The dynamics of molecular evolution of emerging avian reoviruses through accumulation of point mutations and genetic re-assortment

Lisanework E. Ayalew,[1,*] Khawaja Ashfaque Ahmed,[1] Zelalem H. Mekuria,[2] Betty Lockerbie,[1] Shelly Popowich,[1] Suresh K. Tikoo,[3] Davor Ojkic,[4] and Susantha Gomis[1,*]

[1]Department of Veterinary Pathology, Western College of Veterinary Medicine, University of Saskatchewan, 52 Campus Drive, Saskatoon, Saskatchewan, S7N 5B4, Canada, [2]College of Veterinary Medicine and Global One Health Initiative, Infectious Disease Molecular Epidemiology Laboratory, The Ohio State University, 1900 Coffey Road, Columbus, Ohio, 43210, USA, [3]Vaccinology & Immunotherapeutic Program, School of Public Health, University of Saskatchewan, 104 Clinic Place, Saskatoon, Saskatchewan, S7N 5E3, Canada and [4]Animal Health Laboratory, Laboratory Services Division, University of Guelph, 419 Gordon St., Guelph, Ontario, N1H 6R8, Canada

*Corresponding authors: E-mails: lisanework.ayalew@usask.ca; susantha.gomis@usask.ca

Abstract

In the last decade, the emergence of variant strains of avian reovirus (ARV) has caused enormous economic impact in the poultry industry across Canada and USA. ARVs are non-enveloped viruses with ten segments of double-stranded RNA genome. So far, only six genotyping cluster groups are identified worldwide based on sequence analysis of the σC protein encoded by the S1 segment. In this study, we performed deep next generation whole-genome sequencing and analysis of twelve purified ARVs isolated from Saskatchewan, Canada. The viruses represent different genotyping cluster. A genome-wide sequence divergence of up to 25 per cent was observed between the virus isolates with a comparable and contrasting evolutionary history. The proportion of synonymous single-nucleotide variations (sSNVs) was higher than the non-synonymous (ns) SNVs across all the genomic segments. Genomic segment S1 was the most variable as compared with the other genes followed by segment M2. Evidence of positive episodic/diversifying selection was observed at different codon positions in the σC protein sequence, which is the genetic marker for the classification of ARV genotypes. In addition, the N-terminus of σC protein had a persuasive diversifying selection, which was not detected in other genomic segments. We identified only four ARV genotypes based on the most variable σC gene sequence. However, a different pattern of phylogenetic clustering was observed with concatenated whole-genome sequences. Together with the accumulation of point mutations, multiple re-assortment events appeared as mechanisms of ARV evolution. For the first time, we determined the mean rate of molecular evolution of ARVs, which was computed as 2.3×10^{-3} substitution/site/year. In addition, widespread geographic intermixing of ARVs was observed between Canada and USA, and between different countries of the world. In conclusion, the study provides a comprehensive analysis of the complete genome of different genotyping clusters of ARVs including their molecular rate of evolution and spatial distribution. The new findings in this study can be utilized for the development of effective vaccines and other control strategies against ARV-induced arthritis/tenosynovitis in the poultry industry worldwide.

Key words: NGS; avian reovirus; evolution; phylogeography.

1. Introduction

Infection with arthrogenic avian reoviruses (ARVs) can cause tenosynovitis/arthritis syndrome characterized by unilateral or bilateral swelling of the hock joint resulting in lameness in chickens. Depending on the degree of severity, the infection can result in poor growth, poor production, and sometimes death, causing considerable economic losses (Gouvea and Schnitzer 1982; Liu et al. 2003). In the past 8 years, increased incidence and frequent outbreaks of ARV-associated arthritis/ tenosynovitis have been reported in broilers in many geographic locations of USA and Canada. These outbreaks occurred despite vaccination of breeders and broilers with commercial ARV vaccines (Lu et al. 2015). The problem continues to cause significant economic losses to the broiler industry.

ARVs are non-enveloped viruses with an icosahedral double capsid containing segmented double-stranded RNA (dsRNA) genome. The genomic segments are classified as large (L1, L2, and L3), medium (M1, M2, and M3), and small (S1, S2, S3, and S4) based on their electrophoretic mobility on a polyacrylamide gel (Spandidos and Graham 1976). The L1, L2, and L3 segments encode for λA, λB, and λC proteins, respectively. Proteins λA and λB are distributed in the inner core with functions associated with scaffolding and transcription, respectively, while λC protein is distributed in the turrets and functions as a capping enzyme. The M1, M2, and M3 segments express μA (inner core), μB (outer capsid), and μC (non-structural) proteins, respectively. The μB and μC proteins are expressed as precursors and each is post-translationally cleaved into three different proteins. Functions of M1 gene products span from aiding in virus penetration into the host cells to induction of protein recruitment and formation of viral factories. While S1 segment is tri-cistronic and encodes for the σC (outer capsid), P10 (non-structural; with fusogenic and permeabilizing activities), and P17 (non-structural; nuclear protein with unknown function) proteins, the S2, S3, and S4 gene segments express σA (inner core), σB (outer capsid), and σNS (non-structural) proteins, respectively. The σC protein is involved in the virus binding onto the host cells and is the most variable protein encoded by reovirus genome (Guardado Calvo et al. 2005). The σC protein contains specific epitopes, which induce production of neutralizing antibodies (Wickramasinghe et al. 1993).

Previously, we reported the isolation and characterization (phenotypic, genotypic, and antigenic) of emerging ARVs in Canada (Ayalew et al. 2017). Based on the sequence of the most variable σC gene, our isolates were grouped into four distinct genotypes. Interestingly, none of the isolates were grouped in the vaccine cluster. Moreover, we could reproduce the disease in specific pathogen free (SPF) chickens using virus isolates from each cluster group. Interestingly, all virus isolates produced similar lesions with no difference in the degree of gross and histopathological lesions. In addition, antibodies produced against commercial vaccines did not neutralize the variant viruses in a virus neutralization test (Ayalew et al. 2017). Goals of this study were: 1) to perform a deep next-generation sequencing (NGS) of virus isolates from each genotypic cluster; 2) to analyse the genetic make up and nucleotide diversity of each virus including phylogenetic incongruences and re-assortment events between genome segments from different genotyping cluster groups; and 3) to examine the geographic distribution pattern and global evolutionary rate of ARVs based on the most variable σC gene. We effectively demonstrated the mechanisms of evolution, phylogeographic structure, and rate of evolution of emerging ARVs. The results of this study will increase our understanding of the pattern of evolution of arthrogenic ARVs and should help in the selection of candidate viruses for the development of safe and effective broad-spectrum vaccines.

2. Materials and methods

2.1 Virus isolation/purification, RNA extraction, and σC DNA sequencing

Tendon tissue samples collected from broiler chickens with clinical signs of arthritis/tenosynovitis, which were confirmed positive for ARV infection were processed for virus isolation and plaque purification as previously described (Ayalew et al. 2017). Monolayers of Leghorn male hepatoma (LMH CRL-2177) cells (ATCC) propagated in T-75 flasks were used for virus isolation. The infected cells showing complete cytopathic effect were collected, freeze thawed five times, and virus purification was performed by ultracentrifugation (Ayalew et al. 2017). Viral RNA was purified from purified virus particles by phenol-chloroform (24:1) extraction method as described in Labrada et al. (2002). The purity of the RNA extracted from each sample was checked by Nanodrop and quantified using Qubit fluorometer. Reverse transcription, σC gene segment PCR amplification, and Sanger nucleotide sequencing was performed as previously described (Ayalew et al. 2017).

2.2 Sequence library preparation and NGS

Purified RNA from each virus isolate was converted into double-stranded cDNA using first and second strand DNA synthesis modules (New England Biolabs). The sizes of the DNA fragments were analysed using Agilent High Sensitivity DNA kit in a bio-analyzer (Agilent genomics). The newly synthesized double-stranded DNA was quantified using Qubit DNA BR assay kit in a fluorometer (Thermo Fisher Scientific). From each sample, a total of 1 ng of DNA was used to prepare sequence libraries using the Nextera XT DNA library prep kit (Illumina) as per the manufacturer's protocol. The DNA library clean up and size selection was performed using Ampure XP beads (Beckman Coulter) and checked by a bio-analyzer. The indexed libraries were normalized using library normalization beads 1 (LNB1) as per the manufacturer's protocol. The normalized DNAs were pooled, denatured with 0.2 N NaOH, diluted with HT1 hybridization buffer and loaded onto the Miseq reagent kit V3 (600 cycles) cartridge (Illumina). Sequencing was performed on a MiSeq system following the manufacturer's protocol (Illumina). The quality of the run generated by the Real-Time Analysis software was examined real-time using Sequencing Analysis Viewer (SAV) v1.8.

2.3 Full-length viral genome assembly and sequence analysis

Image analysis and base calling data generated as FASTQ files were extracted from the MiSeq. Illumina SAV v1.8 software was used to examine the quality metrics of the sequence data. The FASTQ files were then imported into Geneious v.11 (Bio Matters). The forward and the reverse reads were paired using the Illumina short-read default parameters and mapped to the Reo/PA/Broiler/05682/12 reference genome (Tang et al. 2016). Mapped reads were used to identify single-nucleotide variants (SNVs) from the reference by applying a minimum frequency

filter of 1 per cent and ignoring broken reads and non-specific matches. A technology-associated filter was used to remove any pyro error variant with a minimum length of 3 and a frequency below 0.8. Identified SNVs were analysed for polymorphism type and protein effect. The synonymous SNV (sSNV) and non-synonymous SNV (nsSNV) frequency in each gene segment of virus isolates was summarized and, the difference between group means of sSNVs and snSNVs analysed using one-way ANOVA and Tukey's test (GraphPad Prism 6).

