Research article

Genome-wide mutational analysis of Chikungunya strains from 2016 to 2017 outbreak of central India: An attempt to elucidate the immunological basis for outbreak

Sudheer Gupta a,1, Ashvini Yadav a,1, Sam Stubbs b, Simon Frost b, Kudsia Ansari a, Ram Kumar Nema a, Shashwati Nema a,c, Debasis Biswas a,c,*

a Regional Virology Laboratory, All India Institute of Medical Sciences, Bhopal, India
b Department of Veterinary Medicine, University of Cambridge, UK
c Department of Microbiology, All India Institute of Medical Sciences, Bhopal, India

ARTICLE INFO

Keywords:
Chikungunya
Phylogenetics
Outbreak
Genome sequencing
Nanopore

ABSTRACT

Chikungunya re-emerged in India in 2016–2017, as the first major outbreak since 2006. In our previous study, we undertook partial E1 gene sequencing and phylogenetic/mutational analysis of strains from the 2016–2017 outbreak of Chikungunya in central India and reported important mutations associated with the outbreak. This study was performed to validate the previous findings and to identify key mutations that had emerged throughout the entire genome of Chikungunya virus that could be driving the enormity of this outbreak. The phylogenetic analysis revealed the closeness of our isolates with ECSA genotype, specifically with the Singapore 2015 strain. We found 2 mutations in C and E2 genes, which were present in our isolates but were non-existent during the period of 2010–2016. Furthermore, re-emergence of Arg amino acid in place of stop codon in nsP3 gene and Thr at E2:312 positions was observed after 2011. We also used computational tools to assess the effect of the identified mutations on the T cell and B cell epitopes that could influence the protective immune response against this infection.

1. Introduction

Chikungunya virus (CHIKV) is a vector-borne alphavirus belonging to the Togaviridae family. It is spread by Aedes mosquitoes and is responsible for acute-onset febrile episodes accompanied by significant musculoskeletal manifestations. Since the first report of CHIKV from Tanzania in 1952 [1], there has been a series of outbreaks throughout the globe including East African, West African and Asian countries [2, 3]. In India, CHIKV was first reported in Kolkata in 1963, followed by outbreaks in the 1960s, 1970s and 1980s [4]. After a gap of 32 years, CHIKV re-emerged as a major outbreak in 2006 [5]. Notably, the genotype responsible for this outbreak was of East/Central/South African lineage (ECSA) in contrast to the Asian genotype which had caused the previous outbreaks between 1963 and 1973 [4]. Though the disease remained endemic in India after the outbreak of 2006–2007, there has been a massive upsurge with more than 3-fold rise in confirmed CHIKV cases from 2015 (12% of suspects) to 2016 (41% of suspects) (https://nvbdcp.gov.in). A similar trend was seen in central India where cases rose sharply from 2015 (16% of suspects) to 2016 (37% of suspects) (https://nvbdcp.gov.in).

Previous outbreaks of CHIKV have been associated with the emergence of specific mutations in the CHIKV genome. For example, A226V mutation in the E1 protein was found to be responsible for viral adaptation to Aedes albopictus mosquitoes during the 2006–2007 outbreak [6]. Similarly, K211E in the E1 protein and V264A in the E2 protein have been associated with increased adaptation of the virus in the Aedes aegypti vector during outbreaks in 2009–2010 [7]. In order to explore whether the emergence of any such mutation may have contributed to the 2016–2017 outbreak, we have previously performed a preliminary analysis of partial E1 gene sequences and reported various novel and reemerging mutations [8].

This study expands upon the aforementioned analysis by investigating the entire genome of the 2016–17 outbreak strains, in order to...
identify novel and re-emerging mutations associated with this outbreak. Furthermore, with the availability of whole genome sequences we explored the impact of these mutations on the antigenic repertoire of the virus.

2. Materials and methods

This study has been carried out in Regional Virology laboratory of All India Institute of Medical Sciences, Bhopal, India, which is a tertiary care teaching hospital. The CHIKV suspected cases attending the outpatient department were tested using Chikungunya IgM capture ELISA kit (supplied by National Institute of Virology, Pune, India) as per the protocol provided by the kit manufacturers. IgM ELISA positive samples were delinked from personal identifiers and further confirmed using RT-PCR based viral detection. The RNA extraction was performed using QIAamp viral RNA mini kit (Qiagen, Germany). The PCR was performed using the SS III platinum one step RT-PCR kit (Invitrogen, US) with CHIKV specific primers (CHIK15F and CHIK16R) as previously described [3]. RT-PCR confirmed cases of CHIKV were further utilized for sequencing. A total of 21 samples were randomly selected for whole genome sequencing of CHIKV, among which 4 isolates (RVL-AIIMS-CH01 to RVL-AIIMS-CH04) were from 2016 and 17 samples were from 2017 (RVL-AIMSBPL-CH05 to RVL-AIMSBPL-CH21). The samples were collected from the months of August to October of the respective year, coinciding with the seasonality of Chikungunya cases in our region.

