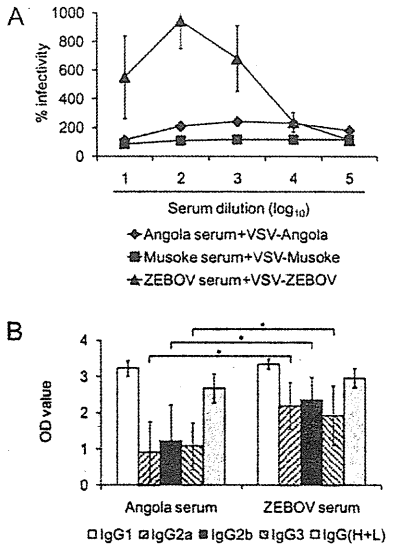


Figure 2. C1q-dependent antibody-dependent enhancement of polyclonal antisera from intraperitoneally immunized mice *A,* and immunoglobulin (Ig) G subclass in the serum samples *B,* Vesicular stomatitis virus (VSV) pseudotyped with Marburg virus glycoproteins (MARV GPs) were incubated with mouse antisera ($1:10^1$ to $1:10^5$ dilutions) in the presence of purified C1q (50 μg/mL) and inoculated into human embryonic kidney 293 cells. Results are expressed as means (± standard deviations) of data for 3 immunized mice. Relative percentages of infected cells are shown as mean values determined based on the number of infected cells in the absence of specific antibodies against MARV GPs (100%). The binding activity of each antibody class in 3 immunized mouse antisera ($1:10^2$) was measured by enzyme-linked immunosorbent assay (ELISA) using histidine-tagged GPs as antigens. Experiments were triplicated and the means and standard deviations were shown. The differences between optical density (OD) values were compared using the Student's *t* test. $^*P < .05$ for differences in OD values between *Angola* and *Zaire ebolavirus* (ZEBOV) GPs.

Characterization of mAbs Specific to MARV GPs

We then generated hybridoma cells producing specific mAbs (47 and 28 clones for Angola and Musoke GPs, respectively). Table 1 lists the number of clones and their characteristics (ie, neutralizing or ADE). Though there was no significant difference by the χ^2 test in the percentage of cell clones producing neutralizing antibodies between the Angola- and Musoke-specific clones, the percentage of clones showing FcR-dependent ADE activity was significantly higher after immunization with Angola GP. The clones producing the enhancing antibodies were further classified into 2 groups; the first (group A) consisted of mAbs that enhanced the infectivity of both VSV-Angola and VSV-Musoke, and the second (group B) consisted of mAbs that enhanced only VSV-Angola infectivity. No mAb showed C1q-dependent ADE activity (data not shown).

Identification of ADE Epitopes on MARV GPs

To identify epitopes involved in ADE activity of mAbs, we tested the representative mAbs (IgG1) from groups A and B to

Table 1. Characterization of Monoclonal Antibody (mAb) Clones Specific to Marburg Virus Glycoprotein (GP)

	mAb Clones, no. (%)	
Clone characteristic	Angola GP	Musoke GP
ELISA positive	47 (100)	28 (100)
Neutralizing	19 (40)	9 (32)
Enhancing	15 (32) [a]	0 (0)

NOTE. Differences in the ratios of the clone numbers between Angola and Musoke GPs were analyzed by χ^2 test, and statistical significance was established at $P < .05$. ELISA, enzyme-linked immunosorbent assay.

[a] $P < .05$.

determine reactivity to synthetic peptides derived from Angola and Musoke GP (PEPscreen; Sigma), and we found that all mAbs belonging to group A bound strongly to peptides corresponding to amino acid positions 330–350 and 410–430 (data not shown). We then constructed mutant GPs lacking amino acid residues 330–350 (AΔ1) or 410–430 (AΔ2) and examined the reactivity of mAbs of group A (Figure 3). In immunostaining of HEK 293T cells expressing the mutant GPs, these mAbs bound to AΔ1 but not AΔ2, suggesting that the epitopes for these mAbs lay between amino acid positions 410–430 where the sequences are partially shared between Angola and Musoke GPs. By contrast, mAbs belonging to group B failed to bind to any synthetic peptide, suggesting that their epitopes were likely in nonlinear conformation. Because mAbs of group B did not bind to Musoke GP, we constructed chimeric GPs between Angola and Musoke GPs (MA1–MA5), as indicated in Figure 3. Based on the reactivity of mAbs to the chimeric GPs expressed in HEK 293T cells, 2 regions on Angola GP were identified as the epitopes of mAbs of group B (Figure 3). mAbs AGP2-1, AGP40-32, and AGP126-10 reacted strongly with MA1 and MA2 but not MA3, MA4, and MA5, indicating that these mAbs recognized an epitope that resided between amino acid positions 369 and 385 in Angola GP. Another mAb, AGP90-28, bound to 4 chimeric GPs, MA1, MA2, MA3, and MA4, but not MA5, indicating that the epitope for this mAb was located between amino acid positions 402 and 418 in Angola GP. All of these identified epitopes were located within the MLR.

Contribution of GP2 Region to ADE

To confirm the contribution of the MLR of MARV in ADE of these mAbs, we constructed chimeric GPs whose MLRs were swapped between Angola and Musoke GPs (AMA and MAM), and the levels of infectivity of VSV pseudotyped with these chimeric GPs (VSV-AMA and VSV-MAM) were compared in K562 cells (Figure 4). Unexpectedly, however, the infectivity of VSV-MAM was not enhanced by Angola MLR-specific mAbs (ie, mAbs A2-1 and A90-28 in group B). Furthermore, in the presence of mAb AGP127-8 belonging to group A, the infectivity of VSV-AMA was still significantly higher than that of

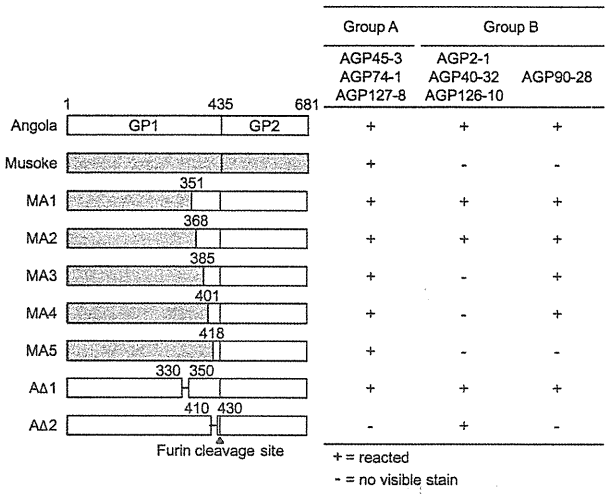


Figure 3. Reactivity of monoclonal antibodies to chimeric and deletion mutant glycoproteins (GPs). Reactivity was tested by immunostaining, as described in Materials and Methods. Numbers indicate amino acid positions of the GP primary amino acid sequence.

VSV-Musoke, and replacement of the MLR of Musoke GP with that of Angola GP (VSV-MAM) had little effect on the enhancement of infectivity compared with the infectivity of VSV-Musoke. We then examined the infectivity of VSV pseudotyped with a series of chimeric GP mutants, as shown in Figure 4, and found that only the infectivity of VSV-MAA, which had the MLR and part of GP2 (shown as GP2C in Figure 4) derived from Angola GP, was reasonably enhanced by mAbs of both groups A and B. These results indicated that both the MLR and the GP2 region of Angola GP were required for maximal ADE activity of these mAbs.

Importance of an Amino Acid at Position 547 for ADE

The GP2C regions of the Angola and Musoke GPs differ in 4 amino acids: positions 504, 547, 596, and 618. To identify which amino acid in GP2C contributed to the differential infectivity enhancement by mAbs AGP127–8, A2-1, and A90-28, the following 8 mutant GPs that contained single-amino-acid substitutions were constructed: 4 Angola-based mutant GPs (A/H504T, A/G547V, A/A596T, and A/R618K) and 4 Musoke-based mutant GPs (M/T504H, M/V547G, M/T596A, and M/K618R) (Figure 5). The infectivity of VSV pseudotyped with these mutant GPs in the presence of these mAbs was compared in K562 cells. We found that the substitution at position 547 clearly switched the viral infectivity in the presence of mAb AGP127-8, as shown by substantially enhanced and reduced infectivity of VSV-M/V547G and A/G547V, respectively, when compared with VSV pseudotyped with wild-type Musoke and Angola GPs. By contrast, the mutations at position 504, 596, and 618 had limited effects on the infectivity of the respective viruses. Interestingly, ADE of VSV-A/G547V infectivity was not observed, even in the presence of Angola GP-specific mAbs, AGP2-1 and AGP90-28. These results indicated that amino acid position 547 played an important role in the ADE activities of these mAbs. To examine whether reduced ADE activities of mAbs resulted from the loss of binding affinity due to the structural change by the mutation at position 547, we measured reactivity to each GP with ELISA (Figure 6). We found that Angola and Musoke GP antisera reacted similarly to all the GP

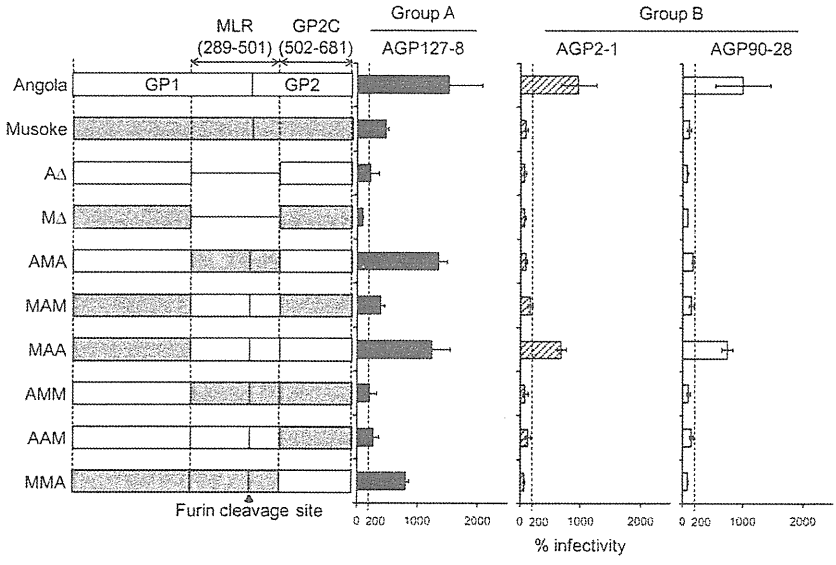


Figure 4. Infectivity of vesicular stomatitis virus pseudotyped with deletion or chimeric mutant glycoproteins (GPs) in K562 cells. VSV pseudotyped with mutant GPs was incubated with monoclonal antibodies AGP127-8, AGP2-1, and AGP90-28 (1 μg/mL) and then inoculated into K562 cells. All experiments were done in triplicate, and means and standard deviations are shown. Numbers indicate amino acid positions of the GP primary amino acid sequence.

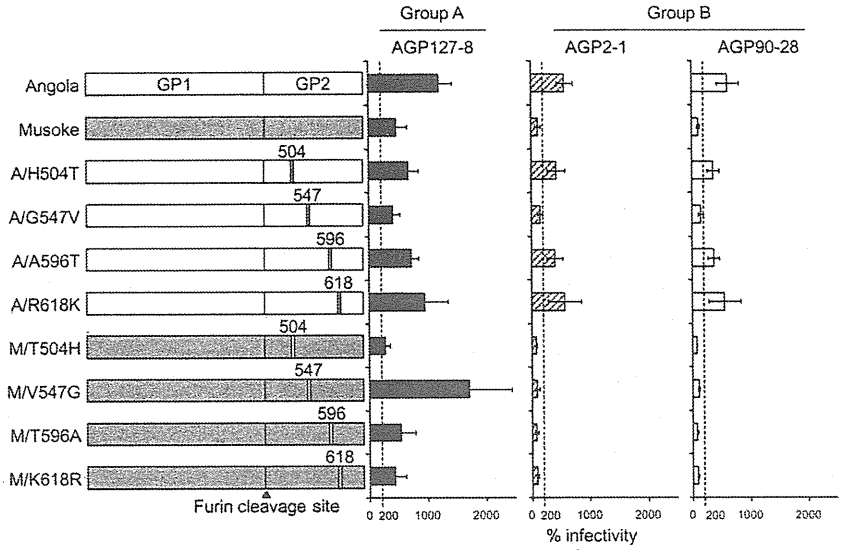


Figure 5. Infectivity of vesicular stomatitis virus pseudotyped with the mutant glycoproteins (GPs) with a single substitution in K562 cells. Experimental conditions were the same as described for Figures 1 and 4. All experiments were done in triplicate, and means and standard deviations are shown.

antigens, suggesting that the substitution at position 547 did not change the overall antigenic structure of MARV GPs. Importantly, this substitution resulted in no significant change in reactivity of these mAbs, indicating that the substitution at position 547 did not affect the binding affinity of these mAbs to GPs.

