Article

Genetic Analysis of West Nile Virus Isolates from an Outbreak in Idaho, United States, 2006–2007

Andriyan Grinev [1,*]**, Caren Chancey** [1]**, Germán Añez** [1]**, Christopher Ball** [2]**, Valerie Winkelman** [3]**, Phillip Williamson** [3]**, Gregory A. Foster** [4]**, Susan L. Stramer** [4] **and Maria Rios** [1,*]

[1] Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD 20892, USA; E-Mails: caren.chancey@fda.hhs.gov (C.C.); german.anez@fda.hhs.gov (G.A.)

[2] Idaho Bureau of Laboratories, Boise, ID 83712, USA; E-Mail: ballc1@dhw.idaho.gov

[3] Creative Testing Solutions, Tempe, AZ 85282, USA; E-Mails: VWinkelman@mycts.org (V.W.); PWilliamson@mycts.org (P.W.)

[4] American Red Cross, Gaithersburg, MD 20877, USA; E-Mails: Gregory.Foster@redcross.org (G.A.F.); Susan.Stramer@redcross.org (S.L.S.)

* Authors to whom correspondence should be addressed; E-Mails: andriyan.grinev@fda.hhs.gov (A.G.); maria.rios@fda.hhs.gov (M.R.); Tel.: +1-301-451-2510 (A.G.); Fax: +1-301-480-7928 (A.G.); Tel.: +1-301-435-4585 (M.R.); Fax: +1-301-480-7928 (M.R.).

Abstract: West Nile virus (WNV) appeared in the U.S. in 1999 and has since become endemic, with yearly summer epidemics causing tens of thousands of cases of serious disease over the past 14 years. Analysis of WNV strains isolated during the 2006–2007 epidemic seasons demonstrates that a new genetic variant had emerged coincidentally with an intense outbreak in Idaho during 2006. The isolates belonging to the new variant carry a 13 nt deletion, termed ID-Δ13, located at the variable region of the 3′UTR, and are genetically related. The analysis of deletions and insertions in the 3′UTR of two major lineages of WNV revealed the presence of conserved repeats and two indel motifs in the variable region of the 3′UTR. One human and two bird isolates from the Idaho 2006–2007 outbreaks were sequenced using Illumina technology and within-host variability was analyzed. Continued monitoring of new genetic variants is important for public health as WNV continues to evolve.

1. Introduction

West Nile virus (WNV; family *Flaviviridae*, genus *Flavivirus*) is a mosquito-borne virus that is maintained in a bird-mosquito enzootic cycle, with occasional infections of humans, horses and other animals [1]. The WNV positive RNA genome is about 11 kb in length, containing a single open reading frame (ORF) encoding one polyprotein processed into three structural and seven nonstructural viral proteins by cellular and viral proteases. It is flanked by 5′- and 3′-untranslated regions (UTR) which are involved in translation and viral RNA replication, and play an important role in genome packaging [2]. Both the 5′ and 3′UTR in the WNV genome form highly conserved secondary and tertiary structures, some elements of which are similar among mosquito-borne flaviviruses. Different functional regions have been described inside both the 5′ and 3′UTR of flaviviruses based on factors such as nucleotide content, degree of sequence conservation, occurrence of repeated sequence motifs, and predicted secondary structure [2–4].

Since its recognition in New York City in 1999, WNV has spread throughout the United States (U.S.) and the Americas, including Canada, Mexico, the Caribbean, Central America and South America [5]. West Nile virus is now one of the most widely distributed arboviruses in the world [1,6]. Most human infections (~80%) are asymptomatic, and symptomatic infections vary from mild influenza-like illness to fatal neuroinvasive disease (~1%). The virus can be transmitted from asymptomatic viremic donors to recipients by organ transplant and by transfusion of infected blood and blood products, affecting the safety of the blood supply [7]. WNV is estimated to have infected ~4 million humans in the U.S. between 1999 and 2012, causing over 37,000 serious illnesses, including more than 16,000 neuroinvasive disease cases with 1,549 deaths reported to the Centers for Disease Control and Prevention (CDC). The first occurrence of WNV in Idaho was reported in 2003, with a single case of WNV fever and no fatalities reported. In the following years 2004 and 2005, there were respectively three and 13 human cases, including one and three cases of neuroinvasive disease, with no fatalities. In contrast, the 2006 outbreak was intense with 969 cases of disease reported to the CDC, including 115 neuroinvasive cases and 21 fatalities. The 2007 outbreak resulted in 132 cases including 10 neuroinvasive disease cases and one fatality [8]. Total numbers of human and non-human WNV infections detected in Idaho, 2005–2012, based on U.S. Geological Survey (USGS) [9] are presented in Figure 1.

Isolates of WNV fall into up to five distinct lineages based on phylogenetic analysis, which correlate well with the geographical point of isolation [10–13]. Clade 1a of lineage I contains isolates from Africa, Europe, the Middle East, Asia, and the Americas and includes all isolates from the U.S. Phylogenetic analyses have demonstrated that WNV strains from this clade have an African origin [14]. In 2001 a new genotype, WN02, emerged in the U.S., becoming increasingly prevalent in 2002, and eventually displacing the ancestor genotype NY99 [15,16].

Figure 1. WNV infections detected in Idaho, 2005–2012. Y axis represents the number of infections in logarithmic scale and the X-axis represents years.
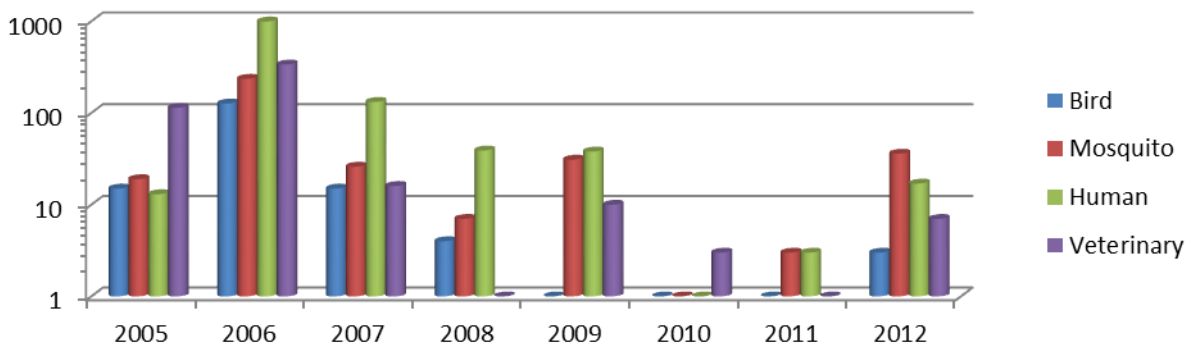
The WN02 genotype is characterized by 13 conserved silent nucleotide mutations and one amino acid substitution in the envelope protein, E-V$_{159}$A, when compared with the original U.S. strain, NY99 (AF196835). This genotype became dominant in the Americas probably because of its ability to disseminate more efficiently in domestic mosquitoes as compared to the initial NY99 genotype which is believed to have come to the New World from the Middle East [17,18]. A third genotype, termed SW/WN03, is co-circulating with WN02. This genotype is characterized by two additional fixed amino acid substitutions, NS4A-A$_{85}$T and NS5-K$_{314}$R [19]. We performed a comprehensive phylogenetic analysis of U.S. WNV isolates from every annual epidemic between 1999 and 2011. We have found that some WNV isolates from Idaho and North Dakota collected during the 2006–2007 outbreaks form a separate cluster within genotype SW/WN03, termed MW/WN06 [20]. Analysis of MW/WN06 isolates demonstrated that they are genetically related and carry a deletion of 13 nucleotides (10,415–10,427), termed ID-Δ13, located in the variable region of the 3′UTR.

In flaviviruses, genetic variation occurs within the linear evolutionary pathway via single base mutations and small insertions and deletions, and also infrequently by recombination [21]. WNV, like other RNA viruses, exists within each individual host as a mixture of viral particles with diverse genomes, which are also known as viral quasispecies. The genetically heterogeneous structure of viral populations allows them to promptly adapt to varying replicative environments and new hosts by selecting preexisting genetic variants with better fitness [22,23]. Minor genomes present in the viral quasispecies spectrum from the particular host in small proportion might be not recognized when a genetic study is performed using classical sequencing techniques, which normally only provides information about the major genome. However, selection acts on the entire viral swarm rather than on the individual fittest sequence. Therefore, minor genomes can potentially play an important role in adaptation to new hosts and in the virulence of circulating strains and need to be investigated.

