

水稻 microRNA 单核苷酸多态性

于 颢 闫 旭 郭卫东 辛德东*
(浙江师范大学 化学与生命科学学院, 浙江 金华 321004; E-mail: xindedong@zjnu.cn)

Single Nucleotide Polymorphism of MicroRNA in Rice

YU Hao, YAN Xu, GUO Wei-dong, XIN De-dong*
(Chemistry and Life Science College, Zhejiang Normal University, Jinhua 321004, China; E-mail: xindedong@zjnu.cn)

YU Hao, YAN Xu, Guo Weidong, et al. Single nucleotide polymorphism of microRNA in rice. Chin J Rice Sci, 2011, 25(5): 467–474.

Abstract: MicroRNAs (miRNAs) are a class of endogenous small (about 22 nucleotides), non-coding RNAs, which play important roles in plant growth, development and stress tolerance by binding to the target sites in mRNAs. Since a single nucleotide change in miRNAs can affect the maturity and regulation of miRNA, study on the distribution of single nucleotide polymorphisms (SNPs) in miRNA gene may be of great importance for functional and evolutionary research. In order to explore the potential functional significance of SNPs in the pre-miRNA in rice, we analyzed the distribution of SNPs in pre-miRNA region and flanking region, including 5' and 3' flanking regions in rice. The results showed that there was no significant difference in the frequency of SNPs between the pre-miRNA region and its flanking region in rice, which was different from the results in other species. Further study on the frequency of SNPs across pre-miRNA secondary structures showed that there was no significant difference between seed region and loop region, which was thought to undergo less negative selection pressure in other species. However, in consistent with other species, the frequency of SNPs in protein-coding gene region of rice was significantly lower than that in gene flanking region, indicating that the flanking region of pre-miRNA may be essential for the function of miRNA. Another reason was the uneven distribution of SNPs in different miRNAs. Some miRNAs were less important to rice function, thus was subjected to relaxation of selective pressure, accumulating more and more mutations to become a new gene or loss of functions. However, we could not rule out the bias in the results because of the unreliable data in current SNP database.

Key words: microRNA; rice; single nucleotide polymorphism

于 颢, 闫 旭, 郭卫东, 等. 水稻 microRNA 单核苷酸多态性. 中国水稻科学, 2011, 25(5): 467–474.
摘 要: microRNA(miRNA)是一类内源性非编码小 RNA, 长度约 22 个核苷酸。miRNA 通过与靶基因 mRNA 特定位点结合, 调控植物生长、发育和胁迫耐受性。pre-miRNA 的单核苷酸突变会影响 miRNA 的成熟过程和调控功能, 因此, 研究 miRNA 基因中的单核苷酸多态性(SNP)分布对研究 miRNA 功能分化和基因进化有重要意义。为探讨水稻 miRNA 基因单核苷酸多态性及其潜在生物学意义, 基于现有水稻 SNP 数据, 分析了水稻 pre-miRNA 区、侧翼区(位于 pre-miRNA 上下游, 与 pre-miRNA 等长的区间)的 SNP 分布。与已经报道过的绝大多数生物不同, SNP 在水稻 pre-miRNA 区的频率与侧翼区无显著差异。进一步研究 SNP 在 pre-miRNA 二级结构中的分布, 发现种子区 SNP 频率跟茎区、环区也无显著差异。然而, SNP 在水稻基因区分布跟其他生物相似, 即基因区 SNP 频率显著低于基因侧翼区。这可能暗示 pre-miRNA 侧翼区功能对 miRNA 非常重要, 也可能与 SNP 在不同 miRNA 基因之间的分布不均衡有关, 一部分 miRNA 基因正经历纯化选择压力放松, 积累较多突变而成为新的基因或失去功能。但是以上结果并不排除是由于当前水稻 SNP 数据可靠性不高而导致的偏差。

关键词: microRNA; 水稻; 单核苷酸多态性
中图分类号: Q74; S511 **文献标识码:** A **文章编号:** 1001-7216(2011)05-0467-08

miRNA(microRNA)是一类来自真核生物基因组的非编码小 RNA, 长度为 21~25 nt, 通过与靶基因 mRNA 特定位点结合起调控作用^[1]。从 1993 年首次发现 miRNA 至今, 已在人、小鼠、果蝇、拟南芥、水稻等 133 个物种中发现 15 000 余条成熟 miRNA(miRBase 15.0)。大多数 miRNA 基因首先转录生成初级转录本 pri-miRNA(primary-microRNA), 通过特定的选择性剪切生成具有茎环结构的前体 pre-miRNA(precursor-microRNA), 而后再次剪切, 最终产生成熟 miRNA。Pre-miRNA 二级结构中成熟区(经过剪切产生成熟 miRNA 的

