



OPEN

# Molecular evolution of coxsackievirus A24v in Cuba over 23-years, 1986–2009

Magilé C. Fonseca<sup>1✉</sup>, Mario Pupo-Meriño<sup>2</sup>, Luis A. García-González<sup>3</sup>, Sonia Resik<sup>1</sup>, Lai Heng Hung<sup>1</sup>, Mayra Muné<sup>1</sup>, Hermis Rodríguez<sup>4</sup>, Luis Morier<sup>5</sup>, Heléne Norder<sup>6,8</sup> & Luis Sarmiento<sup>7,8</sup>

Coxsackievirus A24 variant (CVA24v) is a major causative agent of acute hemorrhagic conjunctivitis outbreaks worldwide, yet the evolutionary and transmission dynamics of the virus remain unclear. To address this, we analyzed and compared the 3C and partial VP1 gene regions of CVA24v isolates obtained from five outbreaks in Cuba between 1986 and 2009 and strains isolated worldwide. Here we show that Cuban strains were homologous to those isolated in Africa, the Americas and Asia during the same time period. Two genotypes of CVA24v (GIII and GIV) were repeatedly introduced into Cuba and they arose about two years before the epidemic was detected. The two genotypes co-evolved with a population size that is stable over time. However, nucleotide substitution rates peaked during pandemics with  $4.39 \times 10^{-3}$  and  $5.80 \times 10^{-3}$  substitutions per site per year for the 3C and VP1 region, respectively. The phylogeographic analysis identified 25 and 19 viral transmission routes based on 3C and VP1 regions, respectively. Pandemic viruses usually originated in Asia, and both China and Brazil were the major hub for the global dispersal of the virus. Together, these data provide novel insight into the epidemiological dynamics of this virus and possibly other pandemic viruses.

Human enteroviruses are small, non-enveloped viruses belonging to the *Enterovirus* genus of the *Picornaviridae* family<sup>1</sup>. The genome is a single-stranded positive sense RNA molecule of approximately 7.4 kb and possesses a long open reading frame (ORF) that is flanked on both ends by the 5' and 3' untranslated regions. The ORF encodes a polyprotein, which is cleaved to form seven non-structural proteins (2A, 2B, 2C, 3A, 3B, 3C, and 3D) and four structural proteins (VP1, VP2, VP3, and VP4). Human enteroviruses comprise four species, namely, species *Enterovirus A–Enterovirus D*. Each of these four species is formed by five to 63 different types, with no cross neutralization, i.e. infection with one type does not infer immunity against another type<sup>1</sup>. Recombination has often been reported either between members of the same or different human enteroviruses species. The recombination “hotspot” regions are located not only between the structural and non-structural coding regions but also between the 5' non-coding region and the protein-encoding region of human enteroviruses<sup>2</sup>.

Coxsackievirus A24v (CVA24v), an antigenic variant of the CVA24 strain (member of species *Enterovirus C*), was first isolated from an outbreak of acute hemorrhagic conjunctivitis (AHC) in Singapore in 1970. Afterwards, CVA24v has been identified as the major causative agent of AHC outbreaks worldwide<sup>3–5</sup>. Currently, four genotypes of CVA24v have been described (I–IV), which have been responsible for major global epidemics of AHC<sup>5–7</sup>. Several studies have investigated the genetic diversity and molecular evolution of CVA24v strains during periods varying from 4 to 20 years<sup>5–7</sup>. Most of these studies were performed on strains of CVA24v that

<sup>1</sup>Virology Department, Center for Research, Diagnosis and Reference, Institute of Tropical Medicine “Pedro Kourí” (IPK), Novia del Mediodía Km 61/2, La Lisa, Marianao 13, P.O. Box: 601, Havana, Cuba. <sup>2</sup>Departamento de Bioinformática, Centro de Matemática Computacional, Universidad de las Ciencias Informáticas (UCI), Havana, Cuba. <sup>3</sup>Departamento de Ciencias de la Computación, Centro de Investigación Científica y de Educación Superior de Ensenada, Ensenada, Baja California, México. <sup>4</sup>Cell Culture Laboratory, Center for Research, Diagnosis and Reference, Institute of Tropical Medicine “Pedro Kourí” (IPK), Havana, Cuba. <sup>5</sup>Department of Microbiology and Virology, Biology Faculty, Havana University, Havana, Cuba. <sup>6</sup>Department of Infectious Diseases/Virology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. <sup>7</sup>Immunovirology Unit, Department of Clinical Sciences, Skåne University Hospital, Lund University, Malmö, Sweden. <sup>8</sup>These authors jointly supervised this work: Heléne Norder and Luis Sarmiento. ✉email: magile@ipk.sld.cu

originated from Asia, however, there have been few phylogenetic studies of CVA24v-related AHC outbreaks in the Americas, as in Brazil, French Guiana and Mexico<sup>8–11</sup>.

The first description of AHC in Cuba is related to an outbreak caused by enterovirus 70 (member of species *Enterovirus D*) in 1981, with more than 800,000 cases<sup>12,13</sup>. From then until 2009, five major outbreaks of AHC due to CVA24v have been documented by the Cuba's National Epidemiologic Surveillance. The first CVA24v AHC outbreak emerged in the municipality of Isla de la Juventud in September 1986 and rapidly spread across the country causing an epidemic of 607,159 cases during the 1986–1987 period. The second nationwide outbreak occurred in 1992–1993 with 90,884 cases<sup>12,13</sup>. It is still unclear how the virus emerged in 1993 because there is no published data on possible AHC outbreaks in the Caribbean islands during this period. Only one sequence of CVA24v isolated in the Dominican Republic during 1993 is available in GenBank. In addition, there were only few reported AHC outbreaks around the world during the early 1990s, those reported occurred in Asia and Africa<sup>14–16</sup>.

The third Cuban AHC outbreak occurred in 1997 spring with 137,136 cases. It began in Havana city and rapidly spread to the entire country<sup>12,13</sup>. During 1997 and 1998, there were numerous outbreaks caused by CVA24v in several Latin American countries, as Antigua/Barbuda, Bahamas, British Virgin Islands, St. Christopher/Nevis, Trinidad and Tobago, Suriname, Puerto Rico and Mexico<sup>17,18</sup>. CVA24v epidemics were also reported in China and India in 1997 and 1999, respectively<sup>19,20</sup>. The fourth large-scale outbreak of AHC (171,910 cases) due to CAV24v in Cuba occurred from July through December 2003<sup>12</sup>. The fifth Cuban outbreak of AHC (72,138 cases) occurred in 2008–2009<sup>21</sup>. There were reports of AHC epidemic in Honduras, Brazil, Taiwan, China and India between 2007 and 2010<sup>8,22–25</sup>.

Given the fact that CVA24v outbreaks has been occurring regularly in Cuba<sup>12</sup>, this study was performed to form basis for the understanding of the evolutionary and epidemiological dynamics of CVA24v in the Americas. The sequences of CVA24v strains from five AHC epidemics in Cuba were used to determine the viral phylodynamics and the phylogenetic relationships between strains involved in the outbreaks and strains isolated from other parts of the world during the same periods.

## Results

**Identity analysis.** A total of 157 partial 3C sequences (507 nt) and 149 partial VP1 sequences (234 nt) were obtained from 159 strains of CVA24v isolated during the epidemic periods from 1985 through 2005 (Supplementary Table S1). Our failure to obtain sequence CVA24v products in 6.3% (VP1) and 1.3% (3C) of CVA24v isolates could be attributed to the degradation of viral RNA due to prolonged storage or multiple freeze–thaw cycles. In addition, failure to achieve amplification in VP1 region to a greater extent than 3C may be due to VP1 primers' reliance on conserved amino acid motifs specific to the *Enterovirus* genus. The 3C and VP1 sequences of the Cuban strains had more than 88.3% nucleotide and 93.0% amino acids identity with the corresponding regions of the prototype EH24\_70\_Singapore 1970 (Supplementary Table S2). The Cuban CVA24v strains isolated during two consecutive years in four outbreaks of AHC (i.e., 1986–1987, 1992–1993, 2003/2005, 2008–2009) were greater than 97.0% identical at the nucleotide level during each outbreak (Supplementary Table S3).