DISCUSSION

It has been shown that in addition to the common receptor- or coreceptor-dependent mechanism of cellular attachment, some viruses use antiviral antibodies for efficient entry into target cells [21]. Although this phenomenon, ADE of EBOV, has been

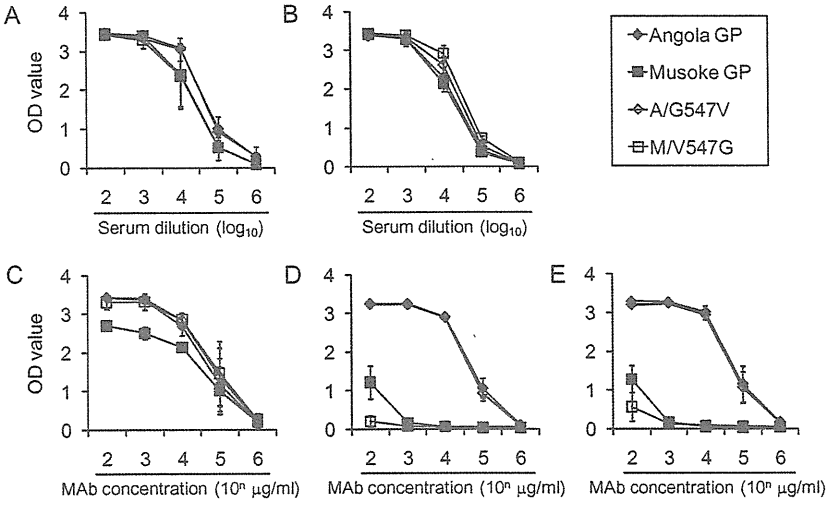


Figure 6. Binding affinities of the mouse antisera and monoclonal antibodies to Marburg virus glycoproteins (MARV GPs). Serial 10-fold dilutions of the mouse antisera to Angola GP *A* and Musoke GP *B* or indicated concentrations of mAbs AGP127-8 *C*, AGP 2-1 *D*, and AGP 90-28 *E*, were tested by enzyme-linked immunosorbent assay (ELISA) for reactivity to wild-type GPs and GPs with single amino acid substitutions. Viruslike particles expressing each GP were used as ELISA antigens. Each optical density (OD) value represents the means and standard deviations of 2 independent experiments.

studied by using mouse mAbs, convalescent human, and experimentally infected nonhuman primate serum samples [19, 20, 22], the possible contribution of MARV-specific antibodies to ADE has not been reported previously to our knowledge.

In the present study, we demonstrated that the infectivity of VSV-Angola in K562 cells was enhanced notably in the presence of Angola GP antisera (ie, FcR-dependent ADE), whereas Musoke GP antisera did not significantly enhance the infectivity of VSV-Angola or -Musoke (Figure 1). This difference between the 2 MARV strains was also supported by the observation that immunization with Angola GP induced significantly higher numbers of B-cell clones with infectivity-enhancing properties than did immunization with Musoke GP (Table 1). These results may suggest that the potential difference in the pathogenicity between the MARV strains Angola and Musoke might be partially explained by the ability to induce infectivity-enhancing antibodies, as was proposed for the distinct pathogenicity seen with ZEBOV and Reston EBOV [20, 22].

However, it has been argued whether the strain Angola is more virulent for humans and nonhuman primates than the strain Musoke. Along with the high case fatality rate in the Angola epidemic, it was also noted that in most earlier outbreaks of Marburg hemorrhagic fever, the case fatality rates did not exceed 50% [2, 4, 5], but there was another outbreak in which the case fatality rate was 83%, quite similar to what was seen in the Angola outbreak [29]. In experimental infection of macaques, 3 animals infected with the strain Angola died after illness that was more rapidly progressive than those caused by other viruses tested in other experiments [8, 9, 30], and necropsy showed infection of the liver with accompanying necrosis. However, because animals infected with the strains Angola and Musoke have not been compared in the same experiment, it could only be concluded that infection with the strain Angola "appeared to progress more rapidly" than other MARVs tested in other studies. In addition, one of the patients infected with the strain Musoke developed a rapidly progressive illness that closely resembled the features of illness seen in macaques experimentally infected with the strain Angola [5, 9]. Thus, although our results may suggest the possible contribution of ADE to differences in pathogenicity between the strains Angola and Musoke, further investigations are needed to conclude that the strain Angola has a uniquely higher pathogenicity.

Compared with ZEBOV GP antisera, Angola GP antisera enhanced the infectivity of VSV-Angola less efficiently in the presence of C1q (Figure 2*A*). We therefore compared the profile of IgG subclasses in the antisera. A significant difference was seen in the levels of IgG2a, IgG2b, and IgG3 between MARV Angola and ZEBOV GP antisera (Figure 2*B*). It is likely that IgG2a and IgG3 play a more prominent role in C1q-dependent ADE activity, because these classes of antibodies are thought to have a higher affinity for C1q molecules than IgG1 [20]. Thus, this difference may account for the lower activity of C1q-dependent ADE in Angola antisera. It will be of interest to elucidate why MARV and EBOV GPs induce a distinct antibody repertoire.

It was shown that most of the ADE epitopes of EBOV GP were located in the MLR [20]. In the present study, we identified 3 distinct ADE epitopes on MARV GP, and these ADE epitopes were mostly located in the MLR of the GP1 subunit (Figure 3). To further ascertain the contribution of the MLR of MARV GP to ADE, we examined the infectivity of VSV pseudotyped with the Angola GP lacking the MLR in K562 cells in the presence of the Angola GP antisera and found that ADE activity was reduced drastically (data not shown). These results suggested that the MLR of MARV GP also contained dominant epitopes for antibodies that enhance viral infectivity.

Two ADE epitopes, amino acid positions 369–385 and 402–418 (group B epitopes), were recognized by Angola GP-specific mAbs, whereas another epitope, amino acids 410–430, was shared between the 2 GPs (group A epitope). Accordingly, mAbs belonging to group A bound to both Angola and Musoke GP and enhanced the infectivity of both VSV-Angola and -Musoke (Figures 4 and 5). These results suggest that Angola and Musoke GPs share some epitopes recognized by the ADE antibodies. However, Musoke GP antisera and the Musoke GP-specific mAbs did not show ADE activity, suggesting that Musoke GP contained fewer ADE epitopes than Angola GP. Thus, we propose that the primary amino acid structure and glycosylation pattern of the MLR may influence epitope exposure on the GP molecule.

It was noted that the substitution at amino acid position 547 in the GP2 region affected ADE significantly but did not change the binding affinity of the mAbs to the Angola and Musoke GPs (Figure 6). Amino acid position 547 is located at the base of the internal fusion loop [24]. We demonstrated elsewhere that a single-amino acid substitution at position 547 affected the efficiency of C-type lectin-mediated entry and the mechanism underlying endosomal entry such as proteolytic processing by endosomal cathepsin [24]. The present findings also suggest that the amino acid at position 547 plays a role in entry mediated through the ADE pathway. When virus particles are internalized in endosomes via the ADE pathway, it is possible that (1) glycine at position 547 weakens the interaction between GP1 and GP2, resulting in increased cathepsin susceptibility or reduced cathepsin dependence in endosomes and (2) the amino acid at position 547 affects the flexibility of the fusion loop and/or the conformational change needed for fusion. It would be interesting to investigate the role of this particular amino acid for proper GP functions.

In this study, we demonstrated in vitro ADE for MARV that might be associated with the distinct pathogenicity of certain MARV strains. The ADE epitopes were localized in the MLR of GP. Furthermore, our data suggest that the efficiency of ADE-mediated entry of MARV is controlled by the epitope structure on GP and the mechanisms underlying endosomal

entry. Further in vitro studies are required to prove the contribution of ADE to the exacerbation of MARV infection; these studies should use serum samples from infected monkeys and patients to confirm whether ADE is induced by actual MARV infection and include animal experiments using a reverse-genetics system to introduce mutations into the ADE epitopes and reduce the production of ADE antibodies. The present study, however, together with our previous study [24], provides new insights into the molecular basis of MARV entry, mediated by interaction between the GP MLR and cellular attachment factors, such as C-type lectins and MLR-specific antibodies.

Funding

This work was supported by Research Fellowships for Young Scientists from the Japan Society for the Promotion of Science, the Takeda Science Foundation, a Grant-in-Aid for Scientific Research on Priority Areas (grant 19041001), and, in part, by the Program of Founding Research Centers for Emerging and Reemerging Infectious Diseases (grant 05021011) and the Global COE Program "Establishment of International Collaboration Centers for Zoonosis Control" (grant F-001) from the Ministry of Education, Culture, Sports, Science and Technology, Japan (http://www.mext.go.jp/english/index.htm). Filovirus work at the Rocky Mountain Laboratories is funded by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health.

Acknowledgments

We thank Hiroko Miyamoto, Ayaka Yokoyama, Teiji Murakami, and Aiko Ohnuma for technical assistance and Kim Barrymore for editing the manuscript.

References

1. Sanchez A, Geisbert TW, Feldmann H. Filoviridae: Marburg and Ebola viruses. In Knipe DM, Howley PM, Griffin DE eds. Fields virology, 5th ed. Philadelphia, PA: Lippincott Williams & Wilkins, **2006**; 1409–48.
2. Martini G. Marburg virus disease. Postgrad Med J **1973**; 49:542–6.
3. Johnson E, Johnson B, Silverstein D, et al. Characterization of a new Marburg virus isolated from a 1987 fatal case in Kenya. Arch Virol Suppl **1996**; 11:101–14.
4. Gear J, Cassel G, Gear A, et al. Outbreak of Marburg virus disease in Johannesburg. Br Med J **1975**; 4:489–93.
5. Smith D, Johnson B, Isaacson M, et al. Marburg-virus disease in Kenya. Lancet **1982**; 1:816–20.
6. Bausch D, Borchert M, Grein T, et al. Risk factors for Marburg hemorrhagic fever, Democratic Republic of the Congo. Emerg Infect Dis **2003**; 9:1531–7.
7. Towner J, Khristova M, Sealy T, et al. Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola. J Virol **2006**; 80:6497–516.
8. Daddario-DiCaprio K, Geisbert T, Geisbert J, et al. Cross-protection against Marburg virus strains by using a live, attenuated recombinant vaccine. J Virol **2006**; 80:9659–6.
9. Geisbert T, Daddario-DiCaprio K, Geisbert J, et al. Marburg virus Angola infection of rhesus macaques: pathogenesis and treatment with recombinant nematode anticoagulant protein c2. J Infect Dis **2007**; 196(Suppl 2):S372–81.
10. Centers for Disease Control and Prevention. Update: filovirus infection in animal handlers. MMWR Morb Mortal Wkly Rep **1990**; 39:221.
11. Fisher-Hoch S, McCormick J. Experimental filovirus infections. Curr Top Microbiol Immunol **1999**; 235:117–43.
12. Dube D, Brecher M, Delos S, et al. The primed ebolavirus glycoprotein (19-kilodalton GP1,2): sequence and residues critical for host cell binding. J Virol **2009**; 83:2883–91.
13. Kuhn J, Radoshitzky S, Guth A, et al. Conserved receptor-binding domains of Lake Victoria marburgvirus and Zaire ebolavirus bind a common receptor. J Biol Chem **2006**; 281:15951–8.
14. Ito H, Watanabe S, Sanchez A, Whitt M, Kawaoka Y. Mutational analysis of the putative fusion domain of Ebola virus glycoprotein. J Virol **1999**; 73:8907–12.
15. Manicassamy B, Wang J, Rumschlag E, et al. Characterization of Marburg virus glycoprotein in viral entry. Virology **2007**; 358:79–88.
16. Yang Z, Duckers H, Sullivan N, Sanchez A, Nabel E, Nabel G. Identification of the Ebola virus glycoprotein as the main viral determinant of vascular cell cytotoxicity and injury. Nat Med **2000**; 6:886–9.
17. Sanchez A, Trappier S, Mahy B, Peters C, Nichol S. The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. Proc Natl Acad Sci U S A **1996**; 93:3602–7.
18. Sanchez A, Trappier S, Stroher U, Nichol S, Bowen M, Feldmann H. Variation in the glycoprotein and VP35 genes of Marburg virus strains. Virology **1998**; 240:138–46.
19. Takada A, Feldmann H, Ksiazek TG, Kawaoka Y. Antibody-dependent enhancement of Ebola virus infection. J Virol **2003**; 77:7539–44.
20. Takada A, Ebihara H, Feldmann H, Geisbert T, Kawaoka Y. Epitopes required for antibody-dependent enhancement of Ebola virus infection. J Infect Dis **2007**; 196(Suppl 2):S347–56.
21. Takada A, Kawaoka Y. Antibody-dependent enhancement of viral infection: molecular mechanisms and in vivo implications. Rev Med Virol **2003**; 13:387–98.
22. Takada A, Watanabe S, Okazaki K, Kida H, Kawaoka Y. Infectivity-enhancing antibodies to Ebola virus glycoprotein. J Virol **2001**; 75:2324–30.
23. Takada A, Robison C, Goto H, et al. A system for functional analysis of Ebola virus glycoprotein. Proc Natl Acad Sci U S A **1997**; 94: 14764–9.
24. Matsuno K, Kishida N, Usami K, et al. Different potential of C-type lectin-mediated entry between Marburg virus strains. J Virol **2010**; 84:5140–7.
25. Noda T, Sagara H, Suzuki E, Takada A, Kida H, Kawaoka Y. Ebola virus VP40 drives the formation of virus-like filamentous particles along with GP. J Infect Dis **2002**; 76:4855–65.
26. Swenson D, Warfield K, Kuehl K, et al. Generation of Marburg virus-like particles by co-expression of glycoprotein and matrix protein. FEMS Immunol Med Microbiol **2004**; 40:27–31.
27. Shahhosseini S, Das D, Qiu X, Feldmann H, Jones SM, Suresh MR. Production and characterization of monoclonal antibodies against different epitopes of Ebola virus antigens. J Virol Methods **2007**; 143:29–37.
28. Nakayama E, Yokoyama A, Miyamoto H, et al. Enzyme-linked immunosorbent assay for the detection of filovirus species-specific antibodies. Clin Vaccine Immunol **2010**; 17:1723–8.
29. Bausch DG, Nichol ST, Muyembe-Tamfum JJ, et al. Marburg hemorrhagic fever associated with multiple genetic lineages of virus. N Engl J Med **2006**; 355:909–19.
30. Daddario-DiCaprio KM, Geisbert TW, Ströher U, et al. Postexposure protection against Marburg haemorrhagic fever with recombinant vesicular stomatitis virus vectors in non-human primates: an efficacy assessment. Lancet **2006**; 367:1399–404.

