

植物KCO基因密码子使用特性

杨平¹, 周峰^{1,2,*}, 张边江¹, 华春¹

¹南京晓庄学院生物化工与环境工程学院, 南京211171; ²山东省黄河三角洲生态环境重点实验室, 山东滨州256603

摘要: 以植物钾离子外排通道(K⁺ channel outward-rectifier, *KCO*)基因为研究对象, 运用CodonW软件分析了75个植物*KCO*基因密码子的使用模式, 探讨密码子的使用模式和影响密码子使用的各种可能因素。结果表明: 碱基组成差异($r=0.961$, $P<0.01$)和自然选择($r=0.568$, $P<0.01$)是影响密码子使用的主要因素, 并且高表达的基因强烈偏爱使用以G或C结尾的密码子。确定了UUC、CUC等26个均以G/C结尾的密码子为植物*KCO*基因的高表达优越密码子。

关键词: *KCO*基因; 密码子偏好性; 自然选择

Characterization of Codon Usage of Plant *KCO* Genes

YANG Ping¹, ZHOU Feng^{1,2,*}, ZHANG Bian-Jiang¹, HUA Chun¹

¹School of Biochemical and Environmental Engineering, Nanjing Xiaozhuang University, Nanjing 211171, China; ²Shandong Provincial Key Laboratory of Eco-Environmental Science for Yellow River Delta, Binzhou, Shandong 256603, China

Abstract: Codons of 75 K⁺ channel outward-rectifier (*KCO*) genes from plant were studied. CodonW software was used to calculate different parameters for the codon usage. The codon usage bias in plant *KCO* genes was analyzed to find the factors affecting the formation of codon bias. The result indicated that nucleotide composition ($r=0.961$, $P<0.01$) and natural selection ($r=0.568$, $P<0.01$) were the main factors on the codon usage of plant *KCO* genes. Moreover, the highly expressed genes had a stronger codon bias of the G or C at the third position of codons. We defined 26 codons end with G/C such as UUC and CUC as the optimal codons in plant *KCO* genes.

Key words: *KCO* genes; codon usage bias; natural selection

KCO (K⁺ channel outward-rectifier)基因是钾离子外排通道基因, 钾离子通道广泛存在于原核和真核生物的体内。Czempinski等(1997)第一次在高等植物拟南芥中克隆了*KCOI*基因, 功能分析发现*KCOI*基因属于两孔钾离子通道基因家族。这是研究植物离子通道的一个重要发现, 在此之前, 两孔钾离子通道基因只有在人类和酵母中发现(Ketchum等1995; Lesage等1996a, b)。拟南芥*KCOI*基因也是第一个克隆的液泡型钾离子通道(Czempinski等2002)。*KCO*基因与植物耐盐有着密切的关系(周峰等2003), Wang等(2012)研究发现在烟草细胞中过量表达白杨两孔钾离子通道基因能增强细胞的耐盐性。

同义密码子使用偏好性与遗传进化、蛋白质功能等有着密切联系, 因此研究同义密码子使用偏好性具有重要的生物学意义(刘汉梅等2010)。目前, 全基因组序列的密码子使用偏好性的研究比较多, 对单一基因家族的密码子使用研究比较少。本研究旨在通过分析植物*KCO*基因全长编码框密码子的使用特性, 为深入研究其遗传进化和表达调控机制等提供重要理论依据。

材料与方法

1 植物*KCO*基因序列的获得

植物*KCO*基因序列来源于NCBI (<http://www.ncbi.nih.gov/>), 通过BLAST比对, 共下载到75条*KCO*基因的CDS (coding DNA sequence)。植物*KCO*基因的进化树是用Mega 5.0软件计算生成, 进化树的名称是物种名缩写加NCBI登录号(图1)。

2 密码子使用特性分析

利用CodonW软件(Peden 2005)分析植物*KCO*家族基因密码子, 密码子参数包括: 氨基酸长度(L_aa)、同义密码子数(L_sym)、疏水性平均值(Gravy)、芳香性值(aromaticity, Aroma)、同义密码子在第3位上腺嘌呤、胸腺嘧啶、鸟嘌呤和胞嘧啶的出现频率(A3s、T3s、G3s和C3s)、GC含量(GC)、编码同一氨基酸的同义密码子的第3位的

