

数据分析

Analysis

小麦淀粉合成酶基因家族的生物信息学分析

李春艳[✉], 李诚[✉], 李卫华[✉], 曹连莆[✉]

石河子大学农学院, 石河子, 832000

✉ 通讯作者: 009917@sina.com ✉ 作者

分子植物育种, 2011年, 第9卷, 第64篇 doi: 10.5376/mpb.cn.2011.09.0064

收稿日期: 2011年03月12日

接受日期: 2011年04月28日

发表日期: 2011年05月27日

这是一篇采用 Creative Commons Attribution License 进行授权的开放取阅论文。只要对本原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

建议的最佳引用格式:

李春艳等, 2011, 小麦淀粉合成酶基因家族的生物信息学分析, 分子植物育种 Vol.9 No.64 (doi: 10.5376/mpb.cn.2011.09.0064)

摘要 为了进一步研究小麦淀粉生物合成机理, 本研究利用生物信息学软件分析了六倍体小麦淀粉合成酶基因家族的组成、结构和进化关系。结果表明, 小麦基因组中至少存在九个编码淀粉合成酶的基因, 分别位于不同的染色体上。九种酶的多肽长度差异较大, 而属于同种同工酶的多肽序列长度相近。根据氨基酸序列的催化和糖原转移两个结构域的起始位置, 可以将小麦淀粉合成酶基因家族分为两个亚家族。

关键词 小麦; 生物信息学; 基因家族; 淀粉合成酶

Bioinformatic Analysis of the Starch Synthase Gene Family in Wheat

Li Chunyan[✉], Li Cheng[✉], Li Weihua[✉], Cao Lianpu[✉]

Shihezi University, Shihezi, 832000, P.R. China

✉ Corresponding author, 009917@sina.com; ✉ Authors

Abstract In order to further investigate the mechanism of starch biosynthesis in wheat, in this study the bioinformatic softwares were employed to analyze the gene constituents, structure and phylogenetic relationship of starch synthase in wheat. The results showed that this gene family were at least composed of nine genes which encode starch-synthesis enzymes and located in different chromosome. The lengths of the amino acid sequences of wheat starch synthases were obviously different but it is similar for the same isoform. Based on the location of catalytic domain and glycosyl transferases domain of the amino acid sequence, the starch synthase gene family in wheat was divided into two subfamilies.

Keywords *Triticum aestivum*. L; Bioinformatics; Gene-family; Starch synthase

研究背景

随着拟南芥和水稻等模式植物全基因组测序的完成(Yu et al., 2002), 推动了人们在基因组水平上对植物重要基因家族进行挖掘、比较分析和功能预测。淀粉的生物合成是一个复杂的过程, 涉及众多淀粉合成酶的共同作用(Tomlinson et al., 2003), 从基因结构的角度可将这些由一个小的多基因家族编码的植物淀粉合成酶划分成GBSS、SS I、SS II、SS III和SS IV五个基因家族(Dian et al., 2005)。目前, 对水稻和拟南芥淀粉合成相关酶基因结构、进化和功能已有众多深入研究(Jeng et al., 2009; Szydlowski et al., 2009)。已报道的小麦基因组中存在九个编码淀粉合成酶的基因, 部分基因已被成功

克隆(Peng et al., 2001; Dian et al., 2003; Leterrier et al., 2008), 其中编码GBSSI蛋白的基因为*gbss I*或*waxy*基因, 有三个高度同源的*waxy*基因*Wx-A1*、*Wx-B1*和*Wx-D1*已被分离, 三个基因序列的相似性达到95.6~96.3%, *gbss II*基因分别位于2B、2D和2A染色体上, 为单拷贝(Murai et al., 1999; Vrinten et al., 2000)。SS I基因位于第7组染色体上, SS II基因也位于第7组染色体上(Li et al., 1999), SS III基因位于第1A染色体上(Li et al., 2000), SS IV基因是一类全新的SS同工型, 位于第1组染色体上(Leterrier et al., 2008), 目前对其研究尚少。本研究以小麦的相关数据为模板分析了淀粉合成酶基因家族的组成、结构和进化关系, 以期为进一步研究小麦淀粉生物

合成和淀粉品质遗传改良提供理论基础。

1 结果与分析

1.1 小麦淀粉合成酶基因家族的组成及结构

小麦基因组中至少存在九个编码淀粉合成酶的基因, 包括2个GBSS (*TaGBSS I*, *TaGBSS II*), 1个SS I (*TaSS I*), 3个SS II (*TaSS II-1*, *TaSS II-2*, *TaSS II-3*), 2个SS III (*TaSS III-1*, *TaSS III-2*)和1个SS IV (*TaSS IV*)。其中SS主要分布在质体基质, 部分也与淀粉粒结合, 是一种葡萄糖转移酶, 催化转移ADP-葡萄糖上的葡萄糖单元, 通过 α -1,4糖苷键不断增加寡聚糖的葡萄糖单位, 最终合成以 α -1,4糖苷键连接的聚糖, 聚糖又将作为淀粉分支酶的底物合成支链淀粉。GBSSI主要与淀粉颗粒结合, 在直链淀粉合成中起关键作用, 将ADP-葡萄糖的葡萄糖残基转移到 α -1,4葡聚糖链的非还原性末端(Vrinten et al., 2000)。为了深入分析小麦淀粉合成酶基因家族的结构与功能, 本研究对已经报道的小麦淀粉合成酶家族的氨基酸序列进行了比对和分析。结果发现, 九种酶的多肽长度差异较大, 如TaSSIII-1的长度几乎是GBSS II的三倍, 而属于同种同工酶的多肽序列长度相近。通过多序列比对结果, 分别属于同种同工酶的氨基酸序列之间相似性(60~95%)高于不同种同工酶(20~40%)之间(表1)。

通过多序列比对发现(图1), 这九种酶均存在两

个结构域: 催化结构域和糖原转移结构域。催化结构域的序列保守性要明显低于糖原转移结构域的保守性。按照两个结构域的起始位置, 可以将九种酶分为两个亚家族: 第一组: TaSS I、TaSS II-1、TaSS II-2、TaSS II-3、GBSS I和GBSS II, 第二组: TaSSIII-1、TaSSIII-2和TaSSIV。

