

## ORIGINAL ARTICLE

# Comparative genomic analysis of two emergent human adenovirus type 14 respiratory pathogen isolates in China reveals similar yet divergent genomes

Qiwei Zhang<sup>1,2</sup>, Shuping Jing<sup>1</sup>, Zetao Cheng<sup>1</sup>, Zhiwu Yu<sup>3</sup>, Shoaleh Dehghan<sup>4,5</sup>, Amirhossein Shamsaddini<sup>5</sup>, Yuqian Yan<sup>1</sup>, Min Li<sup>1</sup> and Donald Seto<sup>5</sup>

Human adenovirus type 14 (HAdV-B14p) was originally identified as an acute respiratory disease (ARD) pathogen in The Netherlands in 1955. For approximately fifty years, few sporadic infections were observed. In 2005, HAdV-B14p1, a genomic variant, re-emerged and was associated with several large ARD outbreaks across the U.S. and, subsequently, in Canada, the U.K., Ireland, and China. This strain was associated with an unusually higher fatality rate than previously reported for both this prototype and other HAdV types in general. In China, HAdV-B14 was first observed in 2010, when two unrelated HAdV-B14-associated ARD cases were reported in Southern China (GZ01) and Northern China (BJ430), followed by three subsequent outbreaks. While comparative genomic analysis, including indel analysis, shows that the three China isolates, with whole genome data available, are similar to the de Wit prototype, all are divergent from the U.S. strain (303600; 2007). Although the genomes of strains GZ01 and BJ430 are nearly identical, as per their genome type characterization and percent identities, they are subtly divergent in their genome mutation patterns. These genomes indicate possibly two lineages of HAdV-B14 and independent introductions into China from abroad, or subsequent divergence from one; CHN2012 likely represents a separate sub-lineage. Observations of these simultaneously reported emergent strains in China add to the understanding of the circulation, epidemiology, and evolution of these HAdV pathogens, as well as provide a foundation for developing effective vaccines and public health strategies, including nationwide surveillance in anticipation of larger outbreaks with potentially higher fatality rates associated with HAdV-B14p1.

*Emerging Microbes & Infections* (2017) 6, e92; doi:10.1038/emi.2017.78; published online 1 November 2017

**Keywords:** adenovirus; China; emergent respiratory pathogen; genome

## INTRODUCTION

Adenoviruses are human pathogens, causing infections and diseases of the respiratory, ocular, gastrointestinal, genitourinary, and metabolic systems, including being linked to obesity.<sup>1</sup> Interestingly, human adenovirus (HAdV) isolates with highly similar genomes are associated with sporadic and limited infections as well as with larger, wide-spread outbreaks, particularly in environments with close-quartered and/or vulnerable populations.<sup>2–14</sup> In terms of public health, although HAdV infections may be highly contagious, they are generally associated with high morbidity rates rather than high mortality rates;<sup>13,15,16</sup> therefore, HAdVs do not command the attention shown to more deadly viral pathogens such as HIV, Ebola virus, and SARS, and MERS coronaviruses. However, in children and immunocompromised individuals, particularly organ transplant recipients, HAdV infections may lead to significant rates of mortality.<sup>13,17–19</sup>

To date, applying high-resolution genomic and bioinformatic approaches has yielded in-depth and clearer views of the natural

variation of HAdVs.<sup>9,12,14,20–37</sup> In particular, the molecular evolution and mechanisms of the genesis of emergent HAdV pathogens are being revealed; for example, recombination appears to be a major driving force<sup>14,20–23,25,26,29,38,39</sup> and zoonosis has been revealed as the origin for at least one important HAdV respiratory pathogen.<sup>36,37,40</sup> Computational analysis of the genomes from these emergent HAdV pathogens has led to laboratory-based experiments to explore consequences of specific mutations and their roles in pathogenic properties;<sup>22,27,29,41</sup> for example, examining the nonsynonymous mutation in the fiber gene that potentially affects host receptor binding, which may contribute to the robustness and altered pathogenic properties of the re-emergent HAdV-14p1.<sup>41</sup>

More than 84 genotypes, including all previously characterized serotypes, have been sequenced, characterized, and classified within seven species A–G (<http://hadv.wgmu.edu/>).<sup>21,24,28–31,42</sup> Of these, species B, C, and E viruses are associated with respiratory diseases.<sup>17,43</sup> Furthermore, species B HAdVs have been divided into

<sup>1</sup>Guangzhou Key Laboratory of Drug Research for Emerging Virus Prevention and Treatment, Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University, Guangzhou, Guangdong 510515, China; <sup>2</sup>Department of Ophthalmology, Howe Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, Massachusetts 02114, USA; <sup>3</sup>Laboratory of Emerging Infectious Diseases and Division of Laboratory Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong 510282, China; <sup>4</sup>Chemistry Department, American University, Washington, D.C. 20016, USA and <sup>5</sup>Bioinformatics and Computational Biology Program, School of Systems Biology, George Mason University, Manassas, Virginia 20110, USA

Correspondence: QW Zhang; Donald Seto

E-mail: zhang.qiwei@yahoo.com; dseto@gmu.edu

Received 26 March 2017; revised 8 August 2017; accepted 21 August 2017

subspecies B1 and B2, based on their genome similarities and restriction enzyme digestion patterns,<sup>44</sup> as well as reports of viruses of these two subspecies showing different tissue tropism.<sup>45</sup> Members of subspecies B1, i.e., HAdV-B3, -B7, -B16, and -B21, are respiratory pathogens, presumably due to their ability to infect cells of the respiratory tract; HAdV-B50 is also a subspecies B1 virus but has not been associated with a specific, if any, disease.<sup>46</sup> Subspecies B2 comprises HAdV-B11, -B14, -B34, -B35, and -B55; of these, HAdV-B11, -B34 and -B35 are recognized as renal or urinary tract pathogens.<sup>17</sup> The remaining two, HAdV-B14 and -B55, are respiratory tract pathogens.<sup>2,14</sup> HAdV-B55 is a 'Trojan Horse' recombinant with a genome that is essentially that of a respiratory pathogen, HAdV-B14, but with the serologic appearance of a renal pathogen.<sup>6–10,14</sup> HAdV-B14 was originally identified in a military trainee population as an acute respiratory disease (ARD) pathogen in The Netherlands in 1955,<sup>2</sup> with a second occurrence reported in a civilian setting in England in the same year (1955).<sup>3</sup> Since then and over the course of approximately fifty years, aside from a few sporadic and limited infections, HAdV-B14 has rarely been reported.

Beginning in 2005, HAdV-B14 re-emerged in the U.S., with multiple outbreaks in both civilian and military settings, as a highly contagious respiratory pathogen that was associated with a 76% hospitalization rate and an 18% fatality rate.<sup>47–49</sup> This fatality rate was alarming and unusual for a HAdV epidemic, particularly in presumably otherwise healthy adults. The same genome type, HAdV-B14p1, appeared in Europe with two ARD outbreaks in Ireland (2009) and the U.K. (2011).<sup>5,11,50</sup> These resulted in the highest fatality rates reported for HAdV-B14 outbreaks, at 33% and 23%, respectively.<sup>5,11,50</sup> In 2011, this pathogen re-emerged in Canada and caused one death in the three patients hospitalized with ARD.<sup>49</sup> In China, there were no HAdV-B14 cases reported until 2010, when HAdV-B14 emerged nearly simultaneously in two geographically distinct locations (Guangzhou and Beijing).<sup>12,32</sup> Following these two reports, at least three additional HAdV-B14-related ARD outbreaks in China were recorded: One occurred in the Gansu Province (2011) in which 43 students in an elementary school setting presented with febrile respiratory illnesses,<sup>51</sup> one occurred in Beijing (2012) in which 30 adults presented with severe symptoms that required hospitalization,<sup>33</sup> and the third occurred in Liaoning Province (2012) in which 24 students in a middle school presented with febrile respiratory illnesses.<sup>52</sup> Unlike the U.S. 303600 strain, these five China strains did not have fatalities associated with them.

Given the severity of the symptoms and the numbers of afflicted individuals in China, the earlier outbreaks, with associated higher morbidity and fatality rates in the U.S., Canada, and Europe, and the putative transmission of HAdV-B14 from U.S. to Asia (South Korea) by military trainees (2007),<sup>53</sup> it is critical to determine the relationships between these China isolates and their global counterparts in order using high resolution genome analysis to understand the epidemiology of this re-emergent pathogen. These data will provide a basis for public health preventive measures, including diagnosis, surveillance, and appropriate protocols for limiting outbreaks and prevention. Specific information of the mutation rates, mutation hotspots, and new variations, particularly at their serologic recognition sites, will play a key role in the development of vaccines against these pathogens. This report presents the bioinformatics analysis of the HAdV-B14p1 Guangzhou strain (GZ01; October 2010), which was the first HAdV-B14 strain isolated in China,<sup>12</sup> and compares it with the contemporaneous Beijing strain (BJ430; December 2010) as well as a subsequent strain isolated in 2012 (CHN2012; February 2012). Provided is a high-resolution view of highly similar

yet intriguingly divergent genomes, linking one of these strains with the globally re-emergent strain isolated in the U.S. in 2005, and suggesting at least two co-circulating lineages of type 14.

## MATERIALS AND METHODS

### Cells, virus stock, and DNA preparation

HAdV-B14p1 GZ01 was isolated from a throat swab of a 17-month-old child hospitalized with acute suppurative tonsillitis in Guangzhou, China (October 2010).<sup>12</sup> Viral DNA was extracted using a Viral DNA Extraction Kit (Omega Bio-Tek Inc Corp; Norcross, GA). The protocol for PCR amplification is described earlier.<sup>54</sup> All the experimental protocols in this study were approved by the institutional ethics committee of Southern Medical University and were carried out in accordance with the approved guidelines. The informed consent for participation in this study was obtained from the guardian of the under-aged participant. Data records of the sample and sample collection are de-identified and completely anonymous.

