



Published in final edited form as:

Vaccine. 2009 March 10; 27(11): 1685–1690. doi:10.1016/j.vaccine.2009.01.036.

Relaxation of Purifying Selection on Live Attenuated Vaccine Strains of the Family *Paxamyxoviridae*

Austin L. Hughes

Department of Biological Sciences, University of South Carolina, Columbia SC 29208

Abstract

In wild-type sequences of three paramyxoviruses (measles virus, mumps virus, and Newcastle disease virus), nucleotide diversity at both non-coding sites and at nonsynonymous sites in coding regions was significantly reduced in comparison to that at synonymous sites. Likewise, both the mean and variance of gene diversity at nonsynonymous polymorphic sites were reduced in comparison to non-coding and synonymous sites. Neither of these patterns, which reflect the action of purifying selection against deleterious mutations at nonsynonymous and non-coding sites, was seen in the case of live attenuated vaccine strains, implying that purifying selection has been substantially relaxed on the latter, potentially affecting their biological properties, including antigenicity and vaccine effectiveness. Since the accumulation of mutations increases as a function of the number of generations of replication, these findings highlight the utility of minimizing the number of generations between the original vaccine master seed and the strains used in vaccination, along with periodic monitoring of the extent of sequence evolution.

Keywords

live attenuated vaccine; measles virus; mumps virus; Newcastle disease virus; paramyxovirus; purifying selection

1. Introduction

Beginning with Jenner's use of cowpox virus as a vaccination for smallpox over 200 years ago and Pasteur's discovery of attenuation a century later, the use of live attenuated vaccines has become an important weapon in the fight against a number of viral diseases [1]. Live attenuated vaccines have the advantage over killed-virus vaccines that the former resemble a natural infection and thus interact with all aspects of the host's immune system; but there are challenges associated with the use of live vaccines [2]. One set of challenges arises from the danger that a live attenuated virus can cause disease in some recipients. For example, in a minority of cases, vaccination with live attenuated virus has caused aseptic meningitis [3]. Likewise, in the case of poliovirus, there is evidence that the live attenuated vaccine virus has mutated in such a way as to regain virulence [4]. Another set of challenges relates to transmission of the live vaccine

Address correspondence to: Austin L. Hughes, Ph.D., Department of Biological Sciences, University of South Carolina, Coker Life Sciences Bldg., 700 Sumter St., Columbia SC 29208 USA, Tel: 1-803-777-9186, Fax: 1-803-777-4002, Email: E-mail: austin@biol.sc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

from a healthy person to one who is immunocompromised [5]. Moreover, studies of live attenuated vaccines often reveal incomplete vaccine effectiveness [6-9].

Live attenuated viruses continue to mutate and thus to evolve. Particularly in the case of RNA viruses, with their inherently high mutation rate, this evolution can be rapid. Evolution of the live vaccine strains can lead to the reacquisition of a virulent phenotype, and evolutionary changes may potentially play a role in the loss of vaccine effectiveness. Because of the importance of these phenomena in the management of vaccination programs, it is important to understand the population processes underlying the evolutionary changes in live attenuated vaccine strains of viruses.

Sequence differences between vaccine strains and their wild-type relatives include those responsible for the attenuated phenotype, which may have been discovered serendipitously [10] or may have arisen after a deliberate series of passages designed to result in attenuation [11]. In addition, there may be certain sequence changes in live attenuated viruses that have been selected because they better adapt the virus to growth under culture conditions [12]. However, all genomes are exposed to deleterious mutations, which natural selection (conservative or “purifying” selection) constantly acts to eliminate [13]. Under the conditions in which live attenuated viruses are cultured, it might be expected that purifying selection will be substantially relaxed, because these viruses are not exposed to many of the selection pressures acting on a free-living virus; for example, the selection pressures arising from natural virus transmission from host to host. In addition, because viruses in culture are expected to have much lower effective population sizes than free-living viruses, it can be expected that selection will be much less effective in removing slightly deleterious mutations in the former than in the latter [14].

Evidence regarding the action of natural selection on protein-coding regions can be obtained by estimating the number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous (amino acid-altering) substitutions per nonsynonymous site (d_N). In most protein-coding regions, d_S substantially exceeds d_N [15]. This pattern occurs because most nonsynonymous mutations are deleterious to protein function and thus are eliminated by purifying selection [15]. Furthermore, analysis of coding sequences reveals two distinct aspects of purifying selection: (1) past selection against strongly deleterious mutations, which are eliminated quickly; and (2) ongoing selection against slightly deleterious mutations, which may be relatively inefficient if the effective population size is small or recombination is limited [14,16]. The presence of slightly deleterious nonsynonymous variants, subject to ongoing purifying selection, can be detected by comparing the population frequency patterns of variants at synonymous and nonsynonymous sites [17-20].

The present study uses statistical analysis of published sequences to test the hypothesis that purifying selection has been relaxed in live attenuated vaccine strains. These analyses were applied to three single-stranded RNA negative sense viruses belonging to the viral family *Paramyxoviridae*: (1) measles virus (MeV; genus *Morbillivirus*), a pathogen of humans; (2) mumps virus (MuV; genus *Rubulavirus*), a pathogen of humans; and (3) Newcastle disease virus (NDV; genus *Avulavirus*), a pathogen of domestic poultry. Applying phylogenetic methods, sequence changes ancestral to clades of vaccine sequences were reconstructed; and patterns of nucleotide sequence polymorphism were compared between samples of wild-type sequences (i.e., those collected from natural infections) and live vaccine sequences.

