

猪瘟病毒编码区基因选择压力及重组分析

贾琳, 姚玉婷, 肖明*

上海师范大学生命与环境科学学院, 微生物分子生物学研究室, 上海 200234

* 联系人, E-mail: xiaoming88@263.net

2013-01-16 收稿, 2013-04-07 接受

国家自然科学基金(31070671)、上海市科委基础研究重点项目(11JC1409300)、上海自然科学基金(12ZR1422100)和上海市教委重点项目(13ZZ102)资助

摘要 为了更全面地探究选择压力和基因重组对猪瘟病毒基因组全编码区进化的影响, 本研究通过生物信息学软件包 PAML 4.5 中的 *codeml* 程序和 RDP 3.44 中的 6 种重组检测算法, 对采集时间 1994~2010 年, 涵盖 3 个基因型中的 6 个基因亚型, 遍布 11 个国家的 50 株猪瘟病毒(CSFV)基因组全编码区进行了选择压力及重组分析. 结果显示编码区基因所编码的全部 12 个蛋白质的选择压力测度参数 ω 均低于 1, 表明 CSFV 所承受选择压力的总体趋势为净化选择, 其中 NS3, NS4A 和 NS4B 蛋白所承受的净化选择压力较大; 而 Npro, E^{ms}, E1, E2 和 NS5A 蛋白含有正选择位点, 存在较大的正选择压力, 表明猪瘟病毒的适应性进化与宿主免疫应答有关. 重组分析的结果共检测出 4 株重组毒株, 其中 3 株为新发现, 分别为 HCLV/99, ALD 和 SWH, 其重组区域所涉及的基因组主要编码与猪瘟病毒免疫和病毒毒力相关的蛋白, 且重组的确改变了重组株所在的进化树分支; 而蛋白 NS3, NS4A 和 NS4B 的编码区均没有发现重组痕迹, 其遗传距离较短, 进化程度较低, 这些结果表明重组可能改变猪瘟病毒毒力, 并影响猪瘟病毒的遗传进化. 由此可见, 猪瘟病毒针对宿主免疫系统的选择压力而产生的突变和重组推动了病毒编码区基因的进化. 本研究为疫苗靶标的研究与日渐趋于复杂的猪瘟病毒流行病学研究提供了理论基础.

关键词

猪瘟病毒
编码区全基因组
适应性进化
重组

猪瘟病毒(classical swine fever virus, CSFV)——猪瘟的病原体^[1], 为有囊膜的单股正链 RNA 病毒, 是严重危害世界养猪业的病毒性传染病. 它与牛病毒性腹泻病毒(BVDV 1 和 BVDV 2)和羊边界病病毒(border disease virus, BDV)同属于黄病毒科瘟病毒属, 猪是其唯一宿主. CSFV 基因组全长 12.3 kb, 5'-端无帽子结构, 3'-端无 poly(A)尾巴, 仅含有 1 个大的开放性阅读框, 编码约含 3898 个氨基酸的多肽. 该多肽经宿主细胞与病毒蛋白酶的共同作用, 裂解为 12 个成熟的蛋白质^[2], 依次为 Npro, C, E^{ms}, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B^[3]. 其中 C, E^{ms}, E1 和 E2 为结构蛋白, 其余为非结构蛋白. 国际上基

于 5' UTR, E2 和 NS5B 等相关核苷酸序列, 将 CSFV 分为 3 个基因型、10 个基因亚型^[4], 这为 CSFV 的分子流行病学研究提供了理论基础.

突变与重组作为驱动 RNA 病毒适应环境的主要分子力量^[5], 成为了研究病毒进化机制的热点. RNA 病毒在进化过程中有 3 个突出特点, 突变速率高、复制时间短、种群规模大^[6]. 其较高的突变速率是由于 RNA 病毒的 RNA 聚合酶在复制过程中缺乏纠错功能, 导致病毒在宿主体内发生变异, 从而产生遗传多样性^[7,8]. 用于检测选择压力的最大似然法检验已经被广泛接受, 其中选择系数 ω (非同义/同义置换速率比值, dN/dS)直观地反映了生物体在密码子水平上的

引用格式: 贾琳, 姚玉婷, 肖明. 猪瘟病毒编码区基因选择压力及重组分析. 科学通报, 2013, 58: 3570-3578

Jia L, Yao Y T, Xiao M. Analysis of selection pressure and recombination in coding-region of classical swine fever virus (in Chinese). Chin Sci Bull (Chin Ver), 2013, 58: 3570-3578, doi: 10.1360/972013-84

进化趋势, $\omega > 1$, $\omega = 1$ 和 $\omega < 1$ 分别代表在进化过程中基因承受正选择、中性选择及净化选择(负选择)^[9]. 目前许多研究表明, 来自宿主特异性免疫的正选择压力有助于病毒蛋白发生变异从而使病毒逃避宿主免疫识别, 发生适应性进化^[10-13], 使得免疫防治效果不佳. 因此, 了解病毒基因多样性的进化过程对于提高疫苗防治效果就显得尤为重要. 有研究表明, 基因组全编码区的选择压力分析对其疫苗的开发具有相当大的指导意义^[14]. 目前已有研究针对 CSFV 毒株特定抗原蛋白 E^{ns} 和 E2 进行选择压力分析, 结果均检测到正选择位点的存在, 但由于样本数据具有时间与地域差异性, 结果不尽相同^[15-17]. 迄今为止未有针对 CSFV 编码区全基因组进行选择压力分析的研究, 且基因重组作为另一个影响病毒基因多样性的原因, 也应该纳入考虑范畴. 基因重组能够直接影响病毒基因变异, 在病毒的进化过程中起到重要作用^[18]. 这一现象已在许多黄病毒科病毒中得到证实, 如牛病毒性腹泻病毒(bovine viral diarrhea virus, BVDV)^[19]、丙型肝炎病毒(hepatitis C virus, HCV)^[20,21]、登革病毒(dengue virus, DEN)^[22]. CSFV 也存在基因重组, 重组现象对其遗传变异产生了影响^[23]. 为了更全面地探究影响病毒基因多样性的因素, 深入了解 CSFV 的进化过程, 针对较长时间段及较大流行区域的病毒基因组进行选择压力及重组分析就变得更为必要. 本研究采用的实验数据采集时间为 1994~2010 年, 涵盖 CSFV 的 3 个基因型中的 6 个基因亚型, 遍布 11 个国家, 应用最大似然法及生物信息学软件 RDP 3.44 对 CSFV 各蛋白质编码序列的选择压力及基因重组情况进行分析.

