

Communication

Genetic Variability and Phylogeny of *European mountain ash ringspot-associated virus RNA3* and *RNA4*

Jenny Roßbach ¹, Heike Luisa Dieckmann ¹, Theresa Büttner ¹, Hans-Peter Mühlbach ², Susanne von Barga ^{1,*} and Carmen Büttner ¹

¹ Division Phytomedicine, Albrecht Daniel Thaer-Institute, Humboldt-Universität zu Berlin; Lentzeallee 55/57, 14195 Berlin, Germany; E-Mails: jenny.robel@hu-berlin.de (J.R.); dieckmann.luisa@yahoo.de (H.L.D.); Roestini@gmx.de (T.B.); carmen.buettner@agrar.hu-berlin.de (C.B.)

² Biocenter Klein Flottbek, Universität Hamburg, Ohnhorststraße 18, 22609 Hamburg, Germany; E-Mail: hans-peter.muehlbach@uni-hamburg.de

* Author to whom correspondence should be addressed; E-Mail: susanne.von.barga@agrar.hu-berlin.de; Tel.: +49-30-2093-46447; Fax: +49-30-2093-46450.

Academic Editors: Jan Stenlid, Jonas Oliva and Audrius Menkis

Received: 31 August 2015 / Accepted: 29 October 2015 / Published: 11 November 2015

Abstract: The *European mountain ash ringspot-associated virus* (EMARaV) is a multipartite RNA virus of negative polarity. It infects *Sorbus aucuparia* (common name—rowan) trees throughout their whole distribution area in North and Central Europe. It causes mottling, chlorotic ringspots and decline of the whole plant. Infected rowans are serious virus sources for rowans and other potential hosts. EMARaV incidence and population structure was investigated in Germany, Finland, Sweden, Scotland, and Norway. Overall, EMARaV variants from 42 rowan trees distributed in 20 different locations were studied with regard to the genetic variability of the p3- and p4-coding genome region, as well as the 5′ and 3′ untranslated regions (UTR) of RNA3. In six of the 42 analyzed samples we found much higher sequence diversities than previously reported at the amino acid level in RNA3 encoded p3 protein sequences as well as at the nucleotide level on the 5′ and 3′ UTR. The other 36 EMARaV variants confirmed the assumed conservation of the nucleocapsid protein coding region. In contrast, the p4-coding genome region showed a high conservation of both nucleotide and amino acid sequences. Both EMARaV proteins were under strong purifying selection pressure, probably acting to maintain the functional integrity of the p3 and p4 proteins.

Keywords: EMARaV; ss(-) RNA virus; European mountain ash; sequence variation

1. Introduction

Little is currently known about viruses affecting trees in forests and urban greenspaces. It is assumed that these viruses can cause significant economic damage [1] given their wide distribution. Chlorotic ringspots and mottling on the leaves of *Sorbus aucuparia* (European mountain ash, synonym rowan) are strictly correlated with infection by the *European mountain ash ringspot-associated virus* (EMARaV) [2,3], which has been observed in Germany since 1960 [4].

S. aucuparia is an important native founder for forestry, because it can grow in low nutrition areas and helps to settle contaminated soils [5]. Besides the native distribution of rowans, *S. aucuparia* or other species of the genus account for up to 20% of planted urban street trees in Nordic cities [6].

It can be assumed that infected rowans can be found in all European countries and several reports document EMARaV infection in Austria, the Czech Republic, Sweden, Finland, Germany, Russia, Norway, and the United Kingdom [7,8]. Symptomatic plants constitute a risk because they serve as the virus reservoir. In addition, host plants of EMARaV other than *S. aucuparia* were identified, such as different *Rosaceae* species and interspecies hybrids [9].

In general, transmission of plant viruses between host plants occur through vectors (insects, nematodes, fungi, mites), mechanically, vegetative as well as through pollen and seed [10]. For at least five *Emaravirus*-related viruses a vector transmission by eriophyid mites is described and transmission by grafting is documented for EMARaV [11]. The pear leaf blister mite (*Phytoptus pyri*) is discussed as a vector of this virus and might contribute to virus dispersal within the same species and transmission to other important members of the family *Rosaceae* such as apples or pears [12].

EMARaV is the type species in the genus *Emaravirus* [13,14]. Recent reports have assigned a number of negative-sensed RNA viruses with at least four genome segments to the genus *Emaravirus*, such as *Fig mosaic virus* (FMV), *Rose rosette virus* (RRV), *Raspberry leaf blotch virus* (RLBV) and *Pigeonpea sterility mosaic virus* (PPSMV) [15–18]. Other viruses, like *Wheat mosaic virus* (WMoV [19]), *Blackberry leaf mottle associated virus* (BLMaV [20]), *Pigeonpea sterility mosaic virus 2* (PPSMV-2 [21]), *Redbud yellow ringspot virus* (RYRV [22]) and viruses infecting *Arctium tomentosum* [23] or *Cordyline frutico* [24] are deemed as *Emaravirus*-related viruses but have not yet officially been accepted as a member of the genus. Elbeaino, *et al.* [25] showed that WMoV and RYRV are tentative emaraviruses by RT-PCR with RNA1 based genus-specific degenerated primer pairs and because they cluster with the assigned species.

The genome of EMARaV consists of four single stranded RNAs (RNA1–RNA4) of negative polarity [3]. Each segment encodes a single protein (p1–p4), which is translated from the complementary strand (Figure 1). The RNA1 encodes a RNA-dependent RNA polymerase (RdRP), RNA2, a glycoprotein precursor, which is processed into two glycoproteins, and RNA3 codes for the nucleocapsid protein (p3). The function of RNA4-encoded p4 protein of EMARaV is so far unknown. Because RNA2 and RNA3 code for proteins that function as a glycoprotein precursor and a nucleocapsid protein, respectively, it is assumed that EMARaV p4 is a non-structural virus protein.

The p4 proteins of other emaraviruses are members of the 30K movement protein (MP) family [26]. However, these p4 proteins are considerably different in sequence and length when compared to EMARaV p4 [13], and the authors concluded that EMARaV p4 is not an orthologous protein [26]. Therefore, the p4 protein of EMARaV is a unique protein within the genus *Emaravirus* and the function remains to be characterized.