OPEN ACCESS Freely available online

PLoS one

Gnarled-Trunk Evolutionary Model of Influenza A Virus Hemagglutinin

Kimihito Ito[1,2], Manabu Igarashi[1], Yutaka Miyazaki[3], Teiji Murakami[1], Syaka Iida[1], Hiroshi Kida[1,4,5,6], Ayato Takada[1,7]*

1 Hokkaido University Research Center for Zoonosis Control, Sapporo, Japan, **2** PRESTO, Japan Science and Technology Agency, Saitama, Japan, **3** Faculty of Liberal Arts and Sciences, Osaka University of Economics and Law, Yao, Japan, **4** Department of Disease Control, Graduate School of Veterinary Medicine, Hokkaido University, Sapporo, Japan, **5** OIE Reference Laboratory for Highly Pathogenic Avian Influenza, Sapporo, Japan, **6** SORST, Japan Science and Technology Agency, Saitama, Japan, **7** School of Veterinary Medicine, The University of Zambia, Lusaka, Zambia

Abstract

Human influenza A viruses undergo antigenic changes with gradual accumulation of amino acid substitutions on the hemagglutinin (HA) molecule. A strong antigenic mismatch between vaccine and epidemic strains often requires the replacement of influenza vaccines worldwide. To establish a practical model enabling us to predict the future direction of the influenza virus evolution, relative distances of amino acid sequences among past epidemic strains were analyzed by multidimensional scaling (MDS). We found that human influenza viruses have evolved along a gnarled evolutionary pathway with an approximately constant curvature in the MDS-constructed 3D space. The gnarled pathway indicated that evolution on the trunk favored multiple substitutions at the same amino acid positions on HA. The constant curvature was reasonably explained by assuming that the rate of amino acid substitutions varied from one position to another according to a gamma distribution. Furthermore, we utilized the estimated parameters of the gamma distribution to predict the amino acid substitutions on HA in subsequent years. Retrospective prediction tests for 12 years from 1997 to 2009 showed that 70% of actual amino acid substitutions were correctly predicted, and that 45% of predicted amino acid substitutions have been actually observed. Although it remains unsolved how to predict the exact timing of antigenic changes, the present results suggest that our model may have the potential to recognize emerging epidemic strains.

Citation: Ito K, Igarashi M, Miyazaki Y, Murakami T, Iida S, et al. (2011) Gnarled-Trunk Evolutionary Model of Influenza A Virus Hemagglutinin. PLoS ONE 6(10): e25953. doi:10.1371/journal.pone.0025953

Editor: Art F. Y. Poon, British Columbia Centre for Excellence in HIV/AIDS, Canada

Received June 7, 2011; **Accepted** September 13, 2011; **Published** October 10, 2011

Funding: This project was supported by the Program of Founding Research Centers for Emerging and Reemerging Infectious Diseases, the Japan Initiative for Global Research Network on Infectious Diseases (J-GRID), the Global COE Program, Grants-in-Aid, all from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, and PRESTO and SORST from Japan Science and Technology Agency (JST) Basic Research Programs. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: atakada@czc.hokudai.ac.jp

Introduction

The hemagglutinin (HA) molecule of influenza A viruses is the prime target of antibodies that neutralize viral infectivity. The strong immune pressure against HA in the human population selects a new variant every 2–5 years [1–6]. Thus influenza A viruses undergo antigenic changes with gradual accumulation of amino acid substitutions on HA, and the antigenic change is one of the primary reasons why vaccination is not a perfect measure to control seasonal influenza. Accordingly, influenza vaccine often requires replacement to avoid antigenic mismatch between vaccine and epidemic strains [7]. The decision of vaccine replacement must be made several months before a minor strain become dominant strain [8]. Thus the prediction of antigenic change of influenza A virus [2,9–15] has been one of the major public health goals.

Phylogenetic analyses of HA genes of human H3N2 viruses have revealed the presence of a long main trunk and short side branches in their evolutionary tree. The main trunk has grown continuously from a pandemic strain in 1968 to recent epidemic strains, and tips of each branch reached a dead end on the evolutionary pathway [13,14,16–18]. This 'cactus-like' phylogenetic tree indicates that the viruses on the side branches do not produce next epidemic strains, while the viruses near the main trunk do contribute to the production of both an epidemic strain and its next epidemic strain. Although two or more antigenically different strains were known to co-circulate in a single epidemic season [13,19,20], the single-trunk phylogenetic tree indicates the diversity of the HA amino acid sequences at any point in time is relatively limited. The reason why only one trunk exists has yet to be fully understood, but several theories have been proposed to explain this phenomenon [21–23].

The aims of the present studies are to establish a practical model enabling us to predict the evolutionary direction of the virus that causes future epidemics and to examine the accuracy of the prediction based on the model. First we analysed relative distances of amino acid sequences among past epidemic strains using a method called multidimensional scaling (MDS) [24]. We found that human influenza viruses have evolved along a gnarled evolutionary pathway with an approximately constant curvature in the MDS-constructed 3D space. The constant curvature was reasonably explained by assuming that the rate of amino acid substitutions varied from one position to another according to a gamma distribution. The estimated parameters of the gamma

distribution allowed us to predict the amino acid substitutions on HA in subsequent years with reasonable accuracy, indicating the potential to select suitable vaccine strains for the subsequent epidemic seasons.

Results

To expose underlying patterns of HA amino acid substitutions in the evolutionary pathway along the main trunk, we conducted multidimensional scaling (MDS) analysis [24] of HA sequences. The fundamental idea for visualizing a large number of sequences in a low dimensional space is based on the same idea described in a recent paper by He and Deem [15]. By performing MDS analysis, one may obtain a visual map of objects where the dissimilarity between objects is represented as the distance between corresponding points. A total of 2,640 unique amino acid sequences of the HA1 [3,25] domain (328 amino acids long) of the H3N2 viruses isolated from humans during the period from 1968 through 2009 were analysed by MDS and visualized in a three-dimensional (3D) space (Figure 1, Movie S1). In the resulting 3D map, each HA sequence was represented as a point, and the number of different amino acids between two HA sequences was represented as the relative distance between two corresponding points. Although the original amino acid sequences provided 328 dimensional data, the numbers of different amino acids among sequences were reasonably approximated by distance in this 3-dimensional map with a root-mean-square error of 1.72 (Figure S1). Consistent with phylogenetic analyses, viruses that were isolated close in time were located near each other, forming a thick main trunk with short branches elongated from the trunk (Figure 1A, Figure 1B). The main trunk grew continuously from a pandemic strain in 1968 to recent epidemic strains. Each branch consisted of epidemic strains isolated during a period of 3–5 years.

It should be noted that the MDS representation revealed a characteristic feature that has not been clear only from phylogenetic trees, the observation of a gnarled trunk constantly curved in the 3D map. Since amino acid substitutions on human virus HA mainly occur in the HA1 domain [3], this result indicated that the evolution of H3N2 virus HA was characterized by this gnarled evolutionary pathway.

We also conducted the same analysis for H1N1 human viruses (Figure S2, Movie S2). The 3D representation of HA sequences of H1N1 seasonal influenza viruses showed the same pattern in their evolutionary pathways, a long gnarled trunk elongated from the pandemic strain in 1918. Another H1N1 pandemic strain, which was introduced into the human population in 2009 [26], was located at the end of a step-wise path that consisted of swine H1N1 influenza viruses isolated from humans from 1976 to 2007 [27].

In the 3D map in Figure 1, the spatial distance between each pair of sequences represents the number of different amino acids between these sequences. Figure 2 illustrates two distinct patterns of amino acid substitutions that produce different spatial arrangements of viruses in this map. If a series of amino acid substitutions all occur in different positions, then the distance from an ancestor to a mutant should be proportional to the number of substitutions. These independent substitutions make a straight arrangement of viruses on the map (Figure 2A, 2B). On the other hand, if amino acids at particular positions were substituted more than once, the distance from an ancestor to a mutant should be less than the number of substitutions. These concentrated substitutions at the same amino acid positions make a curved arrangement of viruses (Figure 2C, 2D). For this reason, the gnarled trunk found in the MDS representation of HA sequences (Figure 1) indicated that HA variants on the trunk favoured multiple amino acid substitutions at the same positions. The fact that most of the amino acid substitutions occurred near antigenic domains A–E [3,5,9,28] was consistent with the observation of the curved trunk in the 3D map.

To investigate the property of the curvature of the main trunk, we analysed the distribution of the number of positions that were substituted on the trunk from 1968 to 2009 (Figure 3A, Table S1). Of the 328 positions on the HA1 sequence, 260 remained unchanged for 41 years. At 36 amino acid positions, residues were substituted once, and at 19 positions twice. The number of positions gradually decreased as their observed frequency of substitutions increased, but there was one position that has substituted eight times. The mean of the substitution frequencies was 0.384, and the variance of the substitution frequencies was 0.904. Given these statistics for the substitution frequency, the shape of the histogram is almost identical to the curve of a gamma distribution [29] having the same mean and variance (Figure 3A). From this result, it is likely that the rate of amino acid substitutions varies from one position to another according to a gamma distribution.

To estimate the parameters of the gamma distribution precisely, non-linear regression analysis was performed. First, 91 HA sequences near the main trunk were selected as trunk sequences. In Figure 3B, the number of different amino acids between two trunk sequences (Figure 3C) is plotted against the difference in their isolation years. It is known that if the amino acid substitution rate varies according to a gamma distribution, the expected number of different amino acids between two sequences can be calculated by the formula: $d = L(1-(a/(a+\bar{r}t))^a)$, where $L$is the length of the sequences, a is the shape parameter of the gamma distribution (gamma parameter), $\bar{r}$ is the mean substitution rate, and t is the difference in the years of the two sequences [30]. By fitting the above formula to the actual numbers of different amino acids on the trunk, the gamma parameter and mean substitution rate were estimated ($a = 0.129$ and $\bar{r} = 0.0118$), showing a good fit to the actual data ($P<0.001$). This result indicated that the number of different amino acids between two sequences could be determined from the difference of the year of isolation, according to the gamma-distribution-based model presented above. Thus, it is reasonable to conclude that the constant curvature on the trunk in the MDS representation (Figure 1) was attributed to this nature.

Next, an attempt was made to apply the gamma-distribution-based substitution model of the trunk to the prediction of amino acid substitutions in subsequent years. The key idea of our prediction method is to select the direct progenitor virus for future epidemics from the surveillance samples of each year. We designate a virus strain that is located near the trunk extending into the next year as a Leading Bud. A Leading Bud can be considered as an potential dominant strain that is not dominant in the given year and become dominant the next year [15]. Using the formula $d = 328 \times (1-(a/(a+\bar{r}t))^a)$under the estimated $a$and $\bar{r}$, one may calculate the expected number of different amino acids between two HA sequences located on the trunk. According to this gamma-distribution-based formula, a virus that will appear on the trunk in a particular year is expected to have 4, 7, 10, 13, ... , 60 different amino acids in HA, when compared with viruses in 1, 2, 3, 4, ... , 42 years before, respectively. Therefore, given the large variety of viruses isolated in a year, the virus that is likely to be located near the extended trunk is the virus to which HA sequence dissimilarities from past viruses have the highest fit to those expected under the gamma-distribution-based formula (Figure 3D).

To examine whether the method correctly selected the Leading Bud in the subsequent year, we conducted retrospective tests for each year from 1997 to 2009. Evaluation was made by comparing

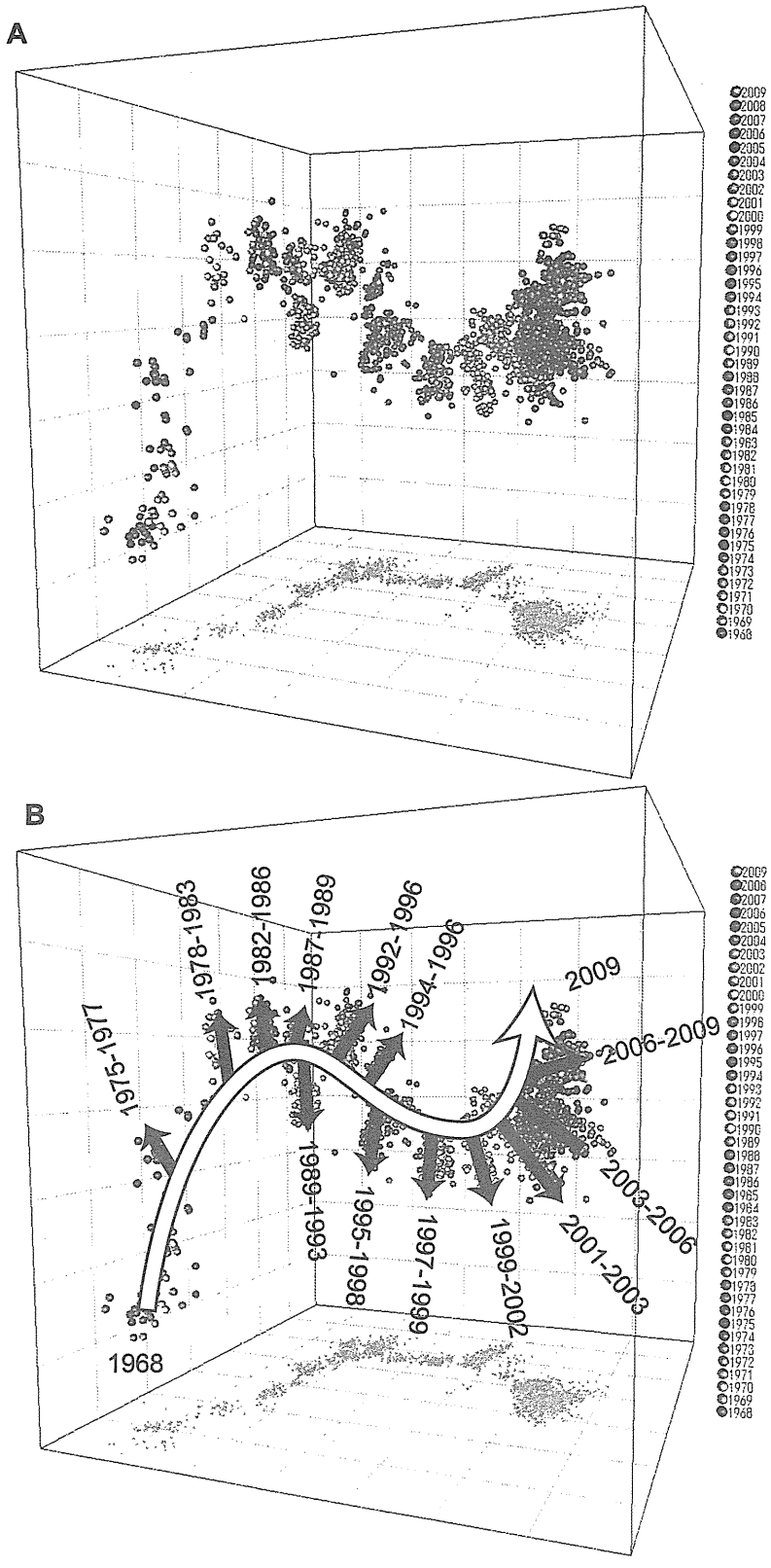

A
2009
2008
2007
2006
2005
2004
2003
2002
2001
2000
1999
1998
1997
1996
1995
1994
1993
1992
1991
1990
1989
1988
1987
1986
1985
1984
1983
1982
1981
1980
1979
1978
1977
1976
1975
1974
1973
1972
1971
1970
1969
1968
B
1968
1975-1977
1978-1983
1982-1986
1987-1989
1988-1993
1992-1996
1994-1996
1995-1998
1997-1999
1999-2002
2001-2003
2003-2006
2006-2009
2009

