

基于水稻基因组序列 SSR 的多态性分析

陈仲中 汪旭升 朱 军*

(浙江大学 生物信息学研究所, 浙江 杭州 310029; E-mail: zhongzhongchen@sohu.com; *通讯联系人, E-mail: jzhu@zju.edu.cn)

Analysis of SSR Polymorphism by Genome-Scale Comparing Between Varieties in Rice

CHEN Zhongzhong, WANG Xu-sheng, ZHU Jun*

(Institute of Bioinformatics, Zhejiang University, Hangzhou 310029, China; E-mail: zhongzhongchen@sohu.com; *Corresponding author, E-mail: jzhu@zju.edu.cn)

Abstract: One thousand four hundred and fifty three loci were identified the same as in two rice varieties, indica rice 93-11 and japonica rice Nipponbare, by using bioinformatics tools based on 6655 sequences released by Monsanto enterprise which consisted of perfect repeats flanked by 100 bp on either side of the simple sequence repeat (SSR). Among those sequence hits, 1449 and 1451 SSRs were detected in 93-11 and Nipponbare, respectively. Among those, 371 SSRs had the same repeat units and the same repeat count, while 804 SSRs had the same special motifs with different repeat counts. In addition, in the same repeat units with different repeat counts SSR class, di-nucleotide SSRs were the most abundant, occupying 62.94% identified, while penta- and hexa-nucleotide SSRs were the least frequent. Polymorphism of SSR between two varieties in the collinear region were identified, revealing that new SSR markers could be developed for the further research of the genetic variation and genome analysis of the two varieties.

Key words: simple sequence repeat; rice; polymorphism

摘要: 根据 Monsanto 公司所发布的 6655 个序列[由简单重复序列(SSR)及其两翼各 100 个碱基组成],通过生物信息学的方法,找到了在籼稻(*Oryza sativa* subsp. *indica*)品种 93-11 与粳稻(*Oryza sativa* subsp. *japonica*)品种日本晴的基因组序列中匹配最好的 1453 个同源位点。在这些位点中,93-11 和日本晴中分别搜索到了 1449 个和 1451 个 SSR。通过对 SSR 的分类比较,发现在 93-11 和日本晴的相同位点存在 1175 个具有相同基序的 SSR,其中 371 个具有相同的基序和重复个数,804 个具有相同的基序但重复次数有差异。具有相同基序但重复次数有差异的 SSR 类型中,频率最高的是 2 个核苷酸的重复子,占 62.94%,基序为 5 个和 6 个核苷酸的 SSR 则较少。对水稻这两个品种相应区域 SSR 的多态性比较分析,揭示了用这些特定类型的 SSR 可以在这两个品种的特定区域发展新的 SSR 标记,并为研究这两个品种之间遗传变异、基因组功能之间的差异等提供依据。

关键词: 简单重复序列; 水稻; 多态性

中图分类号: Q943; S330; S511.03

文献标识码: A

文章编号: 1001-7216(2005)04-0303-05

简单重复序列 (simple sequence repeat, SSR) 又称微卫星 DNA,是一类由 2~6 个碱基组成的串联重复 DNA 序列。SSR 普遍存在于动植物当中,可以根据 SSR 两侧的保守序列设计引物进行 PCR 扩增,由于不同品种 SSR 基序的重复次数不同,导致 PCR 扩增条带差异性,即简单重复序列长度多态性 (simple sequence length polymorphism, SSLP), SSLP 很容易通过 PCR 快速扩增和高分辨率的琼脂糖凝胶或聚丙烯酰胺凝胶电泳检测到。在水稻基因组中估计存在 5 700~10 000 个基序为 2、3、4 个碱基的 SSRs,可用于构建水稻的全基因组图谱^[1]。SSRs 对水稻遗传图谱、物理图谱和基于序列的图谱的整合都非常有用,同时它为育种家和遗传学家提供了一种把表型变异和基因型变异联系起来的工具^[2-5]。

