

· 特邀综述 ·

大豆基因组解析与重要农艺性状基因克隆研究进展

夏正俊*

中国科学院东北地理与农业生态研究所, 哈尔滨 150081

摘要 自2010年大豆(*Glycine max*)基因组论文在*Nature*杂志上正式发表以来, 以中国为主的多国科学家已对众多不同地理来源的野生大豆、农家品种及栽培品种进行了基因组重测序, 并通过比较基因组学等技术手段, 揭示了大豆基因组在进化与驯化过程中的基本规律。利用大豆基因组信息, 科学家已成功克隆了调控大豆生育期、生长(结荚)习性、种子形态性状与组分、结瘤与养分利用效率、生物及非生物胁迫抗(耐)性等重要农艺性状的功能基因, 并初步揭示了其相关的作用机理, 取得了一系列突破性进展, 为今后进一步精细解析大豆基因组与相关性状的基因调控网络奠定了坚实的基础, 同时为通过分子(模块)设计育种来培育新品种创造了必要条件。

关键词 大豆, 研究进展, 基因组, 农艺性状, 克隆, 基因功能

夏正俊 (2017). 大豆基因组解析与重要农艺性状基因克隆研究进展. *植物学报* 52, 148–158.

大豆(*Glycine max*)是我国及世界上主要栽培作物之一, 种植大豆具有共生固氮的生态效应。大豆富含异黄酮及维生素E等多种有益于人体健康的营养成分, 豆制品的加工与食用技术在我国传统的饮食文化中占有重要地位。大豆起源于中国, 从野生大豆(*G. soja*)到农家品种或地方品种经历了5 000年或更长时间的驯化过程(Stupar and Specht, 2013), 而现代栽培大豆(*G. max*)是近代以农家品种或地方品种为基础进行选育与系统改良而成。自大豆基因组正式在*Nature*杂志上发表以来(Schmutz et al., 2010), 在对大量品种资源进行基因组重测序的基础上, 运用比较基因组学等技术手段, 揭示了其在进化与驯化过程中基因组的变异与进化规律; 同时, 在大豆重要农艺性状的基因克隆及基因功能解析等方面也取得了一系列重要进展。迄今已有研究者对大豆不同研究领域或不同时期的进展进行了综述(Chan et al., 2012; Xia et al., 2013; Wang and Tian, 2015; 种康等, 2015)。本文旨在对2010年以来, 尤其是近期各国科学家在大豆相关研究领域所取得的重要成果与进展进行较为全面的跟踪。

1 大豆基因组重测序及基因组进化

大豆首个完成基因组测序的品种为美国现代栽培品种

Williams 82 (Schmutz et al., 2010), 其基因组大小约为1.1 G (Giga bp), 最新版本的基因注释文件中包括56 044个基因, 88 647个预测转录本(Gmax_275_Wm82.a2.v1, <http://phytozome.jgi.doe.gov/pz/portal.html>)。大豆与菜豆(*Phaseolus vulgaris*)共同经历了56.5–60.0 mya (百万年前)全基因组扩增, 大豆属特有的基因组扩增约发生于10–13 mya (Kim et al., 2015), 因此约75%的基因有2个或2个以上的拷贝(Schmutz et al., 2010)。大豆的近缘祖先为野生大豆。现代栽培大豆能与野生大豆经自然或人工杂交而形成有正常繁殖能力的后代, 因此目前普遍认为栽培大豆与野生大豆在分类学上同属于*Glycine max* (Stupar and Specht, 2013)。同在2010年, 野生大豆(系谱IT182932)基因组序列公布于世(Kim et al., 2010), 作者通过比较基因组学提出了野生大豆与栽培大豆的分化始于0.27 mya, 早于人类有记载的驯化历史, 但是该论点有待进一步的证据支持。借鉴来自微生物的泛基因组概念, 中国农科院作物所邱丽娟研究组选择了7份代表性野生大豆进行测序与独立denovo组装, 构建了大豆泛基因组。结果表明, 野生大豆中约48.6%为核心基因; 另51.4%非核心基因主要与野生大豆抗/耐生物和非生物逆境相关, 决定着

收稿日期: 2016-04-19; 接受日期: 2016-07-08

基金项目: 国家重点研发计划(No.2016YFD0100201, No.2016YFD0101902)、中国科学院战略性先导科技专项(No.XDA08010105-03)和国家自然科学基金(No.31471518)

* 通讯作者。E-mail: xiazhj@iga.ac.cn

野生大豆的广适性(Li et al., 2014)。香港中文大学林汉明研究组及合作者于2010年通过对31个野生大豆和栽培大豆全基因组进行低倍重测序,共发现了205 614个单核苷酸多态性(SNP)标签,建立了高密度的分子标记图谱;同时指出较高程度的基因连锁不平衡和较高比例的单核苷酸非同义替换(Ka)/同义替换(Ks)是大豆基因组的重要特征。通过对SNP及插入/缺失(insertion/deletion, indel)突变的多态性分析,发现只有2.99%的位点受到了驯化过程的选择(Lam et al., 2010)。另有研究表明,约4.38%的基因受到了驯化或遗传改良的影响(Li et al., 2013b)。中科院昆明动物所王文研究组与中科院遗传与发育所田志喜研究组,在对302份野生大豆、农家品种及栽培品种等进行重测序(11X)的基础上,在全基因组中检测到230个选择性区域(selective sweeps);通过全基因组关联分析(genome-wide association study, GWAS)发现了10个区段与9个已知的驯化或改良性状相关联,同时报道了13个调控重要农艺性状的新位点(Zhou et al., 2015b)。近期,美国密苏里州立大学Nguyen研究组与我国华大基因合作,成功地对包括野生大豆、农家品种与栽培品种等在内的106份大豆材料进行了基因组重测序,发现了近1 000万个SNP,其中35.34%为首次报道;同时他们还鉴定了159个可能与驯化相关的区域(domestication sweeps),总长约为54.34 Mb,包括4 414个基因(Valliyodan et al., 2016)。由于所选择的资源材料、测序深度(质量)与分析技术(参数)等不尽相同,不同研究所得出的与驯化或品种改良相关的研究结论间存在细微差别。但综合来看,不同研究结果均在不同程度上向人们展示了从野生大豆至栽培大豆的驯化过程中基因组层面的变化与进化规律。

2 大豆基因组组成变异及表达调控规律

长末端重复序列(long terminal repeat, LTR)转座子在开花植物的基因组中广为存在,可分为自主转移与非自主转移。在大豆基因组中,SNARE转座子家族有非自主转移SNRE亚家族及LTR反转座子(SARE(A)与SARE(B))亚家族。SARE(A)与SARE(B)自上一次13 mya全基因扩增后开始独立进化。意外的是,均来源于共同祖先的多个SNRE因子自3 mya前后分别与