2. Methods

2.1. Sequences Analyzed

Complete genome sequences of measles virus (MeV), mumps virus (MuV), and Newcastle disease virus (NDV) were aligned using the CLUSTAL X program [21]. Phylogenetic trees were constructed separately for the complete genomes of each of the three viruses by the neighbor-joining (NJ) method [22] on the basis of the maximum composite likelihood (MCL) distance [23]. The reliability of clustering patterns in phylogenetic trees was tested by bootstrapping [24]; 1000 bootstrap pseudo-samples were used. Accession numbers of sequences used in phylogenetic analyses are shown in Figures 1-3. The NDV sequence AY562985, which was shown to be a naturally occurring recombinant by previous phylogenetic analyses [25], was not included in the phylogenetic analysis; however, AY562985 was included in population genetic analyses, which unlike phylogenetic analyses did not depend on the assumption of an absence of recombination between sequences.

For each of the three viruses, genomic sequences used in analyses were chosen to represent both wild-type isolates – i.e., those collected from natural infections – and vaccine strains. The latter were chosen so that they represented a single clade in each of the viral phylogenies, and that clade received 100% bootstrap support in each case (Figures 1-3). In the case of MeV (Figure 1), vaccine sequences used in analyses all belonged to the Edmondston lineage [26]. For MuV (Figure 2), only vaccine sequences belonging to the Urabe lineage [27] were used; sequences from the Jeryl Lynn vaccine strains were not used because fewer complete genomes were available for the latter. In the case of NDV, naturally isolated strains are characterized as velogenic (highly pathogenic), mesogenic (intermediate), or lentogenic (apatogenic; ref. 28). In analyses of NDV (Figure 3), the vaccine sequences used belonged to a clade including both the B1 and LaSota strains. These two vaccine strains were isolated independently from naturally occurring lentogenic strains [10, 29]. Lentogenic strains from outside this clade that have been used for vaccination (e.g., Ulster and V4; ref. 30) were not included in analyses because few genomic sequences were available.

In each of the three viruses, the genome encodes seven major proteins (listed in 5' to 3' order): NP; V/P; M; F, HN; and L. MuV also encodes SH between F and HN. In MeV and MuV, the V/P gene expresses alternative products with the help of RNA editing. For the purposes of analyzing the pattern of nucleotide diversity in coding sequences as described below, the coding sequence of the longest non-edited open reading frame was analyzed.

2.2. Statistical Analyses

The number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous (amino acid-altering) substitutions per nonsynonymous site (d_N) were estimated by Nei and Gojobori's [31] method. Estimation of these quantities by more complex methods [32,33] yielded essentially identical results, as is expected because the number of substitutions per site was generally low [34]; therefore, we used Nei and Gojobori's method because it has a lower variance [35]. The synonymous nucleotide diversity (symbolized π_S) is defined as the mean of d_S for all pairwise comparisons among a set of sequences, while the nonsynonymous nucleotide diversity (symbolized π_N) is the mean of d_N for all pairwise comparisons among a set of sequences. The Jukes-Cantor method [35] was used to estimate the number of nucleotide substitutions per site (d) in non-coding regions; the mean of all pairwise comparisons of d is the nucleotide diversity (π). Ancestral nucleotide sequences were reconstructed by the maximum parsimony method [36] on the basis of the NJ trees. In each case, the tree was highly consistent, with consistency indices of 0.925 for MeV, 0.905 for MuV, and 0.612 for NDV. By comparison of reconstructed ancestral sequences, d_S , d_N , and d were estimated along branches within the tree. In each phylogeny, there was a wild-type sequence

that formed a sister group to the vaccine clade: AF266288 in MeV (Figure 1); AY309060 in MuV (Figure 2); and DQ060053 in NDV (Figure 3). Mean d_S , d_N , and d were estimated between the members of each vaccine clade and the wild-type sister. The standard errors of mean numbers of nucleotide substitutions, including both nucleotide diversity within sets of sequences and comparisons of between sets of sequences, were estimated by the bootstrap method implemented in the MEGA 4.0 program [23].

Gene diversity (“heterozygosity”) at individual polymorphic nucleotide sites (SNP sites) was estimated by the formula:

$$1 - \sum_{i=1}^n x_i^2$$

where n is the number of alleles and x_i is the population frequency of the i^{th} allele (ref. 37, p. 177). In coding regions, single nucleotide polymorphisms were classified either as synonymous or nonsynonymous depending on their effect of the encoded nucleotide sequence. Ambiguous sites were excluded from these analyses; i.e., sites at which both synonymous and nonsynonymous variants occurred or at which the polymorphism could be considered synonymous or nonsynonymous depending on the pathway taken by evolution. Gene diversity was not normally distributed. Therefore, in testing for differences in the mean and variance of gene diversity among categories (synonymous, nonsynonymous, and non-coding) of SNP sites, randomization tests were used. In each test, 1000 pseudo-data sets were created by sampling (with replacement) from the data; a difference between two categories was considered significant at the α level if it was greater than the absolute value of $100(1-\alpha)$ % of the differences observed between the same categories in the pseudo-data sets.

3. Results

3.1. Nucleotide Diversity

In the wild-type isolates of MeV, MuV, and NDV, both nonsynonymous nucleotide diversity (π_N) and nucleotide diversity in non-coding regions (π) were significantly lower than synonymous nucleotide diversity (π_S ; Table 1). In wild-type sequences of all three viruses, the ratio $\pi_N : \pi_S$ was less than 1:10 (Table 1). This pattern is evidence of past purifying selection both on the amino acid sequence of encoded proteins and on non-coding regions. By contrast, in the vaccine sequences from MeV and MuV, neither π_N nor π in non-coding regions significantly lower than synonymous nucleotide diversity (π_S ; Table 1). The ratio $\pi_N : \pi_S$ was over seven times as high in the vaccine sequences of MeV as in wild-type MeV sequences (Table 1). Similarly, the ratio $\pi_N : \pi_S$ was over 10 times as high in the vaccine sequences of MuV as in wild-type MuV sequences (Table 1).

