



Applications of the Short Tandem Repeat Polymorphisms in Human Biology

SHAN Zhiqiong

State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200433 China

ABSTRACT: The short tandem repeats (STRs) is a kind of unstable short nucleotide repetitive sequence which can be inherited. STRs distribute extensively in human genome. The high polymorphism and heterozygosity, low mutation rate, and convenience for detection make STRs a kind of widely used genetic marker. The examination methods for STRs were also well developed. Today, STRs are applied to gene mapping, linkage analysis, genetic diagnosis and prenatal diagnosis, molecular anthropology, individual identification and paternity identification in forensic sciences, organic transplantation, etc.

Key words: Short tandem repeats; Polymorphism; Genetic marker; Human biology

短串联重复序列在人类生物学中的应用

单志琼

复旦大学生命科学学院遗传工程国家重点实验室, 上海 200433

摘要: 短串联重复序列是一种可遗传的、不稳定的短核苷酸重复序列, 具有基因组中分布广泛、高度多态性、高杂合度、低突变率、检测简便快捷等优点。自开发利用以来, 在检测方法及应用领域等方面都得到了不断的拓展, 现已广泛应用于基因作图、连锁分析、基因诊断与产前诊断、分子人类学、法医学个体识别和亲子鉴定、器官移植等领域, 是目前应用最广泛的遗传标记。

关键词: 短串联重复序列; 多态性; 遗传标记; 人类生物学

短串联重复序列(short tandem repeat, STR), 也称微卫星 DNA (microsatellite DNA), 由于重复单位较简单, 故又称简单重复序列 (simple repeat sequence, SRS; simple sequences of repeats, SSR)。短串联重复序列广泛存在于原核、真核生物基因组中, 核心序列为 2-7bp 的高度串联重复序列, 重复次数从几次到数十次不等。由于核心单位数目的变化, 使 STR 长度在人群中变异较大, 构成了 STR 的遗传多态性。STR 在基因组中分布很广, 分布于所有染色体, 多位于非编码区附近, 也可位于内含子、启动子、Alu 序列中。STR 广泛的存在于人类基因组中, 越来越受到生物科学工作者的重视。近年来, 由于其检测手段和技术方法的改进, 已经广泛应用于遗传制图、基因定位、法医学鉴定、人类学以及遗传病的诊断等许多领域的研究, 是目前应用比较广泛的遗传标记。

一、STR 的发现和物种特异性

1980 年由 Wyme 等人首先在基因文库中发现的一种遗传标记, 是由长度在十几个

到几十个碱基的“核心序列”串联重复构成的, 重复次数一般在 6-100 次。后来 Nakamura 以“VNTR”定名之。1989 年, Litt 和 Luty[1], Weber 和 May[2]在研究心肌肌动蛋白基因时, 各自用 PCR 扩增和测序胶分析的方法在人类基因组中检测出了一些长度差异都在 2bp 的 DNA 片段, 经序列分析得知这些序列是由[CA]这个双碱基“核心序列”构成的串联重复。同年, Edwards 等人应用 DNA 测序的方法分析“自毁容貌综合征”(Lesch-Ny-ham syndrome)患者的基因突变时, 首次发现了以 3 个或 4 个碱基为重复单位构成的串联重复。相对于长“核心序列”VNTR, 他们共同将这类多态性标记称为短串联重复序列。1991 年, Edwards [3]等人又报告了从基因库中发现的 10 个“核心序列”为 3-4bp 的 STR 位点, 确定了各位点的等位基因数及杂合度, 并对其中的 3 个位点 HUMTH01、HUMFABP 和 HUMHPRTB 进行了复合扩增。在此之后, 众多学者纷纷加入搜寻这类遗传标记的行列, 不断有新的 STR 位点被发现。据 GenBank 等数据库资料

统计, 人类 23 对染色体上至少分布着 7901 个 STR 位点, 每对染色体的 STR 位点分别超过 100 个, 性染色体上的已知 STR 位点数在 264 个以上, 平均每 15kb 就有 1 个 STR 位点, 其片段大小在 100-500bp 之间。

与多种基因座的指纹图不同, 大多数 STR 位点具有人类的物种特异性, 至少是具有灵长类的物种特异性。1995 年有学者[4] 调查了 9 个 STR 的人类物种特异性, 结果在被调查的 23 种动物中, FES/FPS 基因座没有扩增产物, 而 CSF1PO、TOX、TH01、HPRTB、vWA、F13A01 等基因座则在灵长类有扩增产物, 但是这些扩增产物的长度均位于这些基因座的 STR 的等位基因 Ladder 范围之外。此后对更多 STR 基因座的调查也得到了相似的结论。

