

Evolutionary trajectory and characteristics of Mpoxy virus in 2023 based on a large-scale genomic surveillance in Shenzhen, China

Received: 8 March 2024

Accepted: 16 August 2024

Published online: 28 August 2024

Shengjie Zhang  ^{1,2,4}, Fuxiang Wang  ^{1,2,4}, Yun Peng ^{1,2,4}, Xiaohua Gong ^{1,2}, Guohao Fan ^{1,2}, Yuanlong Lin ^{1,2}, Liuqing Yang ^{1,2}, Liang Shen ³, Shiyu Niu ^{1,2}, Jiexiang Liu ^{1,2}, Yue Yin ^{1,2}, Jing Yuan  ^{1,2}, Hongzhou Lu ^{1,2} , Yingxia Liu  ^{1,2}  & Yang Yang  ^{1,2} 

 Check for updates

The global epidemic of Mpoxy virus (MPXV) continues, and a local outbreak has occurred in Shenzhen city since June 2023. Herein, the evolutionary trajectory and characteristics of MPXV in 2023 were analyzed using 92 MPXV sequences from the Shenzhen outbreak and the available genomes from GISAID and GenBank databases. Phylogenetic tracing of the 92 MPXVs suggests that MPXVs in Shenzhen may have multiple sources of importation, and two main transmission chains have been established. The combination of phylogenetic relationships, epidemiological features, and mutation characteristics supports the emergence of a new lineage C.1.1. Together with the B.1 lineage diverging from the A.1 lineage, C.1.1 lineage diverging from the C.1 lineage may serve as another significant evolutionary events of MPXV. Moreover, increasing apolipoprotein B mRNA-editing catalytic polypeptide-like 3 (APOBEC3) related mutations, higher rate of missense mutations, and less mutations in the non-coding regions have been shown during MPXV evolution. Host regulation proteins of MPXV have accumulated considerable amino acid mutations since the B.1 lineage, and a lineage-defining APOBEC3-related mutation that disrupts the *N2L* gene encoding a viral innate immune modulator has been identified in the C.1.1 lineage. In summary, our study provides compelling evidence for the ongoing evolution of MPXV with specific features.

Mpoxy (formerly named Monkeypox) is a zoonotic disease caused by mpoxy virus (MPXV)¹, with symptoms similar to smallpox and occasionally severe complications, and a reported case fatality rate of 3.6–10%^{2–5}. Historically confined to Central and West African countries, mpoxy cases have been sporadically reported in these regions, with outbreaks often linked to direct contact with infected animals, such as rodents and non-human primates^{5–7}. Since May 2022, a large-scale,

multinational epidemic of mpoxy has erupted worldwide⁸, and the World Health Organization (WHO) declared the mpoxy outbreak a Public Health Emergency of International Concern (PHEIC) on July 23, 2022⁹. As of August 2, 2024, a total of 99,176 laboratory-confirmed cases and 208 deaths in 116 countries have been reported to the WHO¹⁰. The number of confirmed mpoxy cases reached the peak in the African, Eastern Mediterranean, European, and Americas regions from

¹Shenzhen Key Laboratory of Pathogen and Immunity, Shenzhen Clinical Research Center for Infectious Disease, State Key Discipline of Infectious Disease, Shenzhen Third People's Hospital, Second Hospital Affiliated to Southern University of Science and Technology, Shenzhen, China. ²National Clinical Research Center for Infectious Disease, Shenzhen, China. ³Department of Central Laboratory, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, Xiangyang, Hubei, China. ⁴These authors contributed equally: Shengjie Zhang, Fuxiang Wang, Yun Peng.

 e-mail: luhongzhou@szsystech.edu.cn; yingxiali@hotmai.com; young@mail.sustech.edu.cn

July to October 2022, while from July to October 2023 for the South-East Asia and Western Pacific regions^{8,11}.

The MPXV genome, a linear double-stranded DNA molecule of approximately 197 Kb, encodes nearly 175 nonredundant orthologous poxvirus genes (OPGs), which are involved in various aspects of the virus's life cycle, including replication, assembly, and host interaction^{12,13}. MPXV was classified into two major clades, Clade I (formerly Central African) and Clade II (formerly West African, including IIA and IIB), with further subdivisions into many lineages based on phylogenetic analysis, which reflect the viral evolutionary history and geographical distribution^{14–16}. The analyses on the systematic evolution of MPXV reveal that the clade IIB MPXV circulating in humans may have a single origin^{15,17,18}, and further subdivides into lineages A, B, and C during the global outbreak¹⁴. Lineage A primarily includes the mpox outbreaks in Nigeria in 2017–2018 (Lineage A.1)¹⁹, as well as sporadic cases of other lineages (e.g., A.2 or A.3) detected in other non-endemic countries²⁰, indicating the existence of multiple chains of human-to-human transmission^{18,21}. Comparative genomic analysis suggests that the MPXV B.1 lineage, which caused the multi-country outbreak in 2022, may have a genomic common ancestor with the A.1 lineage from Nigeria^{15,17,21}, and has further diversified into different lineages from B.1.1 to B.1.20 worldwide (<https://nextstrain.org/monkeypox/>). Recent studies indicate that the circulating lineage of MPXV in certain Asian regions is transitioning from B.1.3 to C.1^{22,23}, while the molecular evolutionary mechanisms of MPXV during this process remain poorly understood. In 2022, researchers have observed a mutation rate in the MPXV genome that is much higher than the typical mutation rate in Orthopoxviruses¹⁷, involving C > T or G > A mutations mediated by apolipoprotein B messenger RNA (mRNA) editing catalytic polypeptide-like 3 (APOBEC3) enzymes^{17,18}. Additionally, other studies have reported a possible relationship between deoptimized codon usage in the MPXV and the observed decrease in fatality rates²⁴, as well as evolutionary differences caused by recombination during the evolution of MPXV^{22,25}. However, further research is needed to understand the ongoing evolutionary direction and mechanisms of MPXV after 2022.

According to the data from the Chinese Center for Disease Control and Prevention (China CDC), the total number of confirmed mpox cases in mainland China has exceeded 1610 between June and November, 2023 (Supplementary Fig. 1A). During this period, Shenzhen, an international port city, has also experienced a local outbreak of mpox (Supplementary Fig. 1B)²⁶. Understanding the transmission and evolution dynamics of MPXV in Shenzhen could provide valuable insights into the potential for MPXV to spread across borders and into new populations. However, due to the current global outbreak of mpox primarily occurring in the men-who-have-sex-with-men (MSM) community, it poses certain challenges for epidemiological tracing^{27,28}. Here, we obtained 92 high-quality MPXV genome sequences from the local outbreak in Shenzhen city in 2023, and investigated the transmission characteristics via integrating the epidemiological and molecular evolutionary characteristics. Moreover, in combination with the available complete MPXV genome sequences from GISAID and GenBank databases, we demonstrated the evolutionary trajectory and significant events of MPXV in 2023. This study provides new insights into the ongoing evolutionary direction of the MPXV and important information for the future control of the epidemic.

Results

Baseline characteristics of the cohort

A total of 92 laboratory-confirmed mpox patients who visited or were admitted to Shenzhen Third People's Hospital between June 9 to October 12 were enrolled in this study, with a peak of mpox cases observed during June to August in 2023 (Supplementary Fig. 1B and Supplementary Table 1). Of note, these patients accounted for approximately 30% and 66% of the total reported cases in Guangdong

province and Shenzhen city from June to October, 2023, respectively (Supplementary Fig. 1). The demographic characteristics showed that they were all male with a median age of 30.0 years (IQR 27.0–36.0 years), and 41% of them falling within the age range of 25–30 (Supplementary Table 1). Notably, 88 (95.2%) of the patients reported MSM behaviors, while corroborating epidemiological links were only found in two pairs of sexual partners (SZPMI-001 and SZPMI-003, SZPMI-003 and SZPMI-005). Moreover, nine patients reported high-risk MSM behaviors in Thailand and various locations within China (Hong Kong, Guangzhou, Jiangxi, Beijing, Wuhan, Changsha, Shenyang), indicating the possibilities for multiple importation and exportation events of MPXVs epidemiologically associated with Shenzhen (Fig. 1). About 56.5% (52/92) of enrolled mpox cases were diagnosed with the human immunodeficiency virus (HIV) infection, and 38 patients among them have received HIV anti-retroviral treatment (ART) before the diagnosis of mpox (Supplementary Table 1). Notably, skin lesions were found in multiple body sites of the patients, and genital lesions were found in 62.2% (56/90) of the patients, suggesting the high risk of transmission through sexual contact (Supplementary Table 1).