In the case of NDV, the vaccine sequences showed a distinctive pattern (Table 1). In the coding regions, π_N was significantly lower than π_S , although the ratio $\pi_N : \pi_S$ was greater (0.17) than that seen in wild-type NDV (0.07; Table 1). On the other hand, π in the non-coding regions did not differ significantly from π_S in the case of NDV vaccine strains (Table 1). However, when nucleotide diversities were computed separately for the B1 and LaSota strains, although the numbers of sequences were small, the pattern was similar to that seen in vaccine strains of MeV and MuV (Table 1). For comparisons within B1 and LaSota strains, neither π_N nor π in non-coding regions significantly lower than synonymous nucleotide diversity (π_S ; Table 1). Likewise, the ratios $\pi_N : \pi_S$ computed within B1 and LaSota, were much greater than those for wild-type NDV (Table 1).

3.2. Ancestral Sequence Comparisons

The pattern of nucleotide substitution along the branch ancestral to the clade of vaccine sequences was examined within each of the three viruses. In MeV, d_N was actually greater than d_S on the ancestral branch, although the difference was not significant due to the small number of substitutions (Table 2). In MuV, d_S was slightly greater than d_N on the ancestral branch, but the difference was not significant (Table 2). In both MeV and MuV, d in non-coding regions on the ancestral branch was slightly but not significantly greater than d_S (Table 2). NDV showed a striking difference from the other two viruses, with both d_N in coding regions and d in non-coding regions significantly lower than d_S and on the ancestral branch (Table 2).

In the case of MeV, mean d_N did not differ significantly from mean d_S in comparisons between the members of the vaccine clade and a wild-type sister sequence (Table 2). Likewise, mean d in non-coding regions of MeV did not differ significantly from mean d_S in coding regions in comparisons between the vaccine clade and wild-type sister (Table 2). Likewise in MuV, mean d_N was slightly but not significantly greater than mean d_S in comparisons between the members of the vaccine clade and the wild-type sister sequence (Table 2). On the other hand, mean d in non-coding regions of MuV was actually significantly greater than mean d_S in coding regions in comparisons between the vaccine clade and wild-type sister (Table 2). As with the pattern of reconstructed sequence evolution on the ancestral branch, NDV differed from the other two viruses in the comparison with the nearest wild-type sister sequence, with both d_N in coding regions and d in non-coding regions significantly lower than d_S (Table 2).

3.3. Analysis of SNPs

Gene diversity was computed at individual polymorphic sites in order to test for ongoing purifying selection. In the wild-type populations of MeV, MuV, and NDV, both mean and variance of gene diversity were significantly lower at nonsynonymous SNP sites than at synonymous SNP sites or at non-coding SNP sites (Table 3). In the case of the vaccine strains of MeV and MuV, there was not a significance difference in mean or variance of gene diversity between nonsynonymous SNPs and synonymous or non-coding SNPs (Table 3). In the case of NDV, by contrast, the vaccine strains, like wild-type strains, showed significantly lower mean and variance of gene diversity at nonsynonymous SNPs than at either synonymous or non-coding SNPs (Table 3).

In coding regions of MeV, of 1182 sites that were polymorphic in wild-type sequences but not in vaccine sequences, 296 (25.0 %) were nonsynonymous. By contrast, of 123 sites that were polymorphic in vaccine sequences but not in wild-type sequences, 83 (67.5%) were nonsynonymous. The difference in proportions was highly significant ($\chi^2 = 97.36$; 1 d.f.; $P < 0.001$). There was a similar pattern in the case of MuV, although the number of polymorphic sites in the vaccine sequences was small. Of 1413 coding sites polymorphic in the wild-type but not the vaccine MuV sequences, 210 (14.9%) were nonsynonymous. But 13 of 17 (76.5%) coding sites polymorphic in vaccine but not in wild-type sequences were nonsynonymous.

In MeV, 25 coding sites that were polymorphic in both wild-type and vaccine sequences, 13 (52.0%) were nonsynonymous. The latter proportion differed significantly from that found among coding sites polymorphic only in wild-type sequences ($\chi^2 = 9.34$; 1 d.f.; $P < 0.002$). In the case of MuV, there were only 6 coding sites (only one of which was nonsynonymous) that were polymorphic in both wild-type and vaccine sequences; so comparison with sites polymorphic in wild-type sequences only was not meaningful.

Of coding sites polymorphic in wild-type but not vaccine sequences of NDV, 1255 of 4455 (28.2 %) were nonsynonymous. By contrast, 54 of 82 (65.9 %) coding sites polymorphic in vaccine but not wild-type NDV sequences were nonsynonymous. The difference in proportions

was highly significant ($\chi^2 = 55.70$; 1 d.f.; $P < 0.001$). In this respect, NDV resembled MeV and MuV. On the other hand, of 111 coding sites polymorphic in both wild-type and vaccine sequences of NDV, only 19 (17.1%) were nonsynonymous. The latter proportion was actually lower than that in coding sites polymorphic in wild-type but not vaccine sequences of NDV (Figure 7A); and the difference was statistically significant ($\chi^2 = 6.58$; 1 d.f.; $P = 0.01$).

4. Discussion

Analysis of the patterns of nucleotide sequence polymorphism in wild-type isolates of three species of *Paramyxoviridae* showed reduced nucleotide diversity at nonsynonymous sites in comparison to synonymous sites in coding regions, as is seen in most genes [15]. This pattern is evidence of the action of purifying selection against a substantial fraction of nonsynonymous mutations. In addition, nucleotide diversity in non-coding regions was reduced in comparison to synonymous sites in coding regions, implicating purifying selection on the former. In vaccine strains of measles virus (MeV) and mumps virus (MuV), there was no evidence of reduced nucleotide diversity at either nonsynonymous sites in coding regions or at noncoding sites. Thus, there was evidence that a striking characteristic of the vaccine strains of these viruses is relaxation of purifying selection.

