

• 基础论著 •

甲型流感病毒基因变异与生存选择压力相关性分析

周晓明 赵慧芳 陆嘉良 潘浩 徐眉 赵根明 姜庆五 汪华 俞顺章

【摘要】 目的 研究甲型流感病毒基因变异与生存选择压力的关系,指导保守疫苗靶位点的寻找。方法 选择相同血凝素(HA)血清型具有全基因组序列的分化距离较近的多个流感病毒株系,利用 Blast2 程序计算各株系之间的核酸保守性、蛋白质保守性、变异核苷酸在密码子序位中的频率以及生存选择压力指数,分析与生存选择压力指数的相关性。结果 HA 基因较其他基因的核酸保守性为显著低;HA 抗原蛋白质保守性与 NS 基因同为较低;HA 基因受生存选择压力作用;NS、聚合酶 B1(PB1)基因受选择压力较小;NA、核蛋白(NP)基因受选择压力较大;生存选择压力指数与核苷酸第 3 密码子位点变异频率完全相关。结论 流感病毒基因变异受复制机制形成差异和生存选择压力淘汰差异双重作用影响;各基因在流感病毒生物功能中地位不同,NA、NP 基因保守性较强,适宜作为疫苗靶位点候选,HA 基因相比较具有一定的保守性,有望寻找到保守区段,NS、PB1 基因保守性较弱,不宜作为靶位点。

【关键词】 流感病毒 A 型;变异(遗传学);选择(遗传学);疫苗;密码子;基因,病毒

Correlation analysis of type A influenza virus genetic variation characteristic with survival selective pressure ZHOU Xiao-ming, ZHAO Hui-fang, LU Jia-liang, et al. School of Public Health, Fudan University, Shanghai 200032, China

Corresponding author: ZHOU Xiao-ming, Email: xmzhou@shmu.edu.cn

【Abstract】 **Objective** Study the relationship between type A influenza virus genetic variation with survival selective pressure, help for the finding of possible vaccine conserved antigen target. **Methods** Select seven strains of same HA (Hemagglutinin) serotype, regional and isolation time closely related type A influenza virus with full HA gene coding sequence; use Blast2 program to calculate the parameter of nucleotide conservative, amino acid conservative, mutation ratio of codon 3rd over non 3rd locus, survival selective pressure indicator of these virus strains; analysis the parameters relationship with survival selective pressure indicator. **Results** Nucleotide conservative of HA gene is significantly lower than that of other genes; amino acid conservative of HA gene is similar with NS gene, all lower than that of other genes; genetic variation of HA gene is under survival selective pressure; selective pressure toward NS, PB1 gene is relatively lower than that toward NA, NP gene; survival selective pressure indicator is strongly correlated with mutation frequency upon codon 3rd locus. **Conclusions** Genetic variation of influenza virus is determined both by mechanism of relax replication model and survival selective pressure; genetic conservative of each gene is different, NA, NP gene could be selected as possible vaccine target for their relative high conservative, HA gene possesses medium genetic conservative with prospective of finding more conserved epitope region within its full sequence, NS, PB1 are not recommended as vaccine candidate for their relative low genetic conservative.

基金项目:国家自然科学基金资助项目(30200233)

作者单位:200032 上海,复旦大学公共卫生学院(周晓明、赵慧芳、陆嘉良、徐眉、赵根明、姜庆五、俞顺章);江苏省疾病预防控制中心(潘浩、汪华)

通信作者:周晓明,Email: xmzhou@shmu.edu.cn

【Key words】 Influenza A Virus; Variation(Genetics); Selection (Genetics); Vaccines; Codon

流感病毒可以引起人类上呼吸道疾病,虽然直接的病死率不高,但是发病率极高,发病人群可达 80%~90%^[1],对一些特殊人群因免疫力下降导致并发症致死的比例也可达 1/1500。由于流感病毒经常发生抗原漂移,迄今为止缺少有效的防治措施,因此经常发生规模大小不等的暴发流行,其传播途径尚不清楚,可能有跨物种传播或是区域扩散传播^[2]。