2.1. Whole genome sequencing

Viral RNA was extracted from 200 μL of serum using the QIAamp viral RNA mini kit (Qiagen, Germany). Three μL of RNA was used as input for cDNA synthesis, performed using Superscript III reverse transcriptase (Thermo Fisher Scientific, USA) and 50 ng of random hexamers (Thermo Fisher Scientific, USA) in a 20 μL reaction, as per the enzyme manufacturer's instructions. Two μL of cDNA was used as input for subsequent multiplex-PCR amplification of the CHIKV genome performed using a tiled-amplicon approach [10]. Barcoded, Nanopore sequencing libraries were prepared according to Oxford Nanopore Technologies’. (ONT) 1D Native Barcoding Protocol (available from https://community.nanoporetech.com/protocols), using ONT Native Barcoding kits (EXP-NBDI04 and EXP-NBDI14) and the ONT Ligation Sequencing kit (SQK-LSK109). The resulting barcoded libraries were loaded onto r9.4.1 flow-cells (FLO-MIN106) and sequenced using MinKNOW software v1.13.1. Base-calling of raw FAST5 files was performed using Guppy v3.1.5 with default settings, discarding reads with a q-score below 7. Base-called reads were de-multiplexed using Qcat v1.0.7 with default settings, discarding reads with a q-score below 7. Base-called and individual FASTQ files were aligned to the CHIKV RefSeq sequence (NC_004162.2) using BWA mem v0.7.17 (option-x ont2d) [11]. Primer sequences were trimmed from the aligned reads and draft consensus genome sequences were called using a simple pipeline method as previously described [12], masking regions with a read depth <20X. Draft consensus genomes were aligned using MAFFT v7.427 [13] and visualised using AliView v1.2.6 [14], revealing eight short (200–400 bp) regions with insufficient coverage across multiple samples. Additional PCR primer sets were therefore designed to amplify these regions. The resulting amplicons were again barcoded, sequenced, base-called and de-multiplexed in the same fashion described above.

The two FASTQ datasets were combined and draft consensus sequences were reconstructed using the same method described above. Base-called FASTQ reads were then realigned to the draft consensus sequences, and primer sequences were again clipped prior to consensus correction, which was performed using Nanoplotch variants v 0.1.1 [15] (ploidy 1–min-flanking-sequence 10). Variant calls were filtered to include only those with a quality score of > 200 and a support fraction > 70% before being applied to generate the final consensus genomes. Lastly, polished genome sequences were re-aligned using MAFFT v7.427 and visualised using AliView v1.2.6 [14]. Major discrepancies such as indels or multiple consecutive SNPs were manually verified or corrected by referring back to the BAM read alignment file viewed in Tablet v1.19.09.03 [16].

2.1.1. Phylogenetic and mutational analysis

The full length genomes were annotated using VIGOR software available at VIPR server (https://www.viprbrc.org/brc/vigorAnnotation.jsp?method=ShowCleanInputPage&amp;decorator=toga) [17]. For phylogenetic analysis, one representative genome from each of 7 lineages [18] (KY703988 (American Lineage), MF773562 (Asian Urban Lineage), GU301781 (Indian Ocean Lineage), EU564334 (Eastern African Lineage), MG649978 (South American Lineage), KP003813 (Middle African Lineage), HM045792 (African/Asian Lineage)) and one representative genome each from major regions with reports of CHIKV outbreaks between 2010 and 2015 were downloaded from VIPR database. Similarly, for mutational analysis, all complete genome sequences of Indian origin (n = 30), were retrieved from VIPR database.

Phylogenetic analysis was performed with MEGA-X software by using the Maximum Likelihood method with bootstrap support (1000 replicates) and Tamura-Nei model [19]. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site.

2.2. Immune epitope analysis

In order to assess the modulation in immunological behavior due to sequence variation, the protein sequences from complete Indian CHIKV genomes (from the year 2010–2016) were predicted using in silico prediction algorithms. The immune epitope prediction was performed for B cells, cytotoxic T cells and T-helper cells.

2.2.1. B cell epitope prediction

The prediction of B cell epitopes was performed using LBtope server [20]. LBtope is a Support Vector Machine-based prediction server for B cell epitopes. The window length was set at default value of 15. Residues with probability of correct prediction >60% (default value) were considered as B cell epitopes.

2.2.2. HLA-I binding prediction

HLA-I binding was predicted using IEDB server [21]. The prediction was performed for all 27 reference HLA-I alleles available at the server. The IEDB HLA-I prediction is based on Artificial Neural Network. The window length was set to 9 amino acids. The epitopes having IC50 < 50nM, were considered as strong binders and those having IC50 between 50 and 500 nM, were considered as intermediate binders. Patterns where IC50 was greater than 500 nM were assigned as non-binders as per the instruction given on the server.

2.2.3. HLA-II binding prediction

Similar to the HLA-I binding prediction, HLA-II binders were predicted using IEDB server [21] which is based on a position specific weight matrix. The analysis was performed on default window length of 15 amino acids. The reference set of HLA-II alleles (27 alleles) were selected from the server for analysis. Similar to the HLA-I binding prediction, HLA-II binders with IC50 < 50nM, were considered as strong binders and those having IC50 between 50 and 500 nM, were considered as intermediate binders. Non-binders were those which had IC50 > 500 nM.

The selection criteria for the reference HLA alleles by the authors in IEDB prediction server includes:

1) the most common specificities in the general population, based on data available from DbMHC and allelefrequencies.net
2) representative of commonly shared binding specificities (i.e., supertypes).

In terms of population coverage, the reference sets for class I and II should provide >97% and >99%, respectively.
3. Results

3.1. Phylogenetic and mutational analysis

The phylogenetic analysis performed using representative genome sequences of known lineages and from countries with recent CHIKV outbreaks (since 2010) revealed that the CHIKV isolates from central India during the outbreak of 2016–2017 belonged to ECSA genotype (Figure 1). We had earlier reported on the phylogenetic relationship of CHIKV strains isolated from our region based on partial sequence of the E1 gene [8]. While we observed the phylogenetic proximity of these isolates to recent New Delhi strains of 2015, we wanted to validate these findings based on the whole genome sequences of the CHIKV isolates. The phylogenetic analysis with whole genome sequences revealed that our isolates had closest proximity with strain from Singapore circulating during 2015 and Indian strains of 2014 and 2015 (Figure 1). Furthermore, our isolates also displayed close clustering with strains from Singapore isolated during 2012 and 2013 and from India isolated during 2010, 2012, 2014 and 2015. However, some Indian CHIKV strains, which were also collected in 2016, were found in a separate cluster.