Phylogenetic characteristics of local MPXVs in Shenzhen

To characterize the genomic and molecular evolutionary phylogeny of the MPXVs involved in the outbreak in Shenzhen, whole genome sequences were obtained from the clinical samples of enrolled 92 patients. By integrating the 92 MPXV sequences from Shenzhen and 193 representative reference genomes of each lineage from the GISAID and GenBank databases, we constructed two maximum likelihood phylogenomic trees based on both the whole genome sequences and single nucleotide polymorphisms (SNPs) (Fig. 1 and Supplementary Fig. 2). The results showed that all the 92 Shenzhen MPXVs located within the lineage C.1 in the two phylogenetic trees (Fig. 1 and Supplementary Fig. 2). Then the entire C.1 lineage was retrieved and analyzed in detail. Notably, it can be further diverged into two distinct MPXV sub-lineages of C.1 and C.1.1 in both the whole genomic and the SNP phylogenetic tree (Fig. 2 and Supplementary Fig. 3). Through comparison with the reference genome of clade IIB MPXV (GISAID ID: EPI_ISL_13056282), we depicted the nucleotide mutation spectrum across various lineages of clade IIB. The clustering results of the nucleotide mutation profile aligned with the lineage classification in the phylogenetic trees, further substantiating the stability of the phylogenetic relationships and highlighting specific mutations that have occurred in C.1.1 compared to other lineages (Supplementary Fig. 4). Additionally, except for characteristic mutation sites defined in Nextstrain, a total of 13 distinct unique mutation sites for A.2 (C13298T, C30636T, C34578T, G114659A, G12403A, C149872T, G168205A), A.3 (C51809T, C125492T, C128174T, C134078T, G188889A) and C.1.1 lineages (G152866A) were identified. Particularly, the specific mutation G152866A in the C.1.1 sub-lineage which results in the Asp59Asn mutation in OPG180 protein (A50R, DNA ligase), provides further support for the emergence of the new lineage C.1.1 (Fig. 2C).

The C.1 sub-lineage was predominantly circulating in Asian regions (82 genomes) from January to June in 2023, including Japan, South Korea, Indonesia, and China (Fig. 2B), and a total of six MPXVs from China belonged to this sub-lineage. In detail, three MPXV sequences from Shenzhen (SZPMI-087, SZPMI-090, SZPMI-092) and one from Beijing (GISAID ID: EPI_ISL_18360394) located within the South Korea clusters, while one MPXV sequence from Shenzhen (SZPMI-094) together with one sequence from Hangzhou (GISAID ID: EPI_ISL_17809521) clustered with the sequences from Japan (Fig. 1). Based on the significantly later dates of diseases onset for the four mpox cases in Shenzhen, South Korea and Japan serve as the mostly possible source of these MPXVs (Fig. 2B). The sub-lineage C.1.1 was mainly circulated in Portugal and Ireland in Europe (N=27), as well as China, with a significant number of cases reported in Shenzhen (N=88) (Fig. 2B). In terms of timing, most of the isolates from this sub-

lineage were obtained between June and October in 2023. For the 88 MPXV sequences from Shenzhen in this sub-lineage, they possess closest relationship with those from Portugal in both aspects of evolutionary distance and time of disease onset. Additionally, these 88 sequences formed two genomic clusters (node id of SZPMI-016 to SZPMI-078 and SZPMI-026 to SZPMI-066 in the phylogenetic tree) in C.1.1 sub-lineage, which suggested that there might be two main transmission chains during the outbreak in Shenzhen (Figs. 1 and 2).

The emergence of the new lineage C.1.1 suggests another evolutionary event of MPXV in 2023

In order to investigate the evolutionary dynamic of MPXV among human infection during 2023, particularly the newly emerged C.1.1

lineage, we collected all high-quality genome sequences in GISAID and GenBank databases of clade IIb MPXV. Together with the 92 MPXV sequences from Shenzhen, a total of 5220 MPXV whole genomes were included in this analysis (Supplementary Data 1). Firstly, we performed dimensionality reduction by means of principal coordinates analyses (PCoA) on 7432 unique nucleotide variation data extracted from comparison with the reference genome (GISAID ID: EPI_ISL_13056282). Significant differences (PERMANOVA, $R^2 = 0.66$, $p = 0.001$) with a total explanatory power of 55% was observed in the mutation spectrum between different MPXV lineages. The entire clade IIb MPXV could be roughly divided into four distinct evolutionary process, including the earliest formed A lineage, the A.2 and A.3 lineages derived from the A lineage, the A.1 lineage derived from the A lineage, as well as the B.1 and

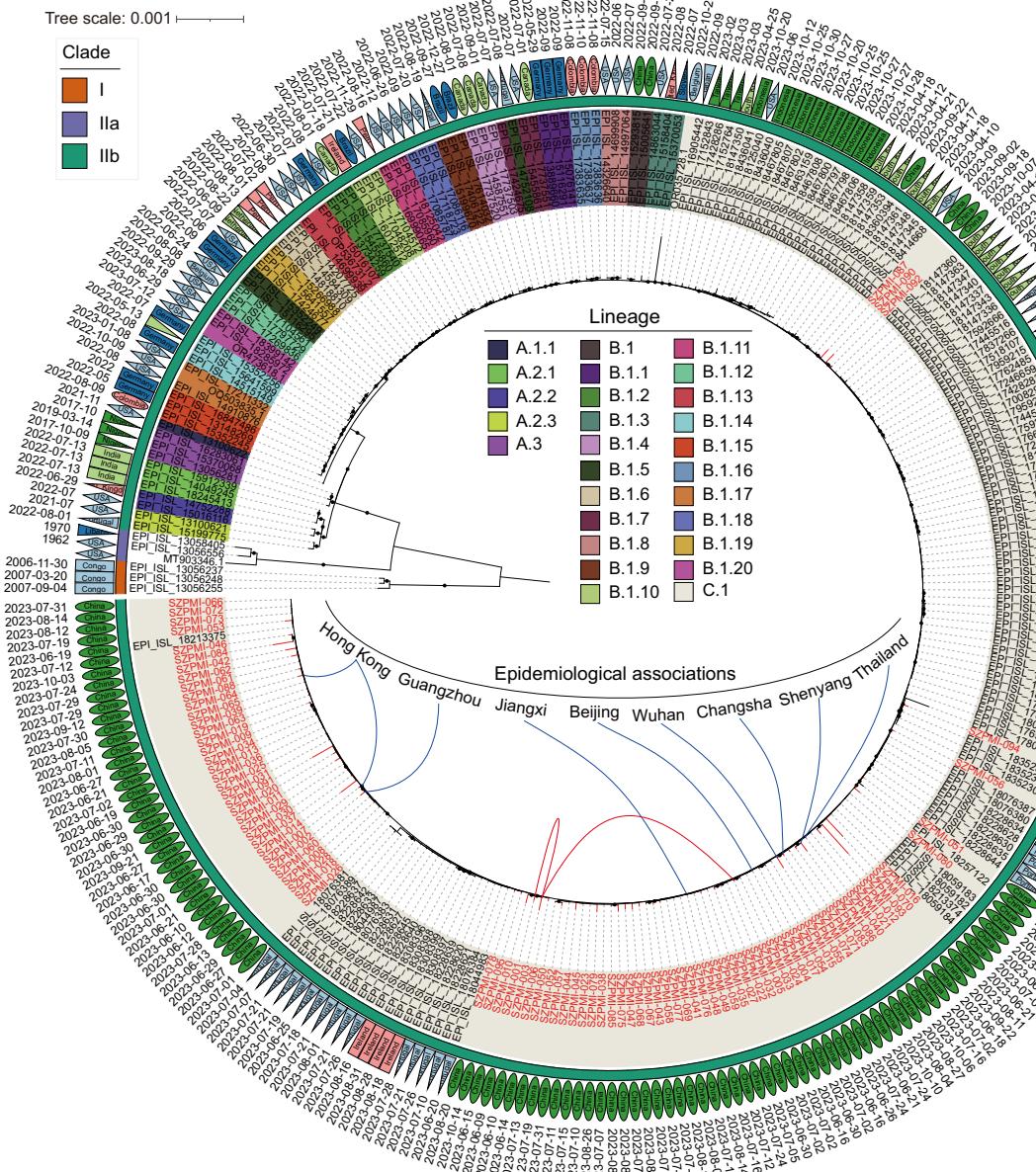


Fig. 1 | Phylogenomic tree of local MPXV genome sequences from Shenzhen in 2023. The maximum likelihood phylogenomic tree was built using whole genome alignment of 92 Shenzhen MPXV sequences with representative reference genomes of MPXV. Red branches with red labels indicate genomes from Shenzhen outbreak. Black labels in the tree indicate sequences from the GISAID or NCBI databases. The tree labels were colored by each lineage. The tree was rooted using clade I as an outgroup. Black dots in the middle of the branch represented

bootstrap support values > 75 . The scale bar indicates the number of substitutions per site. The connection lines between nodes in the tree represented the epidemiological associations. Red lines meant sexual contacts, while blue lines meant high-risk sexual behavior in travel history. The inner annotation ring illustrated three main MPXV clade. The middle annotation ring showed the isolated location of each MPXV. The outer annotation ring represented the isolated date of each MPXV.