SARE(A)或SARE(B)的LTR区域发生了同质化,形成与SARE(A)或SARE(B)相关联的亚家族(Du et al., 2010)。通过对31个完成测序的品种进行序列分析表明,明显不同的偏向性插入可能是在进化选择效率上产生差异的根源,外源LTR-RT与原来存在于基因组中的LTR-RT的分布特征一致,而非同源的DNA反转座子的插入则与插入点附近的重组率呈负相关(Tian et al., 2012)。非参考(nonreference) LTR-RT类转座子的插入亦与插入区域的重组率呈负相关,而DNA转座子未显示出这种相关关系(Tian et al., 2012)。在低水平表达基因的启动子区域,表现出高水平的CG与CHG甲基化程度(H=C、T或者A)和低水平的CHH甲基化频率。对不同转座子(TEs)进行比较研究表明,CHH甲基化水平在II型TEs中高于I型TEs,可能与小RNA多存在于II型TEs中有关(Song et al., 2013)。在发育的子叶中,小RNA丰度与高甲基化区域大致相关联,但与甲基化程度较低的区域呈负相关。胞嘧啶的甲基化对于植物发育过程中的基因表达及转座子的跳跃产生了深远的影响。通过对不同组织中2 162个甲基化程度不同的区域分析表明,甲基化程度较低的区域与邻近基因的高表达水平相关联(Song et al., 2013)。CG甲基化广泛存在于基因组扩增引致的基因中。相对于单拷贝的基因而言,多拷贝数量基因的CG甲基化程度高,且进化速率较慢(Kim et al., 2015)。然而在大豆与菜豆中均为单拷贝的基因甲基化程度亦高,这可能与这类基因的表达量高及表达谱较广相关。很多非CG甲基化与转座子的插入或相邻相关,说明转座子及非CG甲基化在基因组扩增后对基因的表达起重要的调控作用(Kim et al., 2015)。

可变剪接位点(alternative splicing, AS)的变异在高等真核生物中对基因表达起重要的调节作用。通过对28个不同组织样品的转录组分析,发现在大豆基因组中约有63%的多内含子基因可形成可变剪接,在幼嫩的组织中AS发生频率高。AS可分为外显子跳读(exon skipping)、内含子残留(intron retention)、替代供体位点(alternative donor sites)与替代受体位点(alternative acceptor sites) 4种类型。这种可变剪接的数量和剪接方式受到发育时期、基因所在染色体区域的特性、基因自身结构及基因表达量等因素的影响。在基因结构中,内含子长度的减小及外显子数目的减少亦可使大豆重复基因对中可变剪接的频率下降

(Shen et al., 2014)。华南农业大学年海研究组通过对11个大豆组织进行测序,注释了54 132个高可信度基因,包括6 718个新的转录区域,其平均长度为372 bp。AS事件中的15.9%发生在2个或2个以上外显子的基因中,有1 834个基因在不同发育阶段产生AS事件,其中202个为组织偏爱性的外显子跳读(tissue-biased exon-skipping events) (Wang et al., 2014)。

中科院遗传与发育所田志喜研究组分析了产生miRNA的基因(MIRNA genes, MIRs)与miRNA所调控的靶基因间的平行进化关系,发现MIRs较其靶基因进化快,但两者呈现相似的进化模式;MIRs及相应的miRNA靶基因的变异程度低往往与两类基因的高表达水平、较高的基因扩增倍数及较多的互作基因相关联(Liu et al., 2016b)。对野生大豆进行miRNA分析,鉴定了97个已知miRNAs与31个新miRNAs;预测的miRNA的靶基因包括编码ARF、NB-ARC、LRR-TIR、离子转运ATP酶、Myb转录因子及NAM蛋白等参与众多生物学过程的重要基因(Zeng et al., 2012)。

在中心粒附近的染色质区域,基因组片段的扩增状况与异义替换(Ka)及基因表达水平相关。总体上看,在大豆的异染色质区域具有较低同义替换(Ks)与较低的基因表达水平。一般认为,有害的突变会在重组率较低的区域积累。但在大豆中,因在基因组中所处的物理位置不同,基因组扩增所产生的2个同源基因常会发生不对称进化,即处于近染色体中心粒的异染色质区域同源基因与位于常染色质的同源基因相比,Ka较低而表达水平较高,在异染色质区域富集了相对较多的单拷贝基因(Du et al., 2012)。基因所处的不同染色质类型亦是决定基因功能分化方向及基因残留与否的重要因素。

3 大豆重要农艺性状的基因克隆

3.1 生育期基因

早在20世纪20年代人们就观察到一年中分期播种的大豆品种近乎同时开花,于是科学家以大豆及烟草(*Nicotiana tabacum*)等为模式植物,发现了植物光周期现象(Garner and Allard, 1920)。传统遗传学研究发现并定位了E1-E9与J等多个与大豆光周期反应及生育期相关的QTL位点,但其功能基因的克隆自2008年大豆基因组信息在网上(www.phytozome.net)公布之

后才逐步取得了突破性进展(Xia et al., 2012b; 夏正俊, 2013)。截至目前,研究人员已成功克隆了5个大豆生育期基因(E1-E4及E9)。Liu等(2008)首先成功克隆了E4,次年E3基因也得以克隆(Watanabe et al., 2009),这2个基因均转录大豆光敏色素蛋白(phytochrome A)。Watanabe等(2011)成功克隆了E2基因,其功能基因GmGla为拟南芥(*Arabidopsis thaliana*)开花基因GIGANTEA的同源基因。近期克隆的E9基因则为已知的大豆成花素基因GmFT2a (Kong et al., 2014; Zhao et al., 2016)。

控制大豆生育期基因中最大的E1位于6号染色体的中心粒附近,给QTL定位及克隆带来了极大的难度。中科院东北地理与农业生态所夏正俊研究组及合作者利用13 000多株的大容量遗传群体,成功克隆了大豆生育期基因E1,取得了重要进展(Xia et al., 2012a)。E1为豆科植物特有的基因序列,含有双元核定位信号及远缘的B3结构域。E1基因的表达在长日照条件下呈双峰型,在短日照下表达极低或不表达,这是大豆成为短日照作物的主要原因(Xia et al., 2012a)。长日照条件下,黎明后2-4小时E1的表达出现第1次高峰(Zhai et al., 2015; Xu et al., 2015)。通过对在远缘的B3结构域产生了突变的e1-b3a基因型进行研究,证明了该结构域与E1抑制开花的功能紧密相关(Zhai et al., 2015)。此外,研究者还鉴定了GmFT4为E1基因的下流基因,其同样具有抑制大豆开花的功能(Zhai et al., 2014a)。过表达E1基因抑制了大豆成花素基因GmFT2a/5a的表达(Xia et al., 2012a),却促进大豆GmFT4的表达,GmFT4与GmFT2a/5a之间的平衡可能决定着大豆开花时间(Zhai et al., 2014a)。E1的2个同源基因E1La (E1-like-a)与E1Lb的表达亦呈现双峰型,RNAi功能验证推断E1La/b具有与E1相类似的功能(Xu et al., 2015)。综合来看,大豆光周期反应及开花调控的基因网络与模式植物拟南芥或水稻(*Oryza sativa*)间存在着明显的不同,大豆生育期基因E1赋予了大豆独特的光周期反应特性。近期,夏正俊研究组将菜豆中的E1同源基因PvE1L (Phvul.009-G204600)转入大豆,抑制了大豆开花,表现出的功能与E1相似;而将蒺藜苜蓿(*Medicago truncatula*)的E1同源基因Medtr2g058520转入大豆后则对开花无明显影响;蒺藜苜蓿中Medtr2g058520基因的2个Tnt插入突变体与野生型相比呈晚花表型,推测该基因在蒺藜