Figure 1. Three-dimensional map of HA sequences of H3N2 human influenza A viruses. A total of 2,640 amino acid sequences of the HA1 domain of human H3N2 influenza A viruses isolated during the period from 1968 through 2009 are visualized in a 3D space. Each point represents a virus strain. The distance between two viruses in the 3D map represents the number of different amino acids between their HA sequences. The whole coordination is determined by MDS analysis. The root-mean-square error of the 3D map was 1.72. All three axes represent the sequence dissimilarity (spacing between grid lines represents 10 different amino acids), and the configuration can be freely rotated and translated. Shadows represent projections of points onto the coordinate planes. (A) the 3D map colour-coded by the year of isolation of the virus. (B) a schematic diagram of the 3D map.
doi:10.1371/journal.pone.0025953.g001

the predicted Leading Bud with the dominant sequence of the subsequent year, which consisted of the amino acids that constituted the majority at each position in the year. Table 1 shows the results of retrospective tests (for details, see Table S2). The recall, which is the probability that an actual substitution was correctly predicted, was 1.00 in 4 of the 12 calendar years. The overall recall of the prediction was 0.70, indicating that the model had a reasonable ability to predict amino acid substitution in the subsequent year for each year. The precision, which is the probability that a predicted substitution actually occurred, varied from 0.0 to 0.89, and the overall recall of the prediction was 0.45.

To assess the validity of the result of the retrospective tests, we repeated similar retrospective tests with other methods and compared the results (Table 2). First we tested a method that randomly selects an HA sequence for each year. With 100 sets of tests, overall precision and recall were 0.24 ± 0.015 and 0.22 ± 0.012 respectively, showing low predictive ability as expected. Secondly we tested method that selects the HA sequence that has the maximum numbers of substitutions at the 18 positively selected codons identified by Bush et al [10]. Although the overall precision and recall were much higher than random tests, the accuracy of prediction was lower than that of our method. Three methods that select the sequence that has the maximum number of amino acid substitutions from current and past dominant sequences yielded higher recalls. However, the overall precision was lower than the method using 18 positively selected codons. Among all the methods we tested, the gamma-distribution-based method was the only method that yields higher recall and higher precision than Bush's methods.

Discussion

Our study found that the long-term evolution of HA was reasonably characterised by a 'bonsai-like' pathway of which trunk

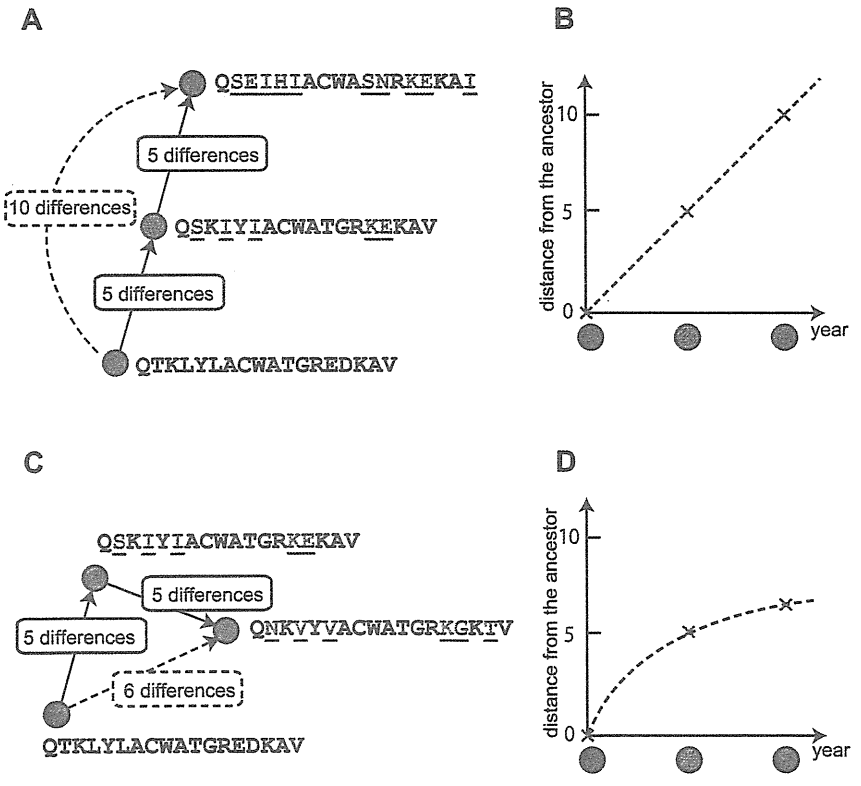


Figure 2. Two distinct patterns of amino acid substitutions that produce different spatial arrangements of viruses. (A) The straight arrangement of viruses. If a series of amino acid substitutions all occur in different positions, then the distance from an ancestor to a mutant should be proportional to the number of substitutions (B). These independent substitutions make a straight arrangement of viruses on the map. (C) The curved arrangement of viruses. If amino acids at particular positions were substituted more than once, the distance from an ancestor to a mutant should be less than the number of substitutions (D). These concentrated substitutions at the same amino acid positions make a curved arrangement of viruses. In both panels, viruses are represented by circles, with illustrative examples of their amino acid sequences and substitutions on them.
doi:10.1371/journal.pone.0025953.g002

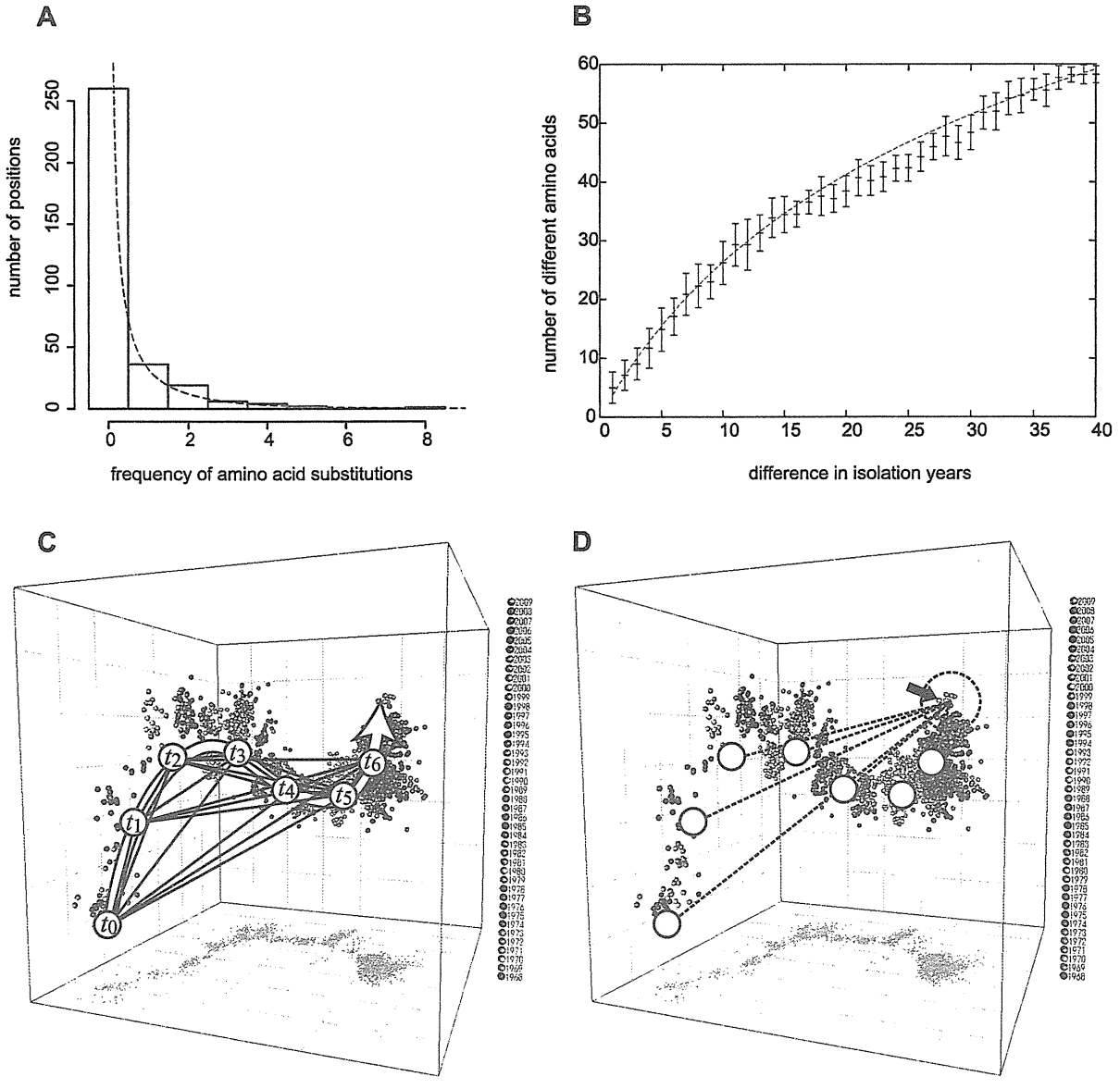


Figure 3. The amino acid substitutions on the gnarled trunk. (A) the distribution of the number of positions that were substituted on the trunk from 1968 to 2009. The gamma distribution that has a mean substitution frequency of 0.418 and a variance of 1.155 is superimposed. (B) The number of different amino acids between the trunk sequences plotted against the difference in their isolation years. The horizontal bars indicate mean values, and vertical lines indicate ±1 standard deviation of the number of different amino acids. A non-linear regression curve using the formula $d = 328 \times (1 - (a/(a+\bar{r}t))^a)$ is superimposed ($a = 0.129$, $\bar{r} = 0.0118$), showing good fit to the actual data ($P<0.001$). (C) A schematic illustration of trunk viruses and the sequence dissimilarities among them. A trunk virus is a virus located near the main trunk. Circles labelled with $t_0, \ldots, t_6$ are examples of trunk viruses. The sequence dissimilarities among trunk sequences are shown by solid lines. (D) The selection of a Leading Bud. Using the formula of $d = 328 \times (1 - (a/(a+\bar{r}t))^a)$, one may calculate the expected sequence dissimilarities between a future trunk virus and past trunk viruses (dotted lines). The bold arrow indicates a candidate for Leading Buds. For the coming influenza season in 2010, A/Thailand/CU-B110/2009(H3N2) was predicted to be the most likely candidate virus (Table S2).
doi:10.1371/journal.pone.0025953.g003

was constantly curved in the MDS-constructed 3D space. This unique property of the sequence evolution indicated that the evolution on the trunk favoured multiple substitutions at the same positions on HA molecules. Our study found that the curvature was relatively constant and reasonably explained by assuming that the rate of amino acid substitutions on HA varied from one position to another according to a gamma distribution. The estimated parameters of the gamma distribution allowed us to predict the amino acid substitutions on HA in subsequent years with reasonable accuracy.

The small value of its gamma parameter estimated in this study suggested that most of amino acids on HA remained unchanged,

Table 1. Results of retrospective tests for the prediction of amino acid substitutions.

Test Year	The number of predicted substitutions	The number of actual substitutions	The number of correctly predicted substitutions	Recall	Precision
1998	9	8	8	1.00	0.89
1999	4	3	1	0.33	0.25
2000	4	5	2	0.40	0.50
2001	10	5	5	1.00	0.50
2002	7	3	1	0.33	0.14
2003	13	12	11	0.92	0.85
2004	4	5	3	0.60	0.75
2005	2	0	0	-	0.00
2006	4	2	2	1.00	0.50
2007	5	2	0	0.00	0.00
2008	4	1	0	0.00	0.00
2009	7	1	0	0.00	0.00
overall	73	47	33	0.70	0.45

Recall was defined as the number of correctly predicted substitutions divided by the total number of actual substitutions. Precision was defined as the number of correctly predicted substitutions divided by the total number of predicted amino acid substitutions.

doi:10.1371/journal.pone.0025953.t001

but amino acid substitutions occurred at a relatively restricted number of positions on the HA. The result was consistent with previous studies identifying several positions that had undergone the positive Darwinian selection, where non-synonymous mutations have been favoured [16,17,28,31,32]. Although the positions that undergo amino acid substitutions could have been moving to different positions over time [33], our analysis indicated that the relative sequence distance between two trunk sequences remained roughly constant with respect to the difference in their isolation years. This stable feature allowed us to predict the relative sequence distance between two viruses located on the trunk, and led us a fully-computerized prediction method.

