

· 研究报告 ·

棉花类结瘤素MtN21基因家族生物信息学分析

袁园园, 王丽, 赵盼盼, 王林嵩*

河南师范大学生命科学学院, 新乡 453007

摘要 结瘤素基因主要参与豆科植物根瘤的形成。非结瘤植物中也存在类结瘤素基因, 主要调控植物的生长发育。MtN21 (*Medicago truncatula* NODULIN 21) 基因家族属于类结瘤素基因家族, 仅少数成员已被鉴定。以拟南芥 (*Arabidopsis thaliana*) MtN21 家族为参考, 对棉花 (*Gossypium hirsutum*) MtN21 基因家族进行了生物信息学分析, 发现棉花与拟南芥的 MtN21 基因同源性较高, 有共同的跨膜结构域 EamA 和 PLN00411; 棉花中仅含 PLN00411 结构域的 MtN21 蛋白等电点低于含 EamA 结构域的蛋白; 亚细胞定位主要在质膜、液泡膜和叶绿体, 少数在细胞核; MtN21 蛋白具有膜内侧的磷酸化位点。研究表明, 棉花 MtN21 为跨膜蛋白, 具有转运活性, 可能在棉花生长发育和病原体免疫方面发挥一定的作用。

关键词 生长素, 保守结构域, 进化分析, MtN21, 结瘤素基因

袁园园, 王丽, 赵盼盼, 王林嵩 (2016). 棉花类结瘤素 MtN21 基因家族生物信息学分析. 植物学报 51, 515–524.

在氮源缺乏时, 根瘤菌可侵染豆科植物形成根瘤。在此过程中, 特异性诱导表达的结瘤素基因负责根瘤的形成与功能 (Hohnjec et al., 2000; Sousa et al., 2001; 何恒斌和贾桂霞, 2013)。近期研究表明, 非结瘤植物如拟南芥 (*Arabidopsis thaliana*)、水稻 (*Oryza sativa*)、棉花 (*Gossypium hirsutum*)、玉米 (*Zea mays*)、杨树 (*Populus*) 和火炬松 (*Pinus taeda*) 等也存在结瘤素基因, 称为类结瘤素基因 (Busov et al., 2004; Ren et al., 2005; Denancé et al., 2014; Xu, 2015), 暗示结瘤素基因在植物生长发育中还发挥着其它功能。

最新研究表明, 类结瘤素蛋白在非结瘤植物生长发育中具有重要功能, 如转运营养物质、氨基酸和激素等 (Denancé et al., 2014)。非结瘤植物中类结瘤素基因被分成 7 个家族 (Denancé et al., 2014), MtN21 基因家族因与蒺藜苜蓿 (*Medicago truncatula*) NODULIN 21 基因同源而命名 (Gamas et al., 1996)。拟南芥 MtN21 基因家族的 44 个成员中仅少数基因的功能被鉴定。目前报道的非结瘤植物中, MtN21 主要参与氨基酸和生长素的转运, 可受生长素、细胞分裂素和乙烯等激素或环境因素的调控, 在植物次生细胞壁的形成、落叶时离层的形成、有机氮的分配和不定根

形成中发挥重要作用 (Busov et al., 2004; Ladwig et al., 2012; Ranocha et al., 2013; Denancé et al., 2014; Xu, 2015)。

棉花是我国重要的经济作物和纺织原料, 棉纤维的发育过程涉及次生细胞壁的形成, 并受多种植物激素的诱导和调控 (Basra and Malik, 1984)。棉花类结瘤素基因 GhNLP 参与棉纤维的发育 (张燕洁等, 2008); 而另一个棉花结瘤素基因仅在棉纤维发育过程中特异表达 (Ren et al., 2005)。鉴于 MtN21 蛋白参与次生细胞壁的形成, 并受多种激素的调控, 推测其在棉纤维发育过程中可能起重要作用。本研究旨在利用生物信息学软件对现有数据库中的棉花 MtN21 基因家族进行分析, 为揭示其在棉花生长发育中的功能奠定基础。

1 材料与方法

1.1 数据提取

本实验从拟南芥数据库 (<https://www.arabidopsis.org/>) 中获取拟南芥 MtN21 基因序列, 以 MtN21 *Gossypium* 为关键词在 GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) 中检索得到雷蒙德氏棉 (*Gossy-*

收稿日期: 2015-08-25; 接受日期: 2016-01-12

基金项目: 河南省高校科技创新团队支持计划、河南省生物学重点学科和河南师范大学博士启动课题 (No.qd14172)

* 通讯作者。E-mail: wls@htu.cn

pium raimondii) *MtN21* 基因序列 (Benson et al., 2011)。并以 Ladwig 等 (2012) 确定的 12 个拟南芥 *MtN21* 基因编码序列作为参考。在几种候选序列中, 如某基因存在选择性剪切, 仅选择最长的编码序列。利用 ProtParm 程序 (<http://web.expasy.org/protparam>) 获取各蛋白质的相对分子质量及等电点预测信息 (Gasteiger et al., 2003)。基因结构信息从 NCBI (<http://www.ncbi.nlm.nih.gov/gene/>) 数据库中获取 (Reddy et al., 2015)。

1.2 基因序列比对、进化分析及染色体定位

利用 Clustal W 软件对序列进行完全比对。用 MEGA5.0 软件, 通过邻接法 (neighbor joining, NJ) 构建基因分子进化树, 并通过 MapInspect 软件绘制染色体定位图。

1.3 蛋白质结构分析

蛋白质结构域信息通过 NCBI Conserved Domain Search (CDD, <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) 平台获取 (Aron et al., 2011)。通过 Phyre 2.0 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) (Kelley and Sternberg, 2009) 和 NetPhos 2.0 Server (<http://www.cbs.dtu.dk/services/NetPhos/>) 并结合 TMHMM Server v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) (Chen et al., 2003) 软件对 3 种保守结构域的三维构象、磷酸化位点以及跨膜区域进行对比分析。利用 WoLF PSORT 程序 (http://www.genscript.com/psort/wolf_psort.html) 进行蛋白质亚细胞定位预测 (Horton et al., 2007), 并运用 MultiLoc (<https://abi.inf.unituebingen.de/Services/MultiLoc>) 进行重复预测验证 (Höglund et al., 2006)。

