HAL
open science

Passage in egg culture is a major cause of apparent positive selection in influenza B haemagglutinin.

Derek Gatherer

To cite this version:

Derek Gatherer. Passage in egg culture is a major cause of apparent positive selection in influenza B haemagglutinin.. Journal of Medical Virology, 2009, 82 (1), pp.123. 10.1002/jmv.21648 . hal-00541147

HAL Id: hal-00541147

https://hal.science/hal-00541147v1

Submitted on 30 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal of Medical Virology

Passage in egg culture is a major cause of apparent positive selection in influenza B haemagglutinin.

Journal:	*Journal of Medical Virology*
Manuscript ID:	JMV-09-1369.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	29-Jul-2009
Complete List of Authors:	Gatherer, Derek; University of Glasgow, MRC Virology Unit
Keywords:	N-glycosylation , non-synonymous substitution, codeml, immune evasion, adaptation

scholarONE™
Manuscript Central

TABLE I

Site 197-199	*Egg-cultured*	*Egg-free*
With N-glycosylation	14	50
Without N-glycosylation	49	9

Site 233-235	*Egg-cultured*	*Egg-free*
With N-glycosylation	29	22
Without N-glycosylation	34	37

Table I: Presence or absence of N-glycosylation site at positions 197-199 and 233-235 in egg-cultured and egg-free strains. Chi-square test significant only for site 197-199 ($p < 0.0001$).

For Peer Review

TABLE II

Site 197-199	*Yamagata*	*Victoria*
With N-glycosylation	46	18
Without N-glycosylation	33	25

Site 233-235	*Yamagata*	*Victoria*
With N-glycosylation	13	38
Without N-glycosylation	66	5

Table II: Presence or absence of N-glycosylation sites at positions 197-199 and 233-235 in Yamagata and Victoria strains. Chi-square test significant only for site 233-235 ($p < 0.0001$).

For Peer Review

TABLE III

	EGG-CULTURED			
M8	Site	Pr(omega>1)	omega	SE
	75	0.994*	9.843	1.092
	116	0.774	7.793	3.827
	129	0.916	9.115	2.638
	197	1.000*	9.902	0.820
	199	1.000*	9.902	0.820
M2	75	0.952*	9.797	2.043
	197	1.000*	10.246	0.522
	199	1.000*	10.246	0.522

	EGG-FREE			
M8	Site	Pr(omega>1)	omega	SE
	48	0.801	1.523	0.613
	56	0.908	1.657	0.527
	73	0.857	1.595	0.583
	75	0.977*	1.729	0.462
	116	0.821	1.549	0.602
	129	0.884	1.629	0.559
	146	0.817	1.544	0.604
	199	0.994*	1.743	0.445
M2	56	0.727	1.662	0.651
	75	0.879	1.835	0.720
	129	0.725	1.676	0.692
	199	0.944	1.880	0.699

For Peer Review

Table III: Positive selection in influenza B HA in egg-cultured and egg-free strains. Pr(omega>1): probability that the site in question is subject to positive selection; SE: standard error of omega. Sites 197 and 199 are part of the N-glycosylation site commonly lost in egg-cultured strains. All sites with a Pr(omega>1) of greater than 0.7 are listed. Asterisks indicate Pr(omega>1) of >0.95

For Peer Review

TABLE IV

M2	*p >= 0.5*	*p >= 0.75*	*p >= 0.95*	*p >= 0.99*	*Max. omega*
Egg-cultured	4	3	3	2	10.246
Egg-free	16	2	0	0	1.880

M8	*p >= 0.5*	*p >= 0.75*	*p >= 0.95*	*p >= 0.99*	*Max. omega*
Egg-cultured	5	5	3	3	9.902
Egg-free	22	8	2	1	1.743

Table IV: Summary of number of candidate positively selected sites and the strength of selection, in egg-cultured and egg-free strains. Model M2 allows for 3 selective categories, namely constraint, neutrality and positive selection, and M8 for a beta distribution of 10 constrained or neutral categories plus an eleventh category of positive selection. *p >= x*: number of sites positively selected at probability of greater than *x*. *Max. omega*: largest sitewise omega value found in the data.

For Peer Review

Passage in egg culture is a major cause of apparent positive selection in influenza B haemagglutinin.