收稿 2012-08-10 修定 2012-10-23

资助 江苏省自然科学基金青年基金项目(BK2012073)、国家高技术研究发展“863”计划(2012AA021701)、江苏省生态学重点学科建设项目(2012)和山东省黄河三角洲生态环境重点实验室开放基金(2011KFJJ02)。

* 通讯作者(E-mail: zfbcas@163.com; Tel: 025-86178262)。

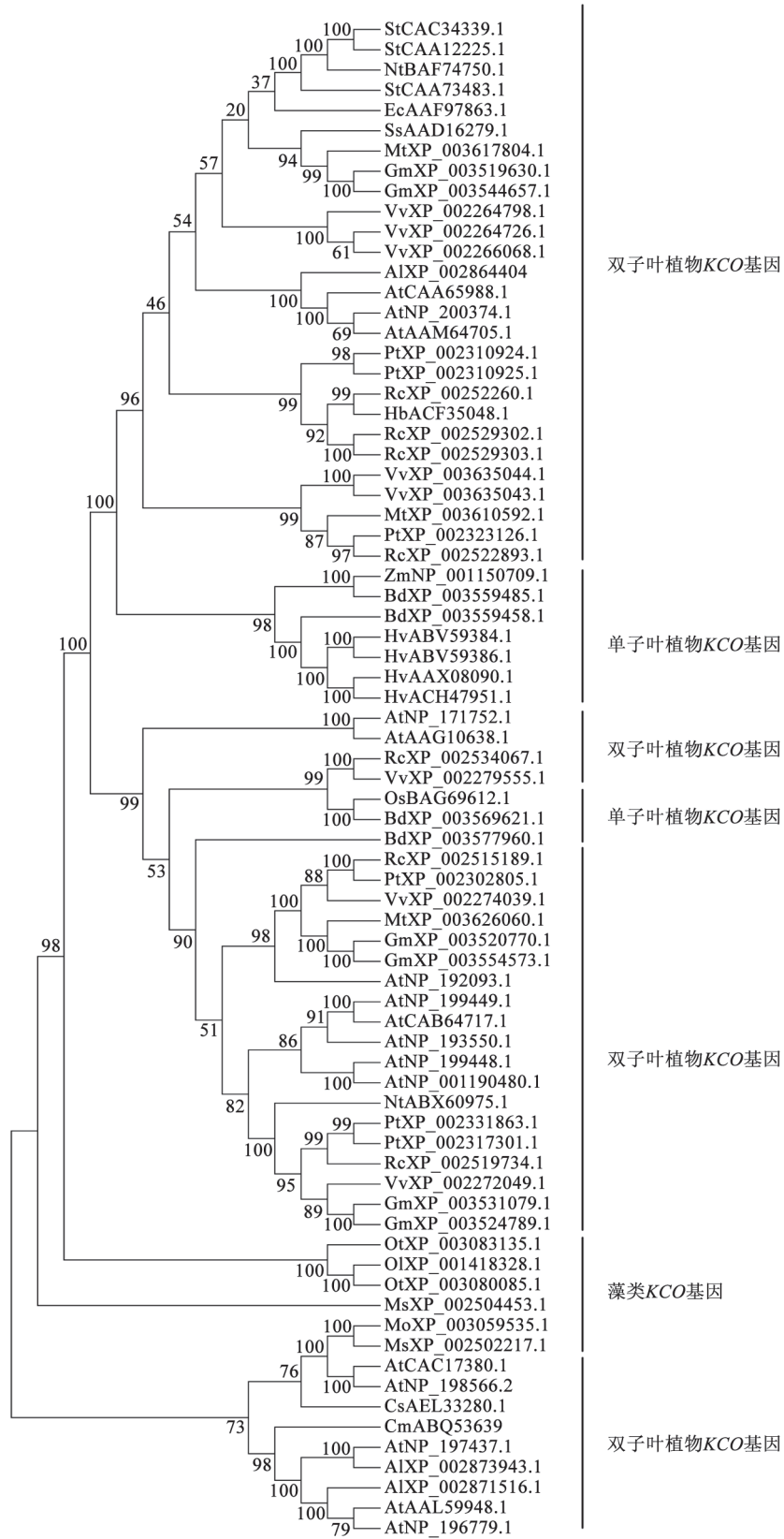


图1 KCO基因进化树

Fig.1 Phylogenetic tree of KCO gene

GC含量(GC3s)、有效密码子数(ENC)、最优密码子使用频率(FOP)、密码子偏爱指数(CBI)、密码子使用的相对概率(RSCU)、密码子适应指数(CAI)等。运用SPSS 17.0统计软件进行密码子组成和使用偏好性各参数间相关性的分析。