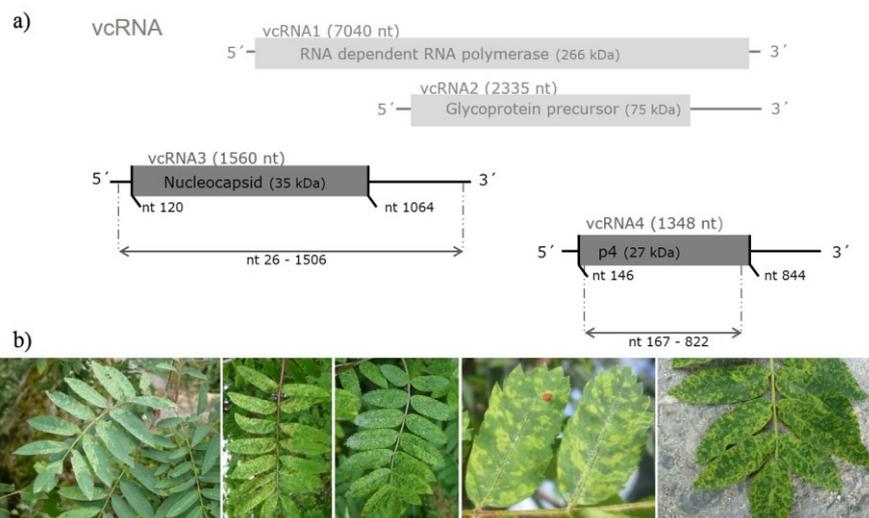


Figure 1. (a) Schematic representation of viral copy RNAs (vcRNA) of the EMARaV genome and nucleotide (nt) position of analyzed genome regions (grey arrows); (b) Leaves of European mountain ash with typical symptoms of EMARaV infection such as chlorotic ringspots and mottling found in different locations at Germany, Scotland, Norway, Sweden and Finland.

It has been shown for some time that plant viruses are characterized by a high mutation rate and, hence, a high level of genetic variation. García-Arenal *et al.* [27], in a meta-review, concluded that the high potential for variation in plant viruses caused by mutation, genetic exchange by recombination, or reassortment of genomic fragments need not necessarily result in high genetic diversity in virus populations. Selection factors, such as the interaction with a specific host or vector, may reduce their genetic diversity [27]. Detailed knowledge of factors forcing genetic variability could be instrumental in disease control [28].

For EMARaV, genetic variability of the nucleocapsid protein was analyzed in sixteen infected rowans from Finland and four samples from Russia (eastern Karelia) [29,30]. The studies showed high conservation of the protein. The analysis of small fragments of all four genome components of ten Swedish EMARaV variants also presented little genetic diversity. None of the genome components demonstrated a higher genetic variability when compared to the others [31]. However, the existence of two different RNA3 variants in one virus preparation was shown for the Wheat mosaic virus (WMoV, [19]). This raises new questions about genetic variability of virus variants from different locations.

In this study, we determined the genetic variability of the EMARaV structural nucleocapsid protein and the non-structural p4 protein. We selected samples from Sweden, Finland, Norway, Scotland and Germany to take the wide distribution area of this virus into account. Samples from 20 representative locations in five European countries were analyzed and directly compared.

2. Experimental Section

Forty-two rowans with chlorotic ringspots and mottling (Figure 1b) were selected from 20 locations in five different countries (Table 1) in order to analyze their genetic diversity. Additionally, a possible correlation of diversity with geographical proximity was assessed.

Table 1. Forty-two *European mountain ash ringspot-associated virus* (EMARaV) variants and the related accession numbers are listed according to the geographic origin of the sampled *S. aucuparia* plants.

	EMARaV Variants	Location of Sampled Rowans	RNA3	RNA4
Sweden (S)	E51605	Lulea	HG799707	HG799749
	E51586	Pitea	HG799704	HG799746
	E51587	Pitea	HG799705	HG799747
	E51594	Örnsköldsvik	HG799706	HG799748
	E53016	Skärssa	HG799711	HG799753
	E53011	Östa	HG799710	HG799752
	E53009	Heby	HG799709	HG799751
	E52165	Stockholm, Vasa	HG799708	HG799750
Finland (FIN)	E51607	Rovaniemi	HG799712	HG799754
	E52278	Rovaniemi	HG799713	HG799755
	E52279	Rovaniemi	HG799714	HG799756
	E52280	Rovaniemi	HG799715	HG799757
Scotland (GB)	E52284	Ullapool Hill	HG799717	HG799759
	E52286	Ullapool Hill	HG799718	HG799760
	E52287	Corrieshalloch	HG799719	HG799761
	E52283	Dunvegan	HG799716	HG799758
	E52288	Kinlochleven	HG799720	HG799762
Germany (D)	E52895	Benz	HG799732	HG799774
	E52991	Hamburg	HG799739	HG799781
	E52992	Hamburg	HG799740	HG799782
	E52993	Hamburg	HG799741	HG799783
	E52994	Hamburg	HG799742	HG799784
	E52995	Hamburg	HG799743	HG799785
	E52996	Hamburg	HG799744	HG799786
	E52997	Hamburg	HG799745	HG799787
	E51609	Berlin	HG799730	HG799772
	E52897	Berlin	HG799733	HG799775
	E52900	Berlin	HG799734	HG799776
	E52901	Berlin	HG799735	HG799777
	E52905	Berlin	HG799736	HG799778
	E52907	Berlin	HG799737	HG799779
	E52990	Berlin	HG799738	HG799780
	E52293	Fichtelgebirge	HG799731	HG799773

Table 1. Cont.

	EMARaV Variants	Location of Sampled Rowans	RNA3	RNA4
Norway (N)	E53111	Mo i Rana	HG799721	HG799763
	E53112	Mo i Rana	HG799722	HG799764
	E53113	Mo i Rana	HG799723	HG799765
	E53114	Mo i Rana	HG799724	HG799766
	E53116	Mosjoen	HG799725	HG799767
	E53117	Mosjoen	HG799726	HG799768
	E53118	Formofossen	HG799727	HG799769
	E53119	Formofossen	HG799728	HG799770
	E53120	Skatval	HG799729	HG799771