### Genome sequencing and annotation

The genome sequence of HAdV-B14p1 strain GZ01 was obtained using the Sanger sequencing method following PCR amplification of targeted overlapping 1–2 kb regions that covered the entire genome, as described earlier.<sup>12</sup> The sequence data, collected with an ABI 3730 Genetic Analyzer, provided an average genome coverage of 3 to 5 folds, with both strands represented. Gaps and ambiguous sequences were PCR-amplified using different primers and resequenced. DNA sequence fragments were assembled using the SEQMAN software from the Lasergene package (DNASar; Madison, WI) into a single contig. The genome was annotated with an annotation protocol used for the HAdV-C1 genome analysis,<sup>35</sup> by first dividing the sequence into contiguous 1 kb non-overlapping segments.

### Sequence analysis

Comparison of sequence differences between and spanning the genomes was performed with Genome Mutation Mapper (GMM; unpublished) software developed by the authors. GMM compares two or more genomes for nucleotide differences, noting SNPs and indels of query genomes relative to a reference; the data were confirmed with DNA Sequencher v5.1 (GeneCodes, Inc.; Ann Arbor, MI). Additionally, pairs of genome sequences were aligned using EMBOSS (<http://www.ebi.ac.uk/Tools/emboss/>) to show sequence identity, which provided a visualization of the alignment. Nucleic acid and amino acid sequence percent identities were calculated using software which was part of the EMBOSS package. Pair-wise comparisons of genomes were performed with the LAGAN (Limited Area Global Alignment of Nucleotides) program of mVISTA (<http://genome.lbl.gov/vista/lagan/submit.shtml>).<sup>55</sup>

Phylogenetic analysis of select genes and whole genomes was performed using MEGA 5.1.0 (<http://www.megasoftware.net/megamac.php>), which accepts FASTA files for sequence alignment and uses a Maximum Composite Likelihood method that generated neighbor-joining and bootstrapped phylogenetic trees with 1000 bootstrap replications; for phylogenetic analysis; all other parameters were set by default.

### Genome type identification

*In silico* restriction maps of available genome sequences from all HAdV-14 strains CHN/GZ01/2010, 303600, CHN/BJ430, CHN2012, and prototype de Wit were generated using the Vector NTI 10.3.0 software (Invitrogen Corp.; San Diego, CA, USA) as described in earlier studies.<sup>56–58</sup> These included profiles from restriction enzymes *Bam*HI, *Bcl*I, *Bgl*II, *Bgl*III, *Bst*EII, *Eco*RI, *Hind*III, *Hpa*I, *Sal*I, *Sma*I, *Xba*I,

**Table 1** The genome, fiber, hexon, and E1A sequences of adenovirus species A–G used in this study

| Type              | Strain       | Year isolated | Country      | Sequence      | GenBank accession no. |
|-------------------|--------------|---------------|--------------|---------------|-----------------------|
| HAdV-A12          | Huie         | 1954          | USA          | Genome        | AC_000005             |
| HAdV-B3           | GB           | 1953          | USA          | Genome        | AY599834              |
| HAdV-B7           | Gomen        | 1952          | USA          | Genome        | AY594255              |
| HAdV-B11p         | Slobitski    | 1956          | USA          | Genome        | NC_011202             |
| HAdV-B11a         | 273          | 1969          | Spain        | E1A           | FJ841916              |
|                   |              |               |              | Fiber         | FJ841908              |
|                   |              |               |              | Hexon         | FJ841900              |
| HAdV-B11a         | 6380         | 1997          | USA          | E1A           | FJ841919              |
|                   |              |               |              | Fiber         | FJ841907              |
|                   |              |               |              | Hexon         | FJ841899              |
| HAdV-B11a         | 2474         | 2001          | Taiwan       | E1A           | FJ841922              |
|                   |              |               |              | Fiber         | FJ841914              |
|                   |              |               |              | Hexon         | FJ841906              |
| HAdV-B11a         | 760          | 2002          | Taiwan       | E1A           | FJ841921              |
|                   |              |               |              | Fiber         | FJ841913              |
|                   |              |               |              | Hexon         | FJ841905              |
| HAdV-B11a         | RKI-2797-04  | 2004          | Turkey       | Fiber         | AY972816              |
|                   |              |               |              | Hexon         | AY972815              |
| HAdV-B11a         | SNG1222      | 2005          | Singapore    | E1A           | FJ841920              |
|                   |              |               |              | Fiber         | FJ841912              |
|                   |              |               |              | Hexon         | FJ841904              |
| HAdV-B14p         | de Wit       | 1955          | Netherlands  | Genome        | AY803294              |
| HAdV-B14p         | 16845        | 1974          | Netherlands  | E1A           | FJ841917              |
|                   |              |               |              | Fiber         | FJ841910              |
|                   |              |               |              | Hexon         | FJ841902              |
| HAdV-B14p1        | 2971         | 2007          | USA          | E1A           | FJ841915              |
|                   |              |               |              | Fiber         | FJ841909              |
|                   |              |               |              | Hexon         | FJ841901              |
| HAdV-B14p1        | Dublin       | 2009          | Ireland      | Fiber         | HQ163916              |
|                   |              |               |              | Hexon         | HQ265808              |
| HAdV-B14p1        | 303600       | 2007          | USA          | Genome        | FJ822614              |
| HAdV-B14p1        | BJ430        | 2010          | China        | Genome        | JN032132              |
| <b>HAdV-B14p1</b> | <b>GZ01</b>  | <b>2010</b>   | <b>China</b> | <b>Genome</b> | <b>JQ824845</b>       |
| HAdV-B14p1        | CHN2012      | 2012          | China        | Genome        | JX892927              |
| HAdV-B14          | Gansu2011-01 | 2011          | China        | Hexon         | JX310315              |
| HAdV-B16          | ch.79        | 1955          | USA          | Genome        | AY601636              |
| HAdV-B21          | AV-1645      | 1956          | Saudi Arabia | Genome        | AY601633              |
| SAdV-21           | Bertha       | 1954          | USA          | Genome        | AC_000010             |
| HAdV-B34p         | Compton      | 1972          | USA          | Genome        | AY737797              |
| HAdV-B35p         | Holden       | 1973          | USA          | Genome        | AY128640              |
| HAdV-B50          | Wan          | 1988          | USA          | Genome        | AY737798              |
| HAdV-B55          | QS-DLL       | 2006          | China        | Genome        | FJ643676              |
| HAdV-C1           | Adenoid 71   | 1953          | USA          | Genome        | AF534906              |
| HAdV-D9           | Hicks        | 1954          | USA          | Genome        | AJ854486              |
| HAdV-E4           | RI-67        | 1952          | USA          | Genome        | AY594253              |
| HAdV-F40          | Dugan        | 1979          | Netherlands  | Genome        | NC_001454             |
| HAdV-G52          | T03-2244     | 2003          | USA          | Genome        | DQ923122              |

The genome of HAdV-B14p1 GZ01 is shown in boldface.

and *Xho*I (TaKaRa Corp.; China), all of which were chosen in order to be consistent with the original nomenclature system devised by Li and colleagues.<sup>59</sup> Genome typing of these strains was determined by comparing its *in silico* RE profiles with other HAdV-14 genome types.<sup>4,34</sup>

#### GenBank accession numbers

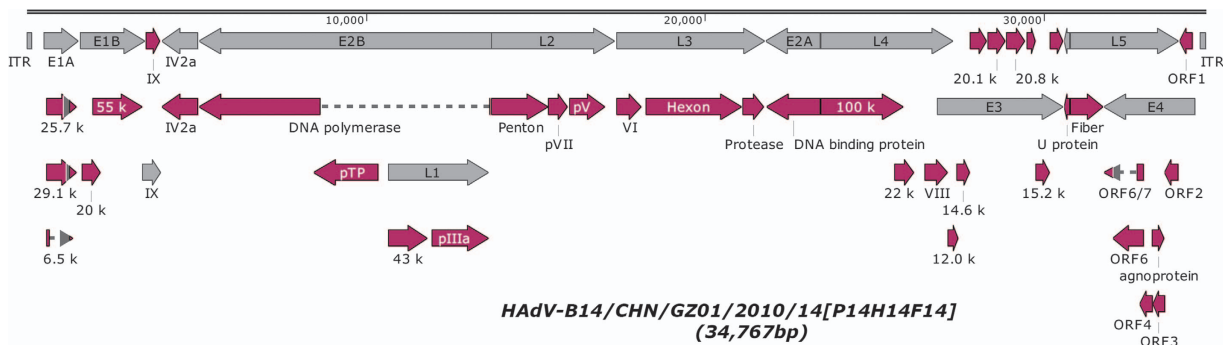
The genome sequences used for phylogenetic analyses are summarized in Table 1, with additional genome typing details given only for the

type 14 strains ('p' denotes prototype; if no further designation is noted, the strain is the prototype). Sequences used for these studies included the E1A and hexon genes, either deposited as single gene entries in GenBank or which were extracted from the genome files.

