



网址:www.bgitechsolutions.com
邮箱:info@genomics.cn

华大基因
BGI

华大基因·总部(深圳)
地址:深圳市盐田区北山工业区综合楼(518083)
电话:400-706-6615

华大基因·美洲(波士顿)
地址:One Broadway, 14th Floor, Cambridge, MA02142, USA
电话:+1-617-500-2741

华大基因·欧洲(哥本哈根)
地址:Ole Maaloes Vej 3, DK-2200 Copenhagen N, Denmark
电话:+45-7026-0806

华大基因·日本(神户)
地址:Kobe KIMEC Center BLDG.8F, 1-5-2 Minatojima-min-amimachi, Chuo-ku,
Kobe City, Hyogo-pref.650-0047, JAPAN
电话:+81-785-996-108

华大基因·亚太(中国香港)
地址:香港新界大埔工业村大富街16号
电话:+852-3610-3510

本手册仅供客户学习、交流和研究使用,请勿用于商业用途,违者必究。

版权声明:本手册版权属于深圳华大基因股份有限公司所有,未经本公司书面许可,任何其他个人或组织均不得以任何形式将本手册中的各项内容进行复制、拷贝、编辑或翻译为其他语言。本手册中所有商标或标志均属于深圳华大基因股份有限公司及其提供者所有。

版次:2019年9月版

华大基因
BGI



产品手册 **动植物研究**

目录

02

华大简介

公司简介	03
平台介绍	04
质量管理	06

07

基因组学

动植物 <i>de novo</i> 测序	09
动植物重测序	20
动植物简化基因组测序	27

33

转录组学

转录组测序	35
全长转录组测序	42
RNA-Seq	48
Small RNA测序	54
长链非编码RNA(lncRNA)测序	60

66

表观组学

全基因组甲基化测序 (WGBS)	67
ChIP-Seq	71

关于
我们

About
Us

华大简介

- 01 | 公司简介 
- 02 | 平台介绍 
- 03 | 质量管理 

公司简介



深圳华大基因股份有限公司(简称华大基因)是华大集团下属子公司,华大基因秉承集团“基因科技造福人类”的愿景,以推动生物研究进展和提高全球医疗健康水平为出发点,基于基因领域研究成果及生物技术在民生健康方面的应用,进行科研和产业布局,致力于助力和加速科学创新,减少出生缺陷,加强肿瘤防控,抑制重大疾病对人类的危害,实现精准治愈感染,助力精准医学。

华大基因依托世界领先的生物信息研发、转化和应用平台,数百台高性能测序仪、质谱仪和强大的服务器存储,为数据的输出、存储、分析提供有力保障。目前华大基因的主营业务为通过基因检测等手段,为医疗机构、科研机构、企事业单位等提供基因组学类的诊断和研究服务。

华大基因总部位于中国深圳,在京、津、汉、沪、穗等主要城市设有分支机构和临床检验中心,并在中国香港、欧洲、美洲、亚太等地区设有核心实验室,已形成“覆盖全国、辐射全球”的网络布局。

公司主要服务于国内外的科研院校、研究所、独立实验室、制药公司等机构,以及国内外的各级医院、体检机构等医疗卫生机构、公司客户和大众客户。

目前公司服务已经覆盖了全球100多个国家和地区。包括国内31个省市自治区的1500多家科研机构和800多家医疗机构,其中三甲医院100多家;欧洲、美洲、大洋洲等地区合作的海外医疗及科研机构超过2000家。

华大科技于2012年完成整合,致力于成为全球生命科学研究机构的首选合作伙伴。为从事生命科学研究的机构和企业提供高质量、行业领先的新一代测序、信息分析、基因分型、蛋白和代谢组质谱检测、Sanger测序、Oligo合成、生物云计算等标准化的生物技术服务,也可依据用户的个性化需求提供定制产品服务。

目前,华大科技已经主导及参与发表了CNNS文章超过100篇,全部科研论文1000余篇。

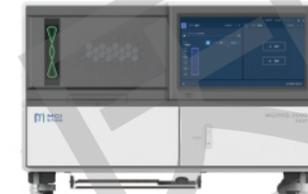
平台介绍

► 核酸测序平台

拥有BGISEQ、Nanopore、PacBio、3730 和 Illumina HiSeq/MiSeq 新一代测序平台,服务产品涵盖动植物、微生物、人及药物研发等多领域的核酸水平研究。



DNBSEQ-T7



MGISEQ-2000



BGISEQ-500



BGISEQ-200



BGISEQ-50



Nanopore平台(PromethION)



Nanopore平台(MinION)



PacBio平台(Sequel I、II)