2.4 Whole-genome nucleotide alignment and phylogenetic analysis

After sequence reads were mapped to Reo/PA/Broiler/05682/12 (Tang et al. 2016), individual consensus sequences were generated for each virus isolate. The most common bases with lowest ambiguity at a position and with highest quality scores (highest calculated Q-score) were used to determine the consensus sequence from the contigs covering each segment of the reovirus genome. Concatenated segments of each virus isolate were aligned with the reference sequence by the mVISTA method (http://genome.lbl.gov/vista/mvista/submit.shtml). Circular representation of the sequence data with sSNVs, nsSNVs, sequence coverage, and protein-coding regions were created using Circa genomics software (OMGenomics). Moreover, concatenated whole-genome sequences of four turkey avian reoviruses (TARV) (Dandár et al. 2015; Tang et al. 2015) were retrieved from the GenBank and compared with the concatenated whole-genome sequences of our isolates after ClustalW alignment with BLOSUM cost matrix technique.

2.5 Genetic diversity analysis of viral genome segments between viral isolates

The genetic diversity between genomic segments of each virus isolate was analysed by calculating the ratio of sSNV and nsSNV. The species diversity index was calculated using the Shannon–Wiener index method (H-index) using the Shannon entropy equation as described in Shannon and Weaver (1949) and Jost (2006). The effective number of species (ENS) representing the true diversity associated with the H-index was calculated using the formula exp. (H-index) (Jost 2006). The calculated H-index and ENS of each segment were tabulated and plotted using GraphPad Prism 6.

2.6 Phylogenetics, phylogeography, and viral evolution analysis

The genomic segments and the concatenated reovirus genome from individual viruses were aligned using Clustal W plugin in Geneious v.11 (Biomatters). For the nucleotide analysis, a default setting was utilized in Clustal W cost matrix, a gap opening cost of 15 and a gap extension cost of 6.6. For translated amino acid sequence alignment, the GONNET cost matrix was used with a gap opening cost of 10 and a gap extension cost of 0.1. The variable N-terminal region of σC sequences of our isolates and σC sequences retrieved from the GenBank (Supplementary Table S1) were used for phylogenetic analysis. The phylogenetic trees were constructed using the same parameters as described by Guindon et al. (2010). The genetic divergence of isolates from the reference viral sequence was performed by constructing maximum likelihood of phylogenies using general time reversible (GTR) substitutions rates with gamma distribution and 1,000 bootstrap replicates. The genetic re-assortment events were inferred by analysing the incongruent topologies among the phylogenetic trees. In addition, the assembled concatenated sequences of the reovirus genome segments were analysed by RDP4 software package using various methods (Martin et al. 2015) for the detection of potential recombination and/or re-assortment events. The estimation of the recombination break points was used to differentiate between recombination versus re-assortment events (Martin et al. 2015). For molecular evolution and phylogeography analyses, 472 σC sequences from ARVs isolated from different parts of the world including USA, Canada, India, China, Germany, Taiwan Japan, Israel, South Korea, Tunisia, Iran, Netherlands, Hungary, Brazil, France, and Australia were retrieved from the GenBank (Supplementary Table S1). The sequences were chosen based on the availability of the information on date and the geographic location of virus isolation. To determine the degree of clock-likeness, linear regression analysis was performed using TempEst-v1.5. For a subset of sequences that have sufficient phylogenetic structure, Bayesian inference of maximum clade credibility (MCC) tree was performed using the Markov Chain Monte Carlo (MCMC) method. Posterior probabilities were calculated for the MCMC trees and chain length generating 1,000,000 trees and sampling every 100th tree. The initial 25 per cent of trees were discarded as a burn-in, and the others were used to construct a majority rule tree. Alternatively, a temporally resolved maximum-likelihood phylogenetic trees were also constructed by considering GTR models and coalescent priors using the algorithm in Tree Time.

3. Results

3.1 Summary of the NGS run and lane metrics

We performed deep paired-end Illumina sequencing of the genome of twelve purified ARV isolates using randomly fragmented library of viral RNAs with an insert size of 300 bp. The viruses were selected from a total of thirty-eight isolates based on their phylogenetic clustering of the most variable ARV gene, σC. NGS of the genomes of twelve reovirus isolates produced a total of 15.93 GB data, which was approximately 1.3 GB sequence data per sample. The respective reads representing individual virus isolates were mapped to the reference reovirus genome (Reo/PA/Broiler/05682/12) (Tang et al. 2016). The result showed that the twelve viruses (D1 to D12) achieved ($>250\times$) read coverages in all the genomic segments with some minor sample-to-sample variabilities (Fig. 1). Overall, average read coverages of each nucleotide of the genome segments were 19,929 for L1, 18,309 for L2, 16,104 for L3, 16,348 for M1, 17,273 for M2, 15,772 for M3, 12,357 for S1, 13,104 for S2, 13,655 for S3, and 11,767 for S4.

3.2 Whole-genome sequence comparison of ARV isolates

The comparative genome analysis of the twelve reovirus isolates indicated the presence of high degree of genetic heterogeneity among ARVs circulating in Canada. In comparison to the reference viral genome sequence, the isolates from this study have more than 25 per cent genome-wide sequence divergence. Concatenated whole-genome nucleotide alignment of virus isolates against the reference is shown in Fig. 2A. Our analyses showed that the different segments in the genome of the twelve ARV isolates underwent comparable as well as contrasting evolutionary path. For example, the L1 segment in D1, D3, D8, D10,
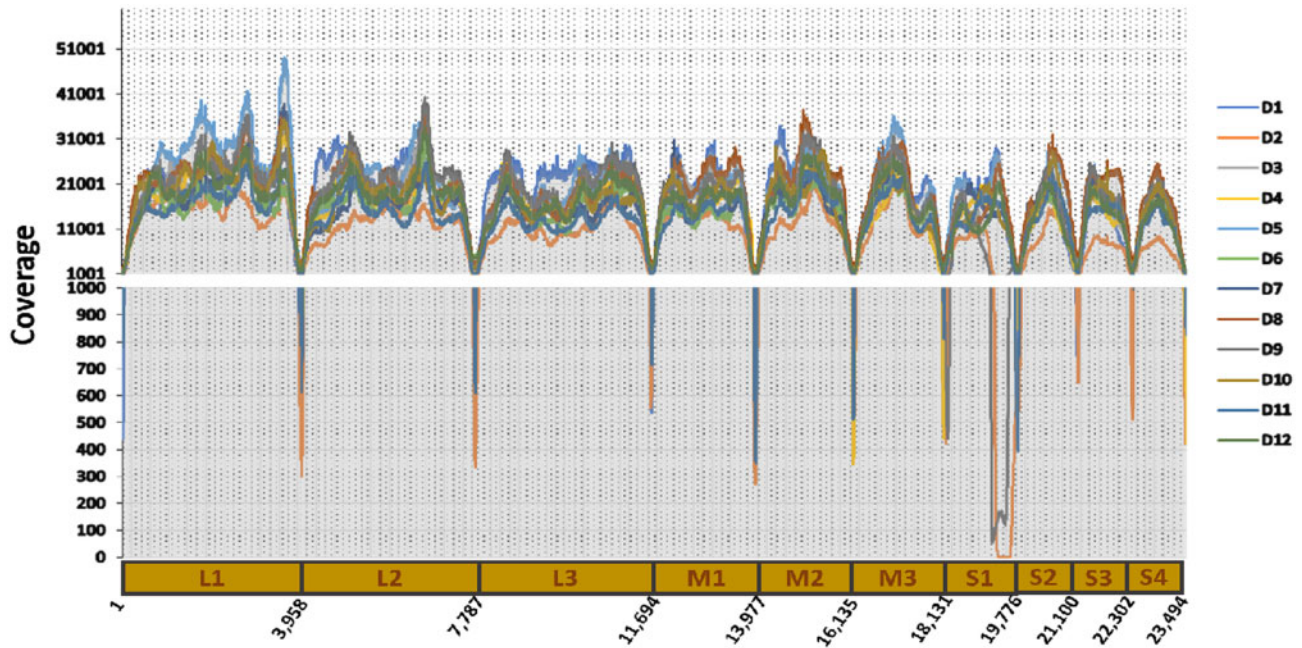
Figure 1. Line plot showing coverage of each nucleotide of all the genome segments. The colour codes indicate the different isolates of ARV. The name of the genomic segments and the length are indicated.

D4, and D11 isolates had a segment-wide nucleotide homology to the reference isolates suggesting a close ancestral relationship. In contrast, homology in L2, L3, and M1 segments was restricted to localized regions. Within each of these localized regions, the percentage of nucleotide identity to the reference viral sequence ranged from 74.6 to 99.9 per cent. The M2 segment was rather unique in its bimodal pattern with the highest segment-wide homology to isolates D1, D2, D3, D6, D8, D9, and D12 and rather divergent clustering to isolates D4, D5, D7, and D10. The latter group only shares 74.7 to 76.9 per cent nucleotide identity with the reference viral genome sequence.

The M3 and S1 genome segments are found to be evolutionarily distant segments between the isolates with a minimum percentage nucleotide identity of 53.8 and 80 per cent for S1 and M3, respectively. Localized differences were noted between D2, D6, D9, and D12 viruses in the S2, S3, and S4 genome segments, which encode proteins that are components of the virus structure. However, in most of the isolates S2, S3, and S4 showed higher degree of nucleotide identity across the entire segment. The overall nucleotide level identities in these regions range from 86.5 to 99 per cent S2, 81.3 to 99.5 per cent S3, and 88.5 to 100 per cent for S4.