通过分析发现, 两个亚家族的不同结构域的氨基酸组分并无显著差别, 均由20种不同的氨基酸组成, 但频率(%)不同(表2)。TaSS II三个同工酶的催化结构域和糖原转移结构域中各氨基酸频率差异不大, GBSS两个同工酶的不同结构域中氨基酸频率差异最大。然而, 通过GBSS I和GBSS II序列比对, 发现其同源性很高, 序列相似度达到60% (图2)。

同时, 本研究对部分淀粉合成酶基因全长做了结构分析, *TaSS II*三个同工酶基因具有相似的内含子/外显子结构, 有8个外显子和7个内含子。*TaSS IV*有16个外显子、15个内含子和2个非翻译区。*GBSS I*基因长度相对较短, 有11个外显子和10个内含子组成(图3)。

1.2 系统树分析

有研究将植物淀粉合成酶(SS)和原核生物糖原合成酶(GS)的氨基酸序列进行比对, 识别出了SS/GS中保守的氨基酸, 并认为这些保守的氨基酸可能形成了淀粉合成酶的催化结构域(Cao et al., 1999)。

表1 小麦淀粉合成酶氨基酸序列比较

Table 1 Comparison of wheat starch synthase protein sequences

酶 Isozyme	蛋白质 Protein	长度 Size a.a.	相似度(%) Similarity to (%)							
			SS II-1	SS II-2	SS II-3	SS III-1	SS III-2	SS IV	GBSS I	GBSS II
SS I	SS I	647	39	39	40	21	24	26	33	31
SS II	SS II-1	799	-	95	95	21	21	20	35	31
	SS II-2	798	-	-	95	22	21	21	35	31
	SS II-3	799	-	-	-	21	21	20	35	31
SS III	SS III-1	1628	-	-	-	-	61	28	21	24
	SS III-2	1271	-	-	-	-	-	26	22	23
SS IV	SS IV	914	-	-	-	-	-	-	22	23
GBSS I	GBSS I	605	-	-	-	-	-	-	-	60
GBSS II	GBSS II	599	-	-	-	-	-	-	-	-

注: *: 相似性数值(%)通过多序列比对软件 Clustal X 得到

Note: *: Values for % similarity of the conserved catalytic domain resulted from multi-sequence alignment analysis using the CLUSTAL X program, Comparisons between proteins belonging to the same isozyme are shown in boldface

TaSSI	-----SIVFVTGEAAPYAKSGGLG-----	158
TaSSII-1	YIGFEFPVEAKDDGWAVADDAGSF EHHQNHDSG PLAGENVMNVVVVAAECSPWCKTGGLG	327
TaSSII-2	YIGFEFPVEAKDDGRAVADDAGSF EHHQNHDSG PLAGENVMNVVVVAAECSPWCKTGGLG	326
TaSSII-3	YIGFEFPVEAKDDGWAVADDAGSF EHHQNHDSG PLAGENVMNVVVVAAECSPWCKTGGLG	327
TaGBSSI	-----AGMNLV FVGAEMAPWSKTGGLG-----	97
TaGBSSII	-----MPIIFVATEVHPWCKTGGLG-----	91
TaSSIII-1	AQEEQNIYTRLLQERREKEETMKRKAERSANIKAEMKAKTMRFFLLSQKHIVYTEPLEIR	1 080
TaSSIII-2	VEEERRIFERLQERTEKEDAGRRAEITARMKAEMKEKTMREFLLSQKHIVYTEPLEVR	723
TaSSIV	CSLVDLLQQKVKLVEDRFQACNSEMHSQIELYEHSIVEFHDTLSKLI EESEKRSLENFTG	348
	: :	
TaSSI	D-----VCGSLPIALAARG----HRVMVVMPPRY	182
TaSSII-1	D-----VAGALPKALAKRG----HRVMVVMPPRY	351
TaSSII-2	D-----VAGALPKALAKRG----HRVMVVMPPRY	350
TaSSII-3	D-----VAGALPKALAKRG----HRVMVVMPPRY	351
TaGBSSI	D-----VLGGLPPAMAANG----HRVMVISPPRY	121
TaGBSSII	D-----VVGGLPPALAAMG----HRVMTIAPRY	115
TaSSIII-1	AGTTVDVLYNPSNTVLNGKSEGWFRCFSNLMHSSGALPPQKMKVSGDGPLLKATVDVPP	1 140
TaSSIII-2	AGTTVDVLYNPSNTVLNGSPEVWFRCFSNLRWTHPSGPLPPQKMNVEVNGSHLQATVRVPL	783
TaSSIV	N-----MPSELWSKISLLIDGWLLEKKISYNDASMLREMVQKRDS	388
	* * .	
TaSSI	LNGSSDKNYAKALYTGKHIKIPCFGG SHEVTF FHEYRDNDVWFVDHPSY-HRPGS----	237
TaSSII-1	G-----DYEEAYDVGV RKYYKAAGQDMEVNYFHAYIDGVDFVFI DAPLFRHRQED----	401
TaSSII-2	G-----DYEEAYDVGV RKYYKAAGQDMEVNYFHAYIDGVDFVFI DAPLFRHRQED----	400
TaSSII-3	G-----DYEEAYDVGV RKYYKAAGQDMEVNYFHAYIDGVDFVFI DAPLFRHRQED----	401
TaGBSSI	D-----QYKDAWDTSVVSEIKVADEYERVRYFH CYKRGVDRVFDHPCFLEKVRGKTKE	175
TaGBSSII	D-----QYKDTWD TNVLVEVIVGDR TETVR FHCYKRGVDRVFDHPMFLEKVVWGKTGS	169
TaSSIII-1	DAYMDFV FSEWEEDGIYDN RNMGDYHIPVSDSIETENYMRI IHIAVEMAPVAKVGGLGD	1 200
TaSSIII-2	DAYMDFV FSEEEGGIYDN RNDGMDYHVPVSDSTAKEPPMHIVHIAVEMAPIAKVGGLGD	843
TaSSIV	R-----LREAYLSYRG TENREVM DNLLKMALPGTSSGLHIAHIAAEMAPVAKVGGLAD	441
	: . : :	
TaSSI	-LYGDNF-GAFGDNQFRYTL L CYAAACEAPL ILELGGY-----IYG-----	275
TaSSII-1	-IYG---GSRQEI M KRMILFCKAAVEVPWHVPCGGV-----PYGD-----	437
TaSSII-2	-IYG---GSRQEI M KRMILFCKAAVEVPWHVPCGGV-----PYGD-----	436
TaSSII-3	-IYG---GSRQEI M KRMILFCKAAVEVPWHVPCGGV-----PYGD-----	437
TaGBSSI	KIYGPDA G TDYEDNQLRFSL LCQAAL EAPRILDLNNN-----PYFSG-----P	218
TaGBSSII	KL YGPTTGTD FRDNQLRFCLLC LAAL EAPRVLN LNNS-----EYFSG-----P	212
TaSSIII-1	VVTSLSRAIQDLGHTVEVILPKYDCLNQSSVKDLHLHY-----QSFSWGGTEIKVWVGRVE	1 255
TaSSIII-2	VVTSLSRAIQDSGHKVEVIFLKYDCLNLSNVKDLHCR-----QSSTWGGTEIKVWFGKVE	898
TaSSIV	VISGLGKALQKKGHLVEIILPKYDCMQVDQVSNLKVLDVLVQSYFEGNMFNKNKIWTGTVE	501
	: . : : :	
TaSSI	--QNCMFVVND---WHASLVPVLLAAKYRPGYVYRDSRSTLVIHNLAHQGV EPASTY PDL	330
TaSSII-1	--GNLVFIAND---WHTALLPVYLKAYYRDHGLMQYTRS IMVIHNI AHQGRGPVDEF PFT	492
TaSSII-2	--GNLVFIAND---WHTALLPVYLKAYYRDHGLMQYTRS IMVIHNI AHQGRGPVDEF PFT	491
TaSSII-3	--GNLVFIAND---WHTALLPVYLKAYYRDHGLMQYTRS IMVIHNI AHQGRGPVDEF PFT	492
TaGBSSI	YGEDVV FVCND---WHTGLLACYLKSNYQSSGIYRTAKVAFCIHNI SYQGRFSFDDFAQL	275
TaGBSSII	YGENVVFVAND---WHTAVLPCYLKSMYKQNGIYVNAKVAFCIHNIAYQGRFPRVDFELL	269
TaSSIII-1	DLTVYFLEPQN---GMFGVGCVYGRNDRRFGFFCHSALEFILQNEFSPI IHCHDWSSA	1 312
TaSSIII-2	GISVYFLEPQN---GMFWVGCVYGKNDSE RFGFFCHSALEFL LQSGSSPDI IHCHDWSSA	955
TaSSIV	GLPVYFIEPQHPAMFFSRAQYYGEHDDFKRFSYFSRAALELLYQSGKKVDI IHCHDWQTA	561
	: . : : : :	
TaSSI	GLPPEWYGALEWVFP EWARRHALDK-----GE	357
TaSSII-1	ELP-EHY-----LEHFRLYDPVG-----GE	511
TaSSII-2	ELP-EHY-----LEHFRLYDPVG-----GE	510
TaSSII-3	ELP-EHY-----LEHFRLYDPVG-----GE	511
TaGBSSI	NLPDRFKSS---FDFIDGYDKPVE-----GR	298
TaGBSSII	NLPESFMPS---FDFVDGHVKPVV-----GR	292
TaSSIII-1	PVAWLYKEHYSQSRMASTRVVF TIHN-----LEF	1 341
TaSSIII-2	PVAWLYKQYVHNGLPNGRVVF TIHN-----LEF	984
TaSSIV	FVAPLYWDVYANLGFNSARICFTCHNF EYQGTAPARDLAWCGLDVEHLDRPDRMRD NSHG	621
	: .	