In view of the separate phylogenetic clustering of our isolates from majority of Indian isolates sequenced post-2010, we were interested in identifying the specific mutations that had emerged in the viral genome in this geographical region since 2010.

Multiple sequence alignment analysis of polyprotein sequences reported from 2010 to 2016 was performed for each position in the protein sequences and four kinds of variations were observed (Table 1). Interestingly, 2 mutations were “novel” as they were non-existent in the intervening period between 2010 and 2016 and were identified in most of our isolates from 2016 to 2017 outbreak. One of these 2 mutations was located in C protein (N79S) and the other was present on E2 protein (A76T). E2:A76T was present in 16 (all from 2017) out of 21 and C79N was present in 14 (all from 2017) out of 21 isolates. Interestingly, both of the novel mutations (C:N79S & E2:A76T) were already present in the Singapore strain (2015) but absent in New Delhi strain (2015) (data not shown) (Figure 1). In addition to these 2 novel mutations, E2:M312Tand nsP3:*524R “re-emerged” in our isolates while they were absent in strains reported between 2012 and 2015. These two mutations had previously been observed in 2011 from Kolkata in the eastern part of India. The E2:M312T mutation was present in 7 out of 21 isolates among which 1 isolate was from 2016 and 6 were from 2017. However, nsP3:*524R was present in all of our isolates. Thirdly 5 additional mutations (E1:I317V, nsP2:H130Y, nsP2:E145D, nsP4:555N and nsP4:R85G) were found to have emerged in 2015 and continue through 2016–2017 outbreak as “stable” mutations. Apart from these novel, re-emerging (n = 2 each) and stable (n = 5) mutations, we observed 31 locations where the mutations appeared in circulating CHIKV strains in 2015 but were not detected in strains from the 2016–2017 outbreak. These “disappearing” or “unstable” mutations included, 2 positions in C protein (11, 82), 3 positions in E1 protein (31, 55 and 104), 1 position in E2 protein (284), 1 position in nsP1 protein (288), 6 positions in nsP2 protein (130, 145, 149, 157, 213 and 375), 14 positions in nsP3 protein (29, 31, 36, 38, 54, 55, 217, 326, 331, 337, 382, 445, 468 and 470) and 4 positions in nsP4 protein (55, 85, 424 and 555).

3.2. Immune epitope analysis

3.2.1. B cell epitope prediction

To appreciate the immunological relevance of these mutations, we undertook a comparison of predicted B cell epitopes across the CHIKV genome. The 2 novel mutations were not found to alter the humoral antigenicity of corresponding proteins (Supplementary Table S1). However, there were three instances where B cell epitopes disappeared as compared to 2015. This loss of B cell antigenicity was observed in nsP2:375, nsP3:29 and nsP3:426. While the reduction in B cell antigenicity at nsP2:375, nsP3:29 could be attributed to the unstable mutations which appeared at these positions during 2015, the disappearance of B cell epitope at nsP3:426 could be because of a change in the neighboring residues. The B cell antigenicity of all the notable mutations has been displayed in Figure 2A. We did not observe any novel B cell epitopes which appeared due to mutation in 2016–2017. However, there were certain amino acid residues such as E1:211 and E2:3212 which were stable as B cell epitopes from 2010 till 2017. Apart from the B cell antigenicity of notable mutations, we also investigated all the B cell epitopic residues. We found 25 such locations which were high confidence B cell epitopes. The epitopic residues and their scores are displayed in Figure 2B.

3.2.2. HLA I binding prediction

The protein sequences were predicted for binding affinity to the reference panel of 27 HLA alleles catalogued in IEDB database. We observed two instances (E2:72; nsP3:471) where the HLA binding affinity for allele HLA-A*30:01 (for E2:72) and HLA-A*23:01 & HLA-A*24:02 (for nsP3:471) were found to be reduced from high to intermediate affinity in our isolates (Supplementary Table S2). While no transition from strong to weak HLA I binding was observed in any of the strains, strong HLA I binding was found to be retained at 8 positions among all variable sites across all the strains sequenced between 2010 and 2016. Four of these 8 positions belonged to the nsP4 protein [nsP4:55 (HLA-A*02:03), nsP4:75 (HLA-A*31:01 & HLA-A*33:01), nsP4:514 (HLA-A*02:03) and nsP4:578 (HLA-B*40:01)]. Similarly, two of these positions belonged to nsP3 protein [nsP3:238 (HLA-A*02:03), nsP3:352 (HLA-A*02:01 & HLA-A*02:06)], one position belonged to C protein [C23 (HLA-A*68:02) and one position belonged to nsP2 protein [nsP2:539 (HLA-B*15:01)]. Beside the HLA I binding affinity of novel, stable, re-emerging and disappearing mutations, we also observed all the strong HLA I bind in whole proteome of our strains. We found 11 such unique HLA I binders. The list of these mutation, their location and the HLA I binding affinity of associated epitope is mentioned in Figure 2C.