Examination of gene diversity at polymorphic nucleotide sites showed that, in the case of wild-type MeV and MuV, both the mean and variance of gene diversity at nonsynonymous SNP sites was reduced in comparison to that at either synonymous SNP sites or those in non-coding regions (Table 3). This pattern is indicative of ongoing purifying selection on nonsynonymous sites because it shows that there are few nonsynonymous polymorphisms occurring at high frequency [8]. Again, no such pattern was seen in the case of the vaccine strains of MeV and MuV (Table 3). Since the polymorphic variants currently found in a vaccine strain must have arisen after the most recent common ancestor of that strain, it is significant that there is evidence of a relaxation of purifying selection against such newly arisen variants. Thus, the relaxation of purifying selection did not simply occur during the attenuation process that gave rise to the vaccine strains of MeV and MuV, but rather it is ongoing.

The results obtained for Newcastle disease virus (NDV) differed from those for MeV and MuV; but these differences are easily explainable by the fact that the NDV vaccine sequences analyzed were derived from two related but independent field isolates, B1 and LaSota [10, 29]. When analyses included both B1 and LaSota strains, they did not reveal evidence of a relaxation of selection (Table 1 and 3). This result can be explained by the fact that the original B1 and LaSota, lineages evolved independently in a natural environment since their most recent common ancestor, prior to the separate isolation and maintenance of each lineage in culture as a vaccine. Thus, sequence comparisons between B1 and LaSota should be expected to show the effects of purifying selection that occurred in the natural state, whereas sequence comparisons within each strain should reveal the relaxation of purifying selection after each lineage was cultured as a vaccine. Consistent with this interpretation, when NDV vaccine sequences were analyzed within strains, they showed a pattern of nucleotide diversity indicating a relaxation of purifying selection similar to that seen in vaccine strains of MeV and MuV (Table 1). Given the independent isolation of the NDV vaccine strains, the observed differences between NDV and the other two viruses strongly support the hypothesis that the relaxation of purifying selection occurs as a consequence of culture of a live vaccine and not for some other reason.

Vaccine strains of MeV, MuV, and NDV all showed an excess of polymorphic nonsynonymous sites in comparison to wild-type sequences, and the bulk of these nonsynonymous polymorphisms were unique to the vaccine strains. This pattern provides further evidence in support of the hypothesis that purifying selection is relaxed in the vaccine strains. Moreover,

because the polymorphisms involved introduced numerous amino acid substitutions unique to the vaccine strains, they suggest that the latter are capable of following evolutionary trajectories substantially different from those of the ancestral viruses.

In the case of MeV, a wild-type sequence close to the Edmonston strain ancestor (AF266288; ref. 38) enabled reconstruction of the small number of changes ancestral to the vaccine clade. A total of six changes were reconstructed, five nonsynonymous and one synonymous; all of these changes corresponded to differences between the Edmonston strain and its wild-type sister sequence that were previously described [38]. Of the five amino acid substitutions, four were fixed in all available sequences in the vaccine clade and in none of the wild-type sequences. These four changes involve the substitutions E→K at position 89 of the M protein; S→G at position 211 and N→Y at position 481 of the H protein; and D→A at position 1717 of the L protein. Because the latter four amino acid changes were fixed in the vaccine clade, they appear likely to contribute to the phenotypic properties selected for during the attenuation process. There were just two reconstructed changes in the non-coding region on the branch ancestral to the MeV vaccine clade, both 5' to the start site of the *NP* gene: A →T at site 23 and C→A at site 43. Since these changes also were fixed in all available members of the vaccine clade, they also are candidates for contributing to the attenuated phenotype.

However, the sequence changes apparently fixed during the attenuation process of MeV vaccine were far fewer in number than the polymorphic sites unique to the vaccine sequences. This implies that most of the evolutionary change observed in the vaccine sequences is due to mutations that occurred after the vaccine strain was established, including many mutations that would be deleterious in a free-living virus and are found in the vaccine strain only because of the relaxation of purifying selection. Thus, the relaxation of purifying selection is likely to be the primary factor giving rise to sequence diversity in the vaccine strain of MeV.

In MuV, nucleotide substitutions in non-coding regions were significantly elevated in comparison to those at synonymous sites in the comparison between vaccine sequences and the nearest wild-type sister sequence (Table 2). This suggests that certain changes in non-coding regions may have been selectively favored during the attenuation process. In contrast to MeV, the nearest available wild-type sister sequence to the MuV vaccine sequence clade was not very closely related; and all available sequences from the MuV Urabe vaccine lineage were very closely related to one another (Tables 1 and 2). Nonetheless, in the vaccine strains of MuV, there were several nonsynonymous polymorphisms unique to the vaccine strains, implying that the relaxation of purifying selection has played a role the diversification of MuV vaccine strain sequences.

In conclusion, the present analyses demonstrate that the extent of continuing molecular evolution in live attenuated vaccine strains of MeV, MuV, and NDV has been substantial. Obviously, empirical studies will be required to test for phenotypic effects of any individual mutation occurring in vaccine lineages. However, the present analyses demonstrate that, as a consequence of the relaxation of purifying selection, vaccine strains can accumulate numerous amino acid sequence differences from free-living strains, thereby potentially affecting their biological properties, including antigenicity and vaccine effectiveness. The accumulation of selectively neutral or nearly neutral mutations increases with the number of generations of replication that the virus undergoes [13]. Therefore, minimizing the number of generations between the original vaccine master seed and the strains used in vaccination would seem a practical step to minimize the accumulation of mutations in vaccine strains, along with monitoring of evolutionary change in live attenuated vaccines by periodic sequence analysis [39].