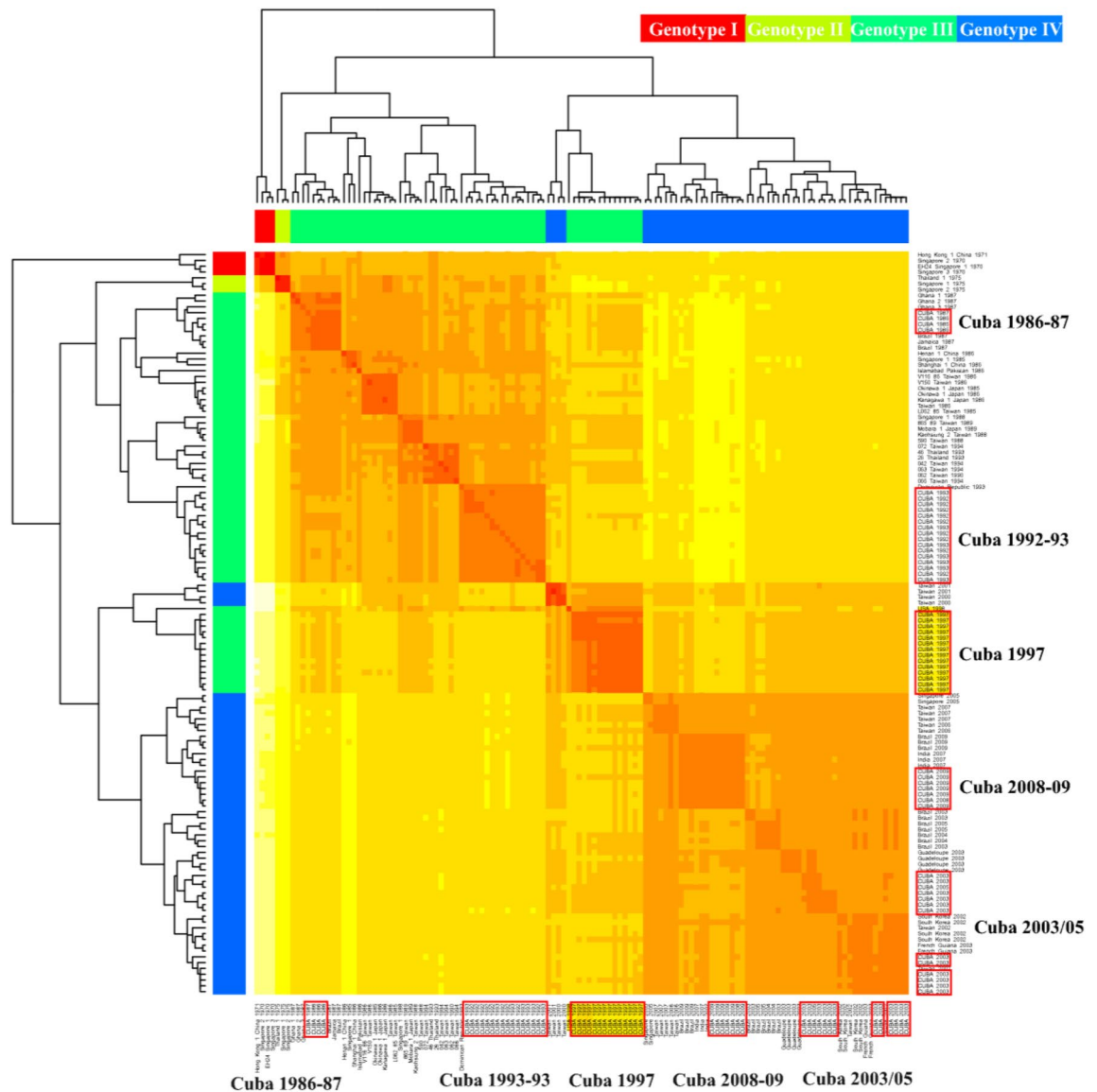
**Molecular epidemiology of CVA24v in Cuba.** The removal of the identical sequences in 3C and VP1 region in a random manner using the ElimDupes tool (<https://www.hiv.lanl.gov>) resulted in 54 Cuban 3C sequences that were compared with 83 partial 3C sequences from strains isolated in 17 countries. Likewise, 35 Cuban VP1 sequences were compared with 95 published sequences of CVA24v strains isolated from 18 countries. Sequences of Cuban strains and globally isolated strains that were obtained from the GenBank database are listed in Supplementary Table S4 and S5.

No saturation was observed neither in the plot of the absolute number of transitions and transversions versus genetic distance nor in the Xia test (Supplementary Fig. S1). The noise analysis showed a good resolution of quartet trees with only 8.4% and 16.5% of points located into partly solved and unresolved quartet area for 3C and VP1, respectively (Supplementary Fig. S2). In addition, no recombination events in the 3C or VP1 regions were demonstrated within each CVA24v sequence selected or between CVA24v and other *Enterovirus C* strains (Supplementary Table S6 and S7).

The GTR nucleotide substitution model with a gamma rate distribution plus invariable sites (GTR + G + I) was identified as the best-fit evolutionary model by jModelTest v2.1.4 program<sup>26</sup>. Heat maps of average nucleotide identity matrix revealed that CVA24v strains isolated from the Cuban AHC epidemics in 1986–1987 and 1992–1993 belonged to genotype III. While this may be expected, given the time periods they were isolated, it is noteworthy that Cuban CVA24v strains isolated in 1997 were clustered into genotype IV. Remarkably, a notably higher intragroup identity than intergroup identity was demonstrated in both coding regions. This suggests that the accepted classification of CVA24v genotypes should be reconsidered (Fig. 1 and Supplementary Fig. S3).

The phylogenetic trees based on the 3C sequences showed that the Cuban CVA24v strains isolated in 1986 and one strain isolated in 1987 formed a clade with CVA24v strains isolated during outbreaks in Jamaica, Brazil and Ghana in 1987 (>98% nucleotide identity) (Fig. 2). These results confirm previous epidemiological data suggesting that 1986 outbreak of AHC in Cuba originated from the introduction of CVA24v by Ghanaian students that arrived to Isla de la Juventud during the summer of 1986<sup>13</sup>. Phylogenetic analysis based on VP1 region revealed a close relation between Cuban strains and Latin-American strains isolated in 1987 (98.7–99.1% nucleotide identity) (Fig. 3). The VP1 region of the Ghanaian strains was not available in GenBank.

Sixteen Cuban CVA24v strains isolated during the 1992–1993 AHC outbreak formed a clade with a Dominican strain isolated in 1993, sharing 97.8–99.8% nucleotide identity in 3C region (Fig. 2). In the VP1 region, the six Cuban strains from 1992 to 1993 formed an independent clade, separated in time to the common ancestor



**Figure 1.** Heat map from nucleotide identity matrix of the 3C region alignment. Sequences were classified by GI-IV Genotypes described by Chu et.al.<sup>7</sup>. Genotypes are indicated by the color legend on the top and in the left side in correspondence with the genotype's clade distribution. Cuban (1997) and USA (1998) sequences are highlighted in yellow. Cuban sequences from five AHC epidemics are highlighted in red rectangles.

of the Dominican strain. Curiously, the Dominican strain was included within the clade formed by the Cuban 1997 sequences (Fig. 3).

The phylogenetic trees based on the VP1 and 3C sequences of Cuban 1997 CVA24v strains showed that these strains formed a separate clade together with a strain isolated in the Unites States in 1998 (Figs. 2, 3).

