RESEARCH ARTICLE

JOURNAL OF
MEDICAL VIROLOGY WILEY

Whole genome sequencing and genome characterization of Aichivirus isolated from Korean adults

Seoyoung Woo[1] | Md Iqbal Hossain[1] | Soontag Jung[1] | Daseul Yeo[1] |
Danbi Yoon[1] | Seongwon Hwang[1] | Huijeong Doh[2] | Seong-il Eyun[2] |
Changsun Choi[1]

[1]Department of Food and Nutrition, College of Biotechnology and Natural Resources, Chung-Ang University, Anseong, Republic of Korea

[2]Department of Life Science, College of Natural Sciences, Chung-Ang University, Seoul, Republic of Korea

Correspondence
Changsun Choi, Department of Food and Nutrition, School of Food Science and Technology, College of Biotechnology and Natural Resources, Chung-Ang University, Anseong-Si, Gyeonggi-do 17546, Republic of Korea.
Email: cchoi@cau.ac.kr

Abstract

The whole-genome sequence (WGS) analysis of Aichivirus (AiV) identified in Korea was performed in this study. Using Sanger and Nanopore sequencing, the 8228-nucleotide-long genomic sequence of AiV (OQ121963) was determined and confirmed to belong to genotype A. The full-length genome of OQ121963 consisted of a 7296 nt open reading frame (ORF) that encodes a single polyprotein, and 5′ UTR (676 nt) and 3′ UTR (256 nt) at 5′ and 3′ ends, respectively. The ORF consisted of leader protein (L), structural protein P1 (VP0, VP1, and VP3), and nonstructural protein P2 (2A, 2B, and 2C) and P3 (3A, 3B, 3C, and 3D). The secondary structure analysis of the 5′ UTR identified only stem-loop C (SL−C) and not SL−A and SL−B. The variable region of the AiV genome was analyzed by MegAlign Pro and re-confirmed by SimPlot analysis using 16 AiV whole genomes known to date. Among the entire regions, structural protein region P1 showed the lowest amino acid identity (96.07%) with reference sequence AB040749 (originated in Japan; genotype A), while the highest amino acid identity (98.26%) was confirmed in the 3D region among nonstructural protein region P2 and P3. Moreover, phylogenetic analysis of the WGS of OQ121963 showed the highest homology (96.96%) with JX564249 (originated in Taiwan; genotype A) and lowest homology (90.14%) with DQ028632 (originated in Brazil; genotype B). Therefore, the complete genome characterization of OQ121963 and phylogenetic analysis of the AiV conducted in this study provide useful information allowing to improve diagnostic tools and epidemiological studies of AiVs.

KEYWORDS
aichivirus, nanopore sequencing, phylogenetic analysis, sanger sequencing, whole genome sequencing

1 | INTRODUCTION

Aichivirus (AiV) was first isolated from the feces of a patient with oyster–associated gastroenteritis in 1989 in Aichi, Japan.[1] The genome of AiV, classified in the *Kobuvirus* genus of the *Picornaviridae* family, comprises an 8.3 kb positive–sense single-stranded RNA.[2,3] The AiVs cause gastroenteritis signs, such as diarrhea, abdominal pain, nausea, vomiting, and fever in humans and spread by fecal-oral transmission in contaminated water or food; however, they may be rarely detected in association with lower respiratory tract disease.[4,5] Infants require more attention owing to higher incidence of AiVs cause gastroenteritis.[6] The seroprevalence of AiV is approximately 60% in children under 10 years of age, reaching 90% later in life.[7–11] The AiVs are distributed worldwide including Europe, Asia, Africa, South America. Some clinical studies observed coinfection with other enteric viruses in patients with diarrhea.[12,13] Moreover, the AiV is detected more frequently in immunosuppressed children with human immunodeficiency virus.[14] Nevertheless, no effective antiviral treatment or vaccine is available. The AiV genome has high GC content and the highest degree of RNA secondary structure found in picornaviruses.[3] The RNA structure complicates the sequencing of the AiV, only sixteen AiV strains whole genome sequences have been reported to date. Therefore, the AiV has been studied less than the other picornaviruses.[15] In the present study, we performed the whole genome sequencing (WGS) of the AiV using two different methods including Sanger sequencing and Next-generation sequencing (NGS) with MinION Nanopore sequencing. WGS using NGS has become an essential tool in understanding molecular epidemiology and pathogen evolution, developing diagnostics and vaccines, and studying the genetic association between viruses and their hosts, especially in the field of viral infectious diseases.[16] Clinical WGS applications are mostly based on Sanger sequencing and on Second-generation sequencing platforms, with Illumina MiSeq/HiSeq.[17] Third-generation sequencers, such as Nanopore sequencing by Oxford Nanopore Technologies (ONT), have recently emerged as a complement or replacement for NGS by enabling the sequencing of long–read single DNA molecules for WGS.[18] The MinION sequencer is attracting considerable interest in the field of infectious diseases, particularly for the pathogen surveillance and clinical diagnostic applications using RNA/DNA-direct sequencing. This technique has been used for the WGS of enteroviruses from human clinical samples using direct RNA sequencing.[19] Furthermore, this technique has been used for the identification of Chikungunya, Ebola, and hepatitis C virus from the human clinical samples without target enrichment, and detection of bacterial pathogens from urine samples and respiratory samples.[20] Therefore, we conducted WGS of AiV in stool sample using MinION Nanopore sequencer. However, ONT reads have higher error rates than Sanger sequencing reads, with accuracy ranging from 65%–88%.[21] Since Sanger sequencing has certainly been a reliable and powerful method that has improved molecular biology with 99.99% accuracy, we confirmed the nanopore data using Sanger sequencing.

To date, WGS of AiV strains has not been extensively done. Therefore, we performed WGS of Korean Aichivirus strain CAU170613 (OQ121963) using Sanger and Nanopore sequencing. In addition, we performed a genetic analysis for understanding molecular epidemiology and pathogen evolution.

2 | MATERIALS AND METHODS

2.1 | Stool samples preparation and virus cultivation

This research was permitted by the Korea Association of Health Promotion Bioethics Committee (Approval no.: 130750-201902-BR-019). In this study, 20 human stool samples were collected from Korea Association of Health Promotion (KAHP) public health care institute, Seoul, Republic of Korea for routine health checkup. Written informed permission from patients was also obtained. Patients whose samples were collected had no gastroenteritis symptoms. Adenovirus, aichivirus, astrovirus, three types of enteroviruses, nine types of influenza A viruses, hepatitis A virus, hepatitis E virus, parvovirus, rotavirus, sapovirus, norovirus GI, norovirus GII were examined. In this study, we used different types of viruses to investigate the status of these viruses in human stool samples, with no further virus selection or screening methods.

Viral RNA of each sample was prepared using the NucleoSpin RNA stool kit (Machery–Nagel, Haerdt, France) according to the manufacturer's instructions. Real-time quantitative PCR (RT–qPCR) was performed according to the AiV universal and genotype-specific assays (Table S1).[22] Overall workflow for WGS of AiV is shown in Figure S1.

