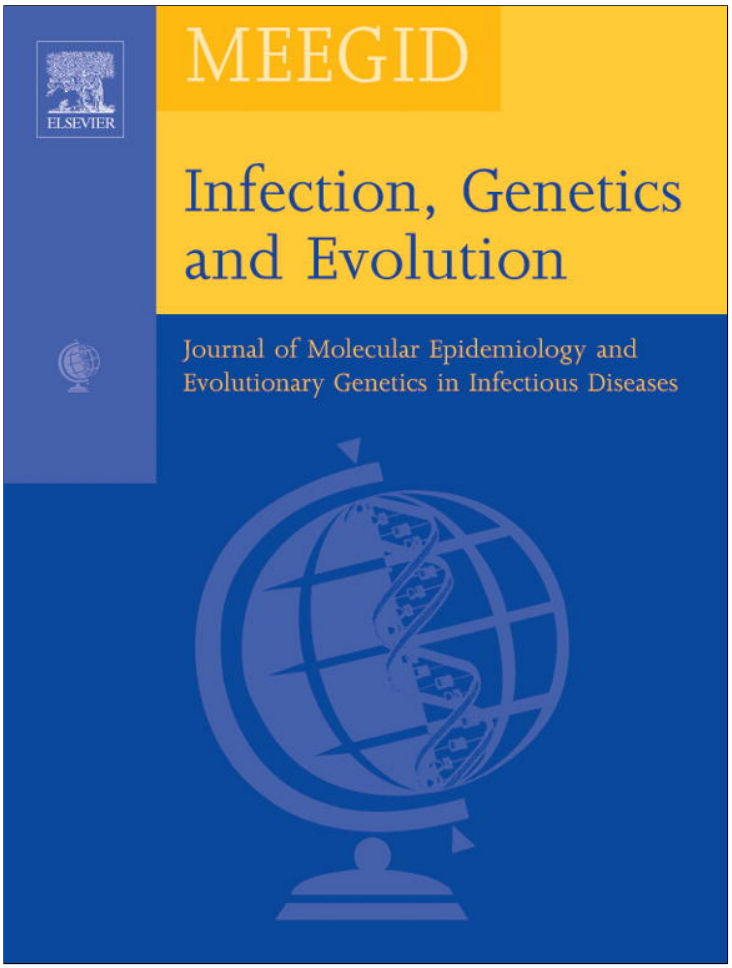

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Infection, Genetics and Evolution 16 (2013) 93–98

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Detection of positive selection eliminating effects of structural constraints in hemagglutinin of H3N2 human influenza A virus

Yoshiyuki Suzuki *

Graduate School of Natural Sciences, Nagoya City University, Japan

ARTICLE INFO

Article history:
Received 2 December 2012
Received in revised form 28 January 2013
Accepted 31 January 2013
Available online 9 February 2013

Keywords:
Positive selection
Structural constraint
Thermodynamic stability
Hemagglutinin
Influenza A virus

ABSTRACT

In the evolutionary studies of proteins, the average effect of natural selection operating on amino acid mutations may be examined by comparing the numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions that have accumulated during the same time period. In this method, destabilizing mutations occurring across protein molecules may interfere with detection of natural selection, particularly positive selection, operating on other mutations. Here an attempt to detect positive selection eliminating effects of structural constraints is demonstrated using hemagglutinin (HA) of H3N2 human influenza A virus as an example. Compatible and incompatible amino acids were inferred at each site from the computational analysis of three-dimensional structure using the thermodynamic stability as an indicator, and natural selection was examined by comparing d_S and d_N among compatible amino acids. In the analysis of 2701 nucleotide sequences for the entire coding region of HA, the new method identified twice as many positively selected amino acid sites as the ordinary method (16 and 4 sites in the former method without and with correction for multiple testing, respectively, and 8 and 2 sites in the latter method). Positively selected sites were involved in epitopes, receptor-binding pocket, epistasis, and stabilization, which appeared to be biologically reasonable. Nevertheless, there still appeared to be several problems, which may largely render this method conservative. It may be effective to analyze many densely sampled sequences in this method.