亚洲栽培稻 (*Oryza sativa* L.) 可以分为两个主要的亚种,籼稻 (*Oryza sativa* subsp. *indica*) 和粳

稻 (*Oryza sativa* subsp. *japonica*)。这两个栽培亚种的种内表型变异非常显著,包括生长、发育和环境适应性。通过国家级研究所 (如美国基因组研究所 TIGR 和北京基因研究中心) 和私人公司 (如 Syngenta 和 Monsanto 公司) 的共同努力,目前已经发布了水稻两个亚种的草图序列^[6-8]。Li 等分析了水稻栽培种日本晴第 4 号染色体上 1844 个 SSRs,并比较了水稻栽培种籼稻 (广陆矮 4 号) 和粳稻 (日本晴) 第 4 号染色体上 SSRs 的分布和长度多态性^[9], McCouch 等根据 Monsanto 公司提供的 6655 个包含重复子的序列开发了 2240 个微卫星标记并整合到水稻遗传图谱中^[10]。然而,对籼稻品种 93-11 和粳稻品种日本晴全基因组 SSR 的多态性研究并未见报道。本文根据 Monsanto 公司提供的 6655 个包含重复子的序列和已经发布的基因组序列,通过

收稿日期: 2004-09-06; 修改稿收到日期: 2004-11-25。

第一作者简介: 陈仲中 (1981 -), 男, 在读硕士研究生。

序列联配和相应位点的比较,分析了籼稻 93-11 和粳稻日本晴在相应区域序列上 2~6 个重复子的 SSRs,揭示了这两个品种之间 SSRs 的差异性,同时对部分 SSRs 进行了电子 PCR(ePCR)验证,以期能为研究这两个品种间基因组功能的差异提供依据,并为 QTL 定位分析和分子标记辅助育种等奠定基础。

1 材料与方法

1.1 基因组序列来源

中国科学院北京基因组研究所发布的籼稻 93-11 基因组序列从 Rise 网站 (<http://rise.genomics.org.cn>) 获得,国际水稻测序中心发布的粳稻日本晴基因组序列从 TIGR 网站 (<http://www.tigr.org>) 下载。所有水稻品种上 SSRs 标记的信息来自 Monsanto 公司 (<http://www.rice-research.org>),同时这些标记的序列可以根据其登录号从 GenBank (GenBank GI # 12700719-12707547) 获取。

1.2 方法

1.2.1 等位标记的获取

首先,把籼稻 93-11 和粳稻日本晴基因组序列分别与由 Monsanto 公司发布的含有 SSR 的序列作 BLASTn^[11], E 值设为 10^{-30} ,在 BLASTn 搜索的结果中取与含 SSRs 标记序列联配最好的 93-11 和日本晴序列;然后把从 93-11 和日本晴中所取得的序列用 Sim4^[12]作联配,能联配上的序列分别作为籼稻(93-11)和粳稻(日本晴)的等位位点,进行 SSRs 多态性分析。

1.2.2 SSRs 分析

通过自编的 Perl 程序分析在籼稻(93-11)和粳稻(日本晴)中都存在的等位标记,获得各种不同类型的 SSRs。检测到重复子后分别列出 2、3、4、5、6 个核苷酸重复子。然后从分类、频率和次数等方面对 SSRs 进行比较,从而得到这两个品种基因组相应位点处 SSRs 多态性。

2 结果与讨论

2.1 籼稻(93-11)和粳稻(日本晴)中 SSRs 的分类比较

在 1453 个相应区域内 93-11 和日本晴的简单重复序列分别为 1449 个和 1451 个。来自这两个品种的 SSRs 可以分为 4 组(表 1):1) 位点、长度和基序都相同的 SSRs;2) 位点、基序相同但长度不同的 SSRs;3) 位点相同但基序不同的 SSRs;4) 只在

表 1 93-11 和日本晴中 SSRs 相应区域多态性检测结果

Table 1. Polymorphism results between 93-11 and collinear Nipponbare.