The AiV positive stool sample (CAU170613) was prepared as a 10% suspension in Dulbecco's Modified Essential Medium (DMEM), 100 U/mL of Antibiotic–Antimycotic (Gibco), and centrifuged at 4000 × g for 30 min. The suspension was filtered through 0.45 and 0.20 μm pore syringe filters (BioFACT). The AiV CAU170613 inoculum was cultivated on a monolayer of Vero cells according to the previous protocol.[7]

2.2 | Primer design and sequencing sample preparation

The complete genome sequence of Japanese AiV strain (A846/88) was obtained from the GenBank database in NCBI (www.ncbi.nlm.nih.gov/genbank). The 27 primer sets were designed using the 'Primal Scheme' tool (http://primal.zibraproject.org). Each amplicon with 400 base pairs (bp) length was overlapped with 60–80 bp of adjacent amplicon (Table S2).

The cDNA was synthesized using 25 ng of random hexamers, 25 μM of Oligo d(T)$_{20}$ primer and RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). The PCR was performed using the Platinum™ SuperFi II DNA Polymerase

(Thermo Fisher Scientific) in T100 Thermal Cycler (Bio-Rad). The PCR products were analyzed using 1.2% agarose gel electrophoresis.

2.3 | Sanger and nanopore MinION sequencing

Multiple overlapping short-fragment PCR amplicons were produced across the entire genome using 27 overlapping primer sets to sequence the whole genome of AiV (Table S2). The PCR products for Sanger sequencing were purified using ExoSAP-IT™ (Thermo Fisher Scientific). Sequencing was performed using the BigDye™ Terminator v3.1 Cycle Sequencing Kit and cleaned up with BigDye XTerminator™ Purification Kit (Applied Biosystems). The purified products were analyzed on the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems).

For Nanopore MinION sequencing, two long-fragment PCR amplicons were produced across the entire genome using two specific primer sets to sequence the whole genome of AiV. The amplicons included a 4972 bp amplicon using Scheme_AiV1 (forward) and Scheme_AiV32 (reverse), and 4338 bp amplicon using Scheme_AiV25 (forward) and Scheme_AiV52 (reverse) (Table S2). A sequencing library was constructed following the manufacturer's protocol (LSK-110; Oxford Nanopore Technology) and sequenced on the R9.4 MinION flow cells using the MinION v. 1.11.5 software.

2.4 | 5′/3′ rapid amplification of cDNA ends (RACE)

The 5′ and 3′ sequences were recovered using a 5′/3′ RACE Kit (2nd Generation; Roche) according to the manufacturer's protocol. We designed two genome-specific primers (GSP) based on the sequencing results of overlapping primers. One primer (GSP1; 5′-GCGCAGGGCGTTGGCACATGAA22; minus sense) was used to synthesize first-strand cDNA, and another primer (GSP2; 5′-AAGACGGGAGAAGGTTACGCACACT25; minus sense) was used for PCR. The PCR products were ligated into pDyne TA Blunt V2 (Dyne), and then recombinant plasmid was purified using an Accu-prep® Plasmid Mini Extraction kit (Bioneer) for the Sanger sequencing.

2.5 | Assembly and annotation

The sequences of genome obtained from Sanger and Nanopore sequencing reads were analyzed with FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Hybrid assemblies were performed with SPAdes in single-cell mode.[23] After aligning to the AiV reference genome (AB040749) using Bowtie 2,[24] the sequence depth was confirmed with Integrative Genomics Viewer (IGV). All nucleotide sequences were edited with the SeqMan Ultra program (DNASTAR v. Lasergene 17). The whole genome sequence was annotated against the complete AiV genome (AB040749) using Viral Annotation DefineR, a reference-based general software tool.[25]

2.6 | Phylogenetic analysis and secondary structure analysis

The multiple genome sequences from 21 AiV including three outgroup species were generated using MAFFT with the L-INS-I algorithm (ver. 7.505).[26] To determine the best-fit substitution models, IQ-TREE (ver. 1.6.12) was performed.[27,28] The phylogenetic tree was reconstructed using the Maximum likelihood and Bayesian inference methods. Maximum likelihood method was performed using the RAxML-NG (ver. 1.1.0) with the GTR + F + I + G4 nucleotide substitution model.[29] Bayesian phylogenetic inference method was performed using the MrBayes package (ver. 3.2.7a) with the GTR + F + I + G4 model.[30] The Markov chain Monte Carlo search was run for 10^6 generations with a sampling frequency of 5×10^2 using three heated and one cold chain. Graphic presentation of the phylogeny was performed with FigTree (ver. 1.4.3) (http://tree.bio.ed.ac.uk/software/figtree). The SimPlot analysis was performed to evaluate and visualize sequence similarity using SimPlot + +.[31] In addition, the secondary structure of the 5′ UTR was analyzed using the MFOLD program.[32]

2.7 | Selection pressure analysis

To examine positive selection sites acting on AiV-1, we used a codon model using codeML program as implemented in PAML software package (ver. 4.10.5).[33,34] We carried out likelihood ratio tests (LRTs) comparing a null model and an alternative model; M0 (one ratio) versus M3 (discrete), M1a (Nearly Neutral) versus M2a (Positive Selection), and M7 (beta) versus M8 (beta & ω).[35,36] Positively selected amino acid sites were identified based on Bayes Empirical Bayes posterior probabilities.[37] All PAML analyses were carried out using the F3X4 model of codon frequency.[38] The level of significance (P) for the LRTs was estimated using a χ2 distribution with given degrees of freedom (d.f.). The test statistic calculated as twice the difference of log-likelihood between the models (Equation 1)

$$2\Delta \ln L = 2[\ln L_1 - \ln L_0], \tag{1}$$

where L_1 and L_0 are the likelihoods of the alternative and null models, respectively.

3 | RESULTS

3.1 | Virus detection and isolation

AiV was detected in only one out of 20 human stool samples using the AiV universal assay. The AiV-positive specimen (CAU170613) was negative for other listed viruses in this study. Genotype-specific

RT-qPCR confirmed that the AiV strain belong to genotype A. The isolation of wild-type AiV strain CAU170613 (OQ121963) was attempted in Vero cells. The reference strain (AB040749) was propagated successfully with cytopathic effect (CPE) evident; however, attempts to propagate the newly identified strain CAU170613 (OQ121963) were unsuccessful (Figure S2). The CPE induced by the reference strain (AB040749) was characterized by the presence of cell lysis and cell debris (Figure S2). No increase in Ct values was detected when the AiV universal assay was performed after each passage (passages were at least three times and several attempts for each passage).

3.2 | Sequence data processing

Nanopore sequencing generated 13,199 reads (130–29,071 bp) and Sanger sequencing 190 reads (1–942 bp), respectively. Nanopore data quality scores ranged from 4 to 28, lower than the Sanger data quality scores from 4 to 58 (Table S3). Sequence alignment to a reference genome (AB040749) was done using the Bowtie 2, and full coverage of the AiV genome was achieved. The generated sequence data for wild-type AiV strain CAU170613 is available from GenBank accession number OQ121963. The OQ121963 showed 8,228 nt (nucleotide) length of the complete genome sequence and 95.18% identity with the reference sequence (AB040749). The sequence contained approximately 19.69% A, 21.44% T, 21.01% G, and 37.86% C nucleotides.