2 结果与讨论

2.1 系统进化和基因结构

本研究共选取了 12 个拟南芥 *MtN21* 基因和 33 个棉花 *MtN21* 基因用于构建系统发育进化树和进行基因结构分析 (表 1; 图 1)。与 Ladwig 等 (2012) 报道的拟南芥进化树相类似, 本研究中棉花和拟南芥 *MtN21* 基因家族也分别分为 6 个类别, 其中 5 个类别 (I、III、IV、V 和

VI) 为 2 种植物共有, 第 II 类为棉花特异性类别, 第 IV 类 (Ladwig 分类) 为拟南芥特有。此外, 采用不同方法构建的进化树, 分类数目不变但进化层次有所变化。棉花和拟南芥的 *MtN21* 基因大多含有 7 个外显子, 少数含有 5–6 个外显子, 而拟南芥 *At1g11450* 和棉花 *Gr-8g62347* 的外显子个数分别为 4 和 2。棉花和拟南芥 *MtN21* 基因长度为 466–8 627 bp 不等, 内含子长度为 84–6 800 bp。棉花和拟南芥 *MtN21* 基因的外显子数目相对保守。

2.2 染色体定位

根据雷蒙德氏棉基因组测序数据, 查找 33 个棉花 *MtN21* 基因在染色体上的定位, 结果 (图 2) 显示, 33 个基因分别定位在第 1、2、4、5、6、7、8、9、10 和 12 号染色体上, 主要定位在第 6、8、9、10 和 12 号染色体上, 而在第 3、11 和 13 号染色体上没有相应的基因定位。

2.3 蛋白质保守结构域及亚细胞定位

利用 NCBI Conserved Domain Search (CDD) 预测 *MtN21* 蛋白的二级结构。拟南芥 12 个 *MtN21* 蛋白中含有 EamA、PLN00411 和 RhaT 三种保守结构域, 构成 5 种模型 (A、B、C、D 和 E) (图 3A–E); 棉花 33 个 *MtN21* 蛋白中含有 EamA 和 PLN00411 两种结构域, 构成 3 种模型 (A、B 和 C) (图 4A–C), 并与拟南芥的前 3 种模型相对应。棉花 *MtN21* 基因家族的第 I 和第 II 类基因编码蛋白以 B 结构模型为主; 第 III、IV 和 V 类均以 C 结构模型为主; 而第 VI 类则以 A 结构模型为主。棉花中仅含有 PLN00411 保守结构域的 *MtN21* 蛋白的等电点约为 4.79–6.56, 含有 1 个 EamA 和 1 个 PLN00411 保守结构域的 *MtN21* 蛋白的等电点约为 8.15–9.43, 含有 2 个 EamA 和 1 个 PLN00411 保守结构域的 *MtN21* 蛋白的等电点约为 8.92–9.65。

采用 2 种方法进行亚细胞定位预测的结果表明, 拟南芥 *MtN21* 基因编码蛋白主要位于质膜、液泡膜和叶绿体, 其比例分别为 33.3%–66.7%、8.3%–25% 和 8.3%; 棉花 *MtN21* 基因编码蛋白主要位于质膜、叶绿体和液泡膜, 其比例分别为 30.3%–51.5%、18.2%–27.3% 和 6.1%–15.2%。由此可知, 拟南芥和棉花的 *MtN21* 蛋白主要定位在质膜上, 并且编码定位于质膜蛋白基因在 6 个家族中广泛分布; 而编码定位于叶绿

表1 本研究所涉及的拟南芥和棉花 *MtN21* 基因Table 1 *MtN21* genes in Arabidopsis and cotton surveyed in this study

Species	Name	GenBank accession	Protein length	pI	Molecular weight (Da)	Subcellular
<i>Arabidopsis thaliana</i> (At)	At1g11450	NP_172612	260	8.70	28532.7	Chlo
	At1g01070	NP_563617	365	9.11	39809.9	Vacu
	At1g25270	NP_173898	355	9.28	38575.9	Plas
	At1g68170	NP_176984	356	9.23	38797.2	Vacu
	At1g44800	NP_175101	370	9.48	39947.4	Plas
	At4g08300	NP_192570	373	9.34	40517.0	Plas
	At1g43650	NP_175030	343	9.24	37715.0	Plas
	At5g64700	NP_201275	359	9.25	39752.5	Plas
	At1g75500	NP_565111	389	9.16	42571.0	Plas
	At3g53210	NP_566981	369	8.11	40804.8	Plas
	At3g28070	NP_189446	360	9.46	39530.9	Vacu
	At3g28100	NP_189449	353	9.68	38560.0	Plas
	<i>Gossypium raimondii</i> (Gr)	Gr09g66187	XP_012440994	108	4.8	12024.8
Gr01g98090		XP_012483464	108	4.79	11943.7	Chlo
Gr08g65377		XP_012439887	371	9.49	41239.0	Plas
Gr07g02450		XP_012489549	358	9.41	39336.6	Plas
Gr07g02306		XP_012489325	375	9.17	41030.6	Vacu
Gr06g00281		XP_012486761	347	8.15	37631.1	Vacu
Gr06g00012		XP_012486312	348	8.97	37166.1	Plas
Gr06g99877		XP_012486123	385	9.21	42284.5	Vacu
Gr06g98291		XP_012483762	353	9.19	38489.8	Vacu
Gr05g94682		XP_012479427	361	9.03	39931.4	Plas
Gr05g94679		XP_012479426	346	9.39	37918.2	Plas
Gr04g93519		XP_012477875	346	9.08	38172.3	Plas
Gr04g92066		XP_012475903	351	5.79	37930.1	Nucl
Gr02g84845		XP_012466223	385	9.05	42037.5	Plas
Gr12g80961		XP_012460967	363	9.65	39961.6	Plas
Gr12g80009		XP_012459521	353	8.94	38181.2	Vacu
Gr12g80008		XP_012459520	369	9.19	40485.9	Nucl
Gr12g78841		XP_012458054	361	8.50	39365.3	Chlo
Gr10g74610		XP_012452611	390	8.98	42451.3	Chlo
Gr10g73856		XP_012451468	386	9.27	41891.2	Plas
Gr10g73069		XP_012450148	384	9.52	41907.6	Plas
Gr10g73068		XP_012450147	384	9.33	41943.8	Chlo
Gr10g71567		XP_012448453	377	9.32	40907.5	Chlo
Gr09g71220		XP_012448086	377	9.39	41028.9	Plas
Gr09g69647		XP_012445870	385	9.36	41619.9	Chlo
Gr09g69646		XP_012445869	385	9.41	41746.2	Plas
Gr09g69000		XP_012444787	384	9.33	41683.1	Plas
Gr01g68549		XP_012443967	396	9.43	43602.6	Plas
Gr09g68216		XP_012443484	377	9.12	41525.5	Plas
Gr09g66102		XP_012440879	351	9.14	37804.9	Chlo
Gr08g65399		XP_012439925	389	9.51	42856.6	Plas
Gr08g62992		XP_012436446	391	8.92	42237.2	Plas
Gr08g62347		XP_012435654	90	6.56	9829.4	Chlo

