



上海伯豪生物技术有限公司
SHANGHAI BIOTECHNOLOGY CORPORATION

作物遗传育种



2019版

目录

■ 种质资源的评价	01
■ 作物复杂性状的遗传解析	04
全基因组关联分析	04
连锁分析	06
BSA定位	07
基因差异表达分析鉴定候选基因	08
■ 全基因组选择育种	10
■ 分子设计育种	12
■ 作物群体遗传进化	14
■ 品种指纹图谱及纯度鉴定	16
■ 伯豪生物技术平台	18
高通量测序平台	18
芯片基因分型平台	19
其它基因分型平台	20
生物信息分析平台	20
样品管理平台	20
■ 伯豪客户农林研究高分文献	21
■ 附：样品准备与运送	22

公司介绍

服务科技创新，护航人类健康！

上海伯豪生物技术有限公司(以下简称“伯豪生物”)2008年12月成立，是一家以科技服务、疾病与健康检测、分子检测产品的开发和生产为主营业务的高新技术企业。公司形成了面向科研和临床的系统技术服务平合，提供全基因组测序、生物信息分析、标志物筛选和分子检测验证、基因功能验证科研一站式服务，同时提供试剂盒开发、生产以及检测应用的转化医学的临床一站式服务。伯豪生物携手各参控股子公司重庆伯豪医学检验所(临床医学检测专业子公司)、伯豪医药(医疗器械经营专业子公司)和迅伯生物(产品研发和GMP生产专业子公司)共同开启“专业服务成就科学发现，产业布局构筑精准医学”的新篇章。



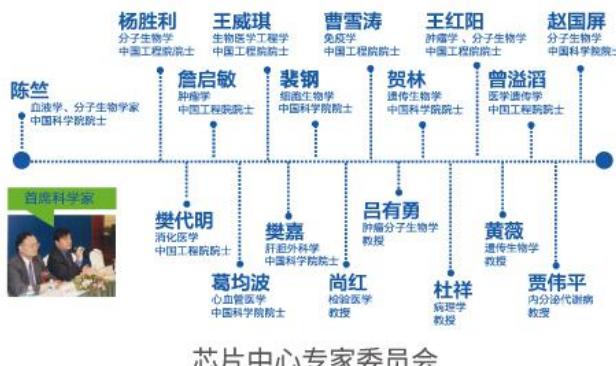
高新技术企业



国家基因检测技术应用示范中心



浦东新区企业研发机构



芯片中心专家委员会

国际一流质量标准

伯豪生物建立了规范化的质量控制体系，通过了ISO9001:2015质量管理体系的认证，并参照GLP-L的标准，通过156个SOP文件对项目、样品、实验过程、生物信息分析过程、数据信息管理、保密等进行严格质控。



多样化的服务平台

伯豪生物多样灵活的服务平台，系统的生物学研究解决方案，全力加速您组学研究的进程！



丰富的项目经验

伯豪生物从事技术服务15年，承接项目数超过15000个，用户单位5000家以上，协助客户发表SCI论文1500余篇（影响因子共计7000多分），全面推动中国基因组学服务产业的发展。



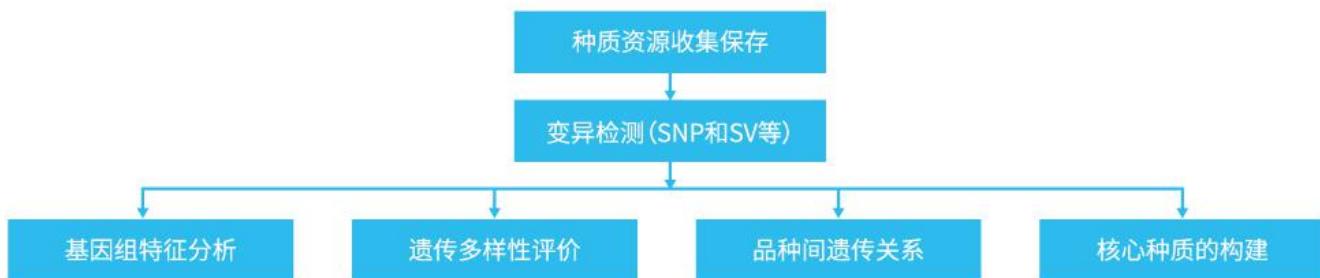
客户分布网络图



伯豪生物协助客户发表相关文章

种质资源的评价

种质资源是指具有特定种质或基因,可供育种及相关研究利用的生物类型,又叫原始材料、遗传资源、品种资源和基因资源。现代育种以种质资源为基础,特异资源对育种成果起到决定性作用,种质资源的多样性程度决定了能否实现育种目标。在育种上稀有特异种质的利用使得农业生产取得重大成就,在水稻和小麦的矮化育种过程中,矮脚乌尖和农林10号的发掘利用引导了世界范围的“绿色革命”。在品质育种中,双低油菜、高赖氨酸玉米、高蛋白高赖氨酸大麦对于作物的品质改良的作用至关重要。应用最多的是抗病方面和杂种优势的利用,大豆的抗根结线虫病、小麦的抗锈病、水稻的杂交种的推广,这些无不反映稀有特异种质的决定性作用。另外,种质资源还是生物学理论研究的载体。随着“人类基因组计划”、“国际千人基因组计划”和“国际癌症基因组计划”的实施,作物各物种的基因组计划也相继开展或酝酿,认识和评价种质资源是作物育种的第一步。在作物的长期演化过程中,遗传物质的改变(或突变)是产生遗传多样性的根本原因,遗传变异越丰富,说明其对环境变化的适应能力就越强,分布范围就越广。遗传多样性的研究主要是指利用遗传标记来反映遗传的多态性研究,其中DNA分子标记具有不受环境影响、数量多、多态性高和共显性的特点,是种质资源评价最有效的遗传标记。



案例分析:3000份水稻基因组计划

文献: Wang W, Mauleon R, Hu Z, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice[J]. Nature, 2018, 557(7703): 43.

单位:中国农业科学院作物科学研究所;国际水稻研究所、上海交通大学、深圳农业基因组研究所等

研究背景

过去20多年,全球科学家已经克隆了多个水稻基因,然而,如何将这些功能基因组的研究成果转化育种家所能利用的信息,则必须对水稻种质资源的基因组信息进行充分了解。“3000份水稻基因组计划”正是在这样的背景下应运而生,解析水稻种群基因组多样性的本质,对于高效挖掘水稻有利基因、实现基于全基因组信息的水稻分子设计育种具有重大的理论和实践意义。

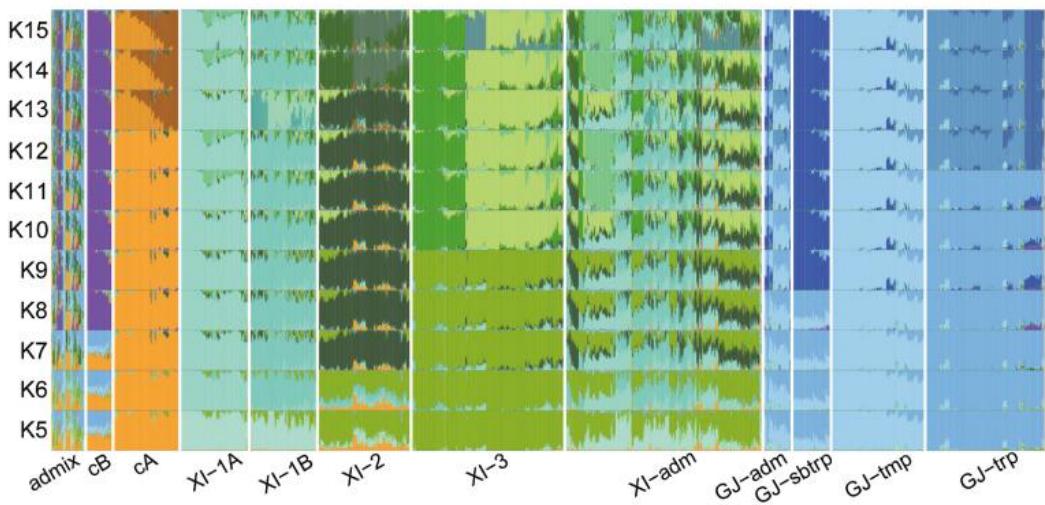
材料和方法

本研究中的3010份水稻是来自全球89个国家代表了78万份核心种质约95%遗传多样性的种质资源,其中2400份来自国际水稻研究所。利用HiSeq2000平台进行重测序,平均测序深度14×,其中用于结构变异检测的453份测序深度大于20×。以Nipponbare RefSeq为参考基因组,BWA数据比对,GATA和Picard检测变异。

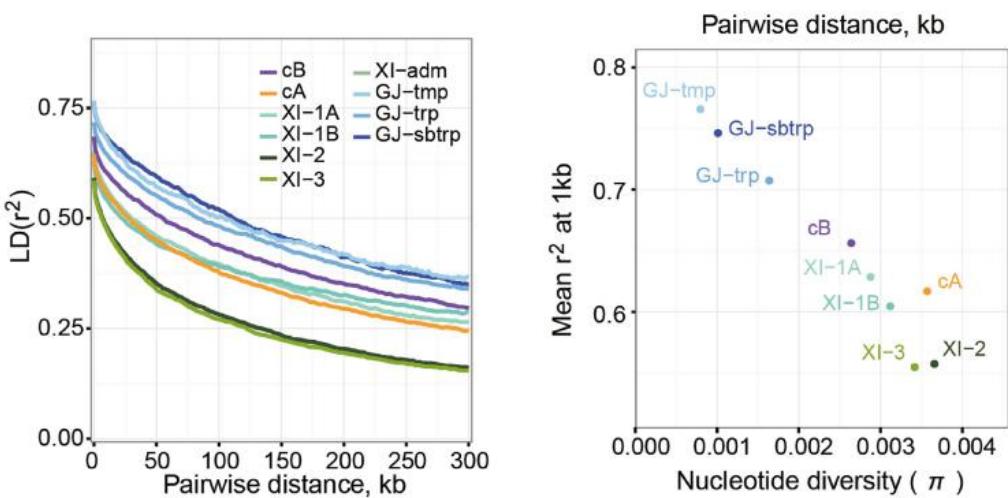
实验结果

1) 对亚洲栽培稻群体的结构和分化进行了更为细致和准确划分

利用检测到的32M的高质量SNPs进行群体结构和聚类分析发现,亚洲栽培稻由传统的5个群体增加到9个,分别是东亚(中国)籼稻(XI-1A)、现代籼稻品种(XI-1B)、南亚籼稻(XI-2)和东南亚籼稻(XI-3)等4个籼稻群体,东南亚温带粳稻(GJ-tmp)、热带粳稻(GJ-trp)、亚热带粳稻(GJ-sbtrp)等3个粳稻群体、以及来自印度和孟加拉的Aus(cA)和香稻(cB)。