1. Introduction

In the evolutionary studies of proteins, the average effect of natural selection operating on amino acid mutations may be examined by comparing the numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions that have accumulated during the same time period under the assumption that synonymous mutations are selectively neutral or nearly neutral; positive, negative, and no selection are inferred when $d_S < d_N$, $d_S > d_N$, and $d_S = d_N$, respectively (Kimura, 1977; Hughes and Nei, 1988; Suzuki and Gojobori, 1999). However, it has been reported in cellular organisms that on average ~36% of amino acid mutations are deleterious and >80% of deleterious mutations affect the thermodynamic stability of proteins (Tokuriki and Tawfik, 2009), which may lead to unfolding or misfolding and aggregation or degradation of proteins (Bloom et al., 2005; DePristo et al., 2005). The stability effect of mutations may be measured with the difference ($\Delta\Delta G$) in the free energy (ΔG) between native and mutant proteins. Since the rate of destabilizing mutations is >20 times greater than that of stabilizing mutations, proteins are only marginally stable; ΔG ranging from −3 to −15 kcal/mol and stability margin from 1 to 3 kcal/mol (DePristo et al., 2005; Tokuriki et al., 2007). It should be noted that a single amino acid mutation may alter ΔG by 0.5–5 kcal/mol (DePristo et al., 2005). Mutations affecting the stability occur across a protein molecule and stability effects of multiple mutations are largely additive (Wells, 1990; DePristo et al., 2005). It has been reported that viral proteins tend to be loosely packed and partially disordered, where the stability as well as the destabilizing effect of mutations may be relatively small compared with cellular organisms (Tokuriki et al., 2009). However, ~50% of mutations were shown to be deleterious in the experimental studies of hemagglutinin (HA) in influenza A virus (Nakajima et al., 2003, 2005).

Amino acid mutations destabilizing proteins are likely to be deleterious, such that they are immediately eliminated from the population and do not contribute to the evolution of proteins (Woo et al., 2010; Williams et al., 2011). In the study of natural selection by comparing d_S and d_N, these mutations may reduce d_N. Since destabilizing mutations are ubiquitous, it is possible that they may interfere with detection of natural selection, particularly positive selection, operating on other mutations. The purpose of the present study was to demonstrate an attempt to detect positive selection eliminating effects of structural constraints using HA of H3N2 human influenza A virus as an example.

* Address: Graduate School of Natural Sciences, Nagoya City University, 1 Yamanohata, Mizuho-cho, Mizuho-ku, Nagoya-shi, Aichi-ken 467-8501, Japan. Tel.: +81 52 872 5821; fax: +81 52 872 5821.
E-mail address: yossuzuk@nsc.nagoya-cu.ac.jp

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.01.017

2. Materials and methods

2.1. Methodology

Natural selection operating at single amino acid sites of proteins may be detected by comparing d_S and d_N at single codon sites of corresponding protein-coding nucleotide sequences. Parsimony (Fitch et al., 1997; Suzuki and Gojobori, 1999), likelihood (Suzuki, 2004a; Kosakovsky Pond and Frost, 2005; Massingham and Goldman, 2005), and Bayesian (Nielsen and Yang, 1998; Yang et al., 2000) methods have been developed for this purpose. In the present study, the parsimony method was modified to detect positive selection eliminating effects of structural constraints. However, the same approach can be incorporated into the likelihood and Bayesian methods as well.

In the ordinary methods for detecting natural selection (Hughes and Nei, 1988; Fitch et al., 1997; Suzuki and Gojobori, 1999), d_S and d_N were computed through dividing the numbers of synonymous (c_S) and nonsynonymous (c_N) differences by those of synonymous (s_S) and nonsynonymous (s_N) sites, respectively. Although c_S and c_N were obtained by simply counting synonymous and nonsynonymous differences, s_S and s_N were computed from the relative rates of synonymous and nonsynonymous mutations, which were defined as the probabilities of occurrence of synonymous and nonsynonymous mutations when 1 nucleotide site in the relevant codon sites was assumed to mutate, regardless of the type of amino acid mutation it may cause. However, if the amino acid mutation destabilizes proteins, it may not contribute to increasing c_N but only to increasing s_N, which may lead to a reduction in d_N and an interference with detection of natural selection, particularly positive selection, operating on other mutations. Therefore, the fraction of s_N corresponding to destabilizing mutations may be discarded from the comparison of d_S and d_N for detecting natural selection other than negative selection operating on structural constraints.