In the past decade, a number of next generation sequencing (NGS) technologies, which are also referred to as massively parallel sequencing, high-throughput or ultra-deep sequencing, have been established. They allow determination of the primary structure of each nucleic acid molecule present in the starting material by generation of extremely large amounts of sequence information from separated individual molecules of nucleic acids [24,25]. These new methods and tools are suitable for analysis of viral quasispecies and have been used to study viruses such as SARS [26], HIV [27], hepatitis B [28], influenza [29], foot-and-mouth disease virus [30], hepatitis C and others [31]. In this article we report

detailed analyses of WNV genetic variability from one human and two bird isolates from the Idaho 2006–2007 outbreaks using data from Illumina paired-end sequencing technology.

The reoccurrence of WNV outbreaks in the New World may be associated with viral adaptation through fixation of spontaneous mutations. The new genetic variant of WNV that appeared in Idaho in 2006 demonstrates the continuing evolution of the virus in North America. The emergence of new genetic variants of WNV raises issues of public health importance because emerging variants may affect the sensitivity of both screening and diagnostic assays, as well as the development of vaccines and drugs.

2. Experimental Section

2.1. Study Sample

This study included sequence analysis of the 3′UTR from 51 WNV isolates produced from human plasma specimens derived from blood donors who tested positive for viral RNA by FDA-approved nucleic acid tests (NAT) used to screen blood donations. Prior to use in our study, all human specimens were anonymized. The studied specimens were collected in the 2006–2007 epidemic seasons from the states of Colorado (CO), Idaho (ID), Illinois (IL), Kansas (KS), Maryland (MD), Minnesota (MN), Nebraska (NE), New York (NY), North Dakota (ND), South Dakota (SD), Texas (TX), and Utah (UT) under IRB-approved informed consent. In addition, one isolate from a mosquito pool and 5 avian specimens collected in Idaho in 2007 were analyzed (Table 1). The studied dataset for analysis of the variable region of the 3′UTR include the sequences of 387 WNV isolates from lineage 1 and 11 isolates from lineage 2 obtained from the GenBank database (Supplemental Table S1).

Table 1. List of WNV isolates and presence of ID-Δ13 detected by RT-PCR. Y = ID-Δ13 deletion present, N = ID-Δ13 deletion absent.

#	Isolate	Host	Year of isolation	State	GenBank accession no.	ID-Δ13
1	ARC1-06	Human	2006	ID	N/A	N
2	ARC10-06	Human	2006	ID	JF957161	Y
3	ARC104-06	Human	2006	ID	N/A	Y
4	ARC105-06	Human	2006	ID	N/A	Y
5	ARC106-06	Human	2006	ID	N/A	Y
6	ARC108-06	Human	2006	ID	N/A	Y
7	ARC112-06	Human	2006	ID	N/A	Y
8	ARC13-06	Human	2006	ID	JF957162	Y
9	ARC17-06	Human	2006	ID	JF957163	Y
10	ARC19-06	Human	2006	ID	N/A	N
11	ARC22-06	Human	2006	ID	N/A	N
12	ARC23-06	Human	2006	ID	JF957164	N
13	ARC26-06	Human	2006	ID	N/A	N
14	ARC27-06	Human	2006	ID	JF957165	Y
15	ARC28-06	Human	2006	ID	N/A	N
16	ARC30-06	Human	2006	ID	N/A	N
17	ARC31-06	Human	2006	ID	N/A	N

Table 1. *Cont.*

| #	Isolate	Host	Year of isolation	State	GenBank accession no.	ID-Δ13
18	ARC32-06	Human	2006	ID	N/A	N
19	ARC41-06	Human	2006	ID	N/A	N
20	ARC42-06	Human	2006	ID	N/A	Y
21	ARC57-06	Human	2006	ID	N/A	N
22	ARC60-06	Human	2006	ID	N/A	N
23	ARC61-06	Human	2006	ID	N/A	Y
24	ARC-Z-06	Human	2006	ID	N/A	N
25	ARC140-07	Human	2007	ID	JF957168	Y
26	ID7mq-07	Mosquito	2007	ID	N/A	Y
27	ID19bd-07	Avian	2007	ID	N/A	Y
28	ID20bd-07	Avian	2007	ID	N/A	Y
29	ID21bd-07	Avian	2007	ID	JF957171	Y
30	ID28bd-07	Avian	2007	ID	JF957172	Y
31	ID29bd-07	Avian	2007	ID	N/A	Y
32	ARC135-06	Human	2006	IL	N/A	N
33	ARC126-06	Human	2006	KS	N/A	N
34	ARC134-06	Human	2006	MD	N/A	N
35	ARC-W-06	Human	2006	MN	N/A	N
36	ARC-Y-06	Human	2006	MN	N/A	N
37	ARC127-06	Human	2006	NE	N/A	N
38	ARC128-06	Human	2006	NE	N/A	N
39	ARC131-06	Human	2006	NE	N/A	N
40	ARC2-06	Human	2006	NE	N/A	N
41	ARC3-06	Human	2006	NE	N/A	N
42	ARC38-06	Human	2006	NE	N/A	N
43	ARC125-06	Human	2006	NY	N/A	N
44	ARC132-06	Human	2006	SD	N/A	N
45	ARC133-06	Human	2006	TX	N/A	N
46	ARC102-06	Human	2006	UT	N/A	N
47	ARC11-06	Human	2006	UT	N/A	N
48	ARC25-06	Human	2006	UT	N/A	N
49	ARC33-06	Human	2006	UT	JF957166	N
50	BSL103-06	Human	2006	SD	N/A	N
51	BSL106-06	Human	2006	ND	JF957167	Y
52	BSL107-06	Human	2006	ND	N/A	N
53	BSL110-06	Human	2006	SD	N/A	N
54	CO2-07	Human	2007	CO	N/A	N
55	CO3-07	Human	2007	CO	N/A	N
56	CO4-07	Human	2007	CO	JF957169	N
57	CO5-07	Human	2007	CO	JF957170	N

2.2. Virus Isolation

Viral isolation was performed in Vero cells (ATCC # CCL-81) as described [32]. Supernatants were harvested when extensive cytopathic effect was observed, centrifuged to remove cell debris and frozen at −80 °C until further analysis. All specimens were subjected to a single passage in Vero cells.

2.3. RNA Extraction and Polymerase Chain Reaction (PCR)

RNA samples from viral passages were isolated from 140 μL of culture supernatants using the QIAamp Viral RNA Mini extraction kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. Total RNA extracts from plasma samples were obtained from 1 mL using Trizol reagent (Invitrogen, Carlsbad, CA, USA). RNA extracts were stored at −80 °C until further analysis. Reverse transcription reactions and PCR amplification were performed for complete genome sequencing as described previously [32]. Briefly, reverse transcription reactions were performed using SuperScript III (Invitrogen) according to the manufacturer's instructions. PCR reactions of the cDNA specimens were performed using the Hi-Fidelity PCR system (Invitrogen) according to the manufacturer's instructions.

Figure 2. (**A**) Scheme of multiplex RT-PCR assay; (**B**) 2% agarose gel stained by ethidium bromide. NC–negative control, no viral RNA added in PCR reaction. C1 and C2–positive controls for F10400-R10630 and F10100-R10630 primers pairs respectively. L–1 kb ladder (Invitrogen). N–no deletion. D–deletion.
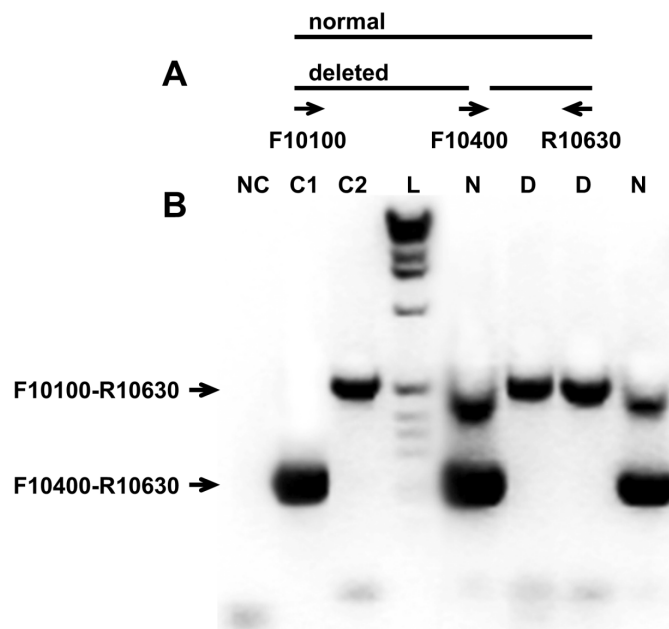
2.4. Multiplex PCR

Multiplex PCR reactions for detection of deletions in 3′UTR (Figure 2) were performed using the One-Step RT-PCR kit (Qiagen, Valencia, CA, USA) and F10100 5′-TCCATGCAGGAG GAGAGTGGATGAC; F10400 5′-CTGTAGATATTTAATCAATTGTAAATAGACAA; R10630 5′-GGGTCCTCCTTCCGAGACGGT primers to amplify fragments that allow visualization of the ID-Δ13 deletion. Amplification was performed under the following cycling conditions: cDNA

synthesis at 50 °C, 30 min; denaturation at 94 °C, 15 min; PCR 30 cycles: 94 °C for 15 s, 55 °C for 30 s and 72 °C for 1 min; final extension 72 °C for 5 min. After amplification, 10 microliters of each reaction mixture were analyzed using 2% agarose gel in TAE buffer.