Derek Gatherer

MRC Virology Unit

University of Glasgow

Church Street

Glasgow

G11 5JR, UK

Correspondence to:

Derek Gatherer, MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UK.

Tel: +44 141 330 6268

Fax: +44 141 337 2236

email: d.gatherer@mrcvu.gla.ac.uk

Running title: Positive selection in Influenza B

For Peer Review

Abstract

Several studies have identified residues apparently under positive selection in influenza B virus haemagglutinin. Host immune evasion is the main mechanism proposed to exert this selection pressure. However, these reports have not considered the culture history of the strains used to calculate positive selection. This paper shows that passage of influenza B virus through egg culture is a strong contributory factor to the strength and statistical significance of positive selection on haemagglutinin. Non-synonymous mutations resulting in the loss of the N-glycosylation site at positions 197-199 of haemagglutinin have been positively selected to a far greater degree in egg-cultured strains than in other strains. Once egg-cultured strains are removed from the analysis, positive selection is found to be far weaker, less statistically significant and more diffusely localised along the protein. Caution should therefore be exercised both in claims for the existence of positive selection in influenza B haemagglutinin, and in attribution of host immune evasion as its cause. The major cause of molecular adaptation in influenza B haemagglutinin proteins may well be laboratory eggs rather than natural hosts.

Keywords: N-glycosylation, immune evasion, non-synonymous substitution, codeml, omega, adaptation

For Peer Review

Introduction

Influenza A (Family *Orthomyxoviridae*, Genus *Influenzavirus A*, Species *Influenza A virus*) and the related influenza B (Family *Orthomyxoviridae*, Genus *Influenzavirus B*, Species *Influenza B virus*) are major public health concerns. A lack of lifelong immunity in human hosts results in seasonal epidemics, with annual vaccination the best current strategy for limiting mortality and the economic impact of these viruses.

The earliest viral isolates confirmed as influenza B by haemagglutinin (HA) gene sequencing date from 1940 (strains B/Lee/0/1940 to B/Lee/3/1940). Other isolates document the continued presence of the virus during the subsequent two decades (strains B/Maryland/0/59 and B/Russia/0/69). Since the late 1980s, systematic seasonal sequencing of new strains has been performed, revealing that all influenza B HA gene sequences of the last 20 years belong to one or other of two lineages, designated Yamagata and Victoria after their original isolates (B/Yamagata/16/88 and B/Victoria/2/87 respectively). HA protein sequences within each of the two lineages are more than 97% identical, with sequence identity in inter-lineage comparisons at 88-90% on average. The original Lee strain HA from 1940 is approximately 91% identical to modern samples of either lineage. By comparison, influenza A HA sequences are only 22-23% identical to influenza B HA.

It has been proposed that influenza B, like influenza A, achieves host immune evasion by rapid evolution of HA. Mutations causing amino acid substitutions at certain residues of HA are hypothesised to afford a selective advantage to the virus, resulting in a statistically significant excess of non-synonymous over synonymous substitutions at the codons corresponding to these residues [Shen et al., 2009]. Positive selection is the term given to such selection for non-synonymous substitutions. The ratio of the rates of non-synonymous to synonymous substitution is termed omega. An omega of less than 1 indicates selective constraint and suppression of non-synonymous substitutions. In

For Peer Review

neutrally evolving proteins omega will be close to 1, and positive selection produces omegas which may be considerably greater than 1. PAML [Yang, 1997] provides a suite of tools for the calculation of omega and determination of its statistical significance. Omega may be calculated as an average over the whole length of a gene, or in a sitewise manner. The latter has proved useful for identifying individual residues that are under selective pressure.

Previous studies have detected codons with omega significantly greater than 1 in HA genes from influenza B, and attribute this positive selection to immune evasion on the grounds that the detected sites are within HA epitope regions [Nunes et al., 2004; Nunes et al., 2008; Pechirra et al., 2005; Shen et al., 2009]. However, some of the candidate positively selected residues are associated with an N-glycosylation site, and mutations causing the loss of that site are often found in laboratory strains passaged in embryonated hen eggs [Robertson et al., 1985; Saito et al., 2004], where the absence of the site can result in increased binding to the plasma membranes of chorio-allantoic cells [Gambaryan et al., 1999]. By contrast, loss of this N-glycosylation site can also occur in strains cultured only in Madin Darby canine kidney (MDCK) cells [Ikonen et al., 2005] and some egg-cultured strains can retain it [Rota et al., 1992]. A full survey of the association of N-glycosylation site presence with egg culture has not been undertaken so far.