3.3 | Annotation

The OQ121963 consisted of an ORF with 7296 nt, encoding a potential polyprotein precursor of 2432 amino acids (aa), preceded by 676 nt and followed by 256 nt and a poly(A) tail. As shown in Figure 1A, a leader protein (L) was encoded at the N-terminus of the polyprotein, followed by structural proteins P1 (VP0, VP3, and VP1), and nonstructural proteins P2 (2A, 2B, and 2C) and P3 (3A, 3B, 3C, and 3D). We confirmed that the genome organization of OQ121963 is identical to AiV with the protein sizes in line with expected values for L protein, structural proteins P1 (VP0, VP3, VP1) and nonstructural proteins P2 (2A, 2B, 2C), P3 (3A, 3B, 3C, 3D; Figure 1A,B).

The Figure 1B indicated the size of the L protein, structural proteins P1 (VP0, VP3, VP1) and nonstructural proteins P2 (2A, 2B, 2C), P3 (3A, 3B, 3C, 3D). We identified that the different region (L, VP3, VP1, 2A, 2B, 2C, 3A, 3B, 3C, 3D) size of OQ121963 were identical with that of AiV genotype A. In addition, the size of the structural protein containing VP0 region was smaller by 1107 nt in genotype A compared to 1110 nt in genotype B.

3.4 | Sequence heterogeneity analysis and phylogenetic analysis of complete sequences

Maximum likelihood and Bayesian inference phylogenetic analyses were performed based on the sequences of twenty-one complete genomes and other *Kobuvirus* reference sequences from GenBank (Figure 2; Table S4). The trees showed that the human *Kobuvirus* strains from same species clustered together with OQ121963 in the AiV-1 lineage. Lineage AiV-1 viruses shared >89.7% nt sequence identity and the AiV-1 strains clustered in two genotypes A and B sharing >94.6% identity within each genotype. Strain OQ121963 shared 90.1%–96.7% identity with AiV-1 sequences, and clustered with genotype A strains (Figure 2; Figure S3).

3.5 | Sequence heterogeneity analysis and phylogenetic analysis of partial region sequences

Phylogenetic trees were generated for individual regions including the structural proteins P1, and the nonstructural proteins P2 and P3 to investigate differences based on function. AiV-1 sequences were consistently clustered in genotypes A and B for all regions investigated, and OQ121963 clustering with genotype A strains.

Sequence analysis of the structural protein region P1 and nonstructural protein regions P2 and P3 showed 94.03%, 96.95%, and 95.43% nt sequence identity with AB040749, respectively (Table 1). Since the lowest identity refers to the region where variation occurs, the amino acid identity of VP0, VP3, and VP1 in the P1 region was compared. The VP3 (98.70%) region had higher amino acid identity compared to VP0 (94.08%) and VP1 (95.43%), and the sequences of VP1 and VP0 were further analyzed. Alignment of the VP0 region showed nt sequence identity of 93.53% with AB040749 (Table 1). Among the 1107 nt of VP0 region, four gaps (nucleotide deletions) were identified at 1296–1297 nt (1 gap), 1299–1300 nt (1 gap), 1381–1382 nt (1 gap), and 1848–1849 nt (3 gaps) in the OQ121963 sequence when compared to AB040749 (Figure S4). Moreover, there was a lot of variation (mismatched nucleotides) across the whole region, which resulted in low identity. The degree of variation was further confirmed through sequence alignment of the VP1 region (Figure S5). Gaps were not identified between a total of 834 nt sequences, but there was a lot of variation (mismatched nucleotides) across the whole region, which resulted in low identity (Figure S5).

OQ121963 in 3B region showed 92.50% and 97.50% sequence identity with AB040749 and JX564249 belonging to genotype A, respectively; however, 80.00% and 78.75% with MK372823 and GQ927704, respectively, showing lower base identity than strains belonging to genotype B (Table 1, Figure 3A). As a result of comparing OQ121963 belonging to genotype A, genotyping was possible with 9 nt (mismatched nucleotides) out of a total of 81 nt sequences (5918–5998 nt) (Figure 3B). The 3D region showed the highest amino acid identity (98.26%) in nonstructural protein region (Table 1).

3.6 | Sequence heterogeneity analysis and phylogenetic analysis of 5′ UTR region sequences

The 5′ UTR, which is non-polyprotein region, had 95.46% nt identity with AB040749. Of the total 629 nucleotides, there were two
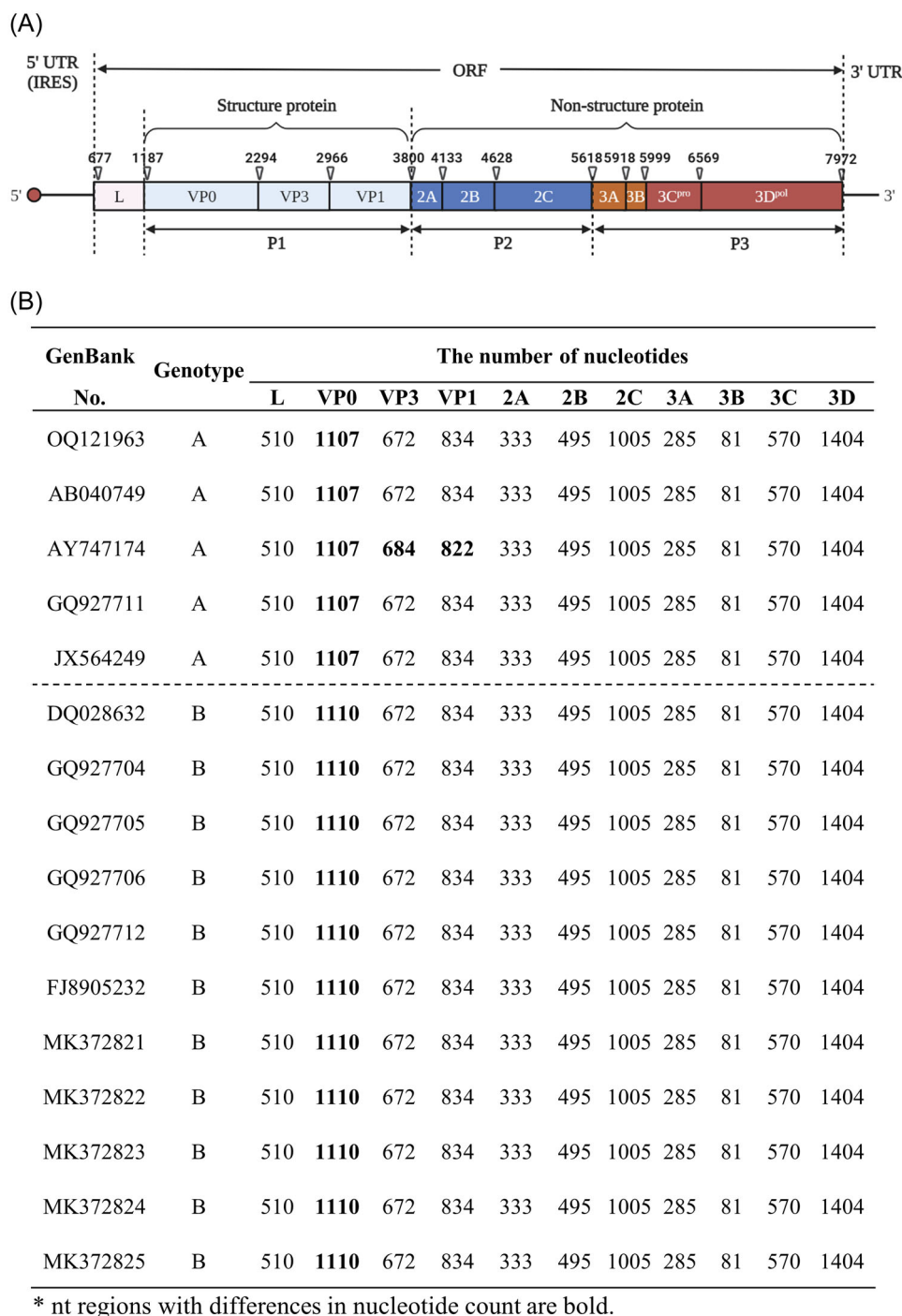
(A)