Genetic variability of EMARaV was investigated by comparison of the p3 encoding region of RNA3 plus 5' and 3' UTR, as well as the RNA4 sequence encoding the non-structural protein p4 (Figure 1). Nearly the complete RNA3 was amplified by three overlapping but independent RT-PCRs using primers described by Kallinen *et al.* [29]. An additional primer pair (NC177f: tta gta aat ata ata tgt aca t/NC810r: gga tat gca tca taa agg aa) was designed to verify deletions found in the 3' UTR of some samples in a second PCR, independent of the previously used primers. With these primers, a fragment of 635 bp was amplified in PCR. The p4-coding region of RNA4 was amplified with two overlapping p4-specific primer sets (p4f: atg gag tcc aac aag atg aag/p4-375r: aca tat tct tct cca taa ata agg; p4-245f: cta ttg gtg aat cca gga tgt t/p4r: tca ttc cat ttg gtc tga tga t). For PCR, a proof-reading polymerase (Velocity, Bionline, London, UK) was used. PCR products were directly sequenced using a “BigDye Terminator v1.1 Ready Reaction Cycle Sequencing Kit” and an ABI PRISM 310 Genetic Analyzer from Applied Biosystems (Carlsbad, NM, USA). Sequences were edited, aligned and primer sequences removed using BioEdit (Ibis Biosciences, Carlsbad, NM, USA) [32]. Sequences reported in this study have been deposited in GenBank and have been assigned the accession numbers displayed in Table 1. Diversity among EMARaV variants was studied using the similarity matrix with function IDENTIFY in the program.

Phylogenetic analyses were carried out using two different methods: (i) A distance method using a neighbor-joining algorithm in ClustalX [33] was applied. Phylogenetic trees of nucleotide and amino acid sequences were drawn in NJplot [34]; (ii) We analyzed the aligned sequences with the character-based method of Maximum Likelihood within MEGA5.2 [35]. Kimura-2-parameter was used for nucleotide sequences and amino acid sequences were analyzed with the Poisson model for amino acid substitutions. The grouping robustness was determined after 1000 pseudo-replicates. Branches supported by less than 50% were discarded as they were considered untrustworthy [36]. Reference sequences from Finland (RNA3: EU885293, EU885287, EU885289), Russia (RNA3: GU563319), and the EMARaV type strain from Hamburg, Germany (RNA3: DQ831831, and RNA4: DQ831828) were included in the constructed trees.

We further examined the type of selection pressures acting on the p3 and p4 protein to better understand evolutionary mechanisms that might affect genetic variability. Selection pressures acting on the coding regions of RNA3 and RNA4 were calculated using the Synonymous/Non-Synonymous Analysis Program (SNAP, Korber [37]) based on the Nei-Gojobori method [38]. Selection pressures

on a single codon were also analyzed with the same algorithm by using the codon based Z-test on MEGA5.2. The dN/dS ratio reflects the relative influence of positive selection and neutral evolution [39]. Protein coding areas are under positive, diversifying selection when dN/dS is >1, neutral when dN/dS is =1, or negative, purifying selection when dN/dS is <1. The dN/dS ratio was calculated pairwise for the different virus variants and then averaged separately for the p3 and p4 coding regions.

3. Results and Discussion

3.1. Sequence Variability of EMARaV RNA3 and the Deduced p4 Protein

Nucleotide sequences of the p3 coding region (945 nt) were 95.3% to 100% identical for 36 of the 42 investigated variants from Scotland, Sweden, Germany, Norway, and Finland (Table 2) as shown by pairwise comparison. Amino acid sequences of the respective samples were 97.4% to 100% identical (Figure S1). Similar studies addressing genetic variability in Finland, Russia and Sweden also revealed high identities between the investigated EMARaV variants [29–31]. Seventeen EMARaV variants from Finland displayed diversities up to 3% within the p3 coding region and 6% diversity in the 3' UTR [29]. The sequence identities of the 36 samples found in the presented study support the assumption of Kallinen *et al.* [29] and von Bargen *et al.* [31] that coevolution processes between the virus, the host and the vector may reduce the genetic variability of the nucleocapsid protein of EMARaV.

In contrast, six out of nine Norwegian EMARaV variants (E53111–E53114, E53119–E53120), which showed identities at the p3 level similar to the variants described above (95.3%–100%), exhibited only sequence identities of 85.7% at the nucleotide level and 92.6%–94.9% at the amino acid level compared to all the other variants (Table 2). Nucleotide substitutions which lead to the lower genetic identity were distributed over the complete RNA3. Every EMARaV RNA3 variant was amplified by three independent overlapping PCR fragments. Substitutions were proven in at least two independent PCR products and by four sequence reactions. The PCR was conducted using a proofreading polymerase with a 50-fold higher fidelity than *Thermus aquaticus* DNA polymerase [40] in order to achieve higher accuracy. It should be noted further that this sequence variation is found in six independently sampled and processed European mountain ashes. Overall, we observed 191 nucleotide substitutions in the EMARaV p3 ORF (945 nt), the major part being 91 nucleotide substitutions that were found in the Norwegian variants. However, most of these nucleotide substitutions were silent mutations. Fifty amino acid substitutions were equally distributed within the p3. Twenty-eight of these exchanges were solely found in the Norwegian virus variants. In the case of the 17 Finnish EMARaV variants, Kallinen *et al.* [29] found only three amino acid substitutions within the nucleocapsid coding region. The effect of the substitutions, including non-conservative exchanges in the nucleocapsid protein of EMARaV, on functionality cannot be predicted because the functional domains of the p3 have not yet been determined. For Wheat mosaic virus (WMoV), another putative *Emaravirus*, two RNA3 sequence variants of different lengths were found within one RNA preparation from purified viruses [19]. These two sequence variants differed significantly (12.5% and 11.1%) on the nucleotide and on the amino acid level, respectively. Tatineni *et al.* [19] excluded the possibility of a mixed infection and found that both RNA3 species were present in a 5:1 ratio. It is still unclear why

WMOV acquired two different RNA3 variants. The analysis of the chromatograms of the sequenced PCR products in our study showed that the RNA3 variants were not different in any sample. Each of the six Norwegian variants with lower identity values were found in individual plants. Similar results are described by Laney *et al.* [15] who investigated the genetic diversity of *Rose rosette virus* (RRV) and found up to 7% sequence diversity among the nucleocapsid or the movement protein (MP) in two of 22 isolates from different U.S. states. Thus, lower identities of the nucleocapsid protein coding region of individual isolates seem to be common in the genus *Emaravirus*, but a possible evolutionary advantage is unknown.