Destabilizing mutations may be inferred from the computational analysis of three-dimensional structure for proteins using the thermodynamic stability as an indicator (Tokuriki and Tawfik, 2009; Williams et al., 2011). In the structure of proteins, each site may be mutated to 20 amino acids and the thermodynamic stability of mutant proteins may be evaluated *in silico*. At each site, the amino acid that provides the least stability required for proper functions of proteins may be determined as the marginal amino acid. Amino acids providing the stability equal to or greater than the marginal amino acid may be considered as compatible to the site, whereas those providing smaller stability incompatible. Mutations from compatible to incompatible amino acids may be considered as destabilizing. It may not be easy to determine the marginal amino acid theoretically (Williams et al., 2011). However, since the rate of destabilizing mutations is much higher than that of stabilizing mutations and proteins are only marginally stable (DePristo et al., 2005; Tokuriki et al., 2007), the marginal amino acid may be inferred empirically as the amino acid that provides the least stability among those observed in samples of sequences and predicted to have existed during evolution.

This strategy is similar to that adopted in examination of natural selection operating on the amino acid mutations that maintain physicochemical properties of amino acids such as charge, polarity, and volume (conservative mutations). In this method, 20 amino acids were classified into several categories according to the physicochemical properties and s_N corresponding to mutations changing the physicochemical properties was discarded from the comparison of d_S and d_N (Hughes et al., 1990; Suzuki, 2007). The new method is equivalent to this method by regarding compatible amino acids as belonging to a single category and each of incompatible amino acids to a separate category. The performance of the latter method has been examined extensively with computer simulation and real data analysis, and it has been concluded that the method was generally conservative and reliable (Suzuki, 2007), which should also apply to the former method.

2.2. Sequence data

Influenza A virus is an etiological agent of influenza (Shope, 1931). HA, together with neuraminidase (NA), constitute envelope glycoproteins of virions, where HA exists ~10 times more abundantly than NA. According to the antigenicity of HA and NA, influenza A virus is classified into subtypes H1-H17 and N1–N10, respectively (World Health Organization, 1980). In the present study, HA of human H3N2 virus was analyzed because of the clinical importance of this virus causing annual epidemics and the availability of sequence data containing a large amount of genetic variation as well as structural data. HA is a homotrimeric type I transmembrane glycoprotein consisting of 566 amino acid sites, and is the sialic acid receptor-binding protein and the major target of humoral immunity (Skehel and Wiley, 2000). The antigenic sites constitute 5 epitopes (A–E) in H3N2 virus (Wilson et al., 1981; Suzuki, 2004b). HA is a good example for evaluating the performance of the method for detecting positive selection empirically, because the functions of amino acid sites have been well characterized. In fact, detection of positively selected amino acid sites has been conducted for this protein in previous studies (Bush et al., 1999; Yang, 2000; Suzuki, 2006; Chen and Sun, 2011; Murrell et al., 2012).

A total of 4167 nucleotide sequences for the entire coding region of HA for human H3N2 virus, excluding laboratory and vaccine strains, were retrieved from the Influenza Virus Resource at the National Center for Biotechnology Information as of May 16, 2012 (Bao et al., 2008). After eliminating sequences with the same strain names as others, sequences identical to others, sequences derived from incidental human infections of swine strains, and sequences with minor gaps, ambiguous nucleotides, and premature termination codons, 2701 sequences were used in the following analysis (Supplementary Table S1). A duck H3N8 virus [A/duck/Hokkaido/10/1985(H3N8); International Nucleotide Sequence Database (INSD) accession number: AB276113] was added as the outgroup to identify the position of the root for the phylogenetic tree of human H3N2 virus. Each sequence consisted of 1698 nucleotide sites.