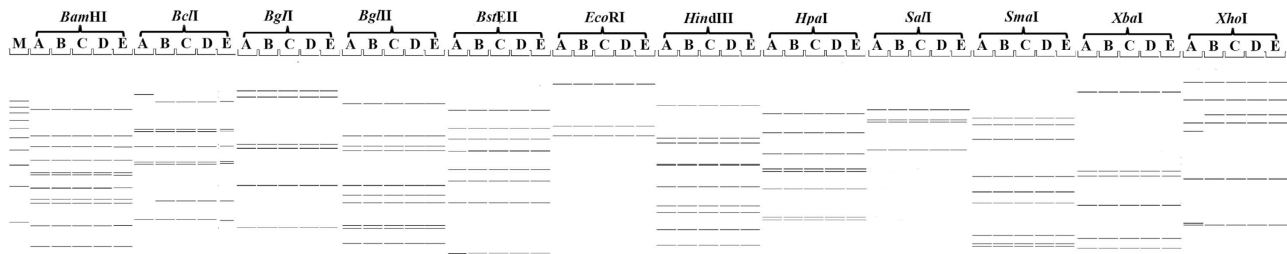
## RESULTS

### Nucleotide sequence analysis of HAdV-B14p1 strain GZ01

The genome of strain GZ01 was sequenced, assembled, annotated, and analyzed using computational methods. Figure 1 presents the genomic



**Figure 1** Genomic organization and transcription map of HAdV-B14p1 strain GZ01. The grey arrows indicate the early, intermediate, and late transcription units; the red arrows indicate coding regions. Arrows reflect the direction of the coding transcripts.



**Figure 2** Identification of the genome types of HAdV-B14 strains GZ01 (Guangzhou), BJ430 (Beijing), and CHN2012 (Beijing). The *in silico* restriction enzyme analysis (REA) digestion patterns of HAdV-14 genomes were generated using the Vector NTI 10.3.0 software (Invitrogen Corp., San Diego, CA, USA). Genome digestion patterns of the strains are as follows: (A) de Wit or prototype (AY803294); (B) GZ01 (JQ824845); (C) 303600, reported as HAdV-B14p1 (FJ822614); (D) BJ430 (JN032132); and (E) CHN2012 (JX892927). These were analyzed with *Bam*HI, *Bcl*I, *Bgl*I, *Bgl*II, *Bst*EII, *Eco*RI, *Hind*III, *Hpa*I, *Sal*I, *Sma*I, *Xba*I, and *Xho*I, as described by Li *et al.*<sup>59</sup> All the three recent China outbreak genomes correspond to the genome type of HAdV-14p1 and have identical REA profiles to each other. They are divergent from the HAdV-B14 prototype (1955). M: 1 kb DNA Sizing Ladder.

organization and transcription map of GZ01. A total of 38 coding sequences were identified. These genome data were deposited in GenBank (accession number JQ824845) under the formal name, preferred by the National Center for Biotechnology Information (NCBI),<sup>24</sup> ‘Human adenovirus 14 isolate HAdV-B/CHN/GZ01/2010/14[P14H14F14]’; in this report, it is shortened to ‘GZ01’. The genome comprises 34 767 bp, with a GC content of 48.83% that is consistent with the other members of subspecies B2 (mean of 49%).<sup>60</sup> A contemporaneous isolate BJ430 contains 34 762 bp.<sup>32</sup> These lengths are very similar to that of other HAdV-B14p1 strains 303600 (Lackland Air Force Base, USA; 2007) and CHN2012 (Beijing, China; 2012),<sup>33</sup> as well as the prototype from 1955; Genome sizes of 34 763 bp, 34 760 bp, and 34 764 bp, respectively.

**Genome type of strains GZ01, BJ430, and CHN2012 as HAdV-B14p1**

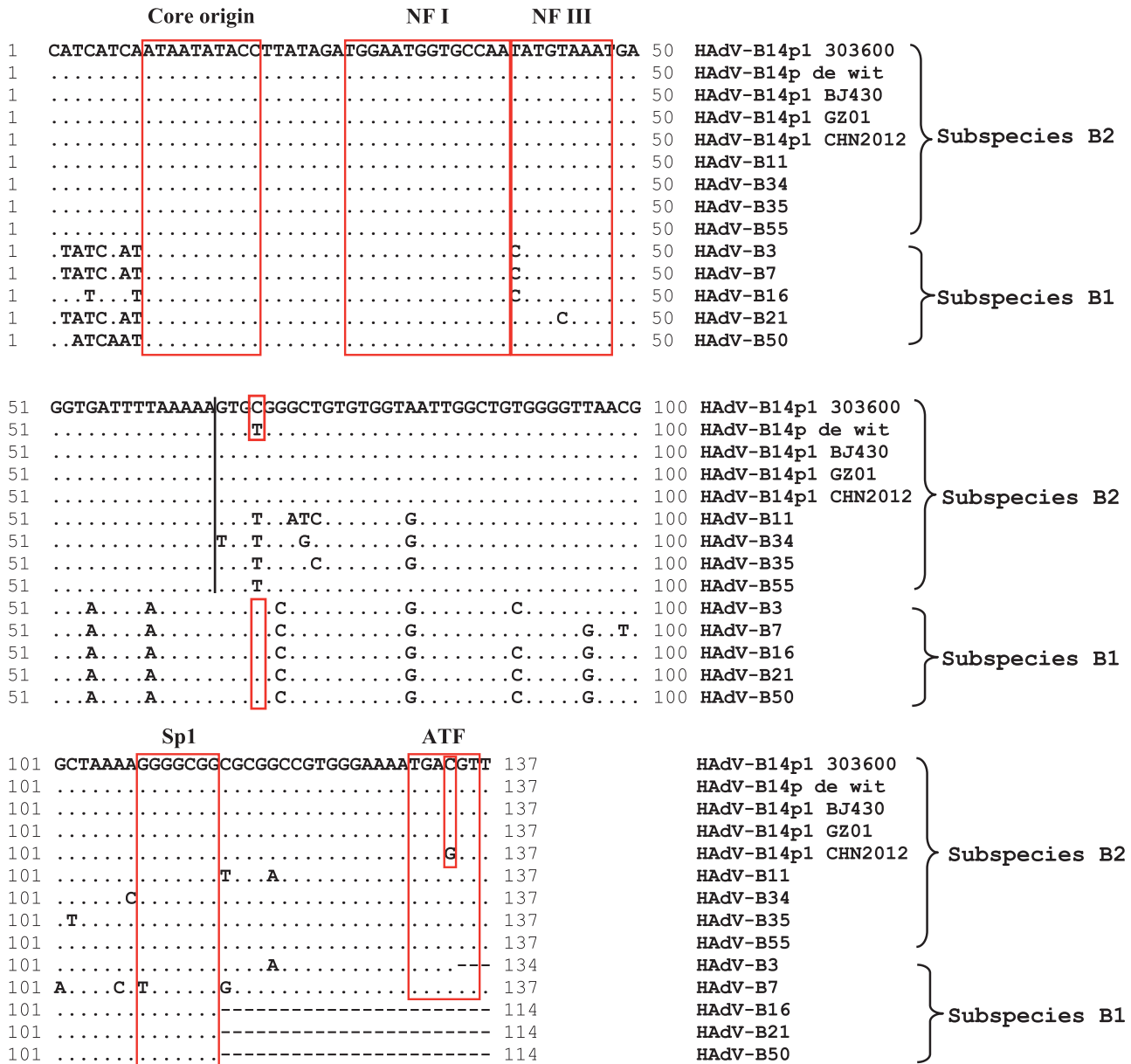
Genome type, based on restriction enzyme analysis (REA), was useful in the past for characterizing strains in the absence of full genome sequence data.<sup>59,61</sup> It has limited use in the era of whole genome data, but may be useful in comparisons with strains reported in the literature with REA maps but are no longer available for further analysis. The genome types of strains GZ01, BJ430, and CHN2012 were determined by *in silico* REA using the Vector NTI 10.3.0 software (Invitrogen Corp.; San Diego, CA, USA) and in comparison with the other HAdV-B14 strains<sup>4,34</sup> (Figure 2). The REA profiles of GZ01, BJ430, and CHN2012 were consistent with and indistinguishable from that of HAdV-14p1 strain 303600, all which were confirmed to be genome type 14p1 according to the nomenclature published previously.<sup>4,34</sup> It is clear, however, that genome type analysis by

REA patterns may be misleading or incomplete as all the three China isolates are highly divergent from HAdV-14p1, as demonstrated by the indel mutations analysis. Furthermore, to support the observation that genome type identification may be misleading, it should be noted that while the initial wet-bench REA data suggested 303600 was a novel HAdV-B14a strain, it was disproven by the whole genome analysis.<sup>4,34</sup>

**Inverted terminal repeat (ITR)**

ITRs contain sequences encoding critical genome replication functions. The HAdV-B14p1 GZ01 genome has a 137 bp ITR that is identical to the ITRs of the other HAdV-B14p1 strains (BJ430, CHN2012, and 303600), except for a C to G mutation at nt 134 of strain CHN2012 (Figure 3). It is nearly identical to the ITRs reported for the other subspecies B2 members. The first 64 bases are completely identical amongst the prototype subspecies B2 ITRs, which is different for divergent subspecies B1. The binding sites for the host transcription factors NF I and NF III were identified in the ITR.<sup>62,63</sup> These sequences have roles in enhancing virus replication and are necessary for efficient virus growth.<sup>62,63</sup>

Transcriptionally-related Sp1 and ATF binding sites were also conserved in subspecies B2 genomes. All four HAdV-B14p1 genomes differ from HAdV-B14p at position 68, with a C present, a transition mutation from the original T; this is also found at the complementary 3'-end, a validation that it is not a sequencing error. This particular T is conserved across all of the subspecies B2 prototypes, including HAdV-B55, but not amongst all the prototypes of the other subspecies, that is, all subspecies B1 members contain a C at this position; this is a marker to distinguish B1 from B2, and also HAdV-



**Figure 3** HAdV-B Inverted terminal repeats (ITR) analysis. An alignment of the left ITRs of select HAdV-B prototypes is presented, with two sets of critical sequence motifs (boxed): (1) Replication- Core origin, NF I and NF III binding sites, and (2) Sp1 and ATF binding sites. Nota bene, ‘.’ represent identical bases and ‘-’ represent deletions or gapping for sequence alignment. HAdV-B1: HAdV-B3, -B7, -B16, -B21, and -B50; HAdV-B2: HAdV-B11, -B14, -B34, -B35, and -B55.

