Next Generation Sequencing shows quasispecies diversification after a single passage in a carrion crow (*Corvus corone*) *in vivo* infection model.

Running title: WNV quasispecies diversification in crows

Content Category: Standard paper, Animal RNA Viruses

Dridi M.[a], Rosseel T.*[a], Orton R.[b,e], Johnson P.[b], Lecollinet S.[c], Muylkens B.[d], Lambrecht B.[a], Van Borm S.[a]*

* Equal contribution.

[a] Operational Direction of Viral Diseases, Veterinary and Agrochemical Research Center (CODA-CERVA-VAR), 99 Groeselenberg, 1180 Brussels, Belgium.

[b] Boyd Orr Centre for Population and Ecosystem Health, Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, G12 8QQ, United Kingdom.

[c] UMR1161 Virologie INRA, ANSES, ENVA, French Agency for Food, Environmental and Occupational Health & Safety (Anses), 23 avenue du Général De Gaulle, 94706 Maisons-Alfort, France.

[d] Integrated Veterinary Research Unit, University of Namur, 61 rue de Bruxelles, 5000 Namur, Belgium.

[e] MRC-University of Glasgow Centre for Virus Research, Institute of Infection Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, G61 1QH, United Kingdom.

Corresponding author: Van Borm, Steven. Operational Direction of Viral Diseases, Veterinary and Agrochemical Research Center (CODA-CERVA-VAR), 99 Groeselenberg, 1180 Brussels, Belgium. Email: steven.vanborm@coda-cerva.be. Tel +32 2 3790 505. Fax +32 2 3790 401.

Summary word count: 209

Text word count excluding summary, acknowledgements and references: 5473

Number of tables: 2

Number of figures: 3

Number of supplementary files: 1

The Genbank accession number for the near complete coding sequence of the WNV inoculum used in this study is KR107956

SUMMARY

West Nile virus (WNV) occurs as a population of genetic variants (quasispecies) infecting a single animal. Previous low resolution viral genetic diversity estimates in sampled wild birds and mosquitoes and in multiple passage adaptation studies *in vivo* or in cell culture, suggest that WNV genetic diversification is mostly limited to the mosquito vector. This study is the first to investigate genetic diversification of WNV in avian hosts during a single passage using next generation sequencing. Wild-captured carrion crowswere subcutaneously infected using a clonal Middle-East WNV. Blood samples were collected on 2 and 4 days post-infection. A RT-PCR approach was used to amplify the WNV genome directly from serum samples prior to next generation sequencing resulting in an average depth of at least 700x in each sample. Appropriate controls were sequenced to discriminate biologically relevant low frequency variants from experimentally introduced errors. The WNV populations in the wild crows showed significant diversification away from the inoculum virus quasispecies structure. On the contrary, WNV populations in intracerebrally infected day-old chickens did not diversify from that of the inoculum. Where previous studies concluded that WNV genetic diversification is only experimentally demonstrated in its permissive insect vector species, we have experimentally shown significant diversification of WNV populations in a wild bird reservoir species.

INTRODUCTION

West Nile virus (WNV) is an arthropod-borne positive sense, single-stranded RNA virus that belongs to the Japanese encephalitis virus antigenic complex in the *Flavivirus* genus (*Flaviviridae* family) (Murray *et al.*, 2010). Sequence comparison of 24 American WNV isolates sampled between 1999 and 2005 revealed that the circa 11 kb WNV genome evolves at a rate of approximately 5 nucleotide substitutions per genome per year (Tang *et al.*, 2008). This exceedingly high mutation rate (10^5- to 10^8-fold the mutation rate of their host's genomic DNA) confers this RNA virus the potential for rapid evolution (Holland *et al.*, 1982) and is due to the high error rate of its RNA-dependent RNA polymerase (RdRp) (Van Slyke *et al.*, 2013).

WNV exists in nature as a quasispecies, i.e. a population of genetic variants hypothesized to undergoing selection as a unit (Jerzak *et al.*, 2005; Van Slyke *et al.*, 2013). The average percentage of mutated nucleotides in naturally infected American crows (*Corvus brachyrhynchos*) and pools of *Culex* mosquitoes amounted respectively to 0.011 % and 0.023% (Jerzak *et al.*, 2005).

WNV is maintained in a bird-arthropod cycle, with ornithophilous *Culex* mosquitoes representing the main vectors, and birds the natural reservoir (Hubálek & Halouzka, 1999; Pfeffer & Dobler, 2010; Turell *et al.*, 2005). Some bird species are sensitive to WNV infection and capable of amplifying the virus, such as birds of the Passeriformes order, especially corvids. Galliformes, Columbiformes and Gruiformes were shown to be more resistant (Komar *et al.*, 2003; Van Der Meulen *et al.*, 2005; Wheeler *et al.*, 2009). Nevertheless, some resistant bird species, including specific pathogen-free (SPF) chickens and Aigamo ducks, are transiently susceptible to WNV infection during a few days following hatching (Dridi *et al.*, 2013; Phipps *et al.*, 2007; Senne *et al.*, 2000; Shirafuji *et al.*, 2009; Totani *et al.*, 2011).

WNV genetic diversification seems to differ depending on the host it colonizes. So far intrahost WNV population dynamics has only been investigated in mosquitoes and chickens (Deardorff *et al.*, 2011; Jerzak *et al.*, 2005, 2007, 2008). From those studies, mosquitoes are vectors that contribute to WNV genetic diversity by relaxing purifying selection, while the more restrictive birds, modeled by young chickens, act as "selective sieves" that maintain a strong purifying selection and shrink the spectrum of WNV genetic variants. This difference in WNV genetic diversity between avian and arthropod hosts might be linked to the antiviral

response prevailing in mosquitoes, namely, RNA interference (RNAi) (Campbell *et al.*, 2008; Myles *et al.*, 2008). It was indeed demonstrated that WNV diversity is lower in RNAi-depleted *Drosophila* cells and higher in RNAi-stimulated cells (Brackney *et al.*, 2009). Moreover, it was shown that RNAi targeting of regions of the WNV genome resulted in increased genetic diversity in these regions compared to more weakly targeted regions (TO BE ADDED: Brackney et al 2015).