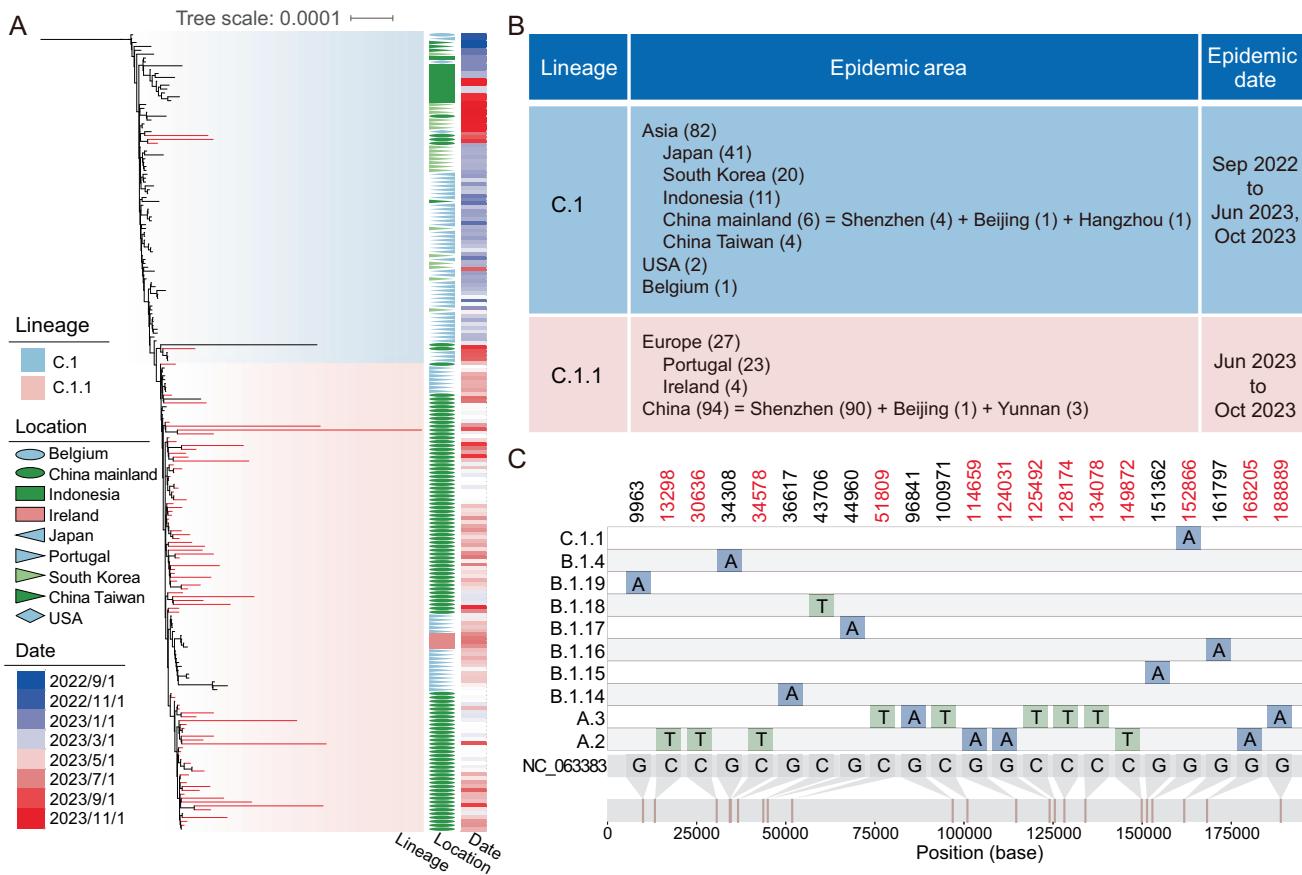


Fig. 2 | Emergence of a novel lineage of MPXV during 2023. A Evolutionary divergence of the C.1.1 Lineage supported by phylogenetic tree based on whole genome alignment. The background colors of the clades on the phylogenetic tree represent different lineages. The isolation location and date were annotated in the right of tree. Red branches indicate genomes from local outbreak of mpox in

Shenzhen. **B** Epidemiological evidence of the new C.1.1 lineage. The numbers in parentheses following the epidemic regions represent the quantity of genomes collected in the database. **C** Unique mutations found in each lineage compared with the reference genome (NC_063383.1) of MPXV. Mutation sites with red font indicate novel mutations compared to the lineage-defining SNPs in Nextclade database.

C.1 lineages derived from the A.1 lineage (Fig. 3A). Specially, the B.1 lineage diverging from the A.1 lineage has shown a new and distinct evolutionary direction, suggesting the occurrence of the first significant evolutionary event during human transmission of MPXV. Then we further conducted a detailed analysis for the MPXV lineages after B.1, and a total of 5115 sequences with 6966 mutations were included. In this scenario, the 21 sub-lineages that differentiated from the B.1 lineage demonstrated apparent clustering characteristics, and could not be distinguished through their mutation spectra (Fig. 3B). However, we observed a significant separation between the B.1 and C.1 lineages (PERMANOVA, $R^2 = 0.33$, $p = 0.001$). Notably, the C.1.1 lineage has diverged from the C.1 lineage and evolved toward another direction, suggesting that it has potentially undergone another significant evolutionary event (Fig. 3B).

To decipher the ongoing mutational trajectory of the MPXV, we explored the patterns and trends of nucleotide mutations (including substitutions, deletions, and insertions) across different lineages and time periods. Although total substitutions, deletions, and insertions of nucleotides showed significant differences across clade IIb MPXV ($p < 2.2e-16$), only nucleotide substitutions exhibited two evolutionary events that corresponded to the results in PCoA (Fig. 3C and Supplementary Fig. 5). In addition, we observed a significantly lower number of substitutions in all lineage A (including A, A.1, A.2, A.3) compared to the overall average number of substitutions in clade IIb MPXV. On the contrary, lineage B.1 exhibited a significant accumulation of nucleotide substitutions (Fig. 3C), which is consistent with the recent report of the significant evolutionary event occurred in lineage B.1¹⁷. Furthermore,

we observed a second significant increase ($p < 0.0001$) of nucleotide substitutions in lineage C.1.1 (Fig. 3C), supporting the emergence of another significant molecular evolutionary event for this lineage. On the other hand, the number of nucleotide substitutions in MPXV has been consistently increasing ($p < 2.2e-16$) over time. MPXV sequences from outbreak before 2021 showed significantly lower number of substitutions, while a significant increase occurred in 2023 ($p < 0.0001$) (Fig. 3D).