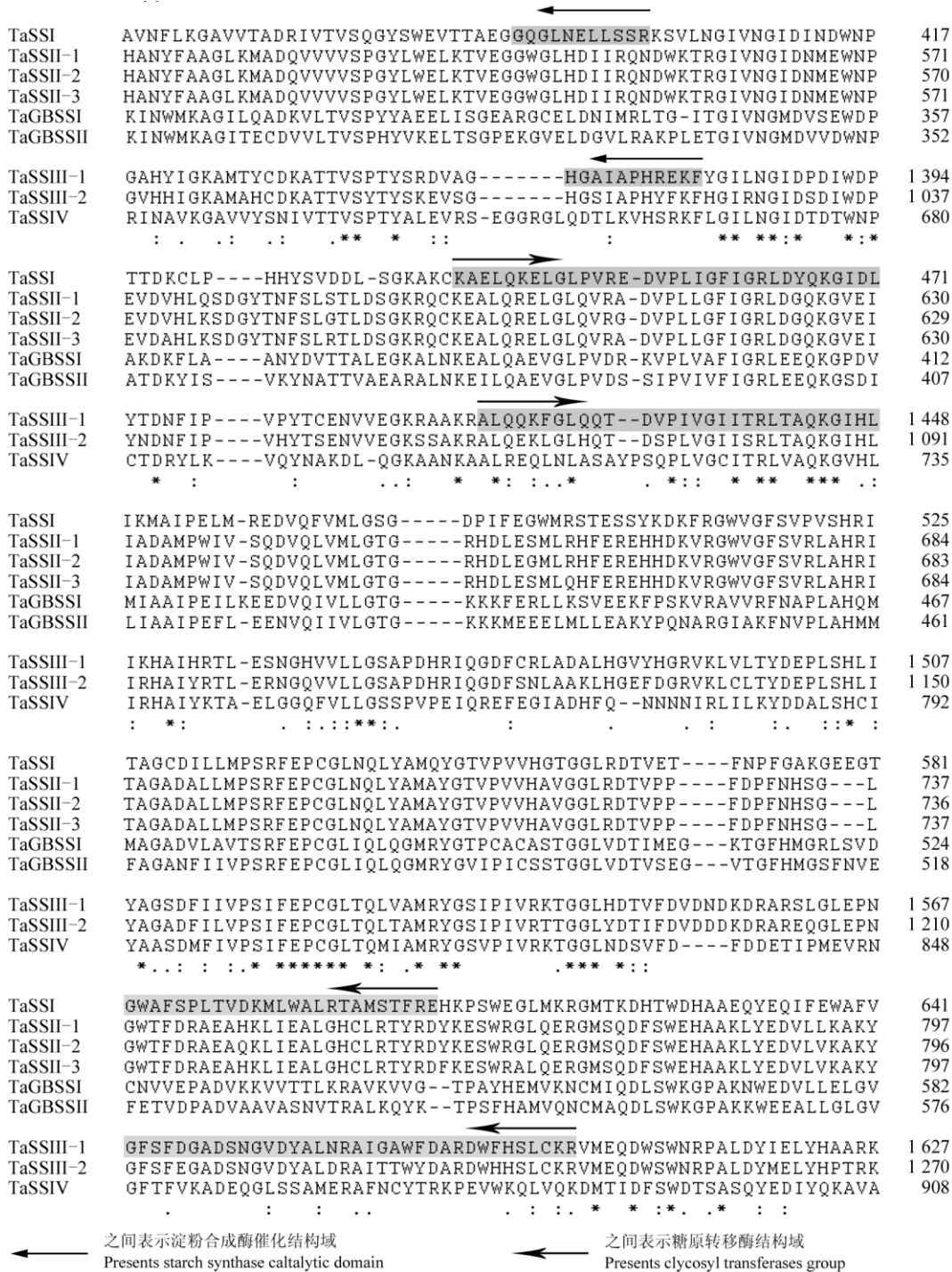


图1 小麦淀粉合成酶的氨基酸序列同源性比较

注: GenBank 登陆号: 小麦(*Triticum aestivum*): TaSS I, CAB99209; TaSS II-1, BAE48798; TaSS II-2, BAE48799; TaSS II-3, BAE48800; TaGBSS I, BAA77351; TaGBSS II, AAF14233; TaSS III-1, AAF87999; TaSS III-2, BAY56823; TaSS IV, AAK97773; 用 CLUSTAL X 得到的多序列比对结果中, “*” 号表示保守性极高的残基位点; “:” 号代表保守性略低的残基位点, 阴影部分表示结构域起始和终止位置

Figure 1 Comparison of the predicted amino acid sequences of wheat starch synthases

Note: The GenBank accession number shown as follow; Wheat (*Triticum aestivum*): TaSS I, CAB99209; TaSS II-1, BAE48798; TaSS II-2, BAE48799; TaSS II-3, BAE48800; TaGBSS I, BAA77351; TaGBSS II, AAF14233; TaSS III-1, AAF87999; TaSS III-2, BAY56823; TaSS IV, AAK97773; The CLUSTAL X program was employed to align the sequences; The identical and conserved residues were marked with “*” and “:”, respectively; The shadow areas mean the locations of starting and terminal domain

TaGBSSI	MAALVTSQLATSGTVLGITDRFRRAGFQGVPRSPADAPLGMRTTGASAAPKQSRKAHR	60
TaGBSSII	-MGSI PNYCSYQ TNSVGLKLSPHI QFQ--QSCNNEVMFLSMRNKTQLAKRRATNYGTHR	57
	. : . . : . . : * . : ** : . . * . . . * : . : **	
TaGBSSI	GTRRCLSMVVRATGSAGMNLVFGAEMAPWVSKTGGLGDVLGGLP PAMAANGHRVMI SPR	120
TaGBSSII	NSSRTPAPIVCSTG--MPIIFVATEVHPWCKTGGLGDVVGGGLPPALAAMGHRVMTIAPR	114
	. : * : : * : ** * : ** . : * : ** : ** : ** : ** : ** : ** : ** : ** : **	
TaGBSSI	YDQYKDAWDTSVVSEIKVADEYERVRYFHICYKRGVDRVFDHPCFLEKVRGKTKEDIYGP	180
TaGBSSII	YDQYKDTWDTNVLVEVIVGDRDETFRFFHICYKRGVDRVFDHMPFLEKVVWGTGSKLYGP	174
	*****:***: : : * . * . * * :*****:*****:*****:*****:****	
TaGBSSI	DAGTDYEDNQLRFSLLCQAAL EAPRI LD LNNNPYFSGPYGEDVVFVNDWHTG LLACYLK	240
TaGBSSII	TTGTDFRDNQLRFCLLCLAALEAPRVLN LNNSEYFSGPYGENVVFVANDWHTAVLPCYLK	234
	:***:*****:*** *****:***:***** *****:*****:*****:*****:****	
TaGBSSI	SNYQSSGIYRTAKVAFCIHNI SYQGRFSFDDFAQLNLPRFKSDFDIDGYDKPVEGRKI	300
TaGBSSII	SMYKQNGIYVNAKVAFCIHNIAYQGRFRVDFELNLPESFMPSFDFVDGHVKPVGGRKI	294
	* : . . *** .*****:*****. ** *****: * .*****:***: **** *	
TaGBSSI	NWMKAGILQADKVLTVSPYYAELISGEARGCELDNIMRLTG-ITGIVNGMDVSEWDPAK	359
TaGBSSII	NWMKAGITECDVVLTVSPHYVKELTS GPEKGVELDGVLRAKPLETGIVNGMDVVDWNPAT	354
	***** : * *****:*.*** ** :* *****:*. ***** :* :**.	
TaGBSSI	DKFLAANYDVTTALEGKALNKEALQAEVGLPVD RKVPLVAFIGRLEE QKGPVMIAAIPE	419
TaGBSSII	DKYISVKYNATTVAEARALNKEILQAEVGLPVDSSIPVIVF IGRLEE QKGS DILIAAIPE	414
	::: . : * . ** . * : *** ***** . : * : . ***** . : * : *****	
TaGBSSI	ILKEEDVQIVLLGTGKKKFERLLKSV EEFPSKVRVRFNAPLAHQMMAGADVLA VTSR	479
TaGBSSII	FL-EENVQIVLGTGKKKMEELMLLEAKYPQNARGIAKFNVPLAHMMFAGANF IIVPSR	473
	: * *****:*****:*. * : * * : . : . : . : . : . ***** : ***** : . : **	
TaGBSSI	FEPGLIQLG MRYGTPCACASTGGLVDTIMEGKTGFHMGRLSVCNVV EPADVKKVVT	539
TaGBSSII	FEPGLIQLG MRYGVIPICSSTGGLVDTVSEGVTGFHMGSFNVEFETVDPADVA AVASN	533
	***** ***** . * :*****: * ***** :. : . : * :***** * . : .	
TaGBSSI	LKRAVKVVGTPAYHEMVKNCMIQDLSWKGPKNWEDV LLELGV E GSEPGVIGEEIAPLAM	599
TaGBSSII	VTRALKQYKTPSFHAMVQNCMAQDLSWKGPAKKWEEALLGLGV EGSQPGIEGEEIAPLAK	593
	. : ** : * ** : * * * : ** ***** : ** : * * ***** : ** : *****	
TaGBSSI	ENVAAP	605
TaGBSSII	QNVATP	599
	:***:*	