In addition to the p3 coding region, we examined the identities of 445 nucleotides of the 3' UTR region of RNA3 (nucleotide position 1.065–1.510). Most samples from Germany, Finland, Sweden and Scotland showed identities between 93.7% and 100% (Figure S2). The EMARaV variants from Norway (E53111–E53114, E53119, and E53120), however, exhibited notably lower identities between 69.6% and 77.5% (Table 2). This result is in contrast to the comparable low 3' UTR diversity reported by Kallinen *et al.* [29] and von Bargen *et al.* [31] that did not exceed 6%. Further, several nucleotide deletions and one insertion in this genome region contributed to the reported higher diversity. This was proofed by an additional PCR of this genome region with the primer pair NC177f/NC810r. The EMARaV variants E53111–E53114, E53119, and E53120 exhibited nucleotide deletions at position 1.065, 1.115, 1.145–1.152, and 1.362. The variants E53111–E53114 showed an insertion of guanine at position 1.227, whereas variants E53119–E53120 contained an adenine insertion. Comparison of the 94 nucleotides of the 5' UTR of RNA3 confirmed genetic variability, as found in the p3 coding region and in the analyzed part of the 3' UTR (Table 2). Deletions in the 3' UTR were also the reason for the different lengths of the two WMOV variants [19]. It can be assumed that the renunciation of a few selected amino acids does not alter the functional integrity for virus replication or protein translation of the emaraviruses. The comparison of both untranslated regions showed that the shorter 5' UTR of the different variants reveals higher identities than the 3' UTR. However, in the 5' UTR lower levels of nucleotide sequence, identities can also be observed for the divergent Norwegian EMARaV variants (84% to 91.4%) as compared to the other 36 EMARaV variants (95.7% to 100%, Figure S3).

Interestingly, Grimová *et al.* [9] recently reported high identities of over 97.1% of the nucleocapsid protein in EMARaV variants from interspecies hybrids in the Czech Republic. This finding indicates that the host plant species does not have an important effect on the variability of the nucleocapsid protein. In light of these findings, it is improbable that small genetic changes in rowan hybrids were the reason for the deviation in the six divergent Norwegian variants from rowan seen in this study, which exhibited significantly lower values of sequence identity. So far, the biological factors influencing the variability are still unknown.

In contrast to the variability observed in the RNA3, we found a strict conservation of the p4 protein in all investigated EMARaV samples (Figure S4). Sequence analysis of the partial p4 coding region revealed identities between 97.4% and 100% for most variants on the amino acid level (Table 2). Nevertheless, it is notable that the highest differences were exhibited by eight of the nine samples from Norwegian trees and were located in the central part of the protein (Figure S5). Here, the EMARaV amino acid sequence identities reached a minimum of 96.7% (Table 2). This fact supports the separate grouping based on RNA3 sequence.

Table 2. Data compilation of sequence analyses using the similarity matrix with function IDENTIFY in BioEdit [32]. Nucleotide or amino acid sequences of two genome regions (RNA3 and p4 protein) of the different EMARaV variants were compared with each other and the values of the identity matrices were summarized (full tables can be found in the supplementary material). GB: Scotland, S: Sweden, FIN: Finland, D: Germany, N: Norway.

Genome Region	Compared EMARaV Variants	Nucleotide Sequence Identity		Amino Acid Sequence Identity	
		Variants from GB, S, FIN, D, and 3 from N (E53116–E53118)	6 of 9 Variants from N (E53111–E53114, E53119–E53120)	Variants from GB, S, FIN, D, and 3 from N (E53116–E53118)	6 of 9 Variants from N (E53111–E53114, E53119–E53120)
RNA3 5' UTR (94 nt)	33 (42) variants from GB, S, FIN, D and 3 from N (E53116–E53118)	95.7% to 100%	84% to 91.4%	-	-
	6 (9) variants from N (E53111–E53114, E53119, E53120)	84% to 91.4%	96.8% to 100%	-	-
p3 (945 nt; 315 aa)	33 (42) variants from GB, S, FIN, D and 3 from N (E53116–E53118)	95.3% to 100%	85.7% to 86.9%	97.4% to 100%	92.6% to 94.9%
	6 (9) variants from N (E53111–E53114, E53119, E53120)	85.7% to 86.9%	98.5% to 100%	92.6% to 94.9%	97.1% to 100%
RNA3 3' UTR (435–445 nt)	33 (42) variants from GB, S, FIN, D and 3 from N (E53116–E53118)	93.7% to 100%	69.6% to 77.5%	-	-
	6 (9) variants from N (E53111–E53114, E53119, E53120)	69.6% to 77.5%	90.1% to 100%	-	-
p4 coding region, primer sequences removed (656 nt; 218 aa)	33 (42) variants from GB, S, FIN, D and 3 from N (E53116–E53118)	97.4% to 100%	97.5% to 100%	97.2% to 100%	96.7% to 99.5%
	6 (9) variants from N (E53111–E53114, E53119, E53120)	97.5% to 100%	97.4% to 100%	96.7% to 99.5%	97.2% to 99.5%

For the non-structural MP (p4) in 21 out of 22 *Rose rosette virus* isolates, a similarly high conservation was shown at the nucleotide as well as at the amino acid level (97%–100% identity) by Laney *et al.* [15]. In addition, for 583 analyzed nucleotides from 37 *Fig mosaic virus* RNA4 variants, a low genetic variation is reported [41]. Therefore, the p4 protein of the *Emaravirus* genus could have high sequence identities to preserve the functionality of the protein, but it is not clear whether the p4 protein of EMARaV is comparable with the p4 proteins of other genus members, which serve as movement proteins. The deviating secondary structure and low sequence identity of EMARaV p4 as compared to p4 proteins of other emaraviruses [16] indicate that it is not an orthologous protein [26]. Therefore, the function of RNA4 encoded p4 is still unknown. However, the high conservation of the p4 protein of EMARaV is indicative of its importance, since proteins that play key roles in virus function are generally conserved through strong selective constraints. Thus, in virus replication or transmission, p4 might play an essential role. This was, for instance, shown for *Tomato spotted wilt virus* (TSWV) [42], which is distantly related to EMARaV.

Consequently, we presume that p4 protein conservation in EMARaV is required to preserve its full functional integrity. Further studies addressing the function of p4 and the identification of the related functional domains are necessary to shed light on the role in EMARaV pathogenicity. Our observations indicate that a more variable core is framed by a conserved N- and C-termini to preserve functional amino acids in these domains. In the p4 coding region, we found 53 variable positions within 656 analyzed nucleotides. Amino acid substitutions were observed in 22 cases. These substitutions were partially non-conservative. Ten of these substitutions occurred between nucleotide positions 108 and 165, marking this region as a hot spot for substitution of amino acids (see supplementary material for schematic representation). This study contributes to information about conserved and variable parts as well as to the characterization of functional domains of the p4 protein.

