

基于甲流病毒突变度的疫情早期预警模型

汪 康, 高 洁*

(江南大学 理学院, 江苏 无锡 214122)

摘要: 通过比较每年与其上一年报道的甲流病毒 HA 蛋白质序列之间的差异, 构建了突变度的算法。对所选的 1916~2014 年 HA 蛋白质序列利用此算法计算出每年的突变度, 观察到这 4 次大型流感爆发年份的突变度值都比相邻的年份要高, 并且突变度值越大, 流感的范围和影响就越大。结果表明这种方法建立的比较合理, 效果不错。这对流感爆发的研究、预警和防御有着重要的意义。

关键词: 甲型流感; HA 蛋白质; 氨基酸序列; 突变度

中图分类号: TS 101.4; C 81 **文献标志码:** A **文章编号:** 1673—1689(2017)05—0542—05

Early Warning of Influenza Virus HA Protein Sequences

WANG Kang, GAO Jie*

(School of Science, Jiangnan University, Wuxi 214122, China)

Abstract: This article constructed the algorithm of mutation degree by comparing the difference of two consecutive year's reports on HA protein sequences of pandemic influenza virus. Based on 1916~2014 HA protein sequences, the mutations of every year were calculated by the algorithm, and thus it was observed that the mutation values of four major flu outbreak years were higher than those of the adjacent years, and the higher of the mutation value, the greater the scope and impact of pandemic influenza. This approach was rational and would benefit the study of influenza outbreak as well as the early warning and defense.

Keywords: influenza, HA protein, amino acid sequences, mutation degree

人类最常见的流行性感病毒是甲型流感病毒, 因为甲流具有多变性, 病毒突变能够感染人类, 并且能够人传人^[1-2]。血凝素(HA)是甲型流感病毒蛋白质最重要的一种, 在甲型流感病毒感染过程中它起着极其重要的作用, 同时它刺激机体产生的体液免疫应答对清除已感染的流感病毒起到了极

其重要的作用^[3-5]。目前, 人们从各个方面对甲型流感病毒进行了研究。He 等^[6]是对禽流感病毒 HA 蛋白质抗原表位的识别进行深入的研究; Chang 等^[6]对流感疫苗进行深入研究, 从而达到预防流感的效果; Banerjee 等^[7]对 H1N1 中的所有 HA 基因序列的结构特征和全球氨基酸的组成进行全面的比较, 进

收稿日期: 2015-06-07

基金项目: 国家自然科学基金项目(11271163)。

* 通信作者: 高 洁(1972—), 女, 江苏无锡人, 理学博士, 副教授, 硕士研究生导师, 主要从事生物信息学方面的研究。

E-mail: ezhun6669@sina.com

引用本文: 汪康, 高洁. 基于甲流病毒突变度的疫情早期预警模型[J]. 食品与生物技术学报, 2017, 36(05): 542-546.

行描绘出将来 H1N1 流感进一步发展的可能性。

本文作者选取了 1916 年至 2014 年之间的流感病毒(HA)蛋白质氨基酸序列,除极个别年份丢失外,通过比较每年与其上一年报道的甲流病毒 HA 蛋白质序列之间的差异,计算出每年中每一个甲流病毒 HA 蛋白质的标准差,再通过突变度的算法计算出每年的突变度。可以观察到标准差越小,则 HA 蛋白质的突变度越大,人类就更容易感染,因此要尤为注意此流感病毒,加强预防措施,从而达到早期预警作用。这样将会大大降低甲型流感病毒的传播速度,并能有效的抑制病情的传播,从而有效的阻止甲型流感在人群中爆发。

1 研究方法

蛋白质是由 20 种不同的氨基酸链接形成的多聚体。假设甲流病毒 HA 中的某个蛋白质 y 是由 t 个氨基酸依次链接形成,它的氨基酸序列表示为 $y = x_1x_2 \cdots x_t$, 其中 $x_i \in \{A, R, N, D, C, E, Q, G, H, I, L, K, M, F, P, S, T, W, Y, V\}; i=1, 2, \cdots, t$ 。设 $s-1$ 年全世界各地报道了 m 个甲流病毒 HA 蛋白质,它们的氨基酸序列分别是 $y_{s-1,1}, y_{s-1,2}, \cdots, y_{s-1,m}$; s 年报道了 n 个甲流病毒 HA 蛋白质,它们的序列分别是 $y_{s,1}, y_{s,2}, \cdots, y_{s,n}$; y_{ij} 的氨基酸个数为 c_{ij} , 其中 $s-1, s; j=1, 2, \cdots, q; q=\max\{m, n\}$ 。依次取出 $y_{s-1,1}, y_{s-1,2}, \cdots, y_{s-1,m}$ 中的第 i 个氨基酸,重新组成一个新的氨基酸序列 $Z_{s-1,i}$, 其中 $i=1, 2, \cdots, k; k=\max\{c_{s-1,1}, c_{s-1,2}, \cdots, c_{s-1,m}\}$ 。