3.2.3. HLA II binding prediction

We observed 12 unique protein positions which were predicted to be strong HLA II binders among all variable sites, in strains reported from the period 2010 to 2016 as well as in our isolates belonging to 2016–2017. These 12 positions included 6K:54 (HLA-DRB1*01:01), C:27 (HLA-DRB5*01:01), E1:291 (HLA-DRB1*04:05), E1:322 (HLA-DRB5*01:01), E2:160 (HLA-DRB1*01:01), E2:375 (HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DRB1*07:01, HLA-DPA1*03:01/DPB1*04:02), E2:377 (HLA-DRB1*01:01, HLA-DPA1*02:01/DPB1*01:01, E3:8 (HLA-DRB1*01:01), nsP2:643 (HLA-DRB1*01:01), nsP3:126 (HLA-DRB1*04:05) and nsP4:500 (HLA-DRB1*01:01). A decrease in HLA II binding affinity was observed for 7 epitopes, where binding affinity changed from intermediate to weak. To enumerate, C:11 (HLA-DRB1*11:01), E2:76 (HLA-DQA1*05:01/DBB1*03:01), E2:312 (HLA-DRB1*01:01), nsP2:374 (HLA-DRB1*04:05), nsP3:217 (HLA-DRB1*11:01), nsP3:352 (HLA-DRB1*07:01) and nsP3:468 (HLA-DRB1*01:01). Notably, E2:76 was novel and E2:312 reemerged in our isolates (2016–2017 outbreak). Furthermore, there were 7 positions where HLA binding affinity altered from weak to intermediate:nsP3:217 (HLA-DRB1*11:01), nsP3:326 (HLA-DRB1*04:01), nsP3:331 (HLA-DQA1*05:01/DBB1*03:01), nsP3:341 (HLA-DRB1*01:01, HLA-DRB4*01:01, HLA-DRB7*01:01, nsP3:468 (HLA-DRB1*01:01), nsP4:95 (HLA-DRB1*13:02) and nsP4:555 (HLA-DRB4*01:01) (Supplementary Table S3). Similar to HLA I binders, we also catalogued all the strong HLA II binders from all proteins of our strains as mentioned in Figure 2D. We observed 16 such locations which were predicted to be strong HLA II binders.

The E1:I317V mutation, which was reported to be unique mutation from central India in our previous study [8], was predicted to lie in a non-B cell epitope during previous years and remained so in 2016–2017. Similar to B cell epitope, the HLA I binding affinity remained unchanged (weak binder) across the years of study including 2016–2017, which
Figure 1. Phylogenetic analysis of 2016–17 isolates from central India, with genomes from recent outbreaks and representative strains of major lineages of Chikungunya (East Central and South African (ECSA) and Asian Urban (AUL)). Our isolates from 2016 to 2017 are shown with blue circles and red squares respectively.
Table 1. Mutational profile of sequences of Chikungunya strains of this study as compared to the whole genome sequences from India (2010-2016). Mutational spots have only been depicted in the table; while positions showing conserved amino acids, compared with the present study, have been marked with “*”. Re-emerging mutations have been represented in bold and italics; disappearing mutations have been shown in bold; and novel mutations have been displayed as bold and underscored.

| Protein | Position | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | Current study |
|---------|----------|------|------|------|------|------|------|------|--------------|
| 6K      | 8        | *    | ^, V | *    | *    | *    | *    | *    | I            |
| 6K      | 54       | *    | L^* | *    | *    | *    | *    | *    | V            |
| C       | 11       | *    | *    | *    | *    | *    | S    | S    | N            |
| C       | 23       | P^* | *    | *    | *    | *    | *    | S    |              |
| C       | 27       | D^* | V    | *    | *    | *    | *    | I    |              |
| C       | 79       | *    | *    | *    | *    | *    | *    | NA            |
| C       | 82       | *    | *    | *    | *    | *    | S    | S    | N            |
| C       | 147      | *    | *    | *    | *    | *    | *    | R    | A            |
| C       | 156      | *    | *    | *    | *    | *    | *    | A^* | R          |
| E1      | 31       | *    | *    | *    | *    | V    | I    | I^* | M            |
| E1      | 55       | *    | *    | *    | V    | *    | *    | V    | I            |
| E1      | 66       | *    | *    | V    | *    | *    | *    | A    |              |
| E1      | 93       | *    | *    | *    | H^* | *    | *    | Y    |             |
| E1      | 104      | *    | *    | *    | A    | *    | *    | V    |             |
| E1      | 211      | *    | *    | K    | *    | *    | *    | E    |             |
| E1      | 226      | *    | *    | *    | V    | *    | *    | A    |             |
| E1      | 269      | *    | M^* | *    | *    | *    | *    | V    |             |
| E1      | 284      | *    | D^* | *    | *    | *    | *    | E    |             |
| E1      | 291      | *    | *    | *    | I^* | *    | *    | V    |             |
| E1      | 317      | I    | I    | I    | I    | I    | I    | V    |             |
| E1      | 322      | *    | ^, V | *    | *    | *    | *    | A    |             |
| E1      | 374      | *    | *    | *    | *    | A    | *    | V    |             |
| E1      | 57       | *    | G^* | *    | *    | *    | G    | K    |             |
| E1      | 72       | *    | *    | *    | S    | *    | *    | N    |             |
| E1      | 74       | *    | L^* | *    | *    | *    | *    | N    |             |
| E1      | 76       | *    | *    | *    | *    | *    | *    | A,T |             |
| E1      | 79       | *    | *    | G    | *    | *    | *    | E    |             |
| E1      | 82       | *    | *    | R    | *    | *    | *    | G    |             |
| E1      | 135      | *    | *    | *    | *    | *    | *    | M,R | V            |
| E1      | 143      | *    | *    | *    | T    | *    | *    | S    |             |
| E1      | 160      | *    | N^* | *    | *    | *    | *    | T    |             |
| E1      | 164      | *    | A^* | *    | *    | *    | *    | T    |             |
| E1      | 181      | *    | L^* | *    | *    | *    | *    | T    |             |
| E1      | 195      | *    | *^R | *    | *    | *    | *    | Q    |             |
| E1      | 211      | *    | L^* | *    | *    | *    | *    | Q    |             |
| E1      | 245      | *    | *    | *    | S    | *    | *    | N    |             |
| E1      | 252      | *    | *    | *    | Q    | *    | *    | K    |             |
| E1      | 264      | *    | *    | V    | *    | *    | *    | A    |             |
| E1      | 267      | *    | M^* | *    | *    | *    | *    | R    |             |
| E1      | 284      | *    | *    | *    | T    | *    | *    | I    |             |
| E1      | 298      | *    | *    | S    | *    | *    | *    | N    |             |
| E1      | 312      | *    | *    | T    | *    | *    | *    | M,T |             |
| E1      | 344      | *    | A^* | *    | *    | *    | *    | T    |             |
| E1      | 357      | *    | *    | Q    | *    | *    | *    | L    |             |
| E1      | 375      | *    | S^* | *    | *    | *    | *    | T    |             |
| E1      | 377      | *    | *    | T    | *    | *    | *    | I    |             |
| E1      | 386      | *    | *    | V    | *    | *    | *    | A    |             |
| E1      | 390      | *    | *    | *    | *    | *    | *    | M    |             |
| E1      | 8        | *    | *    | *    | V    | *    | *    | C    |             |
| E3      | 13       | *    | *    | I^* | *    | *    | *    | T    |             |
| E3      | 23       | *    | L^* | *    | *    | *    | *    | T    |             |
| E3      | 42       | *    | L^* | *    | *    | *    | *    | V    |             |
| nsP1    | 58       | *    | *    | N    | *    | *    | *    | D    |             |
| nsP1    | 128      | *    | *    | T    | *    | *    | *    | K    |             |
| nsP1    | 167      | *    | *    | *    | V    | *    | *    | I    |             |
| nsP1    | 171      | *    | *    | *    | Q    | *    | *    | R    |             |