He and Deem have recently pointed out that an MDS visualization with density estimation allowed us to identify a cluster of 'incipient dominant strains' before it became dominant [15]. They proposed two important criteria for the selection of a new vaccine strain. The first criterion is that a new cluster that does not contain currently circulating strains or vaccine strains is detected. The second criterion is that the current vaccine strain does not provide high protection against strains in the new cluster. The reason why our gamma-distribution-based model achieves high recall and precision can be explained by their two criteria. First of all, the recognition of a Leading Bud conceptually corresponds to the detection of a newly emerging cluster of incipient dominant strains. Since a Leading Bud described in this paper should have a certain amount of different amino acids from dominantly circulating strains, a Leading Bud can be considered as one of the early isolates in a newly emerging cluster. Although we do not consider whether Leading Buds form a cluster or not, this property partially fits their first criterion. Secondly, our gamma distribution-based method recognizes a Leading Bud by finding an HA sequence that has the amino acid substitutions at the same positions as those seen in the past evolution. Since most of past amino acid substitutions are concentrated in the antigenic sites (Table S1), it is highly likely that the Leading Bud having amino acid substitutions at these positions antigenically differ from the dominantly circulating strain and the vaccine strain. This could meet the second criterion.

The overall recall of our prediction method was around 0.70, indicating that the model had a reasonable ability to predict amino acid substitution in the subsequent year for each year. It should be noted that this high recall was achieved by the prediction method that relied only upon the number of different positions in the amino acid sequences and the isolation year of the viruses. The overall precision, on the other hand, was around 0.45. Some of mistaken predictions might be attributable to the delayed

Table 2. Comparison of overall recall and precision with other methods.

Method	Overall Recall	Overall Precision
Select a sequence randomly (n = 100)	0.22±0.012	0.24±0.015
Select the one that has the maximum numbers of substitutions at the 18 codons identified by Bush et al	0.49	0.34
Select the one that has the maximum numbers of substitutions from the current dominant sequence	0.51	0.13
Select the one that has the maximum numbers of substitutions from past dominant sequences	0.55	0.25
Select the one that has the maximum numbers of substitutions from the dominant sequence of two years ago	0.70	0.23
Select the one that has the maximum numbers of substitutions at antigenic domains A–E	0.70	0.32
Select the one that has the minimum errors from the gamma-distribution-based expectation	0.70	0.45

doi:10.1371/journal.pone.0025953.t002

appearance of amino acid substitutions in the dominant sequences on the trunk. For instance, the predicted substitutions for I144N in 1999, R50G in 2000 and 2001, E83K V202I, W222R G225D in 2001, and V226I and S227P in 2002 did not occur in the next years, but rather 2 or 3 years later (Table S2). For these idle periods, the prediction method could have looked too far ahead, and the Leading Buds might be too early to be used as vaccine strains.

In the retrospective tests, we evaluated the prediction methods by comparing the amino acid sequences of the Leading Buds with the dominant sequences in subsequent years. It was also confirmed that WHO-recommended vaccine strains had amino acid residues that were identical to those predicted by our method (Table S2). The major difference lies in the timing. We accept that an overhasty selection of vaccine strains might lead antigenic mismatch between vaccine and epidemic strains. The prediction of the exact timing of the antigenic change could become a subject of future study. In addition to the timing of antigenic change, a careful investigation on their characteristics such as antigenicity and growth in embryonated hen eggs must be practically important for the vaccine selection.

Vaccine strains must be selected in order to match the antigenicity of viruses that will circulate in the influenza season. The antigenic cartography, which was developed by Smith et.al, enables us to accurately predict antigenic similarity between two virus strains based on a large collection of hemagglutination inhibition (HI) assay data [5]. In the period from 2005 to 2006, for instance, the circulating H3N2 viruses changed from A/California/7/2004-like viruses to A/Wisconsin/67/2005-like viruses [15,34]. The Leading Bud found in 2005 was A/Okinawa/18/2005. The antigenic-cartography-based antigenic distance between A/Okinawa/18/2005 and A/Wisconsin/67/2005 was found to correspond to a twofold difference in HI titers of antisera [34]. Therefore, it is likely that the antigenicity of A/Okinawa/18/2005 could match the epidemic strain in 2006. Although the Leading Buds may not be perfect candidates for the vaccine strains, we believe that our prediction method could provide useful information for the formulation of influenza vaccines.

Retrospective tests for 2006–2007, 2007–2008, and 2008–2009 failed to predict the actual amino acid substitutions. The low recall and precision are likely due to the limited number of amino acid substitutions during these periods. Since antigenic changes of H3N2 viruses occur every 3–5 years in a punctuated manner [5], the conservation of dominant sequences in a few contiguous years is common in the evolution of H3N2 viruses. We have not taken such periodicity into account in our evolutionary model, and our method could not predict the exact timing when a dominant strain is replaced by another strain. This result highlighted the need to develop a method that can predict the exact timing of the antigenic change of the virus. However, the HA of the dominant H3N2 virus after the September in 2009 possessed 3 of 6 amino acid substitutions that were predicted by our method using the sequence data before August 2009 (Table S2). A/Perth/16/2009, a similar strain to our Leading Bud (A/Thailand/CU-B110/2009), was recommended as a vaccine strain for H3N2 viruses by WHO on Dec 1st in 2010.

Further understanding of the gnarled trunk might be achieved by combined efforts with experimental studies. Future research direction include the association of the gnarled trunk evolution with the prediction of antigenic evolution [5,35,36], the effect of mutations upon biological activity of the protein [37,38], and the effect of cross immunity to previously circulating dominant strains [21,22]. Finally, the 3D visualization technique we present here enables us to represent the direction of sequence evolution as well as sequence phylogeny, providing additional information that is not obtained via traditional phylogenetic analysis.

Materials and Methods

Sequence Data

Nucleotide sequences for HA genes of H3N2 influenza A viruses isolated from humans during the period from 1968 to 2009 were downloaded from the Influenza Virus Resource at the National Center for Biotechnology Information (NCBI) [39] on Feb 23 in 2010. The isolation date of the latest sequence was July 11th 2009. After eliminating sequences that contained ambiguous nucleotide codes, 6,806 amino acid sequences of the HA1 domain were determined by translating the nucleotide sequences using the standard genetic code. All the amino acid sequences were 328 amino acids long. Nucleotide sequences that gave an identical amino acid sequence were grouped together, and the nucleotide sequence having the fewest mutations from the pandemic strain in 1968 was used as a representative. By removing all identical amino acid sequences except one, 2,640 unique amino acid sequences were obtained.

MDS analysis

For every pair of the 2,640 amino acid sequences of the HA1 domain, the sequence dissimilarity, which is the total number of positions where the two sequences possess different amino acids, was calculated. The resulting 3,483,480 pair-wise dissimilarities were stored in a dissimilarity matrix. The SMACOF algorithm [24] was used to find the optimal coordination of all sequences in the 3D map to minimise the sum of squared errors:

$$\sum_{s_1,s_2 \in HA1} (d(s_1,s_2) - d_M(s_1,s_2))^2,$$

where $s_1$and $s_2$are sequences of the HA1 domain, $d(s_1,s_2)$ is the sequence dissimilarity, $d_M(s_1,s_2)$is the Euclidean distance in the 3D map.

Amino Acid Substitution on the Trunk

A parsimony tree of HA was constructed from a total of 2,640 nucleotide sequences of the HA1 domain. The dnapars program in the PHYLIP package[24] was used to construct the parsimony tree. The main trunk of the tree was defined as the longest path from the HA of the pandemic strain in 1968 to the HA of a strain circulating in 2009. The hypothetical nucleotide sequence on each trunk was translated into an amino acid sequence. For each residue position of the HA1 domain, amino acid substitutions found on the trunk were counted, and then the mean and variance of the substitution frequency were calculated and compared with a gamma distribution having the same mean and variance.

Substitution Model

When assuming the variation of substitution rates follows a gamma distribution, the expected sequence distance between two sequences can be calculated by the formula $d = L(1-(a/(a+\bar{r}t))^a)$, where t is the difference in their isolation years and $L = 328$ is the length of the HA1 domain. To obtain non-hypothetical amino acid sequences located near the trunk, a neighbour-joining tree was constructed from their nucleotide sequences. The tree construction was done using the dnadist and neighbor programs in the PHYLIP package [40] with the Jukes-Cantor distance option. For each trunk node of the neighbour

joining tree, the amino acid sequence having shortest path to the trunk node was selected. Out of 2,640 HA sequences, 91 sequences were selected as trunk sequences (Table S3). For every pair of these selected sequences, their sequence dissimilarity (d) and the difference in isolation years (t) was recorded. By fitting the formula $d=L(1-(a/(a+\bar{r}t))^a)$to the observed relationship between $d$and t, the gamma parameter (a) and mean substitution rate ($\bar{r}$) for our substitution model were estimated. In order to estimate the effect of the selection of trunk sequences on the Figure 3B, we performed a bootstrap resampling analysis. We made 100 datasets each containing 91 trunk sequences obtained by random resampling of the original 91 trunk sequences. Then, errors on the mean number of different amino acids were estimated using the 100 bootstrap datasets. As shown in the Figure S3, the errors of the means were estimated to be around one amino acid, suggesting that the result shown in Figure 3 has moderate robustness to the selection of trunk sequences.

Dominant sequence

To define a representative sequence for each year, we adopted a strategy using the majority vote rule. For each year and each residue position, the dominant amino acid was determined as the amino acid that constituted the majority at the position in the year. The dominant sequence of a year was defined as the concatenation of the dominant amino acids of every position in the year.

Prediction and retrospective tests

For each year, a Leading Bud, which is an amino acid sequence that will be located near the trunk in the next year, was predicted as follows. Let y be a year. For a future HA sequence $s$that would appear in the year $y+1$, the expected number of different amino acids from each past trunk sequence $u$in year y' $(y-y'\geq 0)$can be denoted by the formula $d_E(s,u)=L(1-(a/(a+\bar{r}(y-y'+1)))^a)$. Thus, among sequences isolated in the year y, the sequence that is most likely to become a Leading Bud in the next year is formulated as the sequence s', such that s' has the least sum of squared errors between sequence distance and expected distance from each trunk. The sum of squared errors is calculated by the formula:

$$\sum_{u\in T_{\leq y}} \frac{(d(s',u)-d_E(s',u))^2}{d_E(s',u)},$$

where $T_{\leq y}$ is a set of trunk sequences whose isolation years are earlier than or equal to$y$. The amino acid sequence of the predicted Leading Bud was compared with the dominant sequence for the year, and a set of amino acid substitutions from the dominant sequence to the sequence of the Leading Bud was presented as predicted substitutions. After the prediction was made, the predicted substitutions were compared by the actual amino acid substitutions that occurred in the next year. Recall was calculated as the number of correctly predicted substitutions divided by the total number of actual substitutions. Precision was calculated as the number of correctly predicted substitutions divided by the total number of predicted amino acid substitutions.

Supporting Information

Figure S1 Scatter diagram of numbers of different amino acids vs. corresponding distances in the 3D map. For every pair of two sequences, the actual numbers of different amino acids (X-axis) were plotted against corresponding distances in the 3D map (Y-axis). Horizontal bars show the mean values, and vertical lines indicate ±1SD of distances in the 3D map.
(EPS)

Figure S2 Three-dimensional map of the human H1N1 influenza A viruses. A total of 1228 amino acid sequences of the HA1 domain of human H1N1 influenza A viruses isolated during 1918 to 2010 are visualized in the 3D space. Each point represents an HA sequence, colour-coded by the isolation year of the virus. The whole coordination is determined by MDS analysis. All three axes represent sequence dissimilarity (spacing between grid lines represents 10 different amino acids), and the configuration can be freely rotated and translated. Shadows represent projections of points on a coordinate plane. Bold arrows on the left, middle, and right indicate a seasonal H1N1 virus isolated in 2009, the pandemic H1N1 virus in 1918, and the pandemic H1N1 virus in 2009, respectively.
(EPS)

Figure S3 Boot strap resampling analysis of the mean substitution frequency on the trunk. A total of 100 bootstrap datasets obtained were generated by random resampling of the original 91 trunk sequences. Errors on the mean number of different amino acids were estimated to be around one amino acid. A curve using the formula $d=328\times(1-(a/(a+\bar{r}t))^a)$ is superimposed ($a=0.129$, $\bar{r}=0.0118$).
(EPS)

Table S1 The amino acid positions that were substituted one or more time. The positions on HA where amino amino acid substitution were occurred on the trunk are shown with their frequency. Each alphabet represents the antigenic domain to which the position belongs.
(DOC)

Table S2 Selected Leading Buds, predicted substitutions, and actual substitutions in the retrospective tests. Correctly predicted substitutions are shown in bold-face. The predicted substitutions that did not occur in the next years but occurred 2 or 3 years later are underlined. Different amino acids on HA between a new WHO vaccine strain and preceding vaccine strain are shown in the rightmost column.
(DOC)

Table S3 The HA sequences located near the main trunk.
(DOC)

Movie S1 Movie of 3D map of HA sequences of H3N2 human influenza A viruses.
(MP4)

Movie S2 Movie of 3D map of HA sequences of H1N1 human influenza A viruses.
(MP4)

Acknowledgments

We thank Yuzuru Tanaka, Hiroki Arimura, Chihiro Sugimoto, and Yasumasa Nishiura for discussions and comments.

Author Contributions

Conceived and designed the experiments: AT HK KI. Analyzed the data: KI MI YM. Contributed reagents/materials/analysis tools: SI TM. Wrote the paper: KI AT.