3.2. Phylogenetic Relationships Based on p3 and p4 Protein Sequences

Results obtained by pairwise comparison of sequences from the two different genomic segments of EMARaV were confirmed by phylogenetic studies of the respective regions of RNA3 and RNA4. The analysis of RNA3 showed that the divergent Norwegian variants (E53111–E53114, E53119, and E53120) always clustered separately (Figure 2a,b). This result was independent (i) of the method (distance or character based) used for calculating and (ii) of whether the coding region of RNA3 (Figure 2a) or the 3' UTR (Figure 2b) was examined. The separate cluster existed in 100% of the bootstrap replicates. The other cluster contained two subgroups: one included the variants E53116–E53118 from Norway and E52283–E53284, E53286–E53288 from Scotland and the second consisted of the remaining samples from Germany, Sweden, and Finland. Kallinen *et al.* [29] and Valkonen and Rännäli [30] found three independent clusters for EMARaV nucleocapsid protein by analysis of nucleotide sequence evolution. In our investigations, we involved one member of each cluster. We analyzed them together with our 42 EMARaV variants to figure out the evolutionary distance between the different variants. The incorporated variants cluster with the second subgroup. The divergent Norwegians seem to be split off much earlier from this cluster.

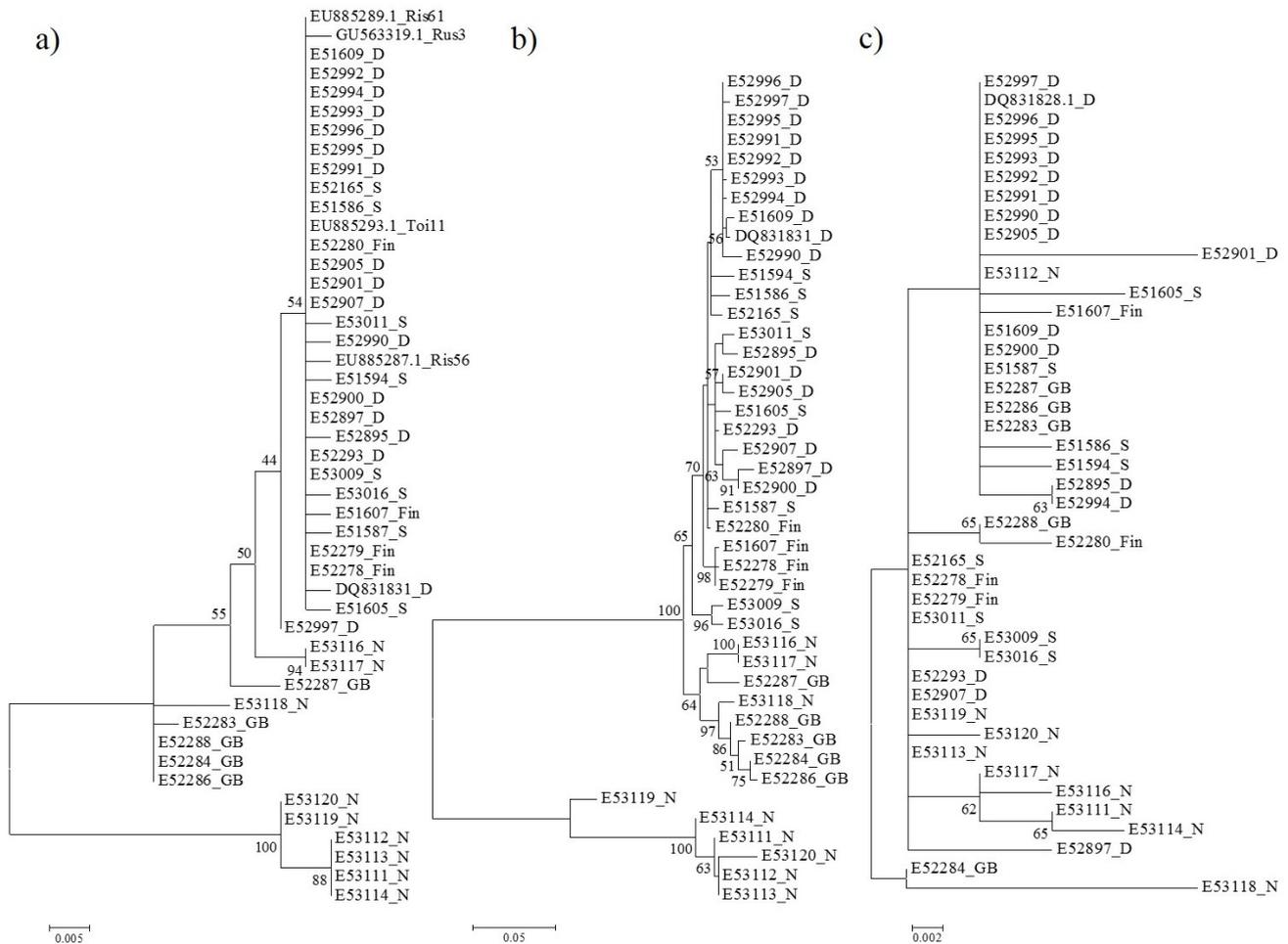


Figure 2. Phylogenetic analysis using nucleotide or amino acid sequences of RNA 3' UTR or p3 and p4 coding region. **(a)** Phylogenetic tree of amino acid sequence analysis of the nucleocapsid protein (p3) coding region; **(b)** Phylogenetic tree of nucleotide sequence analysis of RNA3 3' UTR; **(c)** Phylogenetic tree of amino acid sequence analysis of p4 coding region. The evolutionary history was analyzed with the maximum likelihood method based on the Kimura-2-parameter model for nucleotide sequences and the Poisson correction model for amino acid sequences, respectively. The percentage of replicated trees in which the associated taxa clustered together with 1000 replicates from bootstrapping is shown next to the branches. Bootstrap values lower than 50% were hidden. Forty-two variants from this study, the virus type isolate from Hamburg-Germany (GenBank RNA3: DQ831831 = NC_013108 and RNA4: DQ831828 = NC_013107) and four published RNA3 sequences from Finland (EU885293.1_Toi11, EU885287.1_Ris56, EU885289.1_Ris61) and Russia (GU563319.1_Rus3), which represent the three known phylogenetic clusters were included in this analysis. D: Germany, GB: Scotland, Fin: Finland, N: Norway, S: Sweden.