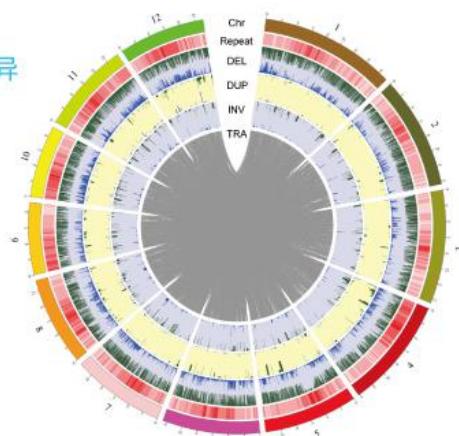
亚洲栽培稻群体的结构。XI-adm, 粳稻品种混合组; GJ-adm, 糜稻品种混合组; admix, 混合组。



各亚群连锁不平衡(左)及与遗传多样性的相关性(右)

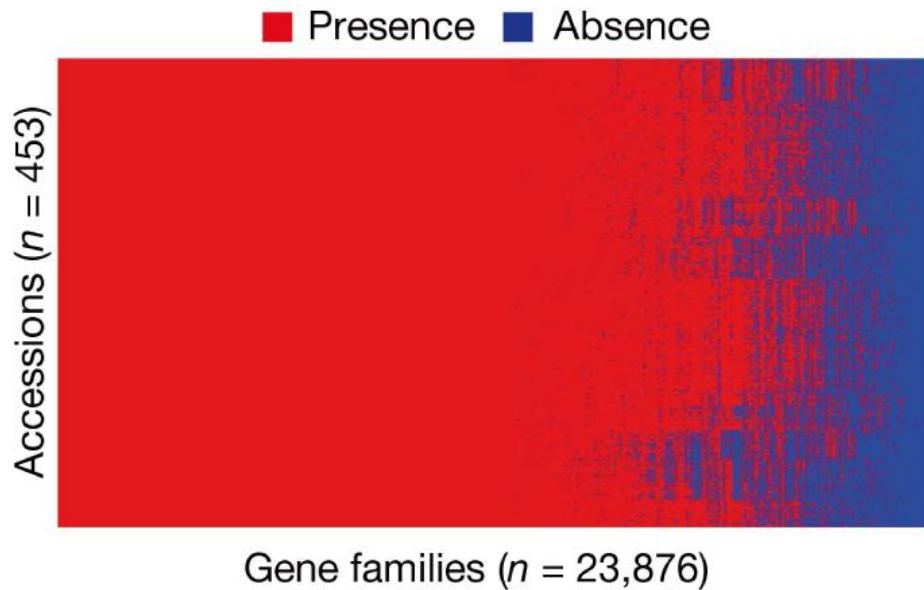
2) 首次揭示了亚洲栽培稻品种间中存在大量微细(>100bp)结构变异

亚洲栽培稻品种的结构变异。Chr, 染色体; Repeat, 红色热图表示500 kb窗口中的重复内容; DEL, 绿色/蓝色的内/外标代表在XI和GJ中检测到的缺失的平均频率; DUP, 绿色/蓝色的内/外标代表在XI和GJ中检测到的复制的平均频率; INV, 绿色/蓝色的内/外标代表在XI和GJ中检测到的倒置的平均频率; TRA, 灰色代表在XI或GJ中平均频率>0.3的基因组上的易位。



3) 构建了亚洲栽培稻的泛基因组

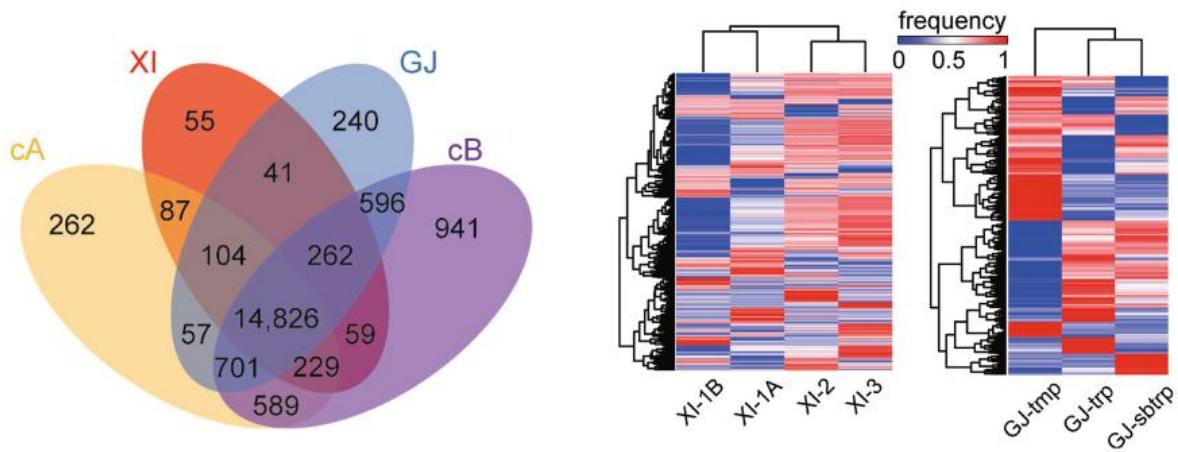
利用泛基因组, 鉴定出12,770个(62.1%)核心(core)基因家族和9,050个(37.9%)分散式(distributed)新基因家族。核心基因比较古老, 大多数的新基因表现更年轻和长度偏短。



基因家族的存在/缺失矩阵图

4) 亚洲栽培水稻起源的新观点和命名恢复

不同地理来源的水稻农家品种群体都带有特异的基因家族。根据这些结果首次提出了籼、粳亚种的独立多起源假说，并恢复使用籼(*Oryza sativa* subsp. *xian*)、粳(*Oryza sativa* subsp. *geng*)亚种的正确命名，使中国源远流长的稻作文化得到正确认识和传承。



不同地理来源的水稻农家品种群体带有特异的基因家族。核心基因家族在各亚群分布的韦恩图(左);籼稻和粳稻亚群内分布不平衡的基因家族(右)。

作物复杂性状的遗传解析

作物的成熟期、株高、籽粒产量及产量构成因素，大豆蛋白质含量，脂肪含量，玉米的赖氨酸含量等在群体中是具有连续变异特征的数量性状，而这些形态的、经济的、生理的和品质的性状是作物育种主要的改良对象。数量性状位点 (quantitative trait locus, QTL) 是控制某个数量性状发生表型变异的基因组区域或基因。新一代测序技术以及大量SNP标记的开发，使得越来越多物种基因组物理图谱被绘制完成，成为了遗传位点的准确定位及发掘的保证。准确定位数量性状的基因位点，阐明它们的效应、上位性以及与环境的互作，是当代遗传育种研究的一个重要方向，同时也是分子遗传育种的首要环节。

全基因组关联分析

全基因组关联分析 (genome-wide association studies, GWAS) 是利用不同基因座位上非等位变异之间的连锁不平衡关系，分析某一群体中目标性状与遗传标记或候选基因的相关性，从而鉴定出与表型变异相关的染色体区间及候选基因。连锁不平衡 (linkage disequilibrium, LD) 是关联分析的基础和前提，决定关联分析的试验方案、结果精度、所选用标记的数量和密度。为了减少GWAS中由群体结构等引起的假阳性，研究者不断优化关联分析方法，较为常用的有Q+K混合线性模型，多位点混合线性模型以及最近开发的限制性二阶段多位点关联分析方法等。GWAS已经成为植物数量性状位点精细定位的重要手段，随着越来越多的物种全基因组测序完成，GWAS将更广泛地应用于育种和进化领域。



案例分析：全基因组关联分析揭示大豆农艺性状的遗传网络

文献：Fang C, Ma Y, Wu S, et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean[J]. Genome Biology, 2017, 18(1):161. (IF=14.028)

单位：中国科学院遗传与发育生物学研究所 华盛顿州立大学

研究背景

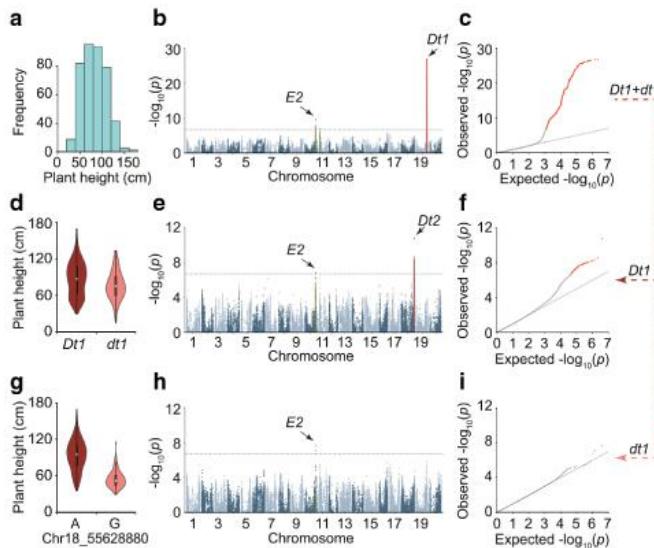
不同复杂性状间的耦合是分子设计育种的关键科学问题。作物的产量、品质等大都是多基因控制的复杂性状，由于受到一因多效和遗传连锁累赘的影响，使某些性状在不同材料和育种后代中协同变化，呈现耦合性相关。解析大豆复杂性状间耦合的遗传调控网络，明确关键调控单元，对于高产优质大豆新品种的培育具有重要意义。

材料和方法

对809份大豆栽培材料进行了重测序（平均深度为8.3×），并对其遗传多样性进行了分析，明确了这些材料的群体结构。进而，对这809份材料的84个产量和品质性状进行了连续多年多点的观测。进一步，该团队利用全基因组关联分析并结合新开发的上位性效应检测方法，对84个性状的调控位点进行了系统的全基因组扫描，解析不同农艺性状之间的内在遗传调控网络。