图2 小麦淀粉粒结合淀粉合成酶的氨基酸序列同源性比较

注: GenBank 登录号: 小麦(*Triticum aestivum*): TaGBSSI, BAA77351; TaGBSSII, AAF14233; 用 CLUSTAL X 得到的多序列比对结果中, “*” 号表示保守性极高的残基位点; “.” 号代表保守性略低的残基位点

Figure 2 Comparison of the predicted amino acid sequences of wheat granule-binding starch synthases

Note: The GenBank accession number shown as above; The sequences were aligned using the CLUSTAL X program; The identical and conserved residues were marked with “*” and “.”, respectively

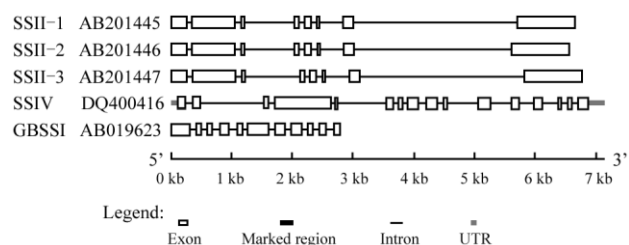


图3 部分小麦淀粉合成酶基因的结构分析

Figure 3 Schematic gene structure of partial wheat starch synthases gene

通过多序列比对和 MEGA 4.0 程序生成植物淀粉合成酶系统进化树(图4)。可将这31个氨基酸序列明显分为3个亚家族: SS I 和SS II 为同一亚家族, GBSS 为一个亚家族, SSIII 和SSIV 属同一个亚家族, 此结果和九个小麦淀粉合成酶的进化关系树相同(图5)。除个别节点外, 其余节点支持率都较高, 说明分枝的可信度高, 进化过程中亲缘关系较近, 如水稻SS IVb 和小麦SS IV 节点支持率为100, 基因同源性达到