(continued on next page)
| Protein | Position | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | Current study |
|---------|----------|------|------|------|------|------|------|------|--------------|
| nsP1    | 172      | *    | L*   | *    | *    | *    | *    | *    | V            |
| nsP1    | 178      | *    | *    | *    | *    | *    | *    | *    | V            |
| nsP1    | 235      | *    | E*   | *    | *    | *    | *    | *    | K            |
| nsP1    | 288      | *    | *    | *    | *    | *    | *    | *    | I*           |
| nsP2    | 377      | *    | T    | *    | *    | T    | *    | *    | M            |
| nsP1    | 384      | *    | *    | M    | *    | *    | *    | *    | L            |
| nsP1    | 385      | *    | L*   | *    | *    | *    | *    | *    | L            |
| nsP1    | 482      | *    | T    | *    | *    | *    | *    | *    | I            |
| nsP1    | 489      | *    | Q*   | *    | *    | *    | *    | *    | R            |
| nsP2    | 508      | *    | L*   | *    | *    | *    | *    | *    | R            |
| nsP2    | 54       | *    | S    | *    | *    | *    | *    | *    | N            |
| nsP2    | 130      | H    | H    | H    | H    | H    | H    | H    | Y            |
| nsP2    | 145      | *    | *    | *    | *    | *    | D*   | D    | D,E          |
| nsP2    | 149      | *    | *    | *    | *    | *    | *    | *    | G            |
| nsP2    | 157      | *    | *    | *    | *    | I*   | *    | T    |              |
| nsP2    | 213      | *    | *    | *    | *    | *    | H*   | *    | N            |
| nsP2    | 374      | *    | H*   | *    | *    | *    | *    | *    | Y            |
| nsP2    | 375      | *    | *    | *    | *    | *    | *    | *    | E            |
| nsP2    | 399      | *    | *    | S    | *    | *    | *    | *    | L            |
| nsP2    | 642      | *    | C*   | *    | *    | *    | *    | *    | Y            |
| nsP2    | 643      | D    | s,S  | *    | *    | *    | *    | *    | N            |
| nsP2    | 692      | *    | *    | T    | *    | *    | *    | *    | I            |
| nsP2    | 704      | *    | *    | *    | *    | R    | *    | *    | K            |
| nsP3    | 11       | *    | *    | *    | *    | F*   | *    | *    | I            |
| nsP3    | 29       | *    | *    | *    | *    | *    | A*   | *    | P            |
| nsP3    | 31       | *    | *    | *    | *    | *    | *    | *    | G            |
| nsP3    | 36       | *    | *    | *    | *    | *    | *    | *    | A            |
| nsP3    | 38       | *    | *    | H    | *    | *    | H*   | H    | Y            |
| nsP3    | 54       | *    | *    | *    | *    | *    | A*   | *    | T            |
| nsP3    | 55       | *    | *    | *    | *    | *    | T    | *    | A            |
| nsP3    | 126      | *    | D*   | *    | *    | *    | *    | *    | N            |
| nsP3    | 175      | *    | *    | V    | *    | *    | *    | *    | I            |
| nsP3    | 217      | *    | Y    | *    | *    | Y    | *    | Y    | H            |
| nsP3    | 238      | *    | *    | N    | *    | *    | *    | *    | S            |
| nsP3    | 247      | S    | *    | *    | *    | *    | *    | *    | P            |
| nsP3    | 326      | *    | P    | *    | *    | P*   | *    | *    | S            |
| nsP3    | 331      | *    | V    | *    | *    | *    | V    | *    | A            |
| nsP3    | 337      | *    | T    | *    | *    | *    | T    | *    | I            |
| nsP3    | 341      | *    | *    | *    | *    | *    | M*   | T    |              |
| nsP3    | 352      | *    | K    | *    | *    | *    | K    | *    | E            |
| nsP3    | 376      | *    | *    | *    | *    | *    | *    | *    | I,T          |
| nsP3    | 382      | *    | A*   | *    | *    | A*   | *    | T    |              |
| nsP3    | 409      | *    | *    | T    | *    | *    | *    | *    | N            |
| nsP3    | 435      | *    | *    | R    | *    | *    | *    | *    | C            |
| nsP3    | 445      | *    | *    | *    | *    | H*   | H*   | R    |              |
| nsP3    | 460      | *    | *    | *    | *    | G    | *    | *    | E            |
| nsP3    | 461      | *    | L*   | *    | *    | *    | *    | *    | P            |
| nsP3    | 462      | *    | S    | *    | *    | *    | *    | *    | N            |
| nsP3    | 468      | *    | *    | *    | *    | *    | Y    | *    | F            |
| nsP3    | 470      | *    | *    | *    | *    | *    | V    | *    | A            |
| nsP3    | 471      | *    | R*   | *    | *    | *    | *    | *    | S            |
| nsP3    | 474      | *    | *    | *    | M    | *    | *    | *    | T            |
| nsP3    | 524      | *    | *    | *    | *    | *    | *    | *    | R            |
| nsP3    | 39       | *    | *    | *    | *    | *    | *    | E*   | K            |
| nsP3    | 55       | *    | *    | *    | *    | N*   | N,R*  | N,S           |
| nsP3    | 75       | *    | T    | *    | *    | *    | *    | *    | A            |
| nsP3    | 85       | R    | R    | R    | R    | R    | R    | *    | G            |
| nsP3    | 177      | *    | *    | S    | *    | *    | *    | *    | A            |
| nsP3    | 254      | *    | *    | T    | *    | *    | *    | *    | A            |