Acknowledgments

This research was supported by grant GM43940 from the National Institutes of Health

References

1. McCullers JA. Evolution, benefits, and shortcomings of vaccine management. *J Managed Care Pharm* 2007;13(7 Suppl B):S2–S6.
2. Melnick, JL. Conventional viral vaccines and their influence on the epidemiology of disease. In: Dimmock, NJ.; Griffiths, PD.; Madeley, CR., editors. *Control of Virus Diseases*. Cambridge: Cambridge University Press; 1999. p. 3-19.
3. Bonnet MC, Dutta A, Weinberger C, Plotkin SA. Mumps vaccine strains and aseptic meningitis. *Vaccine* 2006;24(4950):7037–7045. [PubMed: 16884835]
4. Bonnet MC, Dutta A. World wide experience with inactivated poliovirus vaccine. *Vaccine* 2008;26(39):4978–4983. [PubMed: 18680777]
5. Kamboj M, Sepkowitz KA. Risk of transmission associated with live attenuated vaccines given to healthy persons caring for or residing with an immunocompromised patient. *Infect Control and Hospital Immunol* 2007;28(6):702–707.
6. Cohen C, White JM, Savage EJ, Glynn JR, Choi Y, Andrews N, et al. Vaccine effectiveness estimates, 2004–2005 mumps outbreak, England. *Emerging Infect Dis* 2007;13(1):12–17. [PubMed: 17370510]
7. Marin M, Nguyen HQ, Langidrik JR, Edwards R, Briand K, Papiana MJ, et al. Measles transmission and vaccine effectiveness during a large outbreak on a densely populated island: implications for vaccination policy. *Clin Infect Dis* 2006;42(3):315–319. [PubMed: 16392073]
8. Marin M, Quinlisk P, Shimabukuru T, Sahney C, Brown C, LeBaron CW. Mumps vaccination coverage and vaccine effectiveness in a large outbreak among college students – Iowa, 2006. *Vaccine* 2008;26(2930):3601–3607. [PubMed: 18539365]
9. Velicko I, Müller LL, Pebody R, Gergonne B, Aidyralieva C, Kostiuhenko N, et al. Nationwide measles epidemic in Ukraine: the effect of low vaccine effectiveness. *Vaccine*. 2008 in press
10. Hitchner SB. Serendipity in science: discovery of the B-1 strain of Newcastle disease virus. *Avian Dis* 1975;19(2):215–223. [PubMed: 1098636]
11. Enders JF, Katz SL, Milovanovic ML, Holloway A. Studies of an attenuated measles virus vaccine. *N Engl J Med* 1960;263:153–159. [PubMed: 13820246]
12. Boroskin YS, Yamade A, Kaptsova TI, Skvortsova OI, Sinitsyna OA, Takeuchi K, et al. Genetic evidence for variant selection in the course of dilute passaging of mumps vaccine virus. *Res Virol* 1992;143(4):279–283. [PubMed: 1410827]
13. Kimura, M. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press; 1983.
14. Ohta T. Slightly deleterious mutant substitutions in evolution. *Nature* 1973;246(5428):96–98. [PubMed: 4585855]
15. Hughes, AL. *Adaptive Evolution of Genes and Genomes*. New York: Oxford University Press; 1999.
16. Hughes AL. Near neutrality: leading edge of the neutral theory of molecular evolution. *Ann NY Acad Sci* 2008;1133:162–179. [PubMed: 18559820]
17. Hughes AL, Packer B, Welsch R, Bergen AW, Chanock SJ, Yeager M. Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proc Natl Acad Sci USA* 2003;100(26):15754–15757. [PubMed: 14660790]
18. Hughes AL. Evidence for abundant slightly deleterious polymorphisms in bacterial populations. *Genetics* 2005;169(2):553–558.
19. Hughes AL, Hughes MK. More effective purifying selection on RNA viruses than in DNA viruses. *Gene* 2007;404(12):117–125. [PubMed: 17928171]
20. Hughes AL, Piontkivska H. Nucleotide sequence polymorphism in circoviruses. *Infect Genet Evol* 2008;8(2):130–138. [PubMed: 18093882]
21. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Diggins DG. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25(24):4876–4882. [PubMed: 9396791]

22. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4(4):406–425. [PubMed: 3447015]
23. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24(8):1596–1599. [PubMed: 17488738]
24. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985;39(4):83–791.
25. Han G-Z, He C-Q, Ding N-Z, Ma LY. Identification of a natural multi-recombinant of Newcastle disease virus. *Virology* 2008;371(1):54–60. [PubMed: 18028976]
26. Parks CL, Lerch RA, Walpita PA, Wang H-P, Sidhu MS, Udem SA. Analysis of the noncoding regions of measles virus strains in the Edmonston lineage. *J Virol* 2001;75(2):921–933. [PubMed: 11134305]
27. Amexis G, Fineschi N, Chumakov K. Correlation of genetic variability with safety of mumps vaccine Urabe AM9 strain. *Virology* 2001;287(1):234–241. [PubMed: 11504558]
28. Hanson, RP. Heterogeneity within strains of Newcastle disease virus: key to survival. In: Alexander, DJ., editor. *Newcastle Disease*. Boston: Kluwer Academic Publishers; 1988. p. 112-120.
29. Goldhaft TM. Historical note on the origin of the LaSota strain of Newcastle disease virus. *Avian Dis* 1980;24(2):297–301. [PubMed: 7002145]
30. Miller PJ, King DJ, Afonso CL, Suarez DL. Antigenic differences among Newcastle disease virus strains of different genotypes used in vaccine formulation affect viral shedding after a virulent challenge. *Vaccine* 2007;25(41):7238–7246. [PubMed: 17719150]
31. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986;3(5):418–426. [PubMed: 3444411]
32. Li W-H. Unbiased estimates of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* 1993;36(1):96–99. [PubMed: 8433381]
33. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* 2000;17(1):32–43. [PubMed: 10666704]
34. Hughes AL, French JO. Homologous recombination and the pattern of nucleotide substitution in *Ehrlichia ruminantium*. *Gene* 2007;387(12):31–37. [PubMed: 17005333]
35. Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*. New York: Oxford University Press; 2000.
36. Swofford, DL. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sunderland, Mass.: Sinauer; 2000.
37. Nei, M. *Molecular Evolutionary Genetics*. New York: Columbia University Press; 1987.
38. Baričević M, Forčević D, Gulija TK, Jug R, Mažuran R. Determination of the coding and non-coding nucleotide sequences of genuine Edmonston-Zagreb master seed and current working seed lot. *Vaccine* 2005;23(8):1072–1078. [PubMed: 15620481]
39. Amexis G, Rubin S, Chizhikov V, Pelloquin F, Carbone K, Chumakov K. Sequence diversity of Jeryl Lynn strain of mumps virus: quantitative mutant analysis for vaccine quality control. *Virology* 2002;300:171–179. [PubMed: 12350348]