计算 $Z_{s-1,i}$ 中每个氨基酸的出现次数,如果出现次数最多的氨基酸有 2 个或则 2 个以上,不失一般性我们取它们中的第一个氨基酸记为 x_i , 反之,取出现次数最多的即可,其中 $1 \leq i \leq k$ 。把 x_i 依次链接得到 1 个新的氨基酸序列 $U_{s-1} = x_1x_2 \cdots x_k$ 。 $y_{s,1}, y_{s,2}, \cdots, y_{s,n}$ 分别与 U_{s-1} 对应的氨基酸进行一一比较:如果对应氨基酸种类不同就赋值为 1,反之为 0。因此 s 年得到 n 个只有 1 和 0 的序列,记为 $E_{s,1}, E_{s,2}, \cdots, E_{s,n}$ 。例如:2013 年中有 4 个氨基酸序列: $y_{2013,1} = \text{LMQGSTLPRRS}$; $y_{2013,2} = \text{MEARLLCLLC}$; $y_{2013,3} = \text{EEARLLCLLCAATNAD}$; $y_{2013,4} = \text{NTVSSFQNLLCSG}$; 因此 $U_{2013} = \text{LEARLLCLLCSGATNAD}$ 。2014 年有 3 个氨基酸序列,分别是: $y_{2014,1} = \text{LEARLUCLBG}$; $y_{2014,2} = \text{MLLRGLCS}$; $y_{2014,3} = \text{LMEALRLCULGSAQGSTLPRRS}$; 因此 $E_{2014,1} = 0, 0, 0, 0, 1, 0, 0, 1, 1$; $E_{2014,2} = 1, 1, 1, 0, 1, 0, 0, 1$; $E_{2014,3} = 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1$; 分别计算出 $E_{s,1}, E_{s,2}, \cdots, E_{s,n}$ 的标准

差 $\sigma(s,1), \sigma(s,2), \cdots, \sigma(s,n)$ 。由此定义突变度 $f(s)$ 为:

$$f(s) = 1 - \frac{\sum_{i=1}^n \sigma(s,i)}{n}$$

其中 $\sigma(s,i)$ 表示 s 年第 i 个甲流病毒 HA 蛋白质标准差。

2 甲流 HA 蛋白质序列预警

2.1 数据分析

选择 1916~2014 年全世界各地报道的甲流病毒 HA 蛋白质氨基酸序列见表 1, 通过上述研究方法计算出每年的突变度 $f(s)$ 。但是由于 1920、1921、1922、1923、1926、1928、1929、1932 年和 1941 年的甲流病毒 HA 蛋白质氨基酸序列缺失, 因此无法计算出 1920、1921、1922、1923、1924、1926、1927、1928、1929、1930、1932、1933、1941 年和 1942 年的突变度。如果 $\sigma(s,i)$ 越小, 说明 $E_{s,1}$ 中 1 的个数就越多, 则这个甲流病毒 HA 蛋白质与上一年的差异越大, 基因突变的就越厉害, 人类就越容易感染此病毒。因

此 $\frac{\sum_{i=1}^n \sigma(s,i)}{n}$ 越小, 说明 s 年中甲流病毒 HA 蛋白质总体上与上一年的差异越大, 即突变度 $f(s)$ 越大, 所以在 s 年爆发甲型流感的可能性就越大。反之, 爆发的可能性就越低。

2.2 结果

1918 年爆发了西班牙流感, 在 1918~1919 年期间全世界约有 10 亿人感染了流行性感, 其中大概有 4 000 万人死亡^[1,6,8]。如图 1 所示, 1917、1918 年和 1919 年的突变度分别是 0.757 266、0.824 252、0.764 706, 可以看出 1918 年的突变度比 1917 年和 1919 年的都要高很多, 这说明 1918 年的甲流病毒 HA 蛋白质序列相对于上一年的突变率比 1919 年和 1917 年相对上年的要高, 因此人类在 1918 年更容易感染此甲流病毒。

