The tRNA and rRNA Genes in the *Oryza sativa* Genome

WANG Xi-Yin[1,2,3], SHI Xiao-Li[1,2], HAO Bai-Lin[2,4,5,①]

(1. *Center of Bioinformatics*, *Peking University*, *Beijing* 100871, *China*;
2. *Beijing Genomic Institute*, *Chinese Academy of Sciences*, *Beijing* 101300, *China*;
3. *College of Mathematics*, *Heibei Polytechnic University*, *Tangshan* 063009, *China*;
4. *T-Life Research Center*, *Fudan University*, *Shanghai* 200433, *China*;
5. *Institute of Theoretical Physics*, *Chinese Academy of Sciences*, *Beijing* 100080, *China*)

Abstract: In the recently assembled genomes of rice *Oryza sativa* ssp. *indica* and *japonica*, we identified 564 and 519 tRNA genes, respectively. The modified wobble hypothesis, namely, at least 46 tRNA species must present in order to decode all 61 possible anticodons, is perfectly observed in both subspecies. Among the 46 tRNA species, *indica* and *japonica* have many identical ones in sequence. There are 18 rice tRNA species that have identical counterparts in *Arabidopsis*. In the *indica* superscaffold dataset, 384 5S rRNA genes, dozens of 17S and 5.8S rRNA genes and one 25S rRNA gene were discovered. The incompleteness of observed rRNA genes is mainly caused by the fact that the rRNA genes always exist as tandem arrays in heterochromatic regions that are not successfully sequenced in a whole-genome shotgun approach.

Key words: tRNA; rRNA; rice genome; modified wobble hypothesis; *indica*; *japonica*

水稻基因组中的 tRNA 和 rRNA 基因

王希胤[1,2,3], 史晓黎[1,2], 郝柏林[2,4,5,①]

(1. 北京大学生物信息中心, 北京 100871;
2. 中国科学院北京基因组研究所, 北京 101300;
3. 河北理工大学理学院, 唐山 063009;
4. 复旦大学理论生命科学研究中心, 上海 200433;
5. 中国科学院理论物理研究所, 北京 100080)

摘 要: 在最近完成测序的水稻籼稻和粳稻两个亚种基因组中, 各找到 564 和 519 个较为可靠的 tRNA 基因, 进一步证实了于 2002 年发表的基于基因组序列草图的分析结果。修正的摆动假设, 即至少需要 46 种 tRNA 基因才能译出 61 种可能的反密码子, 在这两个亚种中均准确成立。在这 46 种 tRNA 中, 有些在籼稻和粳稻中的序列均全同。有 18 种水稻 tRNA 与拟南芥中的相应序列全同。在籼稻基因组序列中还发现了 384 个 5S rRNA 基因, 一批 17S 和 5.8S rRNA 基因以及一个 25S rRNA 基因。这些 rRNA 基因的不完备是由于它们通常以串接阵列形式存在于异染色质区域, 而后者在全基因组霰弹法测序中不易完整测出。在 tRNA 和 rRNA 基因序列之间发现了多

收稿日期: 2004 - 07 - 26; 修回日期: 2004 - 08 - 10

基金项目: 国家自然科学基金项目(编号: 30170232)和中国科学院创新工程项目(编号: KJCX1 - 08)资助[Supported by Chinese National Natural Science Foundation (No. 30170232) and the Innovation Project of CAS (No. KJCX1 - 08)]

作者简介: 王希胤(1972 -), 男, 河北滦县人, 博士研究生, 研究方向: 生物信息学。E-mail: wangxy@mail.cbi.pku.edu.cn; Tel: 010-62756730

① 通讯作者。郝柏林(1934 -), 北京人, 教授, 中国科学院院士, 研究方向: 生物信息学。E-mail: hao@itp.ac.cn

处互补片段,这将有助于研究它们的进化和相互作用。

关键词:tRNA;rRNA;水稻基因组;修正的摆动假设;籼稻;粳稻
中图分类号:Q75;S511 文献标识码:A 文章编号:0379-4172(2004)09-0871-07

