- Manning JT, Forrester N, Paessler S. Lassa virus isolates from Mali and the Ivory Coast represent an emerging fifth lineage. *Front Microbiol*. 2015;6:1037.
- Whitmer SLM, Strecker T, Cadar D, et al. New lineage of Lassa virus, Togo, 2016. *Emerg Infect Dis*. 2018;24(3):599–602.
- Yadouleton A, Agolinou A, Kourouma F, et al. Lassa virus in Pygmy mice, Benin, 2016–2017. *Emerg Infect Dis*. 2019;25(10):1977–9.
- Olayemi A, Cadar D, Magassouba N, et al. New hosts of the lassa virus. *Sci Rep*. 2016;6:25280.
- Ehichioya DU, Dellicour S, Pahlmann M, et al. Phylogeography of lassa virus in Nigeria. *J Virol*. 2019;93(21):e00929–19.
- Wiley MR, Fakoli L, Letizia AG, et al. Lassa virus circulating in Liberia: a retrospective genomic characterisation. *Lancet Infect Dis*. 2019;19(12):1371–8.
- Andersen KG, Shapiro BJ, Matranga CB, et al. Clinical sequencing uncovers origins and evolution of lassa virus. *Cell*. 2015;162(4):738–50.
- Sevilla N, de la Torre JC. Arenavirus diversity and evolution: quasispecies in vivo. *Curr Top Microbiol Immunol*. 2006;299:315–35.
- Riviere Y, Oldstone MB. Genetic reassortants of lymphocytic choriomeningitis virus: unexpected disease and mechanism of pathogenesis. *J Virol*. 1986;59(2):363–8.
- Lukashevich IS. Generation of reassortants between African arenaviruses. *Virology*. 1992;188(2):600–5.
- Koma T, Huang C, Coscia A, et al. Glycoprotein N-linked glycans play a critical role in arenavirus pathogenicity. *PLoS Pathog*. 2021;17(3):e1009356.
- Hallam HJ, Hallam S, Rodriguez SE, et al. Baseline mapping of Lassa fever virology, epidemiology and vaccine research and development. *NPJ Vaccines*. 2018;3:11.
- Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2019;20(4):1160–6.
- Lanfear R, von Haeseler A, Woodhams MD, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530–4.
- Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics*. 2001;17(3):282–3.
- Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007;7:214.
- Ayres DL, Cummings MP, Baele G, et al. BEAGLE 3: improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. *Syst Biol*. 2019;68(6):1052–61.
- Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect Math Life Sci*. 1986;17:57–86.
- Flouri T, Jiao X, Rannala B, et al. A bayesian implementation of the multi-species coalescent model with introgression for phylogenomic analysis. *Mol Biol Evol*. 2020;37(4):1211–23.
- Penny D, Drummond AJ, Ho SYW, et al. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006;4(5):e88.
- Rambaut A, Drummond AJ, Xie D, et al. Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst Biol*. 2018;67(5):901–4.
- Bielejec F, Baele G, Vrancken B, et al. Sprea D3: interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol Biol Evol*. 2016;33(8):2167–9.
- Martin DP, Varsani A, Roumagnac P, et al. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evolution*. 2021;7(1):veaa087.
- Lole KS, Bollinger RC, Paranjape RS, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol*. 1999;73:152–60.
- Bowen MD, Rollin PE, Ksiazek TG, et al. Genetic diversity among Lassa virus strains. *J Virol*. 2000;74(15):6992–7004.
- Ehichioya DU, Hass M, Becker-Ziaja B, et al. Current molecular epidemiology of Lassa virus in Nigeria. *J Clin Microbiol*. 2011;49(3):1157–61.
- Oloniyi OK, Unigwe US, Okada S, et al. Genetic characterization of Lassa virus strains isolated from 2012 to 2016 in southeastern Nigeria. *PLoS Negl Trop Dis*. 2018;12(11):e0006971.
- Charel RN, de Lamballerie X, Emonet S. Phylogeny of the genus Arenavirus. *Curr Opin Microbiol*. 2008;11(4):362–8.
- Hoxie I, Dennehy JJ. Intragenic recombination influences rotavirus diversity and evolution. *Virus Evol*. 2020;6(1):vez059.
- Hastie KM, Zandonatti MA, Kleinfelter LM, et al. Structural basis for antibody-mediated neutralization of Lassa virus. *Science*. 2017;356(6341):923–8.
- Cao J, Zhang G, Zhou M, et al. Characterizing the Lassa virus envelope glycoprotein membrane proximal external region for its role in fusogenicity. *Virus Sin*. 2021;36(2):273–80.
- Ngo N, Henthorn KS, Cisneros MI, et al. Identification and mechanism of action of a novel small-molecule inhibitor of arenavirus multiplication. *J Virol*. 2015;89(21):10924–33.
- Sayed SB, Nain Z, Khan MSA, et al. Exploring lassa virus proteome to design a multi-epitope vaccine through immunoinformatics and immune simulation analyses. *Int J Pept Res Ther*. 2020;26(4):2089–2107.
- Glushakova SE, Lukashevich IS, Baratova LA. Prediction of arenavirus fusion peptides on the basis of computer analysis of envelope protein sequences. *FEBS Lett*. 1990;269(1):145–7.
- Glushakova SE, Omelyanenko VG, Lukashevich IS, et al. The fusion of artificial lipid membranes induced by the synthetic arenavirus "fusion peptide." *Biochim Biophys Acta*. 1992;1110(2):202–8.
- Enriquez AS, Buck TK, Li H, et al. Delineating the mechanism of anti-Lassa virus GPC-A neutralizing antibodies. *Cell Rep*. 2022;39(8):110841.

42. Coetzee CG. The biology, behaviour, and ecology of *Mastomys natalensis* in southern Africa. *Bull World Health Organ.* 1975;52(4–6):637–44.
43. Sironi M, de la Torre JC. The chameleonic genetics of Lassa virus. *Lancet Infect Dis.* 2019;19(12):1276–7.
44. Mang S, Xiandan L, Xiao C, et al. The evolutionary history of vertebrate RNA viruses. *Nature.* 2018;556:197–202.
45. Pontremoli C, Forni D, Sironi M. Arenavirus genomics: novel insights into viral diversity, origin, and evolution. *Curr Opin Virol.* 2019;34:18–28.
46. Redding DW, Gibb R, Dan-Nwafor CC, et al. Geographical drivers and climate-linked dynamics of Lassa fever in Nigeria. *Nat Commun.* 2021;12(1):5759.
47. Xu X, Peng R, Peng Q, et al. Cryo-EM structures of Lassa and Machupo virus polymerases complexed with cognate regulatory Z proteins identify targets for antivirals. *Nat Microbiol.* 2021;6(7):921–31.
48. Patricio AR, Herbst LH, Duarte A, et al. Global phylogeography and evolution of chelonid fibropapilloma-associated herpesvirus. *J Gen Virol.* 2012;93(Pt 5):1035–45.
49. Pond SL, Murrell B, Poon AF. Evolution of viral genomes: interplay between selection, recombination, and other forces. *Methods Mol Biol.* 2012;856:239–72.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.