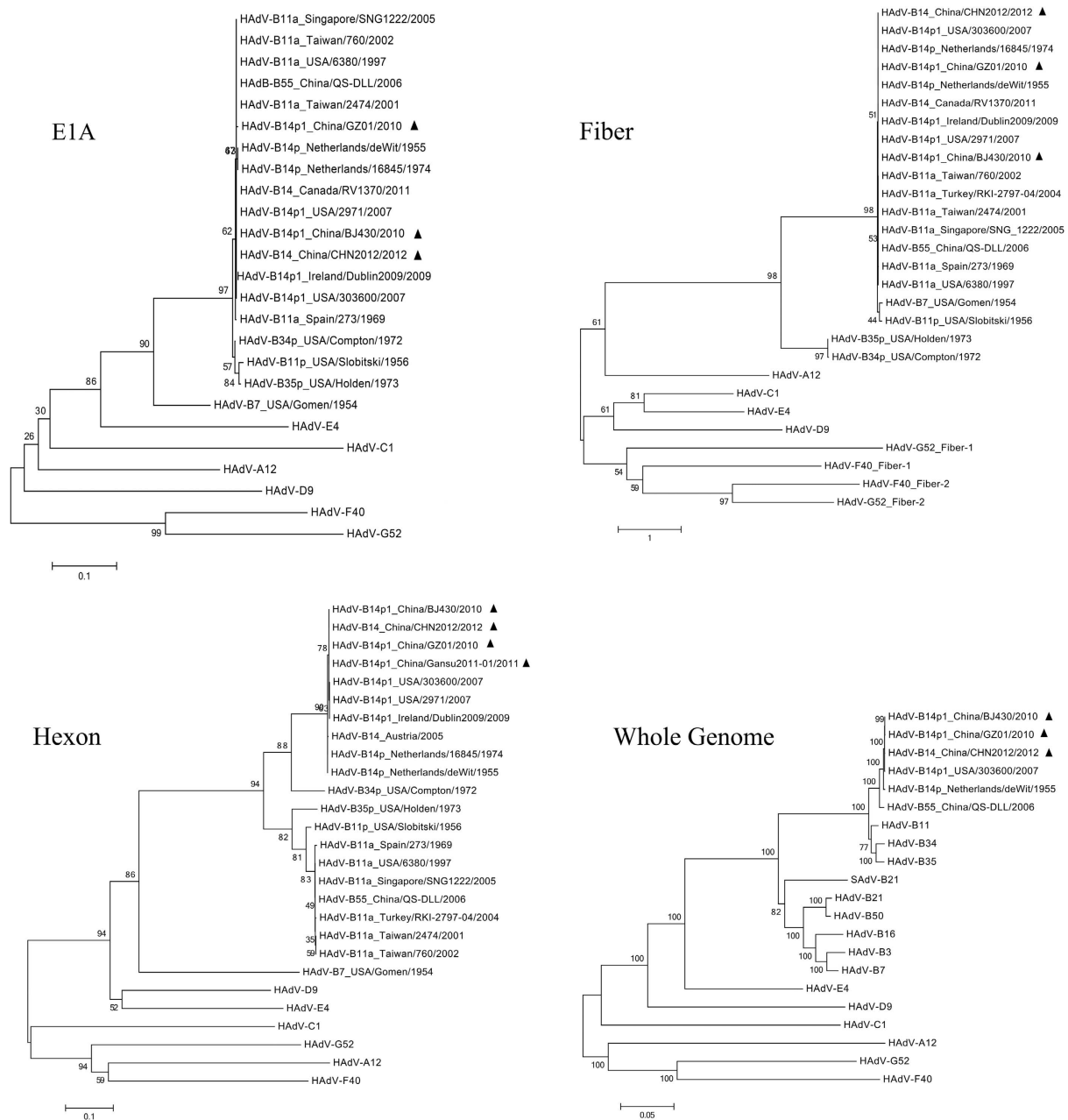
B14p1 from 14p. The nucleotide varies for the other six HAdV species (data not shown).

**Phylogenetic analysis of select genes and the whole genomes**

Of the five recent China strains, only three were reported with whole genome data (GZ01, BJ430, and CHN2012). Phylogenetic analysis of the HAdV-B14 E1A, fiber, and hexon genes shows that the recent HAdV-B14p1 strains (strains GZ01, BJ430, CHN2012, 303600, 2971, and Dublin2009) are closely related to each other as well to the prototype and earlier reported HAdV-B14p genomes (1955 and 1974), as shown in Figure 4. The phylogenetic analysis of the whole genome sequences also confirms the sequence similarity with the prototype genome after approximately 50 years and across large geographical distances (Figure 4).

**Distribution of indel mutations across the genomes**

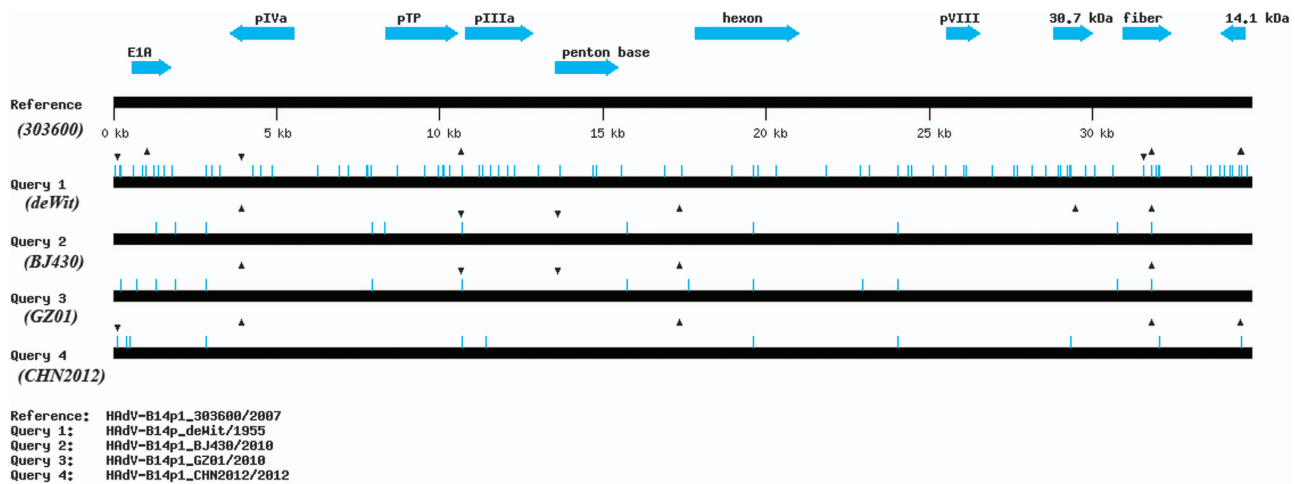
A global visualization of the alignment of genome sequences of GZ01, BJ430, 303600, CHN2012, and de Wit prototype is presented in Figure 5. This was generated using a ‘beta test version’ software, Genome Mutation Mapper (GMM; unpublished) that notes SNPs and indels of query genomes relative to a reference, and confirmed using DNA Sequencer (GeneCodes, Inc.; Ann Arbor, MI). The sequence similarities amongst them indicates that although these HAdV-B14p1 viruses have a common ancestor (type 14 serotype), their insertion/deletion (indel) patterns suggest two lineages. Indels and their subsequently inherited conserved patterns across genomes are key markers in determining lineages and for following molecular evolution, as reversion is highly unlikely.<sup>64</sup> Both strains GZ01 and BJ430 are highly divergent from the U.S. strain 303600, with respect to the indel



**Figure 4** Phylogenetic analysis of the E1A, fiber, and hexon genes, as well as the whole genome sequences, from all HAdV-B14 strains and representatives of other HAdV species. Nucleotide sequences of the E1A, fiber, and hexon genes, as well as limited whole genomes, are available from GenBank. Taxon names include GenBank acc. no., isolation country, strain name, and year of isolation. Phylogenetic trees were generated by using the neighbor-joining method with 1000 replicates and constructed by the MEGA 5.1.0 software (<http://www.ebi.ac.uk/tools/mafft>). In these analyses, default parameters were applied, with a maximum-composite-likelihood model. Bootstrap numbers shown at the nodes indicate the percentages of 1000 replications producing the clade, with values above 80 considered robust. The scale bar is in units of nucleotide substitutions per site. The China HAdV-B14 strains were indicated with filled triangles.

patterns; there is one indel distinguishing between GZ01 and BJ430, a deletion at nt 29 481 in the BJ430 genome, that could be interpreted as GZ01 being more similar to 303600. CHN2012 contains a divergent indel pattern with respect to the other two China isolates. In addition to three deletions in common, it contains two additional indels (one each, insertion and deletion), relative to GZ01 and BJ430, that are uniquely also in the prototype genome. This may be indicative of two sub-lineages and independent introductions into China from abroad, or subsequent divergence from one. The indel numbers and patterns

suggest there were at least two prototypes circulating in 1950s, time of the first prototype isolate, with one leading to 303600 and the second leading to the cluster of recent China strains. This visualization of the mutations alignment is critical to the understanding of the molecular evolution and lineages of pathogens as the low resolution genome percent nucleotide identities and REA (Figure 2) present very subtle and limited differences. GZ01, BJ430, CHN2012, and 303600 have highly similar percent identities in comparison with the prototype de Wit at 99.646%, 99.666%, 99.684%, and 99.689%, respectively.



