

doi:10.3969/j.issn.1672-5565.2013.01.07

人腺病毒 B、D 亚种基因组中简单重复序列分析

田 优, 谭钟扬*, 赵相艳, 牛一乐, 冯海平, 欧阳清检

(湖南大学, 湖南 长沙 410082)

摘要:本文以人腺病毒 B 亚种 31 条基因组序列及 D 亚种 39 条基因组序列为研究材料, 利用 Imperfect Microsatellite Extractor 和 DNAMAN 软件对这些基因组序列中简单重复序列 (SSR) 的分布情况进行了系统性分析和比较。分析结果显示: 人腺病毒 B、D 亚种基因组中简单重复序列的平均相对密度是十分接近的, 但在不同类型 SSR 中分布情况又有所不同。D 亚种中二型 SSR 明显高于 B 亚种, 在两亚种一型 SSR 中 (A)_n、(T)_n 都是比较多的, 而在两亚种二型 SSR 中的 (CG/GC)_n 表现出了较高的偏好性。在同亚种多序列比对分析中, D 亚种表现出了更高的稳定性。B、D 亚种中 SSR 的这种特异性分布可能与它们的进化机制和致病性有关。

关键词:人腺病毒, 亚种, 简单重复序列

中图分类号: Q343.1+5 文献标识码: C 文章编号: 1672-5565(2013)-01-039-05

Analyses of simple sequence repeat in genomes of Human adenoviruses B and D subspecies

TIAN You, TAN Zhong-yang*, ZHAO Xiang-yan, NIU Yi-le, FENG Hai-ping, OUYANG Qing-jian

(Hunan University, Hunan Changsha 410082, China)

Abstract: In our study, the simple sequence repeat in 31 genomes of human adenoviruses B and 39 genomes of human adenoviruses D were analyzed by the software Imperfect Microsatellite Extractor and DNAMAN. The results showed that in human adenoviruses B and D subspecies the average relative density was very similar, but the distribution of different repeat motifs was markedly different. The trinucleotide repeats had higher proportion in the D subspecies. In mononucleotide repeats the (A)_n and (T)_n were predominant and the (CG/GC)_n were more in dinucleotide repeats in Human adenoviruses B and D subspecies. In many sequence alignment analysis of same subspecies, the D subspecies showed higher stability. The specific distribution of SSR may be related to the evolution mechanism and different pathogenicity of Human adenoviruses B and D subspecies.

Key words: Human Adenoviruses, Subspecies, Simple Sequence Repeat

腺病毒 (Adenovirus) 是从手术切除的扁桃体组织分离培养得到的一种 DNA 病毒。根据宿主范围不同, 将腺病毒分为哺乳动物腺病毒 (Mastadenovirus) 和禽类腺病毒 (Aviadenovirus) 两个属^[1]。人腺病毒分为 7 个亚种, 即人腺病毒 A、B、C、D、E、F、G^[2], 主要在细胞核内繁殖, 常引起人上呼吸道和眼部上皮细胞感染。腺病毒基因组 DNA 为线状双链, 长约 35kb, 其 DNA 分子各含有 100bp 反向重复序

列。人腺病毒除 B 和 D 亚种外, 其它亚种的受体都是柯萨奇-腺病毒受体。B 亚种中大多数可与胞膜表面免疫调节分子 CD46 相结合, 其余可与相关免疫分子 CD80 或 CD86 结合, 而 D 亚种可与细胞表面普遍存在的唾液酸受体相结合^[3]。不同的腺病毒亚种会引起不同的疾病, 如人腺病毒 C 亚种主要引起小儿科的上呼吸道感染, 人腺病毒 B 亚种常在成年人中引起流行, 人腺病毒 B 和 E 亚种是军营人

收稿日期: 2012-04-26; 修回日期: 2012-06-25.

基金项目: 湖南省科技计划项目 [2008CK13070]; 长沙市科技计划项目 [2011K1113021/11]。

作者简介: 田优, 男, 硕士研究生, 研究方向: 生物化学与分子生物学, 生物信息学。E-mail: tianyoyo922@163.com.