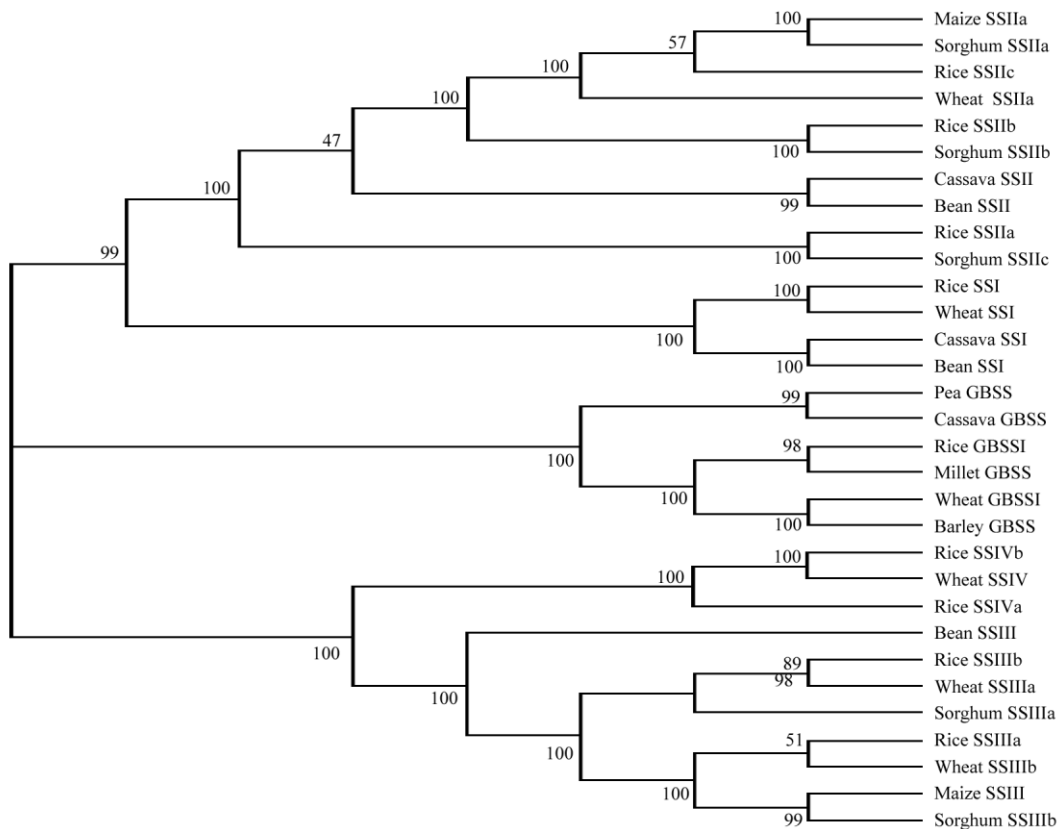


图 4 植物淀粉合成酶的进化关系

注: 利用 CLUSTAL X (Ver. 1.83) 软件对植物淀粉合成酶氨基酸序列进行多序列排列; 用软件 MEGA 4.0 构建邻接 (neighbor-joining, NJ) 树; 选择 Poisson Correction (泊松修正) 这一模型; 分析时不考虑插入和缺失, 节点支持率使用 1 000 次 Bootstrap 检验; 物种名称及蛋白质序列登陆号如下: 水稻: rice (*Oryza sativa*) GBSS I, ACJ86365; GBSS II, BAF26592; SS II c, AAL16661; SS II a, AAK64284; SS II b, AAK81729; SSIIIa, AAL40942; SSIIIb, AAM49811; SSIVa, AAQ82622; SSIVb, AAQ82623; 小麦: wheat (*Triticum aestivum*) GBSS I; BAA77351; GBSS II, SSIa, CAB99209; SS II a, CAB86618; SS II b, SSIIIa, AAF87999; SSIIIb, ABY56823; SSIV, AAK97773; 大麦: barley (*Hordeum vulgare*) GBSS, AAM74052; 玉米: maize (*Zea mays*) SS II a, AAS77569; SS I, ABQ23357; GBSS I, AAF79205; SSIIIb, ABP35815; 高粱: sorghum (*Sorghum bicolor*), SS II a, ACC86844; SS II b, ACC86845; SS II c, ABY27640; SSIIIb, ACC86847; 木薯: cassava (*Manihot esculenta*) SS I, ABV25893; 蚕豆: bean (*Phaseolus vulgaris*) SS I, BAD18845; SS II, BAD18846; SSIII, BAF49176; GBSS, ABU87173; 小米: Italian millet (*Setaria italica*) GBSS, BAC06486; 豌豆: pea (*Pisum sativum*) GBSS, CAC69955; SS, CAA61268

Figure 4 Phylogenetic tree showing the relationship of plant starch synthases based on their deduced amino acid sequences

Note: The MEGA 4.0 and CLUSTAL X (Ver. 1.83) program were employed to align the amino acid sequences of starch synthases; The tree was constructed using the neighbour-joining method and poisson correction model and confidence limits to branch points in the tree were assigned by bootstrapping the alignment with a random number generator seed of 111 and 1 000 trials; Genbank accession numbers of the starch synthases shown in the tree are as follow; Rice (*Oryza sativa*) GBSS I, ACJ86365; GBSS II, BAF26592; SS II c, AAL16661; SS II a, AAK64284; SS II b, AAK81729; SSIIIa, AAL40942; SS III b, AAM49811; SS IV