Stressing the functional importance of quasispecies diversity, Villareal and Witzany (2013) proposed a concept of quasispecies consortia where individual RNAs operate as cooperative modules (Villarreal & Witzany, 2013). The target of selection is the quasispecies rather than the single (molecular) species (Eigen & Schuster, 1978). Using Hela cells and a model of susceptible mice, it was demonstrated that the pathogenicity of poliovirus is not only determined by the selection of adaptive mutations, but that quasispecies diversity *per se* is a major determinant of virulence (Vignuzzi *et al.*, 2006). However, in another study (Deardorff *et al.*, 2011) the reduction in the genetic diversity during twenty consecutive passages of WNV in SPF chickens resulted in an increased viral replication in that host, while the enlarged genetic diversity generated by twenty consecutive passages of WNV in mosquitoes of the *Culex* genus led to a decreased viral replication in SPF chickens.

The SPF chicken is so far the only bird species investigated for its influence on WNV genetic population dynamics. It is not known whether all WNV avian host species are equally restrictive for WNV genetic expansion. Furthermore, the available studies on WNV mutant spectrum diversification relied on Sanger sequencing methods (Deardorff *et al.*, 2011; Jerzak *et al.*, 2005, 2007, 2008), whereby a limited set of consensus sequences were obtained after a plaque-purification or molecular cloning step. Next generation sequencing technologies (NGS) present an opportunity to deepen the analysis of RNA virus genetic variants (Barzon *et al.*, 2011; Van Borm *et al.*, 2015).

The aim of this study was to investigate whether a genetically homogeneous cloned virus (derived from a lineage 1 Israeli strain: IS-98 STD1) would undergo qualitatively and quantitatively different genetic diversification schemes after passage in young SPF chickens or wild juvenile Carrion crows. NGS was used to allow quasispecies resolution analysis. We inoculated day-old SPF chickens and wild-caught juvenile Carrion crows with cloned WNV by the intracerebral (IC) or subcutaneous (SC) route respectively. The effect of sampling time and avian host species on the evolution of WNV quasispecies detected in collected sera was determined.

RESULTS

Animal experiments

In order to investigate WNV quasispecies dynamics in infected birds, chickens and Carrion crows were inoculated. The first clinical signs of WN disease (WND) were observed in one SPF chicken at 2 dpi, which died at 3 dpi (prostration, lethargy, unresponsiveness). By 4 dpi, all the remaining chickens were prostrate and less reactive. At autopsy, spleens of chickens sacrificed on 4 dpi appeared hypertrophied. The infected crows showed clinical signs (reduced activity, prostration, lack of reactivity) of WND only on 4 dpi.

Viral RNA loads in sera of both bird species confirmed the replication of WNV in the hosts (Table 1). In order to ensure sufficient amounts of viral RNA for performing the deep-sequencing analysis, only the three most viremic chickens at 2 and 4 dpi were selected for further deep-sequencing analysis.

WNV full genome sequencing

The thirteen infection samples sequenced in this study (inoculum and RNA from the sera of the Carrion crows and SPF chickens sacrificed at 2 and 4 dpi, Fig. 1) had identical consensus sequences. This consensus sequence (Genbank accession number KR107956) differed by 16 nucleotides from the Is98 strain (accession number: AF481864) from which the cloned virus was derived. One of these single nucleotide variants (SNVs) is a silent mutation intentionally introduced to create a *SnaBI* restriction site (Bahuon *et al.*, 2012). The remaining fifteen SNVs could have arisen during the passaging of the WNV prior to cloning, the amplification and the subcloning of the Is98 genome, the amplification of the Is98-expressing plasmid in E. coli DH5α, or after transfection into Vero cells for the first rounds of viral amplification.

Our NGS sequencing strategy resulted in an average coverage depth of at least 700x in all samples, allowing quantification of minor frequency variants in the RNA virus quasispecies population. However, due to the variation observed in the control sample (the PCR amplified plasmid, see Supplementary Information S1), a cutoff frequency >1.81% was used to validate variants as significantly different from artifactually introduced RT-PCR and sequencing errors.

The detected genetic variations were scattered over the WNV genome (Table 2). Importantly, none of the observed polymorphic sites involved the single nucleotide differences between our inoculum and the Is98 strain (AF481864). Only in the 5’ non-coding region and in the

NS2b-coding region no variants were observed. Most of these substitutions (34/37) were transitions. Overall, the number of non-synonymous mutations exceeded the number of silent mutations (22 vs. 15). Minor variants that were conserved between the inoculum and the chicken and crow samples distributed equally between silent and non-synonymous mutations, and were mostly localized in the region coding for NS1. The ratio of non-synonymous (NS) to silent (S) mutations did not differ between bird groups (NS/S = 7/8 for chickens and NS/S = 11/11 for crows n°1, n°2, n°4 and n°6), except for crows n°3 and n°5 that gave rise to the most frequent non-synonymous mutations (NS/S = 12/5 and NS/S = 6/3 respectively). Overall, the validated genetic variants did not show significant host species bias, and were scattered over the genome with increased incidence in the NS1 encoding region.

Quasispecies characterization

Clustering analysis of all the polymorphic loci (Table 2) combined with a heatmap visualization of variant frequencies showed that the distribution of mutation frequencies over the genome was more similar in chickens and the inoculum which cluster closely together, while the crows formed a separate cluster and showed more variability (Fig. 2). All chicken samples seemed to maintain the diversity pattern (loci & their variant frequency) present in the inoculum. On the contrary, the quasispecies populations in crows tended to drift away from the population in the inoculum, changing the frequency at variant loci already present in inoculum and generating new variant loci (Fig. 2).