区间)相对其他区间在进化上保守性较强^[2-3]。
单核苷酸多态性(single nucleotide polymorphism, SNP)是由基因组核苷酸水平上的变异引起的 DNA 序列多态性, 包括单碱基的转换、颠换以及单碱基的插入、缺失等。SNP 具有在基因组中数量多、分布密度高等特点。DNA 水平的遗传多态性表现为核苷酸序列的差异, 比如单核苷酸的多态性, 而这些多态性有时也会引起性状上的改变。由于 miRNA 基因在进化上具有保守性, pre-miRNA 或

附近区间微小的碱基突变就可能影响 pre-miRNA 的剪切或成熟 miRNA 的碱基组成,进而影响成熟 miRNA 的功能,使突变体产生异常的生命现象。因此,在所有 3' UTRs 顺式元件中,miRNA 结合位点要比其他顺式元件受到更多的纯化选择(negative selection)^[4]。植物中 pre-miRNA 的单核苷酸多态性与植物生长、发育和逆境胁迫存在重要关联^[5-7]。

水稻是世界重要粮食作物之一,广大研究者对水稻在基因组和转录组水平的研究较绝大多数植物物种深入。SNP 在水稻遗传图谱的构建^[8-9]、基因克隆和功能基因组学研究^[10-11]、标记辅助选择育种^[12]、遗传资源分类及物种进化等方面具有巨大的应用潜力^[13-15]。在植物 SNP 研究中,除了大麦^[16]、大豆^[17]、玉米^[18]和甜瓜^[19]等多个物种 SNP 具有高密度分布外,水稻同样存在 SNP 数量多、分布密度高等特点,并且 SNP 频率在水稻各个染色体之间存在显著差异,在每条染色体上分布不均匀,存在明显的 SNP 富集区与稀缺区^[9, 20]。

研究 SNP 在 pre-miRNA 的分布对 pre-miRNA 的预测和 miRNA 基因进化研究有重要意义。SNP 在 pre-miRNA 的分布已经越来越为大家关注,在人类、拟南芥等物种中已经进行了相关研究,结果显示 SNP 在 pre-miRNA 侧翼区的数目和出现频率显著高于 pre-miRNA 区。SNP 在人类 pre-miRNA 区的出现频率为 1.3 个/kb,显著小于其侧翼区的 3.0 个/kb^[21]; SNP 在拟南芥 pre-miRNA 区的频率为 0.7 个/kb,其两端侧翼区的 SNP 频率分别为 1.7 个/kb 和 2.0 个/kb,与 pre-miRNA 区表现出显著差异^[22]。

本研究分析了 SNP 在水稻 pre-miRNA 及其侧翼区、蛋白编码基因及其侧翼区以及 SNP 在 pre-miRNA 二级结构上的分布,发现水稻 pre-miRNA 区的 SNP 平均频率与其侧翼区平均频率并无显著差异。此外,SNP 在 pre-miRNA 二级结构的种子区、成熟区、茎区和环区的平均频率也无显著性差异。这是继发现水稻 SNP 在染色体上和不同基因间具特异分布之后又一水稻特有的分布特点。但是这种结果并不排除由于当前水稻 SNP 数据可靠性不高而导致的偏差。

1 材料与方法

1.1 pre-miRNA 数据、SNP 数据、水稻染色体数据来源

miRBase 数据库(<http://www.mirbase.org/>, Release15.0)。水稻所有的 SNP 数据下载于 NCBI dbSNP 数据库(ftp://ftp.ncbi.nih.gov/snp/organisms/rice_4530/chr_rpts/),由于 SNP 位置坐标是基于水稻染色体数据版本 build:4.1 得到的,本研究所用的水稻染色体数据也采用版本 build:4.1,下载于 NCBI Mapview 数据库(<http://www.ncbi.nlm.nih.gov/mapview/mapsearch.cgi?taxid=4530>)。

1.2 Pre-miRNA 在水稻染色体上的定位

从 Bowtie 网站(<http://www.sourceforge.net/projects/bowtie-bio/fines/bowtie/0.12.7/>)下载 Bowtie 0.12.7 软件。通过 Bowtie 模块使用命令“bowtie -a -v 3”将水稻 pre-miRNA 序列定位到水稻染色体上。命令行中“-a”表示 Bowtie 将输出所有能和染色体匹配的结果;“-v 3”表示在 pre-miRNA 序列与染色体序列的错配碱基数不大于 3 个。用 Perl 程序提取在染色体上只有 1 个定位位点的 miRNA 定位信息。

1.3 Pre-miRNA 及其侧翼区、蛋白编码基因及侧翼区 SNP 分布

将 miRNA 基因上游且与 pre-miRNA 相邻、等长的两个连续区间分别称为 miRNA 基因的远端 5'侧翼区和近端 5'侧翼区;下游与 pre-miRNA 相邻、等长的两个连续区间分别称为近端 3'侧翼区和远端 3'侧翼区。而 pre-miRNA 在染色体上的位置作为 miRNA 基因区。利用 NCBI dbSNP 数据库中的水稻 SNP 定位信息,分别计算上述 5 个区间 SNP 数目和频率。