类型 Item	93-11	日本晴 Nipponbare
SSR 的总数目(12 bp)	1449	1451
位点、长度和基序都相同	371	371
位点、基序相同但是长度不同	804	804
位点相同但基序不同	272	272
只在一个品种中存在	2	4

一个品种中存在的 SSRs。93-11 和日本晴中共有 371 个等位标记检测到具有相同基序和重复次数的 SSRs,占了这些等位标记的近 25%,这表明在 Monsanto 公司发布的序列中,即使 93-11 和日本晴品种都存在相同基序类型的 SSRs,但因其重复次数相同,若在该重复子两侧设计引物,由于扩增条带无明显差异性,可能并不适合用于开发这两个品种特定区段的分子标记。有 804 个等位位点仅存在重复次数的差异。在 93-11 和日本晴的相应位点,这类 SSRs 由于能够在扩增条带中把两个品种区分开来,可能是一种比较好的分子标记来源。位点相同但基序不同的 SSRs 有 272 个,另外有 2 个 SSRs 在 93-11 中发现但不存在于日本晴中,4 个只存在于日本晴中。这些 SSRs 的存在/缺失可能是由于碱基的替换或剪接引起的。例如,93-11 序列的基序 (CCGACG)₁₃ 在日本晴中变成了基序 (CCG)₂₆。这是由于碱基 CCG 被替换成了 ACG,导致脯氨酸 (CCG) 变成了苏氨酸 (ACG)。这类 SSRs 反映了这两个品种的相似性和分化程度,可能是分析这两个品种基因组功能的一种比较好的 SSRs 来源,但并不适合用来开发基于区分这两个品种的分子标记,可能比较适合于对这几个等位标记进行单独分析以研究这两个品种的特殊基因组功能。

2.2 SSR 标记频率和长度的分析

我们主要分析了具相同基序但重复次数不同的 SSR 类型。对此类 SSRs 数目与频率的比较分析表明,频率最高的是二核苷酸的重复子,占 62.94%,5 个核苷酸的 SSRs 则最少(表 2)。频率最高的 2 个、3 个、4 个核苷酸的重复子分别是 CT/GA、CCG/CGG、ATA G。另外基序为 5 个核苷酸的 SSRs 共检测到 2 个,分别是 GCCTC 和 AGAGA;基序为 6 个核苷酸的 SSRs 共检测到 3 个,分别是 CTCCGC、AAGAAC、AAACCA。同时表 2 也列出了 93-11 和日本晴相同位点、相同基序的 SSR 重复次数差异的变化幅度。如在基序为 2 个核苷酸的重复子当

表 2 93-11 和日本晴基因组相应区域 SSRs 数目和频率差异

Table 2. Total numbers of SSRs and frequency variation between 93-11 and collinear Nipponbare.

SSR	发生次数	百分率	重复次数差异	最高频率类型
	Frequency	Percentage/ %	Variation in repeat count	Types with the highest frequency
(NN) <i>n</i>	506	62.94	1 ~ 71	GA
(NNN) <i>n</i>	229	28.48	1 ~ 43	CCG
(NNNN) <i>n</i>	64	7.96	1 ~ 6	ATA G
(NNNNN) <i>n</i>	2	0.25	1	
(NNNNNN) <i>n</i>	3	0.37	1 ~ 2	
总数 Total	804	100.00		

中,在两个品种的不同位点处检测到 18.2% 富含 (AT)_{*n*} 的 SSRs,其中有一等位标记在 93-11 中 AT 连续重复 114 次,而在日本晴中 AT 连续重复 43 次,相差了 71 次。图 1 列出了在相同位点具有相同基序,不同重复次数的 SSRs 重复次数的差异性比较结果。图 1 表明在这 804 个 SSRs 中重复次数相差 1 个的重复子类型最多,占了 30%,重复次数相差越大,所占百分比依次减少,呈指数分布。

图 2 对 804 个 SSRs 中基序为 2、3、4 个核苷酸重复子的平均重复次数的多态性进行了比较。以二核苷酸重复子为例,这里考虑到编码链和模板链,认

为此类的基序如 (AT)_{*n*} 和 (TA)_{*n*} 是等同的。在二核苷酸重复子中,主要为四类, (GA)_{*n*}、(AT)_{*n*}、(AC)_{*n*} 和 (CG)_{*n*},但是在两个品种的相应区域没有发现富含 (CG)_{*n*} 类型的重复子。其中 (GA)_{*n*} 的类型在这两个品种的相应区域中共有 378 个,占了二核苷酸重复子的 74.7%。在水稻中富含 (GA)_{*n*} 的 SSRs 通常 GC 含量接近于 50%,具有较好的 PCR 扩增性。但是在这两个品种相应区域这些重复子的平均重复次数差异并不明显,在 93-11 中的平均重复次数为 18.3 次,在日本晴中为 17.9 次,平均相差 0.4 次; (AT)_{*n*} 的类型占了二核苷酸重复子的 18.2%,重复