In summary, the highest percentage of nucleotide identity (HNI) of each genome segment of viruses in cluster group-4 (C-4) was with viruses within the same cluster group except segment L1 of viruses D1 and D3, which had the HNI with segment L1 of virus D10 (i.e. C-5) (Fig. 2B). Within C-4, all the genome segments of virus D8 had the HNI with the corresponding genome segments of virus D11 (i.e. C-5) and vice versa except segment S1 which had HNI with segment S1 of virus D12 (i.e. C-4). The genome segment S1 of virus D11 had the HNI with virus D4 (i.e. C-6) (Fig. 2B). The genome segments L1, L2, and S1 of virus D5, which was clustered in C-2 had the HNI with L1 and S1 of virus D7 of the same genotyping cluster group. However, the rest of the genome segments had the HNI with the corresponding genome segments of viruses grouped under C-4, C-5, or C-6. Similarly, within C-5, genome segments M1 and S1 of virus D2

had the HNI with the corresponding genome segments of virus D10 grouped under the same cluster. The rest of the genome segments had HNI with genome segments of viruses in the other cluster groups (Fig. 2B). Genome segments L1, S1, and S4 of virus D4 of C-6 had the HNI within the corresponding genome segments of virus D11 of the same cluster group. The rest of the genome segments of virus D4 had the HNI with genome segments of viruses grouped under C-2, C-4, or C-5. Within cluster group C-4, all genome segments of virus D1 had the HNI with the corresponding genome segments of virus D3 and vice versa except genome segments L1 and M3, which had the HNI with segment L1 and M3 of virus D10 and D12, respectively. Similarly, genome segments of virus D6 had the HNI with genome segments of virus D12 and vice versa. The exception was the genome segments L1 of virus D6 and D12, and S4 of virus D6 which had the HNI with the corresponding genome segments of virus D9. The genome segments L1, M1, S2, and S4 of virus D9 had the HNI with the corresponding genome segments of virus D6, whereas genome segments L2, L3, M2, M3, S1, and S3 of D9 had the HNI with the respective genome segments of virus D12 (Fig. 2B). The complete sequence of the genomic segments of each ARV isolate is deposited in the GenBank (Supplementary Table 2). We also examined the whole-genome sequence relatedness between the ARV isolates and TARV sequences retrieved from the GenBank. The nucleotide identity of the Hungary (Dandár et al. 2015) and the Pennsylvania, USA (Tang et al. 2015) TARV isolates and our ARV isolates was between 64.565 and 67.341 per cent and 80.367 and 81.662 per cent, respectively. The Pennsylvania TARV isolates had also a higher degree of nucleotide identity with our ARV isolates than with the TARVs isolated in Hungary (Supplementary Fig. S1).

3.3 sSNVs between the ARV isolates and the reference virus strain

In addition to examining the consensus sequence, the within sample diversity was evaluated by quantifying the sSNVs and
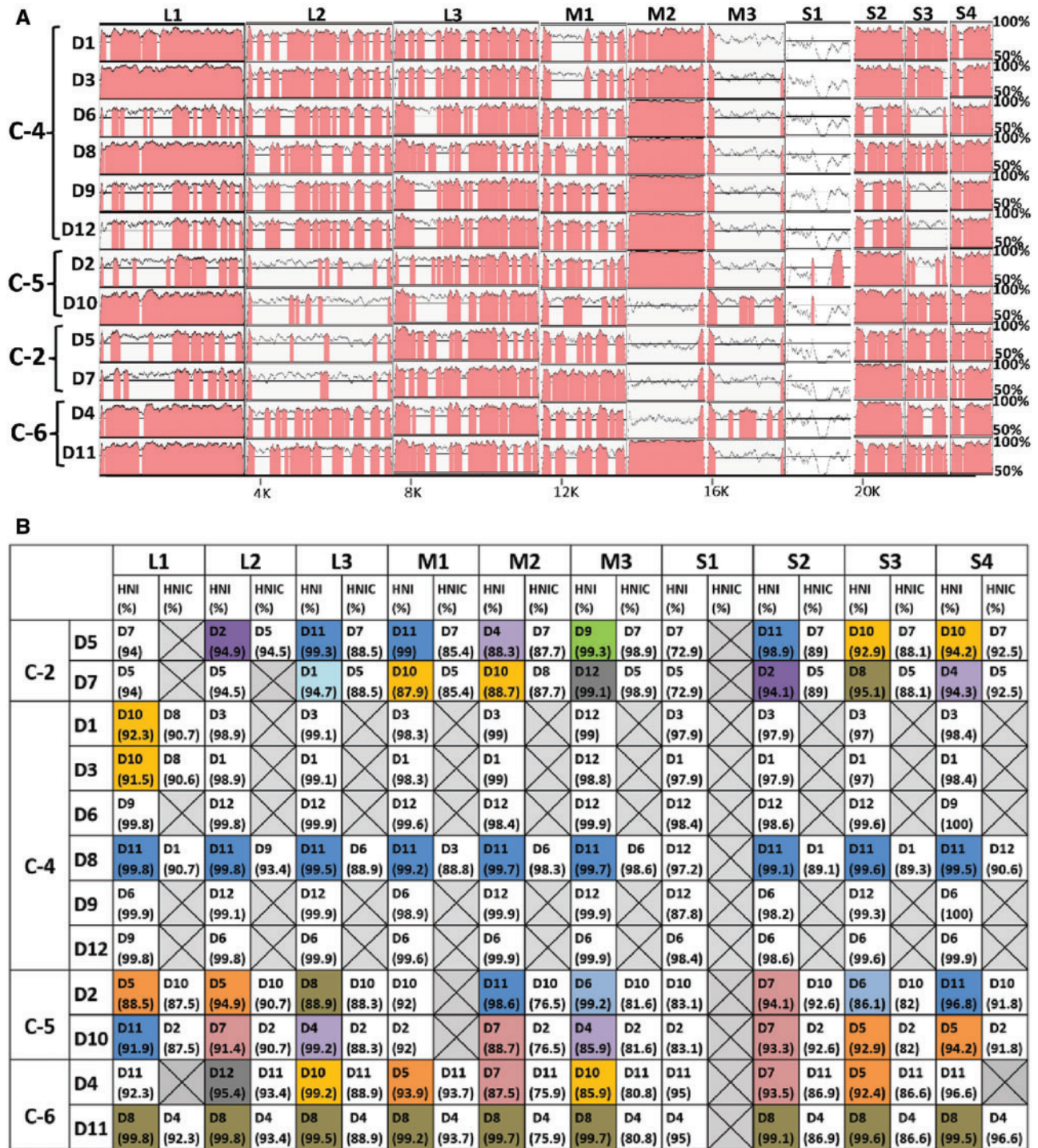
A

B

		L1 HNI (%)	L1 HNIC (%)	L2 HNI (%)	L2 HNIC (%)	L3 HNI (%)	L3 HNIC (%)	M1 HNI (%)	M1 HNIC (%)	M2 HNI (%)	M2 HNIC (%)	M3 HNI (%)	M3 HNIC (%)	S1 HNI (%)	S1 HNIC (%)	S2 HNI (%)	S2 HNIC (%)	S3 HNI (%)	S3 HNIC (%)	S4 HNI (%)	S4 HNIC (%)
C-2	D5	D7 (94)	X	D2 (94.9)	D5 (94.5)	D11 (99.3)	D7 (88.5)	D11 (99)	D7 (85.4)	D4 (88.3)	D7 (87.7)	D9 (99.3)	D7 (98.9)	D7 (72.9)	X	D11 (98.9)	D7 (89)	D10 (92.9)	D7 (88.1)	D10 (94.2)	D7 (92.5)
	D7	D5 (94)	X	D5 (94.5)	X	D1 (94.7)	D5 (88.5)	D10 (87.9)	D5 (85.4)	D10 (88.7)	D8 (87.7)	D12 (99.1)	D5 (98.9)	D5 (72.9)	X	D2 (94.1)	D5 (89)	D8 (95.1)	D5 (88.1)	D4 (94.3)	D5 (92.5)
C-4	D1	D10 (92.3)	D8 (90.7)	D3 (98.9)	X	D3 (99.1)	X	D3 (98.3)	X	D3 (99)	X	D12 (99)	X	D3 (97.9)	X	D3 (97.9)	X	D3 (97)	X	D3 (98.4)	X
	D3	D10 (91.5)	D8 (90.6)	D1 (98.9)	X	D1 (99.1)	X	D1 (98.3)	X	D1 (99)	X	D12 (98.8)	X	D1 (97.9)	X	D1 (97.9)	X	D1 (97)	X	D1 (98.4)	X
	D6	D9 (99.8)	X	D12 (99.8)	X	D12 (99.9)	X	D12 (99.6)	X	D12 (98.4)	X	D12 (99.9)	X	D12 (98.4)	X	D12 (98.6)	X	D12 (99.6)	X	D9 (100)	X
	D8	D11 (99.8)	D1 (90.7)	D11 (99.8)	D9 (93.4)	D11 (99.5)	D6 (88.9)	D11 (99.2)	D3 (88.8)	D11 (99.7)	D6 (98.3)	D11 (99.7)	D6 (98.6)	D12 (97.2)	X	D11 (99.1)	D1 (89.1)	D11 (99.6)	D1 (89.3)	D11 (99.5)	D12 (90.6)
	D9	D6 (99.9)	X	D12 (99.1)	X	D12 (99.9)	X	D6 (98.9)	X	D12 (99.9)	X	D12 (99.9)	X	D12 (87.8)	X	D6 (98.2)	X	D12 (99.3)	X	D6 (100)	X
	D12	D9 (99.8)	X	D6 (99.8)	X	D6 (99.9)	X	D6 (99.6)	X	D6 (99.9)	X	D6 (99.9)	X	D6 (98.4)	X	D6 (98.6)	X	D6 (99.6)	X	D6 (99.9)	X
C-5	D2	D5 (88.5)	D10 (87.5)	D5 (94.9)	D10 (90.7)	D8 (88.9)	D10 (88.3)	D10 (92)	X	D11 (98.6)	D10 (76.5)	D6 (99.2)	D10 (81.6)	D10 (83.1)	X	D7 (94.1)	D10 (92.6)	D6 (86.1)	D10 (82)	D11 (96.8)	D10 (91.8)
	D10	D11 (91.9)	D2 (87.5)	D7 (91.4)	D2 (90.7)	D4 (99.2)	D2 (88.3)	D2 (92)	X	D7 (88.7)	D2 (76.5)	D4 (85.9)	D2 (81.6)	D2 (83.1)	X	D7 (93.3)	D2 (92.6)	D5 (92.9)	D2 (82)	D5 (94.2)	D2 (91.8)
C-6	D4	D11 (92.3)	X	D12 (95.4)	D11 (93.4)	D10 (99.2)	D11 (88.9)	D5 (93.9)	D11 (93.7)	D7 (87.5)	D11 (75.9)	D10 (85.9)	D11 (80.8)	D11 (95)	X	D7 (93.5)	D11 (86.9)	D5 (92.4)	D11 (86.6)	D11 (96.6)	X
	D11	D8 (99.8)	D4 (92.3)	D8 (99.8)	D4 (93.4)	D8 (99.5)	D4 (88.9)	D8 (99.2)	D4 (93.7)	D8 (99.7)	D4 (75.9)	D8 (99.7)	D4 (80.8)	D4 (95)	X	D8 (99.1)	D4 (86.9)	D8 (99.6)	D4 (86.6)	D8 (99.5)	D4 (96.6)