2.3. Data analysis

Multiple alignment of nucleotide sequences for HA of 2702 human and duck viruses was made by using the computer program MAFFT (version 6.901b) (Katoh et al., 2002), which did not contain any gaps. Phylogenetic trees were constructed by the neighbor-joining method (Saitou and Nei, 1987) with the p distance (Nei and Kumar, 2000) and the maximum composite likelihood (MCL) distance (Tamura et al., 2004), which were known to produce reliable phylogenetic trees when a large number of closely related sequences was analyzed, using MEGA (version 5.05) (Tamura et al., 2011). Nucleotide sequences at interior nodes of the phylogenetic tree were inferred by the maximum parsimony method (Fitch, 1971; Hartigan, 1973) using PAML (version 4.4b) (Yang, 2007).

The marginal as well as compatible and incompatible amino acids were inferred at each site of HA using the three-dimensional structures for A/Hong Kong/19/1968(H3N2) (corresponding to amino acid positions 17–344 and 346–520; Protein Data Bank [PDB] ID: 2HMG) and A/Victoria/3/1975(H3N2) (positions 26–341 and 346–516; 4GMS), which were different at 26 (5.339%) amino acid sites. In either structure, each site was mutated to 20 amino acids and ΔG of mutant proteins was predicted using FOLDX (version 3.0 beta 5.1) (Guerois et al., 2002; Schymkowitz et al., 2005).

Table 1
Positively selected amino acid sites detected in HA of H3N2 human influenza A virus.

Position	Function				No marginal amino acid	A/Hong Kong/19/1968(H3N2)			A/Victoria/3/1975(H3N2)			Previous study
	Epitope[a]	RBP[b]	Epistasis[c]	Other		Exterior and interior nodes	Interior nodes	Structural property	Exterior and interior nodes	Interior nodes	Structural property	
3			160		+[e]	+	+	Surface	N.A.[h]	N.A.	N.A.	Yes[ijk,l]
47	C					+		Surface			Surface	
53	C				+	+	+	Surface	+	+	Surface	Yes[k]
94	E				+	+	+	Surface	+	+	Surface	Yes[kl]
112			391		+	+	+	Core	+	+	Core	Yes[j]
135	A	Yes				(+)[g]	(+)	Surface			Core	Yes[i,j,k,l,m,n]
137	A	Yes	73,108,236			+	+	Surface		+	Surface	Yes[i,k,l,m,n]
138	A	Yes			+[f]	+	+	Core	+	+	Core	Yes[i,l,m,n,o]
142	A					+	+	Surface	+	+	Surface	Yes[j,k,m]
145	A				+	+	+	Surface	+	+	Surface	Yes[i,j,k,l,m,n]
157	B						+	Surface		+	Core	Yes[j,k,l,n]
164	B							Core	+	+	Core	
198	B	Yes				(+)	(+)	Surface			Surface	
203	D						+	Core		+	Core	
220				Stabilization[d]			+	Core			Core	Yes[j,l,m,p]
223		Yes	291		+	+	+	Surface	+	+	Core	
225		Yes	41,91,147,171,172 ,205,218,238,243			+	+	Surface		+	Surface	Yes[k]
262	E				+	+	+	Surface	+	+	Surface	Yes[j,k,l,m]

[a]Wiley et al. (1981).
[b]Skehel and Wiley (2000).
[c]Kryazhimskiy et al. (2011).
[d]Vanderlinden et al. (2010).
[e]The + symbol indicates that positive selection was detected using the phylogenetic trees constructed with the p and MCL distances.
[f]The box is colored grey when positive selection was detected with correction for multiple testing.
[g]The + symbol is indicated in parentheses when positive selection was detected using the phylogenetic tree constructed with the p distance but not with the MCL distance.
[h]N.A.: not applicable because the position was not included in the three-dimensional structure for A/Victoria/3/1975(H3N2).
[i]Fitch et al. (1997).
[j]Plotkin and Dushoff (2003).
[k]Shih et al. (2007).
[l]Chen and Sun (2011).
[m]Bush et al. (1999).
[n]Yang (2000).
[o]Suzuki and Gojobori (1999).
[p]Suzuki (2006).

The marginal amino acid was determined as the amino acid that provided the least stability among those observed at exterior nodes and inferred at interior nodes of the phylogenetic tree. It should be noted that, in the phylogenetic tree of HA, exterior nodes may contain slightly deleterious mutations (Pybus et al., 2007), which may be destabilizing. Therefore, only the amino acids inferred at interior nodes were also used in determination of marginal amino acid.