The possibility therefore remains that host immune evasion is being confused with egg culture adaptation as a cause of positive selection in influenza B HA genes. This study analyses positive selection in an extensive set of influenza B HA genes, separating genes from egg-cultured strains from those of cell-cultured strains. Studies published previously of positive selection in influenza B HA genes have not performed this analysis. A survey of the association of N-glycosylation site presence with egg culture is also described. The relative contribution of egg culture to detected positive selection may therefore be more fully understood.

For Peer Review

Materials & Methods

Identification of culture history

259 full length non-redundant influenza B HA coding sequences, dating from 1940 to 2007, were retrieved in GenBank format from the NCBI Influenza Virus Resource (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html). GenBank format includes the "/lab_host" tag, annotating the culture history of the strain. Of the 259 sequences, 133 had a "/lab_host" tag. Of these, 63 sequences were listed as having had contact with egg culture (Supplementary Data Table 1). For instance, entry CY018373 (B/Michigan/04/2006) has "/lab_host" tag: "Egg 2 passage(s)". 70 sequences had no mention of eggs, for example CY018301 (B/Houston/B720/2004) has "/lab_host" tag: "RhMK 1 passage(s)", indicating the use of a rhesus monkey kidney cell line. This initial set of 70 sequences was reduced to 59 by removal of those with unknown or ambiguous history (Supplementary Data Table 2). For example, entry CY018725 (B/Paraguay/636/2003) has "/lab_host" tag: "Unknown 1 MDCK 2 passage(s)", and was therefore omitted from the final set of 59 putative egg-free sequences, on the grounds that "Unknown" could indicate previous egg culture. The egg-cultured strains range in date from 1987 to 2006 and the egg-free strains from 1992 to 2006. The proportions of egg-cultured and egg-free strains are roughly equal for most years after 1997. The lineages of the egg-free strains are 23 Victoria and 36 Yamagata strains. For egg-cultured strains, there are 20 Victoria and 43 Yamagata strains.

N-glycosylation site analysis

Due to indels in HA, the N-glycosylation site discussed in the present paper may elsewhere be referred to as occurring at residues 197 to 199 [Ikonen et al., 2005; Kinnunen et al., 1992; Nunes et al., 2004; Nunes et al., 2008; Pechirra et al., 2005; Rota et al., 1992; Rota et al., 1990; Shaw et al., 2002], 196 to 198 [Robertson et al., 1985; Saito et al., 2004] or 194 to 196 [Shen et al., 2009]. In this study, the majority

For Peer Review

system is adopted: 197-199. For each HA, the presence or absence of the N-glycosylation consensus sequence (NxS/T, where x is any amino acid) was recorded at 197-199 and the nearby 233-235, which has no previous association with positive selection and therefore serves as a control. A 2x2 chi-square test was used to determine the statistical significance of the proportion of N-glycosylation sites in the egg-cultured and egg-free sets, and also the proportion of N-glycosylation sites in the Yamagata and Victoria lineages.

Quantitation of positive selection

Sequences were translated using transeq (EMBOSS) [Rice et al., 2000], then aligned using MAFFT [Katoh et al., 2005]. The protein alignment was then used as a basis for alignment of the DNA sequences. PAML's CODEML mode [Yang, 1997] was used to quantify positive selection in these alignments. Comparisons between models M1 (2 discrete categories: constraint plus neutrality only) and M2 (3 discrete categories: constraint, neutrality and positive selection), between models M7 (10 beta-distributed categories: constraint plus neutrality only) and M8 (11 categories: as M7 with 1 further discrete category for positive selection), and between models M8 and M8a (as M8 but with maximum omega fixed at 1), were used to assess the statistical significance of detected positive selection.