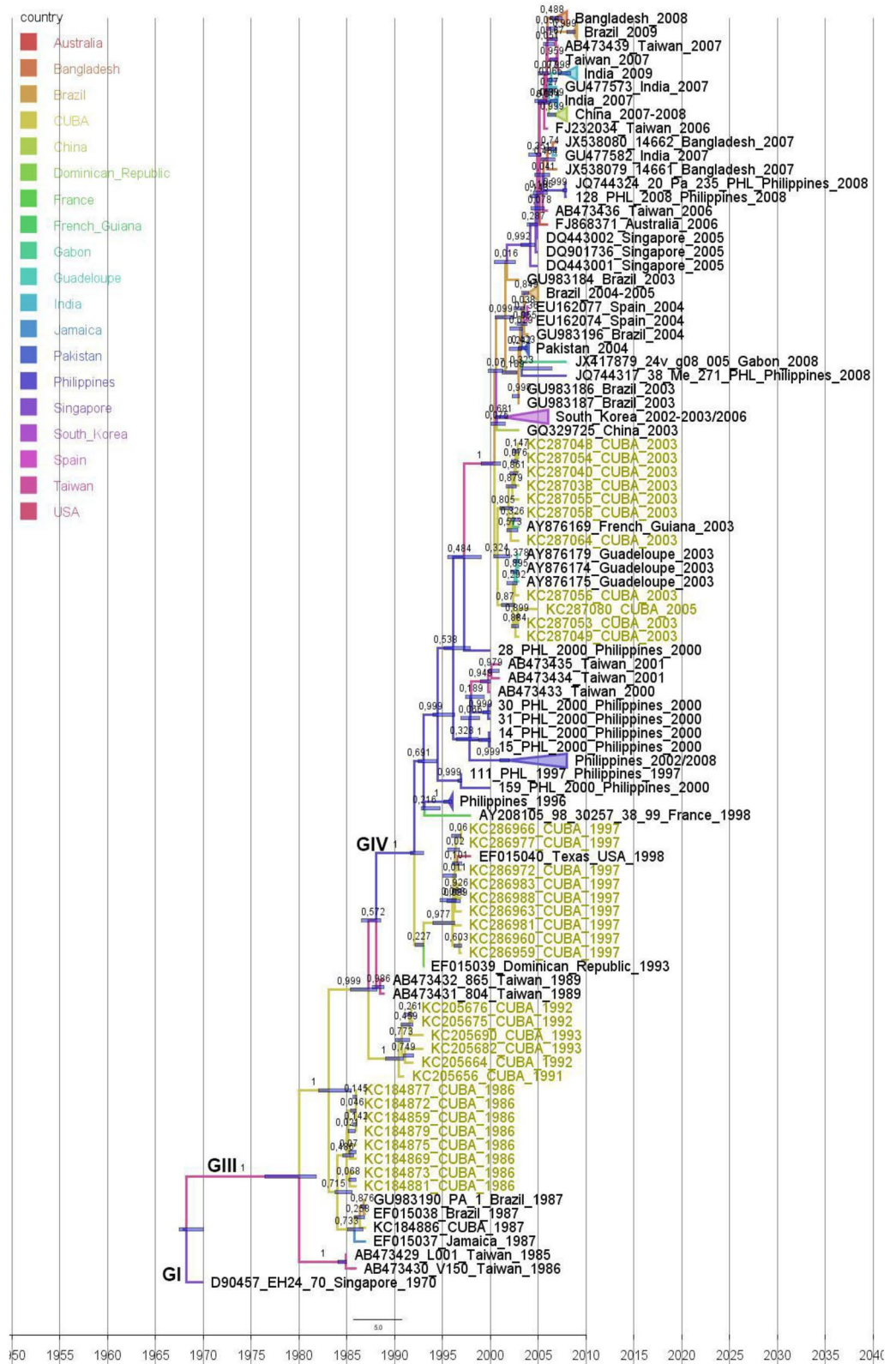
The inconsistency in the 1990s topology suggests the occurrence of at least one intratypic recombination event, which was verified by testing complete sequences of some CVA24v and other *Enterovirus C* strains (Supplementary Table S8) using additional recombination methods included in the RDP4 program<sup>27</sup>. This analysis showed that the Dominican strain isolated in 1993 is a recombinant virus with a major parental sequences (JN228097: Human coxsackievirus A24 variant South Korea 2004 and EF015037: JAM87-10628) that share similarity with Cuban strains isolated the same year or contemporary, which were classified within genotype III and IV (Supplementary Table S9).

Beyond inconsistency between 3C and VP1 trees topologies, the sequences from the late 1990s and early 2000s are located into the same branch forming a single and well-defined clade. This observation in conjunction with the fact that heat map analysis displays a close nucleotide identity between sequences dating from the late-1990s and the early-2000s, support the idea of considering late 1990s sequences as the first occurrence of Genotype IV.

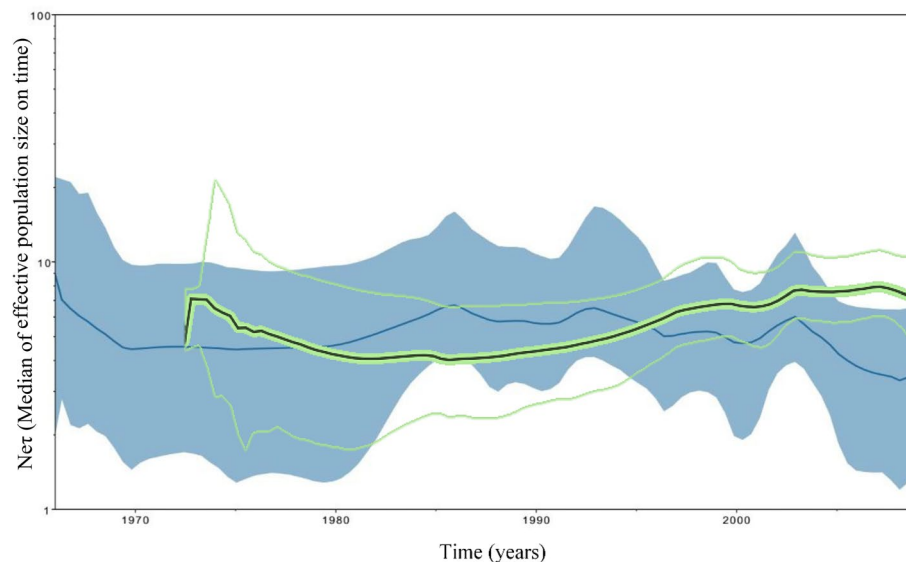
Phylogenetic analysis of the 3C region of 12 Cuban strains isolated in 2003 and one strain in 2005 revealed a separation of the strains into two clades (Fig. 2). One of the clades was formed by six strains from 2003 and the strain from 2005 together with four isolates from an AHC outbreak in Guadeloupe in 2003, sharing 95.6–99.8% nucleotide identity. The other clade was formed by six Cuban strains from the 2003 AHC outbreak in Cuba,



**Figure 2.** Maximum clade credibility (MCC) phylogeny tree of 3C sequences (507 nt) of Cuban (n=54, in red) and worldwide CVA24v strains (n=83). For each branch, the color indicates the most probable location state of their descendent nodes. Bars at nodes indicate 95% HPDs of TMRCA. Branches forming genotypes GI–GIV are shown. The sequences are indicated by GenBank accession number, strain name, country and year of isolation.



**Figure 3.** Maximum clade credibility (MCC) phylogeny of VP1 sequences (234 nt) of Cuban ( $n = 35$  in yellow) and worldwide CVA24v strains ( $n = 95$ ). For each branch, the color indicates the most probable location state of their descendent nodes. Bars at nodes indicate 95% HPDs of TMRCAs. Branches forming genotypes GI–GIV are shown. The sequences are indicated by GenBank accession number, strain name, country and year of isolation.



**Figure 4.** Bayesian skyride plot of Cuban and world-wide of 3C (137 sequences, 507 nt) coding region (blue) and VP1 (130 sequences, 234nt) coding region (green) of CVA24v. The x-axis is the time scale (years) and the y-axis is the logarithmic  $N_{\text{eff}}$  scale ( $N_{\text{eff}}$  is the median of effective population size and  $\tau$  is the generation time). The thick solid line indicates the median estimates and the shaded area indicates the 95% HPD.

two Indian strains isolated in 2003 and CVA24v strains that had caused the Brazilian AHC outbreaks during 2003–2005 (97.4–99.6% nucleotide identity). The analyses of the VP1 region revealed that Cuban strains also formed two clades (Fig. 3). In the first clade, three strains from 2003 and one strain from 2005 clustered with three strains isolated in Guadeloupe in 2003, which were 98.2–99.1% identical at nucleotide level with the Cuban strains. In the other clade, seven of the Cuban strains from 2003 clustered with a strain isolated from an outbreak in French Guiana in the same year, sharing 98.2–100% nucleotide identity. Analysis of 3C gene sequences of Cuban CVA24v strains obtained during the 2008–2009 AHC outbreak<sup>21</sup> showed a clade together with strains isolated from India in 2007 and during the AHC outbreak of 2009 in Brazil, sharing 98.2–99.8% nucleotide identity (Fig. 2).

**Phylogenetic analysis of CVA24v based on the sequenced 3C and VP1 regions.** Topologies based on the sequences of the 3C and VP1 regions were similar to each other (Figs. 2, 3). The analysis revealed a monophyletic origin of all strains that diverged from those isolates that circulated during the first epidemic in Singapore in 1970<sup>5,28</sup>. The CVA24v topologies exhibit an unbalanced tree structure towards a “ladder-like” shape, characterized by a sequential replacement of the existing viral populations by other emergent ones on a global scale, and a limited genetic diversity observed through time. Additionally, short terminal branches with low support values were observed, which may suggest that the grouped sequences can be explained by different topology<sup>5,29–31</sup>.