将蛋白编码基因上下游与它相邻、等长的区间分别称为 5'侧翼区和 3'侧翼区。自 NCBI(ftp://ftp.ncbi.nih.gov/genomes/MapView/Oryza_sativa/objects/BUILD.4.1/initial_release/)下载水稻所有蛋白编码基因的染色体定位数据,提取基因组坐标信息,获取蛋白编码基因区及 5'侧翼区和 3'侧翼区在染色体上的物理位置,分别计算这 3 个区间的 SNP 数目和频率。

1.4 Pre-miRNA 二级结构分区及 SNP 分布

将 pre-miRNA 二级结构分为 4 个区间:1)种子区(即成熟 miRNA 的前 6 个碱基位置);2)成熟区(即成熟 miRNA 在 pre-miRNA 的位置);3)环区;4)茎区(即除以上区间外 pre-miRNA 上的配对区)。通过 Perl 程序,分别计算上述 4 个区间的 SNP 个数及出现频率。

1.5 SNP 位点、pre-miRNA 和蛋白编码基因在染色体上的分布

研究表明,水稻 SNP 在染色体中呈不均衡分布^[9]。为进一步检验水稻 SNP 分布与 pre-miRNA 和蛋白编码基因分布之间是否具有相关性,以 100 kb 为单位,沿每条染色体移动,依次计算每一区间内 SNP、pre-miRNA 和蛋白编码基因在整条染色体上的出现频率,即每一区间内出现数目与整条染色体出现数目的比值。用 Spearman 秩相关检验方法分别检验 pre-miRNA 富集区与 SNP 富集区、蛋白编码基因富集区与 SNP 富集区之间的相关性。

1.6 SNP 分布的显著性检验

将 SNP 频率数据导入 R 软件(<http://www.r-project.org/>),分别检验 pre-miRNA 与侧翼区(远端 5'侧翼区、近端 5'侧翼区、远端 3'侧翼区、近端 3'侧翼区)之间、蛋白编码基因与其侧翼区(5'侧翼区、3'侧翼区)之间以及 pre-miRNA 二级结构不同区域之间(种子区、成熟区、环区和茎区)SNP 平均频率的差异显著性。平均值检验方法分别采用 *t* 检验和非参数 Wilcoxon 检验。柱状图中误差线代表 95% 置信区间,用非参数 Bootstrap 方法获得。

2 结果与分析

2.1 水稻 pre-miRNA 及其侧翼区 SNP 频率无显著差异

用 Bowtie 将水稻 pre-miRNA 定位到染色体上,去掉具有多个定位信息的序列后,得到 412 条在染色体上具有唯一位置的 miRNA 基因。去除数据中位置不确定的 SNP 之后,共得到 4 318 个在 miRNA 基因及其侧翼区具有唯一位置的 SNP 位

点。在远端 5'侧翼区、近端 5'侧翼区、pre-miRNA 区、近端 3'侧翼区、远端 3'侧翼区的 SNP 数目分别为:998、836、885、838 和 761 个,每条 pre-miRNA 对应的上述 5 个区间中 SNP 平均出现频率(在各个区间中将每条 pre-miRNA 上相应区间 SNP 数目与 pre-miRNA 长度的比值相加求平均值)分别为 17.30、14.76、14.83、14.23、12.97 个/kb。为了检验水稻 pre-miRNA 区 SNP 频率是否倾向于比侧翼区低,分别用 *t* 检验和 Wilcoxon 检验,比较 SNP 频率在 pre-miRNA 区与侧翼区差异。结果发现,与人^[21]和拟南芥^[22]不同,水稻 SNP 频率在 pre-miRNA 区与其 5'侧翼区和 3'侧翼区均无显著差异(表 1)。

2.2 水稻蛋白编码基因及基因侧翼区 SNP 分布

与其他物种不同,SNP 在水稻 pre-miRNA 区与其侧翼区的频率无显著性差异。为进一步检验这种现象是否在蛋白编码基因中也存在,通过统计 NCBI 所有水稻 SNP 位点的个数和水稻 12 条染色体长度,计算出水稻染色体 SNP 分布的平均频率为 14.6 个/kb,而后,根据下载的水稻所有蛋白编码基因定位信息,计算水稻蛋白编码基因及其侧翼区的 SNP 数目,共得到 27 714 个 SNP 位点。在基因的 5'侧翼区、基因区、基因 3'侧翼区,其 SNP 出现的平均频率(每条基因及相邻侧翼区 SNP 数目除以相应基因长度,将各个区间所有数据的加和求平均值)分别为 14.8、12.6 和 14.7 个/kb。*t* 检验和 Wilcoxon 检验表明,SNP 在基因区的平均频率与所有侧翼区数据存在显著差异(表 2)。SNP 在基因侧翼区的平均频率与在水稻染色体上的平均分布频率相同,而 SNP 在基因区的频率显著小于水稻染色体平均分

表 1 Pre-miRNA 及其侧翼区的 SNP 平均频率
Table 1. Mean SNP frequency in pre-miRNA region and its flanking region.