References

1. Laver WG, Air GM, Webster RG (1981) Mechanism of antigenic drift in influenza virus : Amino acid sequence changes in an antigenically active region of Hong-Kong (H3N2) influenza virus hemagglutinin. J Mol Biol 145: 339–361.
2. Both GW, Sleigh MJ, Cox NJ, Kendal AP (1983) Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: Multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. J Virol 48: 52–60.
3. Wilson IA, Cox NJ (1990) Structural basis of immune recognition of influenza virus hemagglutinin. Annu Rev Immunol 8: 737–771.
4. Cox NJ, Bender CA (1995) The molecular epidemiology of influenza viruses. Semin Virol 6: 359–370.
5. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, et al. (2004) Mapping the antigenic and genetic evolution of influenza virus. Science 305: 371–376.
6. Wright PF, Neumann G, Kawaoka Y (2007) Orthomyxoviruses. In: Knipe DM, Howley PM, eds. Fields Virology, 5 ed Philadelphia: Lippincott Williams & Wilkins. pp 1691–1740.
7. de Jong JC, Beyer WEP, Palache AM, Rimmelzwaan GF, Osterhaus ADME (2000) Mismatch between the 1997/1998 influenza vaccine and the major epidemic A(H3N2) virus strain as the cause of an inadequate vaccine-induced antibody response to this strain in the elderly. J Med Virol 61: 94–99.
8. Treanor J (2004) Weathering the influenza vaccine crisis. N Engl J Med 351: 2037–2040.
9. Webster RG, Laver WG (1980) Determination of the number of nonoverlapping antigenic areas on Hong Kong (H3N2) influenza virus hemagglutinin with monoclonal antibodies and the selection of variants with potential epidemiological significance. Virology 104: 139–148.
10. Bush RM, Bender CA, Subbarao K, Cox NJ, Fitch WM (1999) Predicting the evolution of human influenza A. Science 286: 1921–1925.
11. Ferguson NM, Anderson RM (2002) Predicting evolutionary change in the influenza A virus. Nat Med 8: 562–563.
12. Plotkin JB, Dushoff J, Levin SA (2002) Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. Proc Natl Acad Sci U S A 99: 6263–6268.
13. Nelson MI, Simonsen L, Viboud C, Miller MA, Taylor J, et al. (2006) Stochastic processes are key determinants of short-term evolution in influenza A virus. PLoS Pathog 2: 1144–1151.
14. Wolf YI, Viboud C, Holmes EC, Koonin EV, Lipman DJ (2006) Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. Biol Direct 1: 34.
15. He J, Deem MW (2010) Low-dimensional clustering detects incipient dominant influenza strain clusters. Protein Eng Des Sel 23: 935–946.
16. Fitch WM, Leiter JME, Li XQ, Palese P (1991) Positive Darwinian evolution in human influenza A viruses. Proc Natl Acad Sci U S A 88: 4270–4274.
17. Fitch WM, Bush RM, Bender CA, Cox NJ (1997) Long term trends in the evolution of H(3) HA1 human influenza type A. Proc Natl Acad Sci U S A 94: 7712–7718.
18. Nelson MI, Holmes EC (2007) The evolution of epidemic influenza. Nat Rev Genet 8: 196–205.
19. Holmes EC, Ghedin E, Miller N, Taylor J, Bao YM, et al. (2005) Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. PLoS Biol 3: 1579–1589.
20. Lavenu A, Leruez-Ville M, Chaix ML, Boelle PY, Rogez S, et al. (2006) Detailed analysis of the genetic evolution of influenza virus during the course of an epidemic. Epidemiol Infect 134: 514–520.
21. Ferguson NM, Galvani AP, Bush RM (2003) Ecological and immunological determinants of influenza evolution. Nature 422: 428–433.
22. Koelle K, Cobey S, Grenfell B, Pascual M (2006) Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. Science 314: 1898–1903.
23. Pybus OG, Rambaut A (2009) Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet 10: 540–550.
24. Borg I, Groenen PJF (2005) Modern multidimensional scaling: Theory and applications. New York: Springer.
25. Palese P, Shaw ML (2007) Orthomyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM, eds. Fields Virology, 5 ed Philadelphia: Lippincott Williams & Wilkins. pp 1647–1689.
26. Neumann G, Noda T, Kawaoka Y (2009) Emergence and pandemic potential of swine-origin H1N1 influenza virus. Nature 459: 931–939.
27. Shinde V, Bridges CB, Uyeki TM, Shu B, Balish A, et al. (2009) Triple-reassortant swine influenza A (H1) in humans in the United States, 2005-2009. New Eng J Med 360: 2616–2625.
28. Bush RM, Fitch WM, Bender CA, Cox NJ (1999) Positive selection on the H3 hemagglutinin gene of human influenza virus A. Mol Biol Evol 16: 1457–1465.
29. Nei M, Chakraborty R, Fuerst PA (1976) Infinite allele model with varying mutation rate. Proc Natl Acad Sci U S A 73: 4164–4168.
30. Ota T, Nei M (1994) Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. J Mol Evol 38: 642–643.
31. Suzuki Y (2004) Three-dimensional window analysis for detecting positive selection at structural regions of proteins. Mol Biol Evol 21: 2352–2359.
32. Kryazhimskiy S, Bazykin GA, Plotkin J, Dushoff J (2008) Directionality in the evolution of influenza A haemagglutinin. Proc R Soc Lond B 275: 2455–2464.
33. Blackburne BP, Hay AJ, Goldstein RA (2008) Changing selective pressure during antigenic changes in human influenza H3. PLoS Pathog 4: e1000058.
34. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, et al. (2008) The global circulation of seasonal influenza A (H3N2) viruses. Science 320: 340–346.
35. Gupta V, Earl DJ, Deem MW (2006) Quantifying influenza vaccine efficacy and antigenic distance. Vaccine 24: 3881–3888.
36. Liao YC, Lee MS, Ko CY, Hsiung CA (2008) Bioinformatics models for predicting antigenic variants of influenza A/H3N2 virus. Bioinformatics 24: 505–512.
37. 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.
38. Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459: 668-U671.
39. Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, et al. (2008) The influenza virus resource at the National Center for Biotechnology Information. J Virol 82: 596–601.
40. Felsenstein J (1989) PHYLIP-phylogeny inference package (version 3.2). Cladistics 5: 164–166.

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Virology

Innate and adaptive immune responses to viral infection and vaccination

Taiki Aoshi[1,2], Shohei Koyama[3], Kouji Kobiyama[1,2], Shizuo Akira[4] and Ken J Ishii[1,2]

Recent accumulating evidence suggests that the human immune system possesses a variety of innate receptors that recognize, distinguish, and respond to viral infections and to vaccination. These include Toll-like receptors, C-type lectin receptors, RIG-I-like receptors, Nod-like receptors and possibly AIM2-like receptors. However, the precise mechanisms by which these receptors exert their critical roles in the induction of virus-specific adaptive immune responses have not been fully elucidated. In this review, we discuss recent advances in our understanding of the innate immune recognition of viruses and the differential connection to the adaptive immune responses induced by infection or vaccination, with a particular focus on the influenza virus.

Addresses
[1] Laboratory of Adjuvant Innovation, National Institute of Biomedical Innovation, Osaka, Japan
[2] Laboratory of Vaccine Science, WPI Immunology Frontier Research Center, Osaka University, Osaka, Japan
[3] Department of Respiratory Medicine, Tohoku University Graduate School of Medicine, Miyagi, Japan
[4] Laboratory of Host Defense, WPI Immunology Frontier Research Center, Osaka University, Osaka, Japan

Corresponding author: Ishii, Ken J (kenishii@biken.osaka-u.ac.jp)

Current Opinion in Virology 2011, 1:226–232

This review comes from a themed issue on Vaccines
Edited by Hildegund Ertl and Bali Pulendran

Available online 30th July 2011

1879-6257/$ – see front matter

DOI 10.1016/j.coviro.2011.07.002

Introduction

Several families of innate immune receptors, including Toll-like receptors (TLRs) [1], C-type lectin receptors [2], RIG-I-like receptors (RLRs) [3], Nod-like receptors (NLRs) [4], and AIM2-like receptors (ALRs) [5] have been identified over the last decade. Generally, these germ-line-encoded receptors recognize 'non-self' molecules derived from a variety of microbes. Some of these receptors also recognize danger signals sent out by damaged cells/tissues [6]. These innate immune receptors are critical for the initiation and regulation of host immune responses against infection and autoimmunity [7]. Furthermore, it is evident that innate immune responses are extremely important for establishing effective adaptive immune responses to infection and vaccination [8•,9,10]; although it is still not clear whether all innate responses contribute equally to the induction of adaptive responses [8•,11•,12•]. In the following sections, we briefly review the current knowledge about virus recognition by innate immune receptors, and discuss the connections between the innate and adaptive immune responses, using influenza virus as an example.

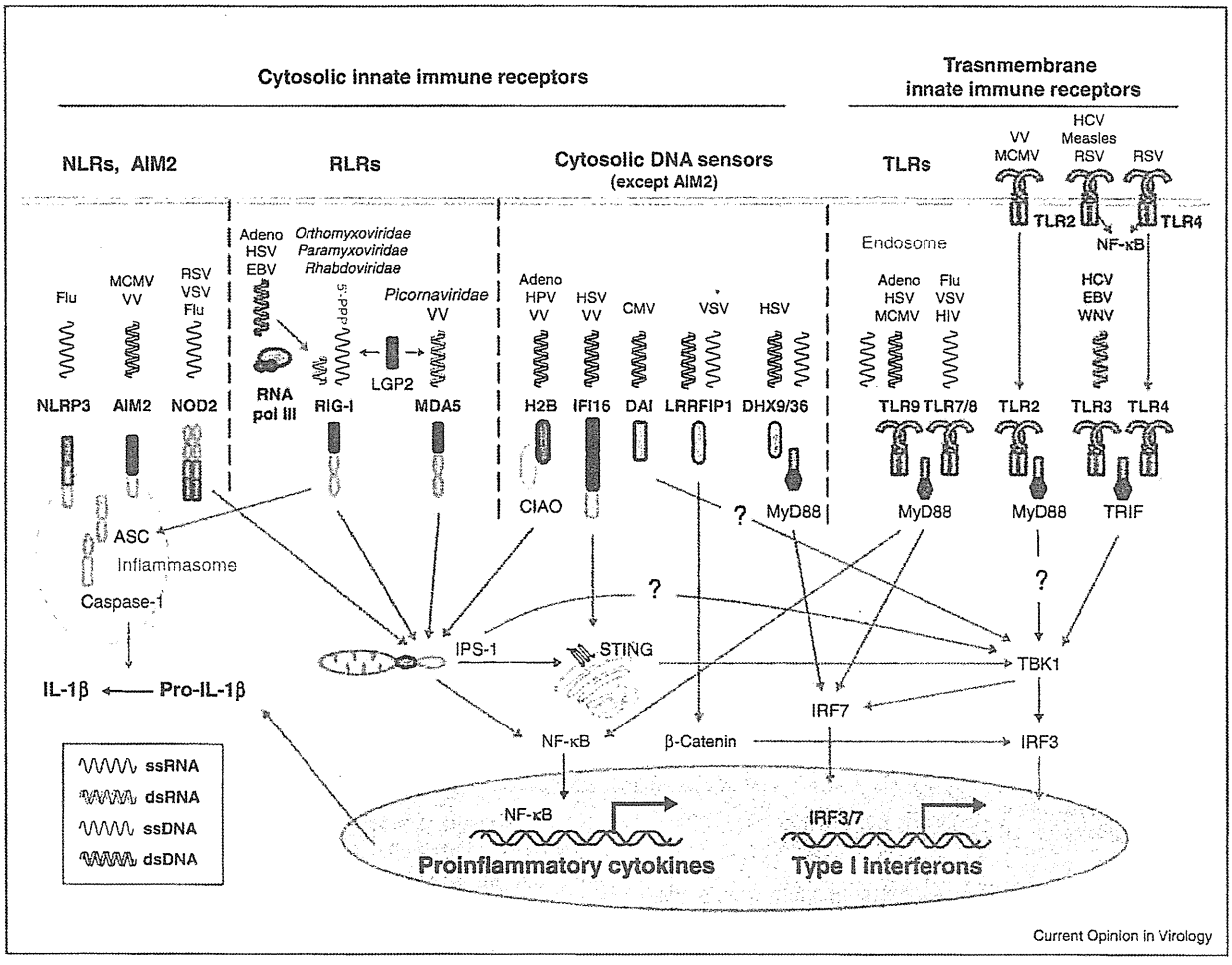
The innate immune system may distinguish between the presence of a virus and viral infection

In addition to bacteria and parasites, viruses are a major cause of infectious diseases. Because of their diverse organ/tissue tropisms, genomic structure (positive or negative stranded, single or double stranded, RNA or DNA) and pathogenic lifecycles, host cells can recognize viruses through a variety of innate immune receptors. Extracellular viruses are detected by transmembrane receptors such as TLRs, and cytosolic viral infections are detected by cytosolic receptors such as RLRs, NLRs, and ALRs (Figure 1). This diverse set of innate receptors may also allow the host immune system to determine viral status — live or dead, replicating or not replicating, pathogenic or non-pathogenic — in a manner similar to that recently proposed for bacterial infection [13]. These innate immune receptors trigger signaling cascades that are generally integrated with innate responses, such as nuclear factor kappa B (NF-κB)-dependent cytokine responses, interferon regulatory factor (IRF)-dependent IFN-α/β responses, and inflammasome/caspase-1-dependent IL-1β responses. IFN-α/β are the major cytokines that limit viral replication, while other cytokines, including IL-6, TNF-α and IL-1β, recruit immune cells to the site of infection and elicit inflammation. NF-κB-dependent and IRF-dependent cytokines are transcriptionally regulated, whereas inflammasome-dependent IL-1β secretion is regulated both transcriptionally and post-transcriptionally (Figure 1). Importantly, many viruses can suppress these innate responses at the 'sensing' and/or transcriptional level upon replication within infected cells [14].