1 材料与方法

(i) CSFV 序列数据. 本研究采用的 50 条 CSFV 全基因组序列均从 GenBank 上获得, 数据的时间跨度为 1994~2010 年, 涵盖 CSFV 的 3 个基因型中的 6 个基因亚型(1.1, 1.2, 2.1, 2.2, 2.3 和 3.4), 遍布 11 个国家(表 1). 应用 CLUSTAL X1.83^[24]对 50 条 CSFV 序列进行对位排列, 剔除基因组非编码区序列, 编码区部分序列对齐比较工整. 仅有病毒株 LPC 在核苷酸序列第 1547 位置缺失 3 个碱基, 位于 E^{ns} 蛋白第 125 位. 本研究只针对 CSFV 编码区进行分析, 共编码 12 个蛋白质.

(ii) 方法. 应用 MEGA 5.0 依据分别编码 12 个

蛋白质的核苷酸序列构建 N-J 进化树, 自举值为 1000, 验证进化树的可信程度, 并采用最大似然复合法(maximum composite likelihood method)计算各个蛋白的进化距离, 调查各个蛋白质编码区的进化程度.

采用 PAML 4.5 软件包中的 codeml 程序^[25], 基于本研究自建进化树及对位比对结果, 分别对 CSFV 的 12 个蛋白质进行选择压力分析. 在分支模型上选择最简单的单一比率模型, 位点模型选用以下 6 种: 模型 M0(单一比率模型)、模型 M1(中性模型)、模型 M2(选择模型)、模型 M3(离散模型)、模型 M7(Beta 模型)、模型 M8(Beta- ω 模型). 利用似然比检验统计来评估成对模型: M0 vs M3, M1 vs M2 和 M7 vs M8, 检测正选择压力, 其结果遵从 χ^2 分布, 自由度为 2 个模型的参数之差.

重组分析采用 RDP 3.44 软件包中的 6 种算法: RDP, GENBCONV^[26], BootScan^[27], Maxchi^[28], Chimaera^[28], SiScan^[29]. 为了提高检测结果的可信度, 将 P 值设为 0.01, Bootscan 方法中的自举值设为 1000, 其他参数采用软件默认参数, 且将至少有 4 种方法支持的重组事件视为有效的潜在重组事件. 为进一步研究潜在重组区域是否影响 CSFV 遗传多样性, 采用 MEGA 5.0 基于最大相似法分别对假定重组毒株的潜在重组区域及非重组区域构建进化树, 自举值为 1000, 验证进化树可信度.

2 结果

本研究基于 CSFV 各个蛋白质构建的进化树树形不尽相同(结果未列), 各个蛋白质的进化距离见图 1. 结果显示, NS3, NS4A 和 NS4B 蛋白的进化距离相对较短, 表明这几个病毒蛋白编码区进化程度较低.

2.1 选择压力分析

CSFV 的 12 个蛋白质的选择压力分析与似然比检验结果见表 2. 在模型 M0 下 12 个蛋白质的 ω 值均小于 1, 表明 CSFV 各个蛋白质均主要承受净化选择(purifying selection)压力. 从表 2 的结果中不难发现各个蛋白质的 ω 值有差异, 如 E2 蛋白的 ω 值最高, 而 NS3 蛋白的 ω 值最低, 说明各个蛋白质所承受的净化选择强度明显不同. 本研究使用似然比检验评估了嵌套模型 M0 vs M3, M1 vs M2, M7 vs M8, 结果显示, 模型 M3 应用于 Npro, E^{ns}, E1, E2, NS2, NS3, NS4B, NS5A 和 NS5B 蛋白显著优于模型 M0; 模型