血凝素(HA)抗原是甲型流感病毒的表面抗原,中和该抗原可导致对流感病毒感染的免疫。不同流行期检测到 HA 抗原存在差异,使前次获得的免疫对本次流行的 HA 抗原无效,形成抗原漂移现象^[3]。对于流感病毒 HA 抗原漂移的原因,一种学说认为归因于 RNA 病毒的松弛复制机制,导致 HA 基因的突变率增高^[4]。由于已检测了较多株系流感病毒的全基因序列,我们分析了这些株系的基因保守性,认为 HA 抗原漂移不但与 RNA 病毒松弛复制机制有关,可能与 HA 基因正常生命功能的生存选择压力也有关系。根据 HA 基因同时受生存选择压力的理论,有望获得针对性较强的抗体用于流感病毒的防治。

材料和方法

一、甲型流感病毒株系与基因组全序列的获得

株系与基因组全序列源于美国国家生物技术信息中心(NCBI)的基因数据库,通过分类检索,选择具有基因组全序列公布的、同一宿主来源的、同一血清型、同一区域以及分离时间接近的甲型流感病毒株系进行分析。

二、核酸同源序列保守性比较

使用 NCBI 网页提供的 Blast 2(Blast n)程序^[5]两两比较同源的核酸片段,列出各序列差异核苷酸的总数以及差异核苷酸的位置。

三、蛋白质同源序列保守性比较

使用 NCBI 网页提供的 Blast 2(Blast p)程序^[6],输入需比对的流感病毒株系基因翻译多肽的蛋白质编码,并选择 Matrix 替代矩阵 Blosum 62 对蛋白质序列保守性进行比对。

四、变异核苷酸位置的确定和统计

根据 Blast 2-Blast n 程序给出两对核酸序列

的比对图,由核酸序列基因库表述格式中的 FEATURES/CDS/Codon_Start 的注释项确定密码子起始位置,根据 FEATURES/CDS 注释项确定密码子的终止位置,统计各变异核苷酸位置在密码子的第 1/2/3 位置出现的频数,在编码序列以外的变异核酸统一计算为非编码变异核酸。总变异碱基数与总核酸长度的比值称为核酸保守性 $cn\%$,总变异氨基酸数与总蛋白长度的比值称为蛋白质保守性 $cp\%$ 。

五、选择压力(粗)指数

无选择压力时,任意核苷酸碱基的变异均会反映到蛋白质的序列变异中,由于三联体密码子通常第三位的变异不影响编码氨基酸(粗),因此 $2(1-cn\%)/(1-cp\%) = 1$;如果存在选择压力,某些导致蛋白质变异的碱基变异会在选择中被淘汰, $2(1-cn\%)/(1-cp\%) > 1$,碱基第三位变异的比率高于一和二;选择压力越高,观察到的总变异频率越低。选择压力指数以 $K = (1-cp\%)/[2(1-cn\%)]$, K 值在 1~0 之间; K 值越小,选择压力越大。

六、统计分析

使用 SPSS 8.0 程序作相关性分析和显著性检验。

结 果

一、选择进行分析的甲型流感病毒株系

根据选择甲型流感病毒株系需具有基因组全序列、同一宿主来源、同一血清型、同一区域来源以及分离时间接近的特点实际选择待分析株系为 H1N2 血清型、猪源、明尼苏达州 2001 年分离获得的 7 个流感病毒株系,分别为 A/SW/MN/16356/01(H1N2)、A/SW/MN/16419/01(H1N2)、A/SW/MN/17138/01(H1N2)、A/SW/MN/22860-S/01(H1N2)、A/SW/MN/22860-T/01(H1N2)、A/SW/MN/23124-S/01(H1N2)、A/SW/MN/23124-T/01(H1N2);分析基因分别为 HA、神经氨酸酶(NA)、核心蛋白(NP)、非结构蛋白(NS)、聚合酶 A(PA)、聚合酶 B1(PB1)、聚合酶 B2(PB2)。