Non-coding RNA genes consist of tRNA, rRNA, snRNA, snoRNA and other small RNA genes. Although they account for only a small fraction of all genes in a genome, they play indispensable role in biological activities. The tRNA and rRNA genes take part in synthesizing the protein both constitutionally and functionally. Transfer RNAs bridge the information flow from mRNA to protein, acting as an adaptor of high fidelity. There are hundreds of, sometimes even more than one thousand, tRNA genes in an organism 's nuclear genome. Francis Crick stated in 1966 that less than 61 species of tRNA genes are needed in an organism through non-Watson-Crick base-pairing rules[1]. This is the well-known wobble hypothesis. In 1982, Guthrie and Abelson revised the wobble hypothesis[2]. They showed that 46 tRNA species would be enough in Eukaryotes via the modified base-pairing rules.

The collection of tRNA genes in the sequenced genome of several *Eukaryotes* has been analyzed[3]. This includes yeast, worm, *Arabidopsis* and human[3,4]. The modified wobble hypothesis is almost perfectly obeyed though 1 to 3 species in excess of 46, which might be pseudogenes, were found in those genomes. These were tabulated in our tRNA study[5] on the *indica* draft sequence not yet assembled into chromosomes. We found then 592 tRNA genes in the indica genome and the modified wobble hypothesis was perfectly followed. In the *japonica* genome a group of 85 tRNA genes were found in chromosome 1[6], 70 tRNA genes were found in chromosome 4[7], including 46 nuclear and 24 chloroplast-derived tRNA genes. The *japonica* chromosome 10 sequencing consortium found 67 tRNA genes, including 28 nuclear and 39 chloroplast-derived tRNA genes[8].

Most of the 17S, 5.8S, 25S rRNA genes reside in the NORs (nucleolar organization regions) and several NORs can cluster into a nucleolus[9]. A 17S rRNA, a 5.8S rRNA and a 25S rRNA form a structure unit and hundreds of these structure units repeat as tandem arrays with spacers between the rRNA genes. These rRNAs are transcribed by polymerase I and the mature products after base modifications construct the ribosomes together with dozens of proteins. Another kind of rRNA taking part in constructing ribosomes is 5S rRNA. Hundreds of 5S rRNA genes cluster together to form several regions in genome and are transcribed by RNA polymerase Ⅲ. The 5S rRNA gene has internal control regions (ICRs), which are indispensable and sufficient for the gene 's transcription. Any base deletion in ICRs dimishes transcript. In the 3′-flanking region, there is an oligo-T string serving as a terminator[9]. The highly conserved segments of tRNA and rRNA genes arouse great concern on how they evolve. To interpret the evolution of rRNA genes, many evolution models have been proposed. However the problem is still open. A few complementary regions between these rRNAs and tRNAs were revealed in *E. coli*. The possible interactions implied by the complementary regions are predicted to be essential in constructing ribosomes and synthesizing proteins.

With several rice genome sequencing projects approaching the finishing stage it makes sense to refine and deepen the study of tRNA and rRNA genes in the assembled chromosome sequences and compare the results with what have been reported before.

1 Materials and Methods

We have searched and analyzed the tRNA and rRNA genes in the BGI *indica* and Syngenta

japonica superscaffolds assembled from whole-genome shotgun reads. These sequences have been uploaded to DDBJ/EMBL/GenBank with project accession numbers AAAA02000000 and AACV01000000 which will be available soon in the above databases and via the BGI-RIS[10] system when the paper on the finished rice genomes sees the light. In order to compare the RNA genes in the monocotyledon and dicotyledon genomes, we have downloaded from GenBank the *Arabidopsis* genomic sequences with the accession numbers NC_003070 ~ 71 and NC_003074 ~ 76 and the tRNA and rRNA genes were obtained by us independently from the original annotation.

To locate the tRNA genes, we run the tRNAscan-SE[11] to find the candidate tRNA genes and manual inspection is then performed with the help of BLAST[12] to obtain the creditable tRNA datasets. The rRNA genes were found using BLAST with known rRNA genes in rice and other species as queries. To find possible interaction sites in rRNA sequences, BLAST and the EMBOSS[13] programs WATER and MATCHER are used to make local alignments.