3 最优密码子的确定

本文计算所有基因的ENC值和RSCU值, 并对ENC值进行排序, 取该有序数据的上下限区域各5%的序列数据, 确定高表达和低表达样本组。比较高/低表达样本组相应密码子的RSCU值, 如果差异大于0.3, 且该密码子的RSCU值在高表达基因样本中大于1, 在低表达基因样本中小于1, 确定为最优密码子(Duret和Mouchiroud 1999)。

实验结果

1 KCO基因密码子的对应性分析

对应性分析结果显示: 第一条轴(Axis1)和第二条轴(Axis2)能解释基因间RSCU变异分别为40.24%和7.54%。各基因密码子在这两轴的分布见图2, 从图2可以看出KCO基因以A和T结尾的密码子除了CGA全部位于第一轴的左侧, 以G和C结尾的密码子除了TTG和AGG全部位于第一轴的右侧。从图3看出, 藻类和单子叶植物的基因主要位于第一条向量轴右侧, 而双子叶植物的基因大部分位于左侧。从基因表达量上分析表明, 第一条向量轴

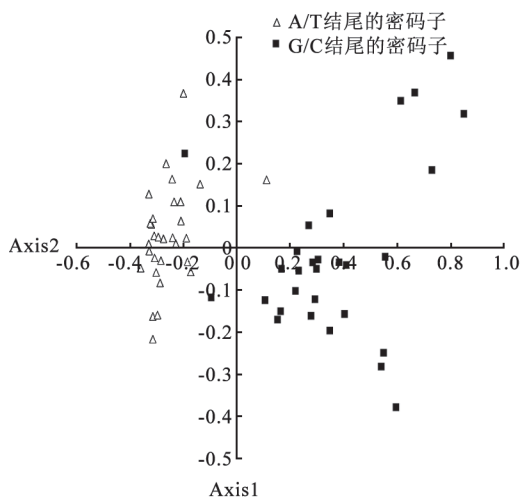


图2 KCO基因所使用的59种同义密码子在对对应分析中的分布
Fig.2 Projection of 59 synonymy codons of KCO genes on the first two principle axes of correspondence analysis

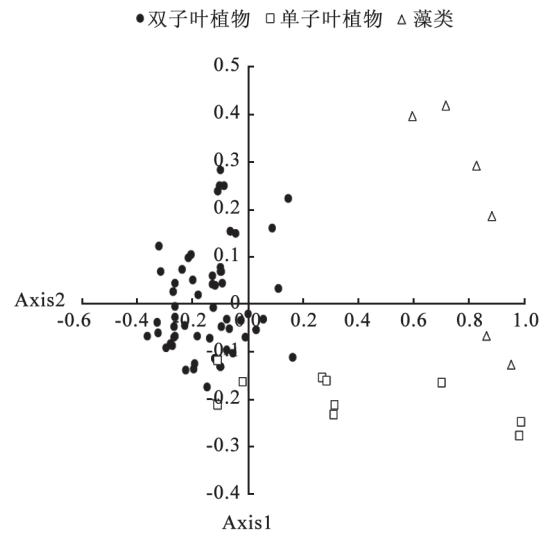


图3 75个KCO基因在前两个向量轴对应分析中的分布
Fig.3 Projection of 75 KCO genes on the first two principle axes of correspondence analysis

右侧是高表达量的基因, 而左侧是低表达量的基因。

第一条向量轴上基因的位置与C3s、G3s、CBI、FOP、CAI、GC、GC3s呈极显著正相关, 而与T3s、A3s、ENC高度负相关(表1), 表明碱基组成差异和自然选择是影响密码子使用的主要因素。并且高表达的基因强烈偏爱使用以G或C结尾的密码子。