Natural selection was examined at each amino acid site of HA by discarding s_N corresponding to destabilizing mutations from the comparison of d_S and d_N (Suzuki and Gojobori, 1999). Here the transition/transversion rate ratio for nucleotide mutation (κ) was required in the computation of s_S and s_N. Using the ratio of the average numbers of transitional/transversional nucleotide substitutions estimated with the 2-parameter model (Kimura, 1980) at 89 fourfold degenerate sites in 2701 sequences for human H3N2 virus, κ was estimated to be 3.994. Therefore, κ was assumed to be 4 in the computation (Suzuki, 2011). Positive selection was detected when the null hypothesis of $d_S = d_N$ was rejected at the significance level of 0.05 (two-tailed test) with $d_S < d_N$. The correction for multiple testing was also conducted with the family-wise significance level of 0.05 (Suzuki, 2011). The relative solvent accessibility of each amino acid site in the three-dimensional structures of HA was obtained by using ASAVIEW (Ahmad et al., 2004). Amino acid sites were judged as surface exposed or core buried when the accessibility was >0.16 or ⩽0.16, respectively (Rost and Sander, 1994).

3. Results and discussion

3.1. Detection of positive selection eliminating effects of structural constraints

When natural selection was examined at each amino acid site of HA using the phylogenetic tree constructed with the p distance under the assumption that 20 amino acids were compatible, positive selection was detected at 8 and 2 sites without and with correction for multiple testing, respectively (Table 1). To eliminate effects of structural constraints, marginal as well as compatible and incompatible amino acids were inferred at each site using the structure for A/Hong Kong/19/1968(H3N2). When the amino acids observed at exterior nodes and inferred at interior nodes of the phylogenetic tree were used in determination of marginal amino acid, the number of compatible amino acids decreased to 9.591 ± 6.577 (Table 2). Consequently, the number of positively selected sites increased to 14 and 3 without and with correction, respectively (Table 1). In addition, by using only the amino acids inferred at interior nodes in determination of marginal amino acid, the number of compatible amino acids further decreased to 6.990 ± 6.261 (table 2) and that of positively selected sites increased to 16 and 4 without and with correction, respectively (Table 1).

Although the exact sets of compatible and incompatible amino acids changed at more than ~70% of sites in HA by using the structure for A/Victoria/3/1975(H3N2), more than ~80% of compatible and incompatible amino acids were shared at each site using different structures (Table 2). Positively selected amino acid sites varied slightly using different structures without correction for multiple testing. However, the same sites were identified as positively selected with correction (Table 1). Similar results were obtained by using the phylogenetic tree constructed with the MCL distance (Tables 1 and 2).

3.2. Functions of positively selected amino acid sites

Most of the amino acid sites identified as positively selected (13 of 18) were located in epitopes (Table 1), which have been charac-

Table 2
Statistics for compatible and incompatible amino acids in HA of H3N2 human influenza A virus.

		Exterior and interior nodes				Interior nodes			
		A/Hong Kong/19/1968(H3N2)		A/Victoria/3/1975(H3N2)		A/Hong Kong/19/1968(H3N2)		A/Victoria/3/1975(H3N2)	
		p distance	MCL distance	*p* distance	MCL distance	*p* distance	MCL distance	*p* distance	MCL distance
A/Hong Kong/19/1968(H3N2)	*p* distance	9.726 ± 6.539[a]	1.000 (1.000)[b]	0.847 (0.238)	0.847 (0.238)	7.105 ± 6.267	0.998 (0.996)	0.790 (0.300)	0.789 (0.300)
	MCL distance	1.000 (1.000)[c]	9.726 ± 6.539	0.847 (0.238)	0.847 (0.238)	0.999 (0.996)	7.076 ± 6.259	0.790 (0.300)	0.791 (0.302)
A/Victoria/3/1975(H3N2)	*p* distance	0.859 (0.238)	0.859 (0.238)	9.595 ± 6.531	1.000 (1.000)	0.889 (0.300)	0.889 (0.300)	6.844 ± 6.194	0.999 (0.996)
	MCL distance	0.859 (0.238)	0.859 (0.238)	1.000 (1.000)	9.595 ± 6.531	0.889 (0.300)	0.890 (0.302)	0.999 (0.996)	6.830 ± 6.183