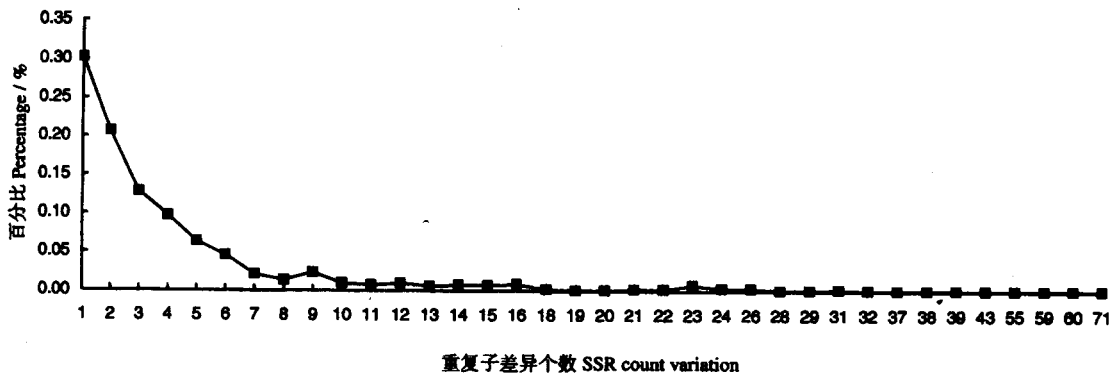


图 1 93-11 和日本晴在相同位点处 SSR 重复次数差异性比较

Fig. 1. SSR repeat count variation between 93-11 and Nipponbare at the same loci.

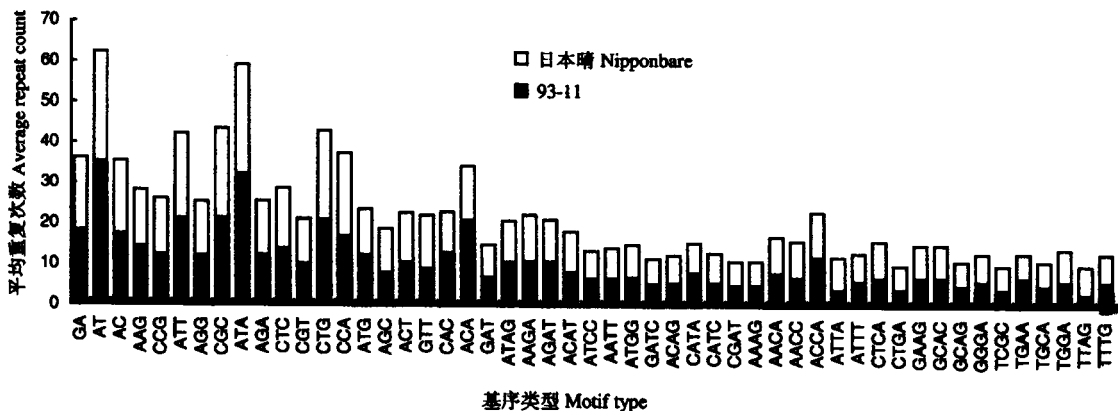


图 2 基序为 2、3、4 个核苷酸重复子平均重复次数的差异性比较

Fig. 2. Average repeat count variation of dimers, trimers, tetramers at the same loci.