As expected, the results obtained from pairwise comparison of sequences of p4 coding regions did not display different clusters with reliable Bootstrap values. P4 sequences of all EMARaV variants formed one big group. Any clustering of the p4 sequences was statistically not supported, either on the nucleotide or on the amino acid level (Figure 2c).

3.3. Selection Pressure on EMARaV Nucleocapsid (p3) and p4 Protein

As an indicator of protein selection pressure, we used the ratio of non-synonymous substitutions per non-synonymous site (d_N) over synonymous substitutions per synonymous sites (d_S) of the coding regions. Both the p3 protein ($d_N/d_S = 0.050$) (Table 3) and the p4 protein ($d_N/d_S = 0.116$) were under negative selection pressure. For viruses of other genera d_N/d_S ratios between 0.013 (*Rice tungro spherical virus*) and 0.257 (*Potato virus V*) of the structural coat protein are reported. Proteins which are involved in the viral replication process show values between 0.095 (*Yam mosaic virus*) and 0.308 (*Citrus tristeza virus*) [27]. As in the case of EMARaV, a strong purifying selection pressure acts on all of these proteins. This seems necessary to maintain the functional integrity of these proteins. A purifying selection pressure may stabilize the virus population and prevent deleterious mutations from dispersing within the population.

Table 3. Synonymous and non-synonymous substitution rates of p3 and p4 coding region of European mountain ash ringspot-associated virus.

	d_N	sd (d_N)	sd (d_N)/ d_N	d_S	sd (d_S)	sd (d_S)/ d_S	d_N/d_S
p3	0.013	0.007	0.492	0.266	0.034	0.128	0.050
p4	0.004	0.001	0.333	0.032	0.008	0.247	0.116

d_N : number of non-synonymous substitutions per non-synonymous site, averaged for all EMARaV variants;
 d_S : number of synonymous substitutions per synonymous site, averaged for all EMARaV variants;
 sd: standard deviation; d_N/d_S is the ratio of non-synonymous substitutions to synonymous substitutions. If $d_N/d_S > 1$, the gene is under positive selection, whereas if $d_N/d_S < 1$ the gene is under purifying selection.

In comparing the number of non-synonymous mutations, we noted a four times higher d_N value for the p3 protein than for the p4 protein (values normalized to the total length of sequence). The standard deviation (sd) of d_N and d_S is a measure of the variability of the pairwise calculated d_N and d_S values. The standard deviation of the average of the p3 protein was higher than that of the p4 protein. Consequently, the diversity in sequence variability was higher in the p3 than in the p4 protein. In summary, these results comprise the findings of identity analyses. The divergent Norwegian virus variants (E53111–E53114, E53119–E53120) showed a lower genetic identity to all the other variants for both p3 and p4. The d_N , d_S , sd, d_N/d_S values are mainly driven by these Norwegian virus variants. The strong negative selection pressure—or purifying selection—for p3 and p4 implies that both proteins are strongly conserved to maintain functional integrity.

Recently, von Bargen, *et al.* [31] showed that coding region fragments of RNA1, RNA2 and RNA4 in Swedish EMARaV variants are under strong purifying selection. They could not confirm that any of the investigated proteins were more variable than the other. For the Finnish EMARaV p3 variants, Kallinen *et al.* [29] reported a d_N/d_S ratio of 0.0123 that was a quarter of our result (0.050). As mentioned above, this discrepancy is most probably due to the higher frequency of amino acid substitutions within the divergent EMARaV variants from Norway. Walia *et al.* [41] documented for FMV higher d_N values for small RNA1 (367 nt) and RNA2 (572 nt) fragments than for RNA3 and RNA4. They concluded that the complete RNA1 (7093 nt) and RNA2 (2252 nt) had higher sequence variabilities. However, this is not the case for EMARaV, as we found a five times higher d_N value of RNA3 as compared to the d_N value calculated for fragments of RNA1 and RNA2 by von Bargen, *et al.* [31]. Thus,

genetic conservation of individual genome segments appears to differ between the members of the genus *Emaravirus*.

3.4. Genetic Diversity in the Nucleocapsid Protein of Norwegian Isolates

García-Arenal *et al.* [27] mentioned that more populations of plant viruses are genetically stable than variable and that population bottlenecks, such as virus adaptation to a vector or a host, maintain the genetic structure, thus keeping virus populations stable. However, our phylogenetic analysis of the nucleocapsid protein and the related 3' UTR identified a separate cluster of some EMARaV variants in Norway suggesting that genetic drift events affected this population.

EMARaV has been reported to infect *Sorbus aucuparia* [7] and interspecies *Sorbus* hybrids naturally [9]. The gall mite, *Phytoptus pyri*, is assumed to be the only vector [12]. Based on studies of Swedish EMARaV variants, the authors suggested a co-evolution between the gall mite and the virus, which acts as a bottleneck for viral genome conservation [31]. Indeed, galls were observed on leaves of *Sorbus aucuparia* trees from Norway, which were infected by the EMARaV variants (E53112–E53114). However, there is presently no proof that these galls are caused by *Phytoptus pyri* and it cannot be ruled out that vectors other than *Phytoptus pyri* are involved in transmission of EMARaV. The Norwegian virus variants may have adapted to different vectors, which might explain the separate cluster.

Besides vector transmission, other mechanisms, such as infection of new host plants or plant species and colonization of new geographical areas, can result in genetic drift. Note that variants E53118 and E53119 were collected at the same location in Formofossen (Norway) but exhibit significantly different RNA3 sequences. This is consistent with reports from Kallinen *et al.* [29] and von Bargaen *et al.* [31], who could not find a correlation between geographical distribution and genetic variability of the analyzed EMARaV variants. Virus variants with high genetic diversity are found at the same location. In contrast, virus variants with low genetic diversity have been collected from locations hundreds of kilometers apart from each other [15,29]. Interestingly, it is reported that the analyzed 261 nucleotides of the RNA2, which encodes the glycoprotein precursor—also a structural protein of the virus particle—are highly conserved (97%–98% identity) for all nine Norwegian EMARaV variants [8]. More sequence information of the Norwegian EMARaV RNA1 and RNA2 is required to assess the genetic variability of the whole genome. Biological factors leading to the new genetic EMARaV RNA3 variant from Norway need to be identified. Knowledge about these factors will be initial steps towards implementation of effective disease management [28].