2 KCO基因Enc-plot分析

以ENC为纵坐标, GC3s为横坐标, 探讨各基因密码子使用偏性情况, 并检测碱基组成对密码子偏性的影响。密码子偏性仅受到突变影响的基因沿标准曲线分布(在无选择压力的情况下)或落在标准曲线附近, 密码子偏性受到选择影响的基因应该落在标准曲线下方较远的位置(Wright 1990)。从图4可以看出, 大多数基因落在标准曲线下方, 且分布偏离期望曲线。藻类和部分单子叶植物的基因分布在标准曲线下方较远的位置, 具有较高的GC3s含量, ENC值偏小, 具有较强的密码子使用偏好性。此外, ENC与CAI、GC、GC3s呈极显著负相关(表1), 表明除碱基组成偏性外, 自然选择等其他因素对密码子的使用也有重要影响。

3 植物KCO基因最优密码子的确定

本研究对植物75个KCO基因进行了初步分析, 表2给出了高/低表达样本组相应密码子的

表1 *KCO*基因密码子各参数之间的相关系数
Table 1 Correlation coefficients among the parameters of codon usage in *KCO* genes

	T3s	C3s	A3s	G3s	CAI	CBI	FOP	ENC	GC3s	GC	L_sym	L_aa	Gravy	Aroma
C3s	-0.960**													
A3s	0.758**	-0.815**												
G3s	-0.861**	0.787**	-0.848**											
CAI	-0.549**	0.606**	-0.460**	0.391**										
CBI	-0.766**	0.819**	-0.728**	0.618**	0.762**									
FOP	-0.807**	0.842**	-0.711**	0.656**	0.800**	0.989**								
ENC	0.516**	-0.546**	0.617**	-0.541**	-0.458**	-0.549**	-0.562**							
GC3s	-0.967**	0.963**	-0.894**	0.917**	0.553**	0.788**	0.814**	-0.593**						
GC	-0.932**	0.923**	-0.891**	0.894**	0.554**	0.750**	0.780**	-0.626**	0.975**					
L_sym	-0.145	0.112	-0.036	0.215	-0.049	0.068	0.096	-0.082	0.131	0.118				
L_aa	-0.152	0.117	-0.045	0.228*	-0.046	0.070	0.100	-0.086	0.139	0.125	0.999**			
Gravy	0.288*	-0.183	0.070	-0.409**	-0.018	-0.004	-0.092	0.137	-0.257*	-0.285*	-0.591**	-0.593**		
Aroma	0.043	0.044	-0.030	-0.083	0.053	0.059	0.025	0.142	-0.016	-0.092	-0.099	-0.097	0.409**	
Axis1	-0.961**	0.944**	-0.875**	0.915**	0.568**	0.796**	0.826**	-0.568**	0.987**	0.961**	0.114	0.123	-0.248*	0.022

*和**分别表示各参数间相关性达0.05和0.01概率的显著。

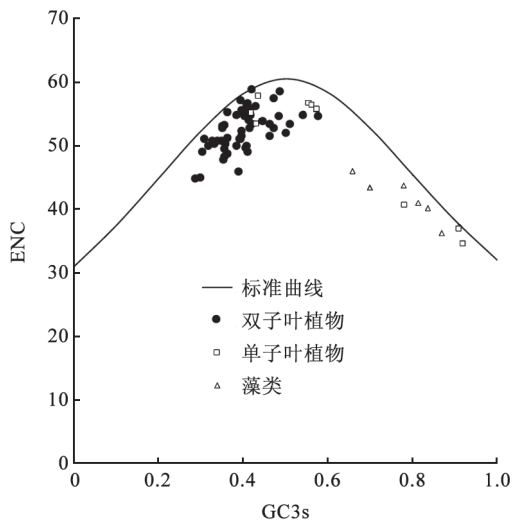


图4 *KCO*基因的ENC-plot曲线
Fig.4 ENC-plot of *KCO* genes

RSCU值和 Δ RSCU值。比较高/低表达样本组相应密码子的RSCU值,如果差异大于0.3,且该密码子的RSCU值在高表达基因样本中大于1,在低表达基因样本中小于1,确定为最优密码子。根据这一方法,并经卡方测验,最终确定了UUC、CUC等26个密码子为植物*KCO*基因的高表达优越密码子(表2中带*的密码子),均以G/C结尾的密码子。

讨 论

本研究通过对75种植物*KCO*基因同义密码子

使用的系统比较发现,碱基组成差异和自然选择是影响植物*KCO*基因密码子使用的主要因素。并且高表达的基因强烈偏爱使用以G或C结尾的密码子,最终确定了UUC、CUC等26个密码子为植物*KCO*基因的高表达优越密码子,也是均以G或C结尾的密码子。