2.5. DNA Sequencing, Assembly and Analysis

PCR products were purified after agarose gel electrophoresis using the MinElute Gel Extraction Kit (Qiagen) according to the manufacturer's protocol, and both strands were subjected to direct sequencing using the amplifying primers and additional internal sequence primers. For validation of multiplex PCR results, sequencing reactions were performed using the F10100 and R10630 primers as described [32]. Nucleotide sequences from each isolate were aligned using the Align X program from Vector NTI (Invitrogen) and compared to the prototype NY99 (AF196835) and to previously published sequences of isolates from different regions of the U.S. and from other countries.

2.6. Illumina NGS

Next generation sequencing was performed on the HiSeq 2000 platform using standard Illumina kits and protocols. We used paired-end technology to sequence both strands. Briefly, DNA samples of isolates ARC13-06, ID21bd-07 and ID28bd-07 were prepared as a mixture of the same overlapping amplicons covering the entire genomes as we used for Sanger sequencing and were then separately sheared into 200–400 bp fragments using Covaris adaptive focused acoustics process (Covaris, Woburn, MA, USA). Libraries were produced and amplified using the Paired-End DNA Sample Preparation kit according to the manufacturer's protocols. The ends of the fragments were repaired and an A-overhang was added. Adapters specific to each sample were A-T ligated onto the ends of the fragments. Clusters were prepared using the TruSeq PE Cluster kit. All three samples were sequenced using one lane of the flow cell generating 100 bp paired-end reads.

2.7. NGS Data Analysis

Preliminary NGS data analysis, conversion and trimming of read sets were performed using Illumina software CASAVA v1.8 and CLC Genomics Workbench v4. Assembly, and aligning, base calls, SNP calls, indel calls and read counts were performed using the High-performance Integrated Virtual Environment (HIVE). HIVE is a cloud-based massively parallel computational environment optimized for the storage and analysis of NGS data [33].

3. Results and Discussion

3.1. Nucleotide Mutations and Amino Acid Substitutions Identified by Sanger Sequencing

Twelve isolates from the 2006–2007 outbreaks were completely sequenced and demonstrated nucleotide divergence in the range of 0.35–0.44% compared to NY99 (AF196835). Most of the nucleotide changes were silent transitions (U↔C, A↔G) and 8 isolates shared the silent transversion $A_{7209}T$ in NS4B. Nucleotide mutations conserved in studied isolates compared to the complete genome of NY99 are shown in Table 2.

Table 2. Nucleotide mutations in studied isolates compared to the complete genome of NY99 (AF196835). Non-silent mutations are shown in bold and resulted in amino acid substitutions E-V$_{159}$A, NS4A-A$_{85}$T and NS5-K$_{314}$R.

Gene	prM	E				NS1			NS2A		NS2B	NS3				NS4A	
nt #	660	1320	1442	1974	2466	2661	3228	3399	3927	4146	4255	4803	6138	6238	6426	6721	6765
NY99	C	A	**T**	C	C	G	T	T	T	A	C	C	C	C	C	**G**	T
ARC10-06	T	G	**C**	T	T	A	C	C	C	G	T	T	T	T	T	**A**	C
ARC13-06	T	G	**C**	T	T	A	C	C	C	G	T	T	T	T	T	**A**	C
ARC17-06	T	G	**C**	T	T	A	C	C	C	G	T	T	T	T	T	**A**	C
ARC23-06	T	G	**C**	T	T			C		G			T	T	T	**A**	C
ARC27-06	T	G	**C**	T	T	A	C	C	C	G	T	T	T	T	T	**A**	C
ARC33-06	T	G	**C**	T	T			C		G			T	T	T	**A**	C
BSL106-06	T	G	**C**	T	T	A	C	C	C	G	T	T	T	T	T	**A**	C
ARC140-07	T	G	**C**	T	T	A	C	C	C	G	T	T	T	T	T	**A**	C
ID21bd-07	T	G	**C**	T	T	A	C	C	C	G	T	T	T	T	T	**A**	C
ID28bd-07	T	G	**C**	T	T	A	C	C	C	G	T	T	T	T	T	**A**	C
CO4-07	T		**C**		T					G			T	T	T		
CO5-07	T	G	**C**	T	T			C		G		T	T	T	T	**A**	C

Gene	NS4B						NS5								3′UTR	
nt #	6936	6996	7015	7209	7245	7269	7938	8550	8621	8811	9264	9352	9660	10062	Δ13	10851
NY99	T	C	T	A	T	T	T	C	**A**	T	T	C	C	T	N	A
ARC10-06	C	T	C	T	C	C	C	T	**G**	C	C	T	T	C	Y	G
ARC13-06	C	T	C	T	C	C	C	T	**G**	C	C	T	T	C	Y	G
ARC17-06	C	T	C	T	C	C	C	T	**G**	C	C	T	T	C	Y	G
ARC23-06	C	T	C			C	C	T	**G**	C	C	T	T	C	N	G
ARC27-06	C	T	C	T	C	C	C	T	**G**	C	C	T	T	C	Y	G
ARC33-06	C	T	C			C	C	T	**G**	C	C	T	T	C	N	G
BSL106-06	C	T	C	T	C	C	C	T	**G**	C	C	T	T	C	Y	G
ARC140-07	C	T	C	T	C	C	C	T	**G**	C	C	T	T	C	Y	G
ID21bd-07	C	T	C	T	C	C	C	T	**G**	C	C	T	T	C	Y	
ID28bd-07	C	T	C	T	C	C	C	T	**G**	C	C	T	T	C	Y	G
CO4-07		T	C				C				C		T		N	G
CO5-07	C	T	C			C	C	T	**G**	C	C	T	T	C	N	G

All isolates from this study with complete viral sequences shared 12 nucleotide mutations, including one non-silent mutation in the envelope gene E-T$_{1442}$C. The number of deduced amino acid substitutions ranged from 3 to 8, most of which were conservative substitutions. In addition to the E-V$_{159}$A amino acid substitution in the envelope protein common to all WN02 genotype viruses, all isolates except CO4-07 shared two substitutions common to the SW/WN03 genotype, NS4A-A$_{85}$T and NS5-K$_{314}$R.

3.2. Analysis of the Variable Region of the 3′UTR

We previously reported the first identified deletion and insertion in the 3′UTR of WNV [32]. Further investigation of 3′UTR variability revealed a 13 nt deletion (10,415–10,427), named ID-Δ13, in isolates from the 2006 Idaho outbreak, in which 996 WNV symptomatic cases and 21 deaths were reported to the CDC. To investigate the penetration of the ID-Δ13 genetic variant, we designed a multiplex PCR (Figure 2) and screened 57 isolates obtained from human, bird, and mosquito specimens collected during the 2006 and 2007 epidemics. The ID-Δ13 deletion was originally identified in isolates obtained from a single passage in Vero cells. Subsequently, we tested plasma samples directly by isolating total RNA from 1 mL of plasma samples using Trizol followed by multiplex one-step PCR. Due to low plasma viral loads, only four plasma samples (ARC1-06, ARC13-06, ARC106-06 and ARC108-06) yielded sufficient RNA for downstream analysis by multiplex one-step PCR, which showed that the ID-Δ13 deletion was present in all samples except ARC1-06. The amplicons covering the region of the 3′UTR including the ID-Δ13 deletion (F10100-R10630) from these four samples were then sequenced by the Sanger method to confirm the presence of the deletion in viral RNA isolated from original plasma. There were no differences between the sequences obtained from the original plasma sample and the sequences obtained after one passage in Vero cells.

The results for the presence of ID-Δ13 deletions in the 3′UTR are shown in Table 1. All samples that had amplicon patterns corresponding to deletions were sequenced to confirm the PCR results (data not shown). Out of 31 specimens (25 humans, one mosquito, and five birds) from Idaho from 2006–2007, 18 (58%) had the ID-Δ13 deletion. The ID-Δ13 deletion was also found in one human isolate from North Dakota in 2006, but not in any isolate from other states included in this study. In addition, a deletion of 3 nt (10,499–10,501) was identified in a Colorado isolate, CO4-07. These deletions and insertion are placed within ~100 bp of the variable region of the 3′UTR located downstream of the stop codon.