次数差异最大,在 93-11 中的平均重复次数为 35.3 次,在日本晴中为 27.2 次,平均相差 8.1 次。93-11 和日本晴相应区域内特殊基序类型 SSRs 的分子标记,将有助于设计更富多态性的分子标记,用于定位染色体上的基因,特别是为不同品种等位基因的差异性研究提供了依据。

2.3 电子 PCR 验证

为了检验特殊类型 SSRs 在 93-11 和日本晴中的多态性,根据 Monsanto 公司所提供的引物,我们对具有相同位点、相同基序但重复次数不同的 804 个 SSR 进行了电子 PCR 验证。分析发现有 74.8% 的引物在 93-11 中扩增到产物,并且在扩增到的产物中有 87.9% 为单拷贝,同时在这些单拷贝中发现有 59.5% 在日本晴中也为单拷贝,且在两个品种中呈现多态性,把这些 SSR 分子标记定位在日本晴的物理图谱中可以发现它们在整个基因组中的分布差异很大,在第 1、2 和 3 染色体上分布比较密集,而在第 10、11 和 12 染色体上则比较疏散(图 3)。通过对部分 SSR 进行电子 PCR 验证表明,在 93-11 和日本晴的相应位点处 SSR 的多态性主要是由于碱基的插入与缺失,或者是由于不同品种之间 SSR 拷贝数的差异造成的。同时也表明 SSR 分子标记是研究水稻品种间多态性的一种重要分子标记来源,为进一步通过大范围寻找基因组内 SSR 的分布并分析品种间特殊类型 SSR 的多态性提供了依据。目前,单纯地开发 SSR 标记已经没有必要,而应结合生物基因组的特定功能,按研究目标区段的序列区域开发 SSR 标记。通过生物信息学的方法所得到的这些 SSR 分子标记,特别是在这两个品种中均为

单拷贝、且扩增条带有明显差异的 SSR 标记,对研究品种中遗传变异、基因组功能之间的差异异常重要。

为了更好地利用禾本科模式植物水稻的基因组,有必要把水稻的遗传资源、功能分析和序列信息结合起来。多态性 DNA 分子标记为基因组序列的遗传图谱和表型变异之间的联系搭建了一座桥梁。对特殊类型 SSRs 进行研究,并将其定位到染色体上,对于揭示水稻基因组的特殊功能具有非常重要的作用。以往研究表明,富含二核苷酸重复子 (DNRs) 的微卫星序列通常位于非编码区,并且是 SSR 多态性分子标记的最好来源^[13]。早期主要是研究基因的 5 端和 3 端,但在研究中并没有发现 DNRs 与转座子有联系^[14,15]。对玉米和人类的基因组研究发现,富含 GC 的三核苷酸重复子 (TNRs) 代表了在 cDNA 中检测到的大部分 SSRs^[16,17]。相反,在拟南芥和酵母菌 (*Sacharomyces cerevisiae*) 中,主要的三核苷酸重复子则富含 AT^[18,19]。这些三核苷酸重复子最初是在酵母基因的转录、信号转导、细胞生长和分裂的机制中发现的,但极少在控制光合作用如糖酵解和呼吸作用的基因中被发现^[13]。不同三核苷酸重复子和某些特定类型基因之间的联系,或许可以表明某种基因组类型。基因特殊功能分类中的 TNRs 频率,对于包含 TNRs 基因的全基因组的注释可能非常有用。而我们在水稻这两个品种相应区域具有差异 SSRs 的三核苷酸重复子中,发现富含 GC 的三核苷酸重复子占了该类重复子的 50.2%,这些富含 GC 的三核苷酸重复子有 50.9% 存在于编码序列中;富含 AT 的三核苷酸重复子占