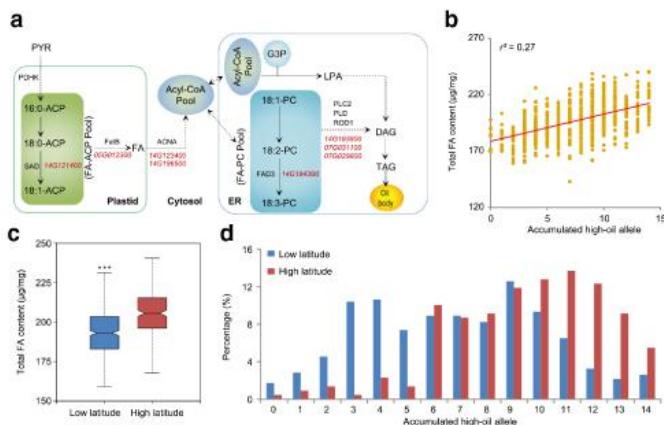
实验结果

1) 鉴定出84个性状的245个显著性关联位点



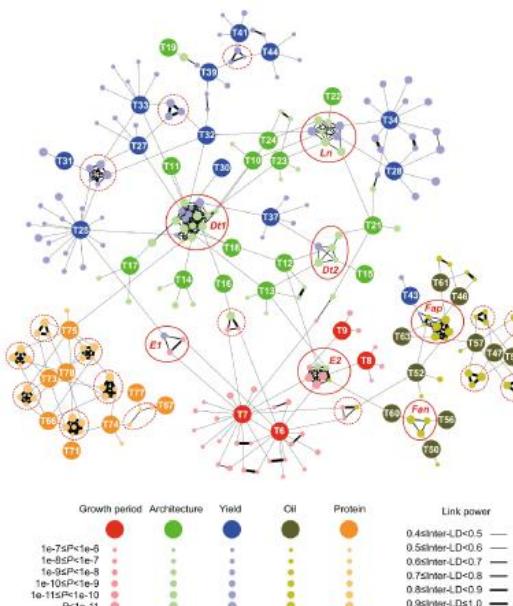
大豆株高的GWAS。a, 809个大豆品种株高的分布；b, 所有品种GWAS的结果，在GWAS结果中，已知的基因DT1和E2都被鉴定出来；c, 株高的Q-Q图；d, DT1等位基因分组后株高的变化。已知的基因DT1将809份亲本分为两个亚组，具有不同的株高平均值；e, 使用DT1亚组进行GWAS的结果；f, DT1亚组进行GWAS的Q-Q图；g, DT1亚群内不同DT2基因型分组后的株高变异；h, 使用dt1亚组进行GWAS的结果；i, dt1亚组进行GWAS的Q-Q图；水平虚线表示GWAS的显著阈值 (2×10^{-7})

2) 脂肪酸和脂代谢相关的基因



大豆脂肪酸含量遗传调控的初步探讨。a, 候选基因参与脂肪代谢途径, 主要参与大豆种子中脂肪酸(FA)合成的变异;虚线代表多个反应步骤;b, 总脂肪含量与高油含量等位基因数目的关系图;c, 低纬度和高纬度地区种质的总FA含量;d, 低纬度和高纬度群体中高油等位基因的比例。对于油含量相关性状, 共鉴定到24个脂肪酸代谢相关和21个脂代谢相关的基因, 它们分别参与了不同的重要酶促反应。深入分析发现这些基因是通过加性效应共同调控多个大豆油脂性状的形成。

3) 大豆不同性状间的遗传调控网络

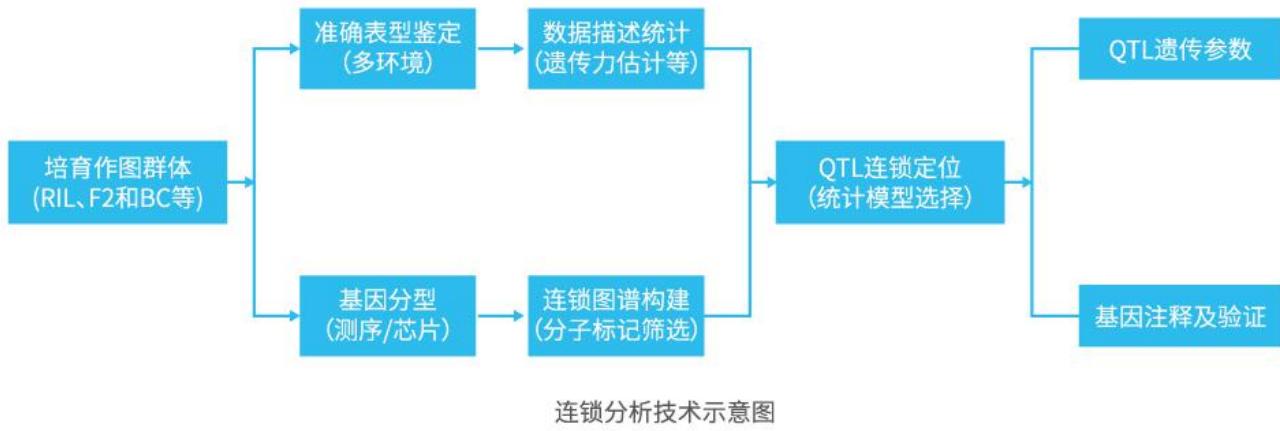


大豆不同性状间的遗传调控网络;节点表示性状及其对应的关联位点;不同性状关联位点之间的边由LD决定;涉及Dt1, Dt2, E1, E2, Ln, Fan和Fap的位点用实线圈表示,未知的位点用虚线圈表示

这些关联位点揭示了不同性状间相互耦合的遗传基础。根据连锁不平衡分析,发现115个关联位点可相互连锁,并将所观测的51个性状联系起来,形成复杂的多性状多位点调控网络,该遗传调控网络很好地解释了不同性状间的耦合关系。研究还发现其中23个关联位点起到了重要作用,对不同性状的形成起到关键调控作用,并对其中部分位点在不同性状耦合中的作用进行了验证。

连锁分析

连锁定位以Morgan的连锁遗传规律为理论依据,分析遍布全基因组的具有高多态性的分子标记和数量性状表型值间的关系,进而分析分子标记与目标性状QTL之间的连锁关系,即利用已知座位的分子标记来定位未知座位的QTL,通过计算分子标记与QTL之间的交换率,从而确定QTL的具体位置并估算其表型贡献率。连锁定位一般步骤包括:选择具有相对性状的纯系进行杂交,获得适宜的作图群体;检测分离世代群体中每一个体的标记基因型和数量性状值;构建遗传连锁图;分析标记基因型和数量性状值的相互关联,确定QTL在染色体上的相对位置,估计QTL的有关遗传参数。其中遗传图谱是通过计算遗传标记之间的重组频率,将所有的分子标记划分为不同的连锁群,对每个连锁群内的标记进行多次三点测验,确定他们的相对距离,一般用厘摩(cM, 即每次减数分裂的重组频率为1%)为单位。QTL作图的统计模型和方法常用的有多区间作图法(MIM)、复合区间作图法(MCIM)、完备区间作图法(ICIM)等。QTL连锁定位是当前最传统也是使用最广泛的对主要农作物的数量性状进行基因定位方法,研究比较深入的作物有水稻、大豆、玉米、棉花和番茄等。



连锁分析技术示意图

案列分析:连锁分析揭示大豆抗裂荚性遗传基础

文献: Han J, Han D, Guo Y, et al. QTL mapping pod dehiscence resistance in soybean (*Glycine max* L. Merr.) using specific-locus amplified fragment sequencing[J]. Theoretical and Applied Genetics, 2019: 1-20. (IF=3.926, Q1)

单位:中国农科院 作物科学研究所

研究背景

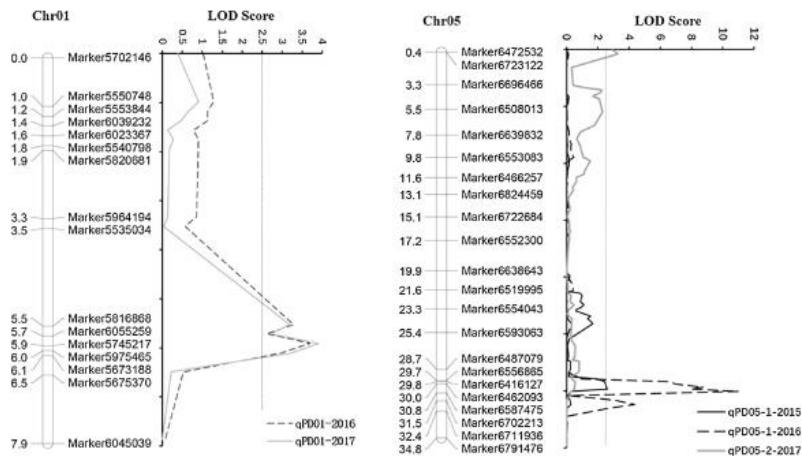
裂荚性是野生大豆繁殖的必要条件,但是在栽培大豆中,裂荚性是导致减产的重要因素。因此,在栽培大豆中,需要人为地选择抗裂荚性,为选育抗裂荚的新品种,需要检测与抗裂荚性有关的数量性状位点。

材料和方法

本研究通过黑河43和黑河18杂交获得的F2,利用单粒传法得到260个F2:7重组自交家系,用改良的烘干法测量三年的裂荚性表型。对260个家系建库(SLAF)后,利用Illumina HiSeq平台测序。两亲本测序深度分别为39.74 × 和 34.87 ×,平均测序深度12.35 ×。用HighMap构建遗传图谱,用完备区间作图法进行QTL定位。

研究结果

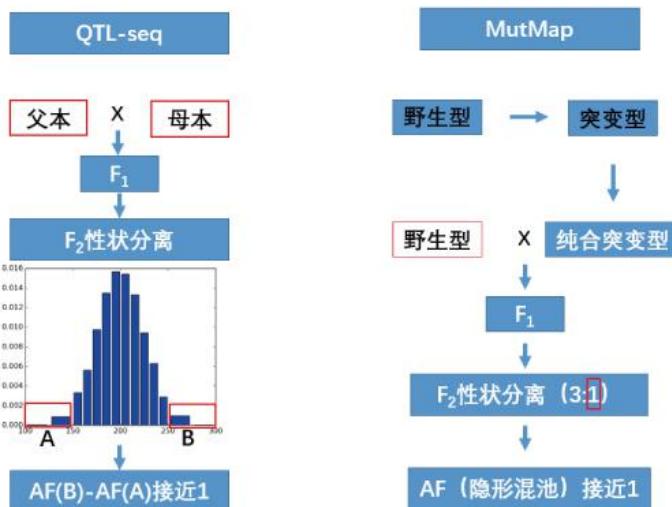
测序质控后获得364,461 SLAF 标记,筛选后用4593个构建了总长1478.86 cM的遗传图谱。在1、5、8和14号共定位到6个QTL与裂荚性相关(先前均未报道)。其中有四个表型解释率大于10%。GO注释到217个基因,其中34个在亲本间存在SNP或Indel,其中9个基因经qRT-PCR验证,表达量在亲本间存在显著差异,并开发了相应的分子标记以辅助育种。