Accumulation of continuous mutations leads to molecular evolution of key proteins in mpox

To further investigate the potential mechanisms responsible for the two evolutionary events in the evolution of MPXV, we explored the frequency of all 12 types of nucleotide substitutions across clade IIb MPXV. The results showed that nucleotide mutations increased along with the evolution of MPXV (27 in A.1, 69 in B.1, 77 in C.1, up to 88 in C.1.1) (Supplementary Table 2), and C>T and G>A mutations occupied the highest proportion (reaching 89%) among all mutation types (Fig. 4A, B), which was consistent with the established APOBEC3 protein-mediated evolution mechanism in MPXV¹⁷. Additionally, the accumulation trend of the two mutations types, APOBEC3-like TC>TT and GA>AA, corresponded with the two evolutionary events (differentiation of B.1 from A.1, differentiation of C.1.1 from C.1) observed in clade IIb MPXV (Fig. 4E, F). Simultaneously, mutations of A>C and G>T also exhibit similar patterns of change, albeit with a much lower total number of nucleotide substitutions (Supplementary Fig. 6). At the level of amino acid, the average mutation number showed a similar

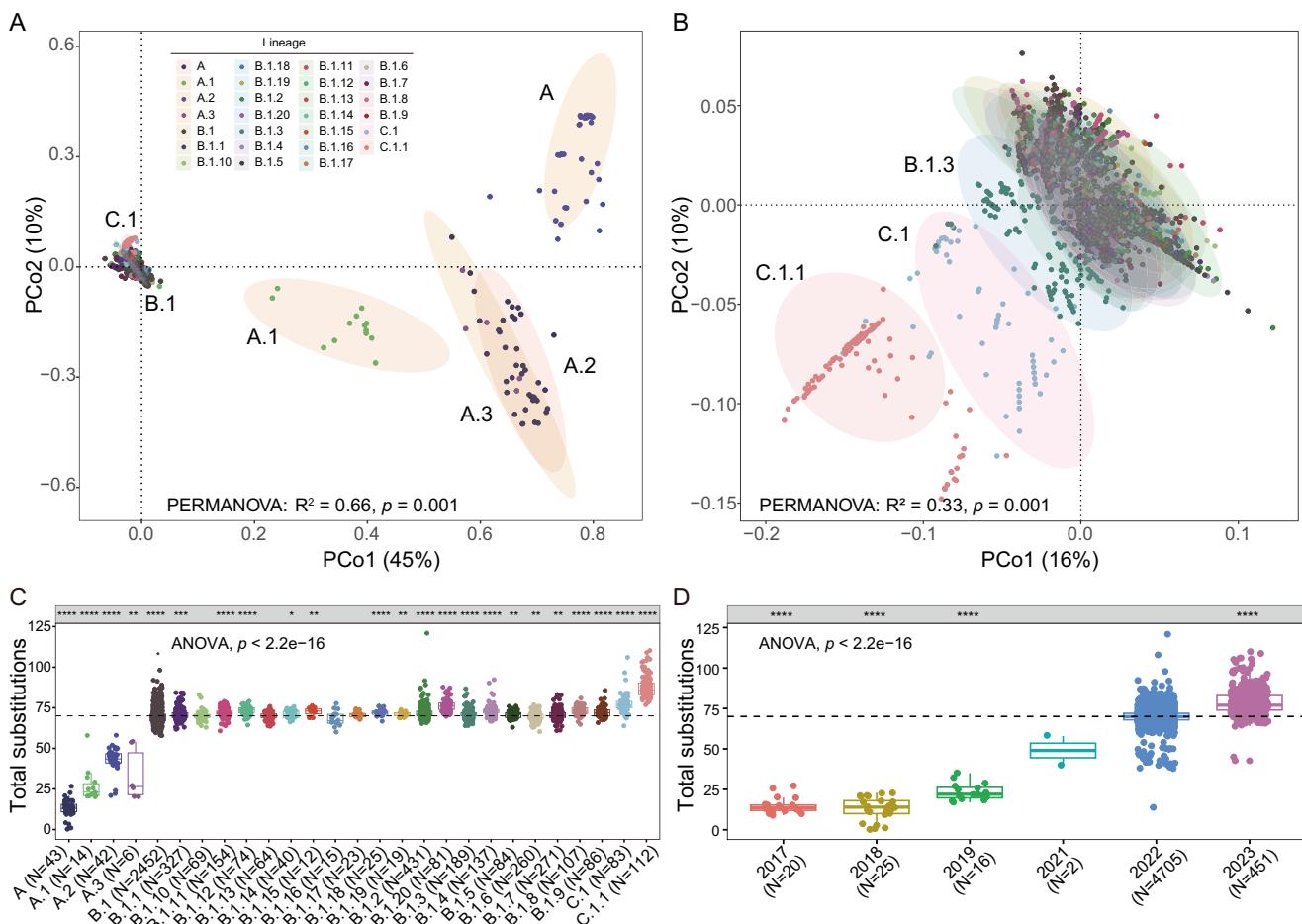


Fig. 3 | Significant evolutionary events among MPXV clade IIb. **A** PCoA showed significant evolutionary divergence of lineage B.1 from lineage A.1. 7432 mutations among 5220 MPXV genome sequences from clade IIb were used to conduct the dimensionality reduction analysis. **B** PCoA showed significant evolutionary divergence of lineage C.1.1 from lineage C.1. 6966 mutations among 5115 MPXV genome sequences from lineage B and lineage C were used to conduct the dimensionality reduction analysis. The dots in panels **A** and **B** were colored by lineage and grouped with 95% confidence intervals. The statistical correlation coefficient and significance were examined by permutational multivariate analyses of variance

(PERMANOVA) between different lineages. Comparison of total substitutions of MPXV genome sequences in clade IIb from the aspect of lineage (**C**) and outbreak (**D**). Boxplots showed the first quartile (minima), the third quartile (maxima) and the median (the solid line in the box plots) value of the substitution number in each lineage. The dashed line in (**C**) and (**D**) represented the mean value for all lineages. Significance was examined using one-way ANOVA among all groups and two-sided Student's *t*-test between two groups. Significance level were represented by black asterisks in the top of boxplot, with * indicating $p < 0.05$, ** indicating $p < 0.01$, *** indicating $p < 0.001$, and **** indicating $p < 0.0001$.

trend of increase among the clade IIb MPXV. Overall, missense mutations (48%), synonymous mutations (40%), and non-coding region mutations (12%) accounted for the majority of the mutations (Fig. 4C, D). Consistent with the patterns of nucleotide substitutions, the three type of amino acid mutations also showed a significant increase ($p < 2.2e-16$) during the two evolutionary events (Fig. 4G–I).

To investigate the key protein changes occurred during the evolutionary process of MPXV, we analyzed the protein mutation landscape among various lineages (Supplementary Data 2). The results showed that the clusters of 27 lineages on protein mutation profile were consistent with the phylogenomic relationships (Fig. 5A). All the five functional classes of the viral proteins have accumulated a significant number of missense mutations, suggesting that an increasing number of proteins have undergone large scale changes as evolution progresses (Fig. 5 and Supplementary Data 2). The A lineages exhibited a clear separation from subsequent evolutionary lineages in the heatmap, as lineages B and C demonstrated an increased accumulation of protein mutations (Fig. 5A). Overall, the MPXV proteins can be categorized into three major groups based on cluster results of euclidean distance of the mutation frequency matrix. The first cluster contains five proteins with the highest mutation number, OPG105, OPG003,

OPG210, OPG145, and OPG056. Among them, OPG105 (J6R, DNA-dependent RNA polymerase) and OPG145 (A18R, DNA helicase) are both involved in the viral replication process. Interestingly, OPG105 is more prone to synonymous mutations accompanied by a small number of missense mutations, while OPG145 predominantly undergoes missense mutations. OPG003 (C19L, Ankyrin repeat protein) and OPG210 (B22R family protein) are both associated with host regulation, but OPG003 tends to undergo synonymous mutations with a small number of missense mutations, whereas OPG210 mostly experiences missense mutations. OPG056 (F12L, EEV maturation protein), a protein associated with assembly, primarily undergoes synonymous mutations with a small fraction of missense mutations (Fig. 5A, B). The second cluster comprises 37 proteins, with the majority (18/37) being related to host modulation, followed by 6 surface proteins, 5 replication and 5 assembly-related proteins. Along with transmission of the lineage B, these proteins have accumulated a substantial number of mutations, mainly the missense and synonymous mutations (82%). Notably, a nonsynonymous mutation C21062T was specifically found in all the isolates of C.1.1 lineage and a few (6%) in the C.1 lineage (Supplementary Data 3), which introduces a c.3 G > A change in the virulence factor gene OPG036 (N2L, innate