苜蓿中具有促进植物开花的功能(Zhang et al., 2016)。豆科特有的*E1*基因家族出现了明显的功能分化,且分化规律与基因组扩增及物种间的进化相一致。

*E1*为大豆中调控开花及成熟的主要功能基因,是我国及世界大豆育种中主要选择的位点之一,其多种基因型在生产上发挥重要作用。*E1*与*e1-as*为最基本的2个基因型。在*e1-as*基因型中,*E1*的单碱基突变位点(精氨酸到苏氨酸的转变)恰好位于核定位信号区域,致使核及细胞的其它器官的定位发生变化,但该等位变异保留着完整的B3结构域,具有一定的抑制大豆开花与成熟的功能(Xia et al., 2012a)。*e1-as*基因型广泛存在于中国、日本及美洲较高纬度地区的栽培品种中,是生产上极为重要的1个基因型(Zhai et al., 2014b)。另外,*E1*基因的等位变异类型*e1-nf*、*e1-b3a*与*e1-nl*均为功能丧失型,其中*e1-nf*与*e1-b3a*分别为单碱基及复合型缺失,因移码提前产生了终止子(Xia et al., 2012a; Zhai et al., 2015);*e1-nl*为缺失型,即整个*E1*基因及调控区域均缺失,这类变异均为极早熟品种类型,多存在于我国东北地区、黄淮海流域夏大豆及日本北海道地区所培育出的栽培品种中(Zhai et al., 2014b)。*E2*基因的分布相对简单,即北方品种大多为隐性基因型,而南方品种大多为显性基因型(Zhai et al., 2014b)。近期,中科院植物所贺超英研究组对*E2*基因(*GmGla*)的研究发现,在野生大豆中有44个不同的单倍体型变异,栽培大豆中有3个(H1、H2和H3)单倍型,其中H1及H2型与正常的*E2*及*e2*基因型相对应,而其中H3类型仅在我国高纬度的东北地区存在(Wang et al., 2016b)。在我国的大豆栽培品种中,*E3*基因的变异类型较为丰富。而我国主要栽培品种的*E4*位点大多为*E4*基因型(Zhai et al., 2014b)。

通过对309个来源于美国与加拿大早熟地区的大豆品种进行GWAS分析,Zhang等(2015)鉴定出27、6、18与27个位点分别与开花期、成熟期、开花至成熟期及株高性状相关联。通过对286个自然品系进行RAD-seq技术研究,Zhou等(2015a)确定了48个与驯化相关的位点,其中12号染色体的4个基因及11号与15号染色体的各1个基因与开花相关。此外,多国科学家还正在致力于克隆调控*E5-E8*及*J*等位点的功能基因。相信在不远的将来,越来越多的QTL位点特别是微效QTL位点将会逐步得以克隆,为进一步全面

解析大豆生育期基因的精细调控网络奠定基础。

3.2 生长(结荚)习性

野生大豆一般具有无限生长(结荚)习性,即边开花边结荚。相反,植株开花后顶端生长点迅速转化为花序,植株停止营养生长,称之为有限生长习性。早在2010年,控制大豆生长的主要基因位点*Dt1*已被成功破译,其功能基因为拟南芥的同源基因*GmTFL1*(Liu et al., 2010; Tian et al., 2010)。同时遗传学研究表明,大豆生长习性还受*Dt2*基因控制,*Dt2/Dt1*基因型组合可形成介于有限与无限生长习性的中间类型——亚有限生长习性。控制生长习性的*Dt2*基因亦得以克隆(Ping et al., 2014)。大豆中的*Dt2*(*Glyma18g50910*)基因为拟南芥中*APETALA1/FRUITFULL*的同源基因,含有MADS-box结构域,可直接与*Dt1*基因的调节区域相结合(Liu et al., 2016c)。同时,在顶端分生组织中*Dt2*还能激发*GmSOC1*的表达,*GmSOC1*蛋白不但能与*Dt2*蛋白结合,还可与*Dt1*的调节区域相结合。*Dt2*与*Dt1*同时在拟南芥*ttl1*背景下异位表达能够产生亚有限生长株型,但是在*ttl1*与*soc1*双突变背景下则为无限生长习性,说明*Dt2*与*SOC1*对于调控*Dt1*的表达均起重要作用(Liu et al., 2016c)。表型遗传分析表明,大豆生育期基因*E1*与开花结荚习性之间存在互作(Xu et al., 2013a)。

3.3 种子形态及组分

南京农业大学喻德跃研究组在对367份大豆资源进行重测序的基础上,设计了高通量基因芯片(JAU 355K SoySNP),并利用该芯片所获得的信息进行连锁不平衡及GWAS分析,在20号染色体得到1个调控大豆百粒重的候选区域(Wang et al., 2016a)。此外,研究者还通过GWAS分析获得了4个位于11与16号染色体的基因,它们可能参与调控大豆种子大小,其中*Glyma11g18720*、*Glyma11g15480*及*Glyma15g35080*与拟南芥相关基因同源。*Glyma11g18720*及*Glyma05g28130*与5个拟南芥调控种子大小的同源基因共表达。*Glyma11g15480*与24个拟南芥种子发育的同源基因共表达(Zhou et al., 2015a)。

豆科作物从野生到栽培种,种子吸水性能是驯化的重要性状。大豆的种子吸水性能由1个单基因(*Gm-Hs1-1*)控制。图位克隆及表达分析表明,该基因编码与钙调磷酸酶相似(calcineurin-like)的金属磷酸酯跨

膜蛋白质(metallophosphoesterase transmembrane protein)。 *GmHs1-1*主要在种皮中的马尔皮基层中表达,同时还与种子中的钙含量相关(Sun et al., 2015)。引起种皮从不透水向透水转变的原因是*GmHs1-1*中人工选择导致的点突变。有意义的是,一些种质逃脱了这种选择,取而代之的是利用种子表皮开裂这一特性。尽管该基因为单位点的点突变,但从等位变异在大豆的群体结构层面来看, *GmHs1-1*很有可能经历了对种子透水性的再选择(Sun et al., 2015)。然而日本学者将该性状的功能基因定位于与*GmHs1-1*相邻的基因,两种研究结论的差异有待进一步验证(Jang et al., 2015)。