Immune recognition of viruses by transmembrane innate receptors

Transmembrane innate receptors, such as TLRs, recognize extracellular viruses, and their activation does not

Figure 1



Innate immune receptors involved in virus recognition. **NLR and AIM2 pathways:** NLRP3 is activated by a wide variety of stimuli, including RNA viruses. Foreign cytoplasmic dsDNA is also detected by AIM2 via the HIN200 domain. Their activation induces the recruitment of the adaptor protein, ASC, via the pyrin domain. Procaspase-1 is also recruited to ASC via the CARD domain (inflammasome formation). This interaction leads to the auto-cleavage of caspase-1 and results in the activation of caspase-1, which cleaves pro-IL-1β. NOD2 is involved in the recognition of ssRNA viruses. NOD2 activates IPS-1, a mitochondrial membrane-anchored protein, through the NBD and LRR domains, which leads to IRF3 activation. **RLR pathway**: RIG-I is essential for IFN responses to several ssRNA viruses such as *Orthomyxoviridae* and *Paramyxoviridae*. However, MDA5 is necessary for responses to a different set of viruses, such as *Picornaviridae*. LGP2 can act as a positive regulator, making viral RNP complexes more accessible to RIG-I and MDA5. Some viral DNAs are transcribed into 5′ tri-phosphate RNA (the RIG-I ligand) by cytosolic RNA polymerase III (pol III). RIG-I and MDA5 signal via the adaptor protein, IPS-1, which leads to type I IFN production through the TBK1-IRF3-dependent pathway, and proinflammatory cytokine production through NF-κB translocation. RIG-I can activate the inflammasome by interacting with the CARD domains of RIG-I and ASC, and produce IL-1β. **Cytosolic DNA sensor pathways**: extra-chromosomal histone H2B binds DNA virus-like HPV through its α-helical region and interacts with IPS-1 via association with the adaptor protein CIAO. IFI16 binds DNA viruses via the HIN200 domains. They then activate the STING-TBK1-IRF3-dependent signaling pathway, resulting in the production of type I IFN. DAI detects DNA viruses and induces TBK1-IRF3-dependent type I IFN production. LRRFIP1 detects both bacterial DNA and viral RNA from VSV and induces type I IFN production via the β-catenin-IRF3 transactivator pathway. The DExD/H box helicase, DHX9/36, detects CpG-ODNs and DNA viruses such as HSV, leading to MyD88-IRF7-dependent type I IFN production. **TLR pathway**: some RNA viruses are detected by cell surface TLR2 and TLR4, which induce MyD88-dependent NF-κB activation. TLR4 is also recruited to the endosome, leading to TRIF-dependent type I IFN production. TLR3 and TLR7/8 recognize dsRNA and ssRNA, respectively, from RNA viruses. TLR3 induces TRIF-TBK1-dependent type I IFN production, whereas TLR7/8 induces NF-κB and IRF7 activation via MyD88. TLR9 detects CpG-ODNs and DNAs derived from DNA viruses, leading to NF-κB and IRF7 activation via MyD88. Some DNA viruses are also recognized by TLR2 in the endosome, which then induces IRF3/7-dependent type I IFN production.

necessarily require infection of the receptor-expressing cells. Based on cellular localization, TLRs can be grouped in two types: cell surface TLRs (TLR1,2,4,5,6) and endosomal TLRs (TLR3,7,8,9) [1]. Cell surface TLRs recognize bacterial/fungal cell wall components. However, many reports show that some viral proteins are also recognized by cell surface TLR2 and TLR4 [15,16]. A recent report by Barbalat *et al.* identified another interesting example of viral recognition by cell surface TLRs. Mouse cytomegalovirus and vaccinia virus (both dsDNA viruses) were recognized via TLR2. This led to the production of IFN-β, which was not observed upon stimulation with Pam3SK4 (a well-known bacterial TLR2 agonist) [17••]. Interestingly, this TLR2-mediated IFN-β production was restricted in Ly6C(hi) inflammatory monocytes, and was dependent on TLR2 recruitment from the cell surface to the endosome [17••]. However, the exact molecular mechanism(s) underlying virus recognition by cell surface TLRs is the subject of future research. The endosomal TLRs, TLR3, TLR7/8, and TLR9 recognize virus-derived dsRNA, ssRNA, and DNA, respectively [18]. Many viruses are recognized by these endosomal TLRs (Figure 1). TLR3 signaling is mediated by the adaptor molecule TRIF, which induces IRF3 phosphorylation leading to IFN-β production. TLR7/8/9 signaling is mediated by another adaptor molecule, MyD88 (an adaptor commonly used by other TLRs, except TLR3) leading to IRF7-mediated IFN-α production. Importantly, expression of these endosomal TLRs is restricted to certain types of dendritic cells (DCs). TLR3 is preferentially expressed by CD8α(+)DCs, and TLR7/9 is preferentially expressed by plasmacytoid DCs (pDCs). Overall, the recognition of the presence of viruses seems to be mediated by limited types of host cells that express these transmembrane innate immune receptors.

Immune recognition of viruses by cytosolic innate receptors

In contrast to transmembrane receptors, cytosolic innate receptors are expressed by all host cells. RLRs and NLRs mainly recognize viral RNAs, and the recently identified ALRs (and other cytosolic DNA sensors) detect viral DNA in the cytosol of infected cells. This cytosolic receptor-mediated virus recognition is critically important for the host innate immune responses to contain viral replication within the infected cells before the adaptive immune responses are fully developed. In contrast, the contribution of this form of cytosolic virus recognition to adaptive responses is varied and more controversial, as discussed later in this review.

RLRs comprise retinoic acid inducible gene-I (RIG-I), melanoma differentiation associated gene 5 (MDA5), and laboratory of genetics and physiology 2 (LGP2). Both RIG-I and MDA5 recognize viral RNAs within the cytoplasm of infected cells. However, the exact molecular signatures of the RIG-I and MDA5 ligands are still not fully understood [19–22]. Owing to the lack of a caspase recruitment domain (CARD), which is important for interactions with IPS-1, LGP2 was assumed to function as a negative regulator of RIG-I and MDA5. However, a recent study suggests that LGP2 positively regulates RIG-I and MDA5 signaling, possibly by modifying the viral RNA structure before detection by these two receptors [23]. Virus recognition by RIG-I and MDA5 is mediated by a single adaptor molecule, IPS-1 (or MAVS), and leads to NF-κB and IRF3/IRF7 activation. Interestingly, a recent report demonstrated that RIG-I can directly activate ASC in an NLRP3-independent manner, leading to caspase-1-dependent IL-1β production during VSV (ssRNA virus) infection [24••].

NLRs comprise a large number of family member proteins that contain a conserved NOD motif [25], and can be classified into two groups. Activation of Nod1 and Nod2 leads to the activation of NF-κB and IRF. Although Nod2 was initially characterized as a cytosolic sensor for the bacterial cell wall component, muramyl dipeptide, which induces NF-κB activation, a recent report suggests that Nod2 also functions as a virus sensor [26••] and activates a non-classical NLR signaling pathway [27]. Sabbah *et al.* showed that Nod2 can directly sense cytosolic ssRNA from RSV and influenza virus, leading to MAVS(IPS-1)-IRF3-mediated IFN-β responses [26••]. Activation of NLRs, such as NLRP1, NLRP3, and NLRC4, leads to inflammasome formation, which results in caspase-1-mediated IL-1β and IL-18 secretion (Figure 1). The NLRP3 inflammasome is one of the best characterized inflammasomes, and is activated by bacterial toxins, LPS, and viral RNAs, as well as uric acid and alum [11•,28]. Interestingly, it appears that many RLRs and NLRs sense virus infections by detecting viral genomic, or replication-intermediate, RNA. This might indicate that the presence of viral nucleic acids provides the stronger proof of active viral infection, rather than general danger signals.

Cytosolic DNA sensors

AIM2 and IFI16 are both recently identified cytosolic DNA sensors and are involved in DNA-dependent inflammasome activation and IFN-β production, respectively [5,29–34]. Because both proteins contain a PYHIN domain [35–37], it has been proposed that they be referred to as ALRs [5]. However, several other molecules are also known to be involved in DNA sensing within the cytosol. DAI (ZBP-1) is the first reported DNA sensor molecule that triggers TBK1-IRF3-dependent IFN-β induction *in vitro* [38]; however, gene knockout mice do not show the same phenotype, suggesting the presence of redundant DNA sensor mechanisms [39]. Lrrfip1 recognizes cytosolic dsDNA (and dsRNA), subsequently interacting with β-catenin and enhancing IRF3-mediated IFN-β expression [40]. DHX36 and DHX9, present in human pDC, are cytosolic CpG-A and CpG-B binding

proteins, respectively. These proteins mediate the MyD88/IRF7-dependent production of IFN-α [41]. AT-rich DNA is also recognized indirectly by RNA polymerase III. AT-rich DNA is transcribed into 5-triphosphate dsRNA, which is then recognized by the RIG-I pathway [42,43]. The cytosolic histone, H2B, is also involved in DNA sensing. The dsDNA/H2B complex activates IPS-1 via CIAO (an adaptor molecule that links histone H2B and IPS-1) in human cell lines leading to IFN-β expression. However, this H2B-mediated dsDNA-dependent IFN-β production is not observed in mice, most likely because of the lack of the interaction between mouse CIAO and mouse IPS-1 [44].

Signaling via which innate immune receptors leads to adaptive immune responses: TLRs, RLRs, NLRs, or others?

Activation of the innate immune system is critical for establishing adaptive immune responses. This is simply demonstrated by the fact that immunization with a highly purified recombinant protein is usually unsuccessful owing to the lack of innate responses [45]. On the other hand, viral infections are usually sensed by multiple innate receptors. The live attenuated yellow fever vaccine 17D has been shown to activate multiple TLRs, resulted in CD8T and a mixed Th1/Th2 immune responses [46,47]. In the case of influenza virus, infection can be detected by three different receptors: TLR7, RIG-I, and NLRP3 [48•,49]. However, conflicting results have been reported, particularly in terms of the adaptive immune responses examined in these receptors/adaptors deficient mice (Table 1).

The involvement of TLR7/MyD88 has been examined by four independent studies [50–53]. It is very difficult to generalize the results, which range from almost nothing to identifying a prominent phenotype, except that all of the studies consistently agreed that CD8T responses were not affected by the absence of the TLR7/MyD88 pathway. However, two independent studies consistently demonstrated that, in contrast to live virus, the immunogenicity of a chemically killed (inactivated) whole virus was completely dependent on TLR7/MyD88 signaling [48•,54•].

The RIG-I/IPS-1 pathway was also examined in two independent studies [52,53]. They concluded that although RIG-I/IPS-1 signaling induces almost overlapping cytokine responses to those induced by TLR7/MyD88 (Figure 1), IPS-1-deficiency had no substantial effect upon adaptive responses to influenza virus infection [52,53]. This may reflect differential cellular expression of these receptors. RIG-I is ubiquitously expressed by most cells, whereas TLR7 is preferentially expressed by pDCs. It may also reflect the fact that RIG-I sensing requires viral replication within the cell, whereas TLR7 recognizes viruses in the endosome, which is not dependent upon virus infection (Figure 1). Differential regulation of adaptive immune responses by TLRs and RLRs has also been reported in another virus infection system. Jung *et al.* demonstrated that during LCMV infection, CD8T responses in MyD88-deficient mice were significantly reduced, whereas IPS-1-deficient mice showed comparable CD8T responses to those of wild-type mice [55].

NLRP3 can be triggered by viral RNA [56] and/or ionic perturbation caused by the influenza M2 protein [57]. NLRP3 triggers ASC-mediated NLRP3 inflammasome formation, leading to caspase-1-dependent IL-1β and IL-18 secretion. Inflammasome involvement in influenza virus infection has been studied by four independent

Table 1

Adaptive immune responses in mice deficient in innate immune receptors/adaptors against influenza virus infection and vaccination.

	Virus	TLRs(Myd88) deficiency	RLRs(IPS-1) deficiency	NLRs(NLRP3, ASC, caspase-1) deficiency
		IFN-I, TNFα/IL-6		Inflammasome (IL-1/IL-18)
Lopez *et al.* [50]	A/PR8	CD4(IFNγ)→, CD4(IL-4)T, CD8→, Ab→	Not examined	Not examined
Heer *et al.* [51]	A/PR8	CD4→, CD8→, Ab(IgG2a)↓	Not examined	Not examined
Koyama *et al.* [52]	A/PR8, A/NC	CD4(IFNγ)↓, CD8→, Ab(IgG2a)↓	CD4(IFNγ)→, CD8→, Ab→	Not examined
Seo *et al.* [53]	A/PR8	CD4(Th1)↓, CD4(Th2)↑, CD8→, Ab→	CD4(Th1)→, CD8→, Ab→	Not examined
Ichinohe *et al.* [58]	A/PR8	Not examined	Not examined	CD4(IFNγ)↓ CD8↓, Ab(IgG, IgA)↓
Allen *et al.* [59]	A/PR8	Not examined	Not examined	Intact adaptive responses (CD8→, Ab→)
Thomas *et al.* [60]	A/PR8	Not examined	Not examined	CD8→, Ab→
Koyama *et al.* [48•]	A/NC	Not examined	Not examined	CD4(IFNγ)→, CD8→, Ab(IgG1)↓
	Inactivated WV(A/NC)	CD4→, Ab↓	CD4→, Ab→	CD4→, Ab→
Geeraedts *et al.* [54•]	Inactivated WV(H5N1)	CD4↓, Ab↓	Not examined	Not examined

groups. Ichinohe *et al.* demonstrated that NLRP3-independent, but ASC-dependent, inflammasome responses were important for both CD4T and CD8T responses, as well as IgA and IgG responses [58] (it is noteworthy that Poeck *et al.* also reported NLRP3-independent, but RIG-I and ASC-dependent, inflammasome activation by VSV [24••]). In contrast, Allen *et al.* and Thomas *et al.* showed that NLRP3 inflammasome responses were not involved in adaptive responses, but play a more important role in the innate phase of host defense and in tissue healing [59,60]. We also examined ASC-deficient mice and found that inflammasome activation had almost no impact on the adaptive response to live influenza virus infection [48•]. At present, the reason for these contradictory results is not clear [11•,61].