1号和5号染色体上的抗裂荚QTL;不同线代表不同年份

BSA定位

在作物性状定位研究中,将家系后代中具有极端相反性状的个体分为两组,以此寻找与性状呈现共分离的分子标记,实现性状定位的方法,称为混合分组分析法(Bulked Segregant Analysis, BSA)。在高通量测序普及的十年内,BSA方法再次成为了性状定位研究的有力工具。在众多方法中,以QTL-seq和MutMap最为常用。QTL-seq可用于应对自然发生的数量性状和质量性状;MutMap是一种特定的应用,分析寻找的是诱变导致的突变性状,而非自然产生的突变。



QTL-seq (左)和MutMap (右)技术示意图,红框内为测序样本

案列分析:基于QTL-seq分析方法加速大豆子叶颜色的基因定位

文献:Song J, Li Z, Liu Z, et al. Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean[J]. Frontiers in plant science, 2017, 8: 919. (IF=4.106)

单位:中国农科院 作物科学研究所

研究背景

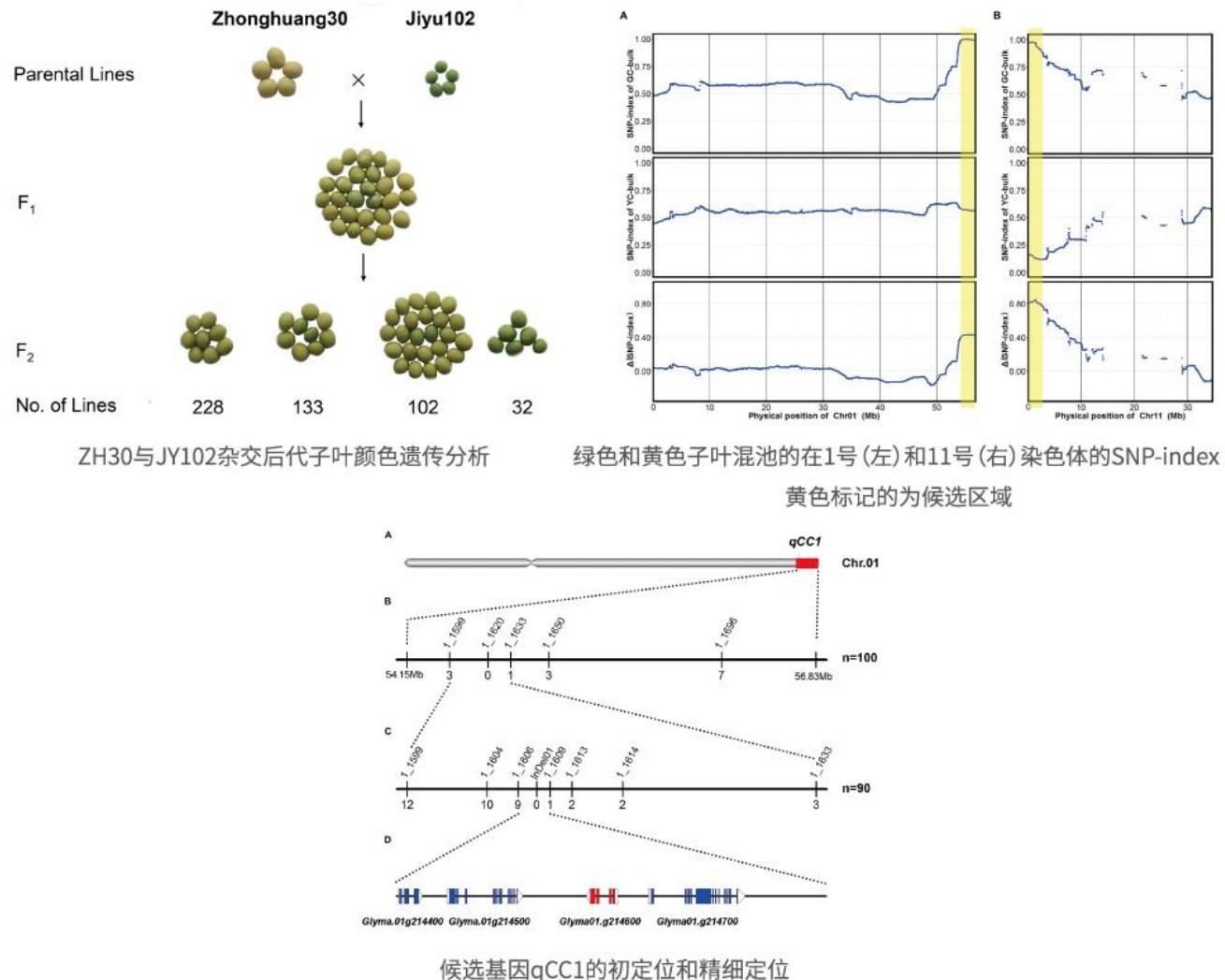
成熟种子中的子叶颜色是大豆育种和种质分类的重要形态特征。大多数大豆品种都有黄色子叶,只有少数有绿色子叶。目前已报道大豆子叶颜色有三种遗传模式(母系遗传、双和单基因遗传),经典遗传学方法表明D1、D2、cytG等基因调控该性状。

研究方法

研究中利用亲本Zhonghuang30 (ZH30, 黄色子叶, YC) 和Jiyu102 (JY102, 绿色子叶, GC) 构建F2群体, 并挑选30个黄色子叶子代混池和30个绿色子叶子代混池。利用Illumina HiSeq 2500对双亲和两个子代混池高通量测序, 亲本测序深度分别为9-12 ×, 混池测序深度为53-59 ×。

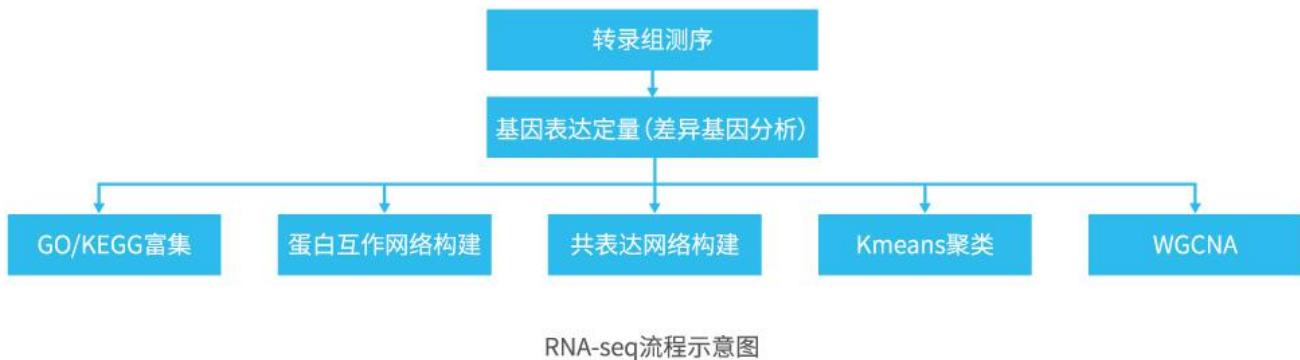
研究结果

两亲本间获得1,084,921个SNP和157,839 Indel, 经过△(SNP-index)分析, 获得qCC1和qCC2两个候选区间。接着用100个家系和候选区域的11个多态SSR标记将qCC1缩小在395Kb区间内, 再用90个家系(70个绿色子叶和20个黄色子叶)和候选区域内6个多态标记将qCC1精细定位至30.7 kb区间, 同理qCC2精细定位在67.7Kb内, 并对候选基因进行测序验证。表明BSA结合高通量测序可用于快速定位大豆性状。



基因差异表达分析鉴定候选基因

RNA-seq技术可以用于比较不同发育阶段或是正常条件下与特定胁迫处理和刺激下全转录本的变化特点, 找出转录水平变化的基因, 即差异表达基因(Different expressed genes, DEGs)。基因差异表达分析是转录组测序技术最基本也是最主要的应用, 通过精确的评估特定基因或剪接体的转录本丰度及变化, 发现基因表达与表型的相关性, 最终来解释和预测基因的功能和作用机理。共表达网络分析是对差异表达基因分析的一个重要补充。对于代谢机制或者复杂的数量性状, 有时需要对全转录组水平进行综合评估。这种网络可利用一种无向图表示出来, 包含一组彼此相连的节点, 其中每个节点对应一个基因, 如果基因之间存在显著的共表达关系, 则两个节点被连接起来。由于转录组测序获得的数据十分庞大, 对基因进行聚类和建立共表达网络, 将有助于数据的梳理, 并建立基因表达水平与表型变化之间的关系, 共表达的基因集也更有利基因功能的注释和分析。



案例分析：玉米胚乳发育和代谢的基因调控网络的研究

题目：Feng F, Qi W, Lv Y, et al. OPAQUE11 is a central hub of the regulatory network for maize endosperm development and nutrient metabolism[J]. The Plant Cell, 2018, 30(2): 375-396.

单位：上海大学 生科院

研究背景

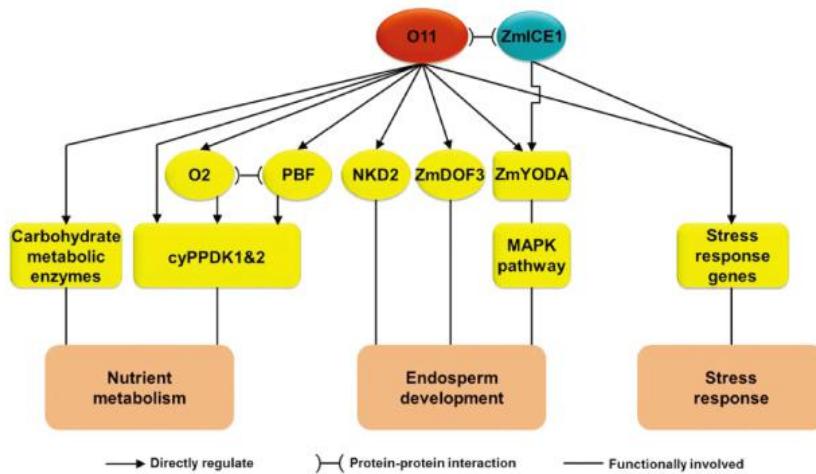
玉米胚乳是营养贮藏的主要组织，在发育过程中分化程度较高。然而，胚乳发育和营养代谢的调控网络在很大程度上仍是未知的。

研究方法

以玉米经典籽粒突变体opaque11为研究对象，通过转录组(RNA-Seq)与染色质免疫共沉淀(ChIP-Seq)分析，全面解析了O11的下游调控网络。

研究结果

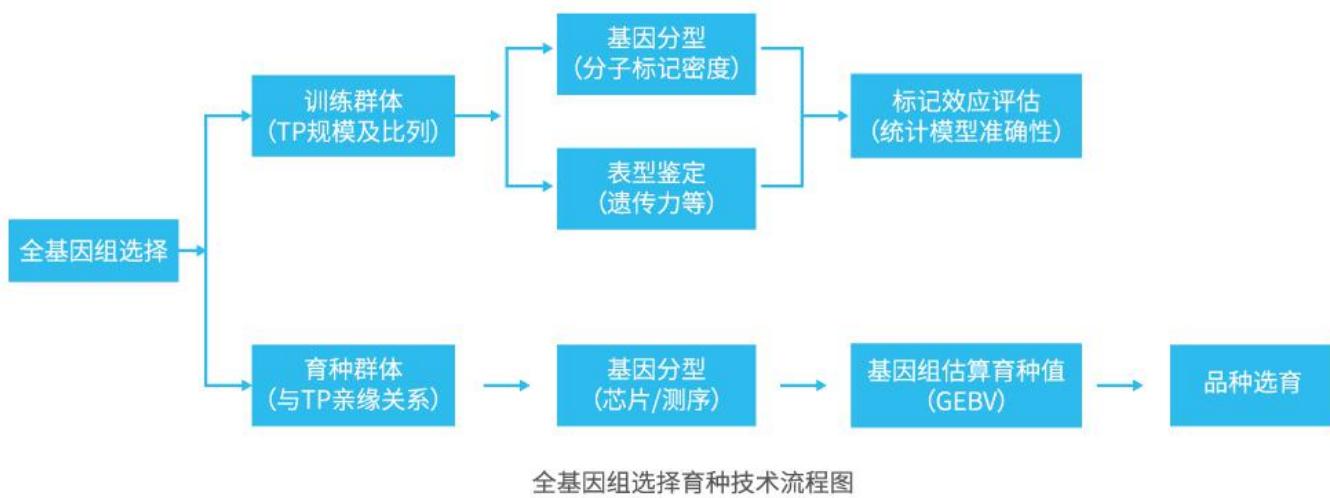
证实突变体opaque11同时影响了籽粒的发育和储藏物的积累。通过图位克隆，发现O11编码了胚乳特异的bHLH转录因子。通过转录组(RNA-Seq)与染色质免疫共沉淀(ChIP-Seq)分析，全面解析了O11的下游调控网络。发现O11不仅调控胚乳发育的关键转录因子(如NKD2和ZmDof3)；还直接调控了多个关键的储藏物代谢关键转录因子(如O2和PBF)。该研究还筛选和鉴定到多个与O11直接互作的蛋白，其中包括与冷胁迫应答相关的关键转录因子ZmICE1。研究发现，O11和ZmICE1协同调控冷胁迫应答相关的基因以及植物发育中重要信号开关ZmYODA。O11作为关键调控因子，协调了胚乳细胞发育、储藏物代谢和逆境响应等生物学过程，是胚乳整体基因调控网络的核心调控节点。



O11调控网络

全基因组选择育种

全基因组选择(Genome-wide selection, GWS),又称基因组选择(Genomic selection, GS),是通过全基因组中大量的分子标记和参照群体的表型数据,建立统计模型,估计出每一标记的育种值,然后仅利用同样的分子标记估计出后代个体的育种值并进行选择。全基因组选择理论主要利用连锁不平衡信息,即假设标记与其相邻的QTL处于连锁不平衡状态,因而由相同标记估计的不同群体的染色体片段效应是相同的,这就要求标记密度足够高以使所有的QTL与标记处于连锁不平衡(LD)状态。全基因组选择在实施过程中主要分为2个步骤:(1)通过训练群体估计出不同染色体片段的效应,主要受表型—基因型效应的影响;(2)预测群体的基因组估计育种值(Genomic EBVs, GEBVs),该步骤可以直接预测无表型记录而有基因型资料作物个体的GEBVs。在GS研究中,其预测准确度(r_{MG})被定义为真实育种值与估算育种值间的相关系数,而预测能力(r_{MP})则被定义为表型数据与EBV间的相关系数,并且预测准确度和预测能力之间的关系定义为 $r_{MP}=r_{MG} \times h$,其中 h 为目标性状遗传率的平方根。随着基因型分型成本的降低和统计方法的发展,作物基因组选择将会逐步完善,将在基因组育种中发挥重要的作用。



案例分析:基于全基因组选择的杂交稻育种

文献:Yanru C, Ruidong Li, Guangwei Li, et al. Hybrid breeding of rice via genomic selection. Plant Biotechnology Journal[J], 2019, 1-11. (IF=6.840)

单位:美国加州大学 植物与植物科学系 安徽农业大学

研究背景

利用杂交种是许多作物特别是水稻和玉米提高产量的主要策略。基因组杂交系育种是利用全基因组标记预测潜在杂交种的技术。预测出的优良杂交系在确认其优良性状后,作为新的杂交种进行田间评价和释放。这大大降低了选到真正优良杂交种的成本。

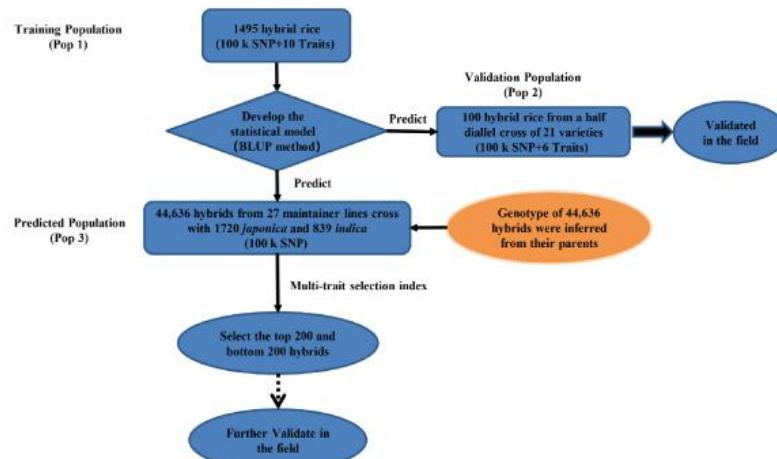
研究方法

利用训练群体(1495个杂交系,Pop1),验证群体(10个粳稻和11籼稻杂交而来的210杂交系,Pop2)和育种群体(3K水稻基因组计划群体,Pop3)三个群体。通过Illumina Hiseq 2000对Pop1和Pop2(测序深度2×)进行基因分型,Pop3基因型从相关网站下载。使用基因组最佳线性无偏估计对3000水稻基因组群体进行杂种优势预测。

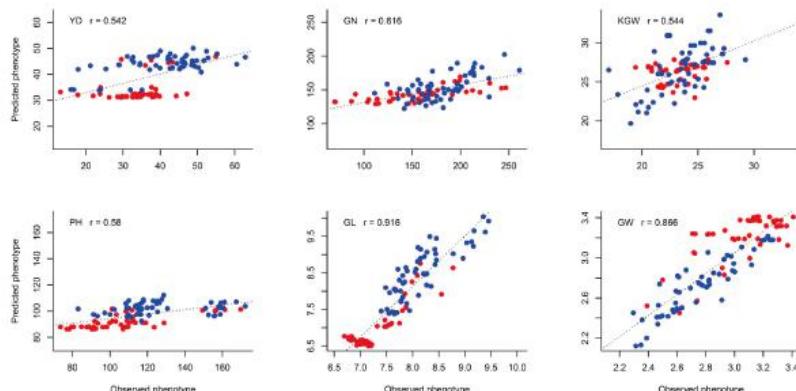
研究结果

利用筛选后的102,795个SNP,迭代10次交叉验证表明,训练群体对10个农艺性状的预测能力在0.35~0.92之间。以Pop1对

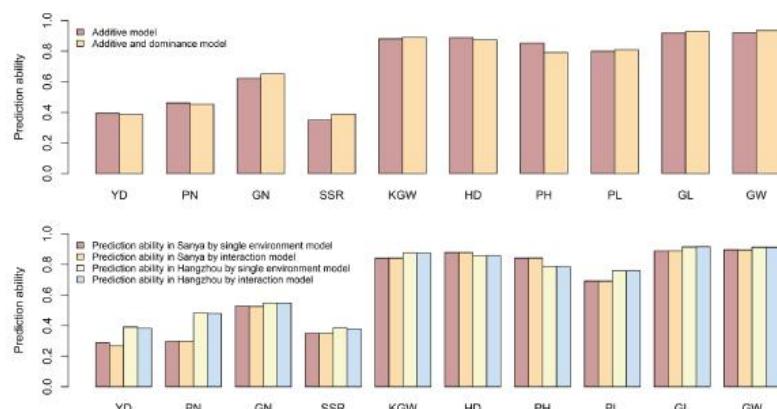
Pop2的6个农艺性状进行了预测，预测能力较高，从0.54(产量)到0.92(粒长)不等。可见，目前的1495个杂交系群体可以用来预测来自不相关的亲本的杂交后代。最终，研究者利用这一训练群体预测了“3K水稻基因组计划”的3010个水稻品种中44,636个细胞质雄性不育系潜在的杂交种。结合10个性状的育种指标，确定了前200名和后200名潜在杂交种。最终，研究者认为，该训练群体的SNP基因型及其估计值可用于水稻杂交系预测的一般应用及进一步验证。



以1495个杂交系为训练群体进行全基因组选择的流程图



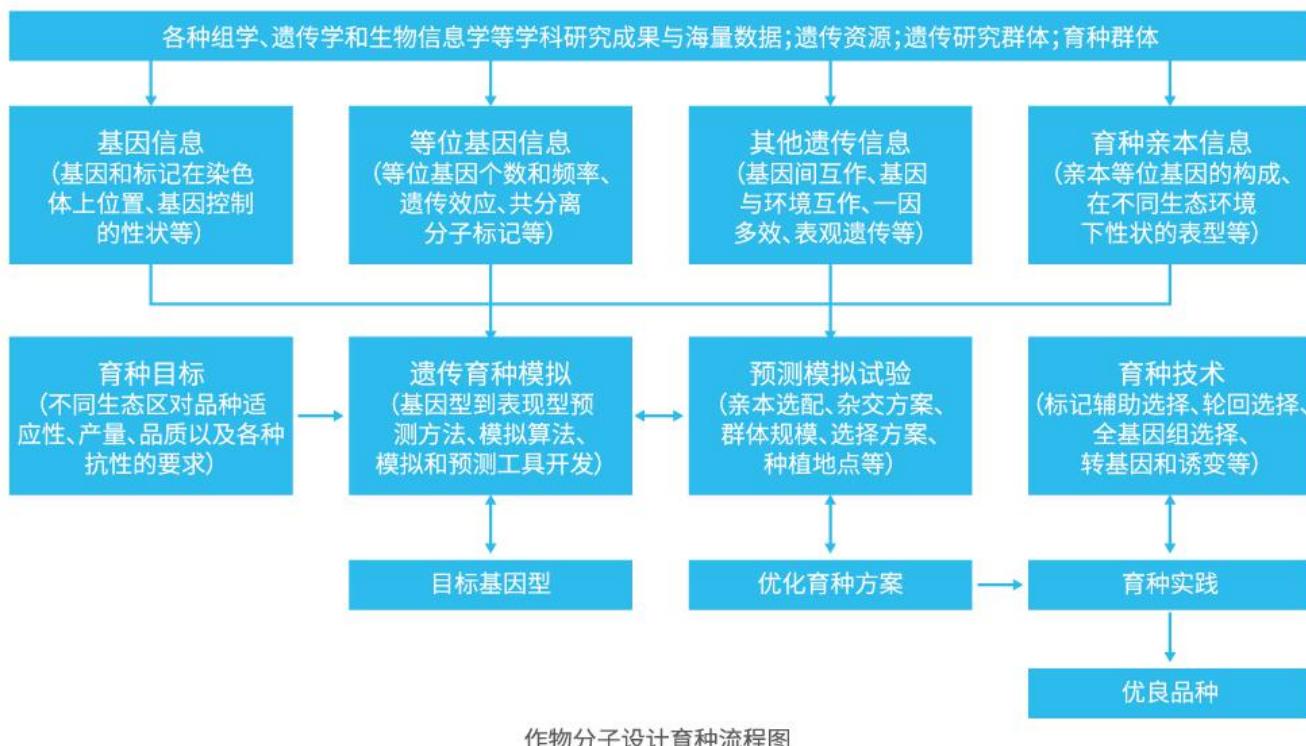
Pop2杂交系预测表型与观测表型对比图。“粳稻”用红色表示，“籼稻”用蓝色表示。



不同模型的预测能力。1495杂交系在加性和显性模型下对10个性状的预测能力(上)；
单环境和环境互作模型对10个性状的预测能力(下)。

分子设计育种

Peleman 和 van der Voort 在 2003 年明确提出设计育种的概念，万建民等在 2006 年又进一步明确分子设计育种应当分三步进行：定位相关农艺性状的基因位点，评价这些位点的等位变异，确立不同位点基因间以及基因与环境间的相互关系；(2)根据育种目标确定满足不同生态条件、不同育种需求的目标基因型；(3)设计有效的育种方案、开展设计育种。分子设计育种提出到现在已成为引领作物遗传改良的研究领域。设计育种的核心是建立以分子设计为目标的育种理论和技术体系，通过各种技术的集成与整合，在育种家进行田间试验之前，对育种程序中的各种因素进行模拟、筛选和优化，确立满足不同育种目标的基因型，根据具体育种目标设计品种蓝图，提出最佳的亲本选配和后代选择策略，结合育种实践培育出符合设计要求的农作物新品种，最终大幅度提高育种效率，实现从传统的“经验育种”到定向、高效的“精确育种”的转变。



案例分析：通过分子设计育种获得高产优质水稻

题目：Zeng D, Tian Z, Rao Y, et al. Rational design of high-yield and superior-quality rice[J]. Nature plants, 2017, 3(4): 17031. (IF=13.297)

单位：中国科学院遗传与发育生物学研究所 中国农科院水稻所 深圳农业基因组所

研究背景

育种学家和稻米种业长期以来致力于培育“高产优质”型超级水稻新品种，但是传统育种进展缓慢。随着水稻功能基因组的发展，“品种设计育种”应运而生，其重要内容之一是将重要农艺性状关键基因的优异等位变异高效聚合，形成超级新品种。然而，品种设计育种实施中优化育种策略的设计面临挑战，尤其对于高产优质复杂性状综合改良，迄今还没有实现从概念到产品的报道。

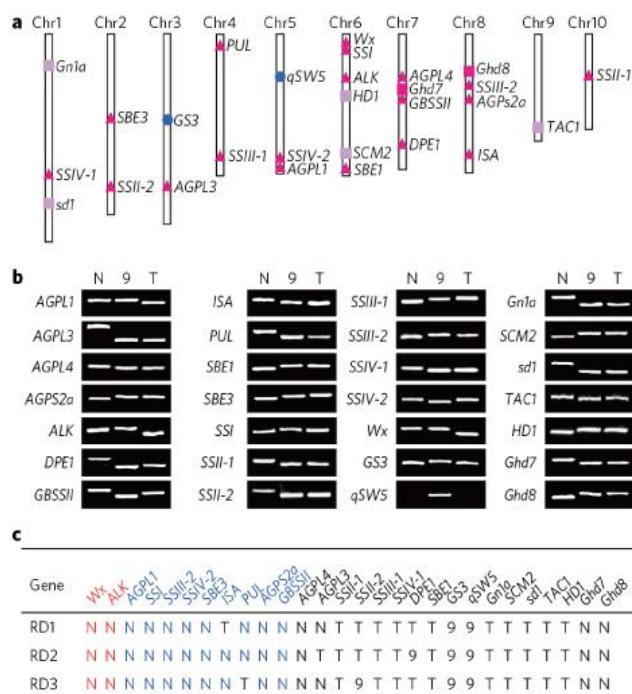
研究方法

利用数百个已经鉴定出的水稻产量相关 QTL 或基因，以及淀粉合成相关基因组成的遗传网络（和水稻蒸煮品质相关）经过精心设计，以超高产但综合品质差的品种“特青”作为受体，以蒸煮和外观品质具有良好特性的品种“日本晴”和“93-11”为供体，对

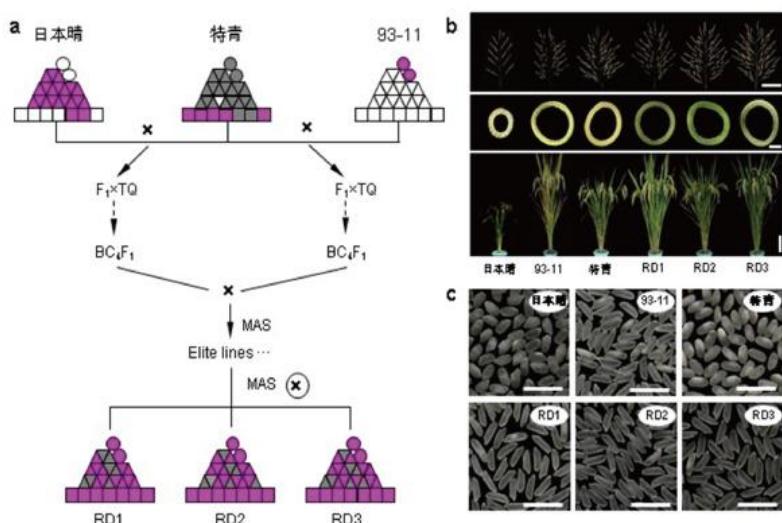
涉及水稻产量、稻米外观品质、蒸煮食味品质和生态适应性的28个目标基因进行优化组合,利用杂交、回交与分子标记定向选择等技术,以求将优质目标基因的优异等位变异聚合到受体材料,并充分保留“特青”的高产特性。

研究结果

这些优异的“品种设计”材料,在高产的基础上,稻米外观品质、蒸煮食味品质、口感和风味等方面均有显著改良,并且以其配组的杂交稻稻米品质也显著提高。这项研究结果将极大推动作物传统育种向高效、精准、定向的分子设计育种转变。



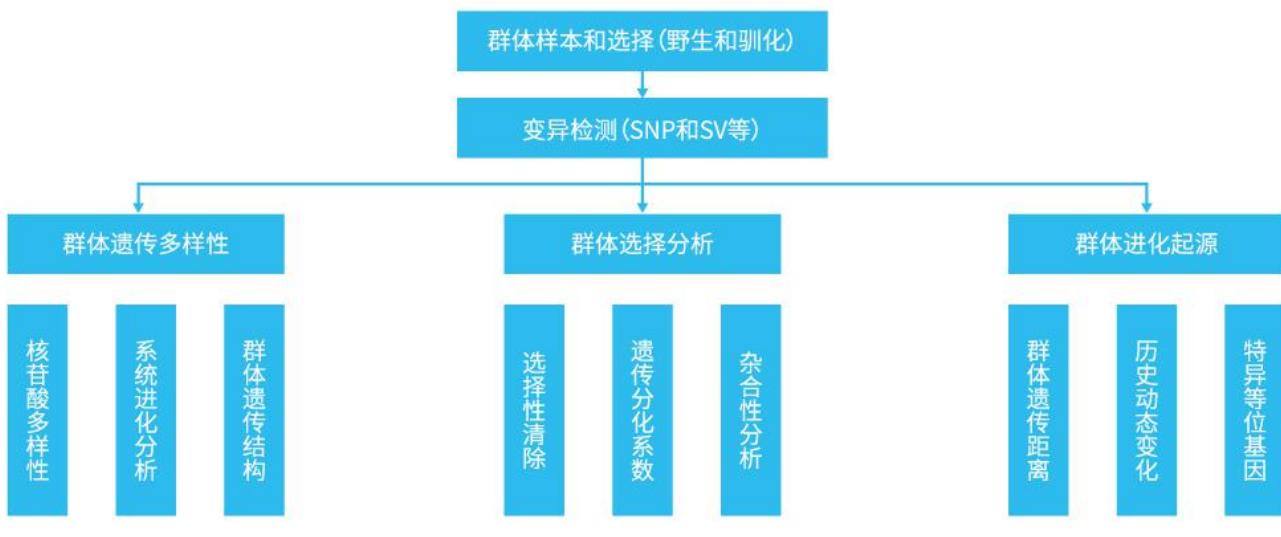
分子设计育种用到的基因和标记。a.设计育种用到的基因在染色体上的分布;三角形,与淀粉合成相关的基因;方型,表示与籽粒产量有关的基因;圆型表示与籽粒外观有关的基因;粉红色, NPB中的有利等位基因;蓝色, 93-11中的有利等位基因;薰衣草色,特青中的优良等位基因。b.设计育种用到的标记在NPB、93-11和Teqing之间的多态性,其中N、9和T分别表示NPB、93-11和Teqing。c.目的基因在RD1、RD2、RD3的基因型。红色,控制蒸煮食味品质的大效应基因;蓝色,控制蒸煮食味品质的小效应基因。