* 通讯作者: 谭钟扬, 男, 副教授, 研究方向: 生物信息学。E-mail: zhongyang@hnu.edu.cn.

员感染的主要病因,人腺病毒 F 亚种会引起感染性腹泻^[4]。

简单重复序列(Short Sequence Repeat, SSR)也称为微卫星(Microsatellite)或短串联重复序列(Short Tandem Repeat, STR),是由 1~6 个碱基对组成的串联重复 DNA 区段,广泛存在于各类生物基因组中并具有较高的突变频率^[5]。SSR 的高度可变性可能是在 DNA 复制、修复或重组过程中链的滑移引起的^[6]。近来,研究发现 SSR 在生物进化中起到非常重要的作用,它能影响生物体中染色质的结构、基因活性的调控、DNA 的重组及错配修复系统等^[7]。SSR 具有很高的可变性、显著的多态性和侧翼序列的相对保守性,一般为共显性,其数量和种类还可能影响翻译活动的水平^[8]。

本研究利用已有的数据库资源,对人腺病毒 B、D 亚种基因组中 SSR 的组成、分布、平均相对丰度及平均相对密度等进行了系统性分析,并把 B、D 亚种相互比较。从而探究 SSR 在人腺病毒 B、D 亚种中的分布规律以及对基因组的影响。

1 材料和方法

1.1 序列下载

在 NCBI (<http://www.ncbi.nlm.nih.gov>) 数据库^[9]中检索人腺病毒全序列,经筛选共有 98 条数据,其中 B 亚种 31 条数据, D 亚种 39 条数据,以 FAST 格式下载至本地硬盘,以作进一步分析。表 1 列出了待测序列的相关信息。

表 1 人腺病毒 B、D 亚种基因组序列的长度及登记号

Table 1 Lengths and accessions of the genome in Human adenoviruses B and D subspecies

No.	Acc. No	Length (bp)	No.	Acc. No	Length (bp)
B - S1	GQ478341	35 306	D - S5	JF799911	35 072
B - S2	HQ659699	32 442	D - S6	AB605246	35 116
B - S3	AY594256	35 240	D - S7	AB605244	35 137
B - S4	AY495969	35 514	D - S8	AB605243	35 137
B - S5	FJ643676	34 755	D - S9	AB605242	35 118
B - S6	DQ099432	35 273	D - S10	AB605241	35 045
B - S7	AF532578	34 794	D - S11	AB605240	35 048
B - S8	FJ597732	34 773	D - S12	HQ910407	35 139
B - S9	AY601636	35 522	D - S13	AB562588	35 067
B - S10	AY601633	35 382	D - S14	AB562587	35 213
B - S11	AY271307	34 794	D - S15	AB562586	35 116
B - S12	DQ086466	35 343	D - S16	HM770721	35 066
B - S13	AY601634	35 198	D - S17	DQ900900	35 213
B - S14	AY599836	35 265	D - S18	FJ824826	35 130
B - S15	AY599834	35 345	D - S19	GQ384080	35 152
B - S16	AY737798	35 385	D - S20	AF108105	35 100
B - S17	AY737797	34 775	D - S21	FJ619037	35 152
B - S18	AY803294	34 764	D - S22	FJ404771	35 166
B - S19	AY598970	34 794	D - S23	EF153474	35 152
B - S20	AY594255	35 306	D - S24	EF153473	35 206
B - S21	AY128640	34 794	D - S25	DQ393829	35 215
B - S22	AY163756	34 794	D - S26	AY875648	35 178
B - S23	DQ105654	35 269	D - S27	FJ169625	34 909
B - S24	EF564601	35 355	D - S28	AB448770	34 920
B - S25	EF564600	35 130	D - S29	AB448778	35 152
B - S26	EF011630	32 327	D - S30	AB448777	35 215
B - S27	FJ822614	34 763	D - S31	AB448774	35 231
B - S28	NC_011202	34 794	D - S32	AB448771	35 125
B - S29	NC_011203	35 343	D - S33	AB448769	34 980
B - S30	AC_000019	34 794	D - S34	AB448768	34 962
B - S31	AC_000018	35 514	D - S35	AB448767	34 980
D - S1	JN162672	35 103	D - S36	NC_010956	35 083
D - S2	JN162671	35 127	D - S37	AC_000006	35 100
D - S3	HQ007053	35 050	D - S38	EF121005	35 231
D - S4	NC_012959	34 920	D - S39	AJ854486	35 083