a, AAQ82622; SSIVb, AAQ82623; wheat (*Triticum aestivum*) GBSS I; BAA77351; GBSS II, SS I a, CAB99209; SS II a, CAB86618; SS II b, SSIIIa, AAF87999; SSIIIb, ABY56823; SSIV, AAK97773; barley (*Hordeum vulgare*) GBSS, AAM74052; maize (*Zea mays*) SSIIa, AAS77569; SS I, ABQ23357; GBSS I, AAF79205; SSIIIb, ABP35815; sorghum (*Sorghum bicolor*), SS II a, ACC86844; SS II b, ACC86845; SS II c, ABY27640; SSIIIb, ACC86847; cassava (*Manihot esculenta*) SS I, ABV25893; bean (*Phaseolus vulgaris*) SS I, BAD18845; SS II, BAD18846; SSIII, BAF49176; GBSS, ABU87173; Italian millet (*Setaria italica*) GBSS, BAC06486; pea (*Pisum sativum*) GBSS, CAC69955; SS, CAA61268

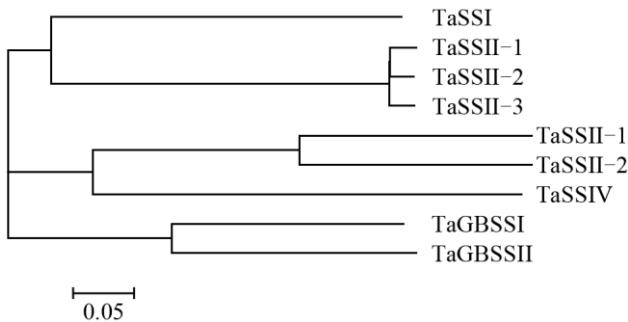


图5 小麦淀粉合成酶进化关系

Figure 5 Phylogenetic relationships of wheat starch synthases

86% (Letierrier et al., 2008)。每个亚家族中又可进一步细化为单子叶植物和双子叶植物。

2讨论

水稻和拟南芥基因组测序的完成推动了对其基因和染色体的研究(Yu et al., 2002), 使得人们在基因组水平上对植物重要基因家族进行挖掘、比较分析和功能预测变为现实。特别是生物信息学相关技术的开发, 在植物大规模的基因分离、基因功能

李春艳等, 2011, 小麦淀粉合成酶基因家族的生物信息学分析, 分子植物育种 Vol.9 No.64 (doi: 10.5376/mpb.cn.2011.09.0064)

妥的作用(土石寺, 2009)。普通小麦是一个开源八倍体物种, 有A、B、D三个不同的基因组。特别是B基因组含有丰富的抗病虫害等有益基因(于海霞和田纪春, 2008)。对小麦各个染色体组进行有关起源、进化、基因定位及基因与产量、生物性状的关联分析, 可更好的发掘和利用各染色体上的有利基因, 拓宽小麦的遗传变异基础, 为在小麦育种中如何引进新的遗传变异及杂优选育提供理论依据。

佃蔚敏等(2004)研究发现所有水稻SS基因家族蛋白质的C-端氨基酸序列保守性均较强, 和细菌的糖原合成酶同源, 与酵母和高等动物的糖原合成酶同源性极低, 并且认为植物淀粉合成酶与酵母和高等动物的糖原合成酶可能源自于不同的祖先。小麦基因组中存在九个编码淀粉合成酶的基因, 分散在不同染色体或染色体的不同位点, 非成簇分布。和小麦相似, 拟南芥和水稻基因组中淀粉合成酶也分别由不同位点基因编码, 说明淀粉合成酶基因家族可能是由染色体大片断倍增形成的(佃蔚敏, 2004)。水稻基因组中淀粉合成酶基因的倍增可能发生在单子叶和双子叶植物分化之后, 禾本科植物分化之前。小麦、玉米、高粱和大麦等禾本科植物基因组

中淀粉合成酶基因也发现有类似的倍增(佃蔚敏, 2004)。

目前, 关于植物淀粉合成酶基因的分子起源和进化研究较少。然而, 研究基因的倍增方式、分化途径可以预测该基因在自然选择的压力下进化过程中的命运(Sankoff, 2001), 这些研究将有利于对植物淀粉合成酶基因的表达特性、调控机理和生理功能的进一步探索, 同时可为深入研究小麦淀粉合成的分子机理和品质的遗传改良提供理论基础。

3试验材料与方法

3.1试验材料

从NCBI (<http://www.ncbi.nlm.nih.gov>)最新数据库下载六倍体小麦相关淀粉合成酶核酸和氨基酸序列, 以及其它相关作物的淀粉合成酶氨基酸序列。