Results

N-glycosylation and egg culture

The presence or absence of N-glycosylation sites at positions 197-199 and 233-235 was examined in HA sequences annotated as egg-cultured and those annotated as mammalian cell-cultured with no mention of "unknown" culture conditions (the "egg-free" strains). The first null hypothesis is that culture conditions do not affect the presence or absence of N-glycosylation sites at either position. Table I shows the

For Peer Review

statistics for the occurrence of the relevant N-glycosylation sites. A 2x2 chi-square test is significant only for site 197-199. Thus the null hypothesis, that culture conditions do not affect N-glycosylation state, can be rejected for site 197-199 only. Egg culture is therefore demonstrated to be a likely cause of loss of N-glycosylation site 197-199 but to be unlikely to have any effect on site 233-235. The second null hypothesis is that a strain's ancestry does not affect the presence or absence of N-glycosylation sites at either position. Table II similarly shows the N-glycosylation site statistics divided between Yamagata and Victoria strains. This time, a 2x2 chi-square test is significant only for site 233-235. Although N-glycosylation is slightly less common in Victoria strains at position 197-199, this is not statistically significant. The second null hypothesis, that lineage does not affect N-glycosylation state, can be rejected for site 233-235 only. At that site, Victoria strains are predominantly N-glycosylated, whereas Yamagata strains are predominantly non-N-glycosylated. Therefore, the presence of the N-glycosylation site of HA at position 233-235 is mainly an aspect of the ancestry of the HA sequence and is not affected by culture conditions. The converse applies for the site at position 197-199 where the relative absence of the N-glycosylation site is a result of selective pressure from culture conditions and there is no correlation with ancestry of the strain. The possibility that the absence of the N-glycosylation site at positions 197-199 is merely a result of greater presence of one or other of the 2 lineages in egg culture is also thereby excluded.

Positively selected sites on HA

Table III shows the PAML output for egg-cultured and egg-free strains. Each set was analysed using Bayes Empirical Bayes (BEB) [Yang et al., 2005], under models M2 (3 discrete categories, omega <1, omega=1 and omega >1) and M8 (10 beta-distributed rate categories from omega of zero to 1, plus 1 discrete category of omega >1). Table IV summarizes the frequency and intensity of positively selected sites for the two models in egg-cultured and egg-free strains.

For Peer Review

For egg-cultured strains, position 75 and the N-glycosylation site 197-199 are strongly positively selected (omega >9) at high probability (p > 0.99), using model M8. Positions 116 and 129 have slightly weaker selection (omegas > 7.5) at lower statistical probabilities (p>0.7 for 116 and p>0.9 for 129). Model M2 detects only positions 75, 197 and 199. This confirms previous results [Nunes et al., 2004; Nunes et al., 2008; Pechirra et al., 2005]. However, in egg-free strains detected positive selection is far weaker. Using model M8, positions 75 and 199 have weak positive selection only (omega < 1.8). With model M2, statistical probabilities of positive selection are also weakened. The apparent positive selection detected in previous studies evaporates to a large degree when egg-cultured strains are excluded from the analysis.

One intriguing further observation (Table IV) is that egg-free strains have more candidate positively selected sites at lower significance levels (0.5 < p < 0.95). Although caution must be exercised in interpretation of this weaker signal, it may indicate more subtle positive selection going on at a wider variety of HA sites in mammalian cell culture or in the human host. In egg-culture, very strong positive selection focuses on a rather small number of sites, of which N-glycosylation site 197-199 is prominent.

Discussion

Detection of the actual N-glycosylation state of an HA protein requires laboratory analysis that is not performed here. However, the assumption is made that the N-glycosylation consensus sequence is a direct indicator of a true N-glycosylation site.

The N-glycosylation site at position 197-199 was chosen as it has previously been described as suppressed in strains cultured in eggs, and also to be a site subject to positive selection. A neighbouring N-glycosylation site at positions 233-235 was also

For Peer Review

chosen as a control since it has not been implicated in either positive selection or culture-dependent loss.

The sites detected in the present study as positively selected at p>0.95: namely 75, 197 and 199, are exactly the same as those detected by Nunes et al. [2004; 2008] and Pechirra et al. [2005], but differ to some extent from those of Shen et al. [2009]. These latter authors point out that their study divides the data set into Yamagata and Victoria lineages, and into time series windows. The residues identified as positively selected differ in each data set as a result of the lineage-specific and time-variable selective pressures.

The possibility cannot be excluded that the egg-free dataset contains some contaminating egg-cultured sequences that have been incorrectly annotated. Therefore, even the residual levels of positive selection seen in egg-free sequences (Table III) may be misleadingly high. If a truly egg-free set were to be obtainable, it is predicted that the signal of positive selection would be even weaker.