通过对302份大豆品种或种质资源进行测序,明确了驯化及改良过程中的选择瓶颈效应。对野生大豆及农家品种的驯化及品种改良过程(农家种至现代栽培品种)中种子大小、种皮颜色、生长习性及其含油量等性状进行分析,发现分别有121及109个强选择信号与驯化和品种改良相关,其中与油脂代谢相关的信号位点明显较多,说明油脂性状受复杂的多基因调控(Zhou et al., 2015b)。

三萜皂苷(triterpene saponins)在许多植物中具有多种生物学功能,其中皂苷一般被烷基化,赋予其结构与功能上的多样性。大豆皂醇A (soyasapogenol A)为大豆皂甙中的关键皂苷配基,其C端22位的末端糖基种类受其遗传学位点*Sg-1*控制,与大豆苦涩味性状相关联。遗传学及克隆结果表明,候选基因*Glyma07g38460*编码二磷酸尿核苷及糖依赖的糖基转移酶(UDP-sugar-dependent glycosyltransferase),该基因的功能通过转基因及功能互补实验得以验证(Sayama et al., 2012)。

3.4 结荚相关性状

改良野生大豆裂荚特性是大豆长期人工驯化的重要目标,相关性状的基因克隆对大豆生产具有重要意义。日本学者通过图位克隆明确了与裂荚相关QTL位点的功能基因为转录dirigent类的家族基因,该基因与木质素合成相关(Funatsuki et al., 2014)。

叶形一般分为宽叶与窄叶,由1个*JAG (JAGGED)*基因控制。拟南芥*JAG*调控器官侧面组织的形成,该基因突变常引发果实形态的各种异常。大豆中控制宽

叶与窄叶的基因为*Gm-jag1 (In)*,该基因能部分恢复拟南芥锯齿状(jagged)突变体的表型。在大豆中*In*还一因多效地调控大豆每荚粒数(4粒荚数量)(Jeong et al., 2011; Fang et al., 2013)。

3.5 生物与非生物胁迫

大豆根结线虫(*Heterodera glycines Ichinohe*)是大豆生产中重要的限制因子。抗大豆线虫的2个重要基因*Rhg4*与*Rhg1*已被成功克隆。*Rhg4*是品种抗该线虫的重要QTL位点。图位克隆结果显示,其功能基因编码丝氨酸羟甲基转移酶(serine hydroxymethyltransferase, SHMT),该酶在生物界广泛分布且具有保守性。已证实该酶在细胞的单碳物质代谢中发挥着促进丝氨酸与甘氨酸间交叉对话(cross-talk)的作用(Liu et al., 2012)。另一方面,大多高度抗病品种均含有*Rhg1 (rhg1-b)*基因。经图位克隆等方法证明在31 kb (拷贝串生)中含有*rhg1-b*,它转录3种类型的功能蛋白,包括1个氨基酸转运蛋白、1个alpha-SNAP蛋白和1个WI12 (wound-inducible domain)蛋白,3者均与抗病性相关。在*rhg1-b*基因型中,10个串生的拷贝存在于31 kb中,而感病品种则只有1个拷贝。过表达实验表明,单个基因的过量表达不足以产生抗病性,而3个基因共同表达赋予了植株的抗病特性,说明增加串生的非同源多基因的拷贝数可以提高抗病性(Cook et al., 2012)。

Xu等(2013b)利用高通量测序定位了南方根结线虫病(southern root-knot nematode)的抗病基因。通过对264个重组自交系(RIL)群体的低倍测序,获得了109 273个SNPs与3 489个重组区间,鉴定出3个QTLs,其中最大的为10号染色体的29.7 kb,内含3个真基因与2个假基因。根据表达分析,确定了2个候选基因*Glyma10g02150*与*Glyma10g02160*,均转录几丁质甲基转移酶抑制因子——几丁质甲基转移酶复合基因。

土壤盐渍化是世界农业生产面临的难题,故大豆抗盐性研究显得尤为重要。栽培大豆的耐盐性由一对显性基因控制,位于3号染色体。我国科学家利用GWAS及图位克隆技术,先后独立完成了该基因的克隆(Qi et al., 2014; Guan et al., 2014)。其功能基因*GmSALT3*编码1个定位于内质网的离子转运蛋白。

4 基因功能鉴定及基因调控网络

4.1 开花相关基因

CO (*CONSTANS*)-*FT* (*FLOWERING LOCUS T*) 调节模式在豆科的部分植物(如豌豆(*Pisum sativum*))中存在一定的保守性(Liew et al., 2014)。大豆基因组中有28个*CO*同源基因与24个*FT*同源基因(Fan et al., 2014)。中国农科院作物所傅永福研究组证明, *GmCOL5/13*与*GmFT1/2/4/5/6*的表达呈日生物钟节律, 只有*GmFTL6*在根中高表达, *GmCOL2*能恢复拟南芥*co*突变体表型, 所有的*GmFTL1/2/3/4/5/6*在拟南芥中均可促进植株开花。*GmCOL5/13*与*GmFT1/2/3/4/5/6*可能形成复杂的*CO-FT*调节元来调控大豆开花(Fan et al., 2014)。美国研究者证实, *GmCOL1a/1b*与*GmCOL2a/2b*尽管功能上具有一定的保守性, 根据拟南芥中的功能推测*GmCOL1a/1b*可能促进大豆开花(Wu et al., 2014), 但通过转基因大豆证明*GmCOL1a/1b*为大豆中开花抑制因子(Cao et al., 2015a)。基因拷贝数量变化及功能分化是植物产生新基因的重要源泉, 可推动植物基因组的进化。通过对大豆与拟南芥中磷酸酯乙醇胺结合蛋白(PEBP)家族成员进行系统性分析, 发现来自大豆的家族成员与大豆基因组的二次基因组扩增相关联; 在祖先的基因组扩增后, *FT*与*TFL1* (*TERMINAL FLOWER1*)亚族与其共同的祖先*MFT* (*MOTHER OF FT AND TFL1*)亚家族在功能上出现分化, 分别进化出具有促进开花与抑制开花的功能。通过对转基因拟南芥进行基因功能分析, 发现大豆和拟南芥中PEBP家族主要成员在进化过程中均能保留所属亚家族的功能, 只有少数基因在功能上出现了显著分化; 同时明确了有关功能保守或导致功能分化的关键氨基酸位点; 此外, 基因表达模式的变化亦是功能分化的重要表现形式(Wang et al., 2015b)。

转基因实验证明miRNA172可通过调控*GmTOE-4a*来影响植株开花(Zhao et al., 2015)。而过表达miRNA156b则可抑制大豆植株的开花(Cao et al., 2015b)。