例如:表中 B - S18 表示 AY803294 属于人腺病毒 B 亚种,且在该文本中为 B 亚种第 18 条序列。

1.2 序列分析

待测序列中的 SSR 分析采用 IMEx^[10] (<http://210.212.215.200/IMEX/index.html>) 在线软件进行,生成原始的数据统计结果。基于待测序列的实际情况,分析时要求 SSR 为完全重复类型,单碱基最小重复单元为 6,二碱基、三碱基、四碱基、五碱基、六碱基最小重复单元均为 3。用 Microsoft Excel 进一步的统计分析待测序列中 SSR 的结构组成、平均相对丰度、平均相对密度、最普遍 SSR 等。最后采用 DNAMAN (versions 6.0) 专用软件进行辅助性分析以及同一亚种内的多序列比对分析。

2 结果与分析

2.1 待测序列碱基组成

人腺病毒基因组大小约为 35kb,在待测序列中,人腺病毒 B 亚种中 A、C、G、T 四种碱基含量均约占 25%,而人腺病毒 D 亚种中 C、G 约占 28%,A、T 约占 22%。B、D 亚种中 GC 含量有着明显的差异(图 1)。相似大小的基因组中 GC 含量的差异可能表明,D 亚种具有更高的稳定性,在物种的进化中更不易突变。

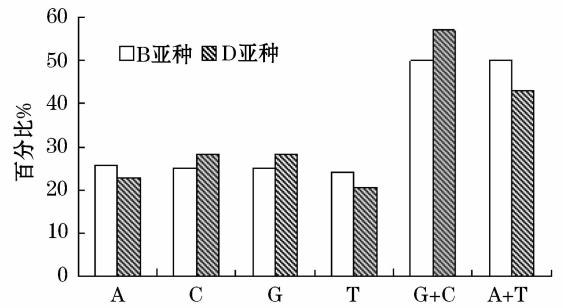


图 1 人腺病毒 B、D 亚种碱基组成分布

Fig. 1 Distribution of base composition in Human adenoviruses B and D subspecies

2.2 各型 SSR 均值比较

根据设定条件,利用 IMEx 分析待测序列。发现在人腺病毒 B、D 亚种中四型、五型、六型中 SSR 个数极少可忽略为零(表 2)。从表 2 中看出,B、D 亚种中平均 SSR 值是一样的,为 162 个。在 B、D 亚种中三型 SSR 所占比例几乎一致,而在 B 亚种中一型和二型比例接近 1,在 D 亚种中二型是一型的两倍。B、D 亚种中一型、二型含量差异明显。

表 2 人腺病毒 B、D 亚种中 SSR 总均值及各型均值

Table 2 The overall average and each average of SSR in Human adenoviruses B and D subspecies

	总计(个)	SSR 重复类型					
		一型	二型	三型	四型	五型	六型
B 亚种	162	60(37%)	69(43%)	33(20%)	0	0	0
D 亚种	162	41(26%)	83(51%)	38(23%)	0	0	0

2.3 平均相对丰度及平均相对密度

SSR 的相对丰度是指单位长度序列中相应类型 SSR 出现的次数,平均相对丰度是指相应类型 SSR 的相对丰度的平均值,单位为 SSR/kb。SSR 的相对密度是指单位长度序列中相应 SSR 的碱基数,平均