(continued on next page)
could be instrumental in evasion of CD8+ T cell mediated cytotoxicity. However, in case of HLA II binding, we observed that although the HLA II binding affinity for HLA-DQA1*05:01/DQB1*03:01 remained unchanged throughout the years of study (strong binder), mutation E1:1317V was found to be intermediate binder for HLA-DRB1*01:01, HLA-DRB1*08:02, HLA-DRB1*11:01 and HLA-DRB5*01:01; and thus

| Protein | Position | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | Current study |
|---------|----------|------|------|------|------|------|------|------|---------------|
| nsP4    | 424      | *    | *    | *    | *    | **   | ___  | ___  | ___           |
| nsP4    | 500      | *    | *    | *    | *    | *    | *    | ___  | ___           |
| nsP4    | 512      | *    | *    | *    | *    | *    | *    | ___  | ___           |
| nsP4    | 514      | *    | *    | *    | *    | *    | *    | ___  | ___           |
| nsP4    | 555      | *    | *    | *    | *    | *    | *    | ___  | ___           |
| nsP4    | 578      | *    | *    | *    | *    | *    | *    | ___  | ___           |
| nsP4    | 604      | *    | *    | *    | *    | *    | *    | ___  | ___           |

Figure 2. Heatmaps showing antigenic properties of various sites in Chikungunya proteins observed in our study; (A) B cell epitopes prediction scores of notable mutations observed in our study; (B) High confidence B cell epitopes and their scores found in the analysis; (C) Strong/intermediate HLA class I binders and (D) Strong/intermediate HLA class II binders, as observed in our study.
could also be associated with evasion of CD4+ T helper cells in individuals possessing the respective HLA II alleles.

4. Discussion

In this paper, we report that the central Indian CHIKV outbreak of 2016–2017 could be attributed to strains belonging to the ECSA genotype, which demonstrated maximum phylogenetic proximity to a Singapore strain and a New Delhi strain, both reported in 2015. More specifically, our isolates displayed two novel amino acid mutations, which were present in the Singapore strain but absent from previously reported Indian strains, suggesting that transmission of CHIKV between Singapore and India may have occurred. We observed a total of four mutations that had emerged in the 2016–2017 outbreak strains compared to the strains reported from 2010 to 2016. These mutations were distributed across E2 (2), C (1) and nsP3 (1) proteins. We attempted to explore the immunological impact of these mutations and observed a number of them that could reduce the affinity of T cell and B cell epitopes and thus enable the virus to subjugate the protective immune response to this infection.

India has been affected continually by CHIKV outbreaks since its first outbreak in 1963. Among the major outbreaks, CHIKV reemerged in 1973, 2005 and 2010 and caused heavy health and economic losses to the country [4, 5]. Against a backdrop of sporadic cases occurring each year since 2010, a massive outbreak was observed in 2016 which continued until 2017 [22]. In this study, we observed many novel mutations as well as several that have been previously implicated with survival of virus in the host/vector. For example, mutations such as K211E in E1 protein and V264A in E2 protein were present in the sequenced genomes which have already been associated with increased adaptation of the virus in the Aedes aegypti vector [7]. Such dual mutation strains were first observed in 2009 in Kerala (India) and continued to be present in outbreaks almost every year since then [23]. Furthermore, the genomes sequenced in this study were found to have codon for Arg in place of opal stop codon in nsP3 gene which is known to attenuate CHIKV virus induced arthritis and pathology. Notably, A226V mutation in E1 protein which is responsible for viral transmission through Aedes albopictus mosquitoes, in 2006–2007 outbreak, was not present in genomes of our isolates.