Chlo: 叶绿体; Plas: 质膜; Vacu: 液泡膜; Nucl: 细胞核

Chlo: Chloroplast; Plas: Plasma membrane; Vacu: Tonoplast; Nucl: Nucleus

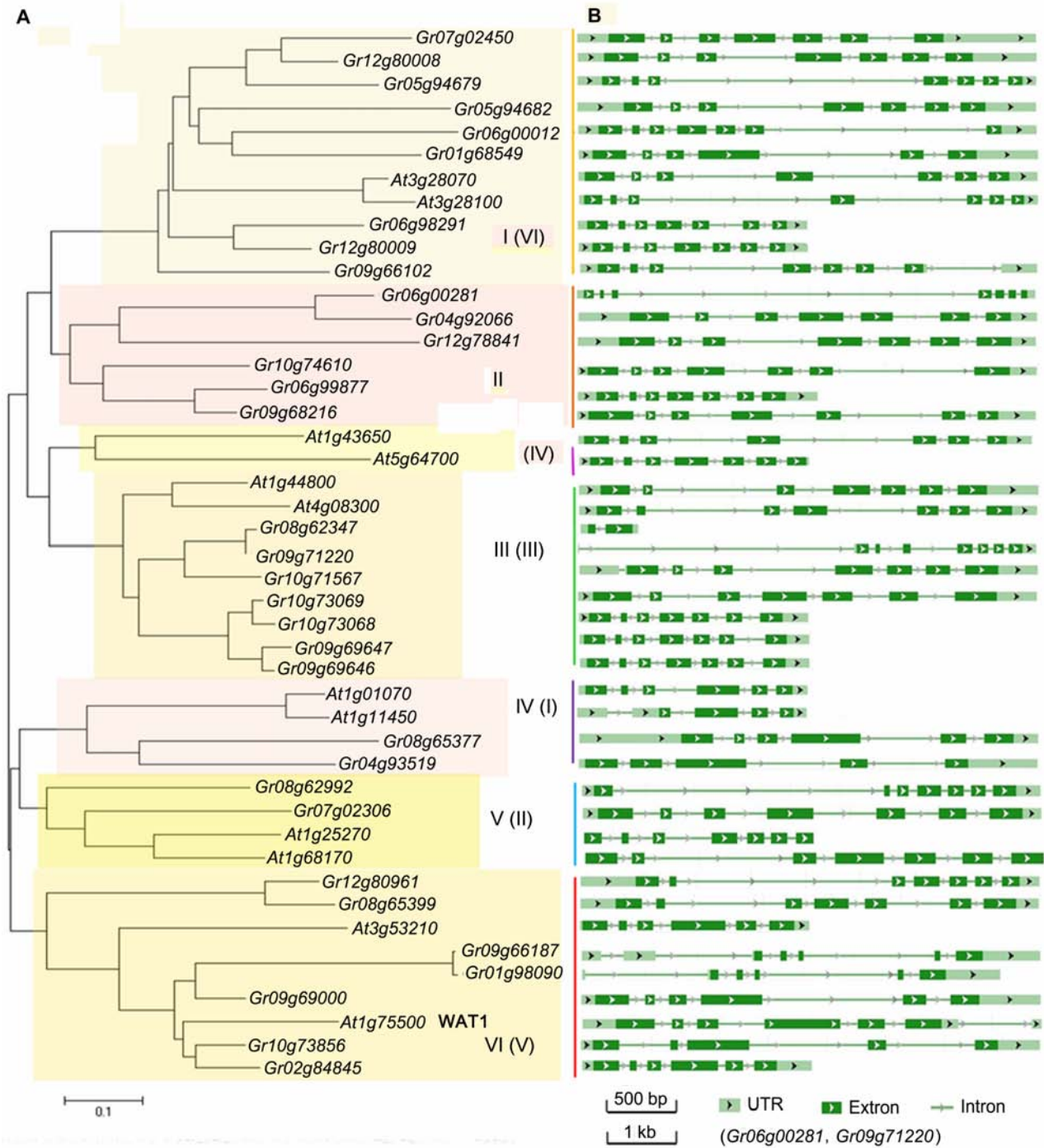


图1 拟南芥和棉花*MtN21*基因系统发育进化树(A)和基因结构(B)
括号内数字为Ladwig等(2012)的分类类别

Figure 1 The phylogenetic tree (A) and gene structure (B) of *MtN21* in Arabidopsis and cotton
The number in the parenthesis represents classification by Ladwig et al. (2012)

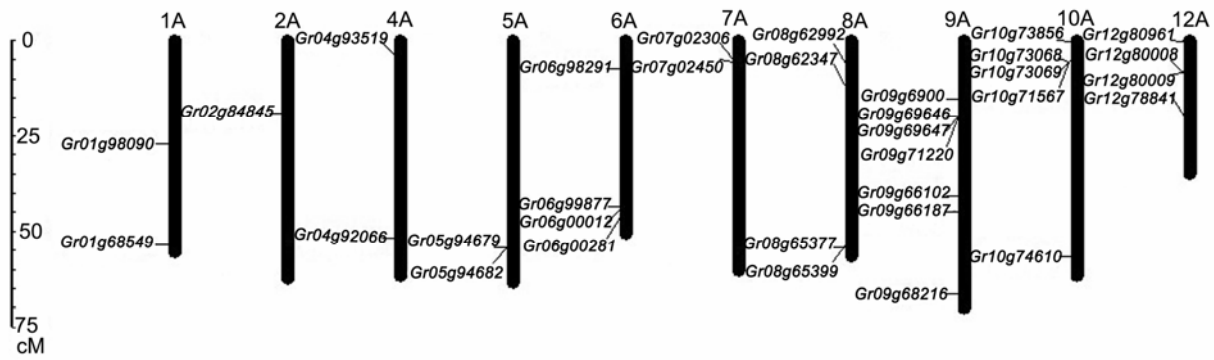


图2 棉花*MtN21*基因染色体定位

Figure 2 The chromosome location analysis of *MtN21* in cotton

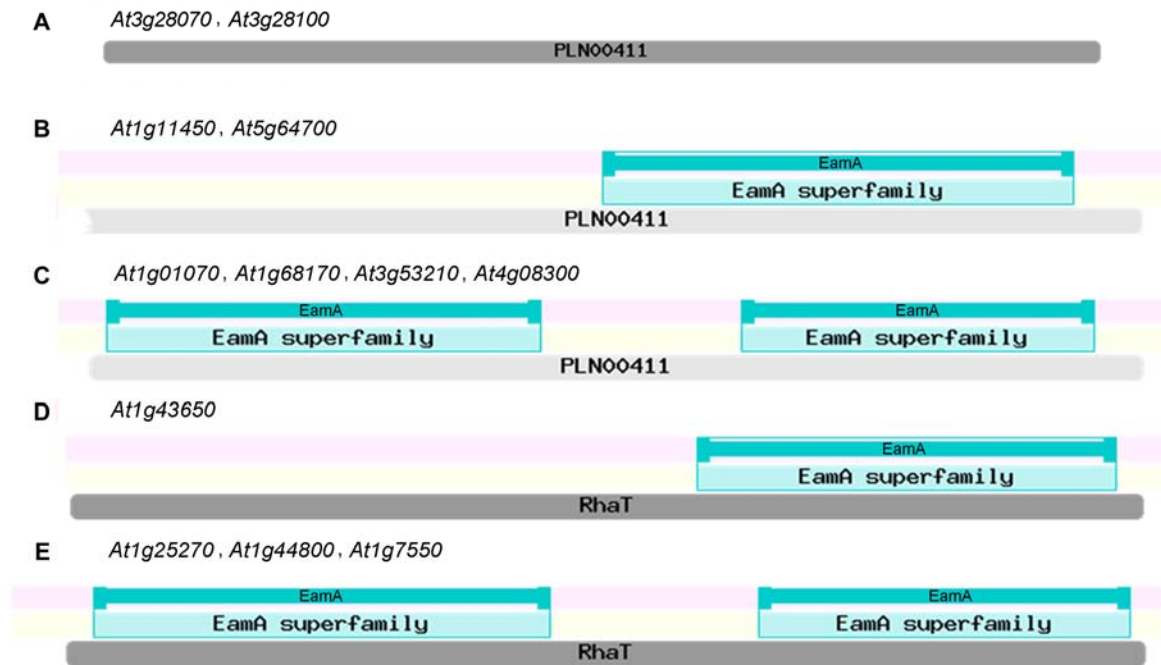


图3 拟南芥*MtN21*蛋白保守结构域预测

Figure 3 The predicted conserved domains of *MtN21* proteins in Arabidopsis

体蛋白的基因多分布在第II、III和VI家族中；编码定位于液泡膜蛋白的基因均分布于I、II和V家族中；编码定位于细胞核蛋白的基因均分布于第I和II家族中。

2.4 结构域空间构象

采用Phyre软件对EamA、PLN00411和RhaT三种结构域的三维构象进行分析，发现三者空间构象上非

常相似，均由3个 α 螺旋和一些无规则卷曲构成；但彼此也存在一些差异，如PLN00411和RhaT结构域中 α 螺旋圈数比EamA的多， α 螺旋之间的相对空间位置及无规则卷曲的长度也存在差异(图5A-C)。

2.5 结构域跨膜结构

利用TMHMM Server v.2.0软件对EamA、PLN00411

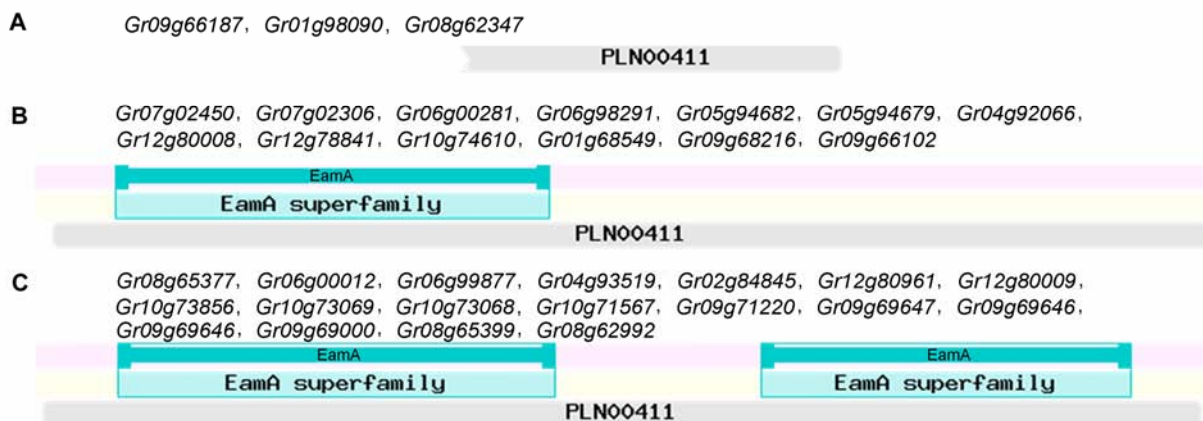


图4 棉花MtN21蛋白保守结构域预测

Figure 4 The predicted conserved domains of MtN21 proteins in cotton

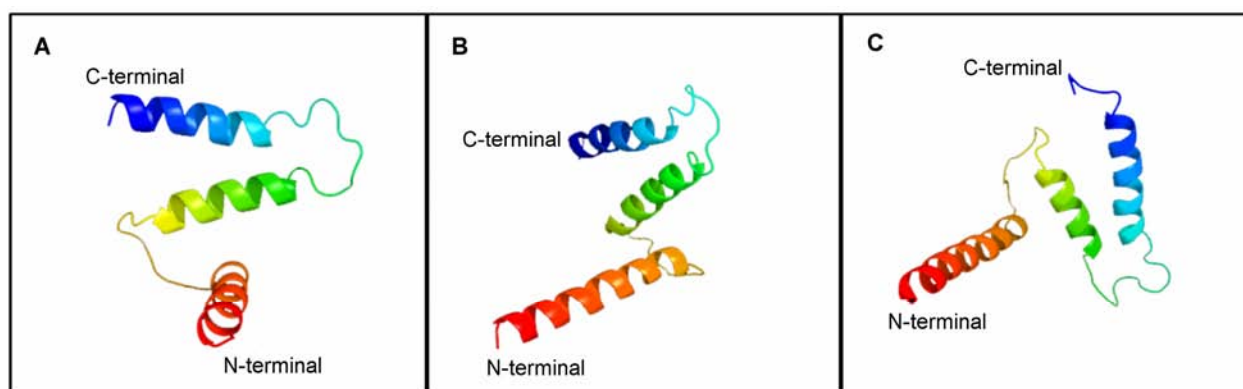


图5 3种结构域的三维构象预测

(A) EamA结构域; (B) PLN00411结构域; (C) RhaT结构域

Figure 5 The tertiary structure of 3 domains

(A) EamA domain; (B) PLN00411 domain; (C) RhaT domain

和RhaT三种结构域的跨膜结构进行分析,发现它们分别有4、10和9个跨膜区域。EamA结构域的N末端和C末端均位于膜外侧(图6A); PLN00411结构域的N末端和C末端均位于膜内侧(图6B); RhaT结构域的N末端位于膜外侧,而C末端位于膜内侧(图6C)。本研究表明,含有EamA、PLN00411和RhaT结构域的MtN21蛋白应为跨膜蛋白。

2.6 结构域磷酸化位点

利用NetPhos 2.0 Server软件对EamA、PLN00411和RhaT三种结构域的磷酸化位点进行分析,结果表明,

EamA结构域含有1个Ser和1个Thr的可能磷酸化位点(图7A); PLN00411结构域含有12个Ser、1个Thr和2个Tyr的可能磷酸化位点(图7B); RhaT结构域含有1个Ser、4个Thr和1个Tyr的可能磷酸化位点(图7C)。进一步对磷酸化位点在二级结构上的位置进行分析,结果表明, EamA结构域中第40位的Thr和56位的Ser均位于膜内(图8A); PLN00411结构域中第69、131、137、340、342和352位的Ser均位于膜内,第37、177、181、182和183位的Ser、第36位的Thr、第170位的Tyr均位于膜外,第234位的Ser、第270位的Tyr均位于跨膜的 α 螺旋上(图8B); RhaT结构域中第175位的Ser、第53

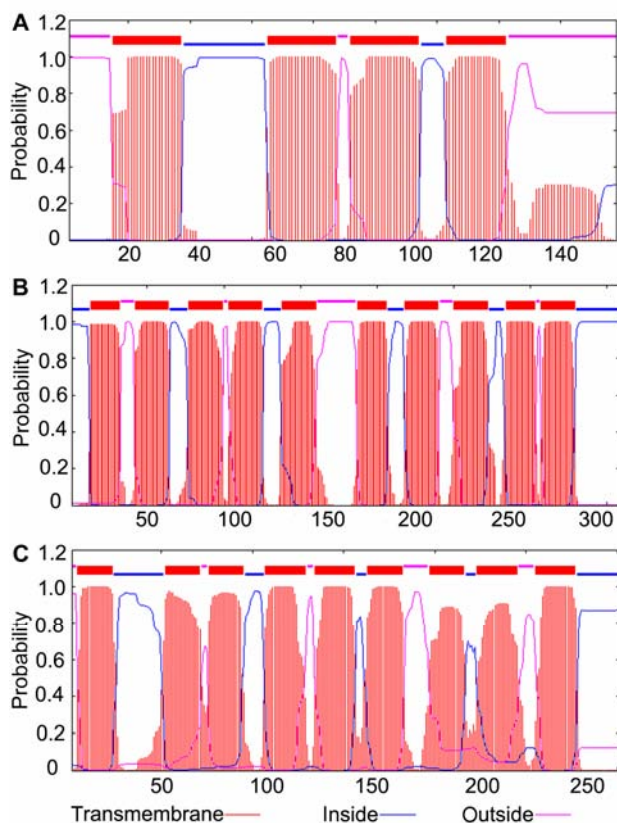


图6 3种结构域的跨膜结构分析

(A) EamA结构域; (B) PLN00411结构域; (C) RhaT结构域。横轴表示氨基酸编号; 纵轴表示可能性的比率。上线代表膜外部分; 下线代表膜内部分; 粗线代表跨膜区。

Figure 6 Analysis of transmembrane helices of 3 domains (A) EamA domain; (B) PLN00411 domain; (C) RhaT domain. The x-axis indicates the position in the amino acid sequence starting at the N-terminal end and the y-axis indicates the probability of residing inside of, outside of, or in the membrane. Different positions relative to the membrane are coded as follows: thin lines outside represent outside of the membrane, thin lines underside represent inside of the membrane, and solid bars represent transmembrane helices.

和58位的Thr均位于膜内, 第85和138位的Thr均位于膜外, 第159位的Tyr位于跨膜的 α 螺旋上(图8C)。

2.7 讨论

结瘤素基因在根瘤植物中主要参与根瘤的形成与结构维持、固氮活性的保护和固氮产物的代谢等过程 (Hohnjec et al., 2000; Sousa et al., 2001)。近年来,

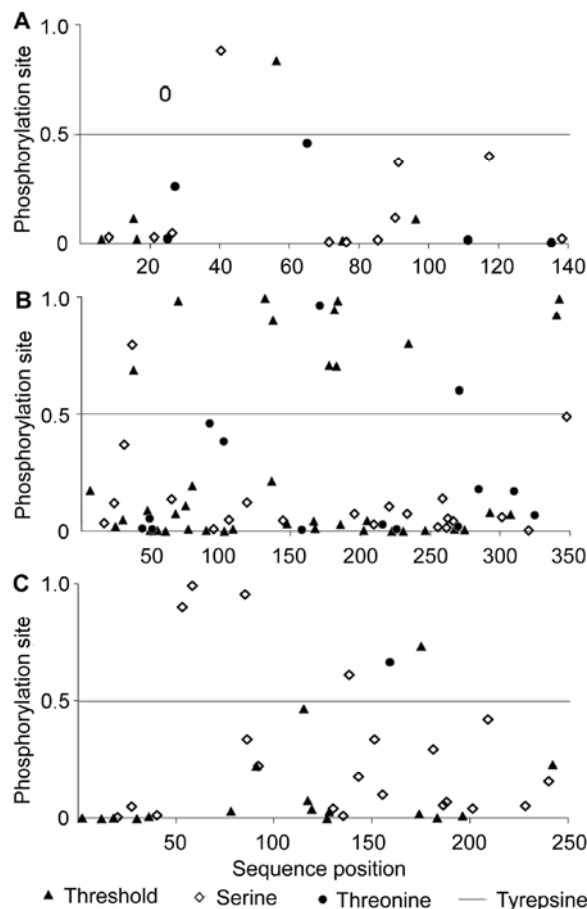


图7 3种结构域的磷酸化位点分析

(A) EamA结构域; (B) PLN00411结构域; (C) RhaT结构域

Figure 7 Phosphorylation site analysis of 3 domains (A) EamA domain; (B) PLN00411 domain; (C) RhaT domain

在非结瘤植物中发现了大量类结瘤素基因, 它们的功能不再是参与根瘤发生, 而是在植物生长发育过程中发挥其它重要作用 (Denancé et al., 2014)。MtN21家族是植物类结瘤素基因家族之一, 广泛存在于非结瘤植物中 (Denancé et al., 2014)。本研究对棉花的MtN21家族基因进行了生物信息学分析, 为进一步探讨其在棉花生长发育中的功能奠定了基础。

目前, 非结瘤植物中模式植物拟南芥的MtN21基因家族包括47个成员, 其编码蛋白可分为6大类 (Ladwig et al., 2012)。水稻中MtN21基因家族有17个成员, 可分为5大类, 与拟南芥中的前5类相同 (Ladwig et al., 2012)。已测序完成的雷蒙德氏棉中MtN21基因家族包括不少于33个成员, 也可分为6大

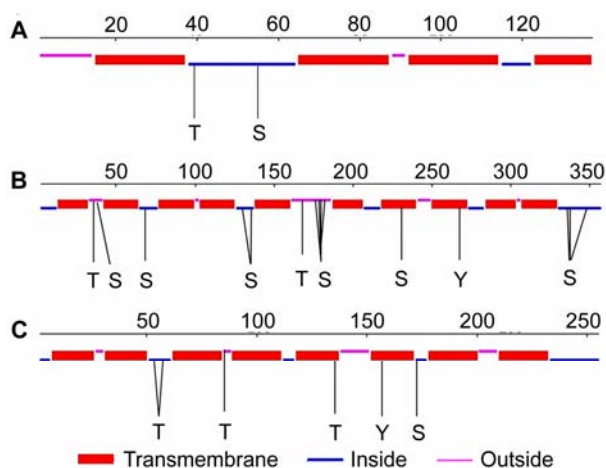


图8 磷酸化位点在3种结构域二级结构上的位置
(A) EamA结构域; (B) PLN00411结构域; (C) RhaT结构域。上线代表膜外部分; 下线代表膜内部分; 粗线代表跨膜区。T: 苏氨酸; S: 丝氨酸; Y: 酪氨酸。

Figure 8 The positions of phosphorylation sites on the 3 domains' secondary structure
(A) EamA domain; (B) PLN00411 domain; (C) RhaT domain. Thin lines outside represent outside of the membrane; thin lines underside represent inside of the membrane, and solid bars represent transmembrane helices. T: Threonine; S: Serine; Y: Tyrosine.