We used the set of 398 WNV sequences obtained from the GenBank (Supplemental Table S1) to study the variable region of the 3′UTR. The partial nucleotide alignment of the 3′UTR of two major lineages of WNV is shown in Figure 3. We observed insertions and deletions within the WNV 3′UTR occurring in the vicinity of two sequence motifs, GTAAGT and YYYTR (Y = C/T; R = A/G), which have been previously described as indel (insertion/deletion) motifs in the human genome [34]. The direct repeat (DR) GTAAATA $(N)_{0-29}$GT appeared once at position 10,435 in lineage 1c, and twice in lineages 1a, 1b, and 2: at positions 10,413 and 10,476 in lineages 1a and 1b; and 10,434 and 10,518 in lineage 2. These DRs are flanked by the sequences GTAA and GT, which in the case of deletion of the internal part of the DR, are able to form the GTAAGT sequence. Another indel motif, YYYTR (Y = C/T; R = A/G), is located at positions 10,449 and 10,495 in lineage 1a; at positions 10,449 and 10,487 in lineage 1b; 10,481 in lineage 1c; and 10,514 and 10,558 in lineage 2. A variety of deletions and insertions in the 3′UTR of the WNV genome occurring within or near these indel motifs are illustrated in Figure 3. Analysis of this portion of the 3′UTR revealed ATTTA pentamers and complementary TAAAT pentamers located within the direct repeats described above. These pentamers are known as classic RNA instability determinants: adenosine and uridine-rich elements (ARE) found in the 3′UTR of short-lived cellular mRNAs [35].

Figure 3. Partial nucleotide alignment of the 3′UTR of two major lineages of WNV isolated from the U.S. and other countries. GTAAATA(N)$_{0-29}$GT repeats underlined by green lines. GTAAGT and YYYTR (Y = C/T; R = A/G) indel motifs are highlighted in green and blue colors respectively. Deletions are shown in red and insertions in yellow. AREs (ATTTA and TAAAT) are underlined. Isolates JF957161–JF957171 contained the ID-Δ13 deletion.
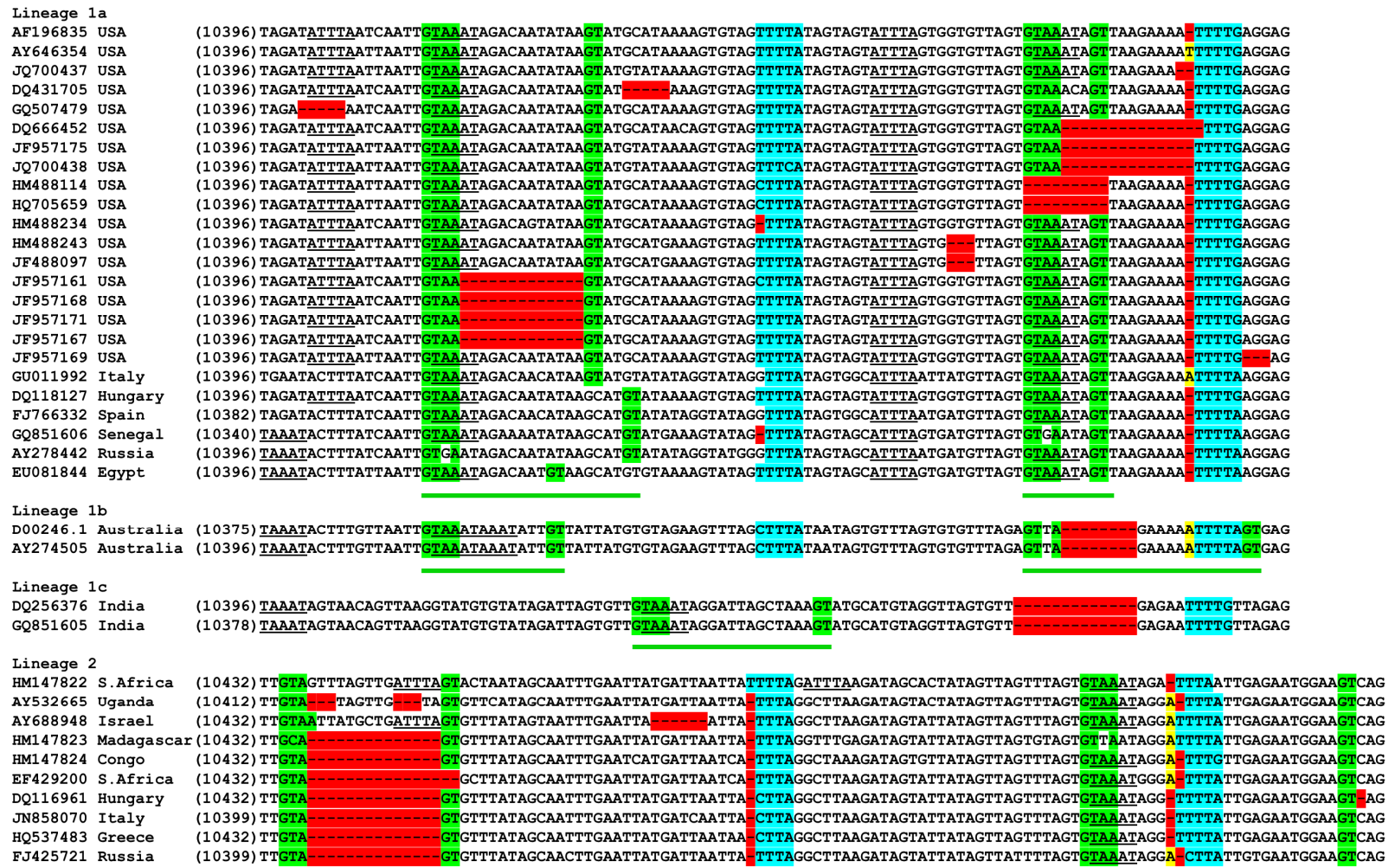
3.3. Illumina NGS Data Analysis

The overlapping amplicons covering the entire genomes of ARC13-06, ID21bd-07 and ID28bd-07, which we used for common Sanger sequencing, were re-sequenced using Illumina NGS technology. NGS generated more than 80 million 100 nt reads per genome. Base and paired-end read counts estimated using the Illumina pipeline software CASAVA v1.8 are shown in Supplemental Table S2. Average quality scores per cycle ranged from 31.4–39.7. Since Phred quality scores are logarithmically linked to error probabilities, base call accuracy was in the range 99.9–99.99%.

Base calling was performed using the sequences of each isolate identified by the common Sanger method for assembly and alignment. Approximately 1 million reads from each isolate were unaligned. Table 3 shows that there was no significant difference in the level of background sequence heterogeneity detected in the studied isolates. The average frequency of insertions and deletions ranged from 0.54–1.29% and per nucleotide variability varied by 0.82–0.89%. We found significantly more (>2 times) transitions than transversions among variable nucleotides. Substantial prevalence of transitions above transversions probably reflects the inherent bias of nucleotide misincorporation common to all known polymerases.

Table 3. Frequencies of mutations identified by HIVE for Illumina NGS data.

|		Mutation frequencies			
Isolate	**Read Count, $\times 10^6$**	**Total nucleotides read, $\times 10^9$**	**Substitutions, %**	**Insertions, %**	**deletions, %**
ARC13-06	80.18	8.02	0.89	0.64	1.29
ID21bd-07	91.23	9.13	0.86	0.78	1.09
ID28bd-07	83.32	8.27	0.82	0.54	1.27

The reads from each isolate were aligned by HIVE Hexagon using NY99 (AF196835) as the reference genome to perform SNP and indel calls. SNP calling results are shown in Supplemental Table S3. Illumina paired-end sequencing detected 70 mutations for isolate ARC13-06; only 46 of those, including the deletion ID-Δ13, were detectable using common Sanger sequencing. For the bird isolates ID21bd-07 and ID28bd-07, the proportions of mutations detected by NGS to mutations detected by common sequencing are 73 to 48 and 71 to 47 respectively. In the region spanning nt 10,414–10,429 where common sequencing recognized ID-Δ13, we found a broad assortment of reads which contained deletions of different sizes (1–14 nt) and some insertions (1–6 nt). Intact genomes are also present in this region in low frequencies ranging from 0.03–0.07.

All three isolates studied using NGS shared nine mutations which are not detectable by common sequencing (frequency range is shown in parentheses): E-G_{2163}A (0.08–0.14); E-C_{2165}G (0.08–0.14); NS4A-A_{6761}G (0.06–0.09); NS4B-T_{7551}A (0.10–0.16); NS5-A_{9060}G (0.06–0.16); NS5-T_{10142}C (0.20–0.40); NS5-T_{10144}C (0.21–0.40); NS5-T_{10148}A (0.20–0.41); NS5-A_{10149}G (0.21–0.42). The bird isolates also shared three transversions: NS3-T_{6278}G (0.085–0.088); NS3-T_{6298}G (0.08–0.10); and NS4B-A_{7558}T (0.08–0.14). In addition, the human isolate ARC13-06 and the bird isolate ID28bd-07 shared seven mutations in the 3′UTR: A_{10432}G (0.26–0.29); C_{10435}T (0.16–0.27); A_{10436}T (0.17–0.27); A_{10438}T (0.16–0.29); A_{10440}T (0.17–0.28); G_{10447}T (0.05–0.09) and T_{10449}A (0.078–0.15).