To determine differences in genetic diversity between host species and sampling times, Euclidian pairwise differences in variant frequencies over all polymorphic sites were calculated to demonstrate differences within and between groups (Fig. 3). In terms of WNV genetic diversification, the chickens were similar to each other and to the inoculum, while crows showed much higher levels of diversification from the inoculum and from each other. Specifically, levels of genetic diversification between the crows and the inoculum and between the crows and the chickens did not differ significantly ($P = 0.51$) but were greater than diversification between the chickens and the inoculum ($P < 10^{-5}$; Fig. 3(a)). Moreover, within-group diversification in variant frequencies over all polymorphic sites was higher within the crow group than within the chicken group ($P < 10^{-5}$; Fig. 3(b)), indicating variable quasispecies diversification in crows compared to a more homogenous quasispecies dynamics in chickens. Time after infection (2 dpi vs 4 dpi) did not seem to affect the observed Euclidian distances (Fig. 3).

Looking more in detail at the low frequency variants detected (Table 2), we observed that the inoculum was characterized by 12 polymorphic sites, with the highest frequency observed at 27.99 % of the viral population. This was surprising given that the inoculum is the product of limited rounds of genomic RNA amplification in Vero cells after electroporation. Although samples derived from chickens showed virtually the same genetic variants (at similar frequencies) as the inoculum, a few exceptions were evident. These included a SNV at position 5,608 that appeared in all the chickens but chicken n°14 (sacrificed at 4 dpi), a SNV at position 10,312 that was only detected in chicken n°3 (sacrificed at 2 dpi), and a SNV at position 9,716 that was present in the inoculum but disappeared from all the six chicken samples (Table 2).

The diversification of WNV quasispecies in crows away from the inoculum quasispecies population resulted in 18 new genetic variants in crows n°1, n°2 and n°3 in comparison with the inoculum or the chicken samples (Table 2). Crow n°3 had the highest number of differing genetic variants (n = 17). The variant at position 9,706 of the inoculum was cleared from the quasispecies population in crows, as it was for chickens. At 4 dpi, 5 additional SNVs appeared in crows n°4, 5 or 6 in comparison with the inoculum or the chicken samples. Some SNVs were present in all samples (inoculum, chickens, crows) at comparable frequencies, while the SNV occurring at position 3,298 of the genome dropped in frequency from an average of 26.79 % in the inoculum and chicken samples to an average of 11.47 % in the crow samples (Table 2).

In conclusion we observed that WNV populations passaged in chickens did not diversify significantly from the virus population of the inoculum. However, the WNV populations in the experimentally infected crows behaved more heterogeneously, and showed significant diversification away from the viral inoculum quasispecies structure.

DISCUSSION

Applying NGS technology, an average coverage of at least 700x was achieved in all samples, ensuring sensitive detection of low frequency variants. The relatively high background level of experimentally introduced error in the quasispecies populations (up to a threshold of 1.81 %) was predominantly attributed to the PCR amplification step, rather than the 454 sequencing step. Indeed, the WNV control plasmid (no PCR amplification) exhibited no

LoFreq-validated minor variants, while the PCR-amplified control resulted in minor variants as frequent as 1.81 %. Although reducing the number of PCR cycles or avoiding PCR amplification may significantly lower this cutoff, the required input DNA concentration of 500 ng for the Roche library preparation used in this study required PCR amplification. Nevertheless, even taking a conservative 1.81 % frequency threshold into account to discriminate meaningful variation from experimentally introduced variation, our data allowed a detailed documentation of the quasispecies dynamics, including the occurrence of new minor variants, the clearance of variants, and variant frequency modulation. Unlike the cloning and Sanger sequencing-based approaches in previous studies, this methodology combines complete genome characterization with high resolution analysis of WNV quasispecies populations.

Previous studies documenting WNV quasispecies (Deardorff *et al.*, 2011; Jerzak *et al.*, 2008) focused on *Culex* mosquitoes as a model for the arthropod vector, and young SPF chickens as a bird reservoir model. Mosquito hosts increased the WNV genetic diversity while birds – represented by the chicken model – acted as selective sieves. Given that young SPF chickens represent an artificial model host for WNV, we investigated WNV quasispecies dynamics in a natural susceptible wild bird, i.e., the Carrion crow. We found that after a single passage, WNV quasispecies populations were significantly more diverse in Carrion crows than in SPF chickens, despite higher average viral RNA loads in the sera of the latter compared to the former. In contrast, the conservation of the inoculum WNV quasispecies population in infected chickens and the overall lower genetic diversity of WNV populations in chickens confirmed previous findings in 2-day-old SPF chickens sampled on 2 dpi and analyzed by Sanger sequencing (Jerzak *et al.*, 2007). Our study represents the first NGS based quasispecies analysis of an experimentally infected WNV susceptible wild bird species. These findings suggest that, apart from the previously documented diversifying role of mosquitoes, wild bird species may also contribute to the genetic diversification of WNV during its natural life cycle.

Although the present study provides clear indications for WNV diversification in wild susceptible birds, our study design was limited by logistical and technical constraints, calling for a careful interpretation of these results. First, our approach identified genetic diversity in the inoculum that was experimentally derived from RNA transcripts representing a cloned virus. This genetic diversity may not be representative of the quasispecies population in the

parental virus. However, we believe our conclusions about the observed evolutionary processes in a single passage in a wild bird species remain valid.

Second, due to sensitivity constraints of the whole genome sequencing approach, we selected samples from the most viremic chickens to allow isolation of sufficient viral RNA for processing by NGS. This may have led to a bias. At 2 dpi, chickens experienced higher viral loads than corvids. However, this higher replication efficacy in chickens did not associate with a higher virus diversification rate.

Finally, we inoculated crows and chickens by two different routes. Chickens were inoculated by the IC route for the sake of comparability with our previous in vivo characterization of the Is98 virus (Dridi *et al.*, 2013), showing a better pathotype discrimination upon IC inoculation in day-old chickens. However, previous studies on WNV quasispecies dynamics used SC inoculation (Deardorff *et al.*, 2011; Jerzak *et al.*, 2007, 2008). For bioethical reasons, and to simulate a more natural infection route, crows were infected subcutaneously. Previous viral challenges in SPF chickens showed equivalent clinical (time to death) and virological (viremia and NS1 secretion kinetics) outcomes irrespective of the inoculation route (SC or IC). We cannot exclude a significant inoculation route effect on quasispecies dynamics as the primary replication sites upon SC vs. IC inoculation represent fundamentally different cell populations. For this reason, we refrain from any comparative conclusions between both infection models as host-specific and inoculation specific factors interfered in the experimental design. However, it should be noted that also previous studies comparing WNV diversity in different hosts suffered from these limitations. For instance, WNV diversity in SC-inoculated chickens and intrathoracically infected mosquitoes was compared (Deardorff *et al.*, 2011; Jerzak *et al.*, 2007, 2008).