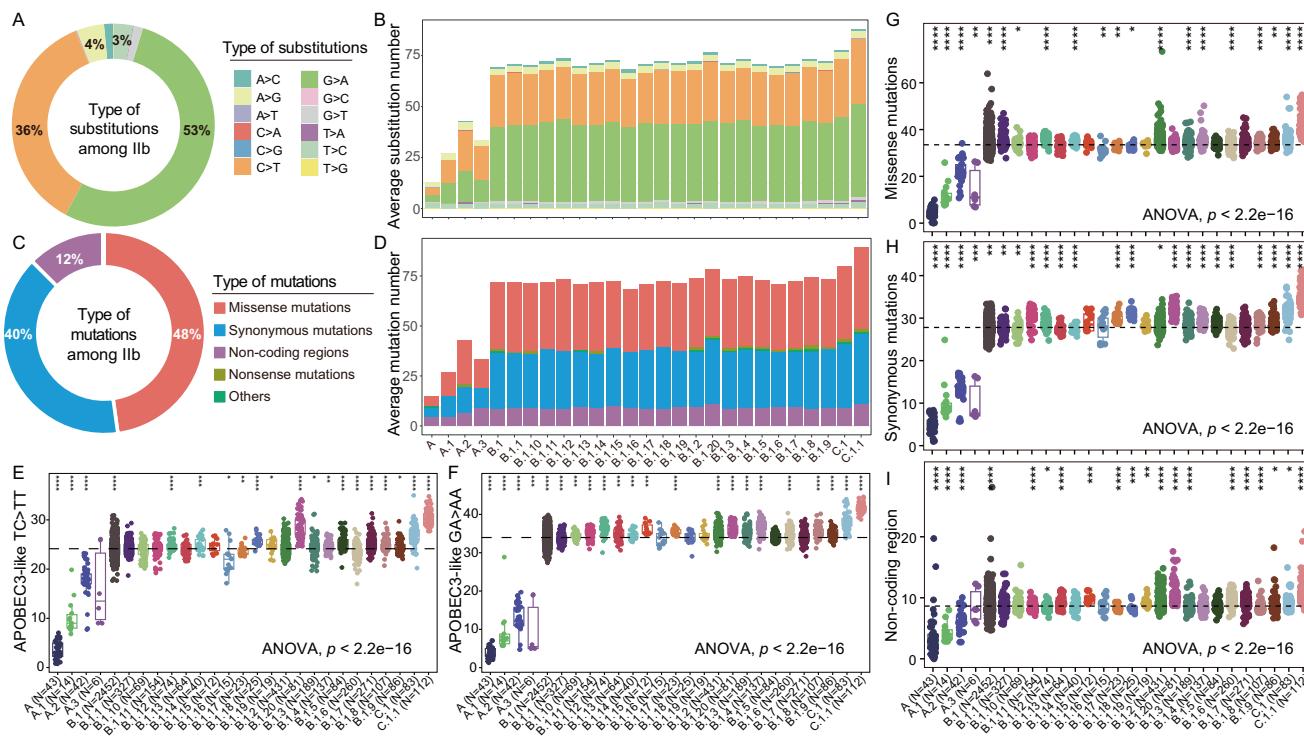


Fig. 4 | Persistent evolutionary characteristics of MPXV in the C.1.1 Lineage.

Total ratio (A) and average number (B) of 12 types of substitutions across all lineages among MPXV clade IIb. Total ratio (C) and average number (D) of 5 types of mutations across all lineages among MPXV clade IIb. Significant differences found in APOBEC3-like TC > TT mutation (E) and APOBEC3-like GA > AA mutation (F) across all clade IIb MPXV genomes. Significant differences found in missense mutations (G), synonymous mutations (H) and non-coding region mutations (I)

across all MPXV clade IIb genomes. Boxplots showed the first quartile (minima), the third quartile (maxima), and the median (the solid line in the box plots) value of mutation number in each lineage. Significance was examined using one-way ANOVA among all groups and two-sided Student's t-test between two groups. Significance level were represented by black asterisks, with * indicating $p < 0.05$, ** indicating $p < 0.01$, *** indicating $p < 0.001$ and **** indicating $p < 0.0001$.

immune modulator) and ultimately leads to the loss of transcription initiation²⁹. The third cluster encompasses the remaining 132 proteins with several lineage-specific protein mutations, which overall exhibit a lower mutation frequency. Among them, OPG074 (O1L, lev morphogenesis protein), OPG016 (Brix domain protein), OPG124 (D12L, mRNA capping enzyme), and OPG192 (B7R, virulence protein) in lineage C have accumulated a greater number of missense mutations (Supplementary Data 2). Moreover, OPG180 (A50R, DNA ligase, predominantly nonsynonymous mutations) and OPG120 (D8L, carbonic anhydrase, predominantly synonymous mutations) in lineage C.1.1 have undergone specific mutations (Fig. 5A, B). After normalization of total mutation number for each OPG by gene length, OPGs with higher mutation counts still exhibit relatively higher standardized mutation numbers (Fig. 5C). In addition, the clustering results of proteins and lineages after normalization also show many similarities, suggesting the consistency and significance of these mutations in the evolution of mpox (Supplementary Fig. 7). These findings suggest that during the evolutionary process of MPXV, different proteins may have been subjected to varying selective pressures.

Discussion

Between 2022 and 2023, the mpox has experienced a shift in the epicenter of its outbreak from Europe and the Americas to the Asia-Pacific region, although the total number of cases in the latter region is relatively low^{8,10,11}. There is still a great deal of uncertainty regarding the transmission trajectory of MPXV in the future, therefore it is crucial to understand the evolutionary laws and directions of MPXV. However, we have noticed only a limited number of comprehensive reports on sustained regional MPXV whole-genome surveillance since 2023^{22,23,30,31}. China, as one of the top ten countries most affected by

mpox¹⁰, reflects the ongoing characteristics and trends of MPXV evolution in 2023, especially in Shenzhen, an open port city located in Guangdong Province with the highest number of reported mpox cases in mainland China. Our large-scale genomic surveillance during the outbreak of mpox in Shenzhen revealed significant molecular evolution of MPXV since 2023.

The mpox outbreaks that occurred since 2022 have exhibited characteristics of accelerated mutations¹⁷ and changes in transmission patterns^{8,32}. Contact tracing in the sexual networks reported by cases poses challenges for tracing the origin of MPXV²⁷, and several studies have shown that large-scale genomic surveillance of MPXV is of great significance for epidemic monitoring, tracing, and the formulation of appropriate public health policies^{16,33-36}. The first case of mpox reported in mainland China occurred in September 2022 in Chongqing, with the virus belonging to the lineage B.1¹⁷. Subsequently, since June 2023, mpox patients have been reported in various parts of China with no genomic relation to that case^{22,23,30,31}, suggesting that mainland China may have experienced another multiple distinct importation event. Our large-scale genomic surveillance studies from June to October of 2023 indicate that the MPXV in Shenzhen region may have three distinct sources (Figs. 1, 2). The importation events within the C.1 lineage demonstrate a clear genomic relationship with the Asian regions of South Korea and Japan. Additionally, these two clusters of MPXV genomes show the highest similarity with sequences from Beijing and Hangzhou in China. This is consistent with the possibility of multiple importation sources from Asia regions that have been reported in Beijing and Guangdong^{22,23,30}. On the other hand, based on the genomic clustering results, MPXVs from Shenzhen within the C.1.1 sub-lineage, together with those from Beijing and Yunnan, are likely to be originated from Portugal. Portugal is one of the countries