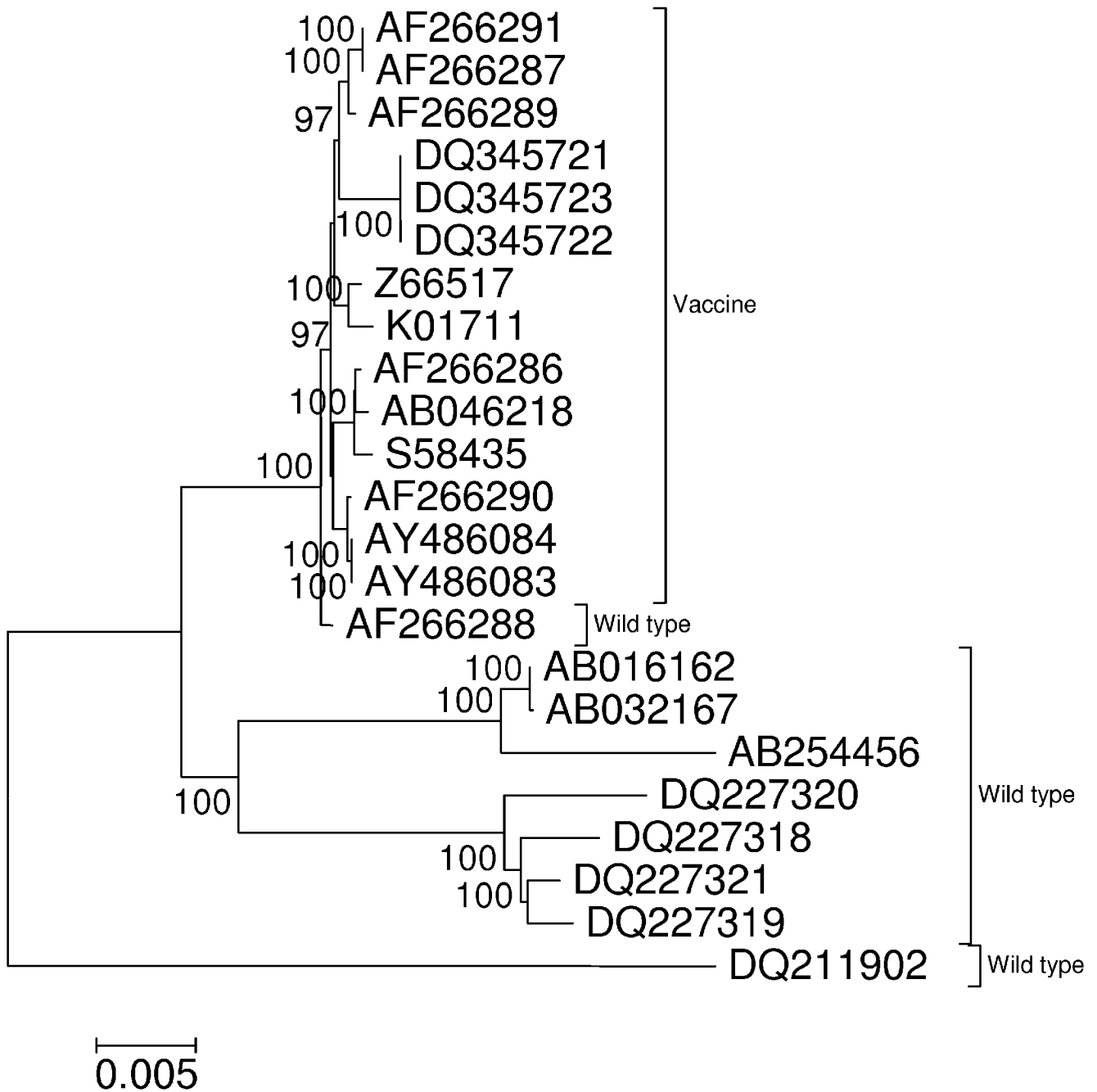


Figure 1. NJ tree based on MCL distance at 15887 aligned nucleotide sites in selected wild-type and vaccine measles virus (MeV) complete genomes. Numbers on the branches represent the percentage of 1000 bootstrap samples supporting the branch; only values $\geq 95\%$ are shown.