Juvenile wild crows (aged almost one year) are expected to mount a full immune response (Smits & Bortolotti, 2008). Hosts immune responses to viral infections drive the emergence of new viral variants. This was shown in mosquitoes where intense RNAi targeting of particular regions of WNV genome associates with increased mutational diversity in these regions compared to more weakly targeted regions(Brackney et al., 2015).. Similarly, in acute hepatitis C virus (HCV) disease, early infection was shown to be first oligoclonal, increasing in quasispecies diversity in the later phase of infection (Manzin *et al.*, 1998). Moreover, in HIV disease, patients with high-titer neutralizing antibodies activity in serum show the

strongest positive selection pressure on the viral envelope glycoprotein early in infection (Bunnik *et al.*, 2008).

Moreover, the while the wild-caught Carrion crows used in this study are expected to exhibit heterogeneous genetic backgrounds. A high diversity in wild birds' genetic backgrounds, specifically if it affects genes involved in the immune response, might modulate the composition of an infecting viral quasispecies. For example, natural populations of great reed warblers (*Acrocephalus arundinaceus*) exhibit a high variability at MHC I loci (Westerdahl *et al.*, 2004), while the house finch (*Haemorhous mexicanus*) transcriptome of genes expressed in the spleen was shown to harbor approximately 85,000 SNPs, a subset of which has functions related to immune responses (Zhang *et al.*, 2014).

NS1 is a multifunctional viral glycoprotein that is highly conserved among flaviviruses (Coia *et al.*, 1988). Interestingly, most of WNV variants that were shared between all samples in this study occurred in the NS1-coding region. Moreover, 6/8 of these substitutions resulted in changes in the amino acid sequence. Previous studies relied on Sanger sequencing, which may explain why they did not detect low frequency genetic variants in NS1. The most abundant genetic variant in this study occurred at position 3,298 of our consensus WNV sequence, within the NS1 coding region. This non-synonymous Thr > Ala mutation contributed to 28.79 % of the viral population in chicken n° 1 (sampled on 2 dpi).Given that the frequency of this genetic variant decreased in crow-derived samples, one could speculate that selective pressure on this 3,298 locus is higher in crows compared to chickens. Host species differential tolerance to WNV mutations was previously evidenced as the modulation of the NS3 helicase activity by an asparagine deletion (position 483) was shown to lead to dramatically lower replicative fitness in chickens than in mosquitoes (Ebel *et al.*, 2011). Another example in our data is the SNV at position 9,716 of Is98 cloned genome which corresponds to a NS5 Val >Ala mutation and is present in 1.84% of the inoculum viral population. This viral subpopulation is cleared from all the chicken and crow samples, suggesting that given the RdRP function of NS5, this mutation might be detrimental to WNV replication fitness in avian hosts. Similarly, some subtle alterations of WNV RdRp affected WNV fitness in a host-dependent manner (Van Slyke et al., 2012).

Crows n°3 and 5, sampled on 2 and 4 dpi respectively, gave rise to the most frequent non-synonymous mutations overall (12/17 and 6/9 respectively), suggesting a relaxed purifying selection in those two birds. Although the primers used for WNV genome amplification did not allow for the complete sequence of the NCR, two mutations emerged in the 3'-NCR isolated from some avian samples: locus 10,312 in chicken n°3 and crows n°4 and 6 and locus

10,4256 in crow n°3, whereas no viral subpopulations could be detected in the 5'NCR of our samples. The conservation of the 5'-NCR length (constant 96 nt) compared to the 3'-NCR (337 to 649 nt depending on strains) was previously demonstrated (Brinton, 2013). On the other hand, the 3'- and 5'-NCRs fold into RNA secondary structures whose deletion is lethal for WNV infectious clones (Anthony *et al.*, 2009; Brinton & Dispoto, 1988; Brinton *et al.*, 1986; Deas *et al.*, 2005, 2007; Elghonemy *et al.*, 2005; Li *et al.*, 2010; Yu & Markoff, 2005). This suggests that the length of the 3'-NCR increases the chance that non-lethal mutations occur in this region, as we seem to observe in some of ours samples.

Previous research led to a paradigm whereby mosquitoes provide a source for viral genetic diversity while birds, as modeled by SPF chickens, act as selective sieves (Deardorff *et al.*, 2011; Jerzak *et al.*, 2008). It should be noted that studies on chickens mostly sampled serum, whereas mosquito samples analyzed represented a more diverse group of organs and tissues. Our findings imply that this paradigm might not be strictly applicable in nature. Indeed, the data we obtained with wild corvids suggest that, besides mosquitoes, at least some of the natural WNV-susceptible avian hosts provide a selective environment during WNV replication that can contribute to genetic diversification. This study shows that the commonly used young SPF chicken IC inoculation model for WNV characterisation may not fully reflect the complexity of WNV quasispecies modulation by its natural avian hosts following natural infection (mimicked by subcutaneous inoculation in the present study). An interesting extension of this work would be to check whether crow-introduced WNV mutations would be retained during further passage in a mosquito host, and whether further alternating passages between crows and mosquitoes would increase WNV infectivity. The application of NGS technologies to the analysis of WNV quasispecies dynamics as the virus cycles between arthropods and avian susceptible hosts might shed light on the mechanisms of WNV endemisation and pathogenic evolution.

MATERIAL AND METHODS

Virus and virus preparation

A wild type cloned virus previously derived (Bahuon et al., 2012) from lineage 1 clade a Is98 virus strain (IS-98 STD1, Genbank accession AF481864) was used in this study. The inocula were unpassaged (P0) as they were prepared by collecting tissue culture supernatant 5 days after electroporation of Vero cells (African green monkey kidney-derived cells, provided by

P. Desprès from the Pasteur Institute of Paris) with WNV genomic RNAs obtained through Sp6-driven *in vitro* transcription of the Is98 infectious clone plasmid. Tissue supernatants were clarified by centrifugation at 600 g for 5 min and stored in aliquots at - 80 °C.