**Figure 5** Global genome mutation map visualization of the alignment of genomes from HAdV-14 strains de Wit, 303600, BJ430, GZ01, and CHN2012. To assess the lineages and molecular evolution relationships of these pathogens, Genome Mutation Mapper (GMM; unpublished beta version) is used to provide a global visualization of the nucleotide differences, including insertions and deletions (indels). The contemporaneously circulating and independently isolated HAdV-B14p1 strains have divergent, yet similar mutation profiles, with the indels defining their lineages. For this analysis, the genome of HAdV-14p1\_303600/USA/2007 is set as the reference to which the other genomes are compared. Query 1: HAdV-B14p\_deWit/The Netherlands/1955 (prototype); Query 2: HAdV-B14p1\_BJ430/China//2010; Query 3: HAdV-14p1\_GZ01/China/2010; and Query 4: HAdV-B14p1\_CHN2012/China/2012. The triangles and inverted triangles indicate deletions and insertions, respectively; blue hash lines indicate base substitutions. Select gene markers are displayed for reference (locations are approximate), as are the genome size markers noted along the bottom.

#### Base substitution analysis of type 14 strains GZ01, BJ430, CHN2012, 303600, and de Wit

To complement and to check the global genome mutation map of the alignments, each point mutation amongst the five genomes are noted and compared, using the 2007 U.S. strain 303600 as reference (Table 2). Base substitutions are indicative of recent and/or essential mutations as reversion may occur, unlike indel mutations.<sup>64</sup> The whole genome sequence of strain GZ01 is very similar to that of strain BJ430 with respect to mutations. Six mutation differences are noted between these genomes. These SNPs include two non-synonymous substitutions in the DNA polymerase (K to N) and protein VI genes (R to K). Four synonymous substitutions are noted for the E1A 29.1 K and 25.7 K genes, the L3 VI gene, and the E2A DNA binding protein gene. One mutation is in a non-coding region (nt 225) and two deletions, three T (nt 10 658) and one A (nt 13 266), in poly (T) and poly (A) regions, are also noted (Table 2). Among these mutations, a deletion of A at nt 29 481 in the BJ430 genome distinguishes it from GZ01 (Figure 5).

When compared with strain 303600, 16 base substitutions were found in 11 coding regions of GZ01, which resulted in three non-synonymous substitutions in the E1A 6.5 K, E1B 54.9 K, and Protein VI coding regions, respectively. All of the indels occurred in non-coding regions, which led to length changes of the corresponding poly (A) or poly(T) sequences. Compared with the prototype de Wit strain, there were more mutations in strain GZ01, as would be expected given the time differences between their isolations: 93 base substitutions in the 31 genes, resulting in 35 non-synonymous substitutions. Although there were nine indels involving 25 nucleotides, only two of these resulted in non-synonymous substitutions: One in E1 29.1 K and 25.7 K coding regions (SV to I), with the other resulting the insertion of two amino acids KE in fiber knob domain, which is the most notable sequence difference between the HAdV-14p de Wit and all the other HAdV-14p1 strains (Table 2). This was believed to have the potential of altering cell receptor binding and tissue tropisms, and hence pathogenicity, as the fiber knob recognizes the host cell receptor.<sup>41</sup>

When compared with strain 303600, 11 base substitutions were found in strain CHN2012, which resulted in two non-synonymous substitutions in the E1B 54.9 K, E3 20.8 K coding regions, respectively. There are four synonymous substitutions in the E2B/L1 43 K, hexon, L4 100 K coding regions, respectively. The other mutations located at NCR, including one C to G mutation in ITR.

#### DISCUSSION

China's large, dense, and generally inaccessible population represents a unique environment for studying viral pathogens once they enter the population. The emergence, re-emergence, and transmission of a particular pathogen may be followed using high-resolution genome sequencing and, through computational methods, the molecular evolution and epidemiology of the pathogen may be revealed in great detail. According to the genome analysis of the three HAdV-B14 strains circulating in China, all three strains appear to be of separate, but related, lineages and may likely have been transmitted from abroad. Unlike the past, Beijing and Guangzhou are now 'open' to unrestricted global travel and this may situate them to be the foci for the introduction of infectious disease agents to and from both overseas and other China provinces. Uniquely, Guangzhou has four additional direct long-standing connections to the international community. First, since the mid-1800s, many China emigrants to overseas destinations have originated from this region; therefore, there has 'always' been direct physical contact through homecoming visits. Second, Guangzhou has hosted several international events recently, bringing in an influx of visitors. For example, the Asian Games (Nov. 2010) was hosted in Guangzhou, bringing in more than 14 000 athletes along with a large number of foreign tourists and regional transient workers. Third, residents of other China provinces migrate to this prosperous region. Finally, Guangzhou is only 119 km from Hong Kong with the populations commingling; both populations are also in close contact with the global community, including Americans from both hemispheres, neighboring Asians, and Europeans, all of which have experienced respiratory diseases linked to the subspecies B2 HAdVs as well as other respiratory pathogens. For instance, from 2002

Table 2 Comparison of Mutations Across the Genomes of four HAAdV-B14 Strains

| Position | Region | Product          | 14p deWit |            |           |          | Non-synonymous Substitution |            |           |          | Synonymous Substitution |            |           |          |
|----------|--------|------------------|-----------|------------|-----------|----------|-----------------------------|------------|-----------|----------|-------------------------|------------|-----------|----------|
|          |        |                  | 14p deWit | 14p1 BJ430 | 14p1 GZ01 | CHN 2012 | 14p deWit                   | 14p1 BJ430 | 14p1 GZ01 | CHN 2012 | 14p deWit               | 14p1 BJ430 | 14p1 GZ01 | CHN 2012 |
| 68       | ITR    | —                | C->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 134      | ITR    | —                | —         | —          | —         | C->G     | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 141      | NCR    | poly(T)          | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 200      | NCR    | —                | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 202      | NCR    | —                | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 209      | NCR    | —                | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 222      | NCR    | —                | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 225      | NCR    | —                | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 403      | NCR    | —                | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 491      | NCR    | —                | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 604      | E1A    | 29.1K 25.7K 6.5K | C->A      | —          | —         | G->A     | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 729      | E1A    | 29.1K 25.7K      | —         | —          | —         | G->A     | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 879      | E1A    | 29.1K 25.7K      | T->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 887      | E1A    | 29.1K 25.7K      | G->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 989      | E1A    | 29.1K 25.7K      | T->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1028     | E1A    | 29.1K 25.7K      | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1028     | E1A    | 29.1K 25.7K      | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1229     | E1A    | Intron           | G->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1299     | E1A    | 29.1K 25.7K      | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1299     | E1A    | 6.5K             | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1392     | E1A    | 29.1K 25.7K      | T->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1537     | NCR    | —                | T->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1538     | NCR    | —                | C->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1553     | NCR    | —                | C->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1801     | E1B    | 20K              | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 1903     | E1B    | 20K              | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 2852     | E1B    | 54.9K            | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 3015     | E1B    | 54.9K            | T->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 3015     | E1B    | 54.9K            | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 3265     | E1B    | 54.9K            | G->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 3927     | IX     | poly(A)          | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 4273     | IV2a   | Iva2             | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 4515     | IV2a   | Iva2             | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 4848     | IV2a   | Iva2             | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 6248     | E2B    | Polymerase       | A->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 6248     | E2B    | Polymerase       | C->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 6913     | E2B    | Polymerase       | C->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 7207     | E2B    | Polymerase       | G->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 7207     | E2B    | Polymerase       | C->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 7757     | E2B    | Polymerase       | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 7772     | E2B    | Polymerase       | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 7877     | E2B    | Polymerase       | C->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 7911     | E2B    | Polymerase       | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 8300     | E2B    | Polymerase       | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 8701     | E2B    | pTP              | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 9532     | E2B    | pTP              | G->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 9943     | E2B    | pTP              | G->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 10085    | E2B    | pTP              | G->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 10135    | E2B    | pTP              | G->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 10303    | E2B    | pTP              | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 10303    | E2B    | pTP              | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 10658    | NCR    | poly(T)          | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 10685    | E2B/L1 | 43K              | T->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 11187    | E2B/L1 | 43K              | C->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 11304    | E2B/L1 | 43K              | T->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 11420    | E2B/L1 | 43K              | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 11532    | E2B/L1 | 43K              | C->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 11805    | E2B/L1 | 43K              | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 12088    | E2B/L1 | pIII             | G->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 12268    | E2B/L1 | pIII             | G->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 13013    | E2B/L1 | pIII             | T->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 13266    | E2B/L1 | poly(A)          | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 13691    | NCR    | —                | C->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 14687    | L2     | PENTON           | A->G      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 14788    | L2     | Penton           | C->T      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 15568    | L2     | VII              | T->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 15739    | L2     | VII              | T->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 16867    | L2     | pV               | T->C      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 17348    | NCR    | poly(A)          | G->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 17408    | L3     | VI               | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 17594    | L3     | VI               | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 17618    | L3     | VI               | —         | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |
| 18944    | L3     | Hexon            | G->A      | —          | —         | —        | —                           | —          | —         | —        | —                       | —          | —         | —        |