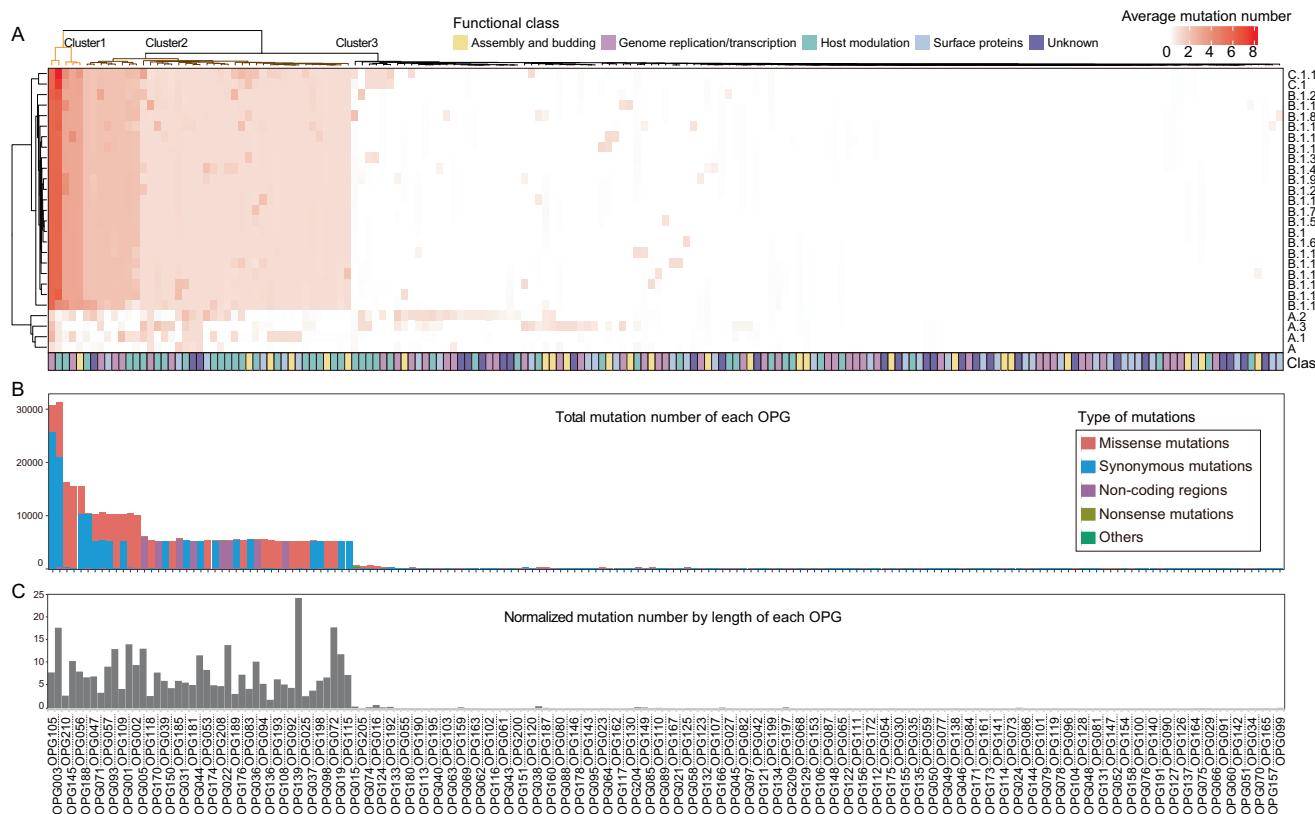


Fig. 5 | Protein mutation profiles across clade IIb MPXV. A Heatmap showed the average mutation numbers across each lineage, and clusters of 175 proteins in MPXV genome across 27 lineages. Proteins were annotated and colored by 5 functional classes. The proteins specifically contained in each cluster can be found

in Supplementary Data 2. **B** Stacked bar chart displayed the total mutation number of each OPG across 5220 MPXV genomes by 5 mutation types. **C** Bar chart presented normalized mutation number by gene length of each OPG.

that first reported mpox cases in 2022 outbreak, and MPXV has undergone multiple lineage evolution processes in Portugal³³. Further observation of the phylogenetic relationship within the C.1.1 sub-lineage reveals clear clustering patterns, indicating the complex transmission chains between regions and within local areas. Nevertheless, without further epidemiological evidence, it is challenging to determine whether these four MPXV sequences in Shenzhen originated from direct importation from overseas or from interprovincial transmission within China. Through travel history, we only found the epidemiological associations of confirmed mpox cases in Shenzhen with Thailand and Hong Kong, as well as several other cities in China (Fig. 1). However, due to the lack of other sequenced MPXV genomes from these regions or the potential hidden transmission^{22,23,31,37,38}, we did not observe clustering of MPXV genomes from these areas on the phylogenetic tree. Our observation highlights the need to strengthen global genomic surveillance of MPXV. This will help us make more accurate assessments of the virus's transmission relationships between different regions, as well as its origins and transmission pathways.

Multiple countries and regions have reported mpox cases with the C.1 lineage of MPXV^{22,23,30,35}, while its origin and evolution remained unclear. One recent study with genomic surveillance of MPXV indicates that the C.1 lineage has been detected in the Netherlands during May 2022 and August 2023 with a limited number of cases³⁵. According to our genomic investigations with global MPXV genomes, MPXVs within the C.1 lineage can be traced back to Japan in September 2022 and Belgium in October 2022, and then predominantly reported in 2023 (Fig. 2). These results suggest that the true global distribution and impact of the C.1 lineage may be more extensive with ongoing

surveillance. Moreover, by utilizing evidence from systematic evolutionary relationships, epidemiological trends, and specific genomic mutations (Figs. 2, 3), we have identified and characterized a potentially novel sub-lineage of MPXV C.1.1 derived from C.1, despite that no correlation between clinical features and the C.1.1 lineage can be established currently. Of particular note, a unique signature mutation site G152866A was found in the C.1.1 sub-lineage (Fig. 2C). Distinct from previously reported mpox mutation sites in different lineages^{14,16,20,21,39}, this signature mutation coupled with phylogenetic tree supports the notion that the MPXV evolved in a specific direction, giving rise to the new lineage C.1.1. Furthermore, the C.1.1 sub-lineage exhibits a unique evolutionary trajectory at the level of whole-genome mutation spectrum, representing another significant evolutionary event of the MPXV following the global outbreak that led to the B.1 lineage diverging from the A.1 lineage during the 2022 outbreak (Fig. 3). The adaptive evolution of the clade II has altered the transmissibility and pathogenicity compared to former clade I of MPXV^{18,24,40}, and the accumulated mutations in the C.1.1 sub-lineage may represent the new evolution direction of transmissibility and pathogenicity of the MPXVs after 2023, which merits further investigation based on clinical and experimental data.

Currently, the reported molecular evolutionary mechanisms for MPXV mainly encompass the accumulation of specific mutations¹⁸, viral recombination²², and the evolution of codon usage²⁴. The phenomenon of mutation accumulation in the MPXV genomes, mediated by the APOBEC3 family of antiviral proteins within human cells, has drawn extensive attention^{17,18,33,41}. Our research further highlighted the important role of APOBEC3 mediated mutations in the evolution of MPXV (Fig. 4), which contributed a lot to the emergence of the

potentially novel lineage C.1.1 in 2023 (Fig. 3). The compressed transmission chain of MPXV during the global outbreak may be a crucial driver of the high incidence of APOBEC3-related mutations observed³⁶, and these results further indicate that MPXV has selected APOBEC3-related mutations with the lowest adaptation cost under the pressure of natural selection¹⁸. Therefore, the impact of APOBEC3 on the MPXV evolution may be a long-term event, which could serve as the primary driving force for its evolution during the MPXV transmission. Notably, our results also showed that mutations are less likely in non-coding regions (Supplementary Fig. 8A, B), which would fit with non-coding regions having regulatory/control functions for gene expression and/or genome replication^{42,43}. This observation suggests that MPXV is able to suppress the generation of harmful mutations during evolution. Moreover, missense mutations are also found to be preferred over synonymous mutations during MPXV evolution (Supplementary Fig. 8C, D), indicating the advantageous nature of these missense mutations for the virus and the potential human adaptation. Our results provide important insights into the molecular evolution mechanisms of mpox during human-to-human transmission.

A profound understanding of the variations in protein sequence and structure resulting from SNPs is crucial for unraveling the changes in pathogenicity and transmissibility of MPXV. Mutations within the first two protein clusters are most prevalent within the lineage B, with most of these mutations occurring in the host regulatory proteins, suggesting the important roles of these proteins for the human transmission of MPXV. Meanwhile, mutations have also been previously confirmed in several replication-associated proteins, such as OPG105, OPG145, and OPG071^{15,34}, which may contribute to the replication and transmission of MPXV. Although mutations in the surface protein of MPXV are comparatively sparse, we have identified mutations in OPG120 (Carbonic anhydrase) within the newly emerged lineage C.1.1 (Fig. 5). Additionally, we have found that approximately 90% of mutations in the virulence protein (OPG192) occurred within the C.1 and C.1.1 lineages, with APOBEC3-related mutations accounting for approximately 83% of the mutations (Fig. 5 and Supplementary Data 2). Moreover, we have found two lineage-specific mutations, C149872T and C21062T, which are potentially of significant impact on the transmission and virulence of MPXV (Supplementary Data 3). The C149872T mutation, which disrupts the *A46R* gene of the MPXV, has been identified as a hallmark adaptation event for the A.2 lineage¹⁵. Notably, our study reveals for the first time that the newly emerged C21062T mutation in the C.1.1 lineage results in the disruption of the transcriptional initiation of the *N2L* gene, a reported innate immune modulator of vaccinia virus²⁹. Meanwhile, mutations in the *N2L* gene have been shown to reduce viral virulence in vaccinia virus²⁹, leading to the speculation that the adaptability and virulence of the C.1.1 lineage may be affected by this mutation, which further supports the emergence of the potentially novel lineage C.1.1. These results indicate that the MPXV has accumulated mutations related to virulence^{44–46} and surface antigenic proteins throughout its ongoing evolution, warranting vigilance for potential changes in its pathogenicity and transmissibility. Further investigation into the biological significance and potential impacts of these mutations will contribute to a better understanding of the evolutionary mechanisms of MPXV and provide scientific evidence for epidemic monitoring and intervention strategies.