Skyride plots were used for reconstructing temporal variations in the genetic diversity of 137 Cuban and world-wide CVA24v strains for 3C coding region as 130 for VP1 coding region (Fig. 4). The analysis showed that the effective population size of the virus is stable over time with slight increases, coinciding with the pandemic periods within the four epidemiological periods of CVA24v<sup>5,7</sup>. This was confirmed when both the 3C and the VP1 regions were analyzed (Fig. 4). The population was emerging by the 1970s when the virus appeared with an effective median population size of 9.52 for VP1 and 7.04 for 3C. Thereafter, it decreased to 4.43 for 3C and to 2.79 for VP1, coinciding with the limited CVA24v circulation in the Southeast Asia and Indian regions prior to 1985. A virulent period occurred after the mid-1980s and was characterized by many outbreaks worldwide. This was represented by a  $N_{\text{eff}}$  increase for 3C ( $N_{\text{eff}} = 6.70$ ) and a weak increase in VP1 ( $N_{\text{eff}} = 2.95–2.99$ ) between 1984 and 1988. Thereafter a silent period occurred during the 1990s with few outbreaks globally, although two peaks in the 3C region ( $N_{\text{eff}} = 6.53$  and  $N_{\text{eff}} = 5.24$ ) were obtained as well as a slight increase in the VP1 region between 1997 and 1999 ( $N_{\text{eff}} = 5.94–6.52$ ; Fig. 4). Importantly, these peaks were in correspondence with the two epidemic periods at the beginning of the 1990s and after 1996<sup>14–17</sup>. Finally, the viral population size increased exponentially between 2002 and 2003 (3C  $N_{\text{eff}} = 6.01$  and VP1  $N_{\text{eff}} = 8.1$ ). Furthermore, other peak for VP1 ( $N_{\text{eff}} = 8.46$ ) was demonstrated between 2007 and 2008, which is consistent with the re-emerging period in the 2000s where the virus rapidly spread to many countries throughout the world<sup>5,7,24</sup>.

The confidence intervals of the skyline plots are so wide that they do not show significant increases or decreases during particular periods. Thus, effective population size is stable over time when both the 3C and the VP1 partial regions were analysed (Fig. 4). The greatest uncertainty is observed in the early 1970s period when the virus appeared and population is supposed to be emerging. The insufficient sampling due to the small number of sequenced strains in this period could bring us to misestimate the effective population size, as population dynamics is mostly reconstructed as a result of the simulation instead of being data-driven. Those confidence

CV-A24v genotypes	TMRCAs (95% HPD)/(N)	
	3C	VP1
GI	1964.8 (1959.5–1965.9)/(4)	1968.1 (1967.5–1970)/(1)
GII	1972 (1968.8–1973.3)/(3)	NA
GIII	1981.7 (1978–1981.7)/(65)	1980 (1976.4–1981.9)/(23)
GIV	1995 (1993–1995.9)/(65)	1996 (1995.1–1997.8)/(106)

**Table 1.** The TMRCAs estimated for CVA24v genotypes (I–IV) described by Chu et al., based on Bayesian sequence analysis of partial 3C and VP1 regions. N, number of sequences; NA, VP1 sequences not available at Genbank.

Cuban AHC epidemics	TMRCAs (95% HPD)/(N)	
	3C	VP1
1986–1987	1985.3 (1984.2–1985.6)/(4)	1984 (1983.8–1985.6)/(9)
1992–1993	1990.6 (1990–1991.2)/(16)	1990.4 (1989–1990.9)/(6)
1997	1995.7 (1995.3–1996.4)/(14)	1996 (1994–1996.2)/(9)
2003/2005 <sup>a</sup>	2002 (2001.1–2002.4)/(6)	2002.3 (2001.2–2002.5)/(7)
	2001.9 (2001.5–2002.6)/(7)	2002.8 (2001–2002.4)/(4)
2008–2009	2007.9 (2007.4–2008)/(7)	ND

**Table 2.** The TMRCAs estimated for Cuban CHA epidemics based on 3C and VP1 coding regions of CVA24v strains. N, number of sequences; ND, non determined, the VP1 region was not sequenced. <sup>a</sup>The TMRCAs for each clade of Cuban sequences obtained.

intervals become narrower in subsequent periods, but still there is a high variability which is most notably in 3C region. No artefacts associated with the convergence of the MCMC runs were detected, so insufficient sampling or the low variability of the virus could have the major role in the uncertainty on estimations.

Based on divergence time analyses, the time for the most recent common ancestor (95% HPD) for the CVA24v strains was estimated around the 1960s (Table 1). The evolution into a new genotype (I–IV) likely emerged between 4 and 5 years before they spread globally, and some caused pandemics. The analyses also indicate that some CVA24v strains may have been present in Cuba for 1 to 2 years prior to the large outbreak (Table 2).

The mean substitution rate was estimated to be  $4.39 \times 10^{-3}$  nucleotide substitutions per site per year (s/s/y) (95% HPD  $3.48 \times 10^{-3}$ – $5.36 \times 10^{-3}$ ) for 3C (137 sequences) and  $5.80 \times 10^{-3}$  (95% HPD  $4.37 \times 10^{-3}$ – $7.32 \times 10^{-3}$ ) for VP1 (130 sequences).

The discrete phylogeographic analysis revealed 25 transmission routes for VP1 and 19 routes for 3C genomic region (Table 3 and Supplementary Fig. S4). The analyses showed that the strains causing outbreaks in Cuba usually originate in Asia, and that both China and Brazil play key roles in reseeding the virus globally (Table 3 and Supplementary Fig. S4).

A reconstructed network analysis based on a random walk based community detection algorithm (Walktrap analysis<sup>32</sup>) was performed to identify molecular signatures for the spreading of CVA24v strains. In agreement with the results obtained in the phylogeographic analysis, strains isolated in Cuba were shown to be involved in one of two major networks, which included strains from Asia (Supplementary Fig. S5).

Both analyses identified four different global transmission routes of CVA24v to Cuba: (1) the route from China to Jamaica for the 1986–1987 Cuban AHC epidemics; (2) four routes from China, South Korea, Thailand and Taiwan to USA, which were obtained when analysing the 3C region of strains isolated during the 1997 AHC outbreak; (3) two routes from China to French Guiana and from China to Guadeloupe for the 2003 outbreak; and (4) a route from China to India for the 2008–2009 Cuban AHC outbreaks (Table 3 and Supplementary Fig. S5).

## Discussion

Cuba has been an active area of CVA24v circulation in the Americas since the mid-1980s. This study identified that each epidemic of AHC in Cuba was caused by a genetically distinct CVA24v strain, and that the epidemic strains had emerged a few years before they gave rise to new epidemics. We also demonstrated a sequential development of the CVA24v genotypes, each involved in pandemics, similar to what has been shown for other viruses causing major outbreaks or global spread such as influenza A virus, HIV and EV-D68<sup>33,34</sup>. Our finding suggests that the CVA24v strains that caused epidemics of AHC in Cuba and the Americas region came from Asia, which is consistent with previous studies on epidemiological data recorded during AHC outbreaks<sup>8,35–38</sup>. Given the close genetic relationship between the Cuban strains and those from common destination countries in Latin America, it is likely that the epidemic strains may have reached some countries in the region before it caused epidemic in Cuba. The strains could have been brought into Cuba by the frequent travelling between Cuba and Latin American countries. Taken together, these results provide insight into the epidemiological dynamics of CVA24v and possibly other pandemic viruses.

3C			VP1		
Viral transmission routes			Viral transmission routes		
From	To	BF	From	To	BF
Brazil	China	11.1	Australia	Singapore	5.5
Brazil	Jamaica	11.4	Australia	Taiwan	5.9
Brazil	Japan	8.7	Bangladesh	India	17.6
China	United State	43.9	Bangladesh	Taiwan	9.8
China	Guadaloupe	2.3	Brazil	China	23.0
China	French Guiana	7.9	Brazil	Jamaica	4.8
China	India	32.9	Brazil	Pakistan	16.4
China	Jamaica	23.6	Brazil	South Korea	3.8
China	Morocco	4.9	Brazil	Spain	7.8
China	Thailand	3.5	Brazil	Taiwan	11.9
South Korea	USA	3.4	China	Guadaloupe	38.6
South Korea	Taiwan	3.0	China	French Guiana	49.2
Cuba	Taiwan	161.7	China	Jamaica	8.8
USA	DR Congo	3.6	China	South Korea	3.0
Jamaica	USA	6.6	China	Taiwan	14.5
Dominican Republic	Singapore	29.4	South Korea	Taiwan	3.2
Thailand	USA	5.62	Cuba	Taiwan	19.4
Taiwan	USA	432	France	Philippines	7.2
Taiwan	DR Congo	23.7	Philippines	Taiwan	39.7
			Gabon	Pakistan	3.3
			India	Taiwan	13.7
			Pakistan	Spain	9.7
			Dominican Republic	Philippines	4.7
			Dominican Republic	Taiwan	3.1
			Singapore	Taiwan	44.1

**Table 3.** Epidemiologic connections routes of CVA24v with BF > 3.0 based on 3C and VP1 genomic regions. Highlighted with the same color the common routes obtained in both analyses.