The two transitions in the ORF, E-$G_{2163}A$ and NS5-$A_{9060}G$, detected using NGS are silent. Others resulted in amino acid substitutions: E-$C_{2165}G$ in E-$S_{400}C$; NS3-$T_{6278}G$ in NS3-$V_{556}G$; NS3-$T_{6298}G$ in NS3-$W_{563}G$; NS4A-$A_{6761}G$ in NS4A-$E_{98}G$; NS4B-$T_{7551}A$ in NS4B-$N_{212}K$; NS4B-$A_{7558}T$ in NS4B-$S_{215}C$. A cluster of neighboring mutations in NS5, NS5-$T_{10142}C$, NS5-$T_{10144}C$, NS5-$T_{10148}A$, and NS5-$A_{10149}G$, resulted in three amino acid substitutions: NS5-$V_{821}A$; NS5-$W_{822}R$ and NS5-$I_{823}K/M$.

The results of the indel calls are shown in Figure 4. We found that all studied isolates demonstrated very similar indel patterns (Figure 4(A)). Detailed analysis of indel profiles of the 3′UTR variable regions (Figure 4(B)) revealed a correlation between the location of insertions and deletions and the identified indel motifs GTAAGT and YYYTR (Y = C/T; R = A/G), shown in Figure 3.

Figure 4. Indel profiles generated by HIVE. Y—percent of mutations; X—nucleotide number. (**A**) Indel profile of the complete genome; (**B**) Indel profile of the variable region of the 3′UTR. Deletions are shown in red, insertions in blue.
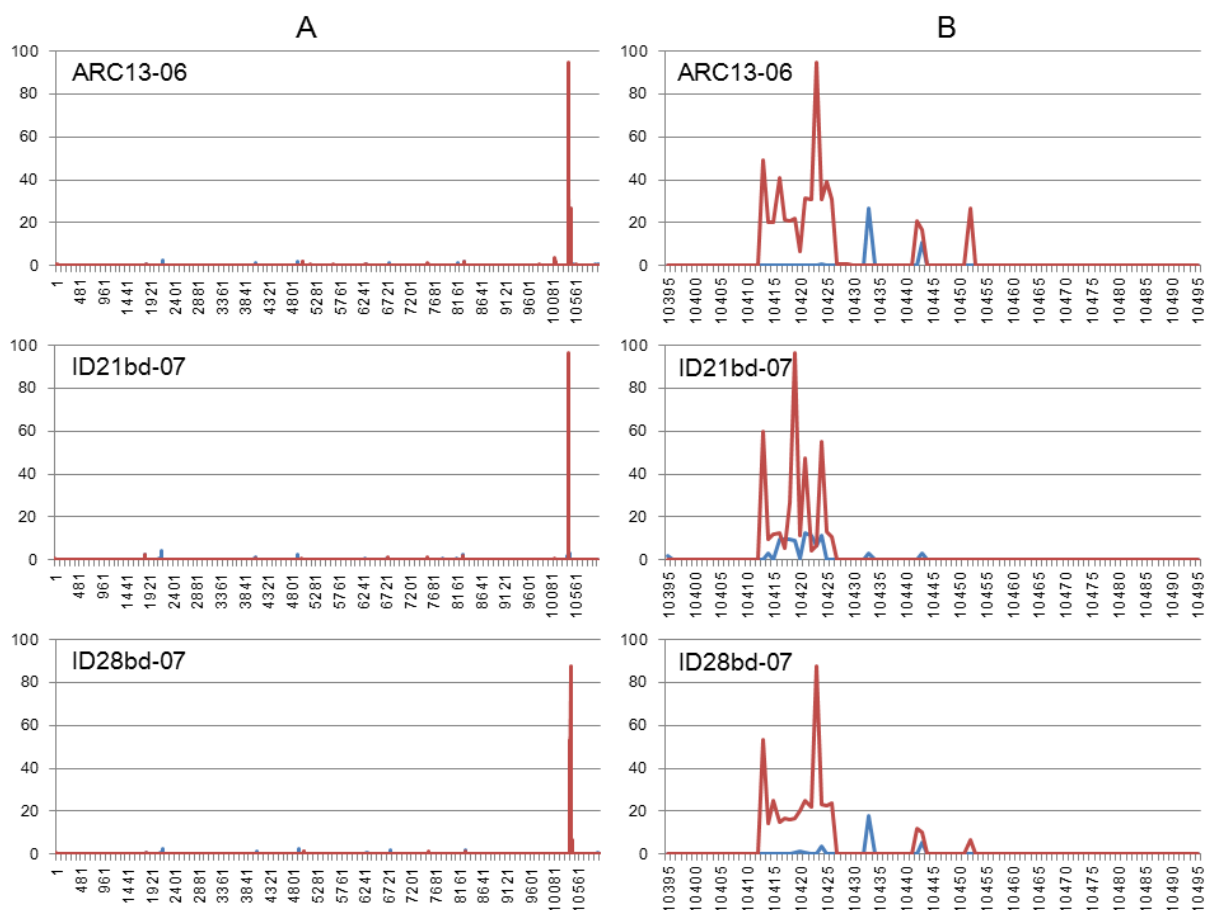
4. Discussion

Several studies have examined the evolutionary dynamics and spread of WNV after its introduction into North America [15,16,19,20,32,36–44]. This study focuses on the analysis of a WNV genetic variant which emerged in Idaho during 2006. Our recent phylogenetic analysis of WNV isolates from the U.S. [20] demonstrates that with few exceptions, WNV strains circulating in the U.S. form phylogenetic trees that are poorly differentiated spatially and temporally. One of these exceptions is the cluster MW/WN06 within the SW/WN03 genotype, which is supported by high bootstrapping and

Bayesian posterior probability values. This cluster was formed by five human WNV strains isolated from Idaho during the 2006–2007 outbreaks (ARC10-06, ARC13-06, ARC17-06, ARC27-06, ARC140-07), one human strain from North Dakota, 2006 (BSL106-06), and two avian WNV strains from Idaho collected in 2007 (ID21bd-07 and ID28bd-07) [20].

From 1999 through 2012, 1,243 human cases were reported to CDC from Idaho including 996 with 21 fatalities reported in 2006. Interestingly, only 13 cases of WNV infection had been reported in Idaho in 2005 before this outbreak. After the 2006 outbreak, decreasing numbers of human cases from Idaho were reported: 132 and one death in 2007; 39 and one death in 2008; 38 and two deaths in 2009; one in 2010; three in 2011; and 17 in 2012 [8]. Genetic analysis of WNV isolates from Idaho during 2006–2007 demonstrated that a new WNV variant had emerged and spread over southwest Idaho counties accounting for 58% of the 31 isolates included in this study (Table 1). These isolates carry the ID-Δ13 deletion (10,415–10,427) located at the variable region of the 3′UTR, and are genetically related. The same deletion was also identified in one human isolate from North Dakota in 2006, but not in any other isolate from that or any other state included in this study. All studied isolates from Idaho from 2007 had the ID-Δ13 deletion. The localized appearance of strains with ID-Δ13 may be due to the initial selection or introduction of one or few genetically similar viruses in the area with simultaneous and rapid spread to mosquitoes and local birds, amplifying the primarily carried genome. Human infections in that area would therefore result from one or very few local WNV colonizing genotypes, and reflect the introduction of one or very few infected vectors followed by rapid localized amplification.

Most studies on WNV evolution to date have focused on the structural genes or on the complete open reading frame. Here we have analyzed the genetic variability of WNV 3′UTR sequences using a dataset from 398 isolates from lineages 1 and 2 collected from different locations worldwide (Supplemental Table S1). Analysis of sequence alignment of WNV genomes revealed the presence of two repeats of GTAAATA $(N)_{0-29}$GT located in the variable region of the 3′UTR (~100 nt in length, located immediately after the ORF). These repeats are present in all WNV isolates from lineages 1 and 2 available in GenBank. It is possible that conserved direct repeats in the 3′UTR of mosquito-borne flaviviruses may function as replication enhancers selected under the constraints on transmission and dissemination imposed by the particular hosts [4]. We also observed that insertions and deletions are located in the vicinity of or within the two motifs GTAAGT and YYYTR (Y = C/T; R = A/G) (Figures 3 and 4). Those motifs had previously been described as indel motifs and reported to be associated with human genetic diseases [34,45,46]. Interestingly, the indel motif GTAAGT is similar to the consensus sequence at the 5′ end of introns (A/C)AG|GT(A/G)AGT in which the cleavage occurs between the G residues [47]. This observation raises suspicion that potentially the same mechanism of RNA posttranscriptional modification is acting on both host and viral genomic RNA, but that remains to be investigated.