In conclusion, we have elucidated the evolutionary trajectory and characteristics of MPXV in 2023 based a large-scale genomic surveillance in Shenzhen, China. Our study confirms that the accumulation of APOBEC3-driven mutations in clade IIb MPXV has not plateaued. This ongoing mutation process continues to provide raw material for mpox evolution, highlighting the significant potential for adaptation. Our findings indicate that the global outbreak strain of mpox presents a continuing international public health challenge. The virus has significant mutational potential that could, in the

future, allow it to spread more widely and pose a threat to the broader population, beyond the currently affected demographic. This finding underscores the importance of ongoing surveillance and research to monitor these evolutionary dynamics and their implications for public health.

Methods

Participants and sample collection

Participants presented in this study were laboratory confirmed mpox patients using quantitative real-time PCR (Macro & Micro Test Co., Ltd.) and were admitted to our hospital during 9 June to 12 October 2023 ($N=92$). Clinical information and laboratory results were collected at the earliest timepoint after hospital admission. Skin lesions samples were collected from the enrolled patients during hospitalization and follow-up, and the samples with highest viral load (indicated as lowest Ct values) were subjected for sequencing. The study protocol was approved by the Ethics Committees of Shenzhen Third People's Hospital. Written informed consent was obtained from all patients.

DNA quality control and library construction

The MPXV DNA was extracted using commercial kit (Qiagen, Germany) and subsequently submitted to BGI China for whole-genome sequencing. The Mpoxy Virus Nucleic Acid Detection Kit (BGI, China) was used to perform multiplexed probe fluorescence quantitative detection on the *F3L* and *B7R* genes, the results of which indicated that the *F3L* and *B7R* genes passed the QC if the Ct values were ≤ 32 , and the Ct value of the internal reference genes was ≤ 35 . Library construction was performed using the Mpoxy Whole Genome Assay Universal Kit (Multiplex Amplification Method) (BGI, China). The kit contains 342 pairs of primers designed according to the whole genome of the MPXV, and the amplicon products with the target sequences of the whole genome of the mpox virus were amplified by multiplex PCR. The amplification products were purified by using magnetic beads, and the purified products were subjected to concentration quantification and fragment quality control. The concentration of the qualified products should be greater than 10 ng/ μ L, and the distribution of target fragments should be in the range of 650 bp–750 bp, with no obvious dimerization and non-specific fragments.

Whole genome sequencing

The amplification product was end-repaired and connected to the Native Adapter kit (NBD-104/114) with the ligation sequencing kit (LSK110). Sequencing was conducted using the MinION platform from Oxford Nanopore Technologies (ONT). The sequencing process was controlled using MinKNOW (v23.04.6, ONT), and basecalling and demultiplexing were performed using guppy (v6.5.7, ONT) in Fast mode. Ultimately, the average sequencing speed for this batch of data was 292 Bases/sec, with a minimum output of 400 K reads or 200 M bases per sample. After acquiring the sequencing data, artic guppyplex v1.2.1 (<https://github.com/artic-network/fieldbioinformatics>) was utilized to filter out reads with sequence lengths below 300 bp or above 1000 bp, as well as sequence average quality lower than 8 in the sequencing data. The filtered clean reads of each samples were mapped to the reference genome of MPXV (NC_063383.1) using minimap2⁴⁷ with default parameters. The base distribution information was extracted using bam-readcount v1.0.1⁴⁸ with default parameters. Subsequently, for each position, we analyzed the base depth and composition. The nucleotide with a sequencing depth greater than 30 \times and a depth proportion greater than 50% was selected as the output for that position. Following these criteria, the consensus genome sequence was generated for each sample. The detailed sequencing information of 92 MPXV genomes from Shenzhen, including reads number, total bases, mapped bases, coverage, and average depth was summarized in Supplementary Table 3.

Phylogenetic and molecular evolution analysis

In order to conduct a systematic molecular evolution study of MPXV, we retrieved 3734 MPXV genomes longer than 196 Kb from the NCBI GenBank database (<https://www.ncbi.nlm.nih.gov/genome>) and extracted 5077 MPXV genomes labeled with “Complete” from the GISAID database (<https://www.epicov.org/epi3/frontend#3abd83>) (as of November 20, 2023). Subsequently, we merged these two databases and removed redundant sequences. The sequences were subjected to quality control using NextClade⁴⁹, resulting in a reference dataset of 5128 high-quality MPXV genomes. Next, we used the NextClade online service platform to preliminarily confirm that all indigenous MPXV genomes in Shenzhen belong to the C.1 lineage. We extracted all genomes belonging to the C.1 lineage from the reference dataset and randomly selected up to three MPXV from each lineage outside of C.1, and conducted the phylogenetic analyses along with the 92 local MPXV genomes from Shenzhen. These MPXV genomes were aligned using MAFFT v7.520 software⁵⁰, and the alignment was trimmed using trimAl v1.4.rev15 software⁵¹. Subsequently, a maximum likelihood phylogenomic tree was constructed using IQ-TREE v2.2.5 software⁵², and the optimal nucleotide substitution model was automatically selected using ModelFinder⁵³. Additionally, 1000 ultrafast bootstrap approximations were performed to calculate the support values for the tree nodes. To further support the results of the phylogenetic analysis, a SNP-based systematic evolution analysis was conducted using Parsnp v1.7.4 software⁵⁴. Finally, the resulting phylogenetic tree and SNP-based evolutionary tree were visualized and edited using the iTOL v5 web tool⁵⁵. Nucleotide and protein mutation profiles were identified using NextClade compared to the reference genome of clade IIb MPXV (accession: NC_063383.1), and the mutations were further annotated by SnpEff v5.2a software⁵⁶. Visualization of specific mutation sites for each lineage was performed using Snipit (<https://github.com/aineniamh/snippet>).

Statistical analysis

All datasets were analyzed using R v4.0.2 software (<https://www.r-project.org>) and visualized using the ggplot2 package (<https://cran.r-project.org/package=ggplot2>). The median and interquartile range (IQR) were reported for continuous variables, while frequencies and percentages were reported for categorical variables in the statistical analysis of epidemiological and clinical characteristics. To assess the distribution of all mutations across the mpox genomes, a principal coordinate analysis (PCoA) was performed using the Bray-Curtis distance matrix. The significance of differences in viral composition among different lineages were assessed using the adonis function from the vegan package (<https://cran.r-project.org/package=vegan>), which performs a non-parametric multivariate analysis of variance (PERMANOVA) based on the Bray-Curtis distance with 999 permutations. Significance was examined using one-way ANOVA among all groups and two-sided Student's *t*-test between two groups, and visualized using the ggpublish package (<https://rpkgs.datanovia.com/ggpublish/>). During ANOVA, when the *p*-value is extremely close to zero, it will be expressed as *p*<2.2e-16 due to the limitations in computer floating-point representation.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The whole genome sequences of 92 Shenzhen Mpox virus have been deposited in the GenBank in National Genomics Data Center under accession number from C_AA061241.1 to C_AA061332.1 that are accessible from <https://ngdc.cncb.ac.cn/genbase>. The 92 raw

sequencing data have been deposited in the Genome Sequence Archive in National Genomics Data Center with accession: CRA015345 under Bioproject: PRJCA024271. The detailed information of MPXV sequences used in this study has been summarized and show in Supplementary Table 3.