3.2淀粉合成酶基因系统发生学分析

应用多序列比对工具Clustal X构建系统进化树(Ramu et al., 2003), 以相关淀粉合成酶氨基酸全序列比对的结果为基础, 用MEGA 4.0程序生成

法, 运行采用如下参数: 选择Poisson Correction(泊松修正)这一模型, 缺口设置为“Pairwise Deletion”, 节点支持率使用1 000次Bootstrap检验。

3.3淀粉合成酶氨基酸序列分析

通过Pfam网站(The protein families database, <http://www.sanger.ac.uk/Software/Pfam/>), 根据淀粉合成酶结构的不同, 将其进行分类。通过Pfam数据库进行系统分析得到的淀粉合成酶蛋白质结构域表现形式, 进一步分析不同结构域的各氨基酸频率。

3.4淀粉合成酶基因结构分析

通过Gene Structure Display Server在线基因结构分析工具, 分析部分淀粉合成酶基因全长的结构。

作者贡献

李春艳、李诚是本研究的实验设计和数据分析的执行人, 并完成论文初稿的写作; 李卫华、曹连甫指导实验设计与论文修改。全体作者都阅读并同意最终的文本。

致谢

本研究由石河子大学高层次人才科研启动资金专项(RCZX201004)和石河子大学科学技术研究发展计划动植物育种专项(GXJS2008-YZ04)共同资助。本文中提到了我们研究

中涉及的有关软件, 这并非我们为这些软件开发商提供推荐。

参考文献

- Cao H.J., Imparl-Radosevich H.G., Keeling P.L., James M.G., and Myers A.M., 1999, Identification of the soluble starch synthase activities of maize endosperm, *Plant Physiol.*, 120(1): 205-215
- Dian W.H., Chen J.Q., Liu F., and Wu P., 2003, Cloning and characterization of the granule-bound starch synthase II gene in rice: gene expression is regulated by the nitrogen level, sugar and circadian rhythm, *Planta*, 218(2): 261-268
- Dian W.H., Jiang W.H., and Wu P., 2005, Evolution and expression analysis of starch synthase III and IV in rice, *J. Exp. Bot.*, 56(412): 623-632
- Dian W.M., 2004, Analysis of the starch synthase gene family in rice: members, structure, evolution, function and expression, Dissertation for Ph.D., Supervisor: Wu Ping, Zhejiang University, pp.28-41 (佃蔚敏, 2004, 水稻淀粉合成酶基因家族基因组成、结构、进化、功能及其表达调控, 博士学位论文, 浙江大学, 导师: 吴平, pp.28-41)
- Jeng T.L., Wang C.S., Tseng T.H., Wu M.T., and Sung J.M., 2009, Nucleotide polymorphisms in the waxy gene of Na₃-induced waxy rice mutants, *J. Cereal Sci.*, 49(1): 112-116
- Leterrier M., Holappa L., Broglie K., and Beckles D., 2008, Cloning, characterisation and comparative analysis of a starch synthase IV gene in wheat: functional and evolutionary implications, *BMC Plant Biol.*, 8: 98-118
- Li Z., Chu X., Mouille G., Yan L., Kosar-Hashemi B., Hey S., Napier J., Shewry P., Clarke B., Appels R., Morell M., and Rahman S., 1999, The localization and expression of the class II starch synthases of wheat, *Plant Physiol.*, 120(4): 1147-1156
- Li Z., Mouille G., Kosar-Hashemi B., Rahman S., Clarke B., Gale K., Appels R., and Morell M., 2000, The structure and expression of the wheat starch synthase III gene. Motifs in the expressed gene define the lineage of the starch synthase III gene family, *Plant Physiol.*, 123(2): 613-624
- Murai J., Taira T., and Ohta D., 1999, Isolation and characterization of the three Waxy genes encoding the granule-bound starch synthase in hexaploid wheat, *Gene*, 234(1): 71-79
- Peng M., Hucl P., and Chibbar R.N., 2001, Isolation, characterization and expression analysis of starch synthase I from wheat (*Triticum aestivum* L.), *Plant Sci.*, 161(6): 1055-1062
- Ramu C., Sugawara H., Koike T., Lopez R., Gibson T.J., Higgins D.G., and Thomppson J.D., 2003, Multiple sequences alignment with the Clustal series of programs, *Nucl. Acids Res.*, 31(13): 3497-3500
- Sankoff D., 2001, Gene and genome duplication, *Curr. Opin. in Gen&Develop.*, 11(6): 681-684
- Szydlowski N., Ragel P., Raynaud S., Lucas M.M., Roldan I., Montero M., Munoz F.J., Ovecka M., Bahaji A., Planchot V., Pozueta-Romero J., D'Hulst C., and Merida A., 2009, Starch granule initiation in Arabidopsis requires the presence of either class IV or class III starch synthases, *Plant Cell*, 21(8): 2443-2457
- Tamura K., Dudley J., Nei M., and Kumar S., 2007, MEGA4: Molecular Evolutionary genetic analysis (MEGA) software version 4.0, *Mol. Bio. Evol.*, 24(8): 1596-1599
- Tomlinson K., Denyer K., and Callow J.A., 2003, Starch synthesis in cereal grains, *Advances in Botanical Research*, Academic Press, 40: 1-61
- Vrinten P., and Nakamura T., 2000, Wheat granule-bound starch synthase I and II are encoded by separate genes that are expressed in different tissues, *Plant Physiol.*, 122(1): 255-263
- Wang Y., Li Z.Y., Tang X.L., Lu S., Xu P., Zhang J., Fang K., and Xi J.H., 2009, Bioinformatic analysis of the NBS-LRR gene family in *Arabidopsis thaliana*, *Zhongguo Nongxue Tongbao (Chinese Agricultural Science Bulletin)*, 25(15): 40-45 (王岩, 李兆阳, 唐心龙, 卢珊, 许鹏, 张静, 方奎, 席景会, 2009, 拟南芥基因组 NBS-LRR 类基因家族的生物信息学分析, 中国农学通报, 25(15): 40-45)
- Yu H. X., and Tian J. C., 2008, Review of genome B in *T. aestivum* L., *Fenzi Zhiwu Yuzhong (Molecular Plant Breeding)*, 6(4): 724-732 (于海霞, 田纪春, 2008, 普通小麦 B 基因组的研究进展, 分子植物育种, 6(4): 724-732)
- Yu J., Hu S., Wang J., Wong G., Li S., 2002, A draft sequence of the rice genome (*Oryza Sativa* L. ssp. Indica), *Science*, 296(5565): 79-92