(B)

GenBank No.	Genotype	The number of nucleotides										
		L	VP0	VP3	VP1	2A	2B	2C	3A	3B	3C	3D
OQ121963	A	510	**1107**	672	834	333	495	1005	285	81	570	1404
AB040749	A	510	**1107**	672	834	333	495	1005	285	81	570	1404
AY747174	A	510	**1107**	**684**	**822**	333	495	1005	285	81	570	1404
GQ927711	A	510	**1107**	672	834	333	495	1005	285	81	570	1404
JX564249	A	510	**1107**	672	834	333	495	1005	285	81	570	1404
DQ028632	B	510	**1110**	672	834	333	495	1005	285	81	570	1404
GQ927704	B	510	**1110**	672	834	333	495	1005	285	81	570	1404
GQ927705	B	510	**1110**	672	834	333	495	1005	285	81	570	1404
GQ927706	B	510	**1110**	672	834	333	495	1005	285	81	570	1404
GQ927712	B	510	**1110**	672	834	333	495	1005	285	81	570	1404
FJ8905232	B	510	**1110**	672	834	333	495	1005	285	81	570	1404
MK372821	B	510	**1110**	672	834	333	495	1005	285	81	570	1404
MK372822	B	510	**1110**	672	834	333	495	1005	285	81	570	1404
MK372823	B	510	**1110**	672	834	333	495	1005	285	81	570	1404
MK372824	B	510	**1110**	672	834	333	495	1005	285	81	570	1404
MK372825	B	510	**1110**	672	834	333	495	1005	285	81	570	1404

* nt regions with differences in nucleotide count are bold.

FIGURE 1 Genome organization of the AiV strains using the "Protein to nucleotide alignment tool" (ProSplign). (A) Genome organization of the OQ121963. The polyprotein of virus is encoded by a leader protein (L) followed by structural proteins P1 (VP0, VP3, and VP1) and nonstructural protein P2 (2A, 2B, and 2C) and P3 (3A, 3B, 3C, and 3D). (B) A comparison of the number of nucleotides for encoding regions between the AiV strains. AiV, Aichivirus.

notable conserved regions with a length >50 nt sequences from 3 to 59 and 349–391 nt. Conversely, the sequence of 123 nt (60–182 nt) out of a total of 629 nucleotides had 4 gaps and 18 nt differences (mismatched nucleotides), suggesting a mutated region of AiV (Figure S6).

In the phylogenetic analysis, all region segments formed two clusters with genotypes A and B; however, the 5′ UTR region was not clustered. The GQ927711, which is genotype A from Germany, clustered with genotype B strains discovered at the same time from Germany (Figure 4). In the 5′ UTR region, the GQ927711 strain had 95%–97% sequence identity with genotype A, and 99.2%–99.7% identity with the German strains (GQ927704, GQ927712, GQ927706, GQ927705). This finding suggests the recombinant polyprotein region of genotype A and the 5′ UTR of the German strains.

FIGURE 2 Phylogenetic tree constructed for whole genome sequences of AiV strain OQ121963 detected in the current study and other representative strains. The tree was generated using the Maximum likelihood and Bayesian inference methods on MAFFT with the L-INS-I algorithm. Country source of AiV strains are characterised by different colours. The AiV strains clustered in two genotypes A and B. The strain that have been sequenced in our study AiV strain (OQ121963) is marked in red in AiV genotype A cluster. AiV, Aichivirus.

TABLE 1 Comparison of amino acid and sequence identity of the isolated Korean AiV strain OQ121963 with reference strains.

Region	Amino acid identity (%) AB040749	Sequence identity (%) AB040749	JX564249	MK372823	GQ927704
WGS	96.59	95.18	96.96	90.51	90.73
UTR	-	97.93	97.62	94.60	95.23
P1	96.07	94.03	95.60	88.67	88.48
P2	96.10	96.95	97.16	91.00	91.05
P3	96.21	95.43	97.95	90.90	91.45
VP0	94.08	93.53	94.96	86.69	87.05
VP3	98.70	93.95	94.94	89.34	89.79
VP1	95.43	94.79	97.47	90.03	87.80
2 A	97.00	96.10	97.30	91.29	91.59
2 B	97.38	94.95	96.57	90.91	90.30
2 C	97.17	95.36	97.39	91.01	91.21
3 A	94.94	95.56	97.74	90.27	90.27
3 B	94.22	92.50	97.50	80.00	78.75
3 C	97.44	96.00	96.00	90.33	91.00
3 D	98.26	97.15	98.01	91.52	92.24

Note: As a phylogenetic tree, AB040749 and JX564249 were selected as the most distant and closest strain from OQ121963 in genotype A. MK372823 and GQ927704 were selected as representative strains for each country in genotype B.
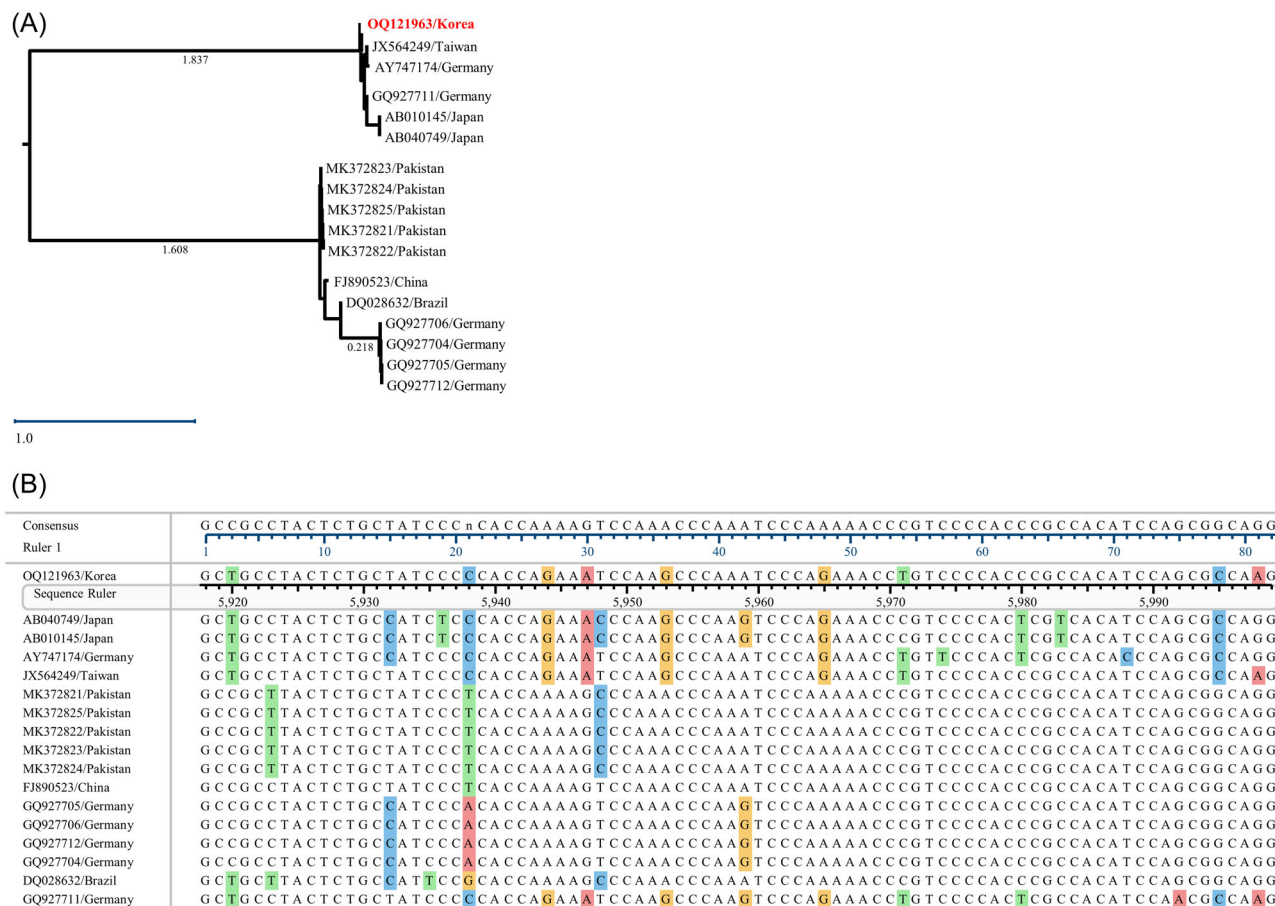
Abbreviation: AiV, Aichivirus.