表1 本研究所用到的50株CSFV病毒株

分离地	毒株名称	年份	基因型	登录号	碱基频率(%)				基因组总碱基数
					G	A	T	C	
中国	HCLV/99	1999	1.1	AF091507	26.4	30.8	22.2	20.7	12310
中国	C/04	2004	1.1	AY663656	26.3	30.8	22.2	20.7	12310
中国	HCLV/08	2008	1.1	EU857642	26.4	30.8	22.2	20.6	12311
中国	SWH	2006	1.1	DQ127910	26.0	31.4	22.1	20.6	12296
中国	Shimen	1999	1.1	AF092448	26.1	31.2	22.0	20.7	12298
中国	cF114	2001	1.1	AF333000	26.0	31.3	22.0	20.6	12297
中国	C/03	2003	1.1	AY382481	26.3	30.8	22.1	20.7	12310
中国	Thiverval	2008	1.1	U490425	26.0	31.2	22.2	20.6	12321
中国	Shimen/HVRI	2007	1.1	AY775178	26.0	31.3	22.1	20.6	12297
中国	C/HVRI	2006	1.1	AY805221	26.3	30.9	22.1	20.7	12310
中国	JL1	2006	1.1	EU497410	26.0	31.3	22.0	20.6	12298
中国	HCLV	2002	1.1	AF531433	26.3	30.9	22.1	20.7	12310
中国	C-ZJ	2008	1.1	HM175885	26.3	30.9	22.1	20.7	12311
中国	CSFV-GZ-2009	2009	1.1	HQ380231	26.0	31.3	22.1	20.6	12298
中国	LPC	2001	1.1	AF352565	26.2	30.8	22.4	20.7	12344
中国	GXWZ02	2002	2.1	AY367767	26.2	31.0	21.7	21.1	12296
中国	SXYL2006	2006	2.1	GQ122383	26.3	30.9	21.5	21.3	12295
中国	Zj0801	2009	2.1	FJ529205	26.4	30.8	21.6	21.1	12296
中国	HEBZ	2010	2.1	GU592790	26.4	31.0	21.5	21.2	12296
中国	SXCDK	2009	2.1	GQ923951	26.4	30.9	21.6	21.0	12296
中国	96TD	2004	2.1	AY554397	26.5	30.8	21.5	21.2	12296
中国	0406/CH	2001	2.1	AY568569	26.3	31.0	21.5	21.3	12296
中国	39	2001	2.2	AF407339	26.1	31.2	21.6	21.2	12297
中国	94.4/IL/94	1994	3.4	AY646427	26.3	31.0	21.7	21.0	12296
日本	ALD	1995	1.1	D49532	26.0	31.2	22.1	20.6	12298
日本	GPE	1999	1.1	D49533	26.0	31.3	22.0	20.7	12298
韩国	LOM	2008	1.1	EU789580	26.1	31.2	22.1	20.6	12298
韩国	flc-LOM	2008	1.1	EU915211	26.1	31.2	22.1	20.6	12298
瑞士	Brescia/98	1998	1.1	AF091661	26.0	31.3	21.9	20.8	12297
瑞士	Riems	2003	1.1	AY259122	26.3	30.9	22.1	20.7	12289
瑞士	Alfort/187	1995	1.1	X87939	26.1	31.2	22.1	20.7	12298
瑞士	CAP	2005	1.1	X96550	26.0	31.3	22.0	20.7	12297
瑞士	Eysrup	2003	1.1	AF326963	25.9	31.3	22.0	20.7	12301
德国	Glentorf	1996	1.1	U45478	26.3	30.9	22.1	20.7	12278
德国	CSF0382	2010	1.1	HM237795	26.0	31.3	22.1	20.6	12298
德国	Riems	1996	1.1	U45477	25.9	31.4	22.0	20.7	12298
德国	CSF857	2006	2.3	GU233731	26.3	30.8	21.9	21.0	12297
德国	XXX0608	2005	2.3	GU233732	26.3	30.8	21.8	21.1	12297
德国	CSF1046	2009	2.3	GU233733	26.4	30.7	21.9	21.0	12297
德国	CSF1045	2009	2.3	GU233734	26.4	30.8	21.9	21.0	12297
德国	XXX0609	2004	2.3	GU324242	26.3	30.9	21.7	21.0	12297
德国	Alfort	2005	2.3	J04358	26.3	30.9	21.7	21.2	12297
俄国	CS	2000	1.2	AF099102	26.0	31.3	21.9	20.8	12310
丹麦	Paderborn/02	2002	2.1	AY072924	26.4	30.9	21.6	21.2	12229
丹麦	Paderborn/10	2010	2.1	GQ902941	26.4	30.8	21.6	21.2	12296
美国	BRESCIAX	2004	1.2	AY578687	25.9	31.4	21.9	20.8	12285
美国	RUCSFPLUM	2004	1.2	AY578688	26.0	31.2	21.9	20.9	12308
西班牙	Sp01	2008	2.2	FJ265020	26.3	30.8	21.8	21.1	12299
荷兰	Brescia/05	2005	1.2	M31768	25.9	31.3	21.9	20.8	12283
法国	Alfort A19	1997	1.1	U90951	26.1	31.2	22.0	20.7	12298

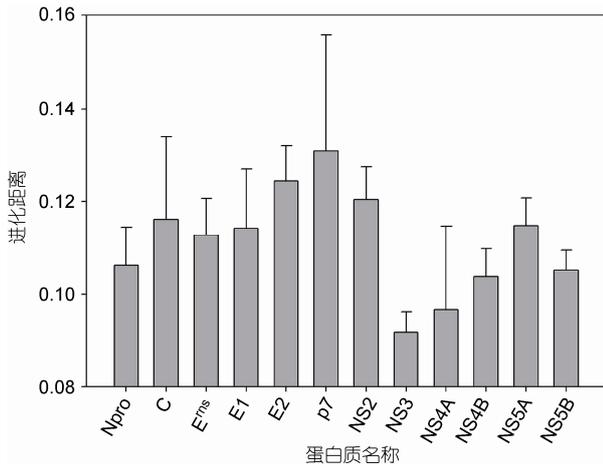


图1 CSFV 各蛋白质进化距离

M8 应用于 Npro, E^{ms}, E1, E2 和 NS5A 蛋白显著优于 M7; 而模型 M1 与模型 M2 的比较中, M2 并不优于 M1. 这说明 Npro, E^{ms}, E1, E2, NS2, NS3, NS4B, NS5A 和 NS5B 蛋白存在正选择压力. 经贝叶斯推论计算后验概率, 在 Npro, E^{ms}, E1, E2 和 NS5A 蛋白中

检测到正选择位点, 分别为 Npro (aa27, aa107), E^{ms} (aa55, aa107, aa119, aa208), E1 (aa80), E2 (aa34, aa72, aa90, aa192, aa197, aa200)和 NS5A (aa163) (表 2).

2.2 重组分析

应用软件 CLUSTAL X1.83 和 RDP 3.44 对 50 条 CSFV 的编码区全基因组序列分别进行对位排列和重组分析后共发现 7 个重组事件, 其中 2 个重组事件仅有 1 种方法支持, 所以将其剔除. 由此得到 5 个潜在重组事件, 并预测出 4 个发生重组的病毒株: HCLV/99, 39, ALD 和 SWH (表 3). 为了进一步确定潜在重组区域对 CSFV 遗传多样性的影响, 采用 MEGA 5.0 基于最大相似法分别对预测出的重组病毒株的重组区域及非重组区域构建了进化树(图 2). 从图 2 可以看出, 在基于 HCLV/99 非重组区域构建的进化树中, HCLV/99 与 C/HVRI 属于同一分支; 而在基于重组区域构建的进化树中, HCLV/99 却与 Shimen 属于同一分支. 其他重组株的情况亦如此, 由此可见, 重组的