4. Conclusions

This is the first study examining the population structure of EMARaV variants from *Sorbus aucuparia* distributed throughout Northern Europe. Leaf samples of rowans from 20 different sites in five different European countries were examined. The nucleocapsid p3 protein and the p4 protein, a non-structural protein encoded by another genomic segment of EMARaV, were investigated. The function of the p4 protein is unknown and previous investigations did not suggest a role in virus movement, as reported for other emaraviruses.

We obtained evidence for a purifying selection pressure on the structural p3 and the non-structural p4 proteins. This implies an intense adaptation of EMARaV to the host (*Sorbus aucuparia*) and to the putative vector (*Phytoptus pyri*). One may also assume that both proteins play a key role in the “life” cycle of the virus, including replication, assembly, movement or interaction with the vector. We found that the genetic structure of EMARaV is not as uniform as expected based on previous studies. Six divergent EMARaV variants from rowans in Norway were identified with a significantly lower nucleocapsid sequence identity as compared to currently known variants. In contrast, analyses of the p4 coding region and of a short fragment of the RNA2 [8] of the divergent Norwegian EMARaV variants showed high sequence identities to all other variants.

Even a single infected rowan tree could play a leading role as a source for virus transmission to other members of *Rosaceae* or other rare native *Sorbus* species. This hypothesis is supported by investigations of Grimová *et al.* [9], who identified new host plants for EMARaV. Thus, possible new modes of transmission have to be taken into consideration. For instance, adaptation to another vector could explain the nucleotide substitutions of the six divergent EMARaV variants. Knowledge about the biological properties which lead to the variability in one genome component is needed.

Follow up studies should address the functional analyses of EMARaV gene products and their involvement in the virus-host-vector interaction. The identification of functional protein domains and pathogenicity determinants are required to assess whether the observed variability of RNA3 indicates the existence of more than one EMARaV strain exhibiting different biological features.

Acknowledgments

We thank Risto Jalkanen from Natural Resources Institute Finland LUKE for providing some samples from Northern Sweden. The project was financially supported by grants from the German Research Foundation (DFG, BU 890/27-1 and MU 559/13-1).

Author Contributions

Susanne von Bargen and Jenny Roßbach conceived and designed the experiments; Heike-Luisa Dieckmann, Theresa Büttner and Jenny Roßbach performed the experiments; Susanne von Bargen and Jenny Roßbach analyzed the data; Hans-Peter Mühlbach and Carmen Büttner as initiators contributed significantly to data interpretation and critically reviewed the manuscript; Jenny Roßbach wrote the paper in close cooperation with Susanne von Bargen.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Büttner, C.; von Bargen, S.; Bandte, M.; Mühlbach, H.-P. Forest diseases caused by viruses. In *Infectious Forest Diseases*; Gonthier, P.N.G., Ed.; CABI: Wallingford, Oxfordshire, UK, 2013; pp. 50–75.

2. Benthack, W.; Mielke, N.; Büttner, C.; Mühlbach, H.P. Double-stranded RNA pattern and partial sequence data indicate plant virus infection associated with the ringspot disease of European mountain ash (*Sorbus aucuparia* L.). *Arch. Virol.* **2005**, *150*, 37–52.
3. Mielke, N.; Mühlbach, H.P. A novel, multipartite, negative-strand RNA virus is associated with the ringspot disease of European mountain ash (*Sorbus aucuparia* L.). *J. Gen. Virol.* **2007**, *88*, 1337–1346.
4. Kegler, H. Das Ringfleckenmosaik der Eberesche (*Sorbus aucuparia* L.). *J. Phytopathol.* **1960**, *37*, 214–216.
5. Hamberg, L.; Malmivaara-Lämsä, M.; Lehvävirta, S.; Kotze, D. The effects of soil fertility on the abundance of rowan (*Sorbus aucuparia* L.) in urban forests. *Plant Ecol.* **2009**, *204*, 21–32.
6. Sæbø, A.; Benedikz, T.; Randrup, T.B. Selection of trees for urban forestry in the Nordic countries. *Urban For. Urban Green.* **2003**, *2*, 101–114.
7. Robel, J.; Bandte, M.; Mühlbach, H.-P.; von Bargen, S.; Büttner, C. Ein neuartiges virus in *Sorbus aucuparia* L.: Nachweis und Verbreitung des *European mountain ash ringspot-associated virus* (EMARaV). In *Jahrbuch der Baumpflege*; Dujesiefken, D., Ed.; Haymarket Media: Braunschweig, Germany, 2013; pp. 47–53.
8. Von Bargen, S.; Büttner, T.; Mühlbach, H.P.; Robel, J.; Büttner, C. First Report of *European mountain ash ringspot-associated virus* in *Sorbus aucuparia* in Norway. *Plant Disease* **2014**, *98*, 700–700.
9. Grimová, L.; Marek, M.; Konrady, M.; Ryšánek, P. Newly identified host range of *European mountain ash ringspot-associated virus* (EMARaV) and its distribution in the Czech Republic. *For. Pathol.* **2015**, *45*, 177–189.
10. Stevens, W.A. Transmission of plant viruses. In *Virology of Flowering Plants*; Stevens, W.A., Ed.; Springer: New York, NY, USA, 1983; pp. 41–68.
11. Führling, M.; Büttner, C. Transmission experiments of viruses to woody seedlings (*Quercus robur* L. and *Sorbus aucuparia* L.) by grafting and mechanical inoculation. *Eur. J. For. Pathol.* **1995**, *25*, 129–135.
12. Mielke-Ehret, N.; Thoma, J.; Schlatermund, N.; Mühlbach, H.P. Detection of *European mountain ash ringspot-associated virus*-specific RNA and protein P3 in the pear leaf blister mite *Phytoptus pyri* (Eriophyidae). *Arch. Virol.* **2010**, *155*, 987–991.
13. Mielke-Ehret, N.; Mühlbach, H.-P. *Emaravirus*: A novel genus of multipartite, negative strand RNA plant viruses. *Viruses* **2012**, *4*, 1515–1536.
14. Mühlbach, H.-P.; Mielke-Ehret, N. *Emaravirus*. In *Virus Taxonomy*, 9th ed.; King, A.M.Q.; Adams, M.J.; Carstens, E.B.; Lefkowitz, E.J., Eds.; Elsevier: Oxford, UK, 2011; pp. 767–770.
15. Laney, A.G.; Keller, K.E.; Martin, R.R.; Tzanetakis, I.E. A discovery 70 years in the making: Characterization of the Rose rosette virus. *J. Gen. Virol.* **2011**, *92*, 1727–1732.
16. McGavin, W.J.; Mitchell, C.; Cock, P.J.; Wright, K.M.; Macfarlane, S.A. Raspberry leaf blotch virus, a putative new member of the genus *Emaravirus*, encodes a novel genomic RNA. *J. Gen. Virol.* **2012**, *93*, 430–437.
17. Elbeaino, T.; Digiario, M.; Martelli, G.P. Complete nucleotide sequence of four RNA segments of fig mosaic virus. *Arch. Virol.* **2009**, *154*, 1719–1727.