In the phylogenetic analysis, the strains are clustered into two groups where 2016 samples are clustered in one group and most of the 2017 samples clustered in other. Since the chikungunya outbreak studied in the paper ranged from the end of 2016 till starting of 2017 as a single outbreak, the reason for two subgroups in phylogeny could be the genetic drift and filtering of the mutation during the course of an outbreak where the landscape of mutations in the beginning of viral outbreak differs slightly from the end of outbreak.

In the past, there have been studies to identify the emerging mutations and their impact on efficacy of potential vaccine candidates in CHIKV strains. However, to the best of our knowledge, there is no study in which the prospect of a CHIKV outbreak has been analyzed in the light of changing antigenicity of the virus. The present study on whole genome sequencing of CHIKV strains builds upon our previous study, which was limited to the partial sequencing of E1 gene during the same outbreak. Our previous study, based on Sanger sequencing, reported the presence of K211E, M269V, D284E, I317V and V322A mutations which were notable because many of these emerged previously in New Delhi strain of 2010. The whole genome sequencing analysis, undertaken in the current study, demonstrated the same mutations in the E1 gene and thus corroborated our earlier observations. Apart from improving the phylogenetic resolution, the approach of whole genome sequencing enabled us to identify mutations in genes outside of the envelope that may also play a role in virus-host interactions.

Our proposed hypothesis, which was based on the emergence of immunomodulatory mutations in this outbreak, is supported by the identification of a range of mutations in several structural and non-structural proteins. Several of these mutations were found to have modest down-regulatory influence on the antigenicity of B-cell and T-cell epitopes and thus collectively evade the protective immune response to the virus and contribute to the massive increase in the number of cases during the 2016–17 outbreak. However, the possible effect of these mutations on improved adaptation to local vector populations also needs to be explored in future entomological studies. Though the signature mutation associated with transmission by Aedes albopictus vector (E1: A226V) was not observed in our study, the role of alternative mutations accounting for vector adaptation can be established only through entomological characterization of locally circulating mosquito populations.

Our study suffered from three limitations. Firstly, the sequencing platform used for sequencing the samples was Oxford Nanopore where error rates are relatively higher. However, it has been shown that when coverage depth is good (i.e. >20x), consensus sequences are generally >99.5% accurate. Sequence correction using the nanopoolish tool (as we have used) can also help to improve accuracy further (up to ~99.99%). So the error rate in these genomes should be minimal but they are, of course, still less reliable than those produced using highly accurate technologies such as Sanger dideoxy sequencing or Illumina. The primer sets used for genome amplification were also designed based on the Asian/American genotype, which may account for low genome coverage in some variable regions. However, additional primer sets were designed to specifically improve coverage of these regions. Secondly, since the assessment of the immunogenicity of CHIKV sequences in terms of B cell epitope and HLA binders, was performed using in silico prediction servers, there is a need to confirm the impact of these mutations in suitable animal models and ex vivo studies. Thirdly, the representation of samples from 2016 was relatively less in this study, owing to the non-availability of adequate samples from the initial months of the outbreak that could meet the quality requirements of the NGS experiment.

5. Conclusion

In this study we have investigated the CHIKV outbreak of 2016–2017 by performing whole genome sequencing of viral samples from central India and analyzed the mutational and immune epitope profile of the viral strains associated with this outbreak with strains from previous years. We report here several novel mutations in structural and nonstructural proteins of CHIKV and also report their potential impact on the occurrence of the outbreak by evading critical B cell and T cell-mediated protective immune responses. The use of the highly accessible and low-cost Nanopore sequencing technology underscores the potential of adopting the same for widespread surveillance of viral pathogens of significant public health interest.

Declarations

Author contribution statement

Sudeer Gupta: Conceptualized and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Ashvini Yadav: Conceptualized and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Sam Stubbs: Conceptualized and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Simon Frost: Conceptualized and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Kudia Ansari, Ram Kumar Nema: Performed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Shashwati Nema: Analyzed and interpreted the data; Wrote the paper.

Debasis Biswas: Conceptualized and designed the experiments; Analyzed and interpreted the data; Wrote the paper.
Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data will be made available on request.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e11400.

Acknowledgements

The authors would like to acknowledge the support provided by the Department of Health Research, Ministry of Health & Family Welfare, Govt. of India, as part of the country-wide network of Viral Research & Diagnostic Laboratories. SK, AKY, RKN and KA are also supported by the same network.

References

[1] E.S. Gudo, J.F. Black, J.L. Cliff, Chikungunya in Mozambique: a forgotten history, PLoS Neglected Trop. Dis. 10 (2016), e0005001.

[2] S.P. Kalantri, R. Joshi, L.W. Riley, Chikungunya epidemic: an Indian perspective, Natl. Med. J. India 19 (2006) 315–322.

[3] E.O. Nsoesie, M.U. Kraemer, N. Golding, D.M. Pigott, O.J. Brady, C.L. Moyes, M.A. Johansson, P.W. Gething, R. Velayudhan, K. Khan, S.I. Hay, J.S. Brownstein, Global distribution and environmental suitability for chikungunya virus, 1952 to 2015, Euro Surveill. 21 (2016).