There has been a great deal of controversy over the last years regarding which region of the genome is more informative from the phylogenetic point of view<sup>7,39,40</sup>. The first phylogenetic studies of the CVA24v genomes were based on the non-structural 3C region, however, recently analysis of this genomic region has been replaced by the analysis of the structural VP1 region<sup>5,6,14,39</sup>. In this work, both regions were shown equally informative for phylogenetic and phylodynamic analyses. Notably, phylogenetic analyses showed similar outcomes for the two genomic regions, except for the Dominican Republic strain isolated in 1993. This data indicates the occurrence of few recombination events on CVA24v within VP4 structural protein and within 2B nonstructural protein, which is in sharp contrast to what has been repeatedly reported for other *Enterovirus C* strains<sup>41</sup>.

The lack of interspecies recombination events for CVA24v is in agreement with previous reports, suggesting a genomic barrier for recombination when complete 3C and VP1 of CVA24 were analysed<sup>28,40</sup>. However, intratypic recombination events in CVA24v may occur in nonstructural rather than structural proteins as shown for most enteroviruses<sup>40,42</sup>. Clearly, further studies are needed to elucidate possible influence of these recombinations on the development of pandemic CVA24v strains.

The fact that CVA24v has been present in Cuba for 1–2 years preceding the epidemic peak raise the possibility that the virus may have circulated undetected in Cuba before the epidemic was detected through epidemiological and laboratory surveillance. We envisage three possible scenarios to explain the viral circulation during inter epidemic periods, (1) CVA24v circulate at low levels and escape the surveillance system since asymptomatic infection is the most common outcome, (2) the virus cause symptoms that are not commonly associated with CVA24-related outbreak of AHC<sup>15,43</sup> and (3) the virus can persist in a human population due to its ability to nearly escape the humoral immune response. In support of this, recurrent AHC epidemics have been associated with a decline in neutralizing antibody titers against CVA24v after epidemics<sup>44–47</sup>. Indeed, this has been observed in



the Cuban population during the 1986 and 1997 AHC outbreaks<sup>48,49</sup>. Thus, the evolution of the persistent strains needs to be investigated to increase our understanding of low-level transmission of the strains. Further work is necessary to determine if there are several strains circulating when a new epidemic occurs or if the previous strain will be completely eradicated due to the occurrence of a new more virulent CVA24v strain. This knowledge will constitute a basis for best management of future AHC outbreaks.

The time between the appearances of the most common ancestors for the four CVA24v genotypes confirms the high epidemic potential of this virus since new variants evolve fast and become epidemic or pandemic variants. The analysis of the CVA24v strains revealed a sequential development of the genotypes, with strains belonging to genotype IV isolated during the AHC outbreak in Cuba in 1997. These strains were similar to strains isolated in 1996 in the Philippines and in France and the United State in 1998, which indicate an earlier occurrence of genotype IV during mid-1990s, than in the beginning of 2000s<sup>7</sup>. Other studies on CVA24v sequences have reported similar topology of the phylogenetic trees as shown in this study, even if these studies did not relate to time of isolation of the strains or ancestral relationships between the sequences<sup>5,10,25,37,38</sup>. Genotypes of CVA24v have been described; however, the definition of a subgenotype is more obscure. Several studies have shown spatial-temporal difference between strains from different countries by using the 3C and VP1 coding regions or complete genome<sup>6,23,38,40,50,51</sup>. Many of these studies have used different definitions of the strains analysed. Therefore, a consensus is necessary in order to obtain a better classification of CVA24v regarding to which genomic region to analyse and which reference sequences can be chosen by periods and geographical origin.

Repeatedly worldwide spreads of CVA24v mainly from Asia to the rest of the world was shown by the analysis of the viral sequences in this study. This may indicate that there is a higher potential for CVA24v variants to evolve into strains with pandemic potential on the Asian continent, where there is a high density of susceptible hosts. The pandemic strains may also have evolved from previous pandemic strains by accumulating considerable genetic changes during its global transmission. In this scenario, new virus variants that escape from human immune surveillance against the prior pandemic virus can emerge when it is re-introduced into Asia causing new outbreaks and pandemics. This circle with global viral transmissions may take 3 to 7 years as shown previously<sup>14–17,19</sup>. Finally, the re-emerging period during the 2000s has been characterized by extremely rapid spread of the virus throughout the world and was represented by exponential growth of the population size in 2002–2003 and 2007–2008. This is consistent with the epidemiological data from the two pandemic periods observed from the early 2000s to 2004 and around 2005<sup>5,7,24</sup>. Even though such behaviour was not supported in the present study in what respect to population size growth, it did in what concerns rapid spread of the virus throughout the world. Further immunological and molecular studies are needed to elucidate mechanisms and factors determining the re-emergence of CVA24v pandemic strains.

This study has some limitations with regard to, (1) the sanger methodology used for sequencing; which might have resulted in an underestimation of the genetic diversity of the CVA24v strains (2) the analysis was performed only on partial and not complete genomes; (3) the few availability of the CVA24v sequences from all affected regions as from all the pandemics periods to get a better understanding of the evolution and dissemination of this virus; (4) the intrinsic limitations of the phylogenetic analysis and software. Despite these limitations, this study is one of the largest studies performed in the Americas regarding the origin, evolution and routes of transmission of CVA24v in the region. It also includes sequences of CVA24v strains isolated between 1986 and 2009 from both the eastern and western hemispheres. Overall, our findings resolve a long-standing question of when and where the epidemics CVA24v could be originated in Cuba. The unexpected detection of strains of CVA24v belonging to genotype IV during mid-1990s highlights the need to revisit outdated origin of genotype IV. Moreover, they underscore the importance of understanding the global evolution of CVA24 and their pandemic threat. The results of the study not only shed light on the genetic diversity, evolution and global transmission of CVA24v but may also help in identifying new control strategies for future epidemics at the national, regional and global level.

## Methods

**Viruses.** Archived CVA24v strains (n = 159) used in this study were obtained from National Reference Laboratory for Enteroviruses at the Pedro Kouri Tropical Medicine Institute. The strains were isolated from conjunctival swabs (n = 125), feces (n = 32), nasal swab (1) and pharyngeal swab (1) of patients with AHC during the Cuban outbreaks of 1986–1987, 1992–1993, 1997 and 2003–2005 (Table 4). All strains were isolated on Hep2 (HeLa derivative, ECACC 86030501) cells and the identity of the isolate was confirmed by neutralization tests with type-specific antiserum and PCR<sup>48,49,52</sup>.

**RNA extraction and cDNAs synthesis.** Nucleic acids were extracted from 250 µl of infected cell culture supernatant using Trizol (Life Technologies. Gibco BRL; Grand Island, N.Y.USA) and precipitated with isopropanol. The pellet was washed with 1 ml of 75% ethanol, dried and suspended in 50 µl of diethylpyrocarbonate (DEPC)-treated water. RNA was reverse transcribed to cDNA with Oligo(dT)<sub>20</sub> using ThermoScript RT-PCR System (Life Technologies. Gibco BRL Inc.) according to the manufacturer's instructions.