2 Results

A total of 564 and 519 tRNA genes were identified in the currently assembled chromosome sequences of *indica* and *japonica*. The tRNA genes are dispersed throughout all 12 chromosomes in each of the rice subspecies (Table 1).

Table 1 The number of tRNA genes in each chromosome found

Chromosome	*Indica*	*Japonica*	Chromosome	*Indica*	*Japonica*
Chr01	85	71	Chr07	34	35
Chr02	57	59	Chr08	45	42
Chr03	79	68	Chr09	34	32
Chr04	45	41	Chr10	28	23
Chr05	58	56	Chr11	23	24
Chr06	38	32	Chr12	38	36

Different from the situation in the human genome, few clusters could be found in the rice genome. According to the tRNA datasets obtained in *indica* and *japonica*, 46 tRNA gene species are found in both rice subspecies and the modified wobble hypothesis is perfectly followed, see Table 2. In Table 2, we show amino acids, their corresponding codons and tRNA anticodons together

Table 2 The tRNA genes and the modified wobble hypothesis in *Oryza sativa* L. ssp. *indica* and *japonica*

Phe	UUU	AAA	0	0	Ser	UCU	AGA	11	10	Tyr	UAU	AUA	0	0	Cys	UGU	ACA	0	0
	UUC	GAA	15	14		UCC	GGA	0	0		UAC	GUA	14	13		UGC	GCA	11	9
Leu	UUA	UAA	4	4		UCA	UGA	7	6	Ter	UAA	UUA	0	0	Ter	UGA	UCA	0	0
	UUG	CAA	10	7		UCG	CGA	7	7		UAG	CUA	0	0	Trp	UGG	CCA	16	14
	CUU	AAG	17	17	Pro	CCU	AGG	14	14	His	CAU	AUG	0	0	Arg	CGU	ACG	14	12
	CUC	GAG	0	0		CCC	GGG	0	0		CAC	GUG	12	11		CGC	GCG	0	0
	CUA	UAG	6	6		CCA	UGG	8	7	Gln	CAA	UUG	14	13		CGA	UCG	5	3
	CUG	CAG	8	8		CCG	CGG	9	8		CAG	CUG	10	10		CGG	CCG	7	6
Ile	AUU	AAU	20	19	Thr	ACU	AGU	12	11	Asn	AAU	AUU	0	0	Ser	AGU	ACU	0	0
	AUC	GAU	0	0		ACC	GGU	0	0		AAC	GUU	19	15		AGC	GCU	10	9
	AUA	UAU	6	5		ACA	UGU	6	6	Lys	AAA	UUU	9	9	Arg	AGA	UCU	9	7
Met	AUG	CAU	28	33		ACG	CGU	5	4		AAG	CUU	19	19		AGG	CCU	11	9
Val	GUU	AAC	22	19	Ala	GCU	AGC	23	21	Asp	GAU	AUC	0	0	Gly	GGU	ACC	0	0
	GUC	GAC	0	0		GCC	GGC	0	0		GAC	GUC	26	24		GGC	GCC	24	22
	GUA	UAC	4	3		GCA	UGC	11	9	Glu	GAA	UUC	11	11		GGA	UCC	12	10
	GUG	CAC	10	11		GCG	CGC	11	10		GAG	CUC	27	25		GGG	CCC	10	9

Note :For each of the 20 amino acids we show its corresponding codon, anticodon, the copy number of the corresponding tRNA gene in *indica* and *japonica*, *respectively*. The wobble hypothesis is displayed by slanting lines between codons and anticodons.