表 2 检测 CSFV 12 个蛋白质编码区序列的选择压力^{a)}

蛋白质	M0 (ω)	似然比检验(2 Δ lnL) ^{a)}			正选择位点(aa) ^{b)}
		M0 vs M3	M1 vs M2	M7 vs M8	
Npro	0.13973	74.270168**	1.087760	9.618292**	27; 107
C	0.10332	5.547734	0	0.001976	未检测到
E ^{ms}	0.11602	99.152688**	0	8.058024*	55; 107; 119; 208
E1	0.11060	69.573916**	0.694538	6.108360*	80
E2	0.15441	283.757308**	0	6.717924*	34; 72; 90; 192; 197; 200
p7	0.09531	2.654936	0.441300	1.198870	未检测到
NS2	0.09681	131.206348**	0	0.000132	未检测到
NS3	0.03148	19.266646**	0.000158	0.040464	未检测到
NS4A	0.03442	0.153904	0.000960	0.000960	未检测到
NS4B	0.07491	62.693810**	0	2.788948	未检测到
NS5A	0.11873	143.351246**	0.000006	7.810782*	163
NS5B	0.08920	125.702082**	0.000156	0.282828	未检测到

a) a, 用于 LRT 的嵌套模型 M0 vs M3, M1 vs M2 和 M7 vs M8 检测自由度分别为 4, 2 和 2; b, 检测所得位点为氨基酸位点; *, $P < 0.05$, **, $P < 0.01$ ($\chi^2_{0.05,2}=5.99$; $\chi^2_{0.01,2}=9.21$; $\chi^2_{0.05,4}=9.48$; $\chi^2_{0.01,4}=13.277$)

表 3 CSFV 重组病毒株^{a)}

重组毒株	亲本毒株 ^{a)}	重组区域	重组区域所涉及蛋白质
HCLV/99	C/HVRI (99.6%); Shimen(98.8%)	2110~2526	E2
39	SXCDK (99.6%); CSFV-GZ-2009(99.3%)	1~178; 8150~11694	Npro, C, NS5A, NS5B
ALD	CSF0382 (98.8%); Alfort/187 (99.9%)	738~2024	C, E ^{ms} , E1
ALD	CSF0382 (99.1%); Alfort A19 (99.5%)	2381~3859	E2, p7, NS2
SWH	CSFV-GZ-2009(99.6%); Glentorf (99.3%)	8331~11086	NS5A, NS5B

a) a, 括号中的百分数为重组毒株与亲本毒株重组区域相似性

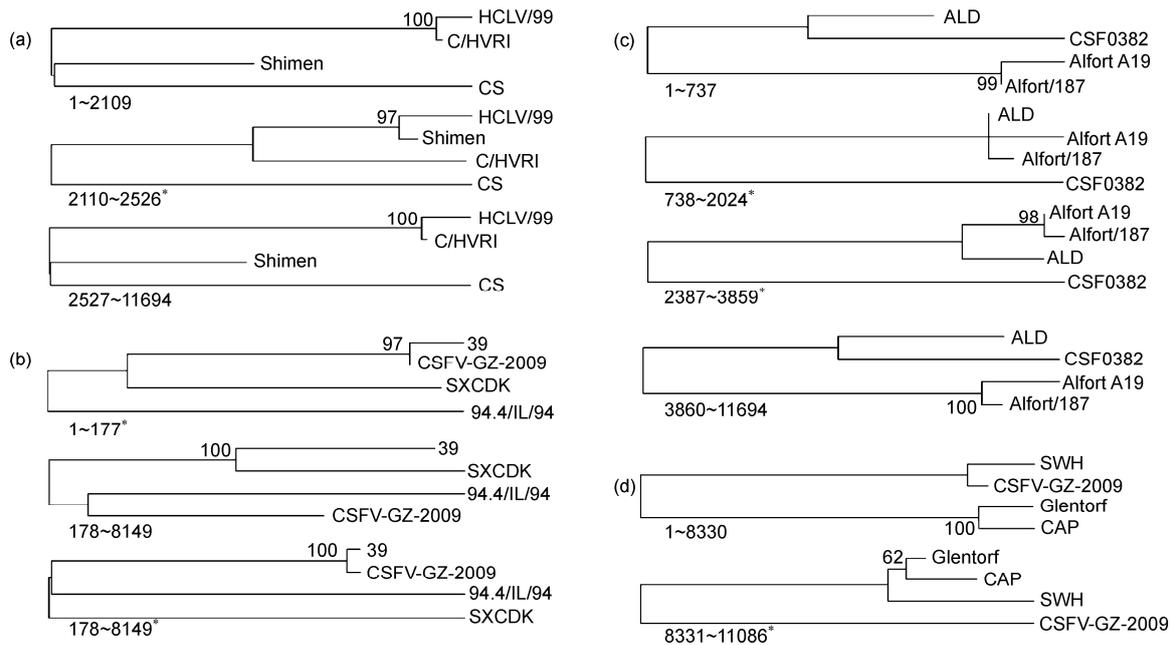


图2 基于重组株与类外群株重组区域及非重组区域的进化树

“*”所示区域为重组区域。(a) HCLV/99, C/HVRI, Shimen 为发生重组株; CS 为类外群株; (b) 39, SXCDK, CSFV-GZ-2009 为发生重组株; 94.4/IL/94 为类群株; (c) ALD, CSF0382, Alfort/187, Alfort A19; (d) SWH, CSFV-GZ-2009, Glentorf 为发生重组株; CAP 为类外群株

确改变了重组株所在的进化树分支,影响了CSFV的遗传多样性。

3 讨论

CSFV 的选择压力分析结果显示, CSFV 各个蛋白质均主要承受净化选择压力。事实上已有研究证明净化选择在 RNA 病毒中比在 DNA 病毒中的作用更加明显^[30]。RNA 病毒的 RNA 聚合酶在复制过程中缺乏纠错功能,导致病毒在宿主体内发生变异,其中大多数是非同义的有害突变^[7,8,30]。RNA 病毒需要一种能抑制有害突变的机制以保证其稳定性,而实际上已有研究表明,净化选择在 RNA 病毒中较大程度上剔除了有害突变^[30]。另外,因为蛋白质功能的重要程度与非同义/同义替换比率呈负相关关系^[31,32],所以 CSFV 各个蛋白质所承受的净化选择强度不同。NS3 蛋白同时具有丝氨酸蛋白酶活性和 RNA 解旋酶活性,在 CSFV 整个病毒周期中起到非常重要的作用^[33],本研究结果也显示其承受较强的净化选择压力。然而 E2 蛋白仅被证明为 CSFV 的主要中和抗原,并不参与病毒的复制过程^[34],其承受的净化选择压力较弱。