ENC取值范围为20到61,内源基因表达量的相对高低与ENC值成正比。其值越低,表明该基因的密码子使用偏好性越强,对应的内源基因往往表达量也越高。ENC值会受到基因氨基酸组成和基因长短的影响(Fuglsang 2004; Peixoto等2003)。本研究中发现第1条向量轴上基因的位置与碱基组成(C3s、G3s、T3s、A3s、GC、GC3s)和CAI均呈极显著正相关(表1),而与基因长度(L_aa)相关性未达到显著水平($r=0.123$)。基因长度(L_aa)与ENC和CAI的相关系数分别为-0.086和-0.046,均未达到显著水平。这表明植物*KCO*基因碱基组成差异和自然选择是影响密码子使用的主要因素,而基因长度与基因的表达水平以及密码子偏性无明显相关。

CAI取值范围在0~1之间,基因表达水平的实际观测值与CAI值非常接近,较高的CAI值的基因具有较高的表达量,反之,表达量较低的基因具有较低的CAI值。CAI的值越高意味着具有较高的密码子偏好性(Peixoto等2003; Romero等2000; 吕林峰等2012)。本研究中CAI与第1条向量轴上基因

表2 KCO基因相对同义密码子使用度

Table 2 RSCU analysis of protein coding region in KCO genes

氨基酸	密码子	高表达序列		低表达序列		ΔRSCU	氨基酸	密码子	高表达序列		低表达序列		ΔRSCU	
		数目	RSCU	数目	RSCU				数目	RSCU	数目	RSCU		
Phe	UUU	4	0.12	35	1.15	-1.03	Ser	UCU	2	0.14	17	1.46	-1.32	
	UUC*	61	1.88	26	0.85	1.03		UCC*	25	1.74	4	0.34	1.4	
Leu	UUA	2	0.08	25	1.06	-0.98	UCA	1	0.07	22	1.89	-1.82		
	UUG	3	0.12	25	1.06	-0.94	UCG*	22	1.53	3	0.26	1.27		
	CUU	4	0.17	55	2.34	-2.17	Pro	CCU	8	0.49	15	1.94	-1.45	
	CUC*	64	2.65	9	0.38	2.27		CCC*	26	1.6	3	0.39	1.21	
	CUA	2	0.08	18	0.77	-0.69		CCA	7	0.43	13	1.68	-1.25	
	Ile	CUG*	70	2.9	9	0.38	2.52	CCG*	24	1.48	0	0	1.48	
		AUU	5	0.42	44	1.67	-1.25	Thr	ACU	2	0.11	33	1.65	-1.54
AUC*		27	2.25	16	0.61	1.64	ACC*		35	2	13	0.65	1.35	
AUA	4	0.33	19	0.72	-0.39	ACA	3		0.17	31	1.55	-1.38		
Met	AUG	35	1	28	1	0	ACG*	30	1.71	3	0.15	1.56		
Val	GUU	2	0.08	41	1.86	-1.78	Ala	GCU	7	0.18	34	1.74	-1.56	
	GUC*	41	1.66	17	0.77	0.89		GCC*	50	1.32	9	0.46	0.86	
	GUA	1	0.04	12	0.55	-0.51		GCA	5	0.13	30	1.54	-1.41	
	GUG*	55	2.22	18	0.82	1.4		GCG*	90	2.37	5	0.26	2.11	
Tyr	UAU	2	0.08	25	1.28	-1.2	Ter	UAA	0	0	1	1	-1	
	UAC*	47	1.92	14	0.72	1.2		UAG	1	1	2	2	-1	
Cys	UGU	3	0.25	14	1.65	-1.4		UGA	2	2	0	0	2	
	UGC*	21	1.75	3	0.35	1.4	Trp	UGG	20	1	9	1	0	
His	CAU	3	0.26	12	1.41	-1.15		Arg	CGU	1	0.07	5	0.7	-0.63
	CAC*	20	1.74	5	0.59	1.15			CGC*	29	1.89	2	0.28	1.61
Gln	CAA	6	0.32	18	0.88	-0.56	CGA		4	0.26	4	0.56	-0.3	
	CAG*	32	1.68	23	1.12	0.56	CGG*	35	2.28	2	0.28	2		
Asn	AAU	5	0.25	19	1.41	-1.16	Ser	AGU	4	0.28	12	1.03	-0.75	
	AAC*	35	1.75	8	0.59	1.16		AGC*	32	2.23	12	1.03	1.2	
Lys	AAA	2	0.1	40	1.23	-1.13	Arg	AGA	1	0.07	19	2.65	-2.58	
	AAG*	39	1.9	25	0.77	1.13		AGG	22	1.43	11	1.53	-0.1	
Asp	GAU	16	0.42	50	1.67	-1.25	Gly	GGU	6	0.28	21	1.22	-0.94	
	GAC*	60	1.58	10	0.33	1.25		GGC*	43	2.02	8	0.46	1.56	
Glu	GAA	9	0.26	26	0.87	-0.61		GGA	11	0.52	24	1.39	-0.87	
	GAG*	60	1.74	34	1.13	0.61	GGG	25	1.18	16	0.93	0.25		