[a] Average numbers of compatible amino acids at each site ± standard deviations are indicated along the diagonal as colored grey.
[b] Average proportions of compatible amino acids at each site shared between the cases of using different phylogenetic trees and structures are indicated above the diagonal. The values in parentheses indicate proportions of amino acid sites where the exact sets of compatible amino acids were shared between the cases.
[c] Average proportions of incompatible amino acids at each site shared between the cases of using different phylogenetic trees and structures are indicated below the diagonal. The values in parentheses indicate proportions of amino acid sites where the exact sets of incompatible amino acids were shared between the cases.

terized as targets of positive selection through promoting mutants to escape from humoral immunity (Fitch et al., 1997; Suzuki and Gojobori, 1999). In addition, 6 sites were associated with the receptor-binding pocket (PBP), which may also be positively selected by facilitating immune evasion of mutants through changing receptor-binding avidity (Laeeq et al., 1997; Hensley et al., 2009). Furthermore, 5 sites have been reported to be involved in epistasis with other sites, which may be important for adaptive evolution of HA (Kryazhimskiy et al., 2011). At least one of these functions could be assigned to all of the positively selected amino acid sites identified in the present study, except for position 220, which appeared to be buried inside the HA molecule (Table 1). It has been reported that position 220 is located at the inter-subunit interface of HA trimer and involved in its stabilization (Vanderlinden et al., 2010). Since the flanking positions of position 220 are involved in epitope D (position 219) and RBP (position 221), it is possible that position 220 plays a role of compensating for destabilizing effects of amino acid mutations at flanking positions, which may be positively selected through inducing immune escape. Some of the positively selected amino acid sites detected in the present study have not been identified in previous studies (Table 1) (Fitch et al., 1997; Bush et al., 1999; Suzuki and Gojobori, 1999; Yang, 2000; Plotkin and Dushoff, 2003; Suzuki, 2006; Shih et al., 2007; Chen and Sun, 2011), which may be due to effects of structural constraints.

3.3. Limitations in the new method

To eliminate effects of structural constraints from the analysis of natural selection operating at the amino acid sequence level, s_N corresponding to destabilizing mutations was discarded from the comparison of d_S and d_N in the present study. Since s_N was decreased and thus d_N was increased, positive selection was identified at a greater number of amino acid sites compared with the ordinary method, where 20 amino acids were assumed to be compatible to each site (Table 1). The positively selected sites identified by the new method in HA of human H3N2 virus appeared to be biologically reasonable, suggesting that this method was useful for detecting positive selection. Nevertheless, there still appeared to be several problems in the new method. First, in this method, the amino acids providing the stability equal to or greater than the marginal amino acid were considered as compatible to each site of proteins. However, mutations increasing the stability of proteins can also be deleterious by affecting dynamics and regulation of proteins (DePristo et al., 2005; Tokuriki et al., 2007). If this was the case, effects of structural constraints may still remain in the new method. Second, destabilizing mutations may spread in the population if they provide advantageous effects on some aspects other than stability that were greater than deleterious effects on stability. In this case, incompatible amino acids may be erroneously considered as compatible if they appear at exterior or interior nodes of the phylogenetic tree. Third, marginal as well as compatible and incompatible amino acids may fluctuate during evolution at each site of proteins according to the occurrence of substitutions at other sites (epistasis) (Nakajima et al., 2003, 2005; Gaucher et al., 2008; Kryazhimskiy et al., 2011; Williams et al., 2011; Breen et al., 2012). In particular, it has been reported that many of the amino acids that were accepted at each site in the long-term evolution of proteins were not acceptable in the short-term evolution (Breen et al., 2012). In the present study, it was observed that the amino acid sites identified as positively selected were largely identical using the structures of HA that were different at ~5% of amino acid sites (Table 1), suggesting that the effect of epistasis on detection of positive selection may be small as long as the divergence of amino acid sequences analyzed was relatively low (at least ~5%). It should be noted that the problems listed above may cause overestimation of s_N and underestimation of d_N, rendering detection of positive selection conservative but reliable (Suzuki, 2010).