FIGURE 3 (A) Phylogenetic tree of the 3B region using AiV strain OQ121963 detected in the current study and other representative strains. The tree was generated using the Maximum-likelihood (RAxML) on MAFFT. The AiV strains formed two clusters with two genotypes A and B. (B) Aligning the sequence of the 3B region using MegAlign Pro. Genotyping was possible with 9 nt (mismatched nucleotides) out of a total of 81 nt sequences (5918–5998 nt). Colored nucleotides show differences from consensus sequence. AiV, Aichivirus

3.7 | SimPlot analysis of AiV

Previous result was confirmed by data similarity analysis (Figure 5). First, the similarity of the structural protein coding region was 5%–10% lower than that of the non–structural protein coding region. Second, the similarity was high in the 3D region of non-structural protein coding region. Third, the similarity of the 5′ UTR region was the highest among all regions; however, by having a variable region up to the first 200 bp, it showed low similarity in the N–terminal 200 bp of 5′ UTR sequence. Finally, the similarity was compared by grouping (genotype A and genotype B) of the AiV, and low similarity in the VP1 and 3B coding regions was confirmed.

3.8 | Secondary structure analysis

The 5′-end (120 nt) of AiV was reported in AB040749 to fold into three stem-loop structure, including SL–A, SL–B, and SL–C, which is an important structure for viral RNA replication.[39,40] The

secondary structure of the 5′ UTR was predicted using the MFOLD program (Figure 6A). The 5′-end sequence of OQ121963 was 66 nt shorter than AB040749. In the secondary structure of the 5′ UTR for AB040749, the position of SL–A, SL–B, and SL–C was in 4–45, 46–69, and 73–104 nt, respectively. In contrast, the OQ121963 was predicted to have only SL–C in 7–37 nt, with no SL–A, SL–B. No significant difference in the secondary structure of internal ribosome entry site (IRES) was observed among the sequences.

3.9 | Selection pressure analysis

Of the 2433 codons that comprise the viral polyprotein, the overall non-synonymous to synonymous substitution rate (d_N/d_S) for the entire AiV coding region was estimated to be 0.05, suggesting mainly negative selection.[41] Only four residues (1077 A, 1097 K, 1117 P, and 1147 P) and two residues (1388D and 1392 T) located in the coding region of 2 A and 2 C proteins, respectively, were detected as being under strong positive selection sites using Naïve Empirical Bayes
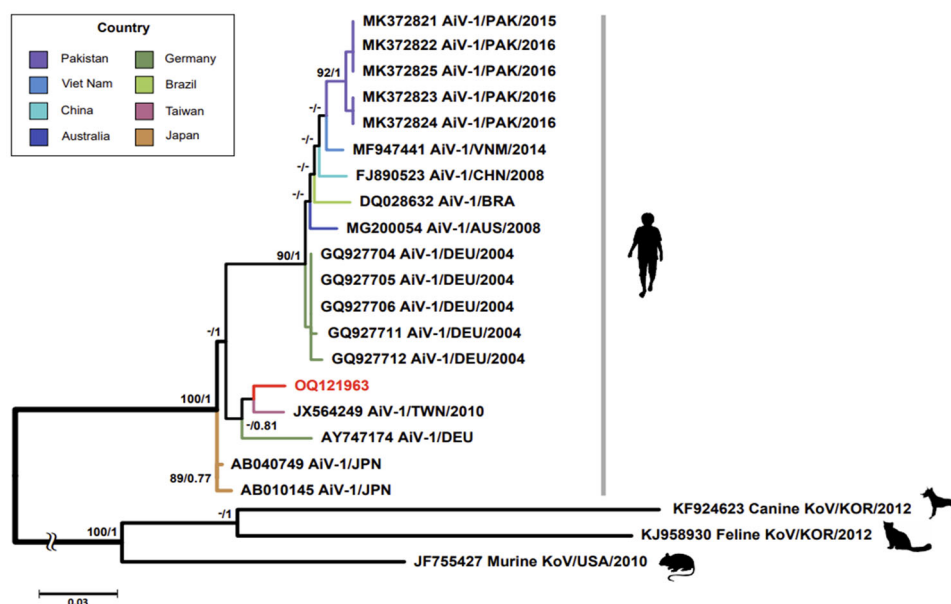
FIGURE 4 Phylogenetic tree of the 5′ UTR region using AiV strain OQ121963 detected in the current study and other representative strains. The tree was generated using the Maximum-likelihood (RAxML) on MAFFT. Dash (-) indicates nodes with PPBI or ML < 70%. Tilde (~) indicates that the outgroups' branch length has been shortened. Country source of AiV strains are characterised by different colours. The strain that have been sequenced in our study AiV strain (OQ121963) is marked in red. AiV, Aichivirus.
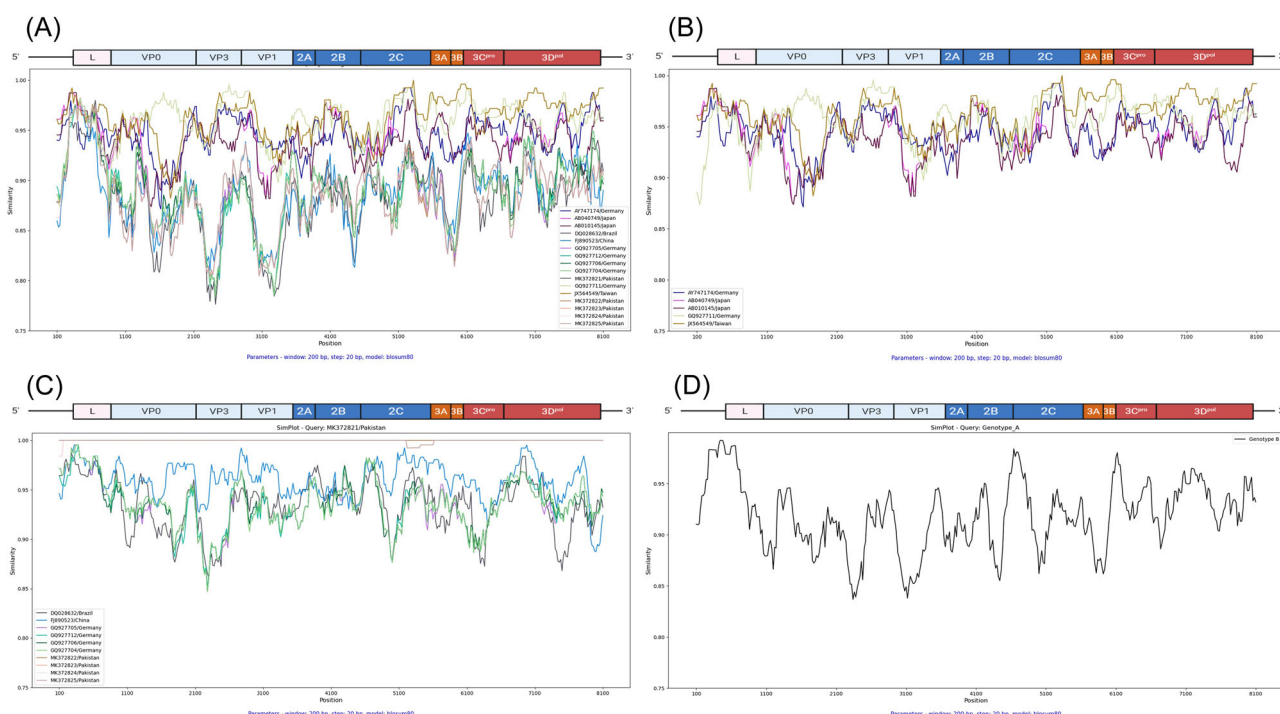

FIGURE 5 SimPlot analysis of whole genome for our study AiV strain (OQ121963) with reference strains. (A) Sixteen reference strains sequence were compared with OQ121963 sequence. (B) Genotype A strains sequence comparison with OQ121963 strain sequence (genotype A). (C) Genotype B strains sequence comparison with MK372821 AiV/Pakistan strain sequence (genotype B). (D) Comparison of genotype A and genotype B. Plotted using SimPlot + + software with a 200 bp window size, 20 bp step, and BLOSUM 80 model. Different colours are used to distinguish different genotypes. The x-axis represents the position of the nucleotide. The y-axis represents bootstrap values. AiV, Aichivirus.

(NEB) and Bayes Empirical Bayes (BEB) analyses (Figure 6B). To evaluate whether putatively positive selection sites occur in functional domains, P2 protein with identified positive selection sites were compared to AB040749. The result showed that all these

codons were adjacent or between the key functional domains, such as transmembrane region and p-loop containing nucleotide triphosphate hydrolases, which are responsible for interaction with host cell and replication (Figure S7).
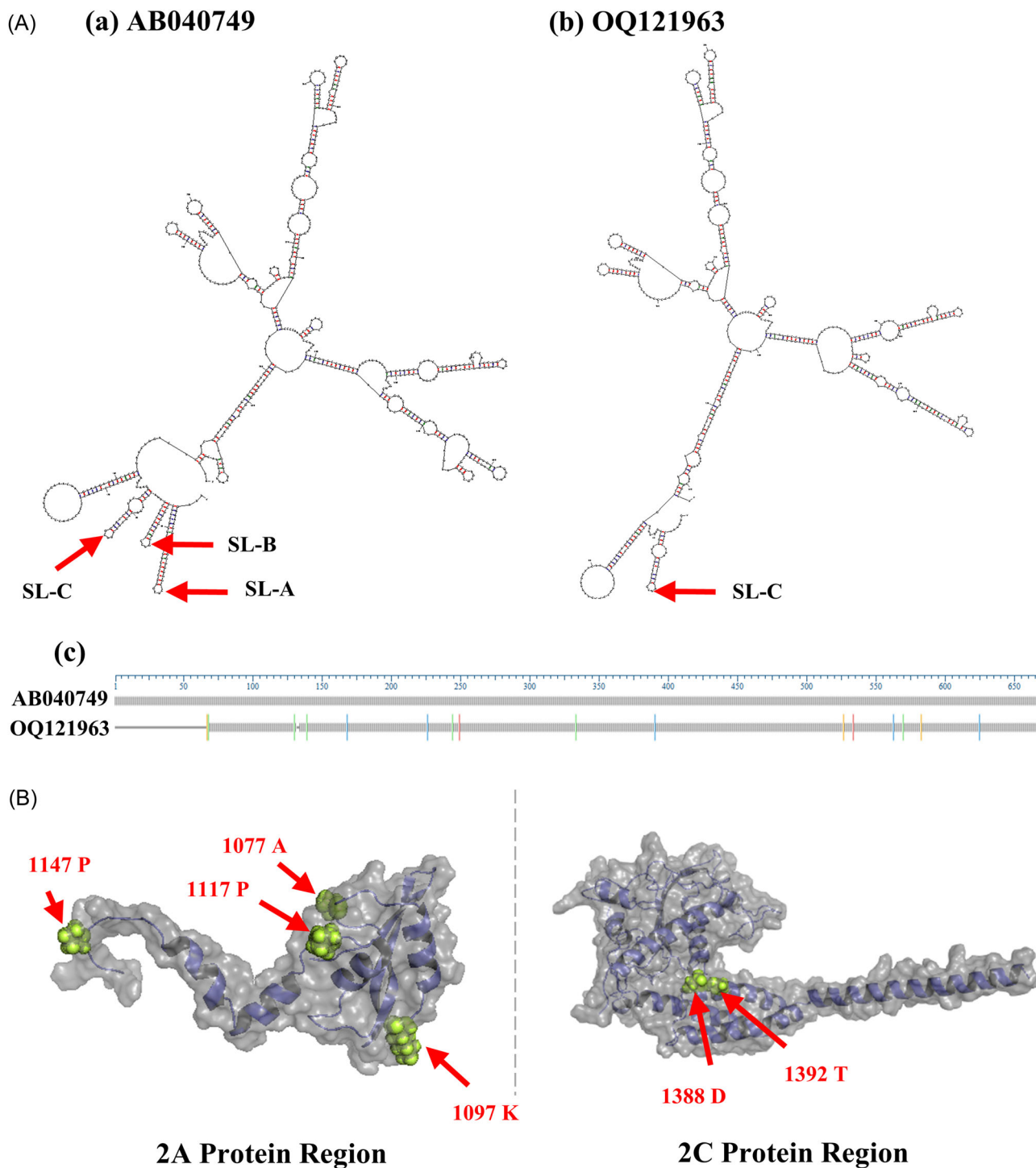
FIGURE 6 (A) Comparison of predicted secondary structure of 5′ UTR region using the MFOLD program. (a) AB040749 sequence (1–713 nt), (b) OQ121963 sequence (1–645 nt). The important structure pointed out in red at Stem-loop (SL) is the stem-loop A (SL–A), SL–B, and SL–C. (c) Snapshots of 5′ UTR region of the aligned AB040749 and OQ121963 sequences were taken using MegAlign Pro. Colored nucleotides show differences from consensus sequence. (B) Positively selected amino acid sites located in the coding region of 2 A and 2 C proteins. Four residues (1077 A, 1097 K, 1117 P, and 1147 P) and two residues (1388 D and 1392 T) located in the coding region of 2 A and 2 C proteins.