Chickens

After hatching, SPF white Leghorn chickens (Lohmann Valo, Cuxhaven, Germany) were kept in biosecurity level 3 (BSL-3) isolators and animal experiments were conducted under the authorization and supervision of the Biosafety and Bioethics Committees at the Veterinary and Agrochemical Research Institute (VAR), following National and European regulations (procedure agreement no. 111202-01). The birds were provided with a commercial diet for poultry and water *ad libitum* throughout the experiments.

Carrion crows

Six juvenile (less than one year of age) Carrion crows were caught in a cereal-baited trap set close to a rubbish tip in central Belgium in April 2013. Ongoing serosurveillance in domestic and wild birds (including carrion crows) did not detect any flaviviral circulation in Belgium during 2012-2013. Birds were transferred to the VAR biosafety level 3 (BSL-3) facilities and kept in isolators located in a room lit 12 h/day. Isolators were kept at 20–22 °C and provided with perches and wood logs. Water and food (cereals and cat food) were provided *ad libitum.* *Crows* were acclimated to isolators 2–3 days before inoculation. Crows could be caught and held captive for scientific purposes after an *ad hoc* derogation granted by the Nature and Forests Department of the General Operational Direction of Agriculture, Natural Resources and Environment (Public Service of Wallonia). Animal experiments were conducted under the authorization and supervision of the Biosafety and Bioethics Committees at the VAR, following National and European regulations (procedure agreement no. 111202-02).

Experimental infection of birds

Twenty day-old SPF chickens were inoculated intracerebrally (IC) with 50 µl inoculum of cloned virus diluted in sterile PBS to 10^3 $TCID_{50}$. Six Carrion crows were inoculated subcutaneously (SC) with 100 µl inoculum of cloned virus diluted in sterile PBS to $10^{3.5}$ $TCID_{50}$. Disease symptoms were observed daily until 4 days post-infection (dpi). On 2 and 4 dpi, ten chickens and three crows were sacrificed. Blood was collected from each animal and allowed to clot for 6 h at room temperature (RT°). After centrifugation at 4000 rpm for 5 min, serum was collected and aliquots stored at -80°C for further use.

Viral RNA load quantification

RNA was extracted from sera samples and inoculum (P0 *Wt Pro* cloned virus) using a QIAamp Viral RNA Mini Kit (Qiagen GmBh, Hilden, Germany) according to the manufacturer's instructions and stored at -80°C. Two µl of each extract were subjected to rRT-PCR using the QuantiTect Probe RT-PCR Kit (Qiagen GmBh, Hilden, Germany), 0.15 µM final concentration of primers (WNNS2-F: 5'-CCTTTTCAGTTGGGCCTTCTG-3' and WNNS2a-R: 5'-GATCTTGGCTGTCCACCTCTTG-3') and0.2 µM final concentration of probe (WNNS2A-6-FAM TAMRA: 5'-TTCTTGGCCACCCAGGAGGTC-3'-TAMRA) in a final reaction volume of 25 µl (Dridi *et al.*, 2013). A reverse-transcription step at 50°C for 30 min was followed by 10 min at 95°C and 50 cycles of 15 sec at 95°C, 34 sec at 54°C and 10 sec at 72°C. Quantification of WNV was realized using standard curves ($R^2 > 0.99$) generated with previously titrated Is98 WNV seed. Only the three most viremic chickens sacrificed at 2 and 4 dpi were selected for the subsequent deep-sequencing analysis to ensure that enough genetic material was available.

WNV genome amplification

RNA was extracted from new aliquots of sera samples and inoculum (P0 cloned virus) using a QIAamp Viral RNA Mini Kit (Qiagen GmBh, Hilden, Germany) from a serum volume of 280 µl instead of 140 µl. RNA was reverse-transcribed using Superscript III Reverse Transcriptase (Invitrogen), 5 µl of extracted RNA, 300 ng of random hexamers, and 0.77 mM dNTPs in a final volume of 13 µl. After 5 min of annealing at 65 °C, 1 µl of 0.1 M DTT, 200 U of Superscript III RT, 4 µl of 5x First-Strand buffer and 1 µl of RNase out were added. Reverse transcription was performed at 25 °C for 5 min followed by 50 °C for 60 min and 70 °C for 15 min. Subsequently, 2 U of RNase H was added followed by incubation at 37°C for 20 min. The WNV genome was PCR amplified using 4 overlapping primer pairs and Platinum Taq DNA Polymerase High Fidelity (Invitrogen) according to the manufacturer's instructions and using 3 µl of cDNA, 20 pmol of each forward and reverse primer, and 2 – 2.5 mM MgSO4 (depending on the primer pairs). Primer pairs 1, 2 and 3 as well as the reverse primer 4 were previously described (Bahuon et al., 2012); forward primer 4 (5'- AAT GTC ACT GGA GGG CAT TC -3') was developed for this study. After denaturing for 1 min at 94 °C, 40 cycles of 15 s at 94 °C, 30 s at 60° C, and 3 min and 16 s at 68 °C were performed, with a final 10-min step at 68 °C. Amplification products (2.3 to 3.3 kb) were analyzed by 1.2 % agarose gel electrophoresis, purified using the High Pure PCR Product Purification Kit (Roche) and quantified using an ND-1000 spectrophotometer (Thermo Scientific).

Deep sequencing

Deep sequencing of the entire WNV genome was performed for the inoculum, the six crows and the six selected SPF chickens (Fig. 1). In order to determine the error rate of both the PCR and sequencing procedures, sequenced controls included the IS98 plasmid clone with or without an initial step of PCR amplification (Fig. 1). For each sample (except for the non-PCR plasmid clone), the four PCR fragments were equimolarly pooled in order to get a total amount of 500 ng of DNA/sample. Subsequently, sequencing libraries for the Genome Sequencer Junior (GS Junior; Roche) were prepared according to the manufacturer's instructions for Titanium Series reagents making use of multiplex identifiers (MID) to identify the different libraries. The resulting libraries were sequenced with a GS Junior with Titanium Series reagents and run protocol (200 cycles). In total, the 15 libraries were spread over 4 runs.