Table 2 (Continued)

| Position | Region | Product | Non-synonymous Substitution |            |           |          |           | Synonymous Substitution |           |          |  |  |
|----------|--------|---------|-----------------------------|------------|-----------|----------|-----------|-------------------------|-----------|----------|--|--|
|          |        |         | 14p deWit                   | 14p1 BJ430 | 14p1 GZ01 | CHN 2012 | 14p deWit | 14p1 BJ430              | 14p1 GZ01 | CHN 2012 |  |  |
| 19593    | L3     | Hexon   | A → G                       | A → G      | A → G     | A → G    | K → K     | K → K                   | K → K     | K → K    |  |  |
| 19748    | L3     | Hexon   | A → G                       | —          | —         | —        | L → L     | —                       | —         | —        |  |  |
| 20309    | L3     | HEXON   | A → G                       | —          | —         | —        | G → G     | —                       | —         | —        |  |  |
| 21810    | NCR    | —       | A → G                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 22886    | E2A    | DBP     | A → G                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 22923    | E2A    | DBP     | A → G                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 23153    | E2A    | DBP     | A → G                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 24008    | L4     | 100K    | G → A                       | —          | —         | —        | P → P     | —                       | —         | —        |  |  |
| 24325    | L4     | 100K    | T → C                       | T → C      | T → C     | T → C    | I → I     | I → I                   | I → I     | I → I    |  |  |
| 24453    | L4     | 100K    | C → T                       | —          | —         | —        | V → V     | —                       | —         | —        |  |  |
| 25117    | L4     | 100K    | G → A                       | —          | —         | —        | L → L     | —                       | —         | —        |  |  |
| 25481    | L4     | 100K    | A → G                       | —          | —         | —        | L → L     | —                       | —         | —        |  |  |
| 26048    | L4     | 100K    | T → C                       | —          | —         | —        | L → L     | —                       | —         | —        |  |  |
| 26112    | L4     | 22K 33K | G → A                       | —          | —         | —        | P → P     | —                       | —         | —        |  |  |
| 26923    | L4     | 22K 33K | G → A                       | —          | —         | —        | P → P     | —                       | —         | —        |  |  |
| 27569    | E3     | VIII    | A → G                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 27698    | E3     | 14.6K   | G → A                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 28128    | E3     | 14.6K   | A → G                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 2814K    | E3     | 18.4K   | C → T                       | —          | —         | —        | F → F     | —                       | —         | —        |  |  |
| 28139    | E3     | 18.4K   | A → C                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 28568    | E3     | 20.1K   | G → A                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 28952    | E3     | 20.8K   | G → A                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 29022    | E3     | 20.8K   | G → A                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 29223    | E3     | 20.8K   | A → G                       | —          | —         | —        | Q → Q     | —                       | —         | —        |  |  |
| 29277    | E3     | 20.8K   | A → T                       | —          | —         | —        | A → A     | —                       | —         | —        |  |  |
| 29312    | E3     | 20.8K   | A → T                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 29335    | E3     | 20.8K   | C → T                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 29481    | NCR    | poly(T) | —                           | ▲T         | —         | T → C    | —         | —                       | —         | —        |  |  |
| 29758    | E3     | 10.1K   | T → C                       | —          | —         | —        | I → I     | —                       | —         | —        |  |  |
| 30054    | E3     | 15.2K   | C → T                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 30057    | E3     | 15.2K   | C → T                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 30621    | U      | U       | G → A                       | —          | —         | —        | F → F     | —                       | —         | —        |  |  |
| 30733    | U      | U       | —                           | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 31546    | L5     | FIBER   | ▲AGAAAA                     | —          | —         | —        | T → T     | —                       | —         | —        |  |  |
| 31561    | L5     | FIBER   | C → T                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 31786    | NCR    | —       | —                           | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 31791    | NCR    | Poly(A) | G → C                       | —          | —         | —        | I → I     | —                       | —         | —        |  |  |
| 31797    | NCR    | —       | ▲A                          | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 31940    | E4     | ORF6/7  | T → C                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 31979    | E4     | ORF6/7  | T → G                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 32017    | E4     | ORF6/7  | G → A                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 32028    | E4     | ORF6/7  | C → A                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 32028    | E4     | ORF6/7  | C → A                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 33005    | E4     | ORF4    | C → T                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 33492    | E4     | ORF3    | G → T                       | —          | —         | —        | R → R     | —                       | —         | —        |  |  |
| 33614    | E4     | ORF2    | C → T                       | —          | —         | —        | L → L     | —                       | —         | —        |  |  |
| 33874    | E4     | ORF2    | C → G                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 34016    | E4     | ORF1    | T → C                       | —          | —         | —        | E → Q     | —                       | —         | —        |  |  |
| 34205    | E4     | ORF1    | G → C                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 34254    | E4     | ORF1    | G → T                       | —          | —         | —        | F → L     | —                       | —         | —        |  |  |
| 34471    | E4     | ORF1    | G → A                       | —          | —         | —        | S → Y     | —                       | —         | —        |  |  |
| 34471    | NCR    | —       | —                           | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 34498    | NCR    | poly(A) | ▲A                          | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 34534    | NCR    | —       | —                           | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 34549    | NCR    | —       | A → C                       | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 34553    | NCR    | —       | ▲AA                         | —          | —         | —        | —         | —                       | —         | —        |  |  |
| 34700    | ITR    | —       | G → A                       | —          | —         | —        | —         | —                       | —         | —        |  |  |

HA8V-B14 prototype (de Wit; 1955), BJ430 (2010), and CHN2012 genomes are analyzed with respect to base substitutions (SNPs) and insertions and deletions (indels), using the 2007 U.S. strain 303600 as reference. Nucleotide and amino acid sequence changes are noted, along with their genome locations and coding consequences. Abbreviations and symbols are as follows: inverted terminal repeat, ITR; non-coding region, NCR; insertion, ▼; deletion, ▲; no change or not applicable, -.

to 2003, the severe acute respiratory syndrome coronavirus (SARS-CoV) spread quickly from Guangzhou to Hong Kong and Beijing, and then across the country and globally, with high morbidity and mortality rates. Guangzhou, therefore, serves as an important focal point for infectious disease pathogen introduction, and surveillance is important in order to limit the public health impact on its population and elsewhere in China.

HAdV are important pathogens that appear to re-emerge after lengthy absences, either due to non-reporting because of diminished pathogenicity and infectivity or to perhaps latency or other mechanisms of cryptic infection. HAdV-B14, for example, re-emerged after approximately fifty years and HAdV-B7d re-emerged after twenty-one years.<sup>56</sup> The 2005 re-emergent HAdV-B14p1 strain was associated with several highly contagious and geographically wide-spread outbreaks of ARD, and included unexpected higher rates of fatalities.<sup>11,47,65</sup> These outbreaks occurred in both civilian (24 communities) and military (nine communities) populations in the United States (2005–2009) and Europe (2009–2011).<sup>4,5,11,47,50,65</sup> Parenthetically and retrospectively, the first case identified was in a child in California (2003)<sup>4</sup> and was presumably the originating point for the subsequent larger military base outbreaks.<sup>65</sup>

In a limited retrospective survey of respiratory infectious disease agents from patients at a hospital in Guangzhou (2010 and 2011), HAdVs were found and characterized with respect to types.<sup>54</sup> One of these isolates, typed, deposited, and noted in GenBank as an emergent and previously unreported HAdV type 14 in China ('human adenovirus 14 isolate HAdV-B/CHN/GZ01/2010/14[P14H14F14]'), was isolated from a throat swab of a 17-month-old child with acute suppurative tonsillitis (October 2010). Another simultaneously emergent HAdV-B14 strain BJ430 was isolated from a six-month-old infant diagnosed with bronchial pneumonia and hospitalized at the Beijing Children's Hospital (December 2010). This is unusual because this condition had not been previously described for infections attributed to either HAdV-B14 or HAdV-B55, with its similar genome and disease symptoms.<sup>11,14,48,66</sup> As HAdV-B14 has re-emerged globally recently, and since it is known as a highly contagious pathogen that has been associated with high hospitalization rates and fatality rates,<sup>11,47,65</sup> the emergence of this virus should set off an alarm that the proper surveillance of this virus is critical for large dense populations with naive immunity to HAdV-B14.

Comparative genomic analysis is leading to discovery of large numbers of novel molecular markers that are proving very helpful in understanding many important aspects of microbial phylogeny;<sup>67,68</sup> of these molecular markers, the conserved indel mutations provide particularly useful means for identifying different groups of microbes in clear molecular terms and for understanding when they have branched off from a common ancestor.<sup>67,69–71</sup> As reported here, indel mutations analysis suggests two co-circulating lineages from at least two co-circulating 'prototype' strains in 1950s, one giving rise to the 303600 strain and the other giving rise to the China cluster of HAdV-B14. Before the genomics era, many important adenovirus strains were characterized by REA. REA is still useful for characterizing current isolates by providing a bridge between their accessible genome data and the REA data that are only available in the literature, as important references and as historical isolates are no longer available for genomic analysis.<sup>72</sup> This report emphasizes that REA data should be carefully considered, as it may be misleading and incorrect. Although the REA analysis confirmed strains GZ01, BJ430, and CHN2012 as belonging to the same genome type 14p1 as strain 303600, high resolution analysis of the genome sequences, including indels, indicate different lineages and introductions into China, along with the absence of fatalities

associated with 303600. Again, based on the genome mutation patterns, especially the indels, the de Wit prototype appears to be very different from the 303600 and is similar to the BJ430, GZ01, and CHN2012 strains. These may represent several lineages co-circulating 'prototype 14' strains, with one reported in the original study and the rest accumulating characteristic indel patterns that have been reported in this study. Most of the indels occur in the poly(A) and poly(T) regions, which have been proposed as high-resolution molecular strain markers for characterizing HAdV-14p1.<sup>71</sup> The recent report that HAdV-B14 dispersed from U.S. to Asia in 2007 via military trainees<sup>53</sup> provides another clue that the emergent HAdV-B14p1 strains in China may have been transmitted from abroad. However, as shown in this report, the whole genome data of that strain is critical for the correct interpretation of that possible transmission pattern.