*为最优密码子。

的位置和碱基组成(C3s、G3s、T3s、A3s、GC、GC3s)均呈极显著正相关(表1), 而与L_aa、L_sym、Gravy、Aroma相关性未达到显著水平(表1)。这说明碱基组成差异和密码子偏性是影响植物KCO基因的表达水平的主要因素, 而基因长度、同义氨基酸数、氨基酸组份指数对植物KCO基因的表达水平影响不明显。

参考文献

刘汉梅, 何瑞, 张怀渝, 黄玉碧(2010). 拟南芥和水稻转录因子 WRKY同义密码子的偏好性分析. 四川农业大学学报, 28 (1):

20~27

吕林峰, 岳建宇, 费忠安, 赵锐, 乔代蓉, 曹毅(2012). 油菜菌核病菌 (*Sclerotinia sclerotiorum*)与细菌间的基因水平转移. 应用与环境生物学报, 18 (1): 42~48
 周峰, 李平华, 王宝山(2003). K⁺营养与植物耐盐性的关系. 植物生理学通讯, 39 (1): 67~70
 Czempinski K, Frachisse JM, Maurel C, Barbier-Brygoo H, Mueller-Roeber B (2002). Vacuolar membrane localization of the *Arabidopsis* 'two-pore' K⁺ channel KCO1. Plant J, 29 (6): 809~820
 Czempinski K, Zimmermann S, Ehrhardt T, Müller-Röber B (1997). New structure and function in plant K⁺ channels: KCO1, an outward rectifier with a steep Ca²⁺ dependency. EMBO J, 16 (10): 2565~2575
 Duret L, Mouchiroud D (1999). Expression pattern and, surprisingly,

- gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. Proc Natl Acad Sci USA, 96 (8): 4482~448
- Fuglsang A (2004). The 'effective number of codons' revisited. Biochem Biophys Res Commun, 317 (3): 957~964
- Ketchum KA, Joiner WJ, Sellers AJ, Kaczmarek LK, Goldstein SAN (1995). A new family of outwardly rectifying potassium channel proteins with two pore domains in tandem. Nature, 376 (6542): 690~695
- Lesage F, Guillemare E, Fink M, Duprat F, Lazdunski M, Romey G, Barhanin J (1996a). A pH-sensitive yeast outward rectifier K⁺ channel with two pore domains and novel gating properties. J Biol Chem, 271 (8): 4183~4187
- Lesage F, Guillemare E, Fink M, Duprat F, Lazdunski M, Romey G, Barhanin J (1996b). TWIK-1, a ubiquitous human weakly inward rectifying K⁺ channel with a novel structure. EMBO J, 15 (5): 1004~1011
- Peden J (2005). Correspondence Analysis of Codon Usage. Available from: <http://codonw.sourceforge.net/>
- Peixoto L, Zavala A, Romero H, Musto H (2003). The strength of translational selection for codon usage varies in three relatives of *Sinorhizobium meliloti*. Gene, 320: 109~116
- Romero H, Zavala A, Musto H (2000). Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. Nucl Acids Res, 28 (10): 2084~2090
- Wang FF, Deng SR, Ding MQ, Sun J, Wang MJ, Zhu HP, Han YS, Shen ZD, Jing XS, Zhang F et al (2012). Overexpression of a poplar two-pore K⁺ channel enhances salinity tolerance in tobacco cells. Plant Cell Tiss Org Cult, DOI: 10.1007/s11240-012-0207-9
- Wright F (1990). The 'effective number of codons' used in a gene. Gene, 87 (1): 23~29