18. Elbeaino, T.; Digiario, M.; Uppala, M.; Sudini, H. Deep sequencing of pigeonpea sterility mosaic virus discloses five RNA segments related to emaraviruses. *Virus Res.* **2014**, *188*, 27–31.
19. Tatineni, S.; McMechan, A.J.; Wosula, E.N.; Wegulo, S.N.; Graybosch, R.A.; French, R.; Hein, G.L. An eriophyid mite-transmitted plant virus contains eight genomic RNA segments with unusual heterogeneity in the nucleocapsid protein. *J. Virol.* **2014**, *88*, 11834–11845.
20. Hassan, M.; Sabanadzovic, S.; Keller, K.; Martin, R.; Tzanetakis, I. A putative new emaravirus associated with blackberry yellow vein disease. In Proceedings of the 22nd International Conference on Virus and Other Transmissible Fruit Crop Diseases, Rome, Italy, 3–8 June 2012; pp. 360–364.
21. Elbeaino, T.; Digiario, M.; Uppala, M.; Sudini, H. Deep sequencing of dsRNAs recovered from mosaic-diseased pigeonpea reveals the presence of a novel *Emaravirus*: Pigeonpea sterility mosaic virus 2. *Arch. Virol.* **2015**, *160*, 2019–2029.
22. Laney, A.G. *Characterization of the Causal Agents of Rose Rosette and Redbud Yellow Ringspot*; University of Arkansas: Fayetteville, NC, USA, 2010.
23. Bi, Y.; Tugume, A.K.; Valkonen, J.P.T. Small-RNA deep sequencing reveals *Arctium tomentosum* as a natural host of *Alstroemeria virus X* and a new putative *Emaravirus*. *PLoS ONE* **2012**, *7*, e42758, doi:10.1371/journal.pone.0042758.
24. Melzer, M.; Sugano, J.; Uchida, J.; Kawate, M.; Borth, W.; Hu, J. Partial characterization of a novel emara-like virus from *Cordyline fruticosa* (L.) with ti ringspot disease, In Proceedings of 2014 APS-CPS Joint Meeting, Minneapolis, MA, USA, 9–13 August 2014; pp. 79–79.
25. Elbeaino, T.; Whitfield, A.; Sharma, M.; Digiario, M. *Emaravirus*-specific degenerate PCR primers allowed the identification of partial RNA-dependent RNA polymerase sequences of Maize red stripe virus and Pigeonpea sterility mosaic virus. *J. Virol. Methods* **2013**, *188*, 37–40.
26. Yu, C.; Karlin, D.G.; Lu, Y.; Wright, K.; Chen, J.; MacFarlane, S. Experimental and bioinformatic evidence that *Raspberry leaf blotch emaravirus* P4 is a movement protein of the 30K superfamily. *J. Gen. Virol.* **2013**, *94*, 2117–2128.
27. García-Arenal, F.; Fraile, A.; Malpica, J.M. Variability and genetic structure of plant virus populations. *Annu. Rev. Phytopathol.* **2001**, *39*, 157–186.
28. Acosta-Leal, R.; Duffy, S.; Xiong, Z.; Hammond, R.W.; Elena, S.F. Advances in plant virus evolution: Translating evolutionary insights into better disease management. *Phytopathology* **2011**, *101*, 1136–1148.
29. Kallinen, A.K.; Lindberg, I.L.; Tugume, A.K.; Valkonen, J.P. Detection, distribution, and genetic variability of *European mountain ash ringspot-associated virus*. *Phytopathology* **2009**, *99*, 344–352.
30. Valkonen, J.P.T.; Rännäli, M. First Report of *European mountain ash ringspot-associated virus* in *Sorbus aucuparia* from Eastern Karelia, Russia. *Plant Disease* **2010**, *94*, 921–921.
31. Von Barga, S.; Arndt, N.; Robel, J.; Jalkanen, R.; Büttner, C. Detection and genetic variability of *European mountain ash ringspot-associated virus* (EMARaV) in Sweden. *For. Pathol.* **2013**, *43*, 429–432.
32. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp.* **1999**, *41*, 95–98.

33. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948.
34. Perrière, G.; Gouy, M. WWW-query: An on-line retrieval system for biological sequence banks. *Biochimie* **1996**, *78*, 364–369.
35. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739.
36. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **1985**, *39*, 783–791.
37. Korber, B. HIV signature and sequence variation analysis. In *Computational and Evolutionary Analysis of HIV Molecular Sequences*; Rodrigo, G.A., Learn, G.H., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2000; pp. 55–72.
38. Nei, M.; Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **1986**, *3*, 418–426.
39. Gonzales, M.J.; Dugan, J.M.; Shafer, R.W. Synonymous-non-synonymous mutation rates between sequences containing ambiguous nucleotides (Syn-SCAN). *Bioinformatics* **2002**, *18*, 886–887.
40. Fujii, S.; Akiyama, M.; Aoki, K.; Sugaya, Y.; Higuchi, K.; Hiraoka, M.; Miki, Y.; Saitoh, N.; Yoshiyama, K.; Ihara, K.; *et al.* DNA replication errors produced by the replicative apparatus of *Escherichia coli*. *J. Mol. Biol.* **1999**, *289*, 835–850.
41. Walia, J.J.; Willemsen, A.; Elci, E.; Caglayan, K.; Falk, B.W.; Rubio, L. Genetic variation and possible mechanisms driving the evolution of worldwide *Fig mosaic virus* isolates. *Phytopathology* **2014**, *104*, 108–114.
42. Tsompana, M.; Abad, J.; Purugganan, M.; Moyer, J.W. The molecular population genetics of the *Tomato spotted wilt virus* (TSWV) genome. *Mol. Ecol.* **2005**, *14*, 53–66.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).