区域 Region	SNP 数目 Number of SNPs	SNP 分布频率 Frequency of SNP		
		平均值 Mean	<i>t</i> 测验 <i>t</i> test	Wilcoxon 测验 Wilcoxon test
		/ (Number · kb ⁻¹)		
远端 5'侧翼区 Far 5' flanking region	998	17.30	0.9720	0.4839
近端 5'侧翼区 5' flanking region	836	14.76	0.6044	0.1067
Pre-miRNA	885	14.83		
近端 3'侧翼区 3' flanking region	838	14.23	0.7997	0.7367
远端 3'侧翼区 Far 3' flanking region	761	12.97	0.8192	0.2041

每两组 SNP 平均频率之间使用 *t* 检验和 Wilcoxon 检验检测。表中的数值为 *P* 值,来自 pre-miRNA 区 SNP 平均频率与侧翼区之间的显著性差异检验。

t test and Wilcoxon test were used to compare two means of SNP frequency. Figures in the table are *P*-values calculated by SNP frequency of pre-miRNA region compared with those of its flanking regions.

表 2 水稻基因区及其侧翼区的 SNP 平均频率
Table 2. Mean SNP frequency in rice gene region and its flanking region.

区域 Region	SNP 平均频率 Mean SNP frequency	<i>t</i> 检验 <i>t</i> test	Wilcoxon 检验 Wilcoxon test
基因 5'侧翼区 5' flanking region	14. 8	8.37e^{-8}	$< 2.2\text{e}^{-16}$
基因 3'侧翼区 3' flanking region	14. 7	3.58e^{-7}	$< 2.2\text{e}^{-16}$
基因区 Gene region	12. 6		

表中的数值为 *P* 值。
Data in the table are *P* values.

布。以上结果表明,与其他物种相似,水稻基因区 SNP 频率显著低于基因侧翼区。

2.3 水稻中 SNP 位点、miRNA 基因及蛋白编码基因在染色体上的分布

水稻 SNP 在基因组上分布不均衡,存在富集区^[9, 20]。本研究为了检验 SNP 频率在 miRNA 基因区的分布特点与在水稻基因组上的分布是否相关,进一步研究了 pre-miRNA 和 SNP 富集区在染色体上的分布特点。根据水稻不同的染色体长度,以 100 kb 为一个长度单位,计算水稻所有染色体上 SNP 位点和 pre-miRNA 数目。发现水稻染色体上 SNP 的分布不均衡,密集的区间 SNP 数目高达 25 831 个,分布密度最小的区间只有 4 182 个。这与在水稻不同品种 SNP 分布频率研究中报道的每条染色体上 SNP 多态性高低呈显著不平衡分布,基因间 SNP 数目占 SNP 总数一半以上的结果相互印证^[9]。另外,pre-miRNA 富集区与 SNP 富集区在水稻 12 条染色体上的分布不存在显著相关性(图 1),蛋白编码基因与 SNP 除第 4、5、8、9 和 12 染色体之外,也不存在显著相关性(第 4、5、8、9 和 12 染色体, *R* 值分别为 0. 26、0. 19、0. 23、0. 27 和 0. 25, *P* 值分别为 6.028e^{-5} 、0. 004742、0. 0004468、 2.779e^{-5} 和 0. 0001662)。

2.4 水稻 SNP 在 pre-miRNA 二级结构不同区间的分布

由于 pre-miRNA 不同的二级结构区间承担的功能不同,并且受到的选择压力也不同,SNP 在 pre-miRNA 不同二级结构区间的分布存在差异^[21, 23-24]。进一步检测 SNP 在每条 pre-miRNA 中不同二级结构区间的分布特点。通过计算 pre-miRNA 区的 SNP 位点,共在 412 条 pre-miRNA 得到 885 个 SNP 位点。其中,在种子区、成熟区、茎区和环区的总个数分别为 39 个、131 个、622 个和 263 个,其对应的平均频率分别为 13. 5、12. 6、14. 0 和 16. 6 个/kb(图 2),各区之间平均频率没有显著差

异。

3 讨论

在人^[21]和拟南芥^[22]中,SNP 在 pre-miRNA 区的频率显著低于其侧翼区。这表明侧翼区在进化过程中所受到的功能限制较少,相对 pre-miRNA 区承受较少纯化选择。本研究通过分析 SNP 在水稻 miRNA 基因中的分布特点,发现 SNP 在 pre-miRNA 区与其侧翼区的分布没有显著差异。这与先前结果不同。为检验 SNP 在水稻蛋白编码基因的分布是否与其他物种也存在差异,对 SNP 在蛋白编码基因的分布做了进一步分析。结果显示,SNP 在蛋白编码基因区的频率显著小于侧翼区,与其他物种 SNP 分布相似。

本研究中,水稻 pre-miRNA 的平均长度与蛋白编码基因平均长度存在较大差异(pre-miRNA, 150 bp; 蛋白编码基因, 3402 bp)。为检验两者长度差别是否是造成 SNP 分布差异的原因,用 pre-miRNA 长度随机在蛋白质编码基因的 5'侧翼区、基因区、3'侧翼区截取三段序列(分别称为随机 5'侧翼区、随机基因内部区和随机 3'侧翼区),统计 SNP 频率在三段序列中的差异。结果发现 SNP 在蛋白编码基因的随机 5'侧翼区,随机基因内部区和随机 3'侧翼区的总个数分别为 63 415 个、54 980 个和 61 685 个,对应的频率分别为 18. 7、16. 1 和 18. 1 个/kb。统计学检验表明,SNP 在随机基因内部区的平均频率显著低于随机 5'侧翼区和随机 3'侧翼区(表 3)。因此,基因长度差别并非是 SNP 在 pre-miRNA 与蛋白编码基因之间存在分布差异的原因。

在上述研究中,蛋白编码基因数目远远大于 pre-miRNA 数目。为检验两组数据之间的样本量差异是否是造成 SNP 分布差异的因素之一,以 pre-miRNA 数目(412 条)为样本量,在蛋白编码基因数据(表 2)中随机抽样,检验 SNP 频率在蛋白编码基因的基因区和侧翼区是否仍具有显著差异。抽样次

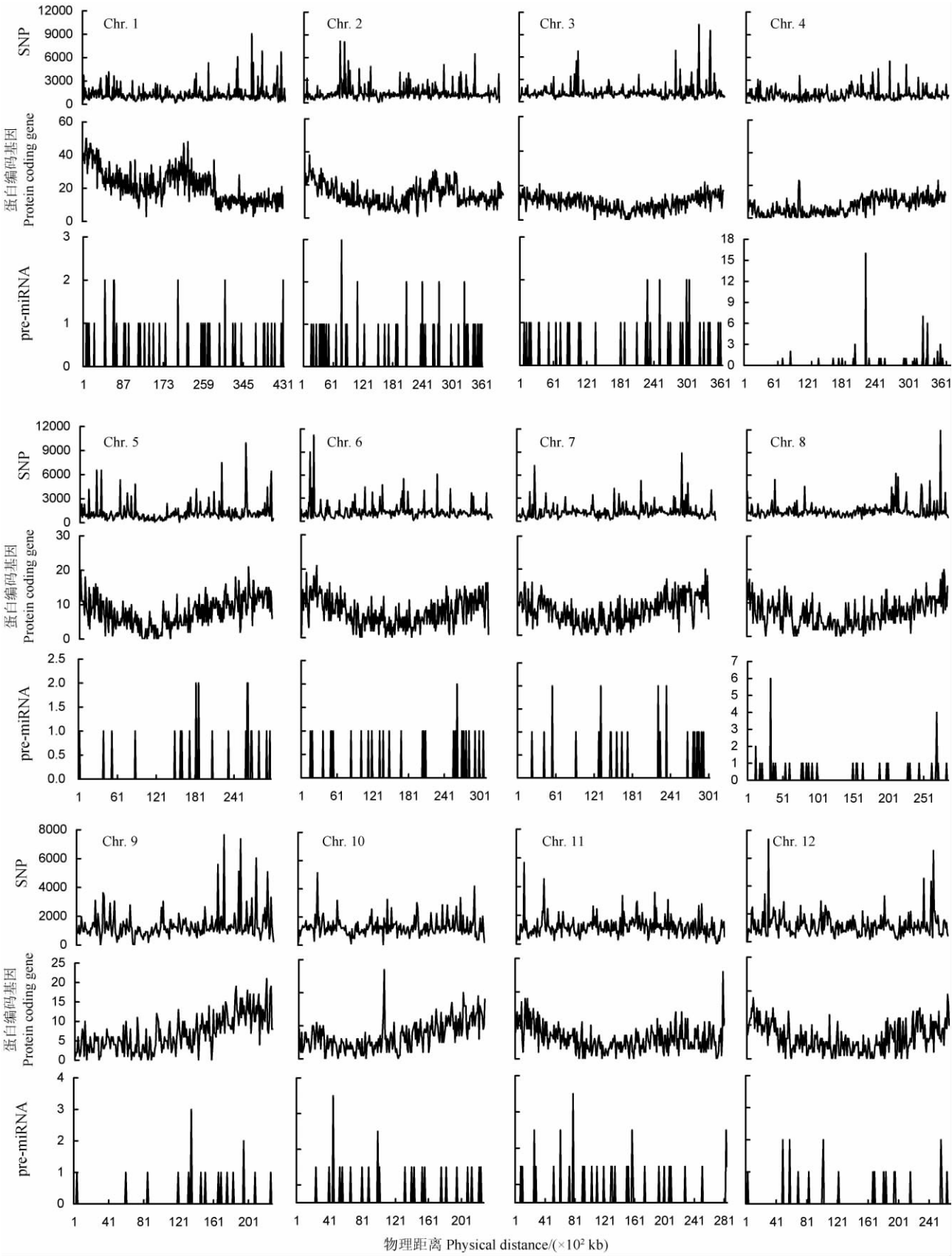


图 1 SNP、pre-miRNA 和蛋白编码基因在水稻染色体上的分布

Fig. 1 The distribution of pre-miRNA, protein-coding gene and SNP across chromosomes.

表 3 水稻随机基因内部区及其侧翼区的 SNP 平均频率的统计学检测

Table 3. Statistics test of mean SNP number in random rice gene region with those of its flanking region.

区域	SNP 数目	SNP 平均频率	t 检验	Wilcoxon 检验
Region	Number of SNPs	Mean SNP frequency	t test	Wilcoxon test
基因 5'侧翼区 5' flanking region	63 415	18.7	7.425e ⁻⁹	< 2.2e ⁻¹⁶
基因 3'侧翼区 3' flanking region	61 685	18.1	5.370e ⁻⁶	< 2.2e ⁻¹⁶
基因区 Gene region	54 980	16.1		

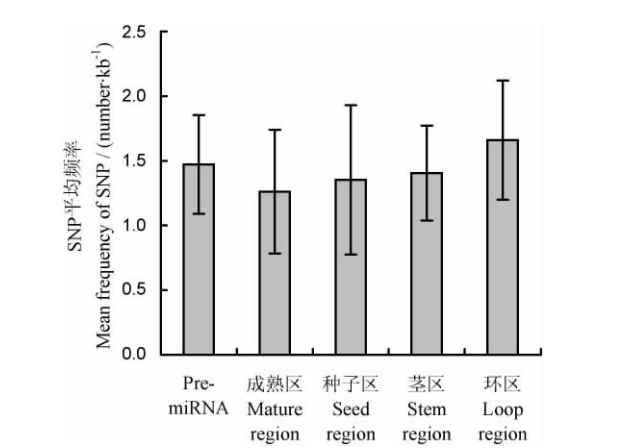


图 2 Pre-miRNA 不同二级结构的 SNP 平均频率

Fig. 2. Mean frequencies of SNPs across different pre-miRNA secondary structure regions.

数设为 1000 和 10000 次,结果在 1000 次随机抽样中,有 997 次存在显著差异;而在 10000 次随机抽样中有 9962 次存在显著差异(表 4)。同样用 412 为样本量,从表 3 数据中抽样。蛋白编码基因的 5'侧翼区、基因区和 3'侧翼区分别来自于对原有区间的随机取样,取样区间长度为 pre-miRNA 长度。因此,抽样结果可以检验样本量及序列长短对 SNP 分布是否具有影响。结果发现,SNP 在基因区和侧翼区的分布存在显著差异的样本数目远远大于无显著差异的样本数目。所以,SNP 在 pre-miRNA 与蛋白编码基因之间的分布差异并非是由于样本量太小和基因长度差异所导致。

由于 SNP 在水稻染色体上分布不均一,存在明显的稀缺区和富集区,同时 SNP 在不同染色体上分布也不同,为了检验 SNP 在 pre-miRNA 的分布特点是否与此有关,进一步分析了水稻 SNP、pre-miRNA 和蛋白编码基因在染色体上的分布特点。结果表明,尽管 SNP、pre-miRNA 和蛋白编码基因在不同染色体上都呈不均一分布,但是 SNP 富集区与 pre-miRNA 富集区在染色体上的分布不具明显相关性。因此,SNP 在 pre-miRNA 及其侧翼区分布无显著差异这一结果与 SNP 在染色体上的不均

表 4 在 1000 和 10000 次随机取样中,蛋白编码基因 SNP 平均频率显著小于其侧翼区的样本数(样本大小为 412)

Table 4. Number of samples (sample size 412) with the mean SNP frequency of gene regions significantly lower than that of the flanking region.

样本 Sample	1000	10000
基因 Gene	997	9962
随机基因内部区 Random gene region	924	9193

“基因”从表 2 数据中随机选取,“随机基因内部区”从表 3 数据中随机取样。对于每个随机样本,计算基因区与侧翼区 SNP 平均频率显著性差异。表中数字代表基因区 SNP 平均频率显著低于侧翼区 SNP 平均频率的样本数目。

‘Gene’ is the samples which were randomly chosen from the data in Table 2. ‘Random gene region’ is the samples which were randomly chosen from the data in Table 3.

匀分布无关。

水稻 pre-miRNA 与其侧翼区 SNP 频率无显著差异,表明植物 miRNA 成熟区对 miRNA 功能具有重要影响^[25],在自然选择过程中受较强纯化选择,因而 SNP 在 pre-miRNA 成熟区的频率小于二级结构其他区域。本研究发现,与其他生物相似,水稻 SNP 在 pre-miRNA 成熟区的频率低于其他二级结构区。这表明水稻 pre-miRNA 成熟区因需要与靶基因识别而承受较多选择压力。但是这种差异不具有统计学意义。

研究表明,水稻 SNP 在染色体不同区间的分布与其他物种有较大差异。在多数物种中,基因组中重组频率低的区间(比如着丝粒区)SNP 频率也低,然而在水稻基因组中,重组频率低的区间 SNP 频率却没有随之降低^[20]。除此之外,水稻 SNP 在不同基因中的分布特点也与其他物种有差异。在水稻和同属禾本科的高粱中,在它们共有的保守基因之间水稻基因 SNP 频率与高粱基因 SNP 频率不具有明显相关性^[20]。因此,水稻 SNP 在 miRNA 基因的分布特点,可能是水稻 SNP 独特分布特点之一。说明跟其他生物不同,水稻 pre-miRNA 侧翼区功能对 miRNA 非常重要,因而在进化上受较强自然选择。进一步研究 pre-miRNA 侧翼区的模体或转录因子

结合位点,将有助于揭示水稻 miRNA 转录调控和成熟过程。而某些 miRNA 在成熟区也积累了大量 SNP,表明这些 miRNA 成熟区功能对 miRNA 不再重要,这些 miRNA 可能逐渐丢失功能或进化出新的功能。

我们发现水稻 SNP 在不同 pre-miRNA 间分布不均衡。近一半的 pre-miRNA 不含有 SNP,而少数水稻 miRNA 具有极高的 SNP 频率。例如 miRNA osa-MIR1871、osa-MIR2931 和 osa-MIR1428f 的 SNP 数目超过 40。这可能是导致 SNP 在不同 pre-miRNA 分布方差过大的原因,进而导致 SNP 频率在 pre-miRNA 与侧翼区之间、不同 pre-miRNA 二级结构区之间的差异不具有统计学差异。

水稻 SNP 在不同 pre-miRNA 之间的不均衡分布表明,一部分 miRNA 基因功能对水稻非常重要,在进化上受较强纯化选择,而另一部分基因由于各种原因,可能正经历纯化选择放松,导致基因功能逐渐丧失或者产生新的基因功能。水稻作为重要粮食作物,经历了漫长的人工驯化改良过程。这部分 miRNA 基因可能是因为在栽培环境下,对水稻的生存已非必要因素,因纯化选择放松而经历快速进化。另外,致使 miRNA 基因功能丢失的突变如果与改良性状相关,也容易在进化过程中受到人工选择作用。通常情况下,重复基因由于功能冗余而经历纯化选择放松。在水稻 412 条 pre-miRNA 中,有 286 条属于多基因家族,126 条属于单基因家族。然而多基因家族的 SNP 平均频率反而比单基因家族的低。这表明水稻部分 miRNA 基因的纯化选择放松不是因为基因重复造成的。

另外,由于水稻 SNP 数据绝大部分没有经过二次实验验证,可能含有大量不可靠 SNP 数据^[20]。也不排除不可靠 SNP 数据导致 SNP 在不同 miRNA 基因之间呈现不均衡分布。

谢辞:香港中文大学邵建林博士对本研究提出许多宝贵建议,特此致谢!

参考文献:

- [1] 金龙国,王 川,刘进元. 植物 MicroRNA. 中国生物化学与分子生物学报, 2006, 22(8): 609-614.
- [2] Reinhar B J, Weinstein E G, Rhoades M W, et al. MicroRNAs in plants. *Genes Dev*, 2002, 16(13): 1616-1626.
- [3] Bartel B, Bartel D P. MicroRNAs: At the root of plant development. *Plant Physiol*, 2003, 132(2): 709-717.
- [4] Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet*, 2006, 38: 1452-1456.
- [5] Park M Y, Wu G, Gonzalez-Sulser A, et al. Nuclear processing and export of microRNAs in *Arabidopsis*. *Proc Natl Acad Sci USA*, 2005, 102: 3691-3696.
- [6] Guo H S, Xie Q, Fei J F, et al. MicroRNA164 directs NAC1 mRNA cleavage to downregulate auxin signals for lateral root development. *Plant Cell*, 2005, 17(5): 1376-1386.
- [7] Bao N L, Ye K W, Barton M K. MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Dev Cell*, 2004, 7: 653-662.
- [8] Jiang M C, Jiang P C, Liao C F, et al. A modified mutation detection method for large-scale cloning of the possible single nucleotide polymorphism sequences. *Biochem Mol Biol*, 2005, 38(2): 191-197.
- [9] Nasu S, Suzuki J, Ohta R, et al. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Res*, 2002, 9: 163-171.
- [10] Monnal L, Kitazawa N, Yoshino R, et al. Positional cloning of rice semidwarfing gene, *sd-1*: Rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res*, 2002, 9(1): 11-17.
- [11] Shen Y J, Jiang H, Jin J P, et al. Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol*, 2004, 135: 1198-1205.
- [12] Jin Q S, Waters D, Cordeiro G M, et al. A single nucleotide polymorphism (SNP) marker linked to the fragrance gene in rice (*Oryza sativa* L.). *Plant Sci*, 2003, 165: 359-364.
- [13] Umenmoto T, Aoki N. Single-nucleotide polymorphisms in rice *starch synthase IIa* that alter starch gelatinisation and starch association of the enzyme. *Funct Plant Biol*, 2005, 32(9): 763-768.
- [14] Hayashi K, Hashimoto N, Daigen M, et al. Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus. *Theor Appl Genet*, 2004, 108: 1212-1220.
- [15] Yamanaka S, Nakamura I, Watanabe K N, et al. Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. *Theor Appl Genet*, 2004, 108: 1200-1204.
- [16] Kanazin V, Talbert H, See D, et al. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). *Plant Mol Biol*, 2002, 48: 529-537.
- [17] Zhu Y L, Song Q J, Hyten D L, et al. Single-nucleotide polymorphisms in soybean. *Genetics*, 2003, 163: 1123-1134.
- [18] Batley J, Barker G, O'Sullivan H, et al. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiol*, 2003, 132: 84-91.
- [19] Morales M, Roig E, Monforte A J, et al. Single-nucleotide polymorphisms detected in expressed sequence tags of melon

(*Cucumis melo* L.). *Genome*, 2004, 47: 352-360.

[20] Feltus F A, Wan J, Schulze S R, et al. An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. *Genome Res*, 2004, 14(9): 1812-1819.

[21] Saunder A M, Liang H, Li W H. Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci USA*, 2007, 104(9): 3300-3305.

[22] Ehrenreich M I, Purugganan D M. Sequence variation of microRNAs and their binding sites in *Arabidopsis*. *Plant Physiology*, 2008, 146: 1974-1982.

[23] Landi D, Gemignani F, Barale R, et al. A catalog of polymorphisms falling in microRNA-binding regions of cancer genes. *DNA & Cell Biol*, 2008, 27(1): 35-43.

[24] Duan R H, Pak C H, Jin P. Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. *Hum Mol Genet*, 2007, 16(9): 1124-1131.

[25] Schwach F, Moxon S, Moulton V, et al. Deciphering the diversity of small RNAs in plants: The long and short of it. *Oxford J*, 2009, 8(6): 472-481.

科学出版社图书介绍

现代蛋白质工程实验指南

现代蛋白质工程是将分子生物学、蛋白质结构与功能分析、理论计算以及生物化学有机结合的学科,其目标是快速及高效率地发展和改进实用的或有价值的蛋白质。本书分为两个部分,第一部分主要介绍了蛋白质理性设计的策略,包括理论计算方法,利用一些很有说服力的例证说明所设计蛋白质的全新特性,阐述了如何设计具有目标特性的蛋白质,并选择了很多如蛋白质-蛋白质相互作用、DNA 结合、抗体模拟以及酶活性设计等具体实例;第二部分蛋白质的定向进化技术主要介绍了包括进化库设计的一般方法、进化库质量的统计评估、核酸混编的新方法,以及不同的选择筛选策略等。同时也给出了不同特性体外定向进化的一些实例,如蛋白质折叠类型、折叠热稳定性以及酶活性等。

本书适合蛋白质科学各个层次的科研工作者,特别是从事相关领域研究的高年级大学生及研究生参考使用。该书已于 2011 年 2 月出版,定价 58.00 元。



农田集雨保水关键技术研究

本书是依托国家旱农项目,在旱作农田集水保水技术多年研究工作的基础上,对研究工作的系统性和阶段性总结,主要内容包括:有机培肥、休闲轮耕、秸秆覆盖种植、粮草带状间作、微集水种植、可降解膜覆盖保墒等旱作技术对农田土壤理化性状及作物生产力的影响。本书的出版对丰富旱作农田集水保水研究理论与实践有重要的意义,对同类农业研究与生产示范具有重要的借鉴与参考价值。

本书可为从事旱地农业研究、节水农业技术开发、农业资源高效利用方面的教学和研究人员提供参考。该书已于 2011 年 1 月出版,定价 98.00 元。



中国至 2050 年农业科技发展路线图(英文版)

本书明确了至 2050 年全球和我国农业发展面临的挑战与机遇,预测我国未来对农业科技的重大需求,提出至 2050 年我国农业科技领域发展战略目标、分阶段目标,提出各阶段农业科技发展的主要方向及可突破的重大科学技术问题,形成未来农业科技发展的总体路线图,并提出为实现以上目标所需要的体制、资源、人才等方面的政策建议。

本报告可作为政府部门、科研机构、大学、企业进行科技战略决策的重要参考,也可供国内外专家、学者研究和参考。该书已于 2011 年 1 月出版,定价 108.00 元。

以上图书均由科学出版社出版,欢迎邮购。联系人:科学出版社科学销售中心 周文宇;电话:010-64031535;E-mail:zhouwenyu@mail.sciencep.com;网上订购: <http://www.shop.sciencepress.cn>。更多精彩图书请登陆网站,欢迎致电索要书目。