NGS Data analysis

Sequencing output files were sorted per sequencing library according to their MID sequences. Primer sequences were first trimmed if present at the end of the reads, then reads were trimmed on low quality (sliding window; QV threshold of 30) and length (threshold of 100 bp) using the Sickle tool (Joshi NA, 2011). Reads of the inoculum sample were mapped with gsMapper (Roche) against WNV Is98 strain reference genome (GenBank accession AF481864) in order to obtain a consensus sequence. Subsequently, reference assemblies were performed with gsMapper for all the other samples using the inoculum consensus genome. Only fully mapped reads were retained in order to avoid including chimeric sequences. To enable quantitative comparison, datasets were normalized so that each sample had the same number of reads by randomly sub-sampling to give 29,665 reads per sample. Subsampled reads were realigned against the consensus genome with gsMapper and single nucleotide variants (SNV) were then called from the BAM alignment file with LoFreq version 0.5 (Wilm *et al.*, 2012) with the recommended settings for RNA viruses (holmbonf strand bias filter and non-incorporation of mapping qualities). LoFreq was found to perform better in distinguishing errors from true viral variants on the plasmid control samples when compared to the variants called by gsMapper and those called by RC454 (Henn *et al.*, 2012) and V-Phaser 2 (Yang *et al.*, 2013) (Supplementary Information S1). The cut-off used to exclude errors caused by PCR amplification was set to 1.81 % as this percentage corresponded to the most abundant variant obtained with the PCR-amplified plasmid. Variants observed in the

plasmid PCR amplified control sample can be attributed to mutation errors generated from the PCR process and base miscalls from the sequencer. However, if a SNV (single nucleotide variant) was present in one sample at a proportion <1.81 %, but exceeded this threshold in another sample, then this SNV was considered as reliable despite being less than the threshold due to the polymorphic site being validated in other samples. Overall, the resulting variant list included 37 validated polymorphic loci across the genome.

Hierarchical Clustering and Pairwise Distances

Samples were hierarchically clustered based on the variant frequencies at all genome positions where a LoFreq approved variant was observed in any one sample, using the *pvclust* function in R 3.1.2 (R Core Team, 2014) with the standard Euclidean distance, complete linkage clustering and 1,000 bootstrap replicates. For visualization, a heatmap was generated by colouring variants based on their frequency, with the heatmap displayed beside the dendrogram to enable quasispecies similarity between samples to be evaluated.

The Euclidean distance between each pair of samples was calculated as the square root of the sum of the squared differences between variant frequencies over all genome positions where a LoFreq approved variant was observed in any one sample, using the *dist* function in R 3.1.2 (R Core Team, 2014). Two-sided Mantel permutation tests of between-group difference in mean pairwise distance were used to determine statistical significance. P-values were estimated by comparing the observed difference in means to the null distribution simulated using 10,000 permutations of the distance matrix.

ACKNOWLEDGEMENTS

This work was funded by Epi-SEQ: a research project supported under the 2nd joint call for transnational research projects by EMIDA ERA-NET (FP7 project nr 219235). The authors are grateful to Christophe Delgrange for assistance in the *in vivo* experiments, as well as to Orkun Ozhelvaci and Marc Boschmans for expert technical assistance. Acknowledgments are also due to Céline Bahuon for her generous gift of Is98 cloned virus and plasmid.

TABLE HEADINGS

Table 1: Individual viral loads in the inoculum and in sera of Carrion crows and SPF chickens infected with WNV. Cells highlighted in gray correspond to SPF chicken-derived serum samples selected for the deep-sequencing analysis. N.A.: not applicable.

Table 2: WNV genetic variants that were detected in the inoculum and in the sera of SPF chickens and Carrion crows. Blank cells correspond to loci where no minor genetic variants could be detected above the 1.81 % variant frequency cut-off. Cells highlighted in gray indicate non-synonymous mutations affecting the amino acid sequence. NCR: non-coding region; N.A.: not applicable. S: silent mutation. NS: non-synonymous mutation. Nucleotide positions are indicated relative to the polyprotein coding sequence of the inoculum consensus sequence (KR107956).

FIGURE AND SUPPLEMENTARY INFORMATION LEGENDS

Fig. 1: **Overview of WNV samples and sources of mutations.** A DNA plasmid containing a full-length WNV genome was used as a control sample to assess the amount of erroneous mutations introduced into the reads by the 454 sequencer alone. The plasmid was then PCR amplified prior to sequencing as a further control to assess the amount of error introduced by the PCR process in addition to the 454 sequencer. These 2 control samples were then used to set a frequency threshold for true viral variant detection. A total of 13 biological samples were sequenced: the inoculum used to infect all birds, 6 SPF chickens (3 sequenced at 2 days post infection [dpi] and 3 at 4 dpi), and 6 wild carrion crows (3 sequenced at 2 dpi and 3 at 4 dpi). A mutation observed in a read of a biological sample could be a real viral mutation or an erroneous mutation introduced during RT or PCR amplification or by 454 sequencer base miscalls. Therefore, only variants observed above the control sample threshold were considered to be real viral mutations enabling the investigation of viral quasispecies dynamics in different hosts.

Fig. 2: Hierarchical clustering of WNV quasipecies. The inoculum along with the 6 chicken and 6 crow samples were hierarchically clustered based on all observed polymorphic loci in the WNV genome. The heatmap displays the frequency of variants in each sample (y-axis) across each polymorphic loci (x-axis) using a colour scale ranging from red for the maximum frequency of ~30% to pale yellow at 0%. The dendrogram is displayed alongside the heatmap to show how samples cluster together. The numbers on branches represent the

Approximately Unbiased (AU) p-value followed by the Bootstrap Probability (BP) p-value in brackets, both represented as a %, from 1,000 bootstrap replicate.