**PCR amplification and sequencing of VP1 and 3C regions.** Published primers of the 5'-half of the VP1-coding region of the genome (primer 222 [5'-CICCIGGIGGIAYRWACAT-3'] and primer 224 [5'-GCI-ATGYTIGGIACICAYRT-3']) were used for PCR amplification and sequencing<sup>53</sup>. The 3C protease region was amplified with primer pair D1 [5'-TACAAACTGTTTGTGGGCA-3'] and U2 [5'-TTCTTTTGATGGTCTCAT-3']<sup>14</sup>. Amplified products were purified using E.Z.N.A Cycle-Pure Kit (Omega Bio-tek. Inc.) according to the manufacturer's instructions. Cycle sequencing reactions with both primers pair were performed by the ABI BigDye Terminator v 3.1 Cycle Sequencing Kit (Applied Biosystems). The ABI PRISM 3100 Genetic Analyser (Applied Biosystems) was used for electrophoresis and data collection. The sequences data were submit-

Cuban AHC epidemics (year)	Number of cases reported	Collection month(s)	Number of stored strains	Sequenced strains	Specimens sources of the strains
1986	596,445	September	27	27	Conjunctival swabs
1987	10,714	August–September	8	8	Conjunctival swabs
1992	3,363	October–December	22	22	Conjunctival swabs
1993	87,807	September	15	15	Conjunctival swabs
1997	137,136	July–September	43	43	Conjunctival swabs (25) Feces (16) Nasal swab (1) Pharyngeal swab (1)
2003	171,910	June–August, November	39	39	Conjunctival swabs (23) Feces (16)
2005	59	November	5	5	Conjunctival swabs
Total			159	159	Conjunctival swabs (125) Feces (32), Nasal swab (1) Pharyngeal swab (1)

**Table 4.** Epidemiological data and number of CVA24v strains from Cuban AHC epidemics. *Source:* National Enterovirus Reference Laboratory, Pedro Kourí Institute of Tropical Medicine (IPK).

ted to GenBank (accession numbers KC128616–KC128648, KC184859–KC184890, KC205656–KC205727, and KC286913–KC287081).

**Identity analysis.** Sequences obtained were edited and the percentage similarity between the Cuban strains and prototype strain EH24\_70\_Singapore 1970 were determined using the Bioedit v7.0.5.3 program<sup>54</sup>.

**Phylogenetic analysis.** The analysis of both coding regions were conducted with the following algorithm: alignment (multiple sequences alignment), quality control, evolutionary model selection, testing phylogeny and phylodynamics (modified from<sup>55</sup>). For the phylodynamic step, inference of population parameters and phylogeographic analysis were performed.

**Data set and multiple alignment.** Sequences from strains isolated from different geographical regions during years or contemporary when outbreaks had occurred in Cuba were obtained from GenBank. Sixteen 3C gene sequences of CVA24v obtained in our previous studies were also included in the phylogenetic analysis. The sequences were derived from CVA24v strains isolated during the Cuban AHC outbreak in 2008 ( $n = 6$ ) and 2009 ( $n = 10$ )<sup>21</sup>. Duplicate sequences were eliminated using the ElimDupes tool from the Alamos HIV database (<https://www.hiv.lanl.gov>). Multiple sequence alignments were performed by MAFFT v7 by using L-INS-I protocol<sup>56</sup>.

**Quality control.** The quality of the sequence alignments of both coding regions resulted from the filtering of duplex sequences. The nucleotide sequences of partial 3C and VP1 regions of Cuban and world-wide CVA24v strains (Supplementary Table S4 and S5) were explored using the following tests. Saturation effects were investigated by plotting the absolute number of transitions and transversions versus genetic distance for all CVA24v selected, using the DAMBE v6.0 software<sup>57</sup>. Additionally, the standard statistical test of Xia et al.<sup>58</sup> was performed to assess whether a set of molecular sequences had experienced substitution saturation. Genetic distances were calculated with the general time reversible (GTR) model at positions 1 + 2 + 3. The noise in the signal was evaluated by the likelihood mapping method implemented at the Tree-puzzle v5.3.rc16 program<sup>59</sup>. RDP4 v4.36 program was used to analyse recombination events within CVA24v alignments resulted from the duplex elimination for 3C and VP1 coding regions and among those strains selected and different strains of enterovirus C species (Supplementary Table S6 and S7 respectively). Potential recombinants were assumed when more than three methods implemented into the program showed significant support for recombination with a Bonferroni-corrected  $P$  value cut-off of 0.05<sup>27</sup>.

**Phylogenetic and phylodynamic analyses.** In order to explore the genotypes of Cuban variants, the unique 3C and VP1 sequences were aligned and used for a heat map analysis by the R statistical package<sup>60</sup>. Column scaling of the heat data was performed to visualise the nucleotide identity matrix taken into account the four CVA24v genotypes as previously described<sup>7</sup>.

Nucleotide substitution models were estimated by jModelTest v2.1.4 program to obtain the best-fit model according to the Bayesian Information Criterion (BIC)<sup>26</sup>. The CVA24v phylogenetic and phylodynamic analysis were inferred by using Bayesian Inference (BI) methodology implemented into Bayesian Evolutionary Analysis Sampling Tree (BEAST) v1.8.4 program<sup>61</sup>. Bayesian Markov Monte Carlo Chain approach (MCMC) was used to analyse the substitution rates per site per year and the time of the most recent common ancestor (TMRCA) with 95% highest posterior density (HPD). It was also used to reconstruct the history of the viral population and spatiotemporal dynamic by using both alignments, 3C and VP1.

The clock model was selected by estimating the marginal (log) likelihood of each model using the path sampling (PS) method described by Baele et al.<sup>62</sup>. This simulation was carried out for models with strict or relaxed molecular clock (using an exponential uncorrelated (UCED) or a lognormal uncorrelated (UCLD) distribution) combined with the Bayesian Skyride Plot (BSP) a nonparametric coalescence model as an epidemiological model for the tree priors<sup>63</sup>. Bayesian MCMC analyses were run with a chain of 70 million of generations for 3C coding region and a chain of 100 million generations for VP1 coding region. Convergence parameters were identified

by Tracer v1.7.1 (<https://tree.bio.ed.ac.uk/software/tracer/>) with the effective sample size (ESS) greater than 200 (ESS > 200). In all cases, the initial 10% of the run was used as “burn-in”.

The Maximum Clade Credibility (MCC) tree was calculated by TreeAnnotator v1.8.4 and then visualized with FigTree v1.4.4. (<https://tree.bio.ed.ac.uk/software/figtree/>).

Phylogeographic analysis was performed using a continuous standard type Markov chain with the Bayesian Stochastic Search Variable Selection method (BSSVS)<sup>64</sup>. A discrete diffusion model was used as a substitution model where the states were the countries where the virus strains were collected. The BSSVS results were visualized with the Spatial Phylogenetic Reconstruction of Evolutionary Dynamics program (SPREAD v0.9.7.1)<sup>65</sup>. The Bayes Factor (BF) test was performed to obtain the statistical data that adequately explained virus dispersal routes. Virus transmission events with BF > 3 were taken as significant. An additional Walktrap analyses was conducted for strains collected within countries with positive transmission routes (BF > 3) in order to detect viral transmission networks of CVA24v<sup>32</sup>.

**Ethics statement.** All procedures performed in the study were in accordance with relevant guidelines and regulations and with the principles of the Declaration of Helsinki. The study protocol was reviewed and approved by the local ethics committee at Pedro Kourí Tropical Medicine Institute under the permit number CEI-IPK 06-17. Patient consent for the use of the samples was waived for this study in view of the fact that the research study was conducted retrospectively from samples obtained for routine diagnostics, which had been de-identified.

## Data availability

A list of NCBI accession number of all sequences analyzed is reported in Supplementary Table S4–S8.