1957 年在中国爆发了大型的流行性感, 俗称亚洲流感, 该病毒在当年迅速传遍了全世界, 并且持续到 1958 年, 最终导致至少 100 万人死亡^[9]。如图 2 所示, 1954 年突变度是 0.679 744, 1955 年是 0.723 869, 1956 年是 0.777 408, 1957 年是 0.700 903, 1958 年是 0.717 463, 1959 年 0.582 661, 1960

表 1 1917~2014 年突变度值

Table 1 Mutation values from 1917~2014

年份	突变度	年份	突变度	年份	突变度	年份	突变度
1917	0.757 266	1952	0.557 580	1973	0.598 993	1994	0.690 963
1918	0.824 252	1953	0.771 734	1974	0.586 253	1995	0.613 733
1919	0.764 706	1954	0.679 744	1975	0.587 718	1996	0.611 473
1925	0.853 910	1955	0.723 869	1976	0.599 587	1997	0.609 230
1931	0.615 930	1956	0.777 408	1977	0.595 550	1998	0.615 971
1934	0.640 717	1957	0.700 903	1978	0.602 281	1999	0.612 545
1935	0.595 277	1958	0.717 463	1979	0.599 949	2000	0.610 610
1936	0.587 628	1959	0.582 661	1980	0.607 628	2001	0.620 487
1937	0.613 120	1960	0.572 100	1981	0.620 394	2002	0.614 625
1938	0.684 987	1961	0.590 653	1982	0.618 911	2003	0.634 767
1939	0.685 452	1962	0.554 003	1983	0.605 570	2004	0.731 033
1940	0.745 931	1963	0.633 216	1984	0.613 530	2005	0.642 960
1943	0.653 382	1964	0.575 193	1985	0.607 043	2006	0.620 501
1944	0.618 729	1965	0.578 476	1986	0.594 308	2007	0.648 464
1945	0.605 941	1966	0.588 730	1987	0.596 478	2008	0.614 772
1946	0.632 589	1967	0.581 692	1988	0.594 618	2009	0.722 447
1947	0.589 532	1968	0.610 796	1989	0.612 478	2010	0.586 937
1948	0.616 410	1969	0.579 518	1990	0.595 437	2011	0.622 589
1949	0.618 356	1970	0.573 542	1991	0.610 981	2012	0.660 562
1950	0.597 627	1971	0.586 072	1992	0.604 166	2013	0.705 504
1951	0.574 381	1972	0.579 077	1993	0.664 612	2014	0.596 943

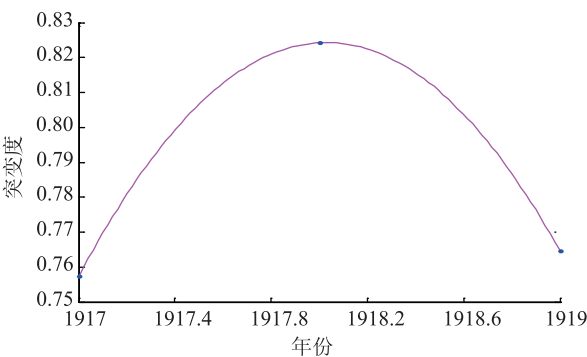


图 1 1917~1919 年甲流 HA 蛋白质突变度趋势图

Fig. 1 Trend chart of pandemic influenza HA protein mutation degree from 1917~1919

年是 0.572 100。可以看出 1954 年和 1955 年的突变度比较高,说明病毒 HA 蛋白质在这两年已经开始突变,但是没有产生大范围的感染,说明病毒已经处在潜伏期。然而 1956 年和 1957 年的突变度也比较高,表明这两年甲流病毒 HA 蛋白质继续处在高突变率下,最终导致流感大爆发。

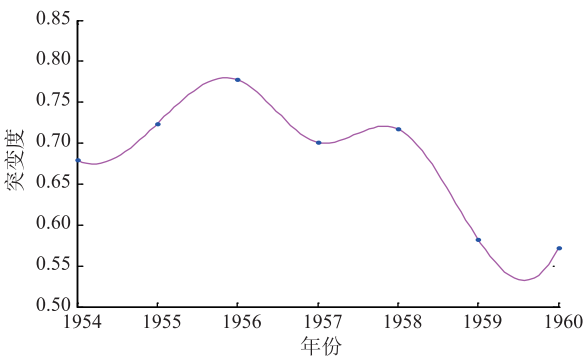


图 2 1954~1960 年甲流 HA 蛋白质突变度趋势图

Fig. 2 Trend chart of pandemic influenza HA protein mutation degree from 1954~1960