二、STR 在人类生物学中的应用

1. STR 与人类遗传病

近年来, 发现基因内外的一些 STR 与遗传病的发病有关。如三核苷酸序列延长可引起 Huntington 病、Kennedy 病、脆性 X 综合征等。在脆性 X 综合征中, 脆性 X 智力低下基因 1(FMR1) 基因中的(CGG) 顺序在正常人中的重复次数少于 60 次, 携带者则有 60-100 次, 而发病者则多于 100 次。也有人用 STR 来研究单基因病, 如苯丙酮尿症等[5]。现在, STR 越来越多用于研究多基因病, 如采用 STR 对 I 型糖尿病进行全基因组扫描, 发现 12 个位点与该疾病高度相关, 其中包括 HLA22 抗原位点。更多的人开始用 STR 方法对高血压、哮喘和 II 型糖尿病进行了研究。同时利用 STR 标记可以追踪肿瘤细胞染色体某一特别类型的畸变在肿瘤发生过程中的传递行为, 找到与畸变发生相关的最狭窄的区域, 为某些肿瘤发生相关基因的定位与搜寻奠定基础。

2. 用于人类基因作图和筛选目的基因

人类基因组遗传图又叫做连锁图, 是指基因在人染色体上的相对位置和遗传距离。在对 DNA 遗传标记的选择方面, 人们经历了从致病基因→ABO 血型、HLA 多态位点→RFLP→STR→SNP 的发展过程。STR 在基因组内分布广泛、多态性程度高、可自动化

检测、成为制作基因组遗传图谱的首选遗传标记。虽然单核苷酸多态性(SNP)的多态性位点是最多的, 比 STR 提供更全面的基因信息, 但单个 SNP 的信息量却没有 STR 高, 普遍认识为 SNP 可以提供的信息量大约仅为 STR 的 20%-30% [6], 或者说两个紧密相邻的 SNP 提供的信息量大约与一个 STR 位点相当[7]。Colette 等[8]在 1994 年采用了 5264 个 STR 标记制作了分辨率为 110 cM 的高密度遗传图; 法国 Genethon 实验室与美国国家卫生研究所几个中心合作于 1996 年建立了以 6000 多个 STR 为主体遗传标记、分辨率达 194kb 的高精密图谱[9]。STR 的出现使遗传图的精度得到进一步提高, 同时也成为物理图上的标记, 从而促进了遗传图与物理图的融合。利用 STR 作为遗传标记, 人类基因组计划中的物理图于 2000 年也顺利完成。通过人类家系和对照研究, 运用连锁和(或)相关分析, 可以找到与疾病高度相关的 STR 位点。一般来说, 目的基因若与 STR 位点有连锁关系, 则其位置必定与 STR 位点邻近, 故对其附近基因进行克隆测序, 就有可能发现目的基因[10]。

3. 用于群体遗传学、人类学和进化生物学

STR 标记多位于非编码区, 变异一般不影响人体的结构和功能, 受自然选择压力较小, 突变在进化过程中以近乎稳定的速率传递且不断积累, 形成多样性。STR 等位片段种类和频率在不同人群中分布可能有差异, 这些差异代表某一种族或人群的特征, 反映不同人群间的遗传差异。在 STR 得以运用之前, 人们常用血型差异及 HLA 抗原多态性通过聚类分析和连锁不平衡分析方法分析人类进化、起源、各民族亲缘关系及遗传距离。但由于血型和 HLA 抗原的检测多是针对基因表达的产物, 而在基因表达过程中会有许多因素的干扰, 因而有一定的局限性。通过研究 STR 多态性、变异速率以及比较序列间差异、人群间差异, 分析不同人群间的遗传距离, 就可从分子生物学角度揭示人类的起源、迁徙、进化等历史进程。

20 世纪 90 年代以来对 Y 染色体 STR 的研究, 更为 STR 标记在人类进化方面开拓了新的研究领域。Y 染色体由于在减数分裂时

不发生重组, 呈稳定的父系遗传, 序列的改变仅是由突变引起。因此, 选择多个 Y-STR 构建合适的单倍型、研究单倍型在不同人群中的分布规律, 可以追溯现代人类的父系祖先, 阐明各人群间的遗传进化关系, 与母系遗传的线粒体 DNA(mtDNA)研究互补。目前, 根据 mtDNA、Y 染色体 DNA、常染色体 STR 标记以及多态性 Alu 序列的研究, 大多数分子遗传学证据支持现代人单起源学说, 认为现代人起源于 20 万年前的非洲原始部落, 然后向世界各地迁徙。但也有少数学者坚持现代人类多起源, 认为除非洲外, 亚洲、欧洲也可能是人类发源地。