Finally, in the new method, the marginal amino acid was determined among those observed at exterior nodes and inferred at interior nodes of the phylogenetic tree. However, the marginal amino acid may not necessarily have existed in nature. If this was the case, s_N may be underestimated and d_N may be overestimated, which may inflate the false-positive rate of detecting positive selection. To avoid this problem, it may be important to observe multiple amino acids at each site of proteins. Taken together with the results obtained above, these observations suggest that it may be effective to analyze many densely sampled sequences for detecting positive selection in the new method.

Acknowledgements

The author thanks Yuki Kobayashi, Manabu Igarashi, and two anonymous reviewers for valuable suggestions and comments. This work was supported by the Grant-in-Aid for Research in Nagoya City University to Y.S.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2013.01.017.

References

Ahmad, S., Gromiha, M.M., Fawareh, H., Sarai, A., 2004. ASAView: database and tool for solvent accessibility representation in proteins. BMC Bioinformatics 5, 51.
Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., Ostell, J., Lipman, D., 2008. The influenza virus resource at the national center for biotechnology information. J. Virol. 82, 596–601.
Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C., Arnold, F.H., 2005. Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. USA 102, 606–611.
Breen, M.S., Kemena, C., Vlasov, P.K., Notredame, C., Kondrashov, F.A., 2012. Epistasis as the primary factor in molecular evolution. Nature 490, 535–538.
Bush, R.M., Fitch, W.M., Bender, C.A., Cox, N.J., 1999. Positive selection on the H3 hemagglutinin gene of human influenza virus A. Mol. Biol. Evol. 16, 1457–1465.
Chen, J., Sun, Y., 2011. Variation in the analysis of positively selected sites using nonsynonymous/synonymous rate ratios: an example using influenza virus. PLoS One 6, e19996.
DePristo, M.A., Weinreich, D.M., Hartl, D.L., 2005. Missense meanderings in sequence space. a biophysical view of protein evolution. Nat. Rev. Genet. 6, 678–687.
Fitch, W.M., 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. 20, 406–416.
Fitch, W.M., Bush, R.M., Bender, C.A., Cox, N.J., 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. Proc. Natl. Acad. Sci. USA 94, 7712–7718.
Gaucher, E.A., Govindarajan, S., Ganesh, O.K., 2008. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature 451, 704–707.
Guerois, R., Nielsen, J.E., Serrano, L., 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J. Mol. Biol. 320, 369–387.
Hartigan, J.A., 1973. Minimum mutation fits to a given tree. Biometrics 29, 53–65.
Hensley, S.E., Das, S.R., Bailey, A.L., Schmidt, L.M., Hickman, H.D., Jayaraman, A., Viswanathan, K., Raman, R., Sasisekharan, R., Bennink, J.R., Yewdell, J.W., 2009. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science 326, 734–736.
Hughes, A.L., Nei, M., 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335, 167–170.
Hughes, A.L., Ota, T., Nei, M., 1990. Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol. Biol. Evol. 7, 515–524.
Katoh, K., Misawa, K., Kuma, K.-I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.
Kimura, M., 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267, 275–276.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208–1222.
Kryazhimskiy, S., Dushoff, J., Bazykin, G.A., Plotkin, J.B., 2011. Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet. 7, e1001301.
Laeeq, S., Smith, C.A., Wagner, S.D., Thomas, D.B., 1997. Preferential selection of receptor-binding variants of influenza virus hemagglutinin by the neutralizing antibody repertoire of transgenic mice expressing a human immunoglobulin μ minigene. 71, 2600–2605.
Massingham, T., Goldman, N., 2005. Detecting amino acid sites under positive selection and purifying selection. Genetics 169, 1753–1762.
Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., Kosakovsky Pond, S.L., 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8, e1002764.
Nakajima, K., Nobusawa, E., Tonegawa, K., Nakajima, S., 2003. Restriction of amino acid change in influenza A virus H3HA: comparison of amino acid changes observed in nature and in vitro. J. Virol. 77, 10088–10098.