4.2 共生结瘤与磷的高效利用

大豆根瘤共生固氮是我国历来重视的重要研究领域, 同时亦与农业生产、生态平衡及环境保护等相关。根瘤共生体系的建立是重要的生物学固氮过程, 在豆科

植物最大的蝶形花亚科的90个属中均有发生, 大多蝶形花种的基因组扩增事件发生在58 mya。约1/4的蝶形花由来的扩增基因被保留, 但在表达水平上差别很大。很多保留的基因与大豆结瘤性状相关, 可以说基因组倍增事件提升了大豆与共生根瘤间的共生关系(Li et al., 2013a)。

miR172c (microRNA172c)在调控大豆根瘤固氮中起重要作用(Wang et al., 2014)。无根瘤菌情况下, 结瘤调控因子*ENOD40*基因的启动子可与大豆转录抑制子*GmNNC1* (*Nodule Number Control 1*)直接结合, 表现为表达量下降; 当根瘤菌侵染大豆根系时, 可诱导miR172c的表达, 进而该小RNA通过剪切*GmNNC1* mRNA, 使*GmNNC1*的蛋白量减少, 相当于去除了转录抑制子*GmNNC1*对*ENOD40*的抑制; *ENOD40*进一步激活开启了结瘤因子诱导的信号转导途径, 从而促进根瘤的产生与发育(Wang et al., 2014)。该研究还进一步阐述了大豆存在着超结瘤的自主调控途径(autoregulation, AON)。例如, 可通过芽伴生抑制因子(shoot-derived inhibitor) (如细胞分裂素)来抑制miR172c的表达, 从而使植株不会发生过度结瘤(Wang et al., 2014)。

在大豆耐低磷的QTL定位研究基础上, Zhang等(2014)新定位了1个调控磷吸收效率的主效QTL, 以自然群体及全基因组关联分析来缩小目标QTL区间, 并明确了酸性磷酸酶基因(*GmACP1*)为影响该性状的功能基因。他们还通过诱导表达分析、原核表达及遗传转化等技术对所克隆的基因进行了功能验证。

4.3 种子及果实发育

中科院遗传与发育所陈受宜与张劲松研究团队对40个野生大豆及栽培大豆不同发育阶段的幼嫩种子组织进行了转录组测定, 明确了2 680个差异表达基因与种子成熟相关, 并建立了2个品种特有的基因共表达网络, 通过整合QTL数据, 鉴定了调控种子发育的2个关键基因*GA20OX*与*NFYA*。超表达*GA20OX*植株具有较大的种子(重量), 过量表达*NFYA*植株的种子含油量较高(Lu et al., 2016)。

中科院植物所王印政研究组发现, 果荚腹缝线维管束细胞壁剧烈增厚的纤维帽细胞(fiber cap cells, FCC)是栽培大豆果实裂荚抗性的关键形态结构, 野生大豆和栽培大豆的离层细胞并没有明显差异。一系

列的研究表明, *SHAT1-5*基因(NAC基因家族)上游4 kb处的1个抑制子在栽培大豆品种中被彻底清除, 可能为驯化过程中人工选择的结果, *SHAT1-5*的表达在栽培大豆纤维帽细胞中上调了15倍以上, 致使其次生壁明显加厚, 可有效阻止果荚的自然开裂(Dong et al., 2014)。

4.4 生物与非生物胁迫

利用转化大豆毛状根的实验体系在分子水平上证明 *WRKY27*能够受多种非生物胁迫的诱导并能显著提高大豆对干旱和盐胁迫的适应性。*WRKY27*蛋白对 *NAC29*的负调控作用通过结合 *NAC29*启动子区的 W-box进行(Wang et al., 2015a)。

近来陈受宜研究组与张劲松研究组明确了 *miR-172a*与大豆抗盐性间的关系, *miR172a*通过切割降解靶基因 *SSAC1*, 解除其蛋白对硫胺素(维生素B)前体合成酶基因 *THI1*启动子的抑制作用, 从而促进了 *THI1*的表达而提高了硫胺素合成, 增强了大豆耐盐性。他们同时发现 *miR172a*可作为长距离信号分子从大豆根转运到地上部并调控靶基因及下游基因的表达(Pan et al., 2016)。

南京农业大学李艳研究组通过对117个MATE (Multidrug and toxic compound extrusion transporters)基因的 *cis*因子及表达等进行系统分析, 提出 *GmMATE-75*为大豆耐铝性相关的候选基因(Liu et al., 2016a)。

5 研究展望

自大豆基因组信息发表以来, 大豆中许多重要农艺性状的基因克隆研究取得了重要突破。虽然近期许多重要突破是以中国科学家或华人团队为主实现的, 但是我国在某些领域(如大豆代谢组学)的研究起步较晚。大豆研究中所取得的研究成果可为其它豆科植物中的基因组学与基因克隆提供重要的参考与借鉴。国外同行已在豆科其它植物如菜豆、蒺藜苜蓿、百脉根(*Lotus japonicus*)及鹰嘴豆(*Cicer arietinum*)等基因组研究中取得了突破性进展(Pandey et al., 2016), 而我国在其它豆类中的研究还不够全面与系统。

我国大豆的栽培历史悠久, 豆类栽培、加工与食用的知识积累深厚。但现今我国大豆生产受到其它粮食作物如玉米(*Zea mays*)与水稻, 以及大量从巴西与

阿根廷等世界新兴大豆生产国进口转基因大豆的双重挤压, 种植面积逐年下降。转基因大豆在国外已成为重要产业, 而因生态及安全性等方面的争议, 致使我国只能进口不能自主生产转基因大豆。鉴于我国大豆在产量上没有取得实质性的突破, 大豆高产及品质性状等系统研究已被列入中科院“分子模块设计育种创新体系”战略先导专项(薛勇彪等, 2015)中。我们相信, 在传统遗传学的基础上, 综合运用基因组学及系统生物学等最新的技术手段, 可解析大豆高产、稳产、优质和高效等重要农艺(经济)性状的主要QTL基因或分子模块, 并明确基因网络及分子模块间的耦合规律; 有望经过长期的努力, 通过“分子模块导航育种”持续培育出高产优质的新品种。

参考文献

- 种康, 王台, 钱前, 王小菁, 左建儒, 顾红雅, 姜里文, 陈之端, 白永飞, 杨淑华, 孔宏智, 陈凡, 萧浪涛 (2015). 2014年中国植物科学若干领域重要研究进展. 植物学报 **50**, 412–459.
- 夏正俊 (2013). 大豆光周期反应与生育期基因研究进展. 作物学报 **39**, 1–9.
- 薛勇彪, 种康, 韩斌, 桂建芳, 王台, 傅向东, 何祖华, 储成才, 田志喜, 程祝宽, 林少扬 (2015). 开启中国设计育种新篇章——“分子模块设计育种创新体系”战略性先导科技专项进展. 中科院院刊 **30**, 393–402.
- Cao D, Li Y, Lu S, Wang J, Nan H, Li X, Shi D, Fang C, Zhai H, Yuan X, Anai T, Xia Z, Liu B, Kong F (2015a). *GmCOL1a* and *GmCOL1b* function as flowering repressors in soybean under long-day conditions. *Plant Cell Physiol* **56**, 2409–2422.
- Cao D, Li Y, Wang J, Nan H, Wang Y, Lu S, Jiang Q, Li X, Shi D, Fang C, Yuan X, Zhao X, Li X, Liu B, Kong F (2015b). *GmMiR156b* overexpression delays flowering time in soybean. *Plant Mol Biol* **89**, 353–363.
- Chan C, Qi X, Li MW, Wong FL, Lam HM (2012). Recent developments of genomic research in soybean. *J Genet Genomics* **39**, 317–324.
- Cook DE, Lee TG, Guo XL, Melito S, Wang K, Bayless AM, Wang JP, Hughes TJ, Willis DK, Clemente TE, Diers BW, Jiang JM, Hudson ME, Bent AF (2012). Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science* **338**, 1206–1209.
- Dong Y, Yang X, Liu J, Wang BH, Liu BL, Wang YZ (2014).

- Pod shattering resistance associated with domestication is mediated by a *NAC* gene in soybean. *Nat Commun* **5**, 3352.
- Du J, Tian Z, Sui Y, Zhao M, Song Q, Cannon SB, Cregan P, Ma J** (2012). Pericentromeric effects shape the patterns of divergence, retention, and expression of duplicated genes in the paleopolyploid soybean. *Plant Cell* **24**, 21–32.
- Du JC, Tian ZX, Bowen NJ, Schmutz J, Shoemaker RC, Ma JX** (2010). Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR swapping in soybean. *Plant Cell* **22**, 48–61.
- Fan CM, Hu RB, Zhang XM, Wang X, Zhang WJ, Zhang QZ, Ma JH, Fu YF** (2014). Conserved CO-FT regulons contribute to the photoperiod flowering control in soybean. *BMC Plant Biol* **14**, 9.
- Fang C, Li W, Li G, Wang Z, Zhou Z, Ma Y, Shen Y, Li C, Wu Y, Zhu B, Yang W, Tian Z** (2013). Cloning of *Ln* gene through combined approach of map-based cloning and association study in soybean. *J Genet Genomics* **40**, 93–96.
- Funatsuki H, Suzuki M, Hirose A, Inaba H, Yamada T, Hajika M, Komatsu K, Katayama T, Sayama T, Ishimoto M, Fujino K** (2014). Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proc Natl Acad Sci USA* **111**, 17797–17802.
- Garner WW, Allard HA** (1920). Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *J Agric Res* **18**, 553–606.
- Guan RX, Qu Y, Guo Y, Yu LL, Liu Y, Jiang JH, Chen JG, Ren YL, Liu GY, Tian L, Jin LG, Liu ZX, Hong HL, Chang RZ, Gilliam M, Qiu LJ** (2014). Salinity tolerance in soybean is modulated by natural variation in *GmSALT3*. *Plant J* **80**, 937–950.
- Jang SJ, Sato M, Sato K, Jitsuyama Y, Fujino K, Mori H, Takahashi R, Benitez ER, Liu B, Yamada T, Abe J** (2015). A single-nucleotide polymorphism in an endo-1, 4- β -glucanase gene controls seed coat permeability in soybean. *PLoS One* **10**, e0128527.
- Jeong N, Moon JK, Kim HS, Kim CG, Jeong SC** (2011). Fine genetic mapping of the genomic region controlling leaflet shape and number of seeds per pod in the soybean. *Theor Appl Genet* **122**, 865–874.
- Kim KD, El Baidouri M, Abernathy B, Iwata-Otsubo A, Chavarro C, Gonzales M, Libault M, Grimwood J, Jackson SA** (2015). A comparative epigenomic analysis of polyploidy-derived genes in soybean and common bean. *Plant Physiol* **168**, 1433–1447.
- Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, Kim DS, Lee YS, Park D, Ma J, Kim WY, Kim BC, Park S, Lee KA, Kim DH, Kim KH, Shin JH, Jang YE, Do Kim K, Liu WX, Chaisan T, Kang YJ, Lee YH, Kim KH, Moon JK, Schmutz J, Jackson SA, Bhak J, Lee SH** (2010). Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc Natl Acad Sci USA* **107**, 22032–22037.
- Kong FJ, Nan HY, Cao D, Li Y, Wu FF, Wang JL, Lu SJ, Yuan XH, Cober ER, Abe J, Liu BH** (2014). A new dominant gene *E9* conditions early flowering and maturity in soybean. *Crop Sci* **54**, 2529–2535.
- Lam HM, Xu X, Liu X, Chen WB, Yang GH, Wong FL, Li MW, He WM, Qin N, Wang B, Li J, Jian M, Wang JA, Shao GH, Wang J, Sun SSM, Zhang GY** (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet* **42**, 1053–1059.
- Li QG, Zhang L, Li C, Zhang YM** (2013a). Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the Papilionoideae. *Mol Biol Evol* **30**, 2602–2611.
- Li YH, Zhao SC, Ma JX, Li D, Yan L, Li J, Qi XT, Guo XS, Zhang L, He WM, Chang RZ, Liang QS, Guo Y, Ye C, Wang XB, Tao Y, Guan RX, Wang JY, Liu YL, Jin LG, Zhang XQ, Liu ZX, Zhang LJ, Chen J, Wang KJ, Nielsen R, Li RQ, Chen PY, Li WB, Reif JC, Purugganan M, Wang J, Zhang MC, Wang J, Qiu LJ** (2013b). Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics* **14**, 579.
- Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang SS, Zuo Q, Shi XH, Li YF, Zhang WK, Hu Y, Kong G, Hong HL, Tan B, Song J, Liu ZX, Wang Y, Ruan H, Yeung CK, Liu J, Wang H, Zhang LJ, Guan RX, Wang KJ, Li WB, Chen SY, Chang RZ, Jiang Z, Jackson SA, Li R, Qiu LJ** (2014). *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat Biotechnol* **32**, 1045–1052.
- Liew LC, Singh MB, Bhalla PL** (2014). Unique and conserved features of floral evocation in legumes. *J Integr Plant Biol* **56**, 714–728.
- Liu BH, Kanazawa A, Matsumura H, Takahashi R, Harada K, Abe J** (2008). Genetic redundancy in soybean photo-