Figure 2. (A) Alignment of concatenated whole-genome segments of ARV isolates in comparison to the reference viral genome (Reo/PA/Broiler/05682/12). Areas shaded with pink and white indicate ≥90 per cent, and <90 per cent sequence identities, respectively. Scale bar: The approximate length of the concatenated genome. (B) The heatmap of the HNI of each segment between the virus isolates and virus isolates within the same cluster group (HNIC).

nsSNVs at a given genomic position. The frequencies of intra-sample variants were computed against the coverage as shown in the Circa plot (Fig. 3A and B). The presence of higher proportion of sSNVs compared to nsSNVs was observed in all genome segments (Fig. 4A and C). With respect to the S1 segments, higher nsSNV proportion was detected only in nucleotide positions 1–977 in all the virus isolates. Furthermore, the patterns of nsSNV and sSNV tallies between the respective genome segments of the virus isolates were examined by a one-way ANOVA analysis. The results showed a statistically significant ($P < 0.05$) difference, with the highest degree of nsSNV observed in genome segment S1 (Fig. 4B and D). The lowest and highest nsSNV frequency of genome segment S1 was detected in virus isolates D2 and D8, respectively. Variants from other genomic segments are also summarized (Fig. 4A and C). A wider variability in the proportion of sSNV was recorded in genome segment
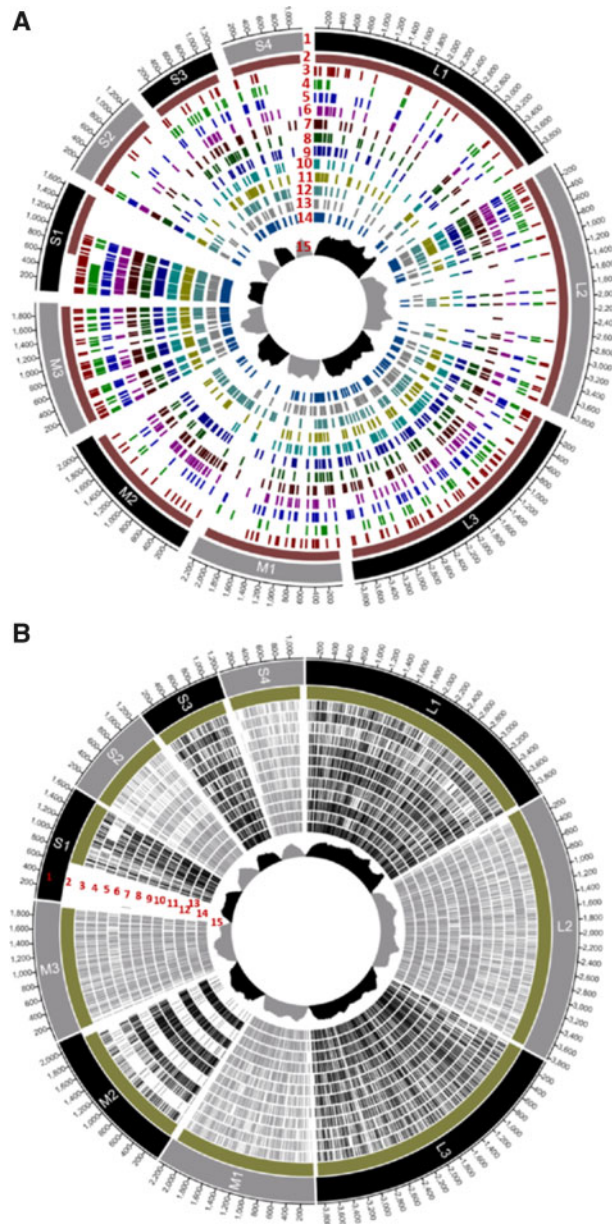
Figure 3. Circa plot of non-synonymous (ns) (A) and synonymous (s) (B) SNPs between the ARV isolates. Lane 1, Reference viral genome sequence (Reo/PA/Broiler/05682/12); Lane 2, protein coding regions; Lanes 3–14, viruses D1–D12, respectively; Lane 15, average coverage.

S2 with the highest variation observed in virus isolate D6 followed by segment M2 with the highest variations detected in virus isolates D5 and D7. Genome segment L3 showed a small difference in the proportion of sSNV between the isolates. When we compared the sSNV of the corresponding genome segments between the virus isolates, genome segment M2 of virus D2, D6, D8, D9, D11, and D12 was the most conserved followed by segment S2 of virus D4 and D7. Analysis of the differences between group means of sSNV between the virus isolates based on Tukey's multiple comparison test showed statistically significant difference (P < 0.05) between genome segments L1 and M3; L2 and L3 or S4; L3 and M1 or M3; M1 and M3; M3 and S1 and S3 or S4 (Fig. 5A). Similarly, the analysis of differences between group means of nsSNV of different genome segments based on Tukey's

multiple comparison test showed statistically significant difference (P < 0.05) between segments L1 and L2, L3, M1, M3, S1, or S3; L2 and L3, M3 or S1; L3 and M1, M3, S1, or S4; M2 and M3 or S1; M3 and S1, S2, S3, or S4; S1 and S2, S3 or S4 (Fig. 5B).

Analyses of the codon positions in the σC protein encoding region of the S1 genomic segment for positive episodic or diversifying selection under a proportion of branches using the Mixed Effects Model of Evolution (MEME) tool in the Datamonkey adaptive evolution server (Murrell et al. 2012) indicated evidence of positive episodic/diversifying selection at five codon positions with a P-value threshold of 0.05 (Fig. 5C).

3.4 Nucleotide diversity between genomic segments

Shannon–Wiener Index (H-index) was used to calculate the nucleotide diversity between genome segments of virus isolates. The highest and lowest ENS in the major variant diversity was observed in genome segment S1 (ENS = 11.94) and M2 (ENS = 8.63), respectively (Fig. 6A). In contrast, the highest and lowest ENS in the minor variant was observed in the genome segment S4 (ENS = 11.42) and M2 (ENS = 5.91), respectively (Fig. 6B).