类, 且其中5类与拟南芥共有, 表明*MtN21*基因在非结瘤植物中较为保守(Busov et al., 2004), 但同时也具有种属特异性。本研究中*MtN21*基因家族中第II类为棉花所特有。此外, *MtN21*基因家族成员在基因结构上也具有一定的保守性, 外显子多为7个。

植物*MtN21*属于药物/代谢物转运蛋白超家族, 经预测此类蛋白为具有10–12个跨膜结构域的膜蛋白(Denancé et al., 2014)。本研究结果显示, 棉花和拟南芥中*MtN21*蛋白的共同结构域为EamA和PLN00411, 此外拟南芥的1个特有结构域为RhaT, 3种保守结构域具有极为相似的空间构象, 均由3个 α 螺旋和一些无规则卷曲构成, 且均有多个跨膜区域, 表明*MtN21*蛋白在跨膜螺旋部分保守性最高(Busov et al., 2004)。EamA结构域命名源于对大肠菌中氨基酸输出蛋白如EmaA (O-acetylserine and cysteine)、RhtA (threonine and homoserine)和YdeD (cysteine, asparagine and glutamine)结构的分析预测, 这些蛋白均具有药物/代谢物转运结构域(Franke et al.,

2003; Livshits et al., 2003)。该结构域是*MtN21*蛋白典型的结构域, 因此*MtN21*蛋白又被称为类EamA转运蛋白(Denancé et al., 2014)。RhaT家族(rhamnose transporter families)是药物/代谢物转运超家族DMT (drug/metabolite transporter superfamily)的14个家族之一, 主要运输糖类、氨基酸及其代谢产物(Jack et al., 2001)。PLN00411结构域的功能目前尚不清楚。棉花中仅含PLN00411结构域的*MtN21*蛋白等电点均明显小于含1个或2个EamA结构域的蛋白, 表明EamA结构域的等电点较大。

目前研究得较为清楚的*MtN21*是拟南芥的WAT1 (the walls are thin1)和SIAR1 (siliques are red1)。前者编码1个液泡膜定位的生长素转运蛋白, 参与茎纤维的次生细胞壁形成(Ranocha et al., 2010, 2013), 同时可通过调控水杨酸和色氨酸交叉代谢途径影响植株对导管类病菌的应答(Denancé et al., 2013); 后者编码一个质膜定位的氨基酸双向转运蛋白, 可调控有机氮的分配(Ladwig et al., 2012)。杨树中与拟南芥同源的WAT1也定位在液泡膜, 参与杨树落叶时离层的形成(Xu, 2015)。本研究中拟南芥*MtN21*蛋白的亚细胞定位预测其主要在质膜, 其次是液泡膜和叶绿体; 而棉花*MtN21*蛋白亚细胞定位也主要在质膜, 其次是叶绿体和液泡膜, 细胞核定位是其特有的, 功能可能也与氨基酸、生长素或其它代谢物的转运有关。

植物中类结瘤素26蛋白可发生翻译后磷酸化修饰, 磷酸化可提高其转运活性并受生长发育和环境因素的调控(Wallace et al., 2006)。本研究结果显示, 拟南芥和棉花的3种保守结构域均可能存在磷酸化修饰位点, 由于使蛋白质发生磷酸化和去磷酸化的酶主要位于细胞内(Hunter, 1995; Nishi et al., 2015), 因此膜内侧的Ser和Thr调节*MtN21*功能的可能性较大。*MtN21*蛋白主要执行转运功能, 其转运活性可能也受磷酸化调控。

目前结瘤素基因在非结瘤植物中的生物功能逐渐受到重视。鉴于植物中结瘤素基因保守性较高, 通过与拟南芥对比, 本研究采用生物信息学方法对棉花类结瘤素*MtN21*家族进行了系统发育进化、基因结构、亚细胞定位、磷酸化位点、功能结构域以及三维构象分析。结果表明, 棉花*MtN21*具有保守跨膜结构域, 为跨膜蛋白, 亚细胞定位主要在质膜, 应具有转运活性, 该活性可能受磷酸化的调控。该结果为进一