A detailed analysis of the variable region of the 3′UTR revealed the presence of AUUUA and complementary UAAAU pentamers in all isolates from lineage 1 and isolates that do not carry any deletion from lineage 2 (Figure 3). The AUUUA pentamer is a typical adenosine-uridine rich element (ARE) present in short-living cellular mRNAs, which usually encode proteins required only for a short period of the cell cycle, such as transcription factors, oncogenes, and cytokines [35]. Many cellular proteins bind to cellular mRNAs that contain AREs, causing either mRNA degradation (e.g.,

ARE/poly(U)-binding/degradation factor 1 (AUF1)) [48] or protecting the mRNA from degradation (e.g., Hu antigen R (HuR)) [49]. AREs have been found in the non-coding parts of several viral genomes including WNV [50,51]. However, the clear function of these elements in the WNV genome is yet to be determined. Studies using a reporter replicon of WNV showed that deletion of most of the 3′UTR containing the variable region did not affect translation efficiency, although the region was essential for virus replication [52]. Small deletions in the variable regions of DENV-1 and JEV do not produce any difference in replication efficiency or plaque size in mammalian or mosquito cells [53,54]. Nevertheless, some reports demonstrated that deletion of most of the variable region of the 3′UTR can affect virus replication [55–58].

The 3′UTR variable region where we identified indel motifs is located near the 5′ end of subgenomic flavivirus RNA (sfRNA). sfRNA is a small, nuclease-resistant fragment produced in significant amounts by all members of the *Flavivirus* genus and is believed to play an important role in viral replication. The production of sfRNA is a result of incomplete genomic RNA degradation after stalling of the cellular exoribonuclease XRN1 on rigid secondary RNA structures in the 3′UTR of the viral genome [59,60]. We performed preliminary studies of sfRNA using WNV genetic variants carrying the insertion and deletions that we identified, and found no significant difference when compared to isolates without deletions (data not shown), which suggests that insertions and deletions in this variable region do not play a significant role in sfRNA formation. However, the relevance and role of the identified motifs and associated genome lesions in viral replication and their evolutionary consequences remain to be explored. The relatively high rate of occurrence of mutations, deletions and insertions in the variable region of the 3′UTR suggests that positive Darwinian selection may have acted on this part of the WNV genome [61].

The appearance of WNV in the New World provides a unique opportunity to understand how an arbovirus adapts and evolves in a new replicative environment. A number of factors might be involved in evolution of WNV including the level of genetic heterogeneity of the viral populations. Selection of preexisting genetic variants in a viral quasispecies swarm can potentially play a key role in adaptation of imported viruses to domestic vectors and hosts and in dissemination of newly emerging viruses [62–64]. Evidence for viral population heterogeneity, where individual sequences differ from the consensus sequence, has been obtained using cloning approaches for different viruses [65–68] including WNV [69]. WNV quasispecies studies based on cloning and sequencing of the 3′ 1,159 nt of the WNV envelope coding region and the 5′ 779 nt of the WNV NS1 coding region suggests that interhost quasispecies dynamics may potentially be less significant for WNV evolution than intrahost quasispecies dynamics [69–72]. However, the cloning technique is laborious and provides only a limited resolution of the quasispecies spectrum within a sample [73]. The well-established Sanger method identifies the consensus or major viral genome present in a particular isolate, when PCR products are directly sequenced, but this method is almost uninformative about minor genetic variants represented within the quasispecies swarm. Next-generation sequencing methods produce a massive amount of sequence data which can be applied for sequencing and re-sequencing of viral genomes to obtain thorough and detailed coverage of minor variants revealing nucleotide substitutions present in only a small part of the viral population [74].

We have used the Illumina NGS technology for analysis of within-host genetic variability of one human and two bird isolates from the Idaho 2006–2007 outbreaks. Our data demonstrate that genetic analysis based on common Sanger sequencing of PCR products provides limited information, and it missed many non-silent mutations in the E, NS4A, NS4B and NS5 genes of studied isolates, even when those mutations were present in relatively high frequencies, such as NS5-A_{10149}G (0.21–0.42). Indel call data generated by HIVE correlate with alignments based on common sequencing, and indel profiles from studied isolates demonstrate the peaks of deletion and insertion frequencies near or within identified indel motifs. However, in the location where the Sanger chromatograms showed only the 13 nucleotide deletion ID-Δ13, we also found numerous reads containing different deletions and insertions, as well as reads with intact sequences. In this study we used the same amplicons for Illumina and Sanger sequencing technologies. Those amplicons were prepared from supernatants from a single passage of original samples in Vero cells. Cultivation, as well as the downstream process used for the generation of the amplicons including synthesis of viral cDNA with subsequent PCR amplification using specific primers, has the potential for selection bias, in which some fractions of within-host viral populations may be lost. Therefore, further work is required to address this point. We are planning to compare the viral swarms present in viral isolates to those obtained from the original samples (*i.e.*, plasma, serum, mosquito pools) when sufficient starting material can be obtained. Nevertheless, our findings strongly suggest that data generated by NGS can provide new insights into WNV evolutionary dynamics at a better resolution than that previously achieved by using traditional sequencing methods.

5. Conclusions

The introduction of WNV into North America represents a unique opportunity to understand how an arbovirus evolves in a new replicative environment. Adaptation to domestic mosquitoes and birds through selection of preexisting genetic variants from a quasispecies swarm may have played a major role in the spread of WNV in the Americas. Genetic analysis of WNV isolates collected during the 2006–2007 epidemics reveals a new genetic variant of the virus. This variant had emerged coincidentally with an intense outbreak in Idaho during 2006. One human and two bird isolates from Idaho, 2006–2007, were re-sequenced using Illumina NGS technology and within-host genetic variability was analyzed. The NGS method produced additional data about mutations presented in minor genetic variants, which were not detectable by common direct Sanger sequencing. NGS technologies have been significantly improved in the past few years allowing for better understanding of viral evolution, fitness, emergence and transmission. Adequate surveillance based on new technologies is essential to public health since emerging mutants of pathogens could potentially affect performance of diagnostic assays, and negatively impact the development of vaccines and specific therapies.

Acknowledgments

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Añez, G.; Chancey, C.; Grinev, A.; Rios, M. Dengue virus and other arboviruses: A global view of risks. *ISBT Sci. Ser.* **2012**, *7*, 274–282.
2. Brinton, M.A. Molecular Biology of West Nile Virus. In *West Nile Encephalitis Virus Infection: Viral Pathogenesis and the Host Immune Response*; Diamond, M., Ed.; Springer: New York, NY, USA, 2009; pp. 97–136.
3. Markoff, L. 5′- and 3′-noncoding regions in flavivirus RNA. *Adv. Virus Res.* **2003**, *59*, 177–228.
4. Gritsun, T.S.; Gould, E.A. Direct repeats in the flavivirus 3′ untranslated region; a strategy for survival in the environment? *Virology* **2007**, *358*, 258–265.
5. Murray, K.O.; Mertens, E.; Despres, P. West Nile virus and its emergence in the United States of America. *Vet. Res.* **2010**, *41*, 67–81.
6. Kauffman, E.B.; Franke, M.A.; Wong, S.J.; Kramer, L.D. Detection of West Nile virus. *Methods Mol. Biol.* **2011**, *665*, 383–413.
7. Hayes, E.B.; Gubler, D.J. West Nile virus: Epidemiology and clinical features of an emerging epidemic in the United States. *Annu. Rev. Med.* **2006**, *57*, 181–194.
8. Centers for Disease Control and Prevention. Available online: http://www.cdc.gov/westnile/statsMaps/ (accessed on 26 June 2013).
9. U.S. Department of the Interior, U.S. Geological Survey. Available online: http://diseasemaps.usgs.gov/wnv_historical.html (accessed on 26 June 2013).
10. Berthet, F.X.; Zeller, H.G.; Drouet, M.T.; Rauzier, J.; Digoutte, J.P.; Deubel, V. Extensive nucleotide changes and deletions within the envelope glycoprotein gene of Euro-African West Nile viruses. *J. Gen. Virol.* **1997**, *78*, 2293–2297.
11. Lanciotti, R.S.; Ebel, G.D.; Deubel, V.; Kerst, A.J.; Murri, S.; Meyer, R.; Bowen, M.; McKinney, N.; Morrill, W.E.; Crabtree, M.B.; *et al.* Complete genome sequences and phylogenetic analysis of West Nile virus strains isolated from the United States, Europe, and the Middle East. *Virology* **2002**, *298*, 96–105.
12. Bakonyi, T.; Hubálek, Z.; Rudolf, I.; Nowotny, N. Novel flavivirus or new lineage of West Nile virus, Central Europe. *Emerg. Infect. Dis.* **2005**, *11*, 225–231.
13. Bondre, V.P.; Jadi, R.S.; Mishra, A.C.; Yergolkar, P.N.; Arankalle, V.A. West Nile virus isolates from India: Evidence for a distinct genetic lineage. *J. Gen. Virol.* **2007**, *88*, 875–884.
14. May, F.J.; Davis, C.T.; Tesh, R.B.; Barrett, A.D. Phylogeography of West Nile virus: From the cradle of evolution in Africa to Eurasia, Australia, and the Americas. *J. Virol.* **2011**, *85*, 2964–2974.
15. Davis, C.T.; Ebel, G.D.; Lanciotti, R.S.; Brault, A.C.; Guzman, H.; Siirin, M.; Lambert, A.; Parsons, R.E.; Beasley, D.W.; Novak, R.J.; *et al.* Phylogenetic analysis of North American West Nile virus isolates, 2001–2004: Evidence for the emergence of a dominant genotype. *Virology* **2005**, *342*, 252–265.