with the cognate tRNA gene numbers in *indica* and *japonica*. The lines in Table 2 show the modified wobble base-pairing rules between codons and their corresponding anticodons. The tRNA genes in *japonica* are like those in *indica* in almost all aspects. The only difference might be the number of tRNA genes, which is partly caused by the smaller size of *japonica* chromosomes assembled so far. Two species of tRNA genes have an intron. Among the 28 trnM genes in *indica*, 16 of them related to elongation of polypeptides have an intron 11 ~ 16 bases long, while the other 12 genes related to the initiation of translation have no intron. The trnY-GUA genes also contain an intron about 14 bases long. Bases are highly even perfectly conserved between rice tRNA genes and their *A. thaliana* counterparts. Among 46 tRNA species, 18 species of rice tRNA genes have identical counterparts in *A. thaliana*. There is 1 or 2 or seldom more point mutations in the other tRNA species. There are about 100 copies of chloroplast and mitochondrion tRNA genes found in the current *indica* genome sequences. All of them are not included in the final set of tRNA genes.

Only three 5.8S rRNA genes, located on chromosome 1, 6 and 12 respectively, and one 17S rRNA gene, located on chromosome 1, were found in current *indica* genome sequence. An 128 base intergenic spacer separates the 17S rRNA gene and the 5.8S rRNA gene on chromosome 1. One 5.8S rRNA gene and one 17S rRNA gene are obtained in chromosome 3 of *japonica* genome sequence. The length of the intergenic spacer between these two genes is 731 bases. The 25S rRNA gene was not found in currently assembled chromosomes of both subspecies. However in the dataset of *indica* scaffolds, some of which are not assembled into chromosomes, we recovered one 8 kb 17S-5.8S-25S rRNA structure unit, fifty 5.8S rRNA genes and seventeen 17S rRNA genes. The lengths of 17S, 5.8S and 25S rRNA genes are 1751, 283 and 3 307 bases, respectively.

We also discovered 71 and 15 5S rRNA genes in the *indica* and *japonica* chromosome sequences, respectively. The number is surely far smaller than expected by about one thousand. The 5S rRNA genes in *indica* are located on 3 chromosomes, forty-nine in two clusters on chromosome 4, twenty in one cluster on chromosome 3 and two on chromosome 2, while the 5S rRNA genes in *japonica* are all located on chromosome 6. On the other chromosomes only small segments of 5S rRNA genes were found. The different distribution on chromosomes between the two genomes may be due to the smaller dataset of 5S rRNA genes. We see that the 5S rRNA genes are mostly clustered on chromosome 4, 3, 6, and 2. In the *indica* scaffolds dataset we obtained 384 copies of 5S rRNA genes. The transcribed regions of all the 5S rRNA genes are typically 120 base long with a non-transcribed intergenic spacer about 203 bases long. They have an oligo-T tail at their 3′-flanking region which is the signal to terminate transcription.

Based on sequence comparison to the 5S rRNA in *Arabidopsis*, the transcription controls regions of rice 5S rRNA were revealed (Fig. 1). The transcription unit is 120 bases long. The internal control regions are conserved between the rice and *Arabidopsis* and the 9 mutation sites are all non-essential ones as the *Arabidopsis* experiments[9] indicate. The major difference between the two species is that they have very different 5′-flanking regions. The Arabidopsis has a TATA-box string which was proved to be an important element for re-initiation of transcription, while in rice no trace of any TATA-box-like string has been seen. The 5′-flanking regions are highly conserved among the rice 5S rRNA genes, which implies possible existence of different control elements for re-initiation of transcription in rice. A GC element exists in the two species 5S rRNA at −11 ~ 12 bases upstream in *Arabidopsis* and −12 ~ 13 bases upstream in rice, which is important for transcription efficiency. The transcription region ends

with a C in *Arabidopsis*, while with two Cs in rice. Though experiments suggested four Ts are enough to terminate transcription, almost all rice 5S rRNA genes terminate with a stretch of five Ts. Another two stretch of four Ts are found in 5′-flanking region (not shown in Fig. 1), which we infer may serve to affirm the terminating the transcription of the last gene unit and beginning transcription of the next gene unit, that is to say, the correct termination of the last gene unit may be significant for starting the transcription of the next unit.

```
-30       -20       -10       -1 1      10        20        30        40        50        60
AAGAAATGACACCCGAATGACCCCAAAGCATGTC GGATGCGATCATACCAGCACTAAAGCACCCGATCCCATCAGAACTCCGAAGTTAAGCGTG
AAGAATAGATACGATTGCATTGCATGTAGTAA C GGATGCGATCATAGCAGCACTAATGCACCAGATCCCATAAGAACTCCGCAGTTAAGTGTG
   ** * ******** *********                *         *    *         *         *    *
     70        80        90        100       110       120
CTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCTGGGAAGTCCTCGTGTTGCATCCCT CCTTTTT
CTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCCCGAAAGTCCTCGTGTTTCATCCCT C-TTTTT
                          * *           *
```