The whole-genome sequences of strains GZ01 and BJ430 are nearly identical with each other, with respect to point mutations. One indel separates them. Strain GZ01 caused acute suppurative tonsillitis, an upper respiratory disease; however, strain BJ430 was associated with bronchial pneumonia, a lower respiratory tract disease. The reason for this difference in pathogenicity may be due to the genomic differences. A major difference between the two genomes is the two non-synonymous substitutions in the DNA polymerase (K to N) and protein VI genes (R to K), respectively (Table 2). Of these two proteins, the 22-kDa cement protein VI is located beneath the peripentonal hexons in the viral capsid. It is identified as an endosomal membrane lytic factor, which is important for adenoviruses to overcome the barrier of the host cell membrane.<sup>73</sup> The non-synonymous substitution of R to K in the protein VI may lead to the function of protein VI being enhanced or weakened, which can presumably further change the HAdV tissue tropism. This is awaiting wet-bench studies.

Given the current ease of travel and global interactions, these ARD outbreaks associated with HAdV-B14 provide insights into the distribution, lineage, and molecular evolution of this pathogen. There had been no reports of HAdV-B14 in China prior to and during the period noted for HAdV-B14 re-emergence and circulation (2005–2009). This report presents the bioinformatics analysis of both HAdV-B14p1 strains in high resolution and detail, and provides putative lineages for these two surprisingly and intriguingly *divergent, but, highly similar* genomes linking Guangzhou and Beijing (2294 kilometers apart). A comparison with the re-emergent strain isolated in the U.S. (2007) as well as with the less similar but still remarkably conserved genome of the prototype virus from The Netherlands (1955) is also presented. The presentation of high resolution nucleotide sequence data and a map of mutations, particularly of the indels, provide detailed insight into two contemporaneously circulating human pathogens with divergent and parallel lineages. Both are noted as genome type HAdV-B14p1 via their low resolution but conveniently available REA patterns. The computational data presented in this report again demonstrates highly conserved but divergent HAdV-B14 genomes, which supports further studies of the epidemiology and molecular evolution of these related pathogens, and which, in turn, will provide a foundation for developing effective vaccines and public health strategies, including nationwide surveillance.

## ACKNOWLEDGEMENTS

This work was funded by the National Natural Science Foundation of China (31370199 and 31570155) and the Excellent Young Teacher Training Plan of Guangdong Province (Yq2013039), as well as by the Young Top-notch Talents of Guangdong Province Special Support Program (2014), and the Guangzhou Healthcare Collaborative Innovation Major Project (201400000002). Portions

of this manuscript were completed at the Department of Ophthalmology, Howe Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School (Boston, Massachusetts, USA) as Q.Z. was funded by the China Scholarship Council (CSC No. 201508440056) as a Visiting Scholar (2015–2016); he thanks James Chodosh and Jaya Rajaiya for providing a stimulating intellectual environment. This project was additionally supported by a summer research grant (2016) to D.S. from the Office of the Vice President for Research at George Mason University. The Funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