3.5 Phylogenetic analyses and phylogenetic incongruences in the genome sequences of the ARV isolates

As shown in the previous sections, genomes of reovirus isolates sequenced in this study exhibited high level of sequence diversity. It was also demonstrated that the relationship of individual isolates to a reference genome varied on the segmental basis, which was a preliminary indication for the occurrence of genetic re-assortment. Hence, to evaluate and measure the extent of genetic diversity as well as to detect the events of segmental re-assortment, phylogenetic analyses of each genomic segment was carried out. The evolutionary relatedness of the isolates was assessed using a neighbour-joining method with Jukes–Cantor genetic distance model. All the sequences included in the analyses were grouped into six Cluster groups based on the variable partial N-terminal σC nucleotide sequence (Fig. 7A). In this article, the same numbering scheme of the ARV genotype clusters as Kant et al. (2003) has been employed. Our virus isolates were grouped into four different cluster groups. As such, the virus isolates D5 and D7 were grouped under C-2. The virus isolates D1, D3, D6, D8, D9 and D12 were grouped under C-4. The virus isolates D2, D10 and the reference virus (Reo/PA/Broiler/05682/12) were grouped under C-5. The virus isolate D4 and D11 were grouped under C-6. None of the sequences from virus isolates in this study was grouped under C-1 or C-3. However, a different genotype clustering was observed when the phylogenetic tree was constructed using concatenated whole-genome sequences (Fig. 7B). In order to get better insight into re-assortment events that shaped our outcomes, we computed segmental phylogenetic trees and inferred the incongruent topologies within them (Fig. 8). Our analysis identified multiple re-assortment scenarios. For example, the L1 segment of virus D2 has a distant relationship to D5 and D7 viruses. However, the L2 segment of D2 clustered closely with D5 and D7 isolates (Fig. 8). Similarly, the D2 virus exhibited incongruent phylogenetic topologies in S1 segment. The D3 virus also seems to be composed of phylogenetically incongruent L1 and L3 genomic segments. On the contrary, the viruses that are colour shaded maintained the phylogenetically congruent clusters across the different segments of the reovirus genome. It appears that virus
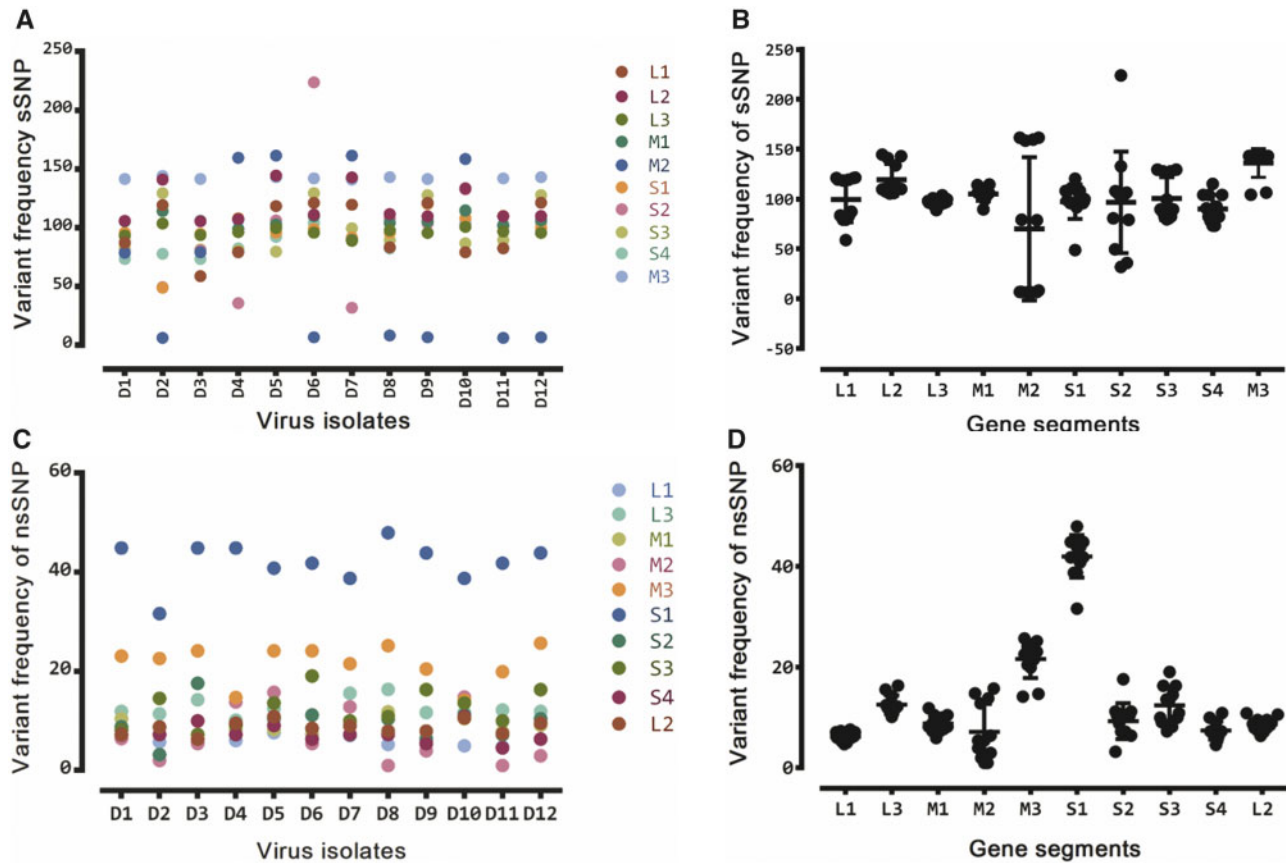
Figure 4. Proportion of sSNP (A) and nsSNP (C) frequency in each genomic segment of each virus. One-way ANOVA analysis of nsSNV (B) and sSNV (D) frequencies between the respective genome segments.

D8 (C-4) and virus D11 (C-6) are co-infection variants (Fig. 8). Neither recombination nor re-assortment events were detected between the sequences of the genome segments of our ARV isolates and TARV sequences retrieved from the GenBank after analysis by different techniques using RDP4 software package.

3.6 Geographic distribution and rate of nucleotide substitution of ARV

The spatial and temporal genetic divergence were analysed to determine the pattern of geographic distribution and molecular rate of evolution of arthrogenic ARV isolates. The results of the molecular clock and phylogeographic analysis are shown (Figs 9A and B and 10). To determine a global evolutionary rate, we analysed the sequence of 471 reovirus σC sequences using maximum-likelihood trees (Fig. 9A). Even though, sequences in the phylogenetic tree exhibit a linearly progressive branch lengths, our molecular-clock regression analysis using the root to tip genetic distance showed lack of good temporal resolution for calculating a global mean evolutionary rate. However, subsets of the sequence data from 1991 to 2016 had a good temporally resolved cluster with a mean rate of molecular evolution of 2.3×10^{-3} substitution/site/year. In order to further elucidate these events, we re-constructed the same data set in the phylogenetic tree to infer the spatial patterns among the global isolates. The result demonstrated extensive spatial inter-mixing of the ARVs between different countries/regions (Fig. 9B). We also observed widespread geographic intermixing of USA and Canadian ARVs (Fig. 10).

4. Discussion

Since 2011, there has been a sudden increase in the outbreaks of arthritis/tenosynovitis in broiler flocks in different parts of Canada and USA. The isolation and molecular characterization of viruses from these outbreaks have been reported (Rosenberger et al. 2013; Lu et al. 2015; Ayalew et al. 2017; Sellers 2017; Palomino-Tapia et al. 2018). Based on the sequence of most variable σC protein, ARVs isolated from different parts of the world are grouped into six groups (Liu et al. 2003; Kort et al. 2013; Lu et al. 2015; Ayalew et al. 2017; Egaña-Labrin et al. 2019). The virus isolates from all the six cluster groups were involved in the recent disease outbreaks in Canada and the USA (Lu et al. 2015; Ayalew et al. 2017; Palomino-Tapia et al. 2018; Egaña-Labrin et al. 2019). Even though pathogenic variant ARVs are genetically highly divergent, no difference in the degree of virulence between pathogenic viruses from different genetic lineages has been observed. As a result, the classification of ARVs based on pathotype has been difficult.

We previously reported characterization of ARVs isolated from clinical cases of arthritis/tenosynovitis in broilers from recent outbreaks (Ayalew et al. 2017). The molecular characterization was based on the ARV σC and σB gene sequences. Structural modelling and antigenicity prediction of the σC protein suggested that there was a difference in the secondary structure of the predicted antigenic epitopes between viruses in the different genotyping cluster groups as a result of amino acid substitutions. Since multiple variant strains co-circulate at the same time, control of ARV-induced arthritis/
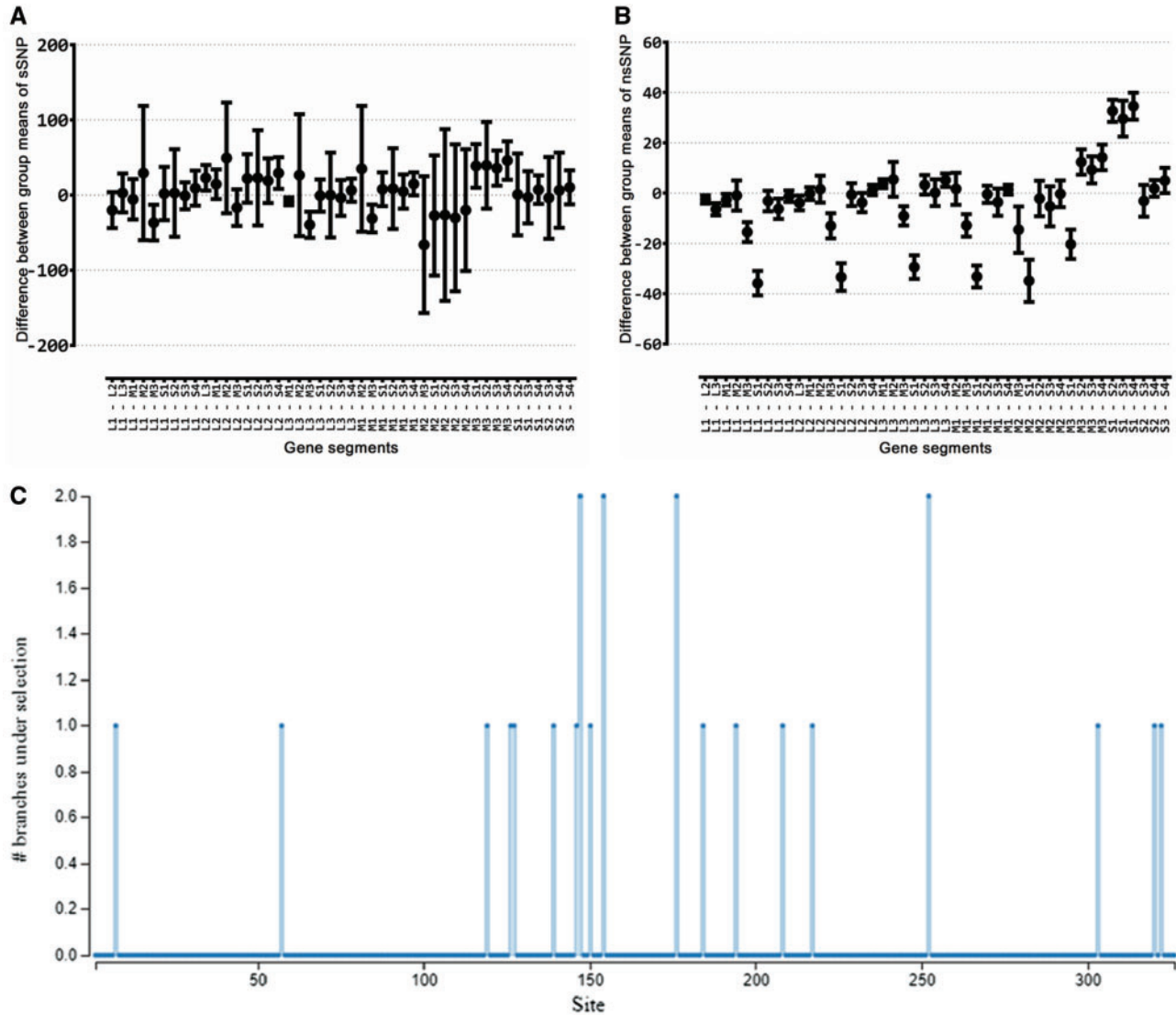
Figure 5. Analysis of differences between group means of sSNP (A) and nsSNP (B) between the virus isolates based on a Tukey's multiple comparison test. (C) Codon positions indicating evidence of positive episodic/diversifying selection (P < 0.05).
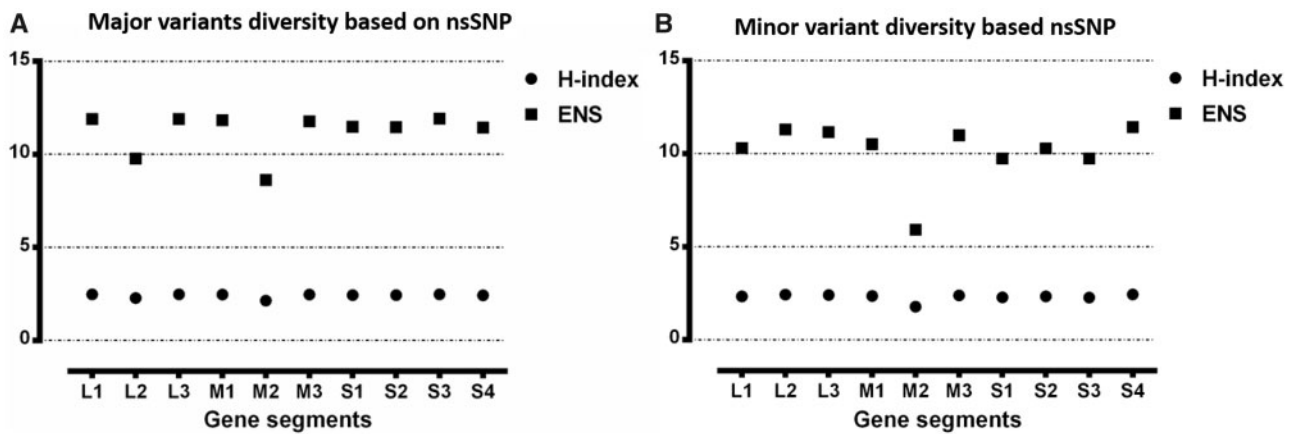