1968 年爆发了香港流感,其中有 75 万人死亡,相对 20 世纪是较不致命的大型流感^[1]。如图 3 所示,1965 年突变度是 0.578 476,1966 年是 0.588 730,1967 年是 0.581 692,1968 年是 0.610 796,1969 年是 0.579 518,1970 年是 0.573 542,1971 年是 0.586 072。可以观察到突变度在 1968 年急剧上升达到最高点,说明在 1968 年甲流病毒 HA 蛋白质突变的相当

厉害,人类很容易感染此病毒。

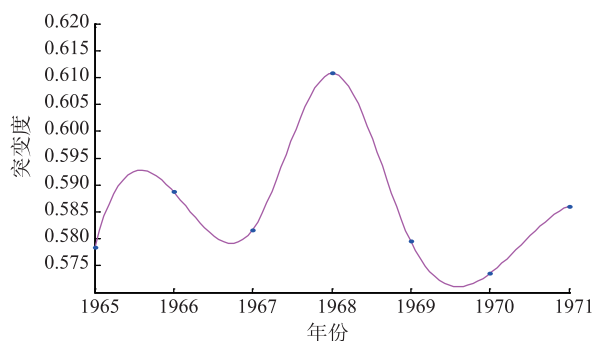


图3 1965~1971年甲流HA蛋白质突变度趋势图

Fig. 3 Trend chart of pandemic influenza HA protein mutation degree from 1965~1971

2009年由一次新型变体甲型H1N1流感病毒所引发的全球性流感,全世界大约有20%的人感染,大约有30万人死亡,因此致死率不高^[7,10]。2013年冬季,墨西哥爆发了大型性流感,共有14 236住院和1 163人死亡的^[10]。如图4所示,2006年标准差是0.620 501,2007年是0.648 464,2008年是0.614 772,2009年是0.722 447,2010年是0.586 937,2011年是0.622 589,2012年是0.660 562,2013年是0.705 504,2014年是0.596 943。2009年和2013年的突变度比相邻两年都要高,说明2009年和2013年甲流病毒HA蛋白质突变率比较高,因此人类容易感染。

这5次流感中,其中1918、1957、1968年和2009年是全球性大型流感。在这4次全球性大型流感中,属西班牙流感的范围最广、影响力最大、突变度最高;香港流感范围最小、影响力最低、突变度最低。因此,流感爆发的强弱跟它的突变度值是有一定的关系的,突变度越大往往流感爆发的范围越广,影响力越大。所以,如果这一年的突变度急剧偏

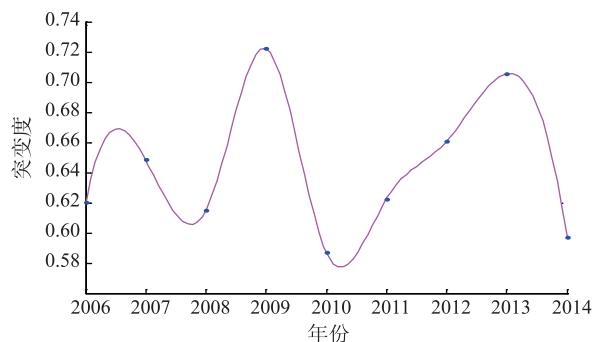


图4 2006~2014年甲流HA蛋白质突变度趋势图

Fig. 4 Trend chart of pandemic influenza HA protein mutation degree from 2006~2014

高,这就预警我们这一年爆发甲型流感的可能性非常大,因此要尤为注意此流感病毒,加强预防措施,防止病情扩散,有效的抑制甲型流感的传播。

3 结 语

通过对每年和其上一年报道的甲流HA蛋白质序列进行对比,计算出每年HA蛋白质序列的突变度。根据这种方法,选取了1916~2014年甲流病毒HA蛋白质氨基酸序列除极个别年份丢失外,分别计算出每年的突变度。从图1、图2、图3和图4中可以观察到,1918、1957、1968年和2009年的突变度分别是0.824 252、0.700 903、0.610 796和0.722 447,发现这4次全球性大型流感爆发年份的HA蛋白质序列的突变度都比较高,并且发现大型流感的范围和影响力越大,它们HA蛋白质序列的突变度往往就越大。这充分的说明每年甲流病毒HA蛋白质氨基酸序列的突变度值与爆发每年甲型流感病毒可能性有着很大的关系,这对甲流病毒的研究和预警有着重要的意义。同时,也表明此方法建立的比较合理,预警效果很好。