步研究MtN21在棉花生长发育中的功能提供了理论依据。

参考文献

- 何恒斌, 贾桂霞 (2013). 豆科植物早期共生信号转导的研究进展. *植物学报* **48**, 665–675.
- 张燕洁, 朱一超, 郭旺珍, 张天真 (2008). 与棉纤维发育相关基因 *GhSAMS*、*GhNLP* 的克隆、鉴定与定位. *中国农业科学* **41**, 2581–2588.
- Aron MB, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, Carol WS, Jessica HF, Lewis YG, Renata CG, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011). CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res* **39**, 225–229.
- Basra AS, Malik CP (1984). Development of the cotton fiber. *Int Rev Cytol* **89**, 65–113.
- Benson DA, Ilene KM, Lipman DJ, Ostell J, Sayers EW (2011). GenBank. *Nucleic Acids Res* **39**, 33–36.
- Busov VB, Johannes E, Whetten RW, Sederoff RR, Spiker SL, Lanz-Garcia C, Goldfarb B (2004). An auxin-inducible gene from loblolly pine (*Pinus taeda* L.) is differentially expressed in mature and juvenile-phase shoots and encodes a putative transmembrane protein. *Planta* **218**, 916–927.
- Chen YJ, Yu P, Luo JC, Jiang Y (2003). Secreted protein prediction system combining CJ-SPHMM, TMHMM and PSORT. *Mamm Genome* **14**, 859–865.
- Denancé N, Ranocha P, Oria N, Barlet X, Riviere MP, Yadeta KA, Hoffmann L, Perreau F, Clement G, Alesandra MG, Berg GC, Savelli B, Fournier S, Aubert Y, Pelletier S, Thomma BP, Molina A, Jouanin L, Marco Y, Goffner D (2013). Arabidopsis *wat1* (*walls are thin1*)-mediated resistance to the bacterial vascular pathogen, *Ralstonia solanacearum*, is accompanied by cross-regulation of salicylic acid and tryptophan metabolism. *Plant J* **73**, 225–239.
- Denancé N, Szurek B, Noël LD (2014). Emerging functions of nodulin-like proteins in non-nodulating plant species. *Plant Cell Physiol* **55**, 469–474.
- Franke I, Resch A, Dassler T, Maier T, Böck A (2003). YfiK from *Escherichia coli* promotes export of O-acetylserine and cysteine. *J Bacteriol* **185**, 1160–1166.
- Gamas P, Niebel FC, Lescure N, Cullimore J (1996). Use of a subtractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. *Mol Plant Microbe Interact* **9**, 233–242.
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**, 3784–3788.
- Höglund A, Dönnies P, Blum T, Adolph HW, Kohlbacher O (2006). MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics* **22**, 1158–1165.
- Hohnjec N, Küster H, Albus U, Frosch SC, Becker JD, Pühler A, Perlick AM, Frühling M (2000). The broad bean nodulin VfeNOD18 is a member of a novel family of plant proteins with homologies to the bacterial MJ0577 superfamily. *Mol Gen Genet* **264**, 241–250.
- Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Collier JA, Nakai K (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res* **35**, 585–587.
- Hunter T (1995). Protein kinases and phosphatases: the Yin and Yang of protein phosphorylation and signaling. *Cell* **80**, 225–236.
- Jack DL, Yang NM, Saier MH Jr (2001). The drug/metabolite transporter superfamily. *Eur J Biochem* **268**, 3620–3639.
- Kelley LA, Sternberg MJ (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363–371.
- Ladwig F, Stahl M, Ludewig U, Hirner AA, Hammes UZ, Stadler R, Harter K, Koch W (2012). Siliques Are Red1 from Arabidopsis acts as a bidirectional amino acid transporter that is crucial for the amino acid homeostasis of siliques. *Plant Physiol* **158**, 1643–1655.
- Livshits VA, Zakataeva NP, Aleshin VV, Vitushkina MV (2003). Identification and characterization of the new gene *rhtA* involved in threonine and homoserine efflux in *Escherichia coli*. *Res Microbiol* **154**, 123–135.
- Nishi H, Demir E, Panchenko AR (2015). Crosstalk between signaling pathways provided by single and multiple protein phosphorylation sites. *J Mol Biol* **427**, 511–520.
- Ranocha P, Denancé N, Vanholme R, Freyrier A, Martinez Y, Hoffmann L, Kohler L, Pouzet C, Renou JP, Sundberg B, Boerjan W, Goffner D (2010). Walls are thin1 (WAT1), an Arabidopsis homolog of *Medicago truncatula* NODULIN21, is a tonoplast-localized protein

- required for secondary wall formation in fibers. *Plant J* **63**, 469–483.
- Ranocha P, Dima O, Nagy R, Felten J, Corratge FC, Novak O, Morreel K, Lacombe B, Martinez Y, Pfrunder S, Xu J, Renou JP, Thibaud JB, Ljung K, Fischer U, Martinoia E, Boerjan W, Goffner D** (2013). Arabidopsis WAT1 is a vacuolar auxin transport facilitator required for auxin homeostasis. *Nat Commun* **4**, 2625.
- Reddy PS, Rao TS, Sharma KK, Vadez V** (2015). Genome-wide identification and characterization of the aquaporin gene family in *Sorghum bicolor* (L.). *Plant Gene* **1**, 18–28.
- Ren MZ, Chen QJ, Li L, Zhang R, Guo SD** (2005). Functional analysis of nodulin-like promoter in transgenic cotton plants. *J Integr Plant Biol* **47**, 1254–1259.
- Sousa C, Johansson C, Charon C, Manyani H, Sautter C, Kondorosi A, Crespi M** (2001). Translational and structural requirements of the early nodulin gene *enod40*, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. *Mol Cell Biol* **21**, 354–366.
- Wallace IS, Choi WG, Roberts DM** (2006). The structure, function and regulation of the nodulin 26-like intrinsic protein family of plant aquaglyceroporins. *Biochim Biophys Acta* **1578**, 1165–1175.
- Xu J** (2015). The Role of Auxin in Abscission of Organs and Tissues. Doctoral Thesis. Swedish: Acta Universitatis Agriculturae Sueciae. pp. 1–56.

Bioinformatics Analysis of the *Medicago truncatula* NODULIN 21-like Gene Family in Cotton

Yuanyuan Yuan, Li Wang, Panpan Zhao, Linsong Wang*
College of Life Sciences, Henan Normal University, Xinxiang 453007, China

Abstract Nodulin genes are involved in nodulation in leguminous species. Nodulin-like genes are also found in non-nodulating species, playing important roles during plant development. The *Medicago truncatula* NODULIN 21 (*MtN21*) family is one member of the nodulin-like gene family. However, only a few *MtN21*-like genes have been characterized. We compared *MtN21*-like genes in cotton with *MtN21*-like genes in Arabidopsis by bioinformatics methods. *MtN21*-like genes in cotton are homologous to those in Arabidopsis. *MtN21*-like proteins in cotton and Arabidopsis both have the same transmembrane domains, such as EamA and PLN00411. In cotton, the isoelectric points are predicted to be lower in *MtN21*-like proteins containing only the PLN00411 domain than those of *MtN21*-like proteins containing the EamA domain. The *MtN21*-like proteins are predicted to localize to the plasma membrane, tonoplast, chloroplast or nucleus in cotton. *MtN21*-like proteins in cotton have phosphorylation sites in the cytoplasmic portion of membrane. *MtN21*-like proteins in cotton may be transmembrane proteins with transporter activities. They may play roles in development and defense against pathogens in cotton.

Key words auxin, conserved domain, evolution analysis, *MtN21*, nodulin gene

Yuan YY, Wang L, Zhao PP, Wang LS (2016). Bioinformatics analysis of the *Medicago truncatula* NODULIN 21-like gene family in cotton. *Chin Bull Bot* **51**, 515–524.

* Author for correspondence. E-mail: wls@htu.cn