Figure 6. Major variant (A) and minor variant (B) diversity of nsSNP based on Shannon–-Wiener (H-index) diversity indices.

tenosynovitis by only using antigens from one or few cluster groups will be very unlikely. Hence, complete characterization of circulating variant strains must be undertaken to design effective and safe vaccines and other ARV control strategies. So far, only a total of nine ARV whole-genome sequences were reported from California (Egaña-Labrin et al. 2019) and
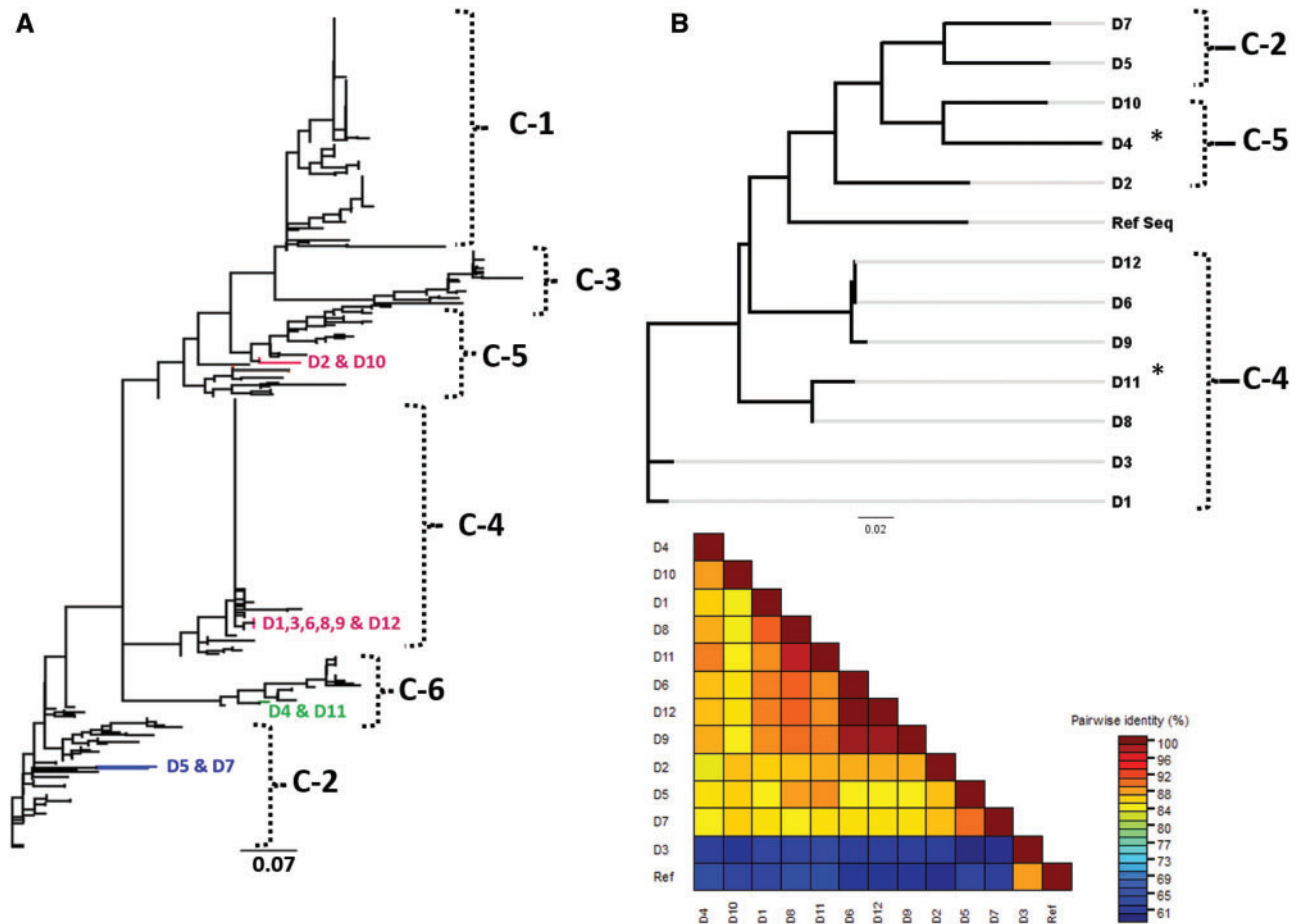
Figure 7. (A) Phylogenetic tree constructed using the N-terminal σC protein coding nucleotide sequence (i.e. nucleotides 45–756) of the twelve ARV isolates including other 500 sequences retrieved from the GenBank. (B) Phylogenetic tree constructed based on the concatenated full-genome sequence of the twelve ARV isolates. The trees were built by neighbour-joining method with Jukes–Canto genetic distance model. The heat map indicates the pair-wise nucleotide identity between the concatenated full-genome sequences. Ref: Reference viral genome sequence.

Pennsylvania (Tang et al. 2015) since the outbreak started in 2011. Here, our study focused on whole-genome NGS and characterization of emerging variant ARV strains chosen based on their virulence in SPF chickens and phylogenetic clustering using σC protein.

Unlike the previous studies (Tang et al. 2015; Egaña-Labrin et al. 2019), our study considered virulence of the virus isolates based on a pathogenicity trial and their phylogenetic clustering based on the amino acid sequence of σC protein (Ayalew et al. 2017), and included a total of twelve viruses selected from each cluster group. The NGS produced deep coverage of all genome segments in all the sequenced virus genomes. The wide sequence divergence observed between the corresponding genome segments indicates the presence of extensive genetic heterogeneity among the ARV isolates circulating in Canada. The nsSNVs and sSNVs were observed in all the regions of all genome segments of the ARV isolates. The exception was genome segment S1 where nsSNV occurred only between nucleotide positions 1–977 as compared with the reference ARV genome. A positive episodic/diversifying selection was also observed in the codon positions of the σC gene using the MEME analysis tool. These findings were important as the S1 segment codes for the structural protein σC, which is essential for virus attachment to the host-cell receptors (Guardado Calvo et al. 2005). The σC protein is also important in inducing the process

of host immune recognition and induction of the production of neutralizing antibodies against ARV (Wickramasinghe et al. 1993). Therefore, this unique phenomenon in genome segment S1 indicates the influence of persuasive diversifying selection of the N-terminus of σC protein as a driver of evolution of ARVs.

Following comparison of nucleotide variability of genome segments, Egaña-Labrin et al. (2019) reported that most of the nsSNVs were observed in S1, M2, and L3 genes. In contrast, we observed the highest proportion of nsSNVs in genome segments S1, M3, and L3. However, the M2 gene segment showed a high degree of conservation between viruses under phylogenetic cluster group C-4 and higher degree of variability between viruses within the other cluster groups. The difference between the findings might be due to the fact that except for one virus isolate, all the viruses included in the Egaña-Labrin et al. (2019) study were only from a single phylogenetic cluster group with no representation from the other phylogenetic cluster groups. Therefore, we can conclude that the extent of genetic divergence of M2 genomic segment varies between viruses within different phylogenetic cluster groups (i.e. conserved in some cluster groups and variable in others).

Our virus isolates and reference sequences from the GenBank were phylogenetically clustered into six genotyping cluster groups based on the σC protein sequence. However, the concatenated whole-genome sequences produced a different
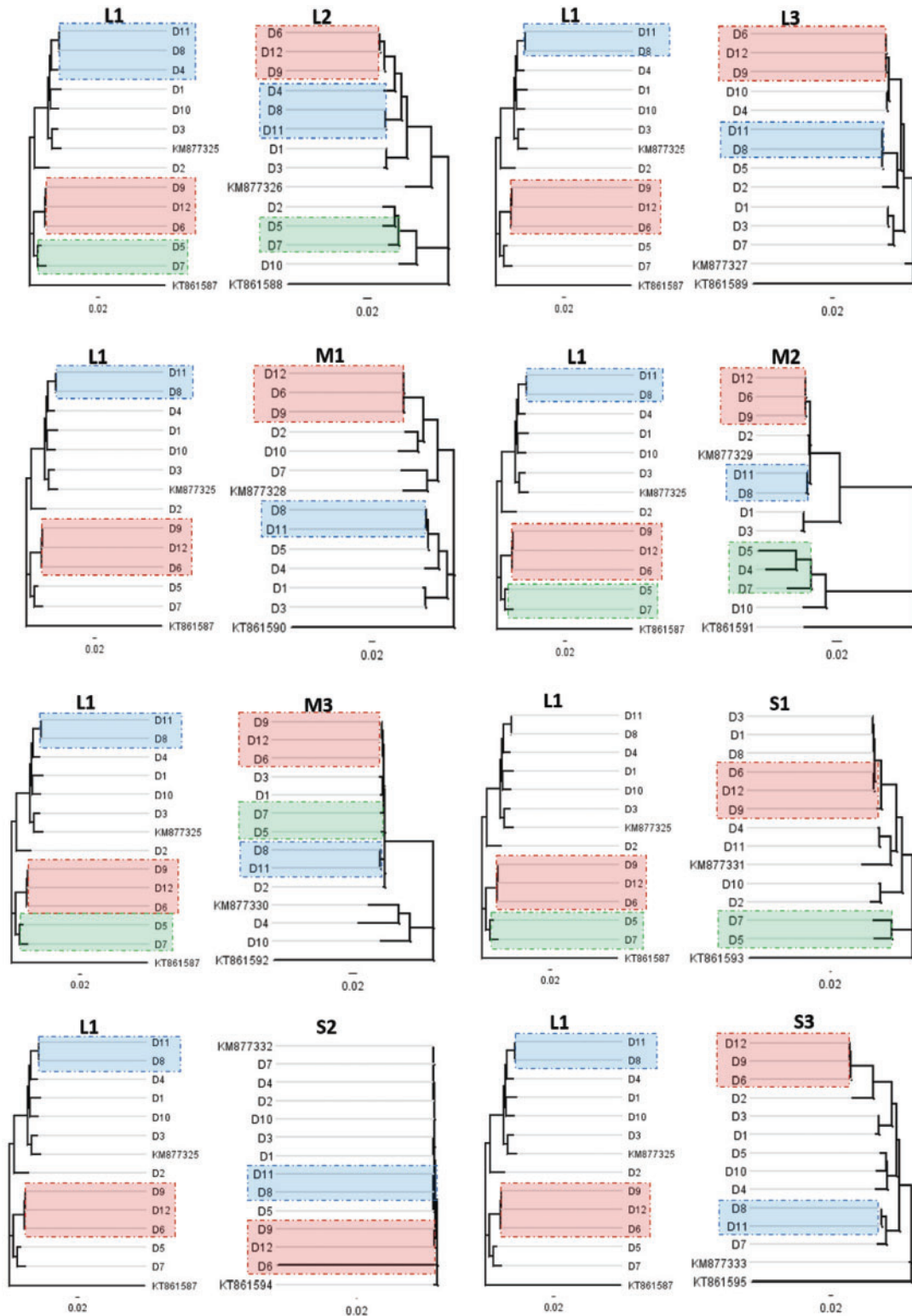
Figure 8. Topologies of segmental phylogenetic trees of the genome of each ARV isolate. The trees were built by neighbour-joining method with Jukes–Cantor genetic distance model.

phylogenetic clustering pattern signifying a phylogenetically in-congruent isolate topology. Evolutionary trees for each genome segment were also distinct with each other. In addition, statistically significant differences were observed in the multiple comparison of group means of nsSNVs and sSNVs between different genome segments. These findings indicate a difference in the polymerase error rate between the various genome segments either within the same class or in different classes, and
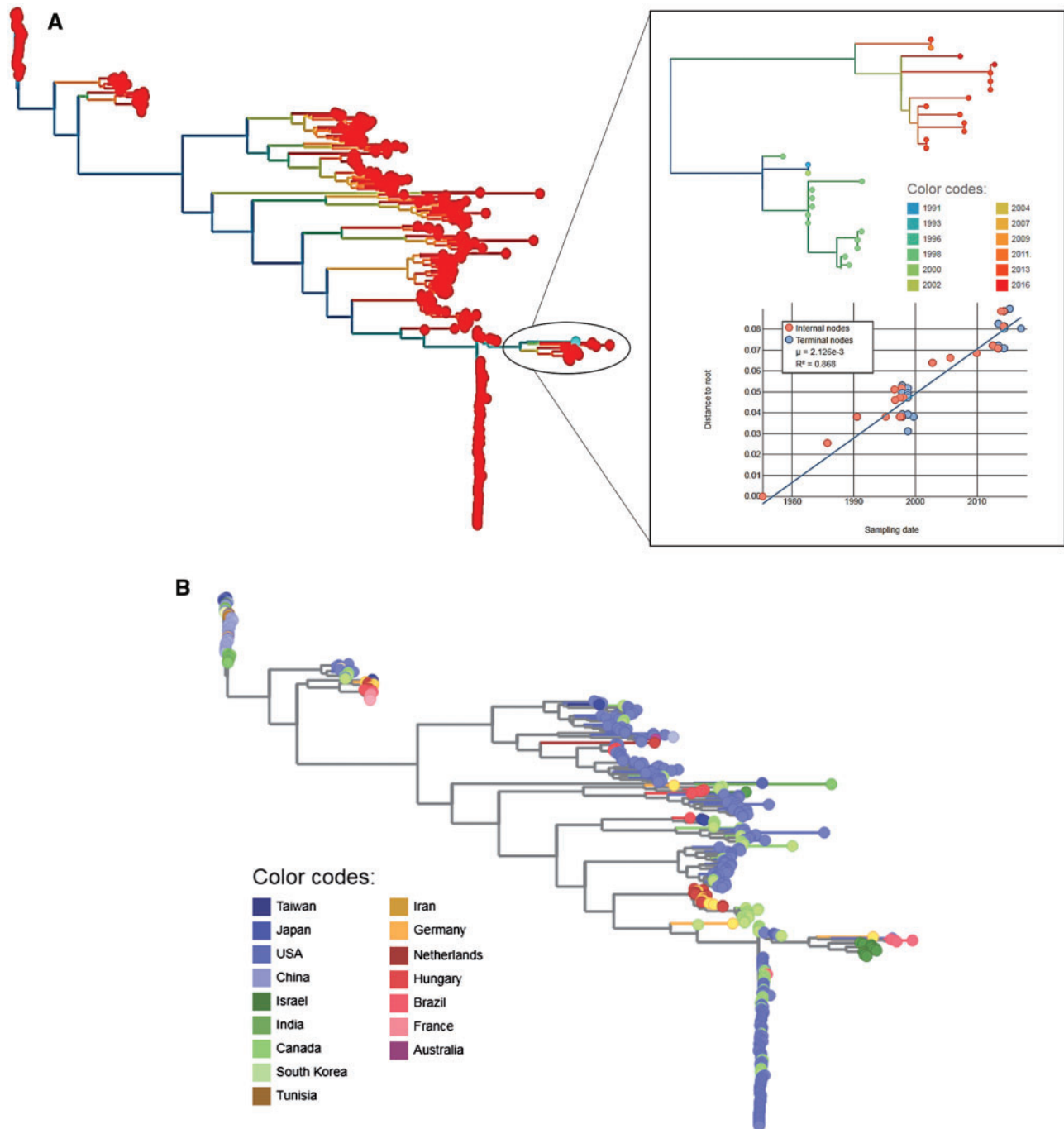
Figure 9. (A) Molecular evolutionary rate of ARV based on the N-terminal σC gene sequence (i.e. nucleotides 45–756) of 471 reovirus isolates retrieved from the GenBank. Our twelve isolates were included in the analysis. For the analysis, the Bayesian inference of MCC tree was performed using the MCMC method with posterior probabilities calculated for the MCMC trees and chain length generating 1,000,000 trees and sampling every 100th tree. (B) Temporally resolved maximum-likelihood phylogenetic tree showing geographic distribution of ARVs based on the σC gene sequence of 471 reovirus isolates retrieved from the GenBank. Our twelve isolates were also included in the analysis. The tree was constructed by considering GTR models and coalescent priors using the algorithm in Tree Time software.

distinct evolution pattern of each gene segment. Similar results were observed by analysis of the S class genes of ARVs (Liu et al. 2003).