相对密度是相应 SSR 的相对密度的平均值,单位为 bp/kb^[11](表 3)。表 3 中数据显示,两亚种中平均相对密度十分接近,约为 32。在各型 SSR 平均相对丰度的分布中表现了与各型 SSR 均值分布一样的规律。

表 3 人腺病毒 B、D 亚种中 SSR 的平均相对丰度及平均相对密度

Table 3 Average relative abundance and average relative density of SSR in Human adenoviruses B and D subspecies

	平均相对丰度					平均相对密度一型
	二型	三型	四型	五型	六型	
B 亚种	1.71	1.99	0.95	0	0	32.63
D 亚种	1.18	2.29	1.05	0	0	31.84

2.4 最普遍的 SSR

基因组中最普遍的 SSR 是指在这个基因组序列中出现次数最多的简单重复序列,研究表明不同的生物体有各自最普遍的 SSR,在哺乳动物和无脊椎动物中为(GT)_n^[12],在植物中为(AT)_n^[13]而在昆虫

中为(CT)_n^[14]。研究中,每一个具体的 SSR 类型代表所有循环排列的 SSR,例如:(AG)_n 代表简单重复序列(AG)_n 和(GA)_n。在人腺病毒 B、D 亚种中一型 SSR 的 A、C、G、T 四种碱基中,A 碱基重复和 T 碱基重复在一型 SSR 中占大部分比例,在二型 SSR 中,

所有的二型重复中 CG/GC 占有极大比例(表 4)。

表 4 人腺病毒 B、D 亚种中一、二型 SSR 中最普遍的 SSR

Table 4 The most common SSR motifs inside Mono - and Di - in Human adenoviruses B and D subspecies

	一型 SSR					二型 SSR				
	A	T	A + T	总 SSR	(A + T) %	CG	GC	CG + GC	总 SSR	(CG + GC) %
B 亚种	25	21	46	60	76%	11	15	26	69	36%
D 亚种	17	9	26	41	63%	24	22	46	83	55%

2.5 同亚种间多序列比对分析

采用 DNAMAN 专用软件对人腺病毒 B 亚种的 31 条序列以及人腺病毒 D 亚种的 39 条序列做同亚种间的多序列比对分析。结果显示 B 亚种 31 条序列的相似性为 67.98%, D 亚种 39 条序列的相似性为 74.12%。在这些保守序列中,由于 SSR 的高度可变性,一型和二型 SSR 较常见,其余类型的 SSR 较少存在。

3 讨论

在 SSR 的研究中,以原核生物和真核生物为对象的研究较多,而以病毒为起点研究 SSR 的相关工作却较少。对于不同的物种来说,相同的是 SSR 广泛的存在整个基因组中,但其不同的 SSR 类型在基因组中的组成、长短以及分布却各具特征,并且在种内保持一定稳定性^[15]。

研究统计的 31 条 B 亚种数据,其长度范围从 32327bp(EF011630) 到 35522bp(AY601636); 39 条 D 亚种数据,其长度范围从 34909bp(FJ169625) 到 35231bp(AB448774) (表 1)。就其基因组长度而言, D 亚种比 B 亚种更为保守,有可能是 B 亚种面临着更大的进化压力,从而产生更多的突变。从图 1 中看出, D 亚种中的 GC 碱基含量明显高于 B 亚种中 GC 碱基的含量,由于 GC 碱基之间的三个氢键的结合力高于 AT 碱基之间两个氢键的结合力,也印证了 D 亚种比 B 亚种在进化上更为保守,更不易突变这一观点。

B 亚种和 D 亚种的总 SSR 的平均数都为 162 个。在各型 SSR 的分布中,四型、五型、六型 SSR 基本为零,相对于真核生物和原核生物这种情况很少见,产生这一现象的原因可能是因为病毒基因组较小或因为四型、五型、六型 SSR 更容易产生变异而不利于病毒的存活。在三型 SSR 中, B 亚种和 D 亚种中的含量及所占基本一致,而且三个碱基作为一个密码子翻译为一个氨基酸,在病毒基因组中,这些三型 SSR 的保守性可能显示了它们对病毒生命活动有所决定性作用。在二型 SSR 中, D 亚种中二型 SSR 为 83 个,占到了总 SSR 的一半(51%), B 亚种中二