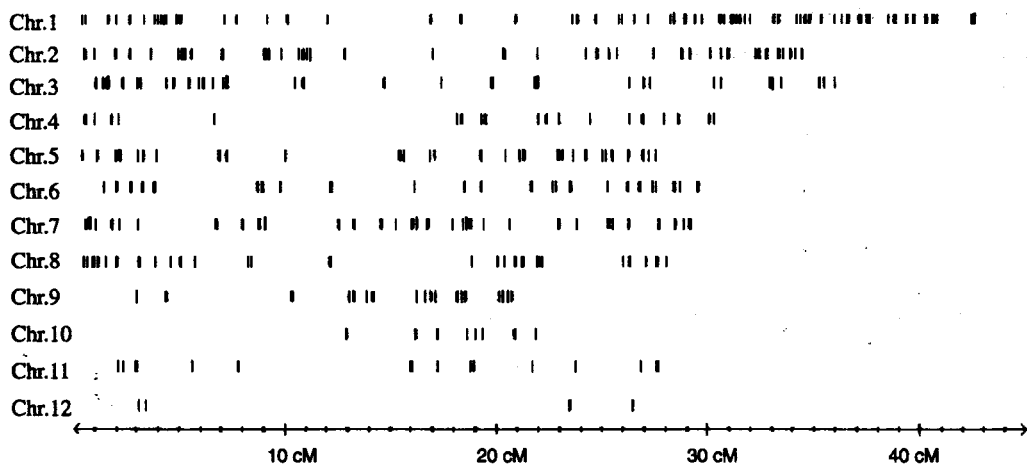


图 3 部分扩增产物为单拷贝的 SSR 分子标记在日本晴物理图谱中的分布

Fig. 3. Distribution of some unique SSR markers on the physical map of Nipponbare.

了该类重复子的 49.8%, 但是却只有 22.1% 在编码序列中检测到。现在还不清楚这些特殊类型三核苷酸重复子是否与水稻基因组的特殊功能相联系, 但若能把这些位点定位到水稻的不同品种之间, 通过比较品种的遗传变异来分析水稻的基因组功能, 可能会取得一些新的突破。

参考文献:

- 1 McCouch S R, Chen X, Panaud O, *et al.* Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Mol Biol*, 1997, 35:89 - 99.
- 2 Akagi H, Yokozeki Y, Inagaki A, *et al.* Microsatellite DNA markers for rice chromosomes. *Theor Appl Genet*, 1996, 93: 1071 - 1077.
- 3 Panaud O, Chen X, McCouch S R. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol Gen Genet*, 1996, 252:597 - 607.
- 4 Temnykh S, DeClerck G, Lukashova A, *et al.* Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency length, variation transposon associations and genetic marker potential. *Genome Res*, 2001, 11:1441 - 1452.
- 5 Wu K S, Tanksley S D. Abundance, polymorphism and genetic mapping of microsatellites in rice. *Mol Gen Genet*, 1993, 241: 225 - 235.
- 6 Barry G F. The use of the Monsanto draft rice genome sequence in research. *Plant Physiol*, 2001, 125:1164 - 1165.
- 7 Goff S A, Ricke D, Lan T H, *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 2002, 296: 92 - 100.
- 8 Yu J, Hu S, Wang J, *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 2002, 296:79 - 92.
- 9 Li C, Zhang Y, Ying K, *et al.* Sequence variations of a simple sequence repeats on chromosome-4 in two subspecies of the Asian cultivated rice. *Theor Appl Genet*, 2004, 108:392 - 400.
- 10 McCouch S R, Teytelman L, Xu Y, *et al.* Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res*, 2002, 9:199 - 207.
- 11 Altschul S F, Gish W, Miller W, *et al.* Basic local alignment search tool. *J Mol Biol*, 1990, 215:403 - 410.
- 12 Florea L, Hartzell G, Zhang Z, *et al.* A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res*, 1998, 8:967 - 974.
- 13 Temnykh S, DeClerck G, Lukashova A, *et al.* Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res*, 2001, 11:1441 - 1452.
- 14 Cho Y G, Ishii T, Temnykh S, *et al.* Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor Appl Genet*, 2000, 100:713 - 722.
- 15 Temnykh S, Park W D, Ayres N, *et al.* Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theor Appl Genet*, 2000, 100:697 - 712.
- 16 Chin E C, Senior M L, Shu H, *et al.* Maize simple repetitive DNA sequences: abundance and allele variation. *Genome*, 1996, 39:866 - 873.
- 17 Jurka J, Pethiyagoda C. Simple repetitive DNA sequences from primates: Compilation and analysis. *J Mol Evol*, 1995, 40: 120 - 126.
- 18 Cardle L, Ramsay L, Milbourne D, *et al.* Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*, 2000, 156: 847 - 854.
- 19 Young E T, Sloan J S, Van Riper K. Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. *Genetics*, 2000, 154: 1053 - 1068.