Fig. 1 Sequence comparison of typical 5S rRNA in rice (upper) and *Arabidopsis* (lower)

An asterisk(*) marks the sites different in *indica* and *japonica* 5S rRNA genes.

The 25S, 5.8S and 5S mature rRNAs constitute the larger ribosomal subunit and the 17S mature rRNA constitutes the small ribosomal subunit. They are predicted to be essential both constitutionally and functionally in translation of the information stored in mRNA to direct the synthesis of peptides. In order to recover whether there are interaction sites between rRNAs, their sequences are locally aligned to find possible complementary interaction sites (Table 3).

Table 3 Possible interaction sites in the rRNA sequences

5S rRNA	55-AGCGTGCTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCTG-GGAAGTCCTCGTG-110
	::::: :: :::: :: : : : : : ::: :::: ::: : ::::: :
5.8S rRNA	282-TCGCAGAAATCCGCACT--GCGGGTCCGTCCGCACGGGAGCCGGCCTACCGGAGCCC-228
5S rRNA	37-ATCAGAACTCCGAAGTTA-54
	::: : :::::::::
25S rRNA	1912-TAGGAATAGGGCTTCAAT-1895
5S rRNA	14-CCAGCACTAAAGCACCGGA-TCC-35
	::: ::: : :: :::::: :::
17S rRNA	922-GGTTGTGTTATCCTGGCCTTAGG-900
5.8S rRNA	218-GCAAGTTGCGCCCGAGGCCATCCGGCCGAGGGCACGCCTGCC-259
	::: ::: ::: :::: :::: :: ::: : : ::::
25S rRNA	1968-CGTCGGACGAGGGTTCCGC-AGGCGGGTTCCGGGTTGCACGG-1928
5.8S rRNA	59-GGTCGGCCCTGGCCGTCCGGCGGCGCGGC-87
	:::::: :: ::::: ::::: ::::
17S rRNA	663-CCAGCCACGAGCGGCACTCCGCCTGGCCG-635
25S rRNA	2214-AACTATGACTCTCTTAAGGTAGCCAAATGCCTCGTCA----TCTAATTAGTG---AC--GCGCATGAATGGATTAA-2280
	::::::: ::: :::: : ::::: :: ::: :: ::::::: :: :: :::: :::: ::::
17S rRNA	97-TTGATATTGACTAAATTAC-TCGGTAAGCGTCAAAGTGTCAAGCTTAATCAAGTATGAACGTCTACGTACCGAATT-23

The possible interaction sites listed are the best matched complimentary regions between the corresponding rRNA sequences. Other interaction regions may also exist. Using BLAST, we found another complementary string between 25S rRNA (2288-cccactgtccctat-2301) and 17S rRNA (860-ccgactgtccctat-847). We can see that the 5S rRNA has well matched segments complementary

with 17S, 25S and 5.8S rRNAs and the segments follow one another immediately. The length of 5S rRNA is 120nt in rice, so it appears that most part of the 5S rRNA sequence, including the ICR regions, serves as a thread linking all the other rRNAs together. The 17S rRNA comprising the small ribosomal subunit has complementary segments with all three rRNAs comprising the large ribosomal subunit. All the four kinds of rRNA genes interact with each other as Table 3 suggests. The interactions between these rRNAs may be very extensive and we infer the extensive interactions between these rRNAs are the crucial base of ribosomal function. The possible interaction regions found may help to predict the 3-D structure of the ribosome and understand how the ribosome works. Interaction sites between rRNAs and tRNAs were searched by local alignment, however the interaction sites between rRNAs and different tRNA species are so diverse that they are suspicious to play any role.