16. Snapinn, K.W.; Holmes, E.C.; Young, D.S.; Bernard, K.A.; Kramer, L.D.; Ebel, G.D. Declining growth rate of West Nile virus in North America. *J. Virol.* **2007**, *81*, 2531–2534.

17. Moudy, R.M.; Meola, M.A.; Morin, L.L.; Ebel, G.D.; Kramer, L.D. A newly emergent genotype of West Nile virus is transmitted earlier and more efficiently by Culex mosquitoes. *Am. J. Trop. Med. Hyg.* **2007**, *77*, 365–370.

18. Vanlandingham, D.L.; McGee, C.E.; Klingler, K.A.; Galbraith, S.E.; Barrett, A.D.T.; Higgs, S. Comparison of oral infectious dose of West Nile virus isolates representing three distinct genotypes in Culex quinquefasciatus. *Am. J. Trop. Med. Hyg.* **2008**, *79*, 951–954.

19. McMullen, A.R.; May, F.J.; Li, L.; Guzman, H.; Bueno, R., Jr.; Dennett, J.A.; Tesh, R.B.; Barrett, A.D. Evolution of new genotype of West Nile virus in North America. *Emerg. Infect. Dis.* **2011**, *17*, 785–793.

20. Añez, G.; Grinev, A.; Chancey, C.; Ball, C.; Akolkar, N.; Land, K.J.; Winkelman, V.; Stramer, S.L.; Kramer, L.D.; Rios, M. Evolutionary dynamics of West Nile virus in the United States, 1999–2011: Phylogeny, selection pressure and evolutionary time-scale analysis. *PLoS Negl. Trop. Dis.* **2013**, *7*, doi: 10.1371/journal.pntd.0002245.

21. Aaskov, J.; Buzacott, K.; Field, E.; Lowry, K.; Berlioz-Arthaud, A.; Holmes, E.C. Multiple recombinant dengue type 1 viruses in an isolate from a dengue patient. *J. Gen. Virol.* **2007**, *88*, 3334–3340.

22. Domingo, E.; Martin, V.; Perales, C.; Grande-Pérez, A.; García-Arriaza, J.; Arias, A. Viruses as quasispecies: Biological implications. *Curr. Top. Microbiol. Immunol.* **2006**, *299*, 51–82.

23. Ruiz-Jarabo, C.M.; Arias, A.; Baranowski, E.; Escarmís, C.; Domingo, E. Memory in viral quasispecies. *J. Virol.* **2000**, *74*, 3543–3547.

24. Rogers, Y.H.; Venter, J.C. Genomics: Massively parallel sequencing. *Nature* **2005**, *437*, 326–327.

25. Mardis, E.R. Next-generation DNA sequencing methods. *Annu. Rev. Genomics. Hum. Genet.* **2008**, *9*, 387–402.

26. Eckerle, L.D.; Becker, M.M.; Halpin, R.A.; Li, K.; Venter, E.; Lu, X.; Scherbakova, S.; Graham, R.L.; Baric, R.S.; Stockwell, T.B.; *et al.* Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog.* **2010**, *6*, doi: 10.1371/journal.ppat.1000896.

27. Rozera, G.; Abbate, I.; Bruselles, A.; Vlassi, C.; D'Offizi, G.; Narciso, P.; Chillemi, G.; Prosperi, M.; Ippolito, G.; Capobianchi, M.R. Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte sub-populations. *Retrovirology* **2009**, *6*, doi: 10.1186/1742-4690-6-15.

28. Margeridon-Thermet, S.; Shulman, N.S.; Ahmed, A.; Shahriar, R.; Liu, T.; Wang, C.; Holmes, S.P.; Babrzadeh, F.; Gharizadeh, B.; Hanczaruk, B.; *et al.* Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. *J. Infect. Dis.* **2009**, *199*, 1275–1285.

29. Baillie, G.J.; Galiano, M.; Agapow, P.M.; Myers, R.; Chiam, R.; Gall, A.; Palser, A.L.; Watson, S.J.; Hedge, J.; Underwood, A.; *et al.* Evolutionary dynamics of local pandemic H1N1/09 influenza lineages revealed by whole genome analysis. *J. Virol.* **2011**, *86*, 11–18.

30. Wright, C.F.; Morelli, M.J.; Thebaud, G.; Knowles, N.J.; Herzyk, P.; Paton, D.J.; Haydon, D.T.; King, D.P. Beyond the consensus: Dissecting within-host viral population diversity of foot-and-mouth disease virus using next-generation genome sequencing. *J. Virol.* **2010**, *85*, 2266–2275.

31. Daly, G.M.; Be xfield, N.; Heaney, J.; Stubbs, S.; Mayer, A.P.; Palser, A.; Kellam, P.; Drou, N.; Caccamo, M.; Tiley, L.; *et al*. A viral discovery methodology for clinical biopsy samples utilising massively parallel next generation sequencing. *PLoS One* **2011**, *6*, doi: 10.1371/journal.pone.0028879.

32. Grinev, A.; Daniel, S.; Stramer, S.; Rossmann, S.; Caglioti, S.; Rios, M. Genetic variability of West Nile virus in US blood donors, 2002–2005. *Emerg. Infect. Dis.* **2008**, *14*, 436–444.

33. HIVE: High-performance Integrated Virtual Environment Provides Solutions for Next-Generation Sequencing Data Storage and Analysis. Available online: https://hive.biochemistry.gwu.edu/ HIVEWhitePaper.pdf (accessed on 26 June 2013).

34. Ball, E.V.; Stenson, P.D.; Abeysinghe, S.S.; Krawczak, M.; Cooper, D.N.; Chuzhanova, N.A. Microdeletions and microinsertions causing human genetic disease: Common mechanisms of mutagenesis and the role of local DNA sequence complexity. *Hum. Mutat.* **2005**, *26*, 205–213.

35. Chen, C.Y.; Shyu, A.B. AU-rich elements: Characterization and importance in mRNA degradation. *Trends. Biochem. Sci.* **1995**, *20*, 465–470.

36. Lanciotti, R.S.; Roehrig, J.T.; Deubel, V.; Smith, J.; Parker, M.; Steele, K.; Crise, B.; Volpe, K.E.; Crabtree, M.B.; Scherret, J.H.; *et al*. Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science* **1999**, *286*, 2333–2337.

37. Beasley, D.W.; Davis, C.T.; Guzman, H.; Vanlandingham, D.L.; Travassos da Rosa, A.P.; Parsons, R.E.; Higgs, S.; Tesh, R.B.; Barrett, A.D. Limited evolution of West Nile virus has occurred during its southwesterly spread in the United States. *Virology* **2003**, *309*, 190–195.

38. Davis, C.T.; Beasley, D.W.; Guzman, H.; Raj, R.; D'Anton, M.; Novak, R.J.; Unnasch, T.R.; Tesh, R.B.; Barrett, A.D. Genetic variation among temporally and geographically distinct West Nile virus isolates, United States, 2001, 2002. *Emerg. Infect. Dis.* **2003**, *9*, 1423–1429.

39. Ebel, G.D.; Carricaburu, J.; Young, D.; Bernard, K.A.; Kramer, L.D. Genetic and phenotypic variation of West Nile virus in New York, 2000–2003. *Am. J. Trop. Med. Hyg.* **2004**, *71*, 493–500.

40. Herring, B.L.; Bernardin, F.; Caglioti, S.; Stramer, S.; Tobler, L.; Andrews, W.; Cheng, L.; Rampersad, S.; Cameron, C.; Saldanha, J.; *et al*. Phylogenetic analysis of WNV in North American blood donors during the 2003–2004 epidemic seasons. *Virology* **2007**, *363*, 220–228.

41. Bertolotti, L.; Kitron, U.; Goldberg, T.L. Diversity and evolution of West Nile virus in Illinois and the United States, 2002–2005. *Virology* **2007**, *360*, 143–149.

42. Bertolotti, L.; Kitron, U.D.; Walker, E.D.; Ruiz, M.O.; Brawn, J.D.; Loss, S.R.; Hamer, G.L.; Goldberg, T.L. Fine-scale genetic variation and evolution of West Nile Virus in a transmission "hot spot" in suburban Chicago, USA. *Virology* **2008**, *374*, 381–389.