型 SSR 所占总 SSR 比例也较高,达到 43%。在一型 SSR 中, B 亚种中一型 SSR 所占总 SSR 的比例为 37%,远高于 D 亚种的 26% 的比例。此外,在 B 亚种中一型和二型 SSR 所占的比例接近 1,而在 D 亚种中二型 SSR 所占的比例是一型 SSR 所占比例的 2 倍(表 2)。这些现象的出项都可能与 B、D 亚种有着不同的病毒受体以及不同的致病性有关。

由于 B、D 亚种的基因组长度相似并出现类似的 SSR 数,因此在对基因组平均相对密度的分析中也出现了相近的结果,均为 32bp/kb 左右(表 3)。在各型 SSR 的平均相对丰度分析中,表现出的规律为三型 SSR 平均相对丰度在 B、D 亚种中数值相似,为 1 左右。在 B 亚种中,一型和二型 SSR 相对平均丰度数值接近,而在 D 亚种中,二型 SSR 的相对平均丰度数值明显高于其余各型。

目前,在对真核生物及原核生物基因组的研究中发现,单碱基的重复中 A 或 T 重复类型最多,而 C、G 较少;两碱基重复中,较低等生物基因组中重复类型以 AT 或 AG 为主,而高等生物基因组中又以 AC 最多^[16]。在人腺病毒基因组中,一型 SSR 的重复中, B、D 亚种均表现出了对 A、T 两种碱基重复的偏好性(表 4)。在 B 亚种中(A)_n 和(T)_n 的单碱基重复占到四种单碱基重复类型的 76%,在 D 亚种中也同样表现出了相应的现象,这一比值达到 63%。在二型的 SSR 重复中,在可能的 16 中二碱基重复中, B、D 亚种对 CG/GC 重复表现了高度的偏好性。在 B 亚种中 CG/GC 所占比例为 36%,在 D 亚种中 CG/GC 所占的比例达到了 55%。出现这一现象的原因可能是有以下几方面。(1)在 B、D 亚种基因组的组成中 D 亚种中 GC 碱基的含量更高;(2)基因组中 CG 的含量与病毒的进化或稳定性有关;(3)较高的 CG 含量可能与 B、D 亚种不同的病毒受体或不同的致病性有关。

利用 DNAMAN 专用软件对同种类的不同序列进行多序列比对分析。在 B 亚种中,序列的相似性为 67.98%,在 D 亚种中,序列的相似性为 74.12%。在 B、D 亚种的保守序列中,其中所含的 SSR 多为一型 SSR,然后为二型 SSR,其余类型的 SSR 几乎为零。这一现象的出现可能是随着 SSR 重复类型的递

增它的突变频率也随之加大,所以一型 SSR 单碱基重复更为保守,更为稳定。

4 结论

本文对人腺病毒 B 亚种的 31 条基因组序列和 D 亚种的 39 条基因组序列的研究结果显示。B、D 亚种基因组大小及所含的 SSR 个数十分相似,但四型、五型、六型的 SSR 却几乎为零,造成这一现象的原因可能是相对于真核或原核生物人腺病毒的基因组相对较小以及与四型、五型、六型的 SSR 中碱基的突变率更高等因素有关。同时,三型 SSR 在 B、D 亚种中含量相似,相对于 B 亚种,D 亚种中二型 SSR 含量占有主导地位。在 B、D 亚种的一型和二型 SSR 表现出了对不同碱基的偏好性,在一型 SSR 中,(A)_n、(T)_n 的单碱基重复占有较高比例,在二型 SSR 中,(CG/GC)_n 的双碱基重复占主导地位,这些现象可能与 B、D 亚种不同的病毒受体以及不同的致病性相关。在碱基的组成及多序列的对方分析方面,相对于 B 亚种,D 亚种显示了更高的稳定性。通过以上对人腺病毒 B、D 亚种基因组中 SSR 的规律分析,从而为人腺病毒 B、D 亚种的进一步科学研究提供了数据上的支持。