4 | DISCUSSION

AiV has been detected in many countries, but several clinical studies confirmed a low occurrence of AiV infection.[13] Some clinical studies have demonstrated the coexistence of AiV with other enteric viruses in fecal samples of patients suffering from gastroenteritis.[12,13,42] Therefore, AiV can circulate unnoticeable and contribute to diverse viral infections that cause intestinal disease. This suggests the need for improved epidemiological survey to detect circulating virus before clinical manifestation. This study reports for the first time molecular

isolation and characterization of complete genome sequence of AiV (OQ121963) in Korea. Of note, only 16 complete genome sequences of AiV have been reported to date, thus we tried to analyze the character of AiV new strain (OQ121963). The results of this study provide valuable information for future studies on the diversity of AiV.

The WGS of AiV was performed using two different methods: Sanger sequencing with RACE, and NGS with Oxford Nanopore Technology or Nanopore-based sequencing technology. Currently, applying Nanopore-based sequencing for genomic study is difficult due to complex errors within reads.[43] The hybrid assembly approach takes advantage by complementing it with shorter, identical target sequences to generate longer, more precise transcription and improved assembly.[44] Hybrid assembly was performed using Nanopore and Sanger sequencing reads, which yield high–quality WGS of AiV (OQ121963). The genome organization of OQ121963 was identical to the complete genome of AiV and showed more than 90% sequence identity.[3] In the phylogenetic analysis, the isolated Korean AiV strain (OQ121963) was close to the German (GQ927711) and Taiwan strains (JX564249). To understand the genetic variation of circulating AiV, a Korean strain (OQ121963) was compared with 16 other AiV complete genomes, which will further help in identifying strains of the virus circulating around the world.

To analyze the variable and conserved region, the sequence and amino acid identity of OQ121963 were confirmed. The structural protein P1 region (96.07%) showed the lowest amino acid identity, and the nonstructural protein P3 region (96.21%) showed the highest amino acid identity. The VP0 region in the P1 was confirmed as the variable region with the lowest amino acid identity between strains. The VP0 region is known as the receptor attachment site for enteroviruses belonging to the *Picornaviridae* family.[45] The amino acid differences in the VP0 region may influence receptor binding, thus suggesting that the receptor-binding motif of AiV requires further investigation. Furthermore, 3D region in the P3 was confirmed as a conserved region with the highest identity. The 3D region of the *Kobuvirus* genome, which is highly conserved, is the recognition region for *Kobuvirus* detection including AiV.[46] Therefore, human AiV can also be identified through the 3D conserved region when *Kobuvirus* is detected. In addition, the 5′ UTR region of AiV has proven to be the most suitable genomic target for molecular detection.[47] In this study, the conserved regions of the 5′ UTR (3–59 and 349–391 nt) were suggested to be suitable for future AiV detection. Moreover, 5′ UTR regions were not clustered by genotype in phylogenetic analysis, unlike other regions. In the phylogenetic analysis of the 5′ UTR region, German strain GQ927711 had genotype A but was clustered together with genotype B strains originating from the same country in Germany (Figure 4); in contrast, genotype A strain (GQ927711) was clustered with the different country-originating genotype A strains during the phylogenetic analysis of the WGS of AiV (Figure 2). Therefore, we believe that the 5′ UTR may be useful for contact tracing. In this study, we analyzed the sequence of the AiV and established that possible variations in the 5′ UTR region may play an important role in viral fitness and pathogenesis. Further investigation of the structural

protein (P1) and nonstructural protein (P2 and P3) is needed to elucidate the possible effects of these variations. The capsid protein VP1, which expresses the antigenicity of the virus, is genetically well-suited for differentiating subtypes of AiV.[48] In addition, the 3B region was confirmed to be divided according to the genotype, suggesting that genotyping was possible.

Growth of the wild-type AiV Korean strain (OQ121963) was attempted in Vero cells, without success. Therefore, a comparative analysis was conducted with a reference strain (AB040749) that can be cultured through amino acid homology to determine why our strain could not be cultured. The amino acid homology between the two strains was calculated for each coding region; low homology was shown in the structural protein VP0 (94.08%) and VP1 (95.43%) encoding regions, and the nonstructural protein 2 A (97.00%) encoding region in the P2. The VP0 and VP1 are regions involved in protein-protein interactions in capsid protein signal transduction, antigen recognition, and cell–to–cell.[49] The 2 A region was involved in both negative- and positive-strand synthesis, and essential for virus replication.[50] The secondary structure at the 5′-end of OQ121963 was further investigated; as seen in the previous studies, folding of the stem-loop at the 5′-end is essential for viral RNA replication and production of infectious viral particles.[39,40] Three stem-loop structures (termed SL–A, SL–B, and SL–C) of AiV genome have been reported at the 5′-end.[39] The loop segment of SL–A (4–45 nt), SL–B (112–115 nt), and downstream of SL–C (112–118 nt) are important for viral RNA replication. Among them, SL–A plays an important role in transfection,[39] and SL–B will enter the entry site with SL–C and pseudokot structure.[40] Furthermore, we predicted the secondary structure at 5′-end of OQ121963 and confirmed the absence of SL–A and SL–B. These results suggested that OQ121963 has a problem with viral RNA replication. Further studies are required to elucidate the real impact of this region on virus cultivation. In future studies, it is important to consider several constraints, such as the insufficient availability of comprehensive demographic and clinical information, variations in the primer sets used, and variances in the groups studied, when analyzing and understanding the findings.

In conclusion, this study reports for the first time the identification and complete of the whole genome characterization of AiV found in Korea. Our analysis results provided useful information for understanding the identification, global spread, and genetic variation of AiV genotype A found in Korea, as well as future studies on AiV diversity.

AUTHOR CONTRIBUTIONS

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. Repository names and accession number(s) are included within the article. The data that supports the findings of this study are available in the supplementary material of this article.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study was approved by the Ethics Committee of the Korea Association of Health Promotion (KAHP), Seoul, Republic of Korea. The institutional committee approved the experiments (IRB No. 130750–201902–BR–019).