In summary, caution is necessary to avoid over-interpretation of positive selection scores in influenza B HA sequences. Although positive selection has now been demonstrated by several studies, it may be considerably weaker and less significant once a correction is made for egg-culture. The correction made in the present study may even be only a partial one, as it cannot be guaranteed that all strains without annotation for egg culture or unknown culture conditions are truly egg-free in their culture history. Recent papers on positive selection in influenza B have failed to address this important issue. Really strong omega values (around 9 or 10) are only found in egg-cultured strains. Inclusion of egg-cultured strains in an analysis may elevate overall significance levels and omega values. Under such conditions it is easy to erroneously attribute a high positive selection score to conditions in the host organism, particularly immune evasion. Perhaps the biggest factor in apparent positive selection in influenza B

For Peer Review

HA is adaptation to egg culture rather than to any host. It is suggested that future analysis of positive selection should only be carried out on sequences free from culture conditions of any kind, e.g. directly amplified by PCR from clinical material. Only then will a true picture be seen of positive selection and its relation to host immune evasion in influenza B.

Acknowledgments

The author thanks Prof. Duncan McGeoch (MRC Virology Unit, Glasgow) for sustained encouragement and support over the last 6 years, and also Prof. Bill Carman and Dr Joy Kean (West of Scotland Specialist Virology Centre, Glasgow) for pointing him in the direction of influenza B.

Gambaryan AS, Robertson JS, Matrosovich MN. 1999. Effects of egg-adaptation on the receptor-binding properties of human influenza A and B viruses. Virology 258(2):232-239.

Ikonen N, Pyhala R, Axelin T, Kleemola M, Korpela H. 2005. Reappearance of influenza B/Victoria/2/87-lineage viruses: epidemic activity, genetic diversity and vaccination efficacy in the Finnish Defence Forces. Epidemiol Infect 133(2):263-271.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33(2):511-518.

Kinnunen L, Ikonen N, Poyry T, Pyhala R. 1992. Evolution of influenza B/Victoria/2/87-like viruses: occurrence of a genetically conserved virus under conditions of low epidemic activity. J Gen Virol 73 (Pt 3):733-736.

Nunes B, Pechirra P, Canto e Castro L, Rebelo-de-Andrade H. 2004. Adaptive evolution on HA1 subunit of influenza B virus. International Congress Series 1263:691-694.

Nunes B, Pechirra P, Coelho A, Ribeiro C, Arraiolos A, Rebelo-de-Andrade H. 2008. Heterogeneous selective pressure acting on influenza B Victoria- and Yamagata-like hemagglutinins. J Mol Evol 67(4):427-435.

Pechirra P, Nunes B, Coelho A, Ribeiro C, Goncalves P, Pedro S, Castro LC, Rebelo-de-Andrade H. 2005. Molecular characterization of the HA gene of influenza type B viruses. J Med Virol 77(4):541-549.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16(6):276-277.

Robertson JS, Naeve CW, Webster RG, Bootman JS, Newman R, Schild GC. 1985. Alterations in the hemagglutinin associated with adaptation of influenza B virus to growth in eggs. Virology 143(1):166-174.

Rota PA, Hemphill ML, Whistler T, Regnery HL, Kendal AP. 1992. Antigenic and genetic characterization of the haemagglutinins of recent cocirculating strains of influenza B virus. J Gen Virol 73 (Pt 10):2737-2742.

For Peer Review

Rota PA, Wallis TR, Harmon MW, Rota JS, Kendal AP, Nerome K. 1990. Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983. Virology 175(1):59-68.

Saito T, Nakaya Y, Suzuki T, Ito R, Saito H, Takao S, Sahara K, Odagiri T, Murata T, Usui T, Suzuki Y, Tashiro M. 2004. Antigenic alteration of influenza B virus associated with loss of a glycosylation site due to host-cell adaptation. J Med Virol 74(2):336-343.

Shaw MW, Xu X, Li Y, Normand S, Ueki RT, Kunimoto GY, Hall H, Klimov A, Cox NJ, Subbarao K. 2002. Reappearance and global spread of variants of influenza B/Victoria/2/87 lineage viruses in the 2000-2001 and 2001-2002 seasons. Virology 303(1):1-8.

Shen J, Kirk BD, Ma J, Wang Q. 2009. Diversifying selective pressure on influenza B virus hemagglutinin. J Med Virol 81(1):114-124.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13(5):555-556.

Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22(4):1107-1118.

For Peer Review