References

1. Ulaeto, D. et al. New nomenclature for mpox (monkeypox) and monkeypox virus clades. *Lancet Infect. Dis.* **23**, 273–275 (2023).
2. Harapan, H. et al. Monkeypox: A comprehensive review. *Viruses* **14**, 2155 (2022).
3. Thakur, M., Das, P., Sobe, R. C. & Kaur, T. Human monkeypox: epidemiology, transmission, pathogenesis, immunology, diagnosis and therapeutics. *Mol. Cell. Biochem.* **478**, 2097–2110 (2023).
4. Magnus, P. V., Andersen, E. K., Petersen, K. B. & Birch-Andersen, A. A Pox-like disease in Cynomolgus monkeys. *Acta Pathol. Microbiol. Scand.* **46**, 156–176 (2009).
5. Bunge, E. M. et al. The changing epidemiology of human monkeypox—A potential threat? A systematic review. *PLOS Negl. Trop. Dis.* **16**, e0010141 (2022).
6. Hutin, Y. J. F. et al. Outbreak of Human Monkeypox, Democratic Republic of Congo, 1996 to 1997. *Emerg. Infect. Dis.* **7**, 434–438 (2001).
7. Durski, K. N. et al. Emergence of Monkeypox — West and Central Africa, 1970–2017. *Mmwr. Morbid. Mortal. Wkly. Rep.* **67**, 306–310 (2018).
8. Kraemer, M. U. G. et al. Tracking the 2022 monkeypox outbreak with epidemiological data in real-time. *Lancet Infect. Dis.* **22**, 941–942 (2022).
9. Wenham, C. & Eccleston-Turner, M. Monkeypox as a PHEIC: implications for global health governance. *Lancet* **400**, 2169–2171 (2022).
10. World Health Organization. 2022-24 Mpx (Monkeypox) Outbreak: Global Trends. accessed 02 August 2024. https://worldhealth.org/shinyapps.io/mpx_global/.
11. Hatmal, M. M. M. et al. Comprehensive literature review of monkeypox. *Emerg. Microbes Infect.* **11**, 2600–2631 (2022).
12. Shchelkunov, S. N. et al. Analysis of the Monkeypox virus genome. *Virology* **297**, 172–194 (2002).
13. Lu, J. et al. Mpox (formerly monkeypox): pathogenesis, prevention, and treatment. *Signal Transduct. Target. Ther.* **8**, 458 (2023).
14. Happi, C. et al. Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. *PLOS Biol.* **20**, e3001769 (2022).
15. Ndodo, N. et al. Distinct monkeypox virus lineages co-circulating in humans before 2022. *Nat. Med.* **29**, 2317–2324 (2023).
16. Molina, I. S. et al. Sub-lineage B.1.6 of hMPXV in a global context: Phylogeny and epidemiology. *J. Med. Virol.* **95**, e29056 (2023).
17. Isidro, J. et al. Phylogenomic characterization and signs of micro-evolution in the 2022 multi-country outbreak of monkeypox virus. *Nat. Med.* **28**, 1569–1572 (2022).
18. O'Toole, Á. et al. APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016. *Science* **382**, 595–600 (2023).
19. Alakunle, E., Moens, U., Nchinda, G. & Okeke, M. I. Monkeypox Virus in Nigeria: Infection biology, epidemiology, and evolution. *Viruses* **12**, 1257 (2020).
20. Luna, N. et al. Phylogenomic analysis of the monkeypox virus (MPXV) 2022 outbreak: Emergence of a novel viral lineage? *Travel Med. Infect. Dis.* **49**, 102402 (2022).
21. Luna, N. et al. Monkeypox virus (MPXV) genomics: A mutational and phylogenomic analyses of B.1 lineages. *Travel Med. Infect. Dis.* **52**, 102551 (2023).

22. Yu, J. et al. Phylogeny and molecular evolution of the first local monkeypox virus cluster in Guangdong Province, China. *Nat. Commun.* **14**, 8241 (2023).

23. Zhang, D. et al. Genetic Features of 84 Genomes of Monkeypox Virus in Recent Circulation — Beijing Municipality, China, 2023. *China CDC Wkly.* **5**, 918–921 (2023).

24. Shan, K.-J. et al. Molecular evolution of protein sequences and Codon usage in monkeypox viruses. *Genom. Proteom. Bioinform.* **22**, qzad003 (2024).

25. Yeh, T.-Y. et al. Recombination shapes the 2022 monkeypox (mpox) outbreak. *Medicine* **3**, 824–826 (2022).

26. Yang, Y. et al. Longitudinal viral shedding and antibody response characteristics of men with acute infection of monkeypox virus: a prospective cohort study. *Nat. Commun.* **15**, 4488 (2024).

27. Vivancos, R. et al. Community transmission of monkeypox in the United Kingdom, April to May 2022. *Eurosurveillance* **27**, 2200422 (2022).

28. Du, M. et al. Global epidemiological features of human monkeypox cases and their associations with social-economic level and international travel arrivals: a systematic review and ecological study. *Int. J. Public Health* **68**, 1605426 (2023).

29. Ferguson, B. J. et al. Vaccinia virus protein N2 is a nuclear IRF3 inhibitor that promotes virulence. *J. Gen. Virol.* **94**, 2070–2081 (2013).

30. Dou, X. et al. Clinical, epidemiological, and virological features of Mpox in Beijing, China – May 31–June 21, 2023. *Emerg. Microbes Infect.* **12**, 2254407 (2023).

31. Jia, H. et al. Genomic and epidemiological perspectives on the first local sporadic cases of Mpox in China. *Emerg. Microbes Infect.* **12**, 2245932 (2023).

32. Thornhill, J. P. et al. Monkeypox virus infection in humans across 16 countries — April–June 2022. *N. Engl. J. Med.* **387**, 679–691 (2022).

33. Borges, V. et al. Viral genetic clustering and transmission dynamics of the 2022 mpox outbreak in Portugal. *Nat. Med.* **29**, 2509–2517 (2023).

34. Patiño, L. H. et al. Phylogenetic landscape of Monkeypox Virus (MPV) during the early outbreak in New York City, 2022. *Emerg. Microbes Infect.* **12**, e2192830 (2023).

35. Schuele, L. et al. Circulation, viral diversity and genomic rearrangement in mpox virus in the Netherlands during the 2022 outbreak and beyond. *J. Med. Virol.* **96**, e29397 (2024).

36. Ulaeto, D. O., Dunning, J. & Carroll, M. W. Evolutionary implications of human transmission of monkeypox: the importance of sequencing multiple lesions. *Lancet Microbe* **3**, e639–e640 (2022).

37. Zhao, H. et al. The first imported case of monkeypox in the mainland of China - Chongqing Municipality, China, September 16, 2022. *China CDC Wkly.* **4**, 853–854 (2022).

38. Chiu, K. H.-Y. et al. The first case of monkeypox in Hong Kong presenting as infectious mononucleosis-like syndrome. *Emerg. Microbes Infect.* **12**, 2146910 (2022).

39. Jia, X., Jia, P., Zhu, Y., Yu, W. & Yang, Q. Evolution and marker mutations in different clades of monkeypox virus. *J. Med. Virol.* **95**, e28596 (2023).

40. Huang, Y., Mu, L. & Wang, W. Monkeypox: Epidemiology, pathogenesis, treatment and prevention. *Signal Transduct. Target. Ther.* **7**, 373 (2022).

41. Suspène, R. et al. APOBEC3F is a mutational driver of the human monkeypox virus identified in the 2022 outbreak. *J. Infect. Dis.* **228**, 1421–1429 (2023).

42. Wang, L. & Lee, C.-W. Sequencing and mutational analysis of the non-coding regions of influenza A virus. *Vet. Microbiol.* **135**, 239–247 (2009).

43. Fortes, P. & Morris, K. V. Long noncoding RNAs in viral infections. *Virus Res.* **212**, 1–11 (2016).

44. Price, N., Tscharke, D. C., Hollinshead, M. & Smith, G. L. Vaccinia Virus Gene B7R encodes an 18-kDa protein that is resident in the endoplasmic reticulum and affects virus virulence. *Virology* **267**, 65–79 (2000).

45. Senkevich, T. G. et al. Ancient gene capture and recent gene loss shape the evolution of orthopoxvirus-host interaction genes. *mBio* **12**, e0149521 (2021).

46. Smith, G. L. et al. Vaccinia virus immune evasion: mechanisms, virulence and immunogenicity. *J. Gen. Virol.* **94**, 2367–2392 (2013).

47. Li, H. & Birn, I. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

48. Khanna, A. et al. Bam-readcount - rapid generation of basepair-resolution sequence metrics. *J. Open Source Softw.* **152**, 1498–1498 (2022).

49. Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).

50. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

51. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

52. Lanfear, R. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

53. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

54. Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **15**, 524 (2014).

55. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

56. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2014).

Acknowledgements

This work was supported by grants from the Shenzhen Clinical Research Center for Emerging Infectious Diseases (No. LCYSSQ20220823091203007), Sanming Project of Medicine in Shenzhen (SZSM202311033), the Shenzhen High-level Hospital Construction Fund (23250G1001 and XKJS-CRGRK-005 and XKJS-CRGRK-004), the Guangdong Science and Technology Plan Project for Construction of High-level Biosafety Laboratories (2021B1212030010) and the Special Funds for Strategic Emerging Industry of Shenzhen (No. F-2022-Z99-502266). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Y. Yang, Y.L. and H.L. conceived and designed the study. S.Z., Y. Yang, X.G., G.F. and L.S. contributed to the analysis and interpretation of data. F.W., Y.L., L.Y., Y. Yin, and J.Y. enrolled patients and collection of samples. Y.P., X.G., S.N., J.L., L.Z. and Y.C. carried out the experiment. S.Z. and Y. Yang drafted the article. All authors reviewed and revised the manuscript and approved the final version. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-51737-4>.

Correspondence and requests for materials should be addressed to Hongzhou Lu, Yingxia Liu or Yang Yang.

Peer review information *Nature Communications* thanks David O. Ulaeto and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024