ORCID

Seong-il Eyun http://orcid.org/0000-0003-4687-1066
Changsun Choi http://orcid.org/0000-0001-7730-8538

REFERENCES

1. Yamashita T, Kobayashi S, Sakac K, et al. Isolation of cytopathic small round viruses with BS-Cl cells from patients with gastroenteritis. *J Infect Dis*. 1991;164(5):954-957.
2. Rivadulla E, Romalde JL. A comprehensive review on human Aichi virus. *Virol Sin*. 2020;35(5):501-516.
3. Kebe O, Fernandez-Garcia MD, Fall A, et al. Prevalence and genetic diversity of Aichi virus 1 from urban wastewater in Senegal. *Intervirology*. 2021;64(2):96-101.
4. Rivadulla E, Varela MF, Romalde JL. Epidemiology of Aichi virus in fecal samples from outpatients with acute gastroenteritis in northwestern Spain. *J Clin Virol*. 2019;118:14-19.
5. Kitajima M, Gerba C. Aichi virus 1: environmental occurrence and behavior. *Pathogens*. 2015;4(2):256-268.
6. Taghinejad M, Ghaderi M, Mousavi-Nasab SD. High frequency of Aichivirus in children with acute gastroenteritis in Iran. *Pediatr Infect Dis*. 2021;39(7):576-579.
7. Nielsen ACY, Gyhrs ML, Nielsen LP, Pedersen C, Böttiger B. Gastroenteritis and the novel picornaviruses Aichi virus, cosavirus, saffold virus, and salivirus in young children. *J Clin Virol*. 2013;57(3): 239-242.
8. Yip CCY, Lo KL, Que TL, et al. Epidemiology of human parechovirus, Aichi virus and salivirus in fecal samples from hospitalized children with gastroenteritis in Hong Kong. *Virol J*. 2014;11:182.
9. Chuchaona W, Khamrin P, Yodmeeklin A, et al. Detection and characterization of Aichi virus 1 in pediatric patients with diarrhea in Thailand. *J Med Virol*. 2017;89(2):234-238.
10. Onosi O, Upfold NS, Jukes MD, Luke GA, Knox C. The first molecular detection of Aichi virus 1 in raw sewage and mussels collected in South Africa. *Food Environ Virol*. 2019;11:96-100.
11. Chen BC, Huang TS, Huang NY, Chen CS, Chen YS, Chang TH. Low seroprevalence of Aichi virus infection in Taiwan. *Pathogens*. 2021;10(5):553.
12. Northill JA, Simmons RJ, Genge D, Moore FA. Molecular characterization of the first reported Aichivirus A in Australia. *Access Microbiol*. 2020;2(4):e000099.
13. Räsänen S, Lappalainen S, Kaikkonen S, Hämäläinen M, Salminen M, Vesikari T. Mixed viral infections causing acute gastroenteritis in children in a waterborne outbreak. *Epidemiol Infect*. 2010;138(9): 1227-1234.
14. Rodrigues Portes S, Mello Volotao E, Rose T, et al. Aichi virus positivity in HIV-1 seropositive children hospitalized with diarrheal disease. *Curr HIV Res*. 2015;13(4):325-331.
15. Lukashev AN. Recombination among picornaviruses. *Rev Med Virol*. 2010;20(5):327-337.
16. John G, Sahajpal NS, Mondal AK, et al. Next-generation sequencing (NGS) in COVID-19: a tool for SARS-CoV-2 diagnosis, monitoring new strains and phylodynamic modeling in molecular epidemiology. *Curr Issues Mol Biol*. 2021;43(2):845-867.
17. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333-351.
18. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc*. 2017;12(6):1261-1276.
19. Grädel C, Terrazos Miani MA, Baumann C, et al. Whole-genome sequencing of human enteroviruses from clinical samples by nanopore direct RNA sequencing. *Viruses*. 2020;12(8):841.
20. Xu Y, Lewandowski K, Lumley S, et al. Detection of viral pathogens with multiplex nanopore MinION sequencing: be careful with crosstalk. *Front Microbiol*. 2018;9:2225.
21. Laver T, Harrison J, O'neill PA, et al. Assessing the performance of the Oxford nanopore technologies minion. *Biomol Detect Quantif*. 2015;3:1-8.
22. Kitajima M, Hata A, Yamashita T, Haramoto E, Minagawa H, Katayama H. Development of a reverse transcription-quantitative PCR system for detection and genotyping of Aichi viruses in clinical and environmental samples. *Appl Environ Microbiol*. 2013;79(13): 3952-3958.
23. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455-477.
24. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359.
25. Schäffer AA, Hatcher EL, Yankie L, et al. VADR: validation and annotation of virus sequence submissions to GenBank. *BMC Bioinformatics*. 2020;21:211.
26. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780.
27. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268-274.
28. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587-589.
29. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 2019;35(21):4453-4455.

30. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)*. 2001;17: 754-755.

31. Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Tatusova T. FLAN: a web server for influenza virus genome annotation. *Nucleic Acids Res.* 2007;35(suppl 2):W280-W284.

32. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003;31(13):3406-3415.

33. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*. 1997;13(5):555-556.

34. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586-1591.

35. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 1998;15(5): 568-573.

36. Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*. 2002;19(6):908-917.

37. Yang Z. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 2005;22(4):1107-1118.

38. Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 1994;11(5): 725-736.

39. Sasaki J, Kusuhara Y, Maeno Y, et al. Construction of an infectious cDNA clone of Aichi virus (a new member of the family *Picornaviridae*) and mutational analysis of a stem-loop structure at the 5′ end of the genome. *J Virol*. 2001;75(17):8021-8030.

40. Nagashima S, Sasaki J, Taniguchi K. The 5′-terminal region of the Aichi virus genome encodes cis-acting replication elements required for positive-and negative-strand RNA synthesis. *J Virol*. 2005;79(11): 6918-6931.

41. Spielman SJ, Weaver S, Shank SD, et al. Evolution of viral genomes: Interplay between selection, recombination, and other forcesIn: Anisimova M, ed. *Evolutionary Genomics. Methods in Molecular Biology*. Vol 1910. Humana; 2019:427-468.

42. Pham NTK, Khamrin P, Nguyen TA, et al. Isolation and molecular characterization of Aichi viruses from fecal specimens collected in Japan, Bangladesh, Thailand, and Vietnam. *J Clin Microbiol*. 2007;45(7):2287-2288.

43. Weirather JL, de Cesare M, Wang Y, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research*. 2017;6:100.

44. Goldberg SMD, Johnson J, Busam D, et al. A sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc Natl Acad Sci USA*. 2006;103(30): 11240-11245.

45. Dang M, Wang X, Wang Q, et al. Molecular mechanism of SCARB2-mediated attachment and uncoating of EV71. *Protein Cell*. 2014;5(9): 692-703.

46. Ribeiro J, Lorenzetti E, Júnior JCR, da Silva Medeiros TN, Alfieri AF, Alfieri AA. Phylogenetic analysis of VP1 and RdRP genes of Brazilian aichivirus B strains involved in a diarrhea outbreak in dairy calves. *Arch Virol*. 2017;162(12):3691-3696.

47. Baumgarte S, de Souza Luna LK, Grywna K, et al. Prevalence, types, and RNA concentrations of human parechoviruses, including a sixth parechovirus type, in stool samples from patients with acute enteritis. *J Clin Microbiol*. 2008;46(1):242-248.

48. Lukashev AN, Drexler JF, Belalov IS, Eschbach-Bludau M, Baumgarte S, Drosten C. Genetic variation and recombination in Aichi virus. *J Gen Virol*. 2012;93(6):1226-1235.

49. Zhu L, Wang X, Ren J, et al. Structure of human Aichi virus and implications for receptor binding. *Nat Microbiol*. 2016;1(11):16150.

50. Sasaki J, Taniguchi K. Aichi virus 2A protein is involved in viral RNA replication. *J Virol*. 2008;82(19):9765-9769.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.