参考文献:

- [1] SUN L,DEPUY G W,EVANC G W. Multi-objective optimization models for patient allocation during a pandemic influenza outbreak[J]. **Computers & Operations Research**,2014,51:350-359.
- [2] REN D,GAO J. Early-warning signals for an outbreak of the influenza pandemic [J]. **Chinese Physical Society and IOP Publishing Ltd**,2011,20:1287011-1287014.
- [3] LUCCHESI G,CAPONE G,KANDUC D. H1N1 versus H5N1 hemagglutinins:A possible differential immunologic impact on neurodevelopment[J]. **Neurology,Psychiatry and Brain Research**,2015,21(1):39-50.
- [4] TSE L V,WHITTAKER G R. Modification of the hemagglutinin cleavage site allows indirect activation of avian influenza virus H9N2 by bacterial staphylokinase[J]. **Virology**,2015,482:1-8.

- [5] HE J L,HSIEH M S,JUANG R H,et al. A monoclonal antibody recognizes a highly conserved neutralizing epitope on hemagglutinin of H6N1 avian influenza virus[J]. **Veterinary Microbiology**,2014,174(3-4):333-341.
- [6] CHANG L Y,SHIH S R,SHAO P L,et al. Novel swine-origin influenza virus A (H1N1);the first pandemic of the 21st century [J]. **Journal of the Formosan Medical Association**,2009,108(7):526-532.
- [7] BANERJEE R,ROY A,DAS S,et al. Similarity of currently circulating H1N1 virus with the 2009 pandemic clone;Viability of an imminent pandemic[J]. **Infection,Genetics and Evolution**,2015,32:107-112.
- [8] BASLER C F,AGUILAR P V. Progress in identifying virulence determinants of the 1918 H1N1 and the Southeast Asian H5N1 influenza A viruses[J]. **Antiviral Research**,2008,79(3):166-178.
- [9] GIRARD M P,TAM J S,ASSOSSOU O M,et al. The 2009 A(H1N1) influenza virus pandemic;A review[J]. **Vaccine**,2010,28(31):4895-4902.
- [10] JAVIER D T,CHOWELL G,VICTOR H B A,et al. Intense seasonal A/H1N1 influenza in Mexico,winter 2013-2014 [J]. **Archives of Medical Research**,2015,46(1):63-70.

科 技 信 息

加拿大禁止部分氢化油用于食品加工

2017年4月7日,加拿大卫生部宣布将禁止食品行业产生反式脂肪的主要来源——部分氢化油(PHO)用于食品,该项禁令将禁止所有使用PHO的食品在加拿大销售。这是继该国要求在食品标签上强制性标识反式脂肪及在加工食品中自愿标识反式脂肪最大含量之后的又一重大举措,旨在将该类脂肪控制在尽可能低的水平。现加拿大卫生部就该项禁令公开征求意见,征求意见期限为2017年4月7日至6月21日。

[信息来源]厦门 WTO 工作站. 加拿大禁止部分氢化油用于食品加工 [EB/OL]. (2017-4-13). <http://www.xmtbt-sps.gov.cn/detail.asp?id=54117>

欧洲允许卵磷脂作为食品添加剂使用

2017年4月7日,欧洲食品安全局(EFSA)食品添加剂及食品营养来源专家组(ANS)发布了对食品添加剂卵磷脂(E322)重新风险评估的结果。

根据委员会法规(EU)第257/2010,专家组认为不需设定卵磷脂每日允许摄入量(ADI)的最大限值,根据精确暴露评估结果,卵磷脂作为食品添加剂对于1岁以上的一般人群不存在安全问题,并进一步认为磷脂胆碱作为食品添加剂用于食品行业的现行使用用量及使用水平亦不存在安全问题(包括12周到11个月的婴儿)。同时对于婴儿及儿童食用的特殊医疗用途食品,专家组认为卵磷脂及磷脂胆碱作为食品添加剂使用在该类产品中也不存在安全问题。

[信息来源]厦门 WTO 工作站. 欧洲允许卵磷脂作为食品添加剂使用 [EB/OL]. (2017-4-12). <http://www.xmtbt-sps.gov.cn/detail.asp?id=54107>