参考文献(References)

- [1] 金天明,武迎红. 腺病毒及腺病毒载体的研究进展[J]. 天津农学院学报,2007,14(2):48-51.
- [2] Jones, M., Harrach, B., Ganac, R. D., Gozum, M. A., Cruz, W. P., Riedel, B., Pan, C., Delwart, E. L., Schnurr, D. P. . New adenovirus species found in a patient presenting with gastroenteritis [J]. *J. Virol*,2007,81(11):5978-5984.
- [3] 陈娜娜,向冬喜,郑丛龙. 腺病毒及其研究进展[J]. 大连医科大学学报,2010,32(5):586-590.
- [4] Soumitra, R., Luk, H. V., Sergey, K., Rebecca, G., Roberto, G., Xin, Y., Martin, k., Arbans, S., Qiang, W., C. Angelica, M., Joshua, B. P., James, M. W. . Isolation and Characterization of Adenoviruses Persistently Shed from the Gastrointestinal Tract of Non - Human Primates[J]. *PLoS Pathogens*,2009,5(7):1-7.
- [5] Ming Chen, Zhongyang Tan, Jianhui Jiang, [ABD2] Mingfu Li, Hongjun Chen, Guoli Shen, Ruqin Yu. Similar distribution of simple sequence repeats in diverse completed Human Immunodeficiency Virus Type 1 genomes[J]. *Fen. Eun. Biol. Sos*,2009,583:2959-2963.
- [6] Van, Belkum, A., Scherer, S., Van, Alphen, L., [ABD3] Verbrage, H. Short - sequence DNA repeats in Prokaryotic genomes [J]. *Microbiol. Mol. Biol. Rev*,1998,62(2):275-293.
- [7] Toth, G., Gaspari, Z., Jurka, J. . Microsatellites in different Eukaryotic genomes; survey and analysis [J]. *Genomes Research*,2000,10:967-981.
- [8] 彭军,谭钟扬,蔡立军,[ABD4] 易丽萍,赵相艳. 甲型流感病毒 HA 基因组的简单重复序列分布分析[J]. 生物信息学,2011,9(1):54-59.
- [9] National Center for Biotechnology Information. Human adenovirus complete genome [DB/OL]. <http://www.ncbi.nlm.nih.gov/nucleore/?term=human%20adenovirus%20complete%20genome>
- [10] Mudunuri, S. B., Nagarajaram, H. A. . IMEX: Imperfect Microsatellite Extractor [J]. *Genome analysis*,2007,23(10):1181-1187.
- [11] 冯海平,谭钟扬,李丹,[ABD6] 赵相艳,李明福,欧阳清检. 乙型肝炎病毒 B2 和 C2 亚型中 SSRs 的比较分布[J]. 生物信息学,2011,9(4):275-280.
- [12] Stallings, R. L., Ford, A. F., Nelson, D., [ABD7] Torney, D. C., Hildebrand, C. E., Moyzis, R. K. . Evolution and distribution of (GT)_n repetitive sequences in mammalian genomes [J]. *Genomics*,1991,10:807-815.
- [13] Lagercrantz, U., Ellegren, H., Andersson, L. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates [J]. *Nucleic. Acids Res*,1993,21:1111-1115.
- [14] Paxton, R. J., Thoren, P. A., Tengo, J., Estoup, A., Pamilo, P. . Mating structure and nestmate relatedness in a communal bee, *Andrena jacobii* (Hymenoptera, Andrenidae), using microsatellites [J]. *Mol. Ecol.*,1996,5:511-519.
- [15] Tautz, D., Schlotterer, C. . Simple sequences [J]. *Curr Opin Genet Dev*,1994,4(6):832-837.
- [16] 高焕,孔杰. 串联重复序列的物种差异及生物功能[J]. 动物学研究,2005,26(5):555-564.