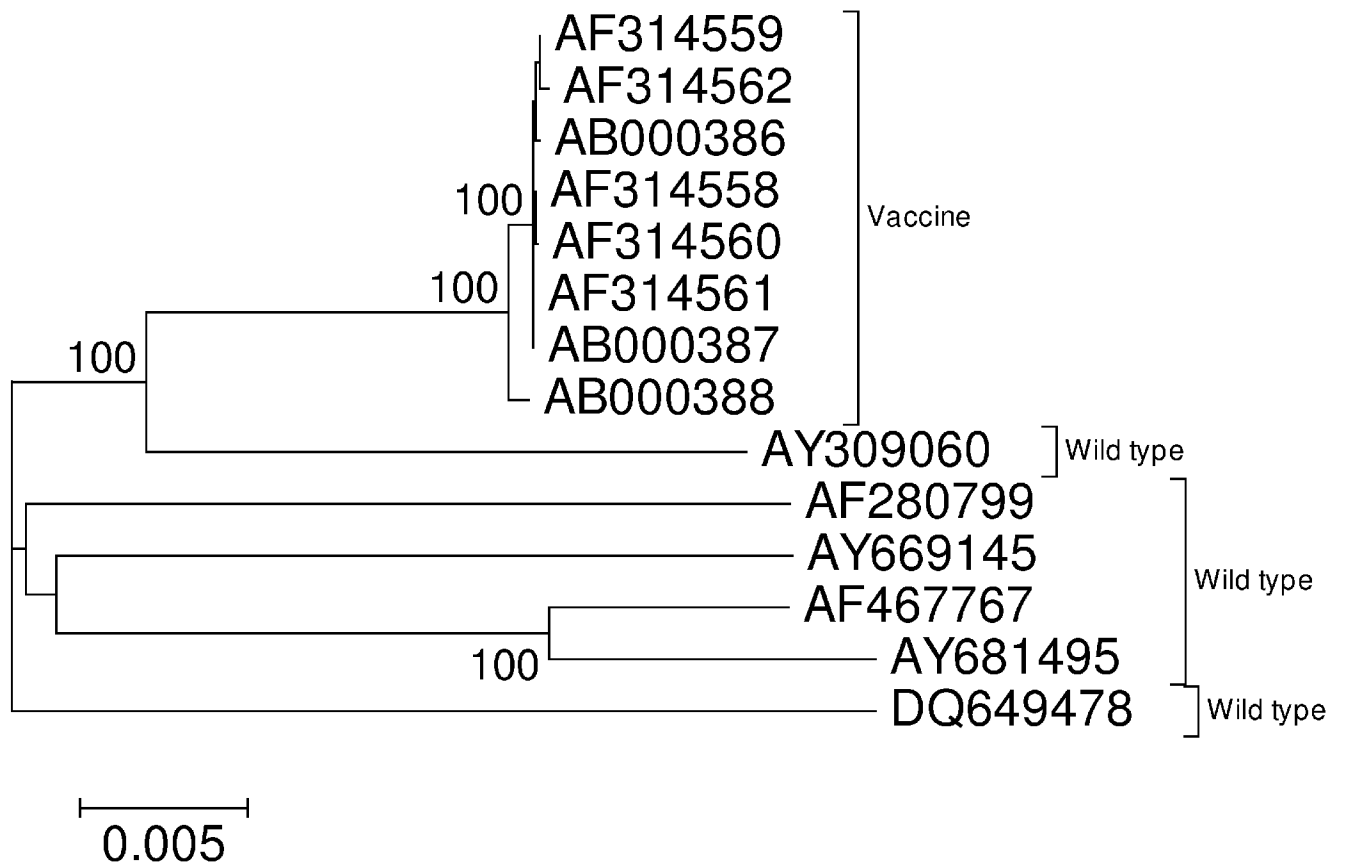


Figure 2. NJ tree based on MCL distance at 15376 aligned nucleotide sites in selected wild-type and vaccine mumps virus (MuV) complete genomes. Numbers on the branches represent the percentage of 1000 bootstrap samples supporting the branch; only values $\geq 95\%$ are shown.