高产优质水稻品种设计及品种培育。a.高产优质水稻品种的流程设计;b.设计培育品系与亲本产量相关特性比较;c.设计培育品系与亲本品质相关特性比较;RD1、RD2和RD3分别为3个设计型新品系。

作物群体遗传进化

群体进化是基于群体遗传学方法研究不同群体之间的结构、多样性和演化规律的一门学科。随着高通量分型技术的发展，群体遗传学研究已从使用单一或少数分子标记，转为使用基于全基因组的分子标记，如使用SNP芯片、全基因组重测序和简化基因组测序等技术得到的SNP标记。群体进化主要讨论以下几方面问题：(1) 群体间的进化关系，群体结构和遗传多样性；(2) 群体间的选择信号，通过遗传分化和选择性清除分析，筛选受到选择的基因，用以解释个体的驯化机制或种群的环境适应性进化；(3) 群体的历史动态变化，例如有效种群大小，基因流，群体分化时间等。



案列分析：棉花驯化过程中不对称的亚基因组选择和顺式调控的分化

文献：Wang M, Tu L, Lin M, et al. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication[J]. Nature Genetics, 2017, 49(4): 579.

单位：华中农业大学 植物科学技术学院

研究背景

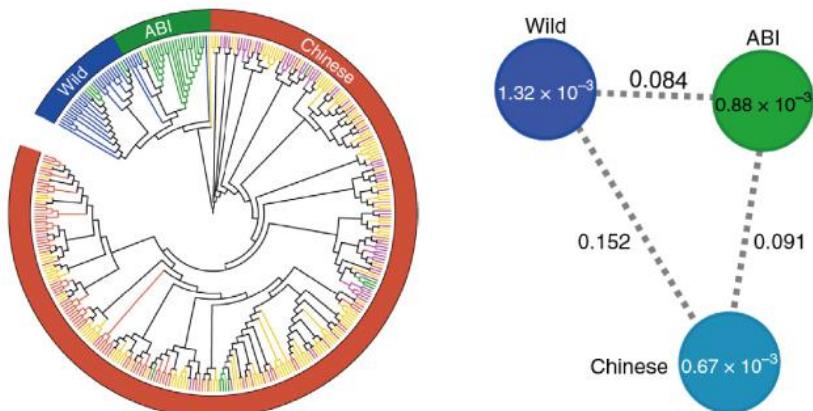
陆地棉驯化和栽培已有5000年以上。生产上主要棉花栽培种为异源四倍体陆地棉(AADD)，其生态适应性广。棉花野生种含有大量遗传变异，而长期的人工驯化选择了一些优异的变异，如栽培种纤维长度较长，能产生白色纤维，野生种纤维略带棕黄色。但是，科学界很少有研究剖析这些性状改变的遗传学基础。

研究方法

对352份陆地棉利用Illumina HiSeq 4000平台重测序，平均测序深度 $6.9\times$ ，主要分析有：群体遗传分化(NJ、PCA、LD、 π 和FST)；驯化选择区域鉴定(XP-CLR和 π_w/π_c ,用QTL hotspots GWAS验证)；不对称的亚基因组选择(同源等位基因对, 纤维长度颜色)；启动子顺式作用原件的驯化(DNA酶I测序、Chip验证及MOTIFS)；增强子的进化(Hi-C、DNA酶I测序和Chip验证)等。

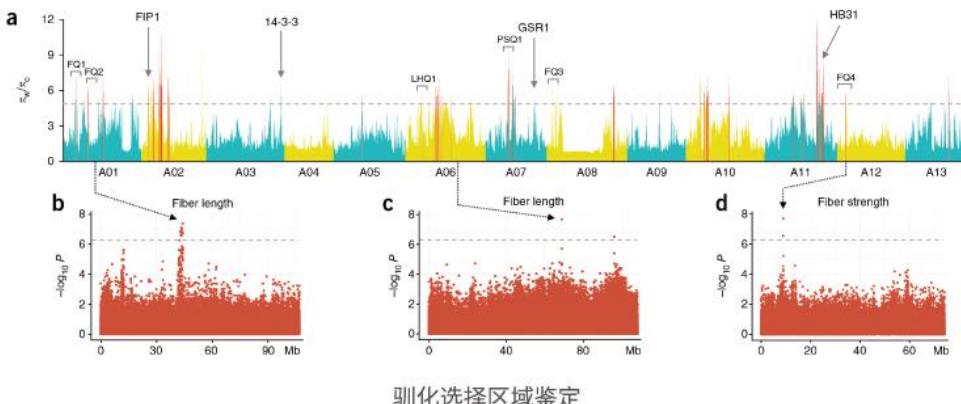
实验结果

1) 群体结构及亚群遗传多样性



352份陆地棉群体结构(左)及亚群遗传多样性(右)

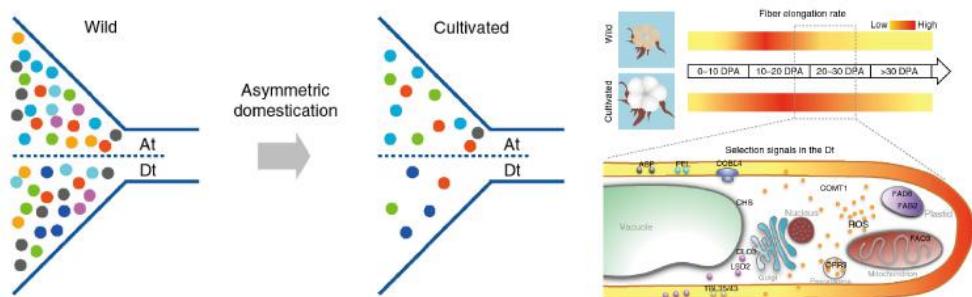
2) 驯化选择区间鉴定



驯化选择区域鉴定

At上存在74Mb驯化选择区域,有549个基因和6个QTL hotspots;Dt上存在104Mb驯化选择区域,有1228个基因和19个QTL hotspots。

3) 纤维长度在驯化过程中的不对称亚基因组选择



驯化过程中存在不对称亚基因组选择。At和Dt之间的不对称驯化模型(左);Dt特异性选择信号对棉花纤维伸长的影响(右),上部分显示了野生和栽培棉花纤维的形态和发育差异,热图显示了二者纤维伸长率,虚线框表示棉花的伸长期延长;下部分显示一个正在发育的纤维模型,图中所示基因为在Dt中有选择信号且在栽培棉纤维发育中下调。

鉴定15,456个同源基因对,其中620受到驯化选择,At有192个,Dt有428个。在驯化选择区域内,与纤维长度相关的18个基因,一个在At,17个在Dt,存在不对称选择。

品种指纹图谱及纯度鉴定

种子是农业生产中最基本的生产资料。选择适宜的分子标记及其配套检测技术，建立一套简便、经济、准确而高效的作物品种鉴定和指纹图谱构建技术体系，是打击假冒伪劣种子、保护作物遗传育种家、种子生产经营企业和农民的合法权益、规范我国种子市场和提高我国种子产业国际市场竞争力的迫切需要。基因是决定农作物形态结构、生理生化特性的最根本因素。因此，通过分子标记技术，从基因水平来区别不同农作物品种是品种身份鉴定最直接最可靠的方法。相对于以往的鉴定方法，SNP指纹鉴定技术具有稳定（不受生长阶段和外界环境的影响）、数量多和易实现自动化的优点，因此，成为品种身份鉴定的首选方案。



建立高通量DNA指纹检测与分析平台的示意图

案例分析：大豆品种分子ID建立和遗传多样性评估所需SNP数量的筛选

文献：Liu Z, Li J, Fan X, et al. Assessing the numbers of SNPs needed to establish molecular IDs and characterize the genetic diversity of soybean cultivars derived from Tokachi nagaha[J]. The Crop Journal, 2017, 5(4): 326-336.

单位：中国农科院 作物科学研究所

研究背景

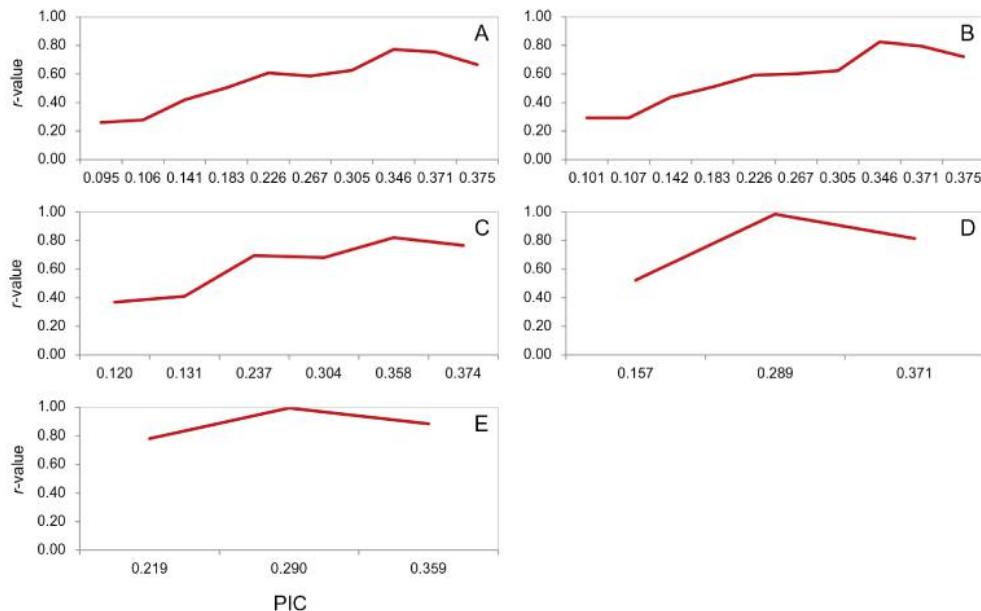
我国保存大豆种质30000余份，其中7%为育成品种。目前假冒伪劣种子禁而不绝，如“克山1号”和“中黄13”等大豆品种均发生过品种权纠纷事件。当前，我国采用36对SSR标记进行大豆品种鉴定，已无法满足检测的需求。SNP标记数量多，分布更均匀，检测手段灵活，能满足不同层次的需求，因此，急需开发一套适合大豆品种身份鉴定的核心SNP标记。