Received: 7 November 2019; Accepted: 21 July 2020

Published online: 13 August 2020

## References

- Zell, R. *et al.* ICTV virus taxonomy profile: Picornaviridae. *J. Gen. Virol.* **98**(10), 2421–2422. <https://doi.org/10.1099/jgv.0.000911> (2017).
- Muslin, C., Mac Kain, A., Bessaud, M., Blondel, B. & Delpyroux, F. Recombination in enteroviruses, a multi-step modular evolutionary process. *Viruses* **11**(9), 859. <https://doi.org/10.3390/v11090859> (2019).
- Chang, C. H. *et al.* The change of etiological agents and clinical signs of epidemic viral conjunctivitis over an 18-year period in southern Taiwan. *Graefes Arch. Clin. Exp. Ophthalmol.* **241**(7), 554–560. <https://doi.org/10.1007/s00417-003-0680-2> (2003).
- Leveque, N. *et al.* Rapid diagnosis of acute hemorrhagic conjunctivitis due to coxsackievirus A24 variant by real-time one-step RT-PCR. *J. Virol. Methods* **142**(1–2), 89–94. <https://doi.org/10.1016/j.jviromet.2007.01.009> (2007).
- Yen, Y. C. *et al.* Phylogenetic characterization of an ocular-tropism coxsackievirus A24 variant. *PLoS ONE* **11**, e0160672. <https://doi.org/10.1371/journal.pone.0160672> (2016).
- Ishiko, H. T. *et al.* Phylogenetic different strains of a variant of coxsackievirus A24 were repeatedly introduced but discontinued circulating in Japan. *Arch. Virol.* **126**, 179–193 (1992).
- Chu, P. Y. *et al.* Molecular epidemiology of coxsackie A type 24 variant in Taiwan, 2000–2007. *J. Clin. Virol.* **45**, 285–291. <https://doi.org/10.1016/j.jcv.2009> (2009).
- Tavares, F. N. *et al.* Molecular characterization and phylogenetic study of coxsackievirus A24v causing outbreaks of acute hemorrhagic conjunctivitis (AHC) in Brazil. *PLoS ONE* **6**, e23206. <https://doi.org/10.1371/journal.pone.2011> (2011).
- Enfissia, A. J. M., Delauneb, D., Delpyrouxb, E., Rousseta, D. & Bessaudb, M. Coxsackievirus A24 variant associated with acute haemorrhagic conjunctivitis cases, French Guiana, 2017. *Intervirology* **60**, 271–275. <https://doi.org/10.1159/000489339> (2018).
- Sousa, I. P. Jr. *et al.* Re-emergence of a coxsackievirus A24 variant causing acute hemorrhagic conjunctivitis in Brazil from 2017 to 2018. *Arch. Virol.* **164**, 1181–1185. <https://doi.org/10.1007/s00705-019-04157-5> (2019).
- Fragoso-Fonseca, D. E. *et al.* Complete genome sequence of a coxsackievirus type A24 variant causing an outbreak of acute haemorrhagic conjunctivitis in southeastern Mexico in 2017. *Arch. Virol.* **165**, 1015–1018. <https://doi.org/10.1007/s00705-020-04552-3> (2020).
- Ministerio de Salud Publica de Cuba. *Conjuntivitis Hemorrágica Epidémica Aguda en Cuba: Caracterización epidemiológica*. RTV 9(1) Enero-Febrero. Preprints at <http://www.sld.cu/galerias/pdf/sitios/vigilancia/rtv0104.pdf> (2004).
- Ministerio de Salud Publica de Cuba. *Conjuntivitis Hemorrágica Epidémica Aguda en Cuba: Caracterización epidemiológica, cronología y situación actual*. 1981–1997. RTV 2(7) Julio 27. Preprints at <https://es.scribd.com/document/50368499/rtv0797> (1997).
- Lin, K. H. *et al.* Genetic analysis of recent Taiwanese isolates of a variant of coxsackievirus A24. *J. Med. Virol.* **64**, 269–274. <https://doi.org/10.1002/jmv.1046> (2001).
- Kosirukvongs, P., Kanyok, R., Sitritantikorn, S. & Wasi, C. Acute hemorrhagic conjunctivitis outbreak in Thailand, 1992. *Southeast Asian J. Trop. Med. Public Health* **27**, 244–249 (1996).
- Kishore, J. & Isomura, S. Detection & differentiation of coxsackie A 24 variant isolated from an epidemic of acute haemorrhagic conjunctivitis in north India by RT-PCR using a novel primer pair. *Indian J. Med. Res.* **115**, 176–183 (2002).
- Centers for Disease Control and Prevention. Acute hemorrhagic conjunctivitis—St. Croix, U.S. Virgin Islands, September–October 1998. *MMWR Morb. Mortal. Wkly. Rep.* **47**(42), 899–901 (1998).
- Centers for Disease Control and Prevention. Acute hemorrhagic conjunctivitis outbreak caused by coxsackievirus A24—Puerto Rico, 2003. *MMWR Morb. Mortal. Wkly. Rep.* **53**, 632–634 (2004).
- Meng, R., Qiu, Y. & Yan, L. Study on etiology of acute hemorrhagic conjunctivitis in Qingdao during 1997. *Zhonghua shiyan he linchuang bingduxue zazhi Chin. J. Exp. Clin. Virol.* **13**(2), 186–187 (1999).
- Madhavan, H. N., Malathy, J. & Priya, K. An outbreak of acute conjunctivitis caused by coxsackie virus A 24. *Indian J. Ophthalmol.* **48**(2), 159 (2000).
- Fonseca, M. C. *et al.* Isolation of coxsackievirus A24 variant from patients with hemorrhagic conjunctivitis in Cuba, 2008–2009. *J. Clin. Virol.* **53**(1), 77–81. <https://doi.org/10.1016/j.jcv.2011.10.006> (2012).
- PAHO/OPS. Conjuntivitis hemorrágica aguda: advertencia—OMS. 24 de octubre, (2009). Preprint at: [https://www.paho.org/hq/index.php?option=com\\_content&view=article&id=1949:2009-alerta-sobre-conjuntivitis-hemorrágica-aguda-only-spanish&Itemid=42346&lang=es](https://www.paho.org/hq/index.php?option=com_content&view=article&id=1949:2009-alerta-sobre-conjuntivitis-hemorrágica-aguda-only-spanish&Itemid=42346&lang=es).
- Kuo, P. C. *et al.* Molecular and immunocytochemical identification of coxsackievirus A-24 variant from the acute haemorrhagic conjunctivitis outbreak in Taiwan in 2007. *Eye* **24**, 131–136. <https://doi.org/10.1038/eye.2009> (2010).