3 Discussion

As we indicated above, the modified wobble hypothesis is perfectly followed in the two rice subspecies. A careful study carried out by us on *A. thaliana* suggests the modified wobble hypothesis may also be perfectly obeyed in this dicotyledon. The fact that the modified base-pairing rules hold in the model organisms of both monocotyledon and dicotyledon hints on that the modified rules may be perfectly obeyed in all land plants. The few exceptional tRNA genes in excess of 46 in yeast and human might be possible pseudogenes. Therefore, Guthrie and Abelson 's hypothesis that 46 species of tRNAs are enough in Eukaryotes is nearly to be verified.

It is evident that the absence or small number of rRNA genes in the current chromosome sequences is caused by the difficulty in assembling highly repetitious regions. The NORs, where 17S, 5.8S and 25S rRNAs reside, are highly heterochromatized that makes it much more difficult to obtain full set of these RNA genes and incorporate them into chromosome sequences. The assembling of highly heterochromatic repetitious regions is a hard job not only for the whole genome shotgun approach but also for the traditional clone-by-clone approach adopted by the Human Genome Consortium and the *Arabidopsis* project. The Human Genome Consortium found few rRNA genes, not a single 17S-5.8S-25S rRNA gene unit at all[4] and the Celera report did not say a word on rRNA[14]. The *Arabidopsis* genome project did not sequence the 17S-5.8S-25S rRNA regions[15]. Being different from rRNA genes, most tRNA genes are successfully assembled into chromosomes. The tRNA genes disperse in the whole genome, not only reside in the heterochromatic region. They are shorter than rRNA, forming only small clusters, most of them having less than 20 copies in each tRNA species. This will not handicap the assembling. Each kind of rRNA genes are highly conserved for concerted evolution. For the tRNA genes things are further complicated by the presence of dozens of different species. The findings in rice and other Eukaryotes show that though the nuclear tRNA genes are distantly located on different chromosomes, the genes from the same tRNA species are highly conserved, even identical. This may also imply the possibility of concerted evolution. However the tRNA evolution mechanism may not be the same as that of rRNA genes.

Acknowledgements: We thank Prof. LUO Jing-Chu for critical reading of the manuscript and helpful advice. We are grateful for the help of NI Pei-Xiang and FANG Lin, both from BGI, in preparing the sequence data.

参考文献(References):

[1] Crick F. Codon-anticodon pairings: the wobble hypothesis. *Journal of Molecular Biology*, 1966, 19: 548 ~ 555.

[2] Guthrie C,Abelson J. Organization and expression of tRNA genes in *Saccharomyces cerevisiae*. In :Strathern J,Jones J R,Broach E W,eds. The Molecular Biology of the Yeast *Saccharomyces* :Metabolism and Gene Expression. Cold Spring Harbor Laboratory,New York,1982,487 ~ 528.

[3] See GtRDB :The Genomic tRNA Database at http ://rna. wustl. edu/GtRDB/

[4] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*,2001,409(6822) :860 ~ 921.

[5] WANG Xi-Yin,SHI Xiao-Li,HAO Bo-Lin. The transfer RNA genes in *Oryza sativa* L. ssp. *Indica*. *Science in China* (Series C). 2002,45(2) :504 ~ 511.
王希胤,史晓黎,郝柏林. 籼稻基因组中的 tRNA 研究. **中国科学(C 辑)**,2002,32(6) :505 ~ 511.