- responses associated with duplication of the phytochrome A gene. *Genetics* **180**, 995–1007.
- Liu BH, Watanabe S, Uchiyama T, Kong FJ, Kanazawa A, Xia ZJ, Nagamatsu A, Arai M, Yamada T, Kitamura K, Masuta C, Harada K, Abe J (2010). The soybean stem growth habit gene *Dt1* is an ortholog of Arabidopsis *TERMINAL FLOWER1*. *Plant Physiol* **153**, 198–210.
- Liu J, Li Y, Wang W, Gai J, Li Y (2016a). Genome-wide analysis of MATE transporters and expression patterns of a subgroup of MATE genes in response to aluminum toxicity in soybean. *BMC Genomics* **17**, 223.
- Liu SM, Kandath PK, Warren SD, Yeckel G, Heinz R, Alden J, Yang CL, Jamai A, El-Mellouki T, Juveale PS, Hill J, Baum TJ, Cianzio S, Whitham SA, Korkin D, Mitchum MG, Meksem K (2012). A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* **492**, 256–260.
- Liu T, Fang C, Ma Y, Shen Y, Li C, Li Q, Wang M, Liu S, Zhang J, Zhou Z, Yang R, Wang Z, Tian Z (2016b). Global investigation of the co-evolution of MIRNA genes and microRNA targets during soybean domestication. *Plant J* **85**, 396–409.
- Liu Y, Zhang D, Ping J, Li S, Chen Z, Ma J (2016c). Innovation of a regulatory mechanism modulating semiterminate stem growth through artificial selection in soybean. *PLoS Genet* **12**, e1005818.
- Liu X, Li QT, Xiong Q, Li W, Bi YD, Lai YC, Liu XL, Man WQ, Zhang WK, Ma B, Chen SY, Zhang JS (2016). The transcriptomic signature of developing soybean seeds reveals genetic basis of seed trait adaptation during domestication. *Plant J* **86**, 530–544.
- Pan WJ, Tao JJ, Cheng T, Bian XH, Wei W, Zhang WK, Ma B, Chen SY, Zhang JS (2016). Soybean miR172a improves salt tolerance and can function as a long distance signal. *Mol Plant* **9**, 1337–1340.
- Pandey MK, Roorkiwal M, Singh VK, Ramalingam A, Kundapa H, Thudi M, Chitikineni A, Rathore A, Varshney RK (2016). Emerging genomic tools for legume breeding: current status and future prospects. *Front Plant Sci* **7**, 455.
- Ping J, Liu Y, Sun L, Zhao M, Li Y, She M, Sui Y, Lin F, Liu X, Tang Z, Nguyen H, Tian Z, Qiu L, Nelson RL, Clemente TE, Specht JE, Ma J (2014). *Dt2* is a gain-of-function MADS-domain factor gene that specifies semi-determinacy in soybean. *Plant Cell* **26**, 2831–2842.
- Qi XP, Li MW, Xie M, Liu X, Ni M, Shao GH, Song C, Yim AKY, Tao Y, Wong FL, Isobe S, Wong CF, Wong KS, Xu CY, Li CQ, Wang Y, Guan R, Sun FM, Fan GY, Xiao ZX, Zhou F, Phang TH, Liu X, Tong SW, Chan TF, Yiu SM, Tabata S, Wang J, Xu X, Lam HM (2014). Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. *Nat Commun* **5**, 4340.
- Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, Hirose A, Sasama H, Ohashi M, Hasegawa H, Terakawa T, Kikuchi A, Kato S, Tatsuzaki N, Tsukamoto C, Ishimoto M (2012). The *Sg-1* glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. *Plant Cell* **24**, 2123–2138.
- Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, Xu D, Hellsten U, May GD, Yu YS, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu SQ, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du JC, Tian ZX, Zhu LC, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010). Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183.
- Shen Y, Zhou Z, Wang Z, Li W, Fang C, Wu M, Ma Y, Liu T, Kong LA, Peng DL, Tian Z (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* **26**, 996–1008.
- Song QX, Lu X, Li QT, Chen H, Hu XY, Ma B, Zhang WK, Chen SY, Zhang JS (2013). Genome-wide analysis of DNA methylation in soybean. *Mol Plant* **6**, 1961–1974.
- Stupar RM, Specht JE (2013). Insights from the soybean (*Glycine max* and *Glycine soja*) genome: past, present, and future. *Adv Agron* **118**, 177–204.
- Sun LJ, Miao ZY, Cai CM, Zhang DJ, Zhao MX, Wu YY, Zhang XL, Swarm SA, Zhou LW, Zhang ZYJ, Nelson RL, Ma JX (2015). *GmHs1-1*, encoding a calcineurin-like protein, controls hard-seededness in soybean. *Nat Genet* **47**, 939–943.
- Tian ZX, Wang XB, Lee R, Li YH, Specht JE, Nelson RL, McClean PE, Qiu LJ, Ma JX (2010). Artificial selection for determinate growth habit in soybean. *Proc Natl Acad Sci USA* **107**, 8563–8568.
- Tian ZX, Zhao MX, She MY, Du JC, Cannon SB, Liu X, Xu X, Qi XP, Li MW, Lam HM, Ma JX (2012). Genome-wide characterization of nonreference transposons reveals evolutionary propensities of transposons in soybean. *Plant Cell* **24**, 4422–4436.
- Valliyodan B, Qiu D, Patil G, Zeng P, Huang J, Dai L, Chen C, Li Y, Joshi T, Song L, Vuong TD, Musket TA, Xu D,