[4] P.N. Vergolkar, B.V. Tandale, V.A. Arankalle, P.S. Sathe, A.B. Sudeep, S.S. Gandhe, M.D. Gokhle, G.P. Jacob, S.L. Hundekar, A.C. Mishra, Chikungunya outbreaks caused by African genotype, India, Emerg. Infect. Dis. 12 (2006) 1580–1583.

[5] D. Cecilia, Current status of dengue and chikungunya in India, WHO South East Asia, J. Public Health 3 (2014) 22–26.

[6] K.A. Tsetsarkin, D.L. Vanlandingham, C.E. McGee, S. Higgs, A single mutation in chikungunya virus affects vector specificity and epidemic potential, PLoS Pathog. 3 (2007), e201.

[7] A. Agarwal, A.K. Sharma, D. Sukumaran, M. Parida, P.K. Dash, Two novel epistatic mutations (E1:K211E and E2:V264A) in structural proteins of Chikungunya virus enhance fitness in Aedes aegypti, Virology 497 (2016) 59–68.

[8] A. Agarwal, S. Gupta, A.K. Yadav, R.K. Nema, K. Ansari, D. Biswas, Molecular and phylogenetic analysis of Chikungunya virus in Central India during 2016 and 2017 outbreaks reveal high similarity with recent New Delhi and Bangladesh strains, Infect. Genet. Evol. 75 (2019), 103946.

[9] S.R. Santhosh, P.K. Dash, M.M. Parida, M. Khan, M. Tiwari, P.V. Lakshmana Rao, Comparative full genome analysis revealed E1:α26V shift in 2007 Indian Chikungunya virus isolates, Virus Res. 135 (2008) 36–41.

[10] J. Quick, N.D. Grubaugh, S.T. Pullan, L.M. Claro, A.D. Smith, K. Gangavarapu, G. Oliveira, R. Robles-Sikisaka, T.F. Rogers, N.A. Beutler, D.R. Burton, L.L. Lewis-Ximenez, J.G. de Jesus, M. Giovanetti, S.C. Hill, A. Black, T. Bedford, M.W. Carroll, M. Nunes, L.C. Alcantara Jr., E.C. Sabino, S.A. Baylis, N.R. Faria, M. Loose, J.T. Simpson, O.G. Pybus, K.G. Andersen, N.J. Loman, Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples, Nat. Protoc. 12 (2017) 1261–1276.

[11] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (2009) 1754–1760.

[12] S.C.B. Stubbs, B.A. Blacklaws, B. Yohan, F.A. Yudhaputri, R.F. Hayati, B. Schwem, E.M. Salvana, R.V. Destura, J.S. Lester, K.S. Myint, R.T. Sammono, S.D.W. Frost, Assessment of a multiplex PCR and Nanopore-based method for dengue virus sequencing in Indonesia, Virol. J. 17 (2020) 24.

[13] K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, Nucleic Acids Res. 30 (2002) 3059–3066.

[14] A. Larsson, Alamino, AliView: a fast and lightweight alignment viewer and editor for large datasets, Bioinformatics 30 (2014) 3276–3278.

[15] N.J. Loman, J. Quick, J.T. Simpson, A complete bacterial genome assembled de novo using only nanopore sequencing data, Nat. Methods 12 (2015) 733–735.

[16] I. Milne, G. Stephen, M. Bayer, P.J. Cock, L. Pritchard, L. Cardle, P.D. Shaw, D. Marshall, Using Tablet for visual exploration of second-generation sequencing data, Briefings Bioinf. 14 (2013) 193–202.

[17] S. Wang, J.P. Sundaram, D. Sprio, VIGOR, an annotation program for small viral genomes, BMC Bioinf. 11 (2010) 451.

[18] A.B. Schneider, R. Ochsenreiter, R. Hostager, I.L. Hofacker, D. Janies, M.T. Wolfsberg, Updated phylogeny of chikungunya virus suggests lineage-specific RNA Architecture, Viruses (2019) 11.

[19] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874.

[20] H. Singh, H.R. Ansari, G.P. Raghava, Improved method for linear B-cell epitope prediction using antigen’s primary sequence, PLoS One 8 (2013), e62216.

[21] Q. Zhang, P. Wang, Y. Kim, P. Haste-Andersen, J. Beaver, P.E. Bourne, H.H. Bui, J.T. Simpson, O.G. Pybus, K.G. Andersen, N.J. Loman, Multiplex PCR method for Zika virus detection using nanopore sequencing, Emerg. Infect. Dis. 21 (2015) 194–200.

[22] R.N.C. Moizeis, T. Fernandes, P. Guedes, H.W.B. Pereira, D.C.F. Lanza, G. Oliveira, R. Robles-Sikisaka, T.F. Rogers, N.A. Beutler, D.R. Burton, L.L. Lewis-Ximenez, J.G. de Jesus, M. Giovanetti, S.C. Hill, A. Black, T. Bedford, M.W. Carroll, M. Nunes, L.C. Alcantara Jr., E.C. Sabino, S.A. Baylis, N.R. Faria, M. Loose, J.T. Simpson, O.G. Pybus, K.G. Andersen, N.J. Loman, Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples, Nat. Protoc. 12 (2017) 1261–1276.

[23] I. Maljkovic Berry, F. Eyase, S. Pollett, S.L. Konongoi, M.G. Joyce, K. Figueroa, V. Oluwa, H. Koka, E. Koke, A. Nyunja, J.D. Mancuso, R.G. Jarmar, R. Sang, Global outbreaks and origins of a chikungunya virus variant carrying mutations which may increase fitness for Aedes aegypti: revelations from the 2016 manda, Kenya outbreak, Am. J. Trop. Med. Hyg. 100 (2019) 1249–1257.