本研究的似然比检验结果发现,所检测出的正选择位点均属于变异程度较大的氨基酸位点,在氨

基酸极性性质上的变异并无规律可循,但亲水氨基酸出现频率较大。正选择位点大多位于 CSFV 的囊膜蛋白。位于病毒囊膜蛋白的氨基酸位点与宿主分子的相互作用介导病毒对宿主细胞的感染过程,通常可作为免疫防治的靶标,这些位点往往经历正选择压力^[15]。E^{ms} 蛋白中包含一个由 3 个重叠抗原区组成的显性免疫区域: 抗原区 1 (AR1)-E^{ms}_{aa 65-145}、抗原区 2 (AR2)-E^{ms}_{aa 84-160}、抗原区 3 (AR3)-E^{ms}_{aa 109-145},可诱导体产生中和反应^[35]。本研究中得到的正选择位点 aa107, aa119 和 aa208 恰好位于 E^{ms} 的显性免疫区域内,并且 aa208 位丝氨酸残基突变为精氨酸残基时,可加强病毒与细胞表面的硫酸乙酰肝素(HS)的相互作用^[36],说明这些位点正承受来自宿主的正选择压力。随后,序列比对结果显示, E^{ms} 的 aa55 位的氨基酸残基与 CSFV 基因多样性具有潜在联系: 基因型 I 型以谷氨酰胺为主,基因型 II 型包含精氨酸或谷氨酸残基,基因型 III 型则为亮氨酸。另外, E^{ms} 含有核苷酸酶活性,属于 T2 核苷酸酶家族^[37],在与已经确定晶体结构的 BVDV E^{ms} 序列^[38]的比对定位中发现, aa55, aa107 和 aa119 分别处于 β2/α2, α5/β3, β3/β4 间的无规卷曲区域,并未位于核苷酸酶的保守活性区域,其氨基酸多变性可能是由于病毒与宿

主细胞的相互作用,并未对酶活性产生影响;而 aa208 处于蛋白 C 端的 $\alpha 6$ 上,虽然氨基酸变异程度很大,但在一定程度上受所在结构的两亲性所约束^[39],出现的氨基酸均为亲水性氨基酸.在 E2 蛋白上检测到 6 个正选择位点(表 2),其中 2 个(aa72, aa200)与先前研究结果相一致,这 2 个正选择位点与病毒在宿主免疫环境压力下潜在的适应性进化有关^[15-17],并且直接关系到强病毒株 Brescia 的毒力衰减^[40].其余 4 个新发现的正选择位点均位于 E2 蛋白的 N 端,恰好位于 E2 蛋白的 4 个重叠抗原区: A, B, C 和 D^[41].随后,根据已经确定三级结构的 BVDV E2 序列^[42]对 6 个位点进行定位研究, BVDV E2 的分子结构共分为 4 个单元域 DA(aa4~aa87), DB (aa88~aa164), DC (aa165~271)和 DD (aa272~aa333),其中单元域 DA 和 DB 暴露于细胞表面,DC 功能不明,DD 为最保守单元域.结果发现,正选择位点 aa34 和 aa72 位于 DA, aa90 位于 DB, aa192, aa197 和 aa200 位于 DC.有报道称 E2 蛋白的 N 端暴露在病毒颗粒表面,与 E1 蛋白形成 E1-E2 二聚体,在诱导病毒进入宿主细胞过程中起到关键作用^[34],这也是病毒蛋白承受正选择压力的一个证据.在 CSFV 主要抗原蛋白上所检测到的这些正选择位点在一定程度上可以成为 CSFV 适应性进化的证据,作为病毒逃避宿主免疫应答的作用靶标.这些正选择位点的分析对识别免疫逃避、抗药性突变、防治病毒性疾病有积极作用^[13,43].

在本研究中, Npro 蛋白作为黄病毒科瘟病毒属成员的特有蛋白,具有相对较高的 ω 值,存在 2 个正选择位点(aa27, aa107) (表 2). Npro 蛋白具有蛋白水解酶活性,裂解 Cys168 与 Ser169 之间的连键^[44],不参与病毒复制过程^[45]. Ruggli 等人^[46]证实 Npro 蛋白的另一个功能与抑制宿主细胞固有免疫有关,可通过抑制 IRF3 活化阻碍干扰素产生,从而抵抗宿主对病毒感染免疫应答^[47].本研究在 NS5A 蛋白上检测到 1 个正选择位点(aa163). Xiao 等人^[48]通过丙氨酸扫描突变研究表明, NS5A 蛋白的突变可直接影响由内部核糖体进入位点(internal ribosome entry site, IRES)所介导的 CSFV 翻译,对病毒翻译与复制之间的转换具有十分重要的作用.此外,另一研究表明,与 CSFV 同属黄病毒科的丙型肝炎病毒(hepatitis C virus, HCV)的 NS5A 蛋白为病毒 RNA 复制的重要元件,直接或间接与宿主细胞相互作用^[49],存在适应性进化^[50],而且这种适应性进化是宿主免疫系统作用直接驱动

的结果.

从表 2 中不难发现,尽管在 CSFV 部分蛋白上检测到正选择压力,但这些蛋白质检测到的正选择压力存在差异,而且还有少数蛋白质,如 C 蛋白、p7 蛋白和 NS4A 蛋白,并没有检测到正选择压力,表明 CSFV 的各个蛋白质所承受的选择压力可能不同. Ruan 等人^[51]在研究 SARS 病毒时也发现功能不同的蛋白质在进化过程中承受了不同强度的选择压力,与本实验室研究结果一致. CSFV 编码区各个蛋白质存在选择压力的差异可能与病毒元件和宿主分子相互作用有关,其进化过程受宿主固有免疫系统和病毒自身的免疫逃避机制的共同影响,通过连续的突变,与宿主形成一个连续的协同进化过程^[52].