- Shannon JG, Shifeng C, Liu X, Nguyen HT (2016). Landscape of genomic diversity and trait discovery in soybean. *Sci Rep* **6**, 23598.
- Wang F, Chen HW, Li QT, Wei W, Li W, Zhang WK, Ma B, Bi YD, Lai YC, Liu XL, Man WQ, Zhang JS, Chen SY (2015a). GmWRKY27 interacts with GmMYB174 to reduce expression of *GmNAC29* for stress tolerance in soybean plants. *Plant J* **83**, 224–236.
- Wang J, Chu S, Zhang H, Zhu Y, Cheng H, Yu D (2016a). Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Sci Rep* **6**, 20728.
- Wang Y, Gu Y, Gao H, Qiu L, Chang R, Chen S, He C (2016b). Molecular and geographic evolutionary support for the essential role of *GIGANTEA* in soybean domestication of flowering time. *BMC Evol Biol* **16**, 79.
- Wang Y, Wang L, Zou Y, Chen L, Cai Z, Zhang S, Zhao F, Tian Y, Jiang Q, Ferguson BJ, Gresshoff PM, Li X (2014). Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate *ENOD40* expression and regulate nodule initiation. *Plant Cell* **26**, 4782–4801.
- Wang Z, Tian Z (2015). Genomics progress will facilitate molecular breeding in soybean. *Sci China Life Sci* **58**, 813–815.
- Wang Z, Zhou Z, Liu Y, Liu T, Li Q, Ji Y, Li C, Fang C, Wang M, Wu M, Shen Y, Tang T, Ma J, Tian Z (2015b). Functional evolution of phosphatidylethanolamine binding proteins in soybean and Arabidopsis. *Plant Cell* **27**, 323–336.
- Watanabe S, Hideshima R, Xia ZJ, Tsubokura Y, Sato S, Nakamoto Y, Yamanaka N, Takahashi R, Ishimoto M, Anai T, Tabata S, Harada K (2009). Map-based cloning of the gene associated with the soybean maturity locus *E3*. *Genetics* **182**, 1251–1262.
- Watanabe S, Xia ZJ, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata S, Kitamura K, Harada K (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the *GI-GANTEA* gene is involved in soybean maturity and flowering. *Genetics* **188**, 395–407.
- Wu FQ, Price BW, Haider W, Seufferheld G, Nelson R, Hanzawa Y (2014). Functional and evolutionary characterization of the *CONSTANS* gene family in short-day photoperiodic flowering in soybean. *PLoS One* **9**, e85754.
- Xia ZJ, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Anai T, Sato S, Yamazaki T, Lu SX, Wu HY, Tabata S, Harada K (2012a). Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proc Natl Acad Sci USA* **109**, E2155–E2164.
- Xia ZJ, Zhai H, Liu BH, Kong FJ, Yuan XH, Wu HY, Cober E, Harada K (2012b). Molecular identification of genes controlling flowering time, maturity, and photoperiod response in soybean. *Plant Syst Evol* **298**, 1217–1227.
- Xia ZJ, Zhai H, Lu SX, Wu HY, Zhang YP (2013). Recent achievement in gene cloning and functional genomics in soybean. *Scientific World J* **2013**, 281367.
- Xu M, Yamagishi N, Zhao C, Takeshima R, Kasai M, Watanabe S, Kanazawa A, Yoshikawa N, Liu B, Yamada T, Abe J (2015). The soybean-specific maturity gene *E1* family of floral repressors controls night-break responses through down-regulation of *FLOWERING LOCUS T* orthologs. *Plant Physiol* **168**, 1735–1746.
- Xu ML, Xu ZH, Liu BH, Kong FJ, Tsubokura Y, Watanabe S, Xia ZJ, Harada K, Kanazawa A, Yamada T, Abe J (2013a). Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated postflowering responses of soybean. *BMC Plant Biol* **13**, 91.
- Xu X, Zeng L, Tao Y, Vuong T, Wan J, Boerma R, Noe J, Li Z, Finnerty S, Pathan SM, Shannon JG, Nguyen HT (2013b). Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. *Proc Natl Acad Sci USA* **110**, 13469–13474.
- Zeng QY, Yang CY, Ma QB, Li XP, Dong WW, Nian H (2012). Identification of wild soybean miRNAs and their target genes responsive to aluminum stress. *BMC Plant Biol* **12**, 182.
- Zhai H, Lü S, Wu H, Zhang Y, Zhang X, Yang J, Wang Y, Yang G, Qiu H, Cui T, Xia Z (2015). Diurnal expression pattern, allelic variation, and association analysis reveal functional features of the *E1* gene in control of photoperiodic flowering in soybean. *PLoS One* **10**, e0135909.
- Zhai H, Lu SX, Liang S, Wu HY, Zhang XZ, Liu BH, Kong FJ, Yuan XH, Li J, Xia ZJ (2014a). *GmFT4*, a homolog of *FLOWERING LOCUS T*, is positively regulated by *E1* and functions as a flowering repressor in soybean. *PLoS One* **9**, e89030.
- Zhai H, Lu SX, Wang YQ, Chen X, Ren HX, Yang JY, Cheng W, Zong CM, Gu HP, Qiu HM, Wu HY, Zhang XZ, Cui TT, Xia ZJ (2014b). Allelic variations at four major maturity *E* genes and transcriptional abundance of the *E1* gene are associated with flowering time and maturity of soybean cultivars. *PLoS One* **9**, e97636.

- Zhang D, Song H, Cheng H, Hao D, Wang H, Kan G, Jin H, Yu D (2014). The acid phosphatase-encoding gene *Gm-ACP1* contributes to soybean tolerance to low-phosphorus stress. *PLoS Genet* **10**, e1004061.
- Zhang JP, Song QJ, Cregan PB, Nelson RL, Wang XZ, Wu JX, Jiang GL (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* **16**, 217.
- Zhang XZ, Zhai H, Wang YY, Tian XY, Zhang YP, Wu HY, Lü SY, Yang G, Li YQ, Wang L, Hu B, Bo QY, Xia ZJ (2016). Functional conservation and diversification of the soybean maturity gene *E1* and its homologs in legumes. *Sci Rep* **13**, 29548.
- Zhao C, Takeshima R, Zhu J, Xu M, Sato M, Watanabe S, Kanazawa A, Liu B, Kong F, Yamada T, Abe J (2016). A recessive allele for delayed flowering at the soybean maturity locus *E9* is a leaky allele of *FT2a*, a *FLOWERING LOCUS T* ortholog. *BMC Plant Biol* **16**, 20.
- Zhao X, Cao D, Huang Z, Wang J, Lu S, Xu Y, Liu B, Kong F, Yuan X (2015). Dual functions of *GmTOE4a* in the regulation of photoperiod-mediated flowering and plant morphology in soybean. *Plant Mol Biol* **88**, 343–355.
- Zhou L, Wang SB, Jian JB, Geng QC, Wen J, Song QJ, Wu ZZ, Li GJ, Liu YQ, Dunwell JM, Zhang J, Feng JY, Niu Y, Zhang L, Ren WL, Zhang YM (2015a). Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Sci Rep* **5**, 9350.
- Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y, Fang C, Shen Y, Liu T, Li C, Li Q, Wu M, Wang M, Wu Y, Dong Y, Wan W, Wang X, Ding Z, Gao Y, Xiang H, Zhu B, Lee SH, Wang W, Tian Z (2015b). Re-sequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol* **33**, 408–414.

Research Progress in Whole-genome Analysis and Cloning of Genes Underlying Important Agronomic Traits in Soybean

Zhengjun Xia*

Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin 150081, China

Abstract Since the soybean genome was published in *Nature*, in 2010, many researchers from different countries, mainly China, have re-sequenced the genomes of wild soybean accessions, landraces and cultivated cultivars of soybean. The general features at the whole-genome level during evolution and domestication have been revealed by comparative genomic analysis. Moreover, several breakthroughs have been achieved by using sequence information of the soybean genome, as evidenced by successful cloning of a number of genes underlying important agronomic traits and understanding their general regulatory mechanisms. The recent progress will lay a solid foundation for studying the elaborate regulatory network of different molecular pathways and fulfill prerequisites for breeding new cultivars with a molecular (model) design.

Key words soybean, research progress, genome, agronomic traits, cloning, gene function

Xia ZJ (2017). Research progress in whole-genome analysis and cloning of genes underlying important agronomic traits in soybean. *Chin Bull Bot* **52**, 148–158.

* Author for correspondence. E-mail: xiazhj@iga.ac.cn

(责任编辑: 朱亚娜)