研究方法

用Illumina SoySNP6k iSelect BeadChip对德胜长叶及其137个后代进行基因分型，使用4044个SNP标记进行分析，目的是基于遗传距离矩阵相关法确定大豆品种分子ID建立和遗传多样性评估所需SNP数量。

实验结果

研究结果：在保持SNPs数目不变的情况下，随着SNPs多态性信息含量(PIC)值的增加，可以区分的品种对数也随之增加。从11个连锁群中选出的20个核心SNP，平均PIC值为0.3703，能够鉴定本研究中99.92%的品种对(9445对)。不能用该核心SNP组鉴定的8对品种均来自同一省份，其中一些具有相同的亲本。利用核心的20个SNP确定了138份材料的分子ID。当选择用于遗传多样性分析用的SNP时，应评估SNP的数量和PIC值。在本研究中，当PIC值为0.3460，SNP数目为200时，SNP在总群体的矩阵与遗传距离矩阵之间的相关系数较高(大于0.800)。高精度、高分辨率的核心SNP组可用于大豆种质资源指纹鉴定和遗传多样性的评估。



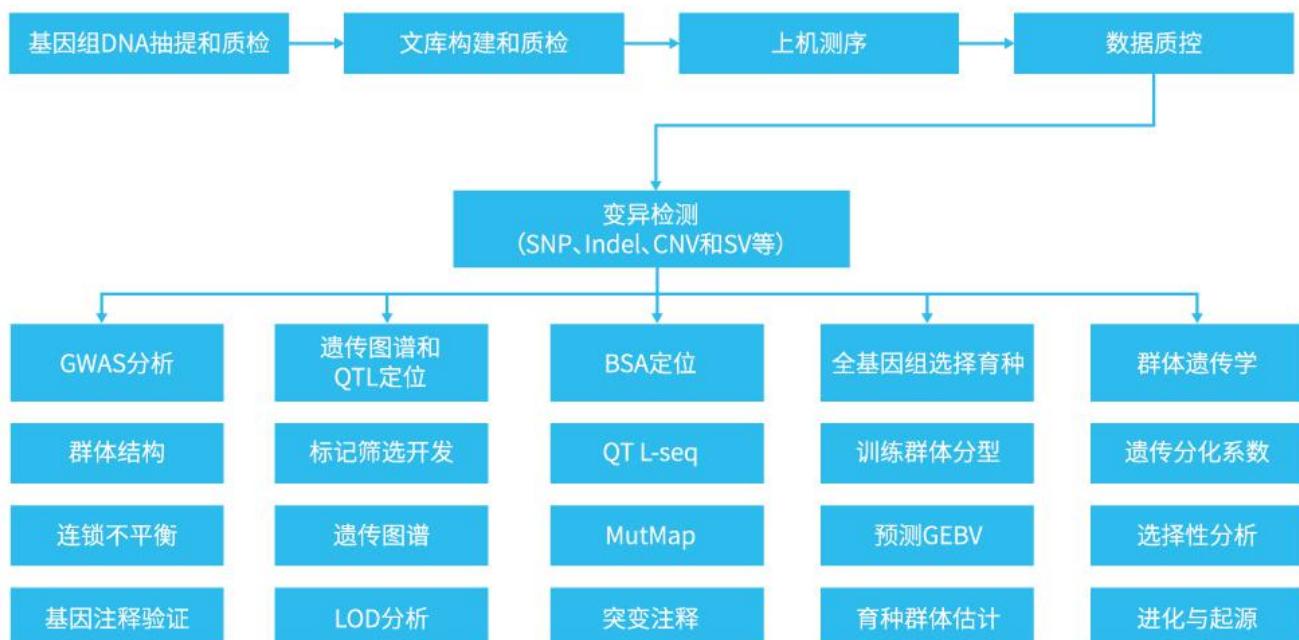
SNP在总群体的矩阵与遗传距离矩阵之间的相关系数。A、B、C、D和E分别用到100、200、500、1000和2000个SNP。

伯豪生物技术平台



高通量测序平台

上海伯豪生物技术有限公司自2009年起先后建立了ABI SOLiD新一代测序服务平台和Illumina新一代测序服务平台,全面推动中国基因组学服务产业的发展。



利用高通量测序技术检测基因组变异的技术路线

芯片基因分型平台

上海伯豪生物技术有限公司作为国内较早从事生物芯片服务的团队，具有多年的研究和服务经验，已建立了SBC点制芯片、Affymetrix芯片、Agilent芯片、Illumina芯片等微阵列技术服务平合，为国内外数千家科研单位和企业提供技术服务和技术支撑。

Affymetrix SNP基因分型芯片

美国Affymetrix公司开发的寡核苷酸原位光刻合成专利技术(light-controlled in situ synthesis of DNA microarrays)，是目前高密度的芯片制备技术。Axiom®基因分型解决方案为您提供多种芯片。您可以选择要研究物种的自定义内容，也可以选择来自Axiom®基因组数据库的基因型经过验证的内容。优点：(1)强大。对任何物种、任何基因组规模和任何倍性水平进行基因分型；Axiom®分析可检测插入或缺失 (InDel) 并保证包含所有候选SNP，与相邻SNP最近可达20 bp，实现了更高效的QTL分析。(2)可靠。低至100 ng DNA，即可获得基因分型结果，适用于各种样本类型；基因型检出率≥99%。(3)扩展。完全自动化的流程，每周可处理最多8张芯片板，而无需增加人工或仪器，一张芯片板上有96个或384个样本，检测每个样品多达260万个变异。

Illumina SNP基因分型芯片

Illumina 公司以测序业务闻名，而其在生物芯片方面的业务也具有不可忽视的优势和竞争力。Illumina SNP Genotyping采用激光共聚焦微珠芯片技术(BeadArrayTM)，可对全基因组或特定SNP位点进行分型，其检测质量可靠，得到业界广泛认可。

芯片产品目录

芯片名称	物种	位点数	最少样本数
植物物种			
Axiom® Soybean Genotyping Array	大豆	181K	192
Axiom® Cotton Genotyping Array	棉花	36K	192
Axiom® Wheat 820k Genotyping Array Plate A	小麦	820K	96
Axiom® Wheat 820k Genotyping Array Plate B	小麦	820K	96
Axiom® Wheat Breeder's Genotyping Array	小麦	35K	768
Axiom® Strawberry Genotyping Array	草莓	90K	192
Axiom® Maize Genotyping Array	玉米	600K	192
Axiom® Rose Genotyping Array	玫瑰	69K	192
Axiom® Rice Genotyping Array	水稻	43K	192
Maize SNP50	玉米	~50K	48
Maize LD BeadChip	玉米	~3K	48
个性化定制			
Axiom® myDesign TG Array	定制	1K-2.6M	>480
iSelect HD Custom Genotyping BeadChips	定制	3k-90k	24
iSelect HTS Custom Genotyping BeadChips	定制	90k-700k	24

注：更多物种相关芯片产品详询伯豪生物。

其它基因分型平台

基于多重PCR的目标区段/SNP的高通量重测序

设计和优化多重PCR引物，扩增目标区域，不同样本加上不同index区分，利用高通量测序平台进行深度测序，鉴定目标区域中的多态或突变位点。优点：自主引物设计及优化平台，保证测序结果特异性及均一性；针对研究目标，灵活定制研究方案，测序有效覆盖度95%以上，准确率>98%。

KASP基因分型技术

KASP (Kompetitive Allele-Specific PCR)，即竞争性等位基因特异性PCR。KASP反应包括Primer Mix和Master Mix两种主要成分。Primer Mix由两条末端碱基不同的等位基因正向引物与一条反向引物构成，两条正向引物5'端分别连有不同序列的检测引物序列。Master Mix包含两条带有不同荧光的检测引物。优点：优异的准确性（独立评价的准确性>99.8%）；超强的灵活性；前所未有的低成本（无需昂贵的双色标记探针）；高效的成熟技术。

Fluidigm技术对SNP位点的验证

Fluidigm的核心技术是微流控阀门技术，Biomark HD系统是目前唯一一种可进行基因分型、基因表达profiling、实时数字定量PCR (qdPCR)和单细胞分析的多用途实时PCR系统。Fluidigm BioMark有三种芯片：96*96, 48*48和192*24。优点：高通量，成本低；灵敏度高，位点特异性探针，荧光标记的延伸引物，与Taqman金标准吻合度高；时间短，从样本处理到拿到数据3-4个小时；灵活性，任意物种，只要知道目的序列，探针可以定制。

MassARRAY质谱分析的基因分型平台

通过PCR将覆盖SNP位点的DNA序列扩增出来，然后应用单一延伸引物(extension primer)扩增PCR产物，然后用飞行质谱进行分析，根据差别进行分型。优点：高性价比，在所有的SNP分型方法中，价格低；高准确性，采用尖端的飞行质谱的分型技术，比RT-PCR技术更精准；高通量，一张芯片可对384个样本进行多重检测，每个体系可实现36重反应；广泛性：可测试各种类型的样本；高效性，全自动分析数据，最快一天内可以完成基因分型报告。

生物信息分析平台

公司定制了天梭TS10000高性能集群，该集群在设计上专注于提高HPC应用的运行效率。伯豪生物还针对广大科研工作者设计工作需要开发向用户开放的个性化实时在线分析系统—SBC Analysis System, SAS系统。同时，软件系统采用商业化和自主开发相结合的方式建立了微阵列芯片数据分析综合解决方案，可针对表达谱芯片、miRNA芯片、SNP芯片、DNA甲基化芯片、CGH芯片等提供各种个性化服务。

样品管理平台

公司样本中心采用LIMS系统进行样本全程管理，年分析样本30000多例。充足的样本储存空间，实时监控，专人管理，严格的样本接收、返还、销毁SOP，确保客户样本资源得到很好的管理和保护。经过十年的经验积累，针对不同物种、类型选择相应的抽提方法，并帮助客户选择稳定的试剂进行抽提，采用了NanoDrop和Agilent 2100核酸质检系统进行样品质检工作，确保RNA/DNA的总量及质量满足后续的实验要求。



服务科技创新，护航人类健康！

咨询热线:800-820-5086/400-880-5086

电话:021-58955370

网址:www.shbio.com

邮箱:market@shbio.com

地址:上海张江高科技园李冰路151号