One of the important mechanisms of evolution of segmented viruses is by the re-assortment of genome segments (McDonald et al. 2016). In this study, extensive re-assortment events were observed from incongruent topologies in segmental phylogenetic trees and heat map of highest segmental nucleotide identities between the virus isolates. Our results provide evidence that in addition to the accumulation of point mutations, the rapid evolution and genetic diversity of ARVs are principally acquired through genetic re-assortment. This phenomenon might be due to co-circulation of multiple ARV variants from different phylogenetic lineages at the same period in the same geographic location. Such outbreaks involving multiple variant ARVs at the same period have been reported (Lu et al. 2015; Ayalew et al. 2017; Palomino-Tapia et al. 2018; Egaña-Labrin et al. 2019). Therefore, the probability of co-
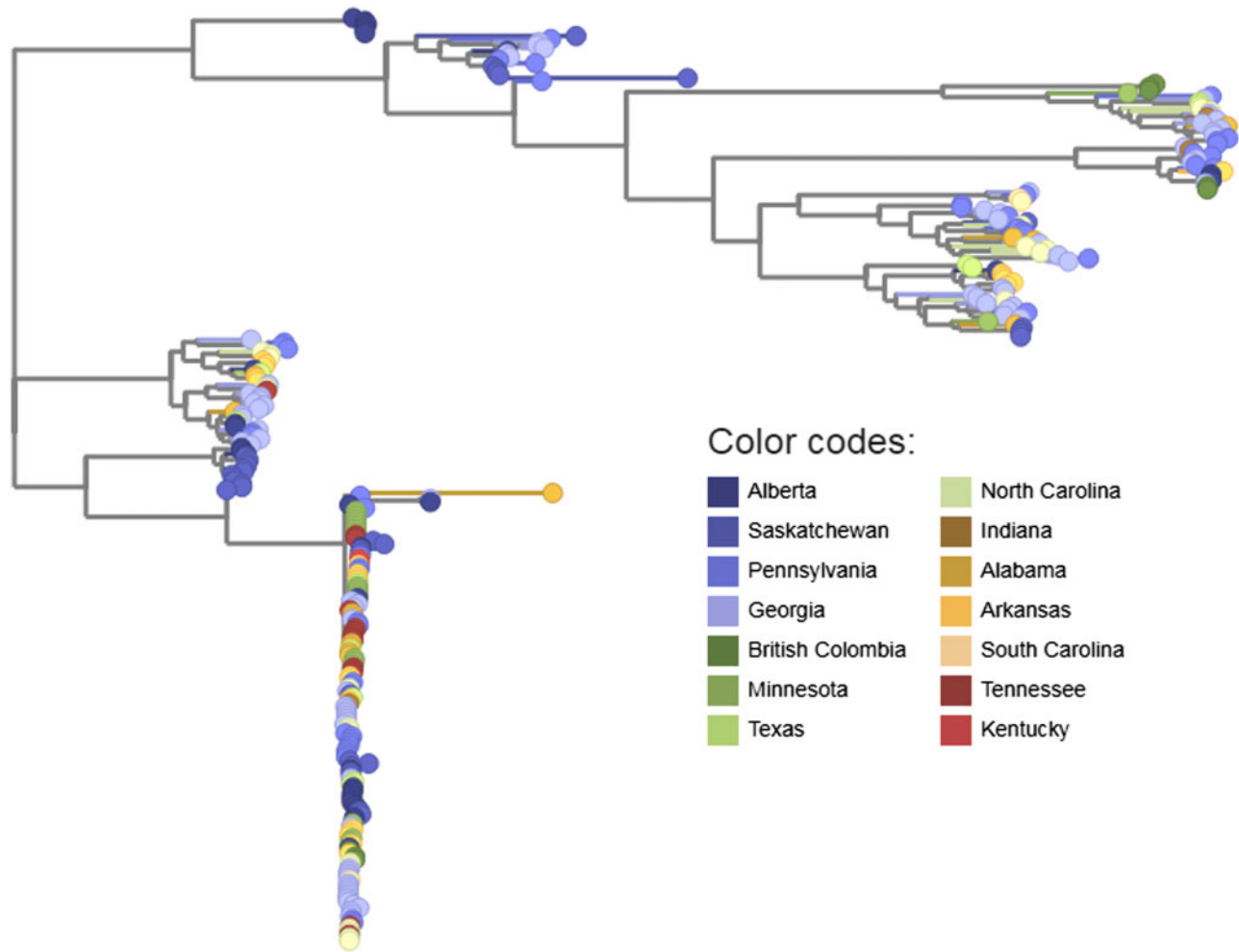
Figure 10. Temporally resolved maximum-likelihood phylogenetic tree showing the geographic distribution of USA and Canadian ARVs.

infection and evolution through exchange of genetic materials between the viruses of different phylogenetic lineages is very high. On the contrary, viruses within C-4 were the most phylogenetically congruent indicating little genomic segmental intermixing. Interestingly, one out of six isolates was a co-infection variant virus in C-4. Similar evidence of naturally occurring co-infection ARV variants has been reported (Tang et al. 2015). However, Egaña-Labrin et al. (2019) found no re-assortment events between genome segments of ARVs isolated in California. This might be because six out of the seven viral genomes sequenced and analysed were from a single cluster group, which would miss important information by excluding the virus isolates from the rest of the cluster groups. In addition, no recombination or re-assortment events were observed between our ARV and TARV sequences. However, a higher percent nucleotide identity was observed between our ARV isolates and Pennsylvania TARV field strain than with the Hungary TARV isolates and the Pennsylvania TARV field strain. Tang et al. (2015) have suggested that the Pennsylvania TARV field strain might have originated through re-assortment of genomic segments between the re-emerging Pennsylvania TARV strains and ARV strains, which explains the closer evolutionary relatedness between our ARV isolates and the Pennsylvania TARV field strain.

In this study, we also determined the molecular evolution rate and pattern of geographic distribution of ARVs for the first

time based on a total of 471 global S1 gene segment sequences retrieved from the GenBank including sequences from our isolates. Except the virus isolates from 1991 to 2016, which had a molecular rate of evolution of 2.3×10^{-3} substitution/site/year, the rest did not produce a good mathematical temporal structure suggesting that reovirus isolates have followed multiple distinct evolutionary events through time. Alternatively, this might be due to inaccurate meta-data information submitted to the GenBank with genome sequences. The results of the combined structured phylogenetic subsets are also good indicators of multiple co-circulating isolates and strains occurring in shorter evolutionary time scale or an outbreak situation in which the collection of sequence data conducted in a very short-time interval. Conversely, the phylogeography results signify widespread global intermixing of ARV variants. Co-clustering of isolates from the USA and Canada also indicate multiple cross-country introductions of variant viruses, which caused past outbreaks. Since ARVs are resistant to disinfection and can survive in the environment for long periods, transfer of variant viruses between the two countries could occur through the importation of eggs, chicken, and other processed chicken products.

In summary, our data demonstrate that there is no association between ARV genotypes and geography. In addition, re-assortment events and accumulation of point mutations are

very important in the evolution of ARVs. The non-specific geographic distribution of all the six genotyping cluster groups of ARV suggests that effective prevention of viral induced arthritis/tenosynovitis cannot be achieved without the inclusion of proper antigens from all the six genotypes in vaccine formulations.

Data availability

All the sequence data are deposited in the GenBank. The GenBank accession numbers are provided in Supplementary Table S1.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Acknowledgements

Funding

Conflict of interest: None declared.

References

Ayalew, L. E. et al. (2017) 'Phenotypic, Genotypic and Antigenic Characterization of Emerging Avian Reoviruses Isolated from Clinical Cases of Arthritis in Broilers in Saskatchewan, Canada', *Scientific Reports*, 7: 3565.

Dandár, E. et al. (2015) 'Genome Sequences of Three Turkey Orthoreovirus Strains Isolated in Hungary', *Genome Announc*, 3: pii: e01333–15.

Egaña-Labrin, S. et al. (2019) 'Genotypic Characterization of Emerging Avian Reovirus Genetic Variants in California', *Scientific Reports*, 9: 9351.

Gouvea, V., and Schnitzer, T. J. (1982) 'Pathogenicity of Avian Reoviruses: Examination of Six Isolates and a Vaccine Strain', *Infection and Immunity*, 38: 731–8.

Guardado Calvo, P. et al. (2005) 'Structure of the Carboxy-Terminal Receptor-Binding Domain of Avian Reovirus Fibre Sigma-C', *Journal of Molecular Biology*, 354: 137–49.

Guindon, S. et al. (2010) 'New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0', *Systematic Biology*, 59: 307–21.

Jost, L. (2006) 'Entropy and Diversity', *Oikos*, 113: 363–75.

Kant, A. et al. (2003) 'Classification of Dutch and German Avian Reoviruses by Sequencing the Sigma C Protein', *Veterinary Research*, 34: 203–12.

Kort, Y. H. et al. (2013) 'Molecular Characterization of Avian Reovirus Isolates in Tunisia', *Virology Journal*, 10: 12.

Labrada, L. et al. (2002) 'Avian Reoviruses Cause Apoptosis in Cultured Cells: Viral Uncoating, but Not Viral Gene Expression, Is Required for Apoptosis Induction', *Journal of Virology*, 76: 7932–41.

Liu, H. J. et al. (2003) 'Molecular Evolution of Avian Reovirus: Evidence for Genetic Diversity and Reassortment of the S-Class Genome Segments and Multiple Cocirculating Lineages', *Virology*, 314: 336–49.

Lu, H. et al. (2015) 'Isolation and Molecular Characterization of Newly Emerging Avian Reovirus Variants and Novel Strains in Pennsylvania, USA, 2011-2014', *Scientific Reports*, 5: 14727.

Martin, D. P. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: vev003.

McDonald, S. M. et al. (2016) 'Reassortment in Segmented RNA Viruses: Mechanisms and Outcomes', *Nature Reviews Microbiology*, 14: 448–60.

Murrell, B. et al. (2012) 'Detecting Individual Sites Subject to Episodic Diversifying Selection', *PLoS Genetics*, 8: e1002764–10.

Palomino-Tapia, V. et al. (2018) 'Molecular Characterization of Emerging Avian Reovirus Variants Isolated from Viral Arthritis Cases in Western Canada 2012-2017 Based on Partial Sigma (σ)C Gene', *Virology*, 522: 138–46.

Rosenberger, J. K. et al. 2013. 'Pathogenicity and Control of Recent (2011–2012) Reovirus Isolates from Broiler and Turkey Flocks Presenting with Viral Arthritis and Tenosynovitis'. In *Proc. 2013 American Association of Avian Pathologist Annual Convention: AAAP Scientific Program*, Chicago, IL. Jacksonville, FL: AAAP (2013, July 23).

Sellers, H. S. (2017) 'Current Limitations in Control of Viral Arthritis and Tenosynovitis Caused by Avian Reoviruses in Commercial Poultry', *Veterinary Microbiology*, 206: 152–6.

Shannon, C. E., and Weaver, W. (1949) *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Spandidos, D. A., and Graham, A. F. (1976) 'Physical and Chemical Characterization of an Avian Reovirus', *Journal of Virology*, 19: 968–76.

Tang, Y. et al. (2015) 'Genomic Characterization of a Turkey Reovirus Field Strain by Next-Generation Sequencing', *Infection, Genetics and Evolution*, 32: 313–21.

—— et al. (2016) 'Detection and Characterization of Two Co-Infection Variant Strains of Avian Orthoreovirus (ARV) in Young Layer Chickens Using Next-Generation Sequencing (NGS)', *Scientific Reports*, 6: 24519.

Wickramasinghe, R. et al. (1993) 'Avian Reovirus Proteins Associated with Neutralization of Virus Infectivity', *Virology*, 194: 688–96.