Viral subversion of innate immune responses may affect adaptive immune responses

These controversies may be explained by differences in the types of virus used; especially the different subversion mechanisms used by the viruses. Influenza virus (and other viruses) possesses an immune evasion protein that modulates the innate immune signaling cascades of the host [14]. Even though most studies used a mouse-adapted PR8 virus, Heynisch *et al.* reported that two variants of A/PuertoRico/8/34 show very different activation patterns for cellular signaling molecules in MDCK cells [62]. This most likely reflects the fact that these variant viruses modulate cytosolic signaling systems in different ways. Influenza NS-1 is the most well-characterized of the proteins that subvert RIG-I mediated IFN-α/β responses at multiple steps [63]. A recent report suggests that the inflammasome is also an evasion target of a herpes virus [64]. Intriguingly, no direct viral mechanism that antagonizes TLR signaling has been described for influenza A virus [63]. Taken together, these data suggest that the same PR8 virus may induce very different host immune responses. Furthermore, they may also suggest that subverting the infection-dependent cytosolic innate system may be easier than subverting the infection-independent TLR system. In line with this hypothesis, once the virus is fixed with formalin (and killed), the host immune response is consistently TLR7/MyD88-dependent [48•,54•].

Conclusions

The existence of diverse innate immune receptors may reflect a redundancy that ensures sensitive detection of viruses in a variety of tissue and cell types, and the subsequent induction of host defense mechanisms. TLRs can detect extracellular viruses (either live or dead), and do not require viral infection of receptor-expressing cells. By contrast, detection by cytosolic receptors requires viral infection and replication, which can be easier evasion targets for many viruses. The innate immune response plays two roles in host defense: (1) it limits (or at least controls) viral replication during initial infection; and (2) it induces adaptive immune responses responsible for viral clearance and maintenance (memory). However, it is still not clear to what extent each innate immune receptor contributes to the adaptive immune responses. Owing to sophisticated immune evasion mechanisms, infection by live viruses may not provide a clear answer. However, immunization with an inactivated whole virion influenza vaccine clearly demonstrates that TLR-mediated innate signaling alone is sufficient to induce adaptive immune responses. Currently, it is difficult to examine the individual contribution of each RLR and NLR to the adaptive immune response because of the lack of selective activators. Recently, Kasturi *et al.* demonstrated that synthetic nanoparticle based vaccines composed of multiple TLR ligands induced persistent antibody and CD8T responses than single TLR activating vaccine [65]. It suggests that activations of multiple innate immune receptors may be required for long lasting memory responses but not necessarily required for mounting temporal effector responses. Further studies will clarify the more detailed coordination between innate and adaptive immune responses, and provide a more rational way of vaccine design.

Conflicts of interest statement

The authors have no conflicts of interest to declare.

Acknowledgements

The authors thank Drs. Cevayir Coban and Fumihiko Takeshita for helpful discussion, and all members of Ishii's and Coban's laboratory. This work was supported by Ministry of Health, Labour and Welfare (MHLW) (KJI), the Knowledge Cluster Initiative (KJI), a Grant-in-Aid for Scientific Research (KAKENHI) (TA, KK and KJI) from the Japanese Ministry of Education, Culture, Sports, Science and Technology, and by CREST, JST (KJI).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Kawai T, Akira S: **The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors.** *Nat Immunol* 2010, **11**:373-384.
2. Geijtenbeek TBH, Gringhuis SI: **Signalling through C-type lectin receptors: shaping immune responses.** *Nat Rev Immunol* 2009, **9**:465-479.
3. Yoneyama M, Fujita T: **RNA recognition and signal transduction by RIG-I-like receptors.** *Immunol Rev* 2009, **227**:54-65.
4. Franchi L, Warner N, Viani K, Nunez G: **Function of Nod-like receptors in microbial recognition and host defense.** *Immunol Rev* 2009, **227**:106-128.
5. Unterholzner L, Keating SE, Baran M, Horan KA, Jensen SB, Sharma S, Sirois CM, Jin TC, Latz E, Xiao TS *et al.*: **IFI16 is an innate immune sensor for intracellular DNA.** *Nat Immunol* 2010, **11**:997-1004.
6. Chen GY, Nunez G: **Sterile inflammation: sensing and reacting to damage.** *Nat Rev Immunol* 2010, **10**:826-837.
7. Takeuchi O, Akira S: **Pattern recognition receptors and inflammation.** *Cell* 2010, **140**:805-820.

8. Palm NW, Medzhitov R: **Pattern recognition receptors and control of adaptive immunity.** *Immunol Rev* 2009, **227**:221-233.
• This (also Ref [11•,12•]) is an insightful review that discusses the connection between innate and adaptive immune responses.

9. Manicassamy S, Pulendran B: **Modulation of adaptive immunity with Toll-like receptors.** *Semin Immunol* 2009, **21**:185-193.

10. Koyama S, Coban C, Aoshi T, Horii T, Akira S, Ishii KJ: **Innate immune control of nucleic acid-based vaccine immunogenicity.** *Expert Rev Vaccines* 2009, **8**:1099-1107.

11. Williams A, Flavell RA, Eisenbarth SC: **The role of NOD-like receptors in shaping adaptive immunity.** *Curr Opin Immunol* 2010, **22**:34-40.
• See annotation to Ref [8•].

12. Joffre O, Nolte MA, Sporri R, Sousa CRE: **Inflammatory signals in dendritic cell activation and the induction of adaptive immunity.** *Immunol Rev* 2009, **227**:234-247.
• See annotation to Ref [8•].

13. Vance RE, Isberg RR, Portnoy DA: **Patterns of pathogenesis: discrimination of pathogenic and nonpathogenic microbes by the innate immune system.** *Cell Host Microbe* 2009, **6**:10-21.

14. Versteeg GA, Garcia-Sastre A: **Viral tricks to grid-lock the type I interferon system.** *Curr Opin Microbiol* 2010, **13**:508-516.

15. Finberg RW, Wang JP, Kurt-Jones EA: **Toll like receptors and viruses.** *Rev Med Virol* 2007, **17**:35-43.

16. Carty M, Bowie AG: **Recent insights into the role of Toll-like receptors in viral infection.** *Clin Exp Immunol* 2010, **161**:397-406.

17. Barbalat R, Lau L, Locksley RM, Barton GM: **Toll-like receptor 2 on inflammatory monocytes induces type I interferon in response to viral but not bacterial ligands.** *Nat Immunol* 2009, **10**:1200-1207.
•• This paper showed that mouse cytomegalovirus and vaccinia virus infection can induce MyD88-mediated IFN responses in Ly6C(hi)CD11b(+)CD11c(−)B220(−) inflammatory monocytes.

18. Blasius AL, Beutler B: **Intracellular Toll-like receptors.** *Immunity* 2010, **32**:305-315.

19. Schlee M, Roth A, Hornung V, Hagmann CA, Wimmenauer V, Barchet W, Coch C, Janke M, Mihailovic A, Wardle G *et al.*: **Recognition of 5′ triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus.** *Immunity* 2009, **31**:25-34.

20. Rehwinkel J, Tan CP, Goubau D, Schulz O, Pichlmair A, Bier K, Robb N, Vreede F, Barclay W, Fodor E *et al.*: **RIG-I detects viral genomic RNA during negative-strand RNA virus infection.** *Cell* 2010, **140**:397-408.

21. Pichlmair A, Schulz O, Tan CP, Rehwinkel J, Kato H, Takeuchi O, Akira S, Way M, Schiavo G, Sousa CRE: **Activation of MDA5 requires higher-order RNA structures generated during virus infection.** *J Virol* 2009, **83**:10761-10769.

22. Schmidt A, Endres S, Rothenfusser S: **Pattern recognition of viral nucleic acids by RIG-I-like helicases.** *J Mol Med* 2011, **89**:5-12.

23. Satoh T, Kato H, Kumagai Y, Yoneyama M, Sato S, Matsushita K, Tsujimura T, Fujita T, Akira S, Takeuchi O: **LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses.** *Proc Natl Acad Sci U S A* 2010, **107**:1512-1517.

24. Poeck H, Bscheider M, Gross O, Finger K, Roth S, Rebsamen M, Hannesschlager N, Schlee M, Rothenfusser S, Barchet W *et al.*: **Recognition of RNA virus by RIG-I results in activation of CARD9 and inflammasome signaling for interleukin 1 beta production.** *Nat Immunol* 2010, **11**:63-1824.
•• This paper showed that cytosolic RNA from VSV sensed by RIG-I can induce ASC/inflammasome activation in an NLRP3-independent manner.

25. Kanneganti TD: **Central roles of NLRs and inflammasomes in viral infection.** *Nat Rev Immunol* 2010, **10**:688-698.

26. Sabbah A, Chang TH, Harnack R, Frohlich V, Tominaga K, Dube PH, Xiang Y, Bose S: **Activation of innate immune antiviral responses by Nod2.** *Nat Immunol* 2009, **10**:1073-1080.
•• This paper showed that Nod2 directly recognizes ssRNA and activates the MAVS(IPS-1)/IRF3/IFN-β pathway before RIG-I mediated sensing.

27. Shaw MH, Kamada N, Warner N, Kim YG, Nunez G: **The ever-expanding function of NOD2: autophagy, viral recognition, and T cell activation.** *Trends Immunol* 2011, **32**:73-79.

28. Martinon F, Mayor A, Tschopp J: **The inflammasomes: guardians of the body.** *Ann Rev Immunol* 2009, **27**:229-265.

29. Burckstummer T, Baumann C, Bluml S, Dixit E, Durnberger G, Jahn H, Planyavsky M, Bilban M, Colinge J, Bennett KL *et al.*: **An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome.** *Nat Immunol* 2009, **10**:266-272.

30. Fernandes-Alnemri T, Yu JW, Juliana C, Solorzano L, Kang S, Wu JH, Datta P, McCormick M, Huang L, McDermott E *et al.*: **The AIM2 inflammasome is critical for innate immunity to *Francisella tularensis*.** *Nat Immunol* 2010, **11**:385-394.

31. Hornung V, Ablasser A, Charrel-Dennis M, Bauernfeind F, Horvath G, Caffrey DR, Latz E, Fitzgerald KA: **AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC.** *Nature* 2009, **458**:514-518.

32. Roberts TL, Idris A, Dunn JA, Kelly GM, Burnton CM, Hodgson S, Hardy LL, Garceau V, Sweet MJ, Ross IL *et al.*: **HIN-200 proteins regulate caspase activation in response to foreign cytoplasmic DNA.** *Science* 2009, **323**:1057-1060.

33. Jones JW, Kayagaki N, Broz P, Henry T, Newton K, O'Rourke K, Chan S, Dong J, Qu Y, Roose-Girma M *et al.*: **Absent in melanoma 2 is required for innate immune recognition of *Francisella tularensis*.** *Proc Natl Acad Sci U S A* 2010, **107**:9771-9776.

34. Rathinam VAK, Jiang ZZ, Waggoner SN, Sharma S, Cole LE, Waggoner L, Vanaja SK, Monks BG, Ganesan S, Latz E *et al.*: **The AIM2 inflammasome is essential for host defense against cytosolic bacteria and DNA viruses.** *Nat Immunol* 2010, **11**:395-403.

35. Goubau D, Rehwinkel J, Sousa CRE: **PYHIN proteins: center stage in DNA sensing.** *Nat Immunol* 2010, **11**:984-986.

36. Choubey D, Duan X, Dickerson E, Ponomareva L, Panchanathan R, Shen H, Srivastava R: **Interferon-inducible p200-family proteins as novel sensors of cytoplasmic DNA: role in inflammation and autoimmunity.** *J Interferon Cytokine Res* 2010, **30**:371-380.

37. Gariglio M, Mondini M, De Andrea M, Landolfo S: **The multifaceted interferon-inducible p200 family proteins: from cell biology to human pathology.** *J Interferon Cytokine Res* 2011, **31**:159-172.

38. Takaoka A, Wang Z, Choi MK, Yanai H, Negishi H, Ban T, Lu Y, Miyagishi M, Kodama T, Honda K *et al.*: **DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response.** *Nature* 2007, **448**:501-505.

39. Ishii KJ, Kawagoe T, Koyama S, Matsui K, Kumar H, Kawai T, Uematsu S, Takeuchi O, Takeshita F, Coban C *et al.*: **TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines.** *Nature* 2008, **451**:725-729.

40. Yang PY, An HZ, Liu XG, Wen MY, Zheng YY, Rui YC, Cao XT: **The cytosolic nucleic acid sensor LRRFIP1 mediates the production of type I interferon via a beta-catenin-dependent pathway.** *Nat Immunol* 2010, **11**:487-494.

41. Kim T, Pazhoor S, Bao MS, Zhang ZQ, Hanabuchi S, Facchinetti V, Bover L, Plumas J, Chaperot L, Qin J *et al.*: **Aspartate-glutamate-alanine-histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells.** *Proc Natl Acad Sci U S A* 2010, **107**:15181-15186.

42. Ablasser A, Bauernfeind F, Hartmann G, Latz E, Fitzgerald KA, Hornung V: **RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate.** *Nat Immunol* 2009, **10**:1065-1072.

43. Chiu YH, MacMillan JB, Chen ZJJ: **RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway.** *Cell* 2009, **138**:576-591.

44. Kobiyama K, Takeshita F, Jounai N, Sakaue-Sawano A, Miyawaki A, Ishii KJ, Kawai T, Sasaki S, Hirano H, Ishii N *et al.*: