

籼稻基因组中的 tRNA 研究*

王希胤^{①②③†} 史晓黎^{①②†} 郝柏林^{②④**}

(① 北京大学生命科学学院, 北京 100871; ② 中国科学院基因组信息中心暨北京华大基因研究中心, 北京 101300; ③ 河北理工学院数理系, 唐山 063000; ④ 复旦大学理论生命科学研究, 上海 200433)

摘要 在籼稻基因组的 127551 个重叠群中, 共发现了 596 个 tRNA 基因, 其中有 3 个携带硒代半胱氨酸的 tRNA 基因和一个抑制型 tRNA 基因. 除了上述 4 个特殊的 tRNA 基因外, 592 个 tRNA 可按密码子分成 45 种, Guthrie 等人提出的修正摆动假设完全成立. 在这些重叠群中, 还发现了 27 个假基因, 但未发现 tmT-CGU 基因, 这可能是由于现有序列数据的不完备. 在上述 592 个 tRNA 基因中, 可能存在冗余, 而且还可能再有新发现. 研究了水稻基因中密码子的使用与 tRNA 基因的数目之间的关系. 从水稻的两个栽培亚种——籼稻和粳稻的叶绿体基因组中, 各发现了 33 个 tRNA 基因, 经多重联配发现, 两组 tRNA 基因完全相同.

关键词 tRNA 基因 摆动假设 密码子偏好 tRNA 假基因 水稻 叶绿体

水稻是人类最主要的谷物之一, 早在 8000 年前中国的江淮地区就有栽植^[1]. 近来, 水稻基因组的测序工作取得了很大进展^[2~4], 不久之后, 将获得水稻基因组的完成图.

tRNA 是一种重要的不编码蛋白的 RNA, 在蛋白质的合成中起关键作用, 它通过反密码子与 mRNA 的密码子之间的相互作用, 读取贮存在 mRNA 中的遗传信息, 完成把各个氨基酸链接成多肽的工作. 共有 61 种编码的密码子, 但并不是相对应地有 61 种带有不同反密码子的 tRNA. Crick 提出了摆动假设^[5]: 通过非 Watson-Crick 碱基配对规则, 只需要较少种类的 tRNA 就可以完成翻译功能. Guthrie 等人^[6]修正了摆动假设, 提出了新的碱基配对规则(表 1): 在一个通常的遗传密码表中, 前两位碱基相同的四个密码子, 如果第 3 位为 U 或 C 的两个密码子与第 3 位为 A 或 G 的两个密码子编码不同的氨基酸, 则称这四个密码子处于一个“二密码格”中; 如果这四个密码子编码相同的氨基酸, 则称它们处于一个“四密码格”中; 在“二密

表 1 Guthrie 等人修正的真核生物中的摆动假设

	密码子第 3 位碱基	反密码子第 1 位碱基
任何密码格	A	U ^{a)}
任何密码格	G	C
二密码格	U&C	G
四密码格	U&C	A ^{b)}
甘氨酸的四密码格	U&C	G

a) 一般地, U 经转录后修饰为其他碱基, 如假尿嘧啶碱基; b) A 转录后几乎一定修饰为 I (次黄嘌呤)

2002-07-30 收稿

* 国家重点基础研究发展规划(批准号: G2000077300)和北京市 248 创新计划资助项目

** 联系人, E-mail: hao@itp.ac.cn

† 同等贡献

码格”中，第 3 位为 U 或 C 的两个密码子都被第 1 位为 G 的反密码子识别；在“四密码格”中，第 3 位为 U 或 C 的两个密码子都被第 1 位为 A 的反密码子识别；只有一个例外，就是甘氨酸所在的“四密码格”中，第 3 位为 U 或 C 的两个密码子都被第 1 位为 G 的反密码子识别，而不是像其他“四密码格”被第 1 位为 A 的反密码子识别；无论在“二密码格”还是“四密码格”中，第 3 位为 A 的密码子都被第 1 位为 U 的反密码子识别，第 3 位为 G 的密码子都被第 1 位为 C 的反密码子识别。

按照修正的碱基配对规则，Guthrie 等人预测，在真核生物中，只要 46 种 tRNA 就足够完成相应的生物功能。已获得完全基因组的真核生物，如人、拟南芥和线虫等，其 tRNA 基因都相当好地服从修正的摆动假设(表 2)^[7](<http://rna.wustl.edu/GtRDB/>)。不过，在上述每种真核生物中，都有 1~3 种违背摆动假设的 tRNA 基因，总共 7 个，在表 2 中用斜体 1 表示。这些例外可能是由于测序错误造成，也可能是假基因。我们的工作就是要阐明在水稻的基因组中有多少 tRNA 基因，修正的摆动假设是否成立。

表 2 3 个真核生物的 tRNA 基因的数目^{a)}

AA 密码子	A	B	C	反密码子	AA 密码子	A	B	C	反密码子	AA 密码子	A	B	C	反密码子	AA 密码子	A	B	C	反密码子						
F	UUU	0	0	0	AAA	S	UCU	37	14	10	AGA	Y	UAU	0	0	1	AUA	C	UGU	0	0	0	ACA		
	UUC	16	16	14	GAA		UCC	1	0	0	GGA		UAC	76	19	11	GUA		UGC	15	13	30	GCA		
	UUA	6	5	8	UAA		UCA	9	7	5	UGA		*	UAA	0	0	1		UUA	*-UGA	0	0	0	UCA	
	UUG	10	7	6	CAA		UCG	4	5	4	CGA			UAG	0	0	1		CUA		W-UGG	14	11	7	CCA
L	CUU	11	18	13	AAG	P	CCU	16	6	11	AGG	H	CAU	0	0	0	AUG	R	CGU	9	18	9	ACG		
	CUC	1	0	0	GAG		CCC	0	0	0	GGG		CAC	10	17	12	GUG		CGC	0	1	0	GCG		
	CUA	10	3	2	UAG		CCA	39	34	10	UGG		Q	CAA	8	18	11		UUG	CGA	6	10	7	UCG	
	CUG	3	5	6	CAG		CCG	5	3	4	CGG			CAG	9	7	21		CUG	CGG	4	3	5	CCG	
I	AUU	20	19	13	AAU	T	ACU	10	17	8	AGU	N	AAU	0	0	1	AUU	S	AGU	0	0	0	ACU		
	AUC	0	0	1	GAU		ACC	0	0	0	GGU		AAC	16	20	33	GUU		AGC	13	9	7	GCU		
	AUA	5	8	5	UAU		ACA	8	11	10	UGU		K	AAA	13	16	16		UUU	R	AGA	9	7	5	UCU
	M-AUG	23	20	17	CAU		ACG	6	7	7	CGU			AAG	18	33	22		CUU		AGG	8	3	4	CCU
V	GUU	15	19	20	AAC	A	GCU	16	21	25	AGC	D	GAU	0	0	0	AUC	G	GGU	1	0	0	ACC		
	GUC	0	0	0	GAC		GCC	0	0	0	GGC		GAC	23	22	10	GUC		GGC	23	14	11	GCC		
	GUA	7	6	5	UAC		GCA	10	10	10	UGC		E	GAA	12	17	14		UUC	GGA	12	33	5	UCC	
	GUG	8	5	19	CAC		GCG	7	4	5	CGC			GAG	13	20	8		CUC	GGG	5	3	8	CCC	

a) 表头中 AA 示氨基酸, A 示拟南芥, B 示线虫, C 示人; 表内*示终止子

假基因是与基因突变相关的序列，在许多生物基因组中也发现有假 tRNA 基因^[7](<http://rna.wustl.edu/GtRDB/>)，即那些在序列上与 tRNA 基因近似，但由于碱基的插入、缺失或缺乏有功能的启动子而失去功能的序列。另一种与 tRNA 基因突变有关的是抑制型 tRNA 基因：一个正常的 tRNA 基因的反密码子的碱基发生了替换，使得 tRNA 基因不再识别其相应的

密码子, 而是识别终止密码子. 另外还有携带硒代半胱氨酸的 tRNA 基因, 它识别一个终止密码子 UGA. 上述几种特殊的 tRNA 基因在籼稻中也存在.

基因组中密码子的使用频率常有显著的差异, 这种差异往往用于寻找基因. 密码子偏好是指编码同一氨基酸的几个密码子的使用频率的差异. 我们试图阐明在水稻的基因组中密码子偏好与 tRNA 基因在基因组中的个数之间的联系.

水稻有两个栽培亚种: 籼稻和粳稻. 已经测得了这两个亚种的叶绿体基因组序列. 我们提取了它们的 tRNA 基因序列并做了比较.

1 材料与方法

籼稻的所有 127551 个重叠群见网站(<http://btn.genomics.org.cn/rice>). 籼稻叶绿体基因组的序列由中国科学院基因组信息中心暨北京华大基因研究中心测定. 粳稻的叶绿体基因组序列^[8]由 GenBank 数据库(<http://www.ncbi.nlm.nih.gov/>)提取.

我们使用了 3 个软件辅助完成研究工作. 用 tRNAscan-SE 软件^[9,10], 从重叠群和叶绿体基因组中找到一组可能的 tRNA 基因. 为了提高搜索到的 tRNA 基因的可信程度, 将这些候选序列作为查询序列用 BLASTN^[11]对 CBI(<http://www.cbi.pku.edu.cn/>)中的非冗余(nr)库进行搜索. 用 ClustalW 软件^[12]做了多重联配.

2 结果与讨论

2.1 tRNA 基因的分类

用 tRNAscan-SE 软件在水稻重叠群中找到了 881 个可能的 tRNA 基因. 根据 BLASTN 的分值和一致性, 可以从中选出 596 个 tRNA 基因, 除 4 个特殊类型的 tRNA 基因(见第 2.5 小节)外, 将 592 个 tRNA 基因分为 3 类: BLASTN 确定的 tRNA 基因、可能的新的 tRNA 基因、推测的新 tRNA 基因. BLASTN 分值高于 115 分的称为 BLASTN 确定的 tRNA 基因; 可能的新的 tRNA 基因指 BLASTN 分值在 100~115 之间的基因; 推测的新 tRNA 基因的 BLASTN 分值在 80~100 之间.

共有 467 个 BLASTN 确定的 tRNA 基因, 它们是最可信的. 对于 74 个可能的新的 tRNA 基因和 51 个推测的新 tRNA 基因, 也有相当大的把握认为其中大部分是真正的 tRNA 基因. 后两类基因未能和非冗余数据库中的其他物种的 tRNA 基因很好地联配, 它们可能是水稻特有的 tRNA 基因.

TIGR 数据库(<http://www.tigr.org/>)中列出了粳稻 596 个 tRNA 基因, 在数量上与籼稻很接近.

2.2 摆动假说

水稻基因组中的 592 个 tRNA 基因(除 4 个特殊的以外)可分为 45 种, 它们很好地符合由 Guthrie 等人提出的修正摆动假说(表 3). 尽管 trnT-CGU 基因未在现有的重叠群中发现, 我们不认为这是对修正的摆动假设的违背(见第 2.7 小节). 在表 3 中, 20 种氨基酸用黑体字给出, 对于每一个密码子, 依次给出了密码子的三联码、其在每 10000 个密码子中的出现频率、与其对应的反密码子的 tRNA 基因的数目、反密码子的三联码. 另外, 配对规则用连线表示. 其中密码子使用频率的数据由 Codon Usage Database(<http://www.kazusa.or.jp/codon/>)提取, 后者共计数了籼稻的 27910 个密码子.

2.3 密码子使用和 tRNA 基因

我们研究了密码子偏好性以及密码子使用和相应 tRNA 基因份数的关系(表 3). 密码子 GAG (393), GGC (240), AAG (325), GAC (292) 和 GCC (279) 的使用频率显著高于密码子 CUA (83), CGA (77), GUA (66), UGU (61) 和 UUA (59). 与以 U 或 A 结尾的密码子相比较, 水

表 3 水稻的密码子和相应的 tRNA 基因^{a)}

F	UUU 137	0 AAA	S	UCU 100	17 AGA	Y	UAU 102	0 AUA	C	UGU 61	0 ACA
	UUC 262	15 GAA		UCC 150	0 GGA		UAC 159	16 GUA		UGC 128	10 GCA
L	UUA 59	7 UAA	P	UCA 110	10 UGA	H	UAA 6	0 UUA	R	-UGA 10	0 UCA
	UUG 152	9 CAA		UCG 115	7 CGA		UAG 3	0 CUA		W-UGG 138	12 CCA
I	CUU 153	19 AAG	T	CCU 114	16 AGG	Q	CAU 120	0 AUG	S	CGU 73	16 ACG
	CUC 276	0 GAG		CCC 106	0 GGG		CAC 157	11 GUG		CGC 141	0 GCG
M	CUA 83	8 UAG	A	CCA 144	11 UGG	N	CAA 124	16 UUG	R	CGA 77	4 UCG
	CUG 216	6 CAG		CCG 153	10 CGG		CAG 225	13 CUG		CGG 106	7 CCG
V	AUU 140	23 AAU	G	ACU 105	9 AGU	K	AAU 134	0 AUU	S	AGU 72	0 ACU
	AUC 229	0 GAU		ACC 161	0 GGU		AAC 198	14 GUU		AGC 166	13 GCU
W	AUA 89	6 UAU	D	ACA 120	8 UGU	E	AAA 144	10 UUU	G	AGA 97	9 UCU
	AUG 249	27 CAU		ACG 113	0 CGU		AAG 325	22 CUU		AGG 142	10 CCU
X	GUU 171	21 AAC	E	GCU 187	25 AGC	F	GAU 241	0 AUC	G	GGU 155	0 ACC
	GUC 223	0 GAC		GCC 279	0 GGC		GAC 292	28 GUC		GGC 340	24 GCC
Z	GUA 66	4 UAC	G	GCA 196	11 UGC	G	GAA 205	15 UUC		GGA 159	13 UCC
	GUG 226	10 CAC		GCG 264	13 CGC		GAG 393	29 CUC		GGG 158	8 CCC

a) 同表 2

稻更倾向使用以 C 或 G 结尾的密码子. 若将前两位字母相同的密码子分为一组, 可得到 16 组

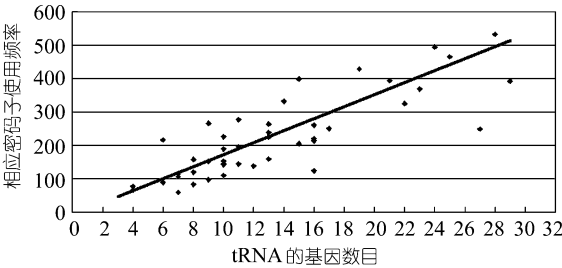


图 1 tRNA 的基因数目与相应的密码子使用频率的关系

密码子. 不考虑 3 个终止密码子和编码色氨酸的 UGG, 在其中的 15 组中, 第 3 位为 C 的密码子使用数目大于第 3 位为 U 的密码子. 在其中的 14 组中, 第 3 位为 G 的密码子使用数目大于第三位为 A 的密码子. 仅有的两对例外分别为 CCU, CCC 和 ACA, ACG.

此外, 密码子使用频率和相应的 tRNA 基因份数存在大致的正相关(图 1), 这与大多数生物是一致的.

按密码子第 3 位碱基将所有密码子分为 3 类 (将第 3 位碱基为 U 和 C 的密码子分为一类), 计算每类中所有的密码子使用频率的和与相应的 tRNA 基因数目的和的比值(表 4). 以 U 和 C 结尾的密码子对应的比值大于以 A 或 G 结尾的密码子的比值, 以 G 结尾的密码子对应的比值大于以 A 结尾的密码子的比值. 一种可能

的解释是与摆动相关的密码子的 tRNA 基因具有更高的转录效率或成熟的 tRNA 可更加有效的执行生物学功能. 对于以 G 和 A 结尾的密码子比值的差异也可以类似地解释.

表 4 tRNA 基因的效率

	密码子频率	相应的 tRNA 基因数	密码子频率/ tRNA 基因数
第 3 位碱基为 U 和 C 的密码子	5332	277	19.25
第 3 位碱基为 A 的密码子	1678	132	12.67
第 3 位碱基为 G 的密码子	2978	183	16.26

2.4 假 tRNA 基因

用软件 tRNAscan-SE 在水稻的重叠群中发现了 71 个可能的假基因, 但经 BLASTN 在公共的非冗余数据库中搜索之后, 只有 27 个被确认与其他物种的 tRNA 基因有一定程度的相似性, 大部分被排除了是假基因的可能. 在这 27 个可能的假基因中, 16 个无法判定反密码子. 假基因不可能用单纯的计算方法确定, 最终需要实验确认.

2.5 携带硒代半胱氨酸的 tRNA 基因、抑制型 tRNA 基因

除 20 种典型的氨基酸之外, 自然界还有很多其他的氨基酸, 但迄今只发现硒代半胱氨酸等 2 种可被编码(另外一种是 pyrrolysine 即吡咯赖氨酸, 被 UAG 编码^[13,14]), 硒代半胱氨酸是由一个终止密码子 UGA 编码的. 在每个域的物种中, 都发现硒代半胱氨酸被编码. 在水稻重叠群中共发现 3 个携带硒代半胱氨酸的 tRNA 基因.

只发现一个可能的抑制型 tRNA 基因, 它识别终止密码子 UAA.

2.6 tRNA 基因中的内含子

共计 36 个 tRNA 基因含有内含子, 它们都只有 1 个内含子. 这些内含子的平均长度为 15 bp. trnS-AGA 基因中的内含子最长, 达 38 bp. 所有 trnY-GUA 基因都有 1 个内含子, 所有在蛋白质延伸过程中使用的 trnM-CAU 基因都有 1 个内含子, 所有和蛋白质合成起始相关的 trnM-CAU 基因都没有内含子. 惟一的抑制型 tRNA 基因包含 1 个内含子.

2.7 关于 trnT-CGU 的缺失

用 tRNAscan-SE 发现了 6 个可能的 trnT-CGU 基因, BLASTN 结果显示, 它们与数据库中的 tRNA 基因匹配很差, 从而否定了其作为 tRNA 基因的可能性. trnT-CGU 缺失可能是由于我们得到的水稻基因组数据并不完整. 现有的水稻重叠群共计 356 Mbp, 而水稻基因组预计有 464 Mbp. 加之 tRNA 基因在染色体上经常聚集成簇, 如在人类基因组中有 140 个 tRNA 基因(占全部 tRNA 基因的 25%)聚集在第 6 号染色体中只有 4 Mbp 的片段上, 水稻有 8 个 trnQ-UUG 基因聚集在同一条染色体上. 数据的不完备和 tRNA 成簇倾向应是目前未发现 trnT-CGU 基因的原因.

2.8 真核生物中 tRNA 基因的数量

真核生物中 tRNA 基因的数量与基因组的大小没有明显的相关性. 基因组的扩增并未导致 tRNA 基因的同步增长(表 5). tRNA 基因的数量与基因组大小的比值随着基因组的增大而迅速减小, 与编码区大小的比值也逐渐减小, 这说明在越高等的生物中 tRNA 基因在转录效率上或成熟的 tRNA 在功能上效率越高.

占水稻 tRNA 基因亚种的四分之一左右的 17 个亚种在拟南芥中有完全相同的亚种; 水稻和人没有完全相同的 tRNA 基因, 最相近的也有一个碱基发生突变; 对各物种内的 tRNA 基因的研究表明, 人和水稻的 tRNA 基因的碱基突变比较多, 而拟南芥的 tRNA 基因要保守得多;

表 5 7 个真核生物中的 tRNA 基因^{a)}

物种	tRNA 基因 数目	基因组大小/ Mbp	基因组中每百万碱基的 tRNA 基因数目	编码区大小/ Mbp	编码区每百万碱基 的 tRNA 基因数目
<i>S. cerevisiae</i>	273	12	22.75	8.45	32
<i>S. pombe</i>	174	14	12.43	6.9	25
<i>C. elegans</i>	584	100	5.84	26.1	22
<i>A. thaliana</i>	620	125	4.96	33.5	18
<i>D. melanogaster</i>	284	180	1.58	24.1	12
<i>O. sativa</i>	596	464	1.48	—	—
<i>H. sapiens</i>	648	3400	0.19	58.5(?)	11(?)

a) 数据部分源于文献[15,16]和网站 <http://rna.wustl.edu/GtRDB/>, <http://www.ncbi.nlm.nih.gov/>

不同的反密码子对应的 tRNA 基因的保守程度是有差异的, 把水稻、拟南芥和人的 tRNA 基因放在一起统计发现, trnP-AGG, trnP-CGG, trnA-AGC, trnA-CGC 基因最为保守, 而 trnE-CTC 和 trnK-TTT 基因突变最多。

2.9 水稻和拟南芥 tRNA 基因在基因组中分布的初步分析

由于籼稻重叠群还没有在染色体上定位, 无法进行籼稻和拟南芥 tRNA 基因分布的全基因组比较. 位于同一个重叠群上的多个 tRNA 基因可以显示 tRNA 基因分布的局部信息. 我们对重叠群中发现的全部 10 个包含 3 个和 3 个以上 tRNA 基因的基因簇进行了研究, 基因最多的一簇中有 8 个 trnQ-TTG, 这些基因的长度均为 71 个碱基, 基因间的距离大致在两千碱基左右. 在拟南芥的 5 条染色体中只有 1 条含有两个以上的 trnQ-TTG. 用 tRNAscan-SE 可在拟南芥第 1 号染色体上找到 4 个 trnQ-TTG, 这些基因的平均间隔大约为 800 万碱基. 对其他 9 类籼稻 tRNA 基因簇的分析表明在拟南芥基因组中没有类似的 tRNA 基因分布。

在拟南芥中发现了 27 个由 trnS-AGA, trnY-GTA 和 trnY-GTA 构成的基因簇, 每个基因簇的基因排列顺序一致, 在基因组中所占长度为几百碱基. 在籼稻的重叠群中未发现任何类似的结构单元. 目前研究表明, 在单子叶植物和双子叶植物分离后的漫长演化过程中, 籼稻和拟南芥的 tRNA 基因在染色体组中的组织和分布已经发生了巨大的变化。

2.10 籼稻和粳稻叶绿体基因组中的 tRNA 基因

用 tRNAscan-SE 在籼稻和粳稻的叶绿体基因组中各发现了 33 个 tRNA 基因. 经过 ClustalW 多重联配, 发现这两组 tRNA 基因完全相同. 这说明两个水稻亚种在经历了 7000 多年的分离后, 叶绿体基因组的 tRNA 基因没有发生突变. 值得注意的是, 粳稻的 tRNA 基因是我们独立提取的, 与原来注释中的 tRNA 基因不完全相同。

先用 tRNAscan-SE 普查, 再借助 BLASTN 搜索来确认, 是从重叠群中寻找 tRNA 基因的有效方法. 我们获得的这组 tRNA 基因可能存在由于测序和拼接错误造成的冗余. 同时, 由于现有基因组数据的不完整性, 得到的这组 tRNA 可能是不完备的. 另外, 这些 tRNA 基因的真实性还需要实验验证。

致谢 感谢中国科学院基因组信息中心暨北京华大基因研究中心的倪培相及所有水稻基因组计划的参与者. 感谢北京大学生物信息中心, 这个工作得到了他们的支持。

参 考 文 献

- 1 张居中, 王象坤, 孔昭宸, 等. 河南贾湖稻作文化的发现与研究. 科学(上海), 2002, 54: 3~6
- 2 于 军, 胡松年, 王 俊, 等. 籼稻(*Oryza sativa* L. ssp. *indica*) 全基因组框架序列. 科学通报, 2001, 46(23): 1937~1941
- 3 Yu J, Hu S N, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science, 2002, 296: 79~92
- 4 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
- 5 Crick F. Codon-anticodon pairings: the wobble hypothesis. Journal of Molecular Biology, 1966, 19: 548
- 6 Guthrie C, Abelson J. Organization and expression of tRNA genes in *Saccharomyces cerevisiae*. In: Strathern J, et al. eds. The Molecular Biology of the Yeast *Saccharomyces*: Metabolism and Gene Expression. New York: Cold Spring Harbor Laboratory, 1982. 487
- 7 International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature, 2002, 409: 860~921
- 8 Hiratsuka J, Shimada H, Whittier R, et al. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Molecular General Genetics, 1989, 217 (2-3): 185
- 9 Eddy S R. Non-coding RNA genes and the modern RNA world. Nature Reviews Genetics, 2001, 2: 919~929
- 10 Lowe T M, Eddy S R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research, 1997, 25: 955~964
- 11 Altschul S F, Madden T L, Schaffer A A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 1997, 25: 3389~3402
- 12 Thompson J D, Higgins D G, Gibson T J. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 1994, 22: 4673~4680
- 13 Srinivasan G, James C M, Krzycki J A. Pyrrolysine encoded by UAG in archaea: Charging of a UAG-decoding specialized tRNA. Science, 2002, 296: 1459~1462
- 14 Hao B, Gong W M, Ferguson T K, et al. A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science, 2002, 296: 1462~1466
- 15 Adams M D, Celniker S E, Holt R A, et al. The genome sequence of *Drosophila melanogaster*. Science, 2000, 287: 2185~2195
- 16 Goffeau A, Barrell B G, Bussey H, et al. Life with 6000 genes. Science, 1996, 274: 546~567