二、核酸碱基的保守性

见表 1。

表 1 流感病毒各基因片段核苷酸保守性

基因名称	统计核苷酸总数	保守核苷酸总数	核苷酸保守率(%)
HA	25 545	23 921	93.6
NA	9 955	9 626	96.7
NP	11 760	11 568	98.4
NS	17 682	17 270	97.7
PA	12 621	12 370	97.5
PB1	11 319	11 119	98.2
PB2	5 754	5 638	98.0

三、蛋白质的保守性

见表 2。

表 2 流感病毒各基因片段蛋白质保守性

基因名称	统计氨基酸总数	同一氨基酸总数	蛋白质保守率(%)
HA	8376	7820	93.3
NA	3297	3251	98.6
NP	3906	3872	99.1
NS	4599	4369	95.0
PA	4200	4107	97.8
PB1	3780	3696	97.8
PB2	1920	1892	98.5

四、核苷酸变异位置及百分比统计

见表 3。

表 3 流感病毒各基因片段核苷酸在密码子各位置变异频率

基因名称	第 1 位置		第 2 位置	
	个数	百分比(%)	个数	百分比(%)
HA	297	19.7	220	14.6
NA	45	24.9	0	0.0
NP	50	26.2	0	0.0
NS	116	35.8	120	37.0
PA	62	19.6	46	14.6
PB1	50	25.0	28	14.0
PB2	18	15.5	10	8.6

基因名称	第 3 位置		非编码区域	
	个数	百分比(%)	个数	百分比(%)
HA	992	65.7	10	0.7
NA	136	75.1	0	0.0
NP	141	73.8	0	0.0
NS	88	27.2	88	27.2
PA	208	65.8	0	0.0
PB1	122	61.0	0	0.0
PB2	88	75.9	0	0.0

五、生存选择压力指数比较

见表 4。

表 4 流感病毒各基因片段生存选择压力指数(%)

基因名称	核苷酸保守率	蛋白质保守率	碱基第 3 位置变异率	生存选择压力指数
HA	93.6	93.3	65.7	0.52
NA	96.7	98.6	75.1	0.21
NP	98.4	99.1	73.8	0.28
NS	97.7	95.0	27.2	0.76
PA	97.5	97.8	65.8	0.44
PB1	98.2	97.8	61.0	0.61
PB2	98.0	98.5	75.9	0.38
与生存压力指数相关性	0.423	0.669	0.996	1.00

讨 论

流感病毒 HA 抗原经常发生漂移,各来源分离获得的甲型流感病毒即使血清型为同一 HA 亚型,免疫作用仍不相同,因此使用减毒活疫苗进行免疫接种为达到较好的效果,需每年监测流感病毒 HA 亚型及具体株系的血清抑制效价,选择相应株系制备当年度的疫苗^[7]。

对于 HA 抗原漂移的原因,根据血清分型的结果有多种假设,包括物种相互间的传播,流感病毒重组等,也有推测源于 RNA 病毒复制的松弛性^[2],但仅根据血清分型的结果不足以进行判断。

HA 抗原漂移如果是由于 RNA 病毒复制松弛性导致,HA 抗原和其他片段基因会发生同样频率的变异;但是 NP 基因等与 HA 基因的保守性是不同的,因此单是 RNA 病毒的复制松弛性不足以解释 HA 抗原漂移特性。可以辅助该假设的为流感病毒各基因还承受生存选择压力;由于 HA 抗原除重要的抗原表位外,有较大片段对其生物活性没有影响,因此可以积累较多的变异点形成漂移,而其他抗原可能生物功能更重要,而不能积累较多的变异。

因为 HA 抗原漂移的原因,迄今为止仍不明白其中保守性的片段。如果 HA 抗原同时也承受生存选择压力,则有望找到 HA 的保守片段,可用于进行基因工程抗原抗体的研究。另外,如 HA 抗原保守性较小,则通过研究流感病毒其他基因的保守性,可研究针对高保守性病毒基因的抗原抗体,虽然有可能不能阻断病毒的感染,但可能以阻断病毒生活史的方式实现对流感病毒的免疫^[7]。