24. Wu, B. *et al.* Genetic characteristics of the coxsackievirus A24 variant causing outbreaks of acute hemorrhagic conjunctivitis in Jiangsu, China, 2010. *PLoS ONE* **9**, e86883. <https://doi.org/10.1371/journal.pone.0086883> (2014).
25. Shukla, D., Kumar, A., Srivastava, S. & Dhole, T. N. Molecular identification and phylogenetic study of coxsackievirus A24 variant isolated from an outbreak of acute hemorrhagic conjunctivitis in India in 2010. *Arch. Virol* **158**, 679–684. <https://doi.org/10.1007/s00705-012-1520-7> (2013).
26. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**(8), 772 (2012).
27. Martin, D. P. *et al.* Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**(1), vev003. <https://doi.org/10.1093/ve/vev003> (2015).
28. Smura, T. *et al.* The evolution of Vp1 gene in enterovirus C species sub-group that contains types CVA-21, CVA-24, EV-C95, EV-C96 and EV-C99. *PLoS ONE* **9**(4), e93737. <https://doi.org/10.1371/journal.pone.0093737> (2014).
29. Volz, E. M., Koelle, K. & Bedford, T. Viral phylodynamics. *PLoS Comput. Biol.* **9**(3), e1002947.32 (2013).
30. Tee, K. *et al.* Evolutionary genetics of human enterovirus 71: origin, population dynamics, natural selection, and seasonal periodicity of the VP1 gene. *J. Virol.* **84**(7), 3339–3350. <https://doi.org/10.1128/JVI.01019-09> (2010).
31. Chu, P. Y. *et al.* Transmission and demographic dynamics of coxsackievirus B1. *PLoS ONE* **10**(6), e0129272. <https://doi.org/10.1371/journal.pcbi.1002947> (2015).
32. Pons, P. & Latapy, M. Computing communities in large networks using random walks. *JGAA* **10**(2), 191–218 (2006).
33. Wasik, B. R. *et al.* Onward transmission of viruses: how do viruses emerge to cause epidemics after spillover?. *Philos. Trans. R. Soc. B* **374**(1782), 20190017. <https://doi.org/10.1098/rstb.2019.0017> (2019).
34. Tokarz, R. *et al.* Worldwide emergence of multiple clades of enterovirus 68. *J. Gen. Virol.* **93**(Pt 9), 1952–1958. <https://doi.org/10.1099/vir.0.043935-0> (2012).
35. Dussart, P. *et al.* Outbreak of acute hemorrhagic conjunctivitis in French Guiana and West Indies caused by coxsackievirus A24 variant: phylogenetic analysis reveals Asian import. *J. Med. Virol.* **75**, 559–565. <https://doi.org/10.1002/jmv.20304> (2005).
36. Chansaenroj, J. *et al.* Epidemic outbreak of acute haemorrhagic conjunctivitis caused by coxsackievirus A24 in Thailand, 2014. *Epidemiol. Infect.* **143**, 3087–3093. <https://doi.org/10.1017/S0950268815000643> (2015).
37. Laxmivandana, R., Yergolkar, P., Rajeshwari, M. & Chitambar, S. D. Genomic characterization of coxsackievirus type A24 strains associated with acute flaccid paralysis and rarely identified Hopkins syndrome. *Arch. Virol.* **159**, 3125–3129. <https://doi.org/10.1007/s00705-014-2129-9> (2014).
38. De, W. *et al.* Phylogenetic and molecular characterization of coxsackievirus A24 variant isolates from a 2010 acute hemorrhagic conjunctivitis outbreak in Guangdong, China. *Virol. J.* **9**, 41. <https://doi.org/10.1186/1743-422X-9-41> (2012).
39. Cabrerizo, M., Echevarria, J. E., Otero, A., Lucas, P. & Trallero, G. Molecular characterization of a coxsackievirus A24 variant that caused an outbreak of acute haemorrhagic conjunctivitis in Spain, 2004. *J. Clin. Virol.* **43**(3), 323–327. <https://doi.org/10.1016/j.jcv.2008.07.017> (2008).
40. Nidaira, M. *et al.* Molecular evolution of VP3, VP1, 3C(pro) and 3D(pol) coding regions in coxsackievirus group A type 24 variant isolates from acute hemorrhagic conjunctivitis in 2011 Okinawa, Japan. *Microbiol. Immunol.* **58**, 227–238. <https://doi.org/10.1111/1348-0421.12141> (2014).
41. Bessaud, M., Joffret, M. L., Blondel, B. & Delpeyroux, F. Exchanges of genomic domains between poliovirus and other cocirculating species C enteroviruses reveal a high degree of plasticity. *Sci. Rep.* **6**, 3883. <https://doi.org/10.1038/srep38831> (2016).
42. Nikolaidis, M. *et al.* Large-scale genomic analysis reveals recurrent patterns of intertypic recombination in human enteroviruses. *Virology* **526**, 72–80. <https://doi.org/10.1016/j.virol.2018.10.006> (2019).
43. Pallansch, M., Oberste, M. S. & Whitton, L. Enteroviruses: polioviruses, coxsackieviruses, echoviruses, and newer enteroviruses. In *Fields Virology* 6th edn (eds Knipe, D. M. & Howley, P. M.) 490–529 (Lippincott Williams & Wilkins, Philadelphia, 2013).
44. Yin-Murphy, M. Simple tests for the diagnosis of picornavirus epidemic conjunctivitis (acute haemorrhagic conjunctivitis). *Bull. World Health Organ.* **54**(6), 675–679 (1976).
45. Christopher, S., Theogaraj, S., Godbole, S. & John, T. J. An epidemic of acute hemorrhagic conjunctivitis due to coxsackievirus A24. *J. Infect. Dis.* **146**(1), 16–19 (1982).
46. Goh, K. T., Ooi, P. L., Miyamura, K., Ogino, T. & Yamazaki, S. Acute haemorrhagic conjunctivitis: seroepidemiology of coxsackievirus A24 variant and enterovirus 70 in Singapore. *J. Med. Virol.* **31**(3), 245–247 (1990).
47. Langford, M. P., Anders, E. A. & Burch, M. A. Acute hemorrhagic conjunctivitis: anti-coxsackievirus A24 variant secretory immunoglobulin A in acute and convalescent tear. *Clin. Ophthalmol.* **9**, 1665–1673. <https://doi.org/10.2147/OPHT.S85358> (2015).
48. Comellas, M., Más, P., Goyenechea, A. & Varcárcel, M. Estudio de un brote de conjuntivitis hemorrágica, Cuba 1986. *Rev. Cub. Hig. Epidemiol.* **27**, 71–79 (1989).
49. Redon, I. A., Lago, P. J., Perez, L. R., Puentes, P. & Corredor, M. B. Outbreak of acute haemorrhagic conjunctivitis in Cuba. *Mem. Inst. Oswaldo Cruz.* **94**(4), 467–468 (1999).
50. Yan, D. *et al.* Molecular identification and phylogenetic analyses of coxsackievirus A24v causing an outbreak of acute hemorrhagic conjunctivitis in Jiangxi, China, in 2010. *Bing Du Xue Bao* **31**(3), 251–257 (2015).
51. Zhang, L. *et al.* Molecular epidemiology of acute hemorrhagic conjunctivitis caused by coxsackie A type 24 variant in China, 2004–2014. *Sci. Rep.* **7**, 45202. <https://doi.org/10.1038/srep45202> (2017).
52. Más, P. L. *et al.* Diagnóstico etiológico de un brote de conjuntivitis hemorrágica. *Bol. Epid. INHEM* **8**, 1–2 (1987).
53. Nix, W. A., Oberste, M. S. & Pallansch, M. A. Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. *J. Clin. Microbiol.* **44**(8), 2698–2704. <https://doi.org/10.1128/jcm.00542-06> (2006).
54. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**, 95–98 (1999).
55. Norström, M., Karlsson, A. & Salemi, M. Towards a new paradigm linking virus molecular evolution and pathogenesis: experimental design and phylodynamic inference. *New Microbiol.* **35**(2), 101–111 (2012).
56. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**(4), 772–780. <https://doi.org/10.1093/molbev/mst010> (2013).
57. Xia, X. & Xie, Z. DAMBE: software package for data analysis in molecular biology and evolution. *J. Hered.* **92**(4), 371–373. <https://doi.org/10.1093/jhered/92.4.371> (2001).
58. Xia, X., Xie, Z., Salemi, M., Chen, L. & Wang, Y. An index of substitution saturation and its application. *Mol. Phylogenet. Evol.* **26**(1), 1–7. [https://doi.org/10.1016/s1055-7903\(02\)00326-359](https://doi.org/10.1016/s1055-7903(02)00326-359) (2003).
59. Strimmer, K. & von Haeseler, A. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**(7), 964. <https://doi.org/10.1093/oxfordjournals.molbev.a025664> (1996).
60. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, 2018. Preprint at <https://www.R-project.org/>.
61. Drummond, A. J., Rambaut, A. & Suchard, M. A. *Bayesian evolutionary analysis sampling trees BEAST v1.8.4, 2002–2016*. Preprint at <https://beast.community/>.
62. Baele, G. *et al.* Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29**(9), 2157–2167. <https://doi.org/10.1093/molbev/mss084> (2012).

63. Minin, V. N., Bloomquist, E. W. & Suchard, M. A. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25**(7), 1459–1471. <https://doi.org/10.1093/molbev/msn090> (2008).
64. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* **5**(9), e1000520. <https://doi.org/10.1371/journal.pcbi.1000520> (2009).
65. Bielejec, F., Baele, G., Rambaut, A., Suchard, M. A. & Lemey, P. SpreadD3: interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol. Biol. Evol.* **33**(8), 2167–2169. <https://doi.org/10.1093/molbev/msw082> (2016).

## Acknowledgements

This work was partially supported by Wood-Whelan Research Fellowships 2007 from the International Union of Biochemistry and Molecular Biology (IUBMB). This work was also partially funded by the VLIR-UOS project “High Performance Computing Software for Bioinformatics Applications” (code ZEIN2014-Z152), coordinated by Department of Computer Science, KU Leuven, Belgium which allowed the computational infrastructure for the analysis on the UCI. The authors thank the researchers who were part of the Enterovirus laboratory in each epidemic period. The authors gratefully acknowledge the Department of Virology, Institute for Infectious Disease Control (SMI), Solna, Sweden, for their valuable technical assistance on the sequence of the strains. We are grateful to Orlando Martínez, MSc, Dr Carlos Schrago and Dr Beatriz Mello, for her bioinformatics support in the early stages of the study. We also thank to professor Lea Necitas G Apostol, RMT, MPH, PhD from the Research Institute for Tropical Medicine-Department of Health and Dr. Oshitani from Tohoku University Graduate School of Medicine of Philippines to provide the Philippines sequences included in the study.

## Author contributions

M.C.F. Proposed the original idea of the study, conceived and designed the study, acquired data, wrote and edited the manuscript. She is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. M.P.M. Contributed to the original idea and the design of the bioinformatics analysis, researched data, contributed to the interpretation of the experiments, revised and edited the manuscript. L.A.G.G. Researched data, contributed to the bioinformatics analysis and interpretation of the experiments, revised and edited the manuscript. S.R., L.H.H., M.M., H.R., L.M.: researched data, revised and edited the manuscript. H.N. and L.S. contributed to the interpretation of the experiments, wrote and edited the manuscript. All authors reviewed and approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-70436-w>.

**Correspondence** and requests for materials should be addressed to M.C.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](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 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020