表 2 小麦淀粉合成酶不同结构域氨基酸组分分析

Table 2 Amino acid frequencies of starch synthases different domain. All frequencies are given in percent

		Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
A	TaSS I	9.27	1.93	4.63	5.02	4.25	10.42	3.86	3.09	3.09	9.27	1.16	3.86	5.41	1.93	4.63	6.18	3.86	9.27	2.32	6.56
	TaSS II-1	8.61	1.64	6.15	5.74	4.10	10.25	4.92	4.92	4.10	6.56	2.87	2.46	5.33	2.87	5.33	1.64	2.05	11.48	2.05	6.97
	TaSS II-2	8.61	1.64	6.15	5.74	4.10	10.25	4.92	4.51	4.10	6.97	2.87	2.46	5.33	2.87	5.33	1.64	2.05	11.48	2.05	6.97
	TaSS II-3	8.61	1.64	6.15	5.74	4.10	10.25	4.92	4.51	4.10	6.97	2.87	2.46	5.33	2.87	5.33	1.64	2.05	11.48	2.05	6.97
	TaGBSS I	7.28	2.68	8.05	5.36	5.36	8.05	1.92	4.51	5.75	7.66	1.92	4.60	4.60	2.68	5.75	5.75	2.68	7.66	1.53	6.13
	TaGBSS II	5.36	2.68	5.75	4.98	5.75	8.43	3.07	4.60	5.36	8.43	2.30	4.60	6.13	1.53	4.98	3.07	5.75	12.26	1.92	4.21
	Avg	7.93	2.05	6.15	5.42	4.63	9.58	3.90	3.45	4.43	7.67	2.31	3.44	5.35	2.45	5.22	3.37	3.11	10.58	1.98	6.28
B	TaSSIII-1	7.07	2.53	5.05	5.05	5.05	8.08	6.06	6.06	3.54	6.57	2.02	2.53	3.54	3.03	5.05	7.58	5.05	9.60	2.02	4.55
	TaSSIII-2	5.56	3.03	4.04	4.55	5.56	9.60	7.07	6.06	5.56	6.57	1.52	3.54	3.03	3.03	2.02	9.09	3.54	10.10	2.53	4.04
	TaSSIV	8.64	2.06	6.17	4.53	4.94	7.41	4.53	4.94	4.94	8.23	2.06	4.12	3.70	4.12	4.94	5.35	3.70	9.05	1.65	4.94
	Avg	7.20	2.50	5.16	4.69	5.16	8.29	5.79	5.63	4.69	7.20	1.88	3.44	3.44	3.44	4.07	7.20	4.07	9.55	2.03	4.54
C	TaSS I	4.49	1.12	5.06	7.30	5.62	11.24	1.69	4.49	5.62	9.55	5.06	1.12	6.18	2.81	5.62	5.62	5.62	6.74	2.81	2.25
	TaSS II-1	8.48	1.21	6.67	5.45	4.24	10.30	5.45	3.64	2.42	12.12	3.03	1.21	4.85	3.64	8.48	3.03	3.64	8.48	1.82	1.82
	TaSS II-2	7.88	1.21	6.67	5.45	4.24	11.52	4.85	3.64	2.42	12.12	3.03	1.21	4.85	4.24	8.48	2.42	3.64	8.48	1.82	1.82
	TaSS II-3	8.48	1.21	6.67	5.45	4.24	10.30	5.45	3.64	2.42	12.12	3.03	1.21	4.85	4.24	7.88	3.03	3.64	8.48	1.82	1.82
	TaGBSS I	8.65	2.70	4.32	7.57	3.24	8.11	1.62	4.32	8.65	9.19	4.32	1.62	5.41	3.78	4.86	2.70	4.86	12.97	0.00	1.08
	TaGBSS II	8.70	1.63	3.26	8.15	5.43	8.15	1.63	8.15	4.89	8.15	4.35	3.80	5.43	4.89	2.72	5.43	3.80	9.78	0.00	1.63
	Avg	7.77	1.54	5.37	6.62	4.51	9.88	3.36	4.70	4.51	10.46	3.84	1.73	5.28	3.93	6.24	3.74	4.22	9.21	1.34	1.73
D	TaSSIII-1	7.57	1.62	8.65	2.16	4.86	9.73	5.41	7.57	3.78	10.81	0.54	2.70	3.78	3.78	6.49	5.41	4.32	7.03	1.08	2.70
	TaSSIII-2	7.57	1.62	9.19	4.32	3.78	9.19	4.32	7.03	3.24	11.89	0.54	2.16	3.78	3.78	7.03	5.41	5.95	4.32	1.08	3.78
	TaSSIV	7.69	2.20	5.49	6.04	5.49	6.59	2.20	7.14	4.40	8.79	2.75	4.40	4.95	5.49	5.49	6.04	3.85	7.14	0.55	3.30
	Avg	7.61	1.81	7.79	4.17	4.71	8.51	3.99	7.25	3.80	10.51	1.27	3.08	4.17	4.35	6.34	5.62	4.71	6.16	0.91	3.26

注: A: 第一组(SS I, SS II 和 GBSS)淀粉合成酶催化结构域氨基酸组分分析; B: 第二组(SSIII和 SSIV)淀粉合成酶催化结构域氨基酸组分分析; C: 第一组(SS I, SS II 和 GBSS)淀粉合成酶糖原转移结构域氨基酸组分分析; D: 第二组(SSIII和 SSIV)淀粉合成酶糖原转移结构域氨基酸组分分析

Note: A: Amino acid frequencies of starch synthase catalytic domain of group 1 (SS I, SS II and GBSS); B: Amino acid frequencies of starch synthase catalytic domain of group 2 (SSIII and SSIV); C: Amino acid frequencies of glycosyl transferases domain of group 1 (SS I, SS II and GBSS); D: Amino acid frequencies of glycosyl transferases domain of group 2 (SSIII and SSIV)