本研究中,选择同血清型、同一宿主来源、同一区域来源以及同年度来源的病毒株系目的为排除其他因素的影响,着重研究生存选择压力与基因核酸和蛋白质保守性的关系。根据本文采取的生存选择压力指数的计算方式,以病毒株系分化亲缘关系较近的计算较为准确;因亲缘分化距离远离后,核酸的突变率不能视为自主突变率,因此不能用于计算生存选择压力。根据基因库中所能找到的满足要求的流感病毒株系,选择其中最大的 1 组 7 个流感病毒株系进行分析。

根据核酸保守性分析,HA 的保守性较低,为 93.6%,其他 6 个基因保守性为(97.75±0.6)%,显示各基因遵从同样的复制机制,HA 可能有不同的复制原理;蛋白质保守性分析,HA 与 NS 的保守性较低,平均为(94.1±1.2)%,其他 4 个基因保守性为(98.4±0.6)%,与 HA 和 NS 的有显著差异;计算生存选择压力,NA 最高为 0.21,显示其在流感病毒生命史中的核心作用,NS 以及 PB1 分别为 0.76 和 0.61,还大于 HA 基因的 0.52,显示 HA 仍具有一定的保守性,PA 以及 PB2 的选择压力接近于 HA 基因,NP 基因选择压力为 0.28 与 NA 基因并列列为流感病毒最重要的基因之一。

生存选择压力越高,则观察到的蛋白质保守性以及核酸在第 3 密码子位置的变异频率越高,因核酸在其他密码子位置的突变会导致蛋白质改变以及在生存选择中被淘汰。生存选择压力与核酸保守率无关,其取决于复制机制的错误率。上述研究表明生存选择压力与核酸保守率相关性为 0.423 没有关系,与核酸第 3 密码子位置变异频率相关性为 0.996 有显著关系,与蛋白质保守率相关性为

0.669 有一定的关系。

根据本文研究指出,HA 抗原基因较流感病毒其他基因可能具有不同的复制机制,并受生存选择压力限制,具有一定的保守性;流感病毒最保守的基因为 NA 基因和 NP 基因,是有希望的流感病毒疫苗的靶位点。NS 及 PB1 基因不适宜作为疫苗位点考虑。流感病毒的抗原漂移主要源于 RNA 病毒复制的松弛机制导致的变异和生存选择压力对部分非保守性序列的淘汰达成的平衡所决定,对 HA 抗原可能找到其中重要的保守性的抗原表位片段。

参 考 文 献

- 1 Couch RB. Advances in influenza virus vaccine research. *Ann NY Acad Sci*. 1993, 685:803-812.
- 2 Gibbs MJ, Armstrong JS, Gibbs AJ. Recombination in the hemagglutinin gene of the 1918 "Spanish Flu". *Science*. 2001, 293:1842-1845.
- 3 Hay AJ, Gregory V, Douglas AR, et al. The evolution of human influenza viruses. *Philos Trans R. Soc Lond B Biol Sci*. 2001, 356:1861-1870.
- 4 Webster RG, Bean WJ, Gorman OT, et al. Evolution and ecology of influenza A viruses. *Microbiol Rev*. 1992, 56:152-179.
- 5 Altschul SF, Gish W, Miller M, et al. Basic local alignment search tool. *J Mol Biol*. 1990, 215:403-410.
- 6 Gish W, States DJ. Identification of protein coding regions by database similarity search. *Nat Genet*. 1993, 3:266-272.
- 7 Brown EG. Influenza virus genetics. *Biomed Pharmacother*. 2000, 54:196-209.

(收稿日期:2004-04-14)

(本文编辑 沈漱瑜)

· 书讯 ·

《医院感染管理与监控》系列教材(VCD)出版

由解放军总医院制作的《医院感染管理与监控》系列教材已由中华医学电子音像出版社出版发行。

本系列教材为医学操作教学片。本套光盘共 7 盘,分别为中心供应室感染管理与监控;产房、婴儿室及母婴同室感染管理与监控;血液净化中心感染管理与监控;手术室感染管理与监控;重症监护病房感染管理与监控;骨髓移植室感染管理与监控;抗菌药物的合理应用。

订购地址:北京东四西大街 42 号 中华医学会电子音像出版社 邮编:100710。联系人:刘峰,电话:010-65133608,传真:010-65133609