- 1 Pasarica M, Loiler S, Dhurandhar NV. Acute effect of infection by adipogenic human adenovirus Ad36. *Arch Virol* 2008; **153**: 2097–2102.
- 2 Van Der Veen J, Kok G. Isolation and typing of adenoviruses recovered from military recruits with acute respiratory disease in The Netherlands. *Am J Hyg* 1957; **65**: 119–129.
- 3 Kendall EJ, Riddle RW, Tuck HA *et al*. Pharyngo-conjunctival fever; school outbreaks in England during the summer of 1955 associated with adenovirus types 3, 7, and 14. *Br Med J* 1957; **2**: 131–136.
- 4 Kajon AE, Lu X, Erdman DD *et al*. Molecular epidemiology and brief history of emerging adenovirus 14-associated respiratory disease in the United States. *J Infect Dis* 2010; **202**: 93–103.
- 5 O'Flanagan D, O'Donnell J, Domegan L *et al*. First reported cases of human adenovirus serotype 14p1 infection, Ireland, October 2009 to July 2010. *Euro Surveill* 2011; **16**: pii = 19801.
- 6 Hierholzer JC, Pumarola A. Antigenic characterization of intermediate adenovirus 14-11 strains associated with upper respiratory illness in a military camp. *Infect Immun* 1976; **13**: 354–359.
- 7 Chmielewicz B, Benzler J, Pauli G *et al*. Respiratory disease caused by a species B2 adenovirus in a military camp in Turkey. *J Med Virol* 2005; **77**: 232–237.
- 8 Zhu Z, Zhang Y, Xu S *et al*. Outbreak of acute respiratory disease in China caused by B2 species of adenovirus type 11. *J Clin Microbiol* 2009; **47**: 697–703.
- 9 Yang Z, Zhu Z, Tang L *et al*. Genomic analyses of recombinant adenovirus type 11a in China. *J Clin Microbiol* 2009; **47**: 3082–3090.
- 10 Kajon AE, Dickson LM, Metzgar D *et al*. Outbreak of febrile respiratory illness associated with adenovirus 11a infection in a Singapore military training camp. *J Clin Microbiol* 2010; **48**: 1438–1441.
- 11 Carr MJ, Kajon AE, Lu X *et al*. Deaths associated with human adenovirus-14p1 infections, Europe, 2009-2010. *Emerg Infect Dis* 2011; **17**: 1402–1408.
- 12 Zhang Q, Seto D, Zhao S *et al*. Genome sequence of the first human adenovirus type 14 isolated in China. *J Virol* 2012; **86**: 7019–7020.
- 13 Kajon AE, Mistchenko AS, Videla C *et al*. Molecular epidemiology of adenovirus acute lower respiratory infections of children in the South Cone of South America (1991–1994). *J Med Virol* 1996; **48**: 151–156.
- 14 Walsh MP, Seto J, Jones MS *et al*. Computational analysis identifies human adenovirus type 55 as a re-emergent acute respiratory disease pathogen. *J Clin Microbiol* 2010; **48**: 991–993.
- 15 Takeuchi S, Itoh N, Uchio E *et al*. Adenovirus strains of subgenus D associated with nosocomial infection as new etiological agents of epidemic keratoconjunctivitis in Japan. *J Clin Microbiol* 1999; **37**: 3392–3394.
- 16 Louie JK, Kajon AE, Holodny M *et al*. Severe pneumonia due to adenovirus serotype 14: a new respiratory threat? *Clin Infect Dis* 2008; **46**: 421–425.
- 17 Echavarría M. Adenoviruses in immunocompromised hosts. *Clin Microbiol Rev* 2008; **21**: 704–715.
- 18 Stalder H, Hierholzer JC, Oxman MN. New human adenovirus (candidate adenovirus type 35) causing fatal disseminated infection in a renal transplant recipient. *J Clin Microbiol* 1977; **6**: 257–265.
- 19 Kajon AE, Wadell G. Molecular epidemiology of adenoviruses associated with acute lower respiratory disease of children in Buenos Aires, Argentina (1984-1988). *J Med Virol* 1992; **36**: 292–297.
- 20 Kaneko H, Aoki K, Ohno S *et al*. Complete genome analysis of a novel intertypic recombinant human adenovirus causing epidemic keratoconjunctivitis in Japan. *J Clin Microbiol* 2011; **49**: 484–490.
- 21 Kaneko H, Aoki K, Ishida S *et al*. Recombination analysis of intermediate human adenovirus type 53 in Japan by complete genome sequence. *J Gen Virol* 2011; **92** (Pt 6): 1251–1259.
- 22 Walsh MP, Chintakuntlawar A, Robinson CM *et al*. Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLoS ONE* 2009; **4**: e5635.
- 23 Robinson CM, Rajaiya J, Walsh MP *et al*. Computational analysis of human adenovirus type 22 provides evidence for recombination between human adenoviruses species D in the penton base gene. *J Virol* 2009; **83**: 8980–8985.
- 24 Seto D, Chodosh J, Brister JR *et al*. Using the Whole-Genome Sequence To Characterize and Name Human Adenoviruses. *J Virol* 2011; **85**: 5701–5702.
- 25 Liu EB, Ferreyra L, Fischer SL *et al*. Genetic analysis of a novel human adenovirus with a serologically unique hexon and a recombinant fiber gene. *PLoS ONE* 2011; **6**: e24491.
- 26 Liu EB, Wadford DA, Seto J *et al*. Computational and Serologic Analysis of Novel and Known Viruses in Species Human Adenovirus D in Which Serology and Genomics Do Not Correlate. *PLoS ONE* 2012; **7**: e33212.
- 27 Robinson CM, Zhou X, Rajaiya J *et al*. Predicting the next eye pathogen: analysis of a novel adenovirus. *mBio* 2013; **4**: e00595-12.
- 28 Matsushima Y, Shimizu H, Kano A *et al*. Novel human adenovirus strain, Bangladesh. *Emerg Infect Dis* 2012; **18**: 846–848.
- 29 Zhou X, Robinson CM, Rajaiya J *et al*. Analysis of human adenovirus type 19 associated with epidemic keratoconjunctivitis and its reclassification as adenovirus type 64. *Invest Ophthalmol Vis Sci* 2012; **53**: 2804–2811.
- 30 Jones MS II, Harrach B, Ganac RD *et al*. New adenovirus species found in a patient presenting with gastroenteritis. *J Virol* 2007; **81**: 5978–5984.
- 31 Hage E, Liebert UG, Bergs S *et al*. Human Adenovirus type 70: A novel, multiple recombinant species D adenovirus isolated from diarrheal faeces of a haematopoietic stem cell transplantation recipient. *J Gen Virol* 2015; **96**: 2734–2742.
- 32 Tang L, An J, Xie Z *et al*. Genome and Bioinformatic Analysis of a HAdV-B14p1 Virus Isolated from a Baby with Pneumonia in Beijing, China. *PLoS ONE* 2013; **8**: e60345.
- 33 Mi Z, Butt AM, An X *et al*. Genomic analysis of HAdV-B14 isolate from the outbreak of febrile respiratory infection in China. *Genomics* 2013; **102**: 448–455.
- 34 Houng HS, Gong H, Kajon A *et al*. Genome sequences of human adenovirus 14 isolates from mild respiratory cases and a fatal pneumonia, isolated during 2006-2007 epidemics in North America. *Respir Res* 2010; **11**: 116.
- 35 Lauer KP, Llorente I, Blair E *et al*. Natural variation among human adenoviruses: genome sequence and annotation of human adenovirus serotype 1. *J Gen Virol* 2004; **85** (Pt 9): 2615–2625.
- 36 Dehghan S, Seto J, Liu EB *et al*. Computational analysis of four human adenovirus type 4 genomes reveals molecular evolution through two interspecies recombination events. *Virology* 2013; **443**: 197–207.
- 37 Dehghan S, Seto J, Jones MS *et al*. Simian adenovirus type 35 has a recombinant genome comprising human and simian adenovirus sequences, which predicts its potential emergence as a human respiratory pathogen. *Virology* 2013; **447**: 265–273.
- 38 Seto J, Walsh MP, Mahadevan P *et al*. Genomic and bioinformatics analyses of HAdV-14p, reference strain of a re-emerging respiratory pathogen and analysis of B1/B2. *Virus Res* 2009; **143**: 94–105.
- 39 Singh G, Robinson CM, Dehghan S *et al*. Overreliance on the Hexon Gene, Leading to Misclassification of Human Adenoviruses. *J Virol* 2012; **86**: 4693–4695.
- 40 Purkayastha A, Ditty SE, Su J *et al*. Genomic and bioinformatics analysis of HAdV-4, a human adenovirus causing acute respiratory disease: implications for gene therapy and vaccine vector development. *J Virol* 2005; **79**: 2559–2572.
- 41 Wang H, Tuve S, Erdman DD *et al*. Receptor usage of a newly emergent adenovirus type 14. *Virology* 2009; **387**: 436–441.
- 42 Singh G, Robinson CM, Dehghan S *et al*. Homologous Recombination in E3 Genes of Human Adenovirus Species D. *J Virol* 2013; **87**: 12481–12488.
- 43 Lion T. Adenovirus infections in immunocompetent and immunocompromised patients. *Clin Microbiol Rev* 2014; **27**: 441–462.
- 44 Wadell G, Hammarskjöld ML, Winberg G *et al*. Genetic variability of adenoviruses. *Ann N Y Acad Sci* 1980; **354**: 16–42.
- 45 Wadell G. Molecular epidemiology of human adenoviruses. *Curr Top Microbiol Immunol* 1984; **110**: 191–220.
- 46 De Jong JC, Wermenbol AG, Verweij-Uijterwaal MW *et al*. Adenoviruses from human immunodeficiency virus-infected individuals, including two strains that represent new candidate serotypes Ad50 and Ad51 of species B1 and D, respectively. *J Clin Microbiol* 1999; **37**: 3940–3945.
- 47 Lewis PF, Schmidt MA, Lu X *et al*. A community-based outbreak of severe respiratory illness caused by human adenovirus serotype 14. *J Infect Dis* 2009; **199**: 1427–1434.
- 48 Prevention CfDca. Acute respiratory disease associated with adenovirus serotype 14-four states, 2006-2007. *MMWR Morb Mortal Wkly Rep* 2007; **56**: 1181–1184.
- 49 Girouard G, Garceau R, Thibault L *et al*. Adenovirus serotype 14 infection, New Brunswick, Canada, 2011. *Emerg Infect Dis* 2013; **19**: 119–122.
- 50 Parcell BJ, McIntyre PG, Yirrell DL *et al*. Prison and community outbreak of severe respiratory infection due to adenovirus type 14p1 in Tayside, UK. *Journal of Public Health* 2014; **37**: 64–69.
- 51 Huang G, Yu D, Zhu Z *et al*. Outbreak of febrile respiratory illness associated with human adenovirus type 14p1 in Gansu Province, China. *Influenza Other Respi Viruses* 2013; **7**: 1048–1054.
- 52 Yu W, Sun H, Tian J *et al*. Molecular epidemic characteristics analysis of an adenovirus infection outbreak. *Chin J Public Health* 2014; **7**: 972–974.
- 53 Trei JS, Johns NM, Garner JL *et al*. Spread of adenovirus to geographically dispersed military installations, May-October 2007. *Emerg Infect Dis* 2010; **16**: 769–775.
- 54 Han G, Niu H, Zhao S *et al*. Identification and typing of respiratory adenoviruses in Guangzhou, Southern China using a rapid and simple method. *Virol Sin* 2013; **28**: 103–108.
- 55 Brudno M, Do CB, Cooper GM *et al*. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 2003; **13**: 721–731.
- 56 Zhao S, Wan C, Ke C *et al*. Re-emergent Human Adenovirus Genome Type 7d Caused an Acute Respiratory Disease Outbreak in Southern China After a Twenty-one Year Absence. *Scientific reports* 2014; **4**: 7365.
- 57 Zhang Q, Su X, Seto D *et al*. Construction and characterization of a replication-competent human adenovirus type 3-based vector as a live-vaccine candidate and a viral delivery vector. *Vaccine* 2009; **27**: 1145–1153.
- 58 Yu Z, Zeng Z, Zhang J *et al*. Fatal community-acquired pneumonia in children caused by re-emergent human adenovirus 7d associated with higher severity of illness and fatality rate. *Sci Rep* 2016; **6**: 37216.

- 59 Li QG, Zheng QJ, Liu YH *et al*. Molecular epidemiology of adenovirus types 3 and 7 isolated from children with pneumonia in Beijing. *J Med Virol* 1996; **49**: 170–177.
- 60 Seto J, Walsh M, Mahadevan P *et al*. Applying Genomic and Bioinformatic Resources to Human Adenovirus Genomes for Use in Vaccine Development and for Applications in Vector Development for Gene Delivery. *Viruses* 2010; **2**: 1–26.
- 61 Li QG, Wadell G. Analysis of 15 different genome types of adenovirus type 7 isolated on five continents. *J Virol* 1986; **60**: 331–335.
- 62 Mul YM, Verrijzer CP, van der Vliet PC. Transcription factors NFI and NFIII/oct-1 function independently, employing different mechanisms to enhance adenovirus DNA replication. *J Virol* 1990; **64**: 5510–5518.
- 63 Hatfield L, Hearing P. The NFIII/OCT-1 binding site stimulates adenovirus DNA replication in vivo and is functionally redundant with adjacent sequences. *J Virol* 1993; **67**: 3931–3939.
- 64 Nomura M, Benzer S. The nature of the 'deletion' mutants in the rII region of phage T4. *J Mol Biol* 1961; **3**: 684–692.
- 65 Metzgar D, Osuna M, Kajon AE *et al*. Abrupt emergence of diverse species B adenoviruses at US military recruit training centers. *J Infect Dis* 2007; **196**: 1465–1473.
- 66 Zhang Q, Seto D, Cao B *et al*. Genome Sequence of Human Adenovirus Type 55, a Re-emergent Acute Respiratory Disease Pathogen in China. *J Virol* 2012; **86**: 12441–12442.
- 67 Gupta RS. Applications of conserved indels for understanding microbial phylogeny. In: Oren A, Papke RT (eds). *Molecular Phylogeny of Microorganisms*. Norwich: Caister Academic Press, 2010: 135–150.
- 68 Naushad HS, Lee B, Gupta RS. Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: identification of molecular signatures that are specific for the phytopathogenic genera *Dickeya*, *Pectobacterium* and *Brenneria*. *Int J Syst Evol Microbiol* 2014; **64** (Pt 2): 366–383.
- 69 Ajawatanawong P, Baldauf SL. Evolution of protein indels in plants, animals and fungi. *BMC Evol Biol* 2013; **13**: 140.
- 70 Gao B, Gupta RS. Microbial systematics in the post-genomics era. *Antonie Van Leeuwenhoek* 2012; **101**: 45–54.
- 71 Hong HS, Lott L, Gong H *et al*. Adenovirus microsatellite reveals dynamics of transmission during a recent epidemic of human adenovirus serotype 14 infection. *J Clin Microbiol* 2009; **47**: 2243–2248.
- 72 Zhang Q, Dehghan S, Seto D. Pitfalls of restriction enzyme analysis in identifying, characterizing, typing, and naming viral pathogens in the era of whole genome data, as illustrated by HAdV type 55. *Virology* 2016; **51**: 448–453.
- 73 Wiethoff CM, Wodrich H, Gerace L *et al*. Adenovirus protein VI mediates membrane disruption following capsid disassembly. *J Virol* 2005; **79**: 1992–2000.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017