对于 CSFV 的重组情况,先前 He 等人^[23]使用 Bootscan 法证实病毒株 39 是重组病毒株,其亲本株为 GXWZ02 和 cF114.本研究采用 6 种方法检测重组事件,并将至少有 4 种方法支持的重组事件视为有效的潜在重组事件,检测出 4 株重组病毒株,其中包括病毒株 39,但其来自亲本株 SXCDK 和 CSFV-GZ-2009,重组区域涉及 Npro, C, NS5A 和 NS5B 蛋白的编码区.分别对病毒株 39 亲本株的核苷酸序列进行比较后,发现 GXWZ02 和 SXCDK 同属于基因型 2.1,其重组区域的相似性都在 93.3%以上; cF114 和 CSFV-GZ-2009 同属于基因型 1.1,其重组区域的相似性都在 99.4%以上.由此看出,本研究得到的亲本株与 He 等人^[23]推测的亲本株实际上是相似性很高的病毒株.事实上,毒株 cF114 自 2001 年被首次分离以来很少再次出现,推测毒株 cF114 已逐渐被其他形式的强毒株(如 CSFV-GZ-2009)所取代.值得注意的是,重组病毒株 39 的重组区域所涉及的蛋白中, Npro 为毒力相关蛋白^[46],且 NS5A 和 NS5B 已经被证实为病毒复制所必需^[53],这些蛋白质都与病毒毒力有密切关系;还发现病毒株 39 与 SXCDK 的相似性(99.6%)高于 39 与 CSFV-GZ-2009 的相似性(99.3%),且 39 与 SXCDK 同属中等毒力毒株,表明 CSFV 在这里的重组可能是病毒毒力发生改变的原因之一.此外,重组病毒株 SWH 的亲本株为 CSFV-GZ-2009 和 Glentorf,与 39 享有一个相同的亲本病毒株 CSFV-GZ-2009,重组区域也涉及与病毒毒力密切相关的 NS5A, NS5B 蛋白的编码区(表 3),进一步说明了重组影响了 CSFV 病毒毒力.

重组株 ALD 重组发生于 C-E1 区域,亲本株是

CSFV0382 和 Alfort/187; 重组发生于 E2-NS2 区域, 亲本株是 CSFV0382 和 Alfort A19. 本研究发现在 C-E1 的重组区域, ALD 与 Alfort/187 的相似性(99.9%) 高于 ALD 与 CSFV0382 的相似性(98.8%); 在 E2-NS2 重组区域, ALD 与 Alfort A19 的相似性(99.5%)高于 ALD 与 CSFV0382 的相似性(99.1%). 已有研究称以 ALD 为亲本株在宿主体内传代培养得到的减毒株 GPE^[54], 与高毒株 Alfort/187 和 Alfort A19 在序列相似上更为相近^[55]. 这些都说明病毒株 ALD 和 Alfort/187, Alfort A19 之间可能确实存在基因重组, 其重组区域涉及几个主要的毒力蛋白 E^{ms}, E1 和 E2, 也说明重组在 CSFV 病毒毒力变异中起到重要作用.

从表 3 可以看出, HCLV/99 可能由亲本毒株 Shimen 和 C/HVRI 重组而来, 重组断裂位点是基因组的第 2110 和 2526 个碱基处(从编码区开始计数), 这 2 个断裂位点刚好落在 E2 蛋白的编码区. 猪瘟疫化弱毒株 HCLV/99 为中国标准减毒活疫苗, 早已被广泛应用于控制猪瘟的传播. 先前的研究指出, 强毒株 Shimen 是猪瘟疫化弱毒株的亲本毒株, 最早在 1945 年分离得到, 经过在兔子体内传代培养, 最终得到可用作疫苗的猪瘟疫化弱毒株(HCLV/99)^[56], 与本研究的预测结果一致. 对重组毒株和亲本毒株的重组区域进行了相似性分析, 发现在 HCLV/99 与 C/HVRI 重组区域内的相似性为 99.6%, 甚至高于 HCLV/99 与 Shimen 的相似性(98.8%), 且 HCLV/99 与 C/HVRI 同为弱毒株. 这些都支持 HCLV/99 可能由亲本毒株 Shimen 和 C/HVRI 重组而来, 又发现重组区域位于

CSFV 的 E2 蛋白的编码区. 已有研究表明, 猪瘟疫化弱毒株与强毒株位于 E2 蛋白上的抗原表位并不相同^[57]. 因此, 推测这种重组可能源于宿主免疫压力.

基于重组病毒株的重组区域及非重组区域构建的进化树结果显示(图 2), 重组能够改变重组毒株所在的进化分支, 所以基于某个单一基因区段来推断 CSFV 的进化过程是不全面的. 对于日渐趋于复杂的 CSFV 进化情况, 重组已成为不可忽略的影响因素. 另外, 进化距离的结果显示(图 1), NS3, NS4A 和 NS4B 蛋白的进化距离相对较短, 表明这几个病毒蛋白编码区进化程度较低. 本研究在重组分析研究结果中仅有 NS3, NS4A 和 NS4B 蛋白未检测到重组事件. 此外, 选择压力分析结果显示 NS3, NS4A 和 NS4B 蛋白所承受的净化选择压力较大, 正选择压力较小. 由此可以看出, CSFV 针对宿主免疫系统选择压力而产生的突变和重组对病毒编码区基因进化有较大的影响, 在一定程度上增加了 CSFV 各个蛋白质进化的复杂程度.

综上所述, 本研究对 50 株 CSFV 编码区全基因组序列选择压力和重组分析的结果证实宿主选择压力及重组均为 CSFV 进化的主要影响因素. 选择压力共检测出 14 个正选择位点, 分布于 Npro, E^{ms}, E1, E2 和 NS5A 蛋白, 都与宿主免疫应答有直接关系. 另外, 重组分析的结果显示, 重组事件涉及病毒的遗传变异及毒力改变, 且在一定程度上使 CSFV 的进化程度日渐趋于复杂. 本研究实验结果为今后 CSFV 疫苗靶标和流行病学研究提供了一定理论依据.