Karafet 等[11]通过美洲土著 Y 染色体奠基单倍型的分析揭示美洲存在多个父系奠基者单倍型, 美洲土著具有亚洲起源。而 Y 染色体上进化比较快的 STR 标记可以揭示新近的历史事件, 这些在以前其它的标记分析中未能检测到。如 Roewer L 等[12]通过分析 Y 染色体上进化比较快的 STR 标记推断出了欧洲来自 91 个地区的 12700 人, 这些基于 Y-STR 基因座的单倍型频率聚类分析, 将 Y 染色体上进化比较快的 STR 标记分为两个亚群即东欧亚群和西欧亚群, 并可以确认中欧为过度区, 这里单倍型频率随经度的变化快于随纬度的变化, 这些和其它观察到的 Y-STR 都表明这些和一些特定历史事件如德国南部法兰哥尼亚公国和土耳其帝国扩张有关。在中国, 对 Y 染色体 STR 标记的研究, 也有着广泛的意义, 特别是在民族起源、法医鉴定等方面。如李辉[13]等对来源极具争议的客家人进行了很有力的推断。其通过对福建长汀 148 个客家男子从父系遗传的 Y 染色体 SNP 主成分分析及各族 M7 个体 Y-STR 单倍型的网络结构分析, 最终判断客家人的主要成分应是中原汉人, 畬族是对客家人影响最大的外来因素。另外, Bowcock 等对全球 14 个人群进行了 30 个 STR 位点分析, 根据遗传距离以邻接法构建了高分辨率进化树, 先分出非洲人群, 再分出欧洲、大洋洲、美洲印第安、亚洲人群, Calafell 等的研究结果亦遵循该趋势, 因此, 遗传多样性体现了非洲以外人群应为非洲人群的分支。支持了现代人起源的替代假说[14,15]。

三、STR 用于细胞鉴定

从事人类生物学研究不可避免要运用到细胞培养, 而目前实验室中细胞系混乱和细胞交叉污染的现象非常普遍。据国外统计数据显示, 超过 20% 的建株细胞系被错误分类。这种错误对研究工作产生非常严重后果, 不仅浪费了人力和物力, 浪费宝贵的时间, 更产生了不可信的科研结果。这一问题对干细胞研究和肿瘤研究领域尤为关键。因此, 国外的一些机构推荐对各种细胞系(新建株或外部接收)进行定期的鉴定确认。有许多方法可以用于鉴定在细胞培养中产生的交叉污染, 如同功酶分析、染色体组分型、人淋巴细胞抗原(HLA)分型和扩增片段长度多态性(AFLP)的特征分析。但是, 比较好的方法是 STR 图谱法, 此方法是在以 DNA 为基础的法医鉴定领域成功建立起来的。如在 ATCC 全球生物资源中心, 就是通过多重 PCR 对 STR 分析来鉴定细胞系, 可以同时扩增 8 个 STR 位点加 1 个性别决定位点 Amelogenin。每个被分析的人源细胞系都有其独特的 DNA 重复模式, 通过与基础图谱进行比对, 即可对每一批新细胞系的身份进行确证。这种 STR 方法曾防止了 ATCC 将其 6 个“错误”细胞系进一步传播出去—因为经过 STR 分型后, 发现原本来自女性的细胞系中存在着 Y 染色体特异的扩增产物。

STR 作为新一代的遗传标记, 具有广阔的应用前景。目前, 对于 STR 的产生机制和用途还不是很清楚, 随着 STR 检测分型技术的不断更新和进步以及人类后基因组时代的不断推进, STR 的作用会日渐明晰, 也将发现更多的基因座及基因型, 人类所探讨的问题将越来越全面。STR 将会应用于更多方面的研究, 而且在人类学、法医学等领域将体现出难以限量的价值。

参考文献

1. Litt M, Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat with in the cardiac muscle actin gene. *Am J Hum Genet* 44: 397-401.
2. Weber JL, May PE (1989) A abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44: 388-396.
3. Edwards AL, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 49:746-752.

4. Crouse CA, Schumm J (1995) Investigation of species specificity using nine PCR-based human STR systems. *Forensic Sci Int* 143: 47-52.
5. 刘洋, 姜丽华, 汪运山(2003) 微卫星 DNA 的测定方法. *中国误诊学杂志* 3:994-995.
6. 郑秀芬(2002) 法医 DNA 分析. 北京: 中国人民公安大学出版社: 383-388.
7. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077-1082.
8. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Missasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of human genome based on 5264 micro-satellites. *Nature* 380:14-17.
9. Kang S, Jaworski A, Ohshima K, Wells RD (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nat Genet* 10:213-218.
10. 王冰梅, 庄文越, 王雪松, 苏铁克(2006) 短串联重复序列的研究. *北华大学学报(自然科学版)* 7: 43-46.
11. Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am J Hum Genet* 64: 877-881.
12. Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet* 116: 279-291.
13. 李辉, 潘悟云, 文波, 杨宁宁, 金建中, 金力, 卢大儒(2003) 客家人起源的遗传学分析. *遗传学报*. 30: 873-880.
14. Bowcock AM, Ruiz-Linares A, Tonfrohde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368: 455-457.
15. Calafell F, Shuster A, Speed WC, Kidd JR, Kidd KK (1998) Short tandem repeat polymorphism evolution in humans. *Eur J Hum Genet* 6:38-49.