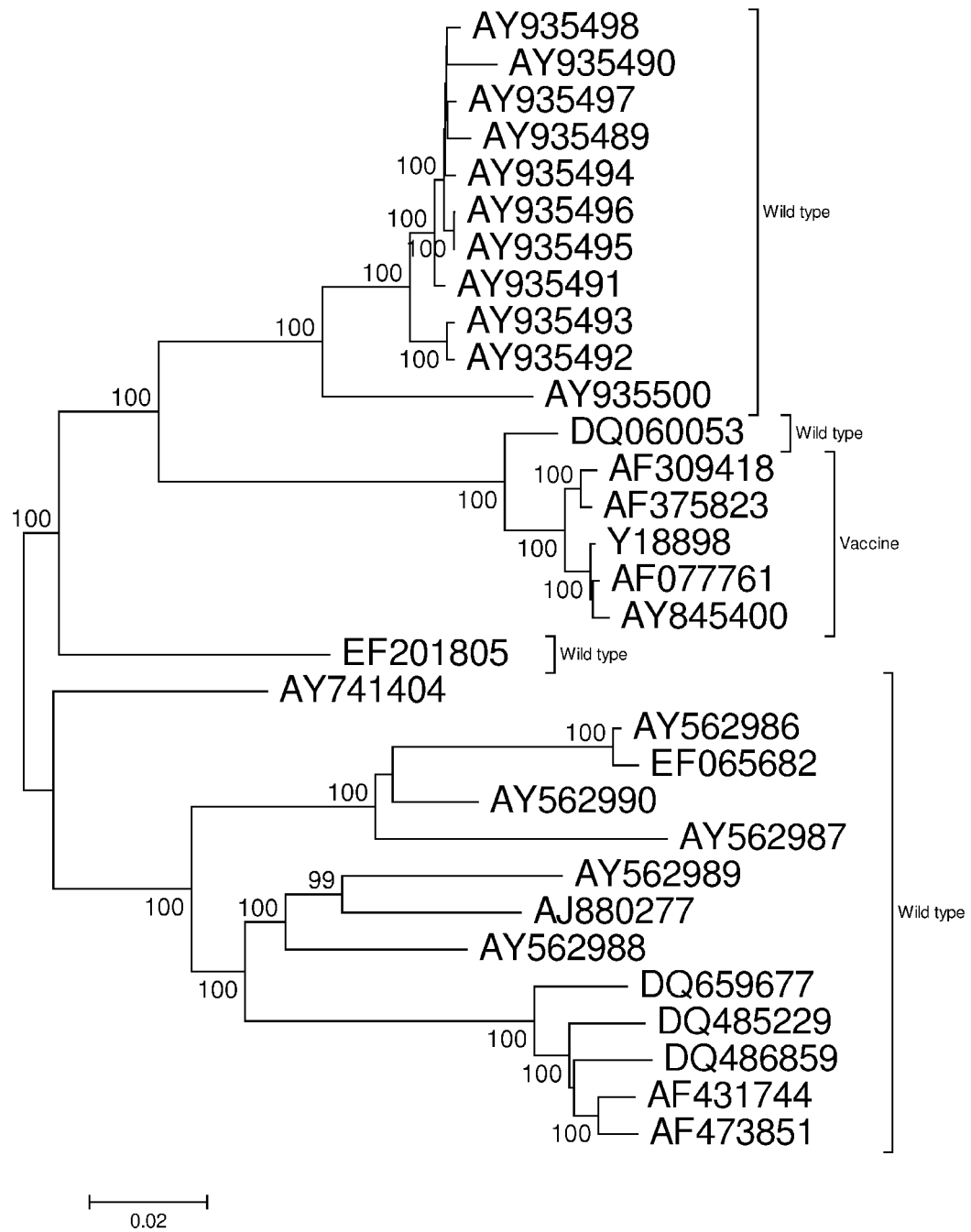


Figure 3. NJ tree based on MCL distance at 15182 aligned nucleotide sites in selected wild-type and vaccine Newcastle disease virus (NDV) complete genomes. Numbers on the branches represent the percentage of 1000 bootstrap samples supporting the branch; only values $\geq 95\%$ are shown.

Table 1

Synonymous (π_S) and nonsynonymous (π_N) nucleotide diversity in coding regions and nucleotide diversity (π) in non-coding regions in vaccine and wild-type (wt) genomic sequences of paramyxoviruses.

Virus	Genome set (no.)	Coding		Non-coding	
		$\pi_S \pm \text{S.E.}$	$\pi_N \pm \text{S.E.}$	$\pi_N : \pi_S$	$\pi \pm \text{S.E.}$
MeV	wt (9)	0.0984 \pm 0.0038	0.0089 \pm 0.0006 ^{***}	0.09	0.0708 \pm 0.0039 ^{***}
	vaccine (14)	0.0042 \pm 0.0005	0.0027 \pm 0.0003	0.64	0.0050 \pm 0.0010
MuV	wt (6)	0.1616 \pm 0.0048	0.0078 \pm 0.0006 ^{***}	0.05	0.0914 \pm 0.0057 ^{***}
	vaccine (8)	0.0007 \pm 0.0003	0.0004 \pm 0.0001	0.57	0.0018 \pm 0.0006
NDV	wt (27)	0.4533 \pm 0.0103	0.0308 \pm 0.0010 ^{***}	0.07	0.0914 \pm 0.0057 ^{***}
	vaccine (5)	0.0200 \pm 0.0019	0.0033 \pm 0.0004 ^{***}	0.17	0.0153 \pm 0.0023
	B1 (2) Lasota (3)	0.0070 \pm 0.0015	0.0029 \pm 0.0005	0.41	0.0098 \pm 0.0026
		0.0043 \pm 0.0010	0.0026 \pm 0.0004	0.61	0.0065 \pm 0.0015

Tests of the hypothesis that π_N or π equals the corresponding value of π_S :

P < 0.001 (Z-test).

Estimated number of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions per site in coding regions and nucleotide substitutions per site (d) in non-coding regions on the branch leading to the ancestor of paramyxovirus vaccine clades; and mean d_S , d_N , and d in comparisons between vaccine strains and the nearest wild-type sister sequence.

Table 2

Virus	Genome set (no.)	Coding		Non-coding	
		$d_S \pm$ S.E.	$d_N \pm$ S.E.	$d_N : d_S$	$d \pm$ S.E.
MeV	branch to ancestor	0.0003 \pm 0.0003	0.0005 \pm 0.0002	1.67	0.0012 \pm 0.0008
	vs. sister group	0.0036 \pm 0.0007	0.0024 \pm 0.0003	0.67	0.0054 \pm 0.0014
MuV	branch to ancestor	0.0114 \pm 0.0019	0.0103 \pm 0.0011	0.90	0.0189 \pm 0.0041
	vs. sister group	0.0241 \pm 0.0029	0.0260 \pm 0.0016	1.08	0.0523 \pm 0.0069 ^{***}
NDV	branch to ancestor	0.0363 \pm 0.0034	0.0034 \pm 0.0006 ^{***}	0.09	0.0196 \pm 0.0037 ^{***}
	vs. sister group	0.0735 \pm 0.0043	0.0072 \pm 0.0007 ^{***}	0.10	0.0430 \pm 0.0047 ^{***}

Tests of the hypothesis that d_N or d equals the corresponding value of d_S .

*** P < 0.001 (Z-test).

Table 3

Means and standard deviations (S.D.) of gene diversity at non-coding, synonymous, and nonsynonymous sites in wild-type and vaccine genomic sequences of paramyxoviruses.

Virus	Wild-type														
	Non-coding					Synonymous					Nonsynonymous				
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.			
MeV	357	0.294 ^d	0.130 ^e	899	0.295 ^b	0.128 ^e	310	0.264			0.122				
MuV	239	0.338 ^b	0.094 ^e	1209	0.326 ^d	0.086 ^c	211	0.309			0.072				
NDV	965	0.306 ^b	0.175 ^e	3298	0.282 ^b	0.172 ^d	1274	0.209			0.165				
	Vaccine														
Virus	Non-coding					Synonymous					Nonsynonymous				
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.			
MeV	27	0.294	0.116	53	0.248	0.094	97	0.279			0.116				
MuV	9	0.219	0.000	9	0.236	0.052	15	0.256			0.080				
NDV	44	0.396 ^d	0.088 ^e	120	0.431 ^b	0.077 ^e	79	0.347			0.063				

Randomization tests of the hypothesis that the mean of gene diversity at non-coding or synonymous polymorphic sites equals that at nonsynonymous sites:

^a $P < 0.01$;

^b $P < 0.001$.

Randomization tests of the hypothesis that the variance of gene diversity at non-coding or synonymous polymorphic sites equals that at nonsynonymous sites:

^c $P < 0.05$;

^d $P < 0.01$;

^e $P < 0.001$.