参考文献

- 1 Paton D J, Greiser-Wilke I. Classical swine fever—An update. *Res Vet Sci*, 2003, 75: 169–178
- 2 Rice C. Flaviviridae: The viruses and their replication. *Fields Virol*, 1996, 3: 931–959
- 3 Moennig V, Plagemann P G W. The pestiviruses. *Adv Virus Res*, 1992, 41: 53–98
- 4 Paton D J, McGoldrick A, Greiser-Wilke I, et al. Genetic typing of classical swine fever virus. *Vet Microbiol*, 2000, 73: 137–157
- 5 Barr J N, Fearn R. How RNA viruses maintain their genome integrity. *J General Virol*, 2010, 91: 1373–1387
- 6 Manrubia S C, Escarmis C, Domingo E, et al. High mutation rates, bottlenecks, and robustness of RNA viral quasispecies. *Gene*, 2005, 347: 273–282
- 7 Weiss R A. Virulence and pathogenesis. *Trends Microbiol*, 2002, 10: 314–317
- 8 Reaney D C. The evolution of RNA viruses. *Annu Rev Microbiol*, 1982, 36: 47–73
- 9 Yang Z, Bielawski J P. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*, 2000, 15: 496–503
- 10 O'Connor D H, McDermott A B, Krebs K C, et al. A dominant role for CD8⁺-T-lymphocyte selection in simian immunodeficiency virus sequence variation. *J Virol*, 2004, 78: 14012–14022
- 11 Suzuki Y, Gojobori T. Positively selected amino acid sites in the entire coding region of hepatitis C virus subtype 1b. *Gene*, 2001, 276: 83–87
- 12 Suzuki Y. Natural selection on the influenza virus genome. *Mol Biol Evol*, 2006, 23: 1902–1911

- 13 Seoghe C, Ketwaroo F, Pillay V, et al. A model of directional selection applied to the evolution of drug resistance in HIV-1. *Mol Biol Evol*, 2007, 24: 1025–1031
- 14 Chen X, Zhang Q, Li J, et al. Analysis of recombination and natural selection in human enterovirus 71. *Virology*, 2010, 398: 251–261
- 15 Tang F, Pan Z, Zhang C. The selection pressure analysis of classical swine fever virus envelope protein genes E^{ms} and E2. *Virus Res*, 2008, 131: 132–135
- 16 Shen H, Pei J, Bai J, et al. Genetic diversity and positive selection analysis of classical swine fever virus isolates in South China. *Virus Genes*, 2011, 43: 234–242
- 17 Perez L J, Diaz de Arce H, Perera C L, et al. Positive selection pressure on the B/C domains of the E2-gene of classical swine fever virus in endemic areas under C-strain vaccination. *Infect Genet Evol*, 2012, 12: 1405–1412
- 18 叶琳, 张原, 梅旖, 等. 应用比较基因组分析检测乙型肝炎病毒中可能的重组事件. *科学通报*, 2010, 55: 1322–1328
- 19 Becher P, Orlich M, Thiel H J. RNA recombination between persisting pestivirus and a vaccine strain: Generation of cytopathogenic virus and induction of lethal disease. *J Virol*, 2001, 75: 6256–6264
- 20 Colina R, Casane D, Vasquez S, et al. Evidence of intratypic recombination in natural populations of hepatitis C virus. *J Gen Virol*, 2004, 85: 31–37
- 21 Cristina J, Colina R. Evidence of structural genomic region recombination in Hepatitis C virus. *Virol J*, 2006, 3: 53
- 22 Holmes E C, Worobey M, Rambaut A. Phylogenetic evidence for recombination in dengue virus. *Mol Biol Evol*, 1999, 16: 405–409
- 23 He C Q, Ding N Z, Chen J G, et al. Evidence of natural recombination in classical swine fever virus. *Virus Res*, 2007, 126: 179–185
- 24 Jeanmougin F, Thompson J D, Gouy M, et al. Multiple sequence alignment with Clustal X. *Trends Biochem Sci*, 1998, 23: 403–405
- 25 Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*, 2007, 24: 1586–1591
- 26 Padidam M, Sawyer S, Fauquet C M. Possible emergence of new geminiviruses by frequent recombination. *Virology*, 1999, 265: 218–225
- 27 Salminen M O, Carr J K, Burke D S, et al. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retrov*, 1995, 11: 1423–1425
- 28 Posada D, Crandall K A. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc Natl Acad Sci USA*, 2001, 98: 13757–13762
- 29 Gibbs M J, Armstrong J S, Gibbs A J. Sister-scanning: A Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*, 2000, 16: 573–582
- 30 Hughes A L, Hughes M A. More effective purifying selection on RNA viruses than in DNA viruses. *Gene*, 2007, 404: 117–125
- 31 Söderholm J, Ahlen G, Kaul A, et al. Relation between viral fitness and immune escape within the hepatitis C virus protease. *Gut*, 2006, 55: 266–274
- 32 Neumann-Haefelin C, Killinger T, Timm J, et al. Absence of viral escape within a frequently recognized HLA-A26-restricted CD8⁺ T-cell epitope targeting the functionally constrained hepatitis C virus NS5A/5B cleavage site. *J Gen Virol*, 2007, 88: 1986–1991
- 33 Wen G, Chen C, Luo X, et al. Identification and characterization of the NTPase activity of classical swine fever virus (CSFV) nonstructural protein 3 (NS3) expressed in bacteria. *Arch Virol*, 2007, 152: 1565–1573
- 34 Wang Z, Nie Y, Wang P, et al. Characterization of classical swine fever virus entry by using pseudotyped viruses: E1 and E2 are sufficient to mediate viral entry. *Virology*, 2004, 330: 332–341
- 35 Lin M, Trottier E, Pasick J, et al. Identification of antigenic regions of the Erns protein for pig antibodies elicited during classical swine fever virus infection. *J Biochem*, 2004, 136: 795–804
- 36 Hulst M M, van Gennip H G, Moormann R J. Passage of classical swine fever virus in cultured swine kidney cells selects virus variants that bind to heparan sulfate due to a single amino acid change in envelope protein E^{ms}. *J Virol*, 2000, 74: 9553–9561
- 37 Hulst M, Panoto F, Hoekman A, et al. Inactivation of the RNase activity of glycoprotein E^{ms} of classical swine fever virus results in a cytopathogenic virus. *J Virol*, 1998, 72: 151
- 38 Krey T, Bontems F, Vonrhein C, et al. Crystal structure of the pestivirus envelope glycoprotein E^{ms} and mechanistic analysis of its ribonuclease activity. *Structure*, 2012, 20: 862–873
- 39 Fetzter C, Tews B A, Meyers G. The carboxy-terminal sequence of the pestivirus glycoprotein E^{ms} represents an unusual type of membrane anchor. *J Virol*, 2005, 79: 11901–11913
- 40 Risatti G R, Holinka L G, Fernandez Sainz I, et al. N-linked glycosylation status of classical swine fever virus strain Brescia E2 glycoprotein influences virulence in swine. *J Virol*, 2007, 81: 924–933
- 41 van Rijn P A, Miedema G K, Wensvoort G, et al. Antigenic structure of envelope glycoprotein E1 of hog cholera virus. *J Virol*, 1994, 68: 3934–3942
- 42 El Omari K, Iourin O, Harlos K, et al. Structure of a pestivirus envelope glycoprotein E2 clarifies its role in cell entry. *Cell Rep*, 2013, 3: 30–35
- 43 Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, 1998, 148: 929–936