Nakajima, K., Nobusawa, E., Nagy, A., Nakajima, S., 2005. Accumulation of amino acid substitutions promotes irreversible structural changes in the hemagglutinin of human influenza AH3 virus during evolution. J. Virol. 79, 6472–6477.
Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, New York.
Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.
Plotkin, J.B., Dushoff, J., 2003. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. Proc. Natl. Acad. Sci. USA 100, 7152–7157.
Pybus, O.G., Rambaut, A., Belshaw, R., Freckleton, R.P., Drummond, A.J., Holmes, E.C., 2007. Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. Mol. Biol. Evol. 24, 845–852.
Rost, B., Sander, C., 1994. Conservation and prediction of solvent accessibility in protein families. Proteins 20, 216–226.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Schymkowitz, J.W.H., Rousseau, F., Martins, I.C., Ferkinghoff-Borg, J., Stricher, F., Serrano, L., 2005. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. Proc. Natl. Acad. Sci. USA 102, 10147–10152.
Shih, A.C.-C., Hsiao, T.-C., Ho, M.-S., Li, W.-H., 2007. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. Proc. Natl. Acad. Sci. USA 104, 6283–6288.
Shope, R.E., 1931. Swine influenza. III. Filtration experiments and etiology. J. Exp. Med. 54, 373–380.
Skehel, J.J., Wiley, D.C., 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu. Rev. Biochem. 69, 531–569.
Suzuki, Y., 2004a. New methods for detecting positive selection at single amino acid sites. J. Mol. Evol. 59, 11–19.
Suzuki, Y., 2004b. Three-dimensional window analysis for detecting positive selection at structural regions of proteins. Mol. Biol. Evol. 21, 2352–2359.
Suzuki, Y., 2006. Natural selection on the influenza virus genome. Mol. Biol. Evol. 23, 1902–1911.
Suzuki, Y., 2007. Inferring natural selection operating on conservative and radical substitution at single amino acid sites. Genes Genet. Syst. 82, 341–360.
Suzuki, Y., 2010. Statistical methods for detecting natural selection from genomic data. Genes Genet. Syst. 85, 359–376.
Suzuki, Y., 2011. Positive selection for gains of *N*-linked glycosylation sites in hemagglutinin during evolution of H3N2 human influenza A virus. Genes Genet. Syst. 86, 287–294.
Suzuki, Y., Gojobori, T., 1999. A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. 16, 1315–1328.
Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. USA 101, 11030–11035.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Tokuriki, N., Tawfik, D.S., 2009. Stability effects of mutations and protein evolvability. Curr. Opin. Struct. Biol. 19, 596–604.
Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L., Tawfik, D.S., 2007. The stability effects of protein mutations appear to be universally distributed. J. Mol. Biol. 369, 1318–1332.
Tokuriki, N., Oldfield, C.J., Uversky, V.N., Berezovsky, I.N., Tawfik, D.S., 2009. Do viral proteins possess unique biophysical features? Trends Biochem. Sci. 34, 53–59.
Vanderlinden, E., Goktas, F., Cesur, Z., Froeyen, M., Reed, M.L., Russell, C.J., Cesur, N., Naesens, L., 2010. Novel inhibitors of influenza virus fusion: structure-activity relationship and interaction with the viral hemagglutinin. J. Virol. 84, 4277–4288.
Wells, J.A., 1990. Additivity of mutational effects in proteins. Biochemistry 29, 8509–8517.
Wiley, D.C., Wilson, I.A., Skehel, J.J., 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature 289, 373–378.
Williams, S.G., Madan, R., Norris, M.G.S., Archer, J., Mizuguchi, K., Robertson, D.L., Lovell, S.C., 2011. Using knowledge of protein structural constraints to predict the evolution of HIV-1. J. Mol. Biol. 410, 1023–1034.
Wilson, I.A., Skehel, J.J., Wiley, D.C., 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 289, 366–373.
Woo, J., Robertson, D.L., Lovell, S.C., 2010. Constraints on HIV-1 diversity from protein structure. J. Virol. 84, 12995–13003.
World Health Organization, 1980. A revision of the system of nomenclature for influenza viruses: a WHO memorandum. Bull. WHO 58, 585–591.
Yang, Z., 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. J. Mol. Evol. 51, 423–432.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.-M.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.