Fig. 3: Dot-plot of pairwise Euclidean distance in mutation frequency between sample groups (**a**) and within sample group (**b**). P-values are from two-sided Mantel permutation tests of between-group difference in mean pairwise distance, indicated by horizontal lines (see text for details). Inoc: inoculum; dpi: days post-infection.

Supplementary information S1: Evaluation of different low frequency variant calling algorithms using a plasmid and PCR-amplified plasmid dataset.

REFERENCES

Anthony, K. G., Bai, F., Krishnan, M. N., Fikrig, E. & Koski, R. A. (2009). Effective siRNA targeting of the 3' untranslated region of the West Nile virus genome. *Antiviral Res* **82**, 166–168.

Bahuon, C., Desprès, P., Pardigon, N., Panthier, J. J., Cordonnier, N., Lowenski, S., Richardson, J., Zientara, S. & Lecollinet, S. (2012). IS-98-ST1 West Nile Virus Derived from an Infectious cDNA Clone Retains Neuroinvasiveness and Neurovirulence Properties of the Original Virus. *PLoS One* **7** (10), e47666.

Barzon, L., Lavezzo, E., Militello, V., Toppo, S. & Palù, G. (2011). Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci* **12**(11),7861-7884.

Van Borm S, Belák, S., G, F., A, F., F, G., D, H., DP, K., I, M., R, O. & T, R. (2015). Next-generation sequencing in veterinary medicine: how can the massive amount of information arising from high-throughput technologies improve diagnosis, control, and management of infectious diseases? *Methods Mol Biol* **1247**, 415–36.

Brackney, D. E., Beane, J. E. & Ebel, G. D. (2009). RNAi targeting of West Nile virus in mosquito midguts promotes virus diversification. *PLoS Pathog* **5**(7), e1000502.

Brinton, M. A. & Dispoto, J. H. (1988). Sequence and secondary structure analysis of the 5'-terminal region of flavivirus genome RNA. *Virology* **162**, 290–299.

Brinton, M. A., Fernandez, A. V & Dispoto, J. H. (1986). The 3'-nucleotides of flavivirus genomic RNA form a conserved secondary structure. *Virology* **153**, 113–121.

Brinton, M. A. (2013). Replication cycle and molecular biology of the west nile virus. *Viruses* **6**, 13–53.

Buckley, A., Dawson, A. & Gould, E. A. (2006). Detection of seroconversion to West Nile virus, Usutu virus and Sindbis virus in UK sentinel chickens. *Virol J* **3**, 71.

Bunnik, E., Pisas, L., van Nuenen, A. & Schuitemaker, H. (2008). Autologous neutralizing humoral immunity and evolution of the viral envelope in the course of subtype B human immunodeficiency virus type 1 infection. *J Virol* **82**, 7932–41.

Campbell, C. L., Keene, K. M., Brackney, D. E., Olson, K. E., Blair, C. D., Wilusz, J. & Foy, B. D. (2008). Aedes aegypti uses RNA interference in defense against Sindbis virus infection. *BMC Microbiol* **8**, 47.

Coia, G., Parker, M. D., Speight, G., Byrne, M. E. & Westaway, E. G. (1988). Nucleotide and complete amino acid sequences of Kunjin virus: Definitive gene order and characteristics of the virus-specified proteins. *J Gen Virol* **69**, 1–21.

Deardorff, E. R., Fitzpatrick, K. A., Jerzak, G. V. S., Shi, P. Y., Kramer, L. D. & Ebel, G. D. (2011). West nile virus experimental evolution in vivo and the trade-off hypothesis. *PLoS Pathog* **7** (11), e1002335.

Deas, T. S., Binduga-Gajewska, I., Tilgner, M., Ren, P., Stein, D. A., Moulton, H. M., Iversen, P. L., Kauffman, E. B., Kramer, L. D. & Shi, P.-Y. (2005). Inhibition of flavivirus infections by antisense oligomers specifically suppressing viral translation and RNA replication. *J Virol* **79**, 4599–4609.

Deas, T. S., Bennett, C. J., Jones, S. A., Tilgner, M., Ren, P., Behr, M. J., Stein, D. A., Iversen, P. L., Kramer, L. D. & other authors. (2007). In vitro resistance selection and in vivo efficacy of morpholino oligomers against west nile virus. *Antimicrob Agents Chemother* **51**, 2470–2482.

Dridi, M., Rauw, F., Muylkens, B., Lecollinet, S., van den Berg, T. & Lambrecht, B. (2013). Setting up a SPF Chicken Model for the Pathotyping of West Nile Virus (WNV) Strains. *Transbound Emerg Dis* **60**, 51–62.

Ebel, G. D., Fitzpatrick, K. A., Lim, P.-Y., Bennett, C. J., Deardorff, E. R., Jerzak, G. V. S., Kramer, L. D., Zhou, Y., Shi, P.-Y. & Bernard, K. A. (2011). Nonconsensus West Nile Virus Genomes Arising during Mosquito Infection Suppress Pathogenesis and Modulate Virus Fitness In Vivo. *J Virol* **85**(23), 12605-13.

Eigen, M. & Schuster, P. (1978). The Hypercycle - A principle of natural self-organization Part B: The abstract hypercycle. *Naturwissenschaften* **65**, 7–41.

Elghonemy, S., Davis, W. G. & Brinton, M. A. (2005). The majority of the nucleotides in the top loop of the genomic 3??? terminal stem loop structure are cis-acting in a West Nile virus infectious clone. *Virology* **331**, 238–246.

Henn, M. R., Boutwell, C. L., Charlebois, P., Lennon, N. J., Power, K. A., Macalalad, A. R., Berlin, A. M., Malboeuf, C. M., Ryan, E. M. & other authors. (2012). Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* **8**(3), e1002529.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982). Rapid evolution of RNA genomes. *Science* **215**, 1577–1585.

Hubálek, Z. & Halouzka, J. (1999). West Nile fever--a reemerging mosquito-borne viral disease in Europe. *Emerg Infect Dis* **5**, 643–650.