[6] Sasaki T,Matsumoto T,Yamamoto K,Sakata K,Baba T,Katayose Y,Wu Jianzhong,Niimura Y,Cheng Zhukuan,Nagamura Y,Antonio B A,Kanamori H,Hosokawa S,Masukava M,Arikawa K,Chiden Y,Hayashi M,Okamoto,M,Ando T,Aoki H,Arita K,Hamada M,Harada C,Hijishita S,Honda M,Ichikawa Y,Idonuma A,Iijima M,Ikeda M,Ikeno M,Ito S,Ito T,Ito Y,Ito Y,Iwabuchi A,Kamiya K,Karasawa W,Katagiri S,Kikuta A,Kikuta A,Kobayashi N,Kono I,Machita K,Maehara T,Mizuno H,Mizubayashi T,Mukai Y,Nagasaki H,Nakashima M,Nakama Y,Nakamichi Y,Nakamura M,Namiki N,Negishi M,Ohta I,Ono N,Saji S,Sakai K,Shibata M,Shimokawa T,Shomura A,Song J Y,Takazaki Y,Terasawa K,Tsuji K,Waki K,Yamagata H,Yamane H,Yoshiki S,Yoshihara R,Yukawa K,Zhong H S,Iwama H,Endo T,Ito H,Hahn J H,Kim H,Eun M Y,Yano M,Jiang J M,Gojobori T. The genome sequence and structure of rice chromosome 1. *Nature*,2002,420 :312 ~ 316.

[7] Feng Q,Zhang Y J,Hao P,Wang S Y,Fu G,Huang Y C,Li Y,Zhu J J,Liu Y L,Hu Xin,Jia P X,Zhang Y,Zhao Q,Ying K,Yu S L,Tang Y S,Weng Q J,Zhang L,Lu Y,Mu J,Lu Y Q,Zhang L S,Yu Z,Fan D L,Liu X H,Lu T T,Li C,Wu Y R,Sun T G,H Y Lei,Li T,Hu H,Guan J P,Wu M,Zhang R Q,Zhou B,Chen Z H,Chen L,Jin Z Q,Wang R,Yin H F,Cai Z,Ren S X,Lv G,Gu W Y,Zhu G F,Tu Y F,Jia J,Zhang Y,Chen J,Kang H,Chen X Y,Shao C Y,Sun Y,Hu Q P,Zhang X L,Zhang W,Wang L J,Ding C W,Sheng H H,Gu J L,Chen S T,Ni L,Zhu F H,Chen W,Lan L F,Lai Y,Cheng Z K,Guk M H,Jiang J M,J Y Li,Hong G F,Xue Y B,Han B. Sequence and analysis of rice chromosome 4. *Nature*,2002,420 :316 ~ 320.

[8] The Rice Chromosome 10 Sequencing Consortium. In-depth view of structure,activity,and evolution of rice dhromosome 10. *Science*,2003,300 :1566 ~ 1569.

[9] Cloix C,Yukawa Y,Tutois S,Sugiura M,Tourmente S. In vitro analysis of the sequences required for the transcription of the Arabidopsis thaliana 5S rRNA genes. *The Plant Journal*,2003,35(2) :251 ~ 261.

[10] Zhao W M,Wang J,He X M,Huang X B,Jiao Y Z,Dai M T,Wei S L,Fu J,Chen Y,Ren X Y,Zhang Y,Ni P X,Zhang J G,Li S G,Wang J,Wong G K,Zhao H Y,Yu J,Yang H M,Wang J BGI-RIS :an integrated information resource and comparative analysis workbench for rice genomics. *Nucl Acids Res*,2004,32(Database Issue) :D377 ~ D382.

[11] Lowe T M,Eddy S R. tRNAscan-SE :A program for improved detection of transfer RNA genes In genomic sequence. *Nucl Acids Res*,1997,25(5) :955 ~ 964.

[12] Altschul S F,Madden T L,Schaffer A A,Zhang J,Miller W,Lipman D J. Gapped BLAST and PSI-BLAST :a new generation of protein database search programs. *Nuc Acids Res*,1997,25(17) :3389 ~ 3402.

[13] Rice P,Longden I,Bleasby A. EMBOSS :The European molecular biology open software suite. Trends Genet,2000,16(6) :276 ~ 277.

[14] Venter J C,Adams M D,Myers E W,…,Xia A,Zandieh A,Zhu X H. The sequence of the human genome. *Science*,2001,291(5507) :1304 ~ 1351.

[15] The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*,2000,408(6814) :796 ~ 815.

(责任编辑 :李绍武)