- 44 Rüménapf T, Stark R, Heimann M, et al. N-terminal protease of pestiviruses: Identification of putative catalytic residues by site-directed mutagenesis. *J Virol*, 1998, 72: 2544–2547
- 45 Mayer D, Hofmann M A, Tratschin J D. Attenuation of classical swine fever virus by deletion of the viral N^{pro} gene. *Vaccine*, 2004, 22: 317–328
- 46 Ruggli N, Tratschin J D, Schweizer M, et al. Classical swine fever virus interferes with cellular antiviral defense: Evidence for a novel function of N^{pro}. *J Virol*, 2003, 77: 7645–7654
- 47 La Rocca S A, Herbert R J, Crooke H, et al. Loss of interferon regulatory factor 3 in cells infected with classical swine fever virus involves the N-terminal protease, N^{pro}. *J Virol*, 2005, 79: 7239–7247
- 48 Xiao M, Wang Y, Zhu Z, et al. Influence of NS5A protein of classical swine fever virus (CSFV) on CSFV internal ribosome entry site-dependent translation. *J Gen Virol*, 2009, 90: 2923–2928
- 49 Tellinghuisen T L, Marcotrigiano J, Gorbalenya A E, et al. The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J Biol Chem*, 2004, 279: 48576–48587
- 50 Brass V, Bieck E, Montserret R, et al. An amino-terminal amphipathic alpha-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. *J Biol Chem*, 2002, 277: 8130–8139
- 51 Ruan Y J, Wei C L, Ee A L, et al. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet*, 2003, 361: 1779–1785
- 52 Lobo F P, Mota B E, Pena S D, et al. Virus-host coevolution: Common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS One*, 2009, 4: e6282
- 53 Chen Y, Xiao J, Xiao J, et al. Classical swine fever virus NS5A regulates viral RNA replication through binding to NS5B and 3'UTR. *Virology*, 2012, 432: 376–388
- 54 Ishikawa K, Nagai H, Katayama K, et al. Comparison of the entire nucleotide and deduced amino acid sequences of the attenuated hog cholera vaccine strain GPE- and the wild-type parental strain ALD. *Arch Virol*, 1995, 140: 1385–1391
- 55 Cao S, Zhang C Y, Yi G H. Recombinant analysis of classical swine fever virus. *J Genet Genomics*, 2005, 32: 52–56
- 56 Wu H, Wang J, Zhang C, et al. Attenuated lapinized Chinese strain of classical swine fever virus: Complete nucleotide sequence and character of 3'-noncoding region. *Virus Genes*, 2001, 23: 69–76
- 57 Edwards S, Sands J J. Antigenic comparisons of hog cholera virus isolates from Europe, America and Asia using monoclonal antibodies. *Dtsch Tierarztl Wochenschr*, 1990, 97: 79–81

Analysis of selection pressure and recombination in coding-region of classical swine fever virus

JIA Lin, YAO YuTing & XIAO Ming

Biology Department, College of Life and Environment Sciences, Shanghai Normal University, Shanghai 200234, China

We detected the selection pressure and recombination of all 50 strains Classical Swine Fever Virus (CSFV) available complete coding region, based on the maximum likelihood estimation and six algorithms within Recombinant Detect Program 3.44, to provide the first complete overview of the evolution of CSFV. The results showed the overall ω (nonsynonymous/synonymous) of 12 proteins under one-radio model were all lower than 1, indicating that the selection pressure of CSFV was subject to strong purifying selection. Among the 12 proteins, NS3, NS4A and NS4B proteins underwent the stronger purifying selection; meanwhile N^{pro}, E^{1ns}, E1, E2 and NS5A proteins were detected the positive selection sites, underwent the stonger positive selection, indicating that the adaptive evolution of CSFV were related to the host immune response. Furthermore, 4 putative recombination virus strains, and the newly found HCLV/99, ALD and SWH, were detected in 50 sample sequences by recombination analysis. The proteins related to the genome recombination were involved in the immunity and virulence of CSFV, and the recombination changed the phylogenetic tree branch of recombination strain. However, the recombination loci were detected in most proteins except NS3, NS4A and NS4B proteins which processed the shorter genetic distance. Our research suggested that the mutation and recombination under the selection pressure of the host immune system promoted the evolution of coding region in CSFV. These studies would provide the foundation for vaccine targets and the more and more complex situation of CSFV epidemiological research.

classical swine fever virus, coding region, adaptive evolution, recombination

doi: 10.1360/972013-84