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

580 **Jerzak, G. V. S., Bernard, K., Kramer, L. D., Shi, P. Y. & Ebel, G. D. (2007).** The West Nile virus
581 mutant spectrum is host-dependant and a determinant of mortality in mice. *Virology* **360**,
582 469–476.

583 **Jerzak, G. V. S., Brown, I., Shi, P. Y., Kramer, L. D. & Ebel, G. D. (2008).** Genetic diversity and
584 purifying selection in West Nile virus populations are maintained during host switching.
585 *Virology* **374**, 256–260.

586 **Joshi NA, F. J. (2011).** Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ
587 files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

588 **Komar, N., Langevin, S., Hinten, S., Nemeth, N., Edwards, E., Hettler, D., Davis, B., Bowen, R. &**
589 **Bunning, M. (2003).** Experimental infection of North American birds with the New York 1999
590 strain of West Nile virus. *Emerg Infect Dis* **9**, 311–322.

591 **Li, X.-F., Jiang, T., Yu, X.-D., Deng, Y.-Q., Zhao, H., Zhu, Q.-Y., Qin, E.-D. & Qin, C.-F. (2010).** RNA
592 elements within the 5' untranslated region of the West Nile virus genome are critical for RNA
593 synthesis and virus replication. *J Gen Virol* **91**, 1218–1223.

594 **Manzin, A., Solforosi, L., Petrelli, E., Macarri, G., Tosone, G., Piazza, M. & Clementi, M. (1998).**
595 Evolution of hypervariable region 1 of hepatitis C virus in primary infection. *J Virol* **72**, 6271–
596 6276.

597 **Van Der Meulen, K. M., Pensaert, M. B. & Nauwynck, H. J. (2005).** West Nile virus in the
598 vertebrate world. *Arch Virol* **150**(4), 637-57

599 **Murray, K. O., Mertens, E. & Desprès, P. (2010).** West Nile virus and its emergence in the United
600 States of America. *Vet Res* **41**(6), 67.

601 **Pfeffer, M. & Dobler, G. (2010).** Emergence of zoonotic arboviruses by animal trade and
602 migration. *Parasit Vectors* **3**, 35.

603 **Phipps, L. P., Gough, R. E., Ceeraz, V., Cox, W. J. & Brown, I. H. (2007).** Detection of West Nile
604 virus in the tissues of specific pathogen free chickens and serological response to laboratory
605 infection: a comparative study. *Avian Pathol* **36**, 301–305.

606 **R Core Team. (2014).** R: A language and environment for statistical computing. *R Found Stat*
607 *Comput Vienna, Austria* URL http://www.R–project.org/.

608 **Senne, D., Pedersen, J., Hutto, D., Taylor, W., Schmitt, B. & Panigrahy, B. (2000).** Pathogenicity of
609 West Nile virus in chickens. *Avian Dis* **44**, 642–9.

610 **Shirafuji, H., Kanehira, K., Kubo, M., Shibahara, T. & Kamio, T. (2009).** Experimental West Nile
611 virus infection in aigamo ducks, a cross between wild ducks (Anas platyrhynchos) and
612 domestic ducks (Anas platyrhynchos var. domesticus). *Avian Dis* **53**, 239–244.

613 **Van Slyke, G. A., Jia, Y., Whiteman, M. C., Wicker, J. A., Barrett, A. D. T. & Kramer, L. D. (2013).**
614 Vertebrate attenuated West Nile virus mutants have differing effects on vector competence
615 in Culex tarsalis mosquitoes. *J Gen Virol* **94**, 1069–1072.

Smits, J. E. G. & Bortolotti, G. R. (2008). Immunological development in nestling American kestrels Falco sparverius: Post-hatching ontogeny of the antibody response. *Comp Biochem Physiol - A Mol Integr Physiol* **151**, 711–716.

Tang, Y., Liu, B., Hapip, C. A., Xu, D. & Fang, C. T. (2008). Genetic analysis of West Nile virus isolates from US blood donors during 2002-2005. *J Clin Virol* **43**, 292–297.

Totani, M., Yoshii, K., Kariwa, H. & Takashima, I. (2011). Glycosylation of the Envelope Protein of West Nile Virus Affects Its Replication in Chicks. *Avian Dis* **55**(4), 561-8.

Turell, M. J., O'Guinn, M. L., Jones, J. W., Sardelis, M. R., Dohm, D. J., Watts, D. M., Fernandez, R., Travassos da Rosa, A., Guzman, H. & other authors. (2005). Isolation of viruses from mosquitoes (Diptera: Culicidae) collected in the Amazon Basin region of Peru. *J Med Entomol* **42**, 891–898.

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344–348.

Villarreal, L. P. & Witzany, G. (2013). Rethinking quasispecies theory: From fittest type to cooperative consortia. *World J Biol Chem* **4**, 79–90.

Westerdahl, H., Wittzell, H., von Schantz, T. & Bensch, S. (2004). MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. *Heredity (Edinb)* **92**, 534–542.

Wheeler, S. S., Barker, C. M., Fang, Y., Armijos, M. V., Carroll, B. D., Husted, S., Johnson, W. O. & Reisen, W. K. (2009). Differential Impact of West Nile Virus on California Birds. *Condor* **111**(1),1-20.

Wilm, A., Aw, P. P. K., Bertrand, D., Yeo, G. H. T., Ong, S. H., Wong, C. H., Khor, C. C., Petric, R., Hibberd, M. L. & Nagarajan, N. (2012). LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* **40**, 11189–11201.

Yang, X., Charlebois, P., Macalalad, A., Henn, M. R. & Zody, M. C. (2013). V-Phaser 2: variant inference for viral populations. *BMC Genomics* **14**, 674.

Yu, L. & Markoff, L. (2005). The topology of bulges in the long stem of the flavivirus 3' stem-loop is a major determinant of RNA replication competence. *J Virol* **79**, 2309–2324.

Zhang, Q., Hill, G. E., Edwards, S. V & Backström, N. (2014). A house finch (Haemorhous mexicanus) spleen transcriptome reveals intra- and interspecific patterns of gene expression, alternative splicing and genetic diversity in passerines. *BMC Genomics* **15**, 305.