43. Gray, R.R.; Veras, N.M.; Santos, L.A.; Salemi, M. Evolutionary characterization of the West Nile Virus complete genome. *Mol. Phylogenet. Evol.* **2010**, *56*, 195–200.

44. Armstrong, P.M.; Vossbrinck, C.R.; Andreadis, T.G.; Anderson, J.F.; Pesko, K.N.; Newman, R.M.; Lennon, N.J.; Birren, B.W.; Ebel, G.D.; Henn, M.R. Molecular evolution of West Nile virus in a northern temperate region: Connecticut, USA 1999–2008. *Virology* **2011**, *417*, 203–210.

45. Chuzhanova, N.A.; Anassis, E.; Ball, E.; Krawczak, M.; Cooper, D.N. Meta-analysis of indels causing human genetic disease: Mechanisms of mutagenesis and the role of local DNA sequence complexity. *Hum. Mutat.* **2003**, *21*, 28–44.

46. Kondrashov, A.S.; Rogozin, I.B. Context of deletions and insertions in human coding sequences. *Hum. Mutat.* **2004**, *23*, 177–185.

47. Rogozin, I.B.; Sverdlov, A.V.; Babenko, V.N.; Koonin, E.V. Analysis of evolution of exon-intron structure of eukaryotic genes. *Brief. Bioinform.* **2005**, *6*, 118–134.

48. DeMaria, C.T.; Brewer, G. AUF1 binding affinity to A + U-rich elements correlates with rapid mRNA degradation. *J. Biol. Chem.* **1996**, *271*, 12179–12184.

49. Peng, S.S.; Chen, C.Y.; Xu, N.; Shyu, A.B. RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J.* **1998**, *17*, 3461–3470.

50. Li, W.; Li, Y.; Kedersha, N.; Anderson, P.; Emara, M.; Swiderek, K.M.; Moreno, G.T.; Brinton, M.A. Cell proteins TIA-1 and TIAR interact with the 3′ stem-loop of the West Nile virus complementary minus-strand RNA and facilitate virus replication. *J. Virol.* **2002**, *76*, 11989–12000.

51. Nadar, M.; Chan, M.Y.; Huang, S.W.; Huang, C.C.; Tseng, J.T.; Tsai, C.H. HuR binding to AU-rich elements present in the 3′ untranslated region of Classical swine fever virus. *Virol. J.* **2011**, *8*, doi: 10.1186/1743-422X-8-340.

52. Tilgner, M.; Deas, T.S.; Shi, P.Y. The flavivirus-conserved penta-nucleotide in the 3′ stem-loop of the West Nile virus genome requires a specific sequence and structure for RNA synthesis, but not for viral translation. *Virology* **2005**, *331*, 375–386.

53. Tajima, S.; Nukui, Y.; Ito, M.; Takasaki, T.; Kurane, I. Nineteen nucleotides in the variable region of 3′ non-translated region are dispensable for the replication of dengue type 1 virus *in vitro*. *Virus Res.* **2006**, *116*, 38–44.

54. Kato, F.; Kotaki, A.; Yamaguchi, Y.; Shiba, H.; Hosono, K.; Harada, S.; Saijo, M.; Kurane, I.; Takasaki, T.; Tajima, S. Identification and characterization of the short variable region of the Japanese encephalitis virus 3′NTR. *Virus Genes* **2012**, *44*, 191–197.

55. Men, R.; Bray, M.; Clark, D.; Chanock, R.M.; Lai, C.J. Dengue type 4 virus mutants containing deletions in the 3′ noncoding region of the RNA genome: Analysis of growth restriction in cell culture and altered viremia pattern and immunogenicity in rhesus monkeys. *J. Virol.* **1996**, *70*, 3930–3937.

56. Yun, S.I.; Choi, Y.J.; Song, B.H.; Lee, Y.M. 3′ cis-acting elements that contribute to the competence and efficiency of Japanese encephalitis virus genome replication: Functional importance of sequence duplications, deletions, and substitutions. *J. Virol.* **2009**, *83*, 7909–7930.

57. Alvarez, D.E.; de Lella Ezcurra, A.L.; Fucito, S.; Gamarnik, A.V. Role of RNA structures present at the 3′UTR of dengue virus on translation, RNA synthesis, and viral replication. *Virology* **2005**, *339*, 200–212.

58. Tajima, S.; Nukui, Y.; Takasaki, T.; Kurane, I. Characterization of the variable region in the 3′ non-translated region of dengue type 1 virus. *J. Gen. Virol.* **2007**, *88*, 2214–2222.

59. Funk, A.; Truong, K.; Nagasaki, T.; Torres, S.; Floden, N.; Balmori Melian, E.; Edmonds, J.; Dong, H.; Shi, P.Y.; Khromykh, A.A. RNA structures required for production of subgenomic flavivirus RNA. *J. Virol.* **2010**, *84*, 11407–11417.

60. Pijlman, G.P.; Funk, A.; Kondratieva, N.; Leung, J.; Torres, S.; van der Aa, L.; Liu, W.J.; Palmenberg, A.C.; Shi, P.Y.; Hall, R.A.; *et al*. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe* **2008**, *4*, 579–591.

61. Hughes, A.L.; Piontkivska, H.; Foppa, I. Rapid fixation of a distinctive sequence motif in the 3′ noncoding region of the clade of West Nile virus invading North America. *Gene* **2007**, *399*, 152–161.

62. Holmes, E.C. The RNA virus quasispecies: Fact or fiction? *J. Mol. Biol.* **2010**, *400*, 271–273.

63. Fishman, S.L.; Branch, A.D. The quasispecies nature and biological implications of the hepatitis C virus. *Infect. Genet. Evol.* **2009**, *9*, 1158–1167.

64. Pfeiffer, J.K.; Kirkegaard, K. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog.* **2005**, *1*, doi: 10.1371/journal.ppat.0010011.

65. Martell, M.; Esteban, J.I.; Quer, J.; Genescà, J.; Weiner, A.; Esteban, R.; Guardia, J.; Gómez, J. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: Quasispecies nature of HCV genome distribution. *J. Virol.* **1992**, *66*, 3225–3229.

66. Plyusnin, A.; Cheng, Y.; Lehväslaiho, H.; Vaheri, A. Quasispecies in wild-type Tula hantavirus populations. *J. Virol.* **1996**, *70*, 9060–9063.

67. Quiñones-Mateu, M.E.; Albright, J.L.; Mas, A.; Soriano, V.; Arts, E.J. Analysis of pol gene heterogeneity, viral quasispecies, and drug resistance in individuals infected with group O strains of human immunodeficiency virus type 1. *J. Virol.* **1998**, *72*, 9002–9015.

68. Cottam, E.M.; King, D.P.; Wilson, A.; Paton, D.J.; Haydon, D.T. Analysis of foot-and-mouth disease virus nucleotide sequence variation within naturally infected epithelium. *Virus Res.* **2009**, *140*, 199–204.

69. Jerzak, G.; Bernard, K.A.; Kramer, L.D.; Ebel, G.D. Genetic variation in West Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection. *J. Gen. Virol.* **2005**, *86*, 2175–2183.

70. Ciota, A.T.; Ngo, K.A.; Lovelace, A.O.; Payne, A.F.; Zhou, Y.; Shi, P.Y.; Kramer, L.D. Role of the mutant spectrum in adaptation and replication of West Nile virus. *J. Gen. Virol.* **2007**, *88*, 865–874.

71. Jerzak, G.V.; Bernard, K.; Kramer, L.D.; Shi, P.Y.; Ebel, G.D. The West Nile virus mutant spectrum is host-dependant and a determinant of mortality in mice. *Virology* **2007**, *360*, 469–476.

72. Ciota, A.T.; Lovelace, A.O.; Jia, Y.; Davis, L.J.; Young, D.S.; Kramer, L.D. Characterization of mosquito-adapted West Nile virus. *J. Gen. Virol.* **2008**, *89*, 1633–1642.

73. Liang, B.; Luo, M.; Scott-Herridge, J.; Semeniuk, C.; Mendoza, M.; Capina, R.; Sheardown, B.; Ji, H.; Kimani, J.; Ball, B.T.; *et al*. A comparison of parallel pyrosequencing and Sanger clone-based sequencing and its impact on the characterization of the genetic diversity of HIV-1. *PLoS One* **2011**, *6*, doi: 10.1371/journal.pone.0026745.

74. Chin-inmanu, K.; Suttitheptumrong, A.; Sangsrakru, D.; Tangphatsornruang, S.; Tragoonrung, S.; Malasit, P.; Tungpradabkul, S.; Suriyaphol, P. Feasibility of using 454 pyrosequencing for studying quasispecies of the whole dengue viral genome. *BMC Genomics* **2012**, *13*, doi: 10.1186/1471-2164-13-S7-S7.