HiSeq 2500/4000



MiSeq

质谱平台

拥有包括Orbitrap Fusion Lumos、Q Exactive/HF/HF-X、QTRAP 6500+、Triple TOF 5600+、Xevo G2-XS、TQS等涵盖轨道阱、飞行时间、三重四级杆以及各类离子源和碎裂模式的40余套质谱仪系统，配套高效样本制备和分子分离平台，能实现工业规模和科研层次的蛋白质组学、多肽组学、代谢组学研究及目标分子检测。



Orbitrap Fusion Lumos



Q-Exactive/ Q-Exactive HF/
Q-Exactive HF-X



QTRAP 6500+



Xevo-G2-XS



TQS/TQD



TSQ Altis

技术平台

拥有深圳、香港、北京、武汉、杭州等数个大型生物信息学超级计算中心，总峰值计算能力达到 820.1028 T flops，总内存容量达到 243.144 TB，总存储能力达到 104.3152 PB(截至 2018 年 3 月)。其中位于深圳和香港的集群的峰值计算能力分列国内生物信息领域第一和第二位，有能力为海量生物信息学数据的存储、处理和分析提供稳定而高效的资源保障。下图所示为华大基因在中国各地所部署的超级计算集群。



深圳平台



天津平台



武汉平台



杭州平台



香港平台

分型平台

拥有芯片分型和质谱分型平台，可应用在疾病研究、药物筛选、动植物群体研究、分子育种等方面。



芯片分型平台-iScan



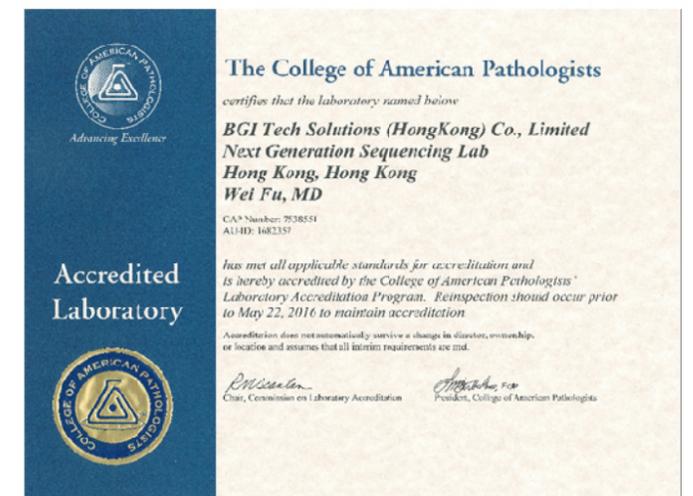
芯片分型平台-Affymetrix Genetian



质谱分型平台-MassARRAY

质量管理

华大科技于2010年通过了ISO9001质量管理体系标准认证；于2011年12月通过了ISO14001环境管理体系认证和ISO18001职业健康安全管理体系认证；2012年6月通过了ISO/IEC 27001信息安全管理体系认证；2015年7月华大基因香港高通量测序实验室获美国病理学会(College of American Pathologists, 以下简称“CAP”)颁发的CAP认可证书，成为中国首家在高通量测序服务和基因检测行业具有CAP认可资质的医学实验室。



基因组学

- 01 | 动植物 *de novo* 测序 
- 02 | 动植物全基因组重测序 
- 03 | 动植物简化基因组测序 

产品
简介

Product
Profile

动植物de novo测序

★ 品牌底蕴足

▶ 产品概述

从头测序即de novo测序,不需要任何参考序列资料即可对某个物种进行测序,用生物信息学分析方法进行拼接、组装,从而获得该物种的基因组序列图谱。利用全基因组从头测序技术,可以获得动植物的全基因组序列,带动这个物种下游一系列研究的开展,从而推进该物种的研究。全基因组序列图谱完成后,可以构建该物种的基因组数据库,为该物种的后基因组学研究搭建一个高效的平台,为后续的基因挖掘、功能验证提供DNA序列信息。华大基因结合短读长和长读长测序技术,可以高效、低成本地完成所有物种的基因组序列图谱。

▶ 产品优势

- 提取能力强**
提取成功率高达75%,针对很多疑难物种的优化提取获得了成功。
- 测序质量好**
已成功执行各种类型的物种的建库和测序,执行严格质控,保证测序质量优于行业标准。
- 分析经验多**
分析经验丰富,已成功组装多类物种,为项目的顺利交付保驾护航。
- 测序通量高**
2台PromethION/12台Sequel I/1台Sequel II测序仪保证超高测序通量,可实现快速交付。
- 品牌底蕴足**
华大基因已经成功完成1000多个物种的全基因组从头测序,合作发表顶级期刊文章160+,其中24篇为封面文章。

★ 提取能力强

对于长读长测序来说,得到一份完整度高、纯度高的DNA是测序成功最关键的一步。华大基因专门针对长片段提取进行研发优化,提取成功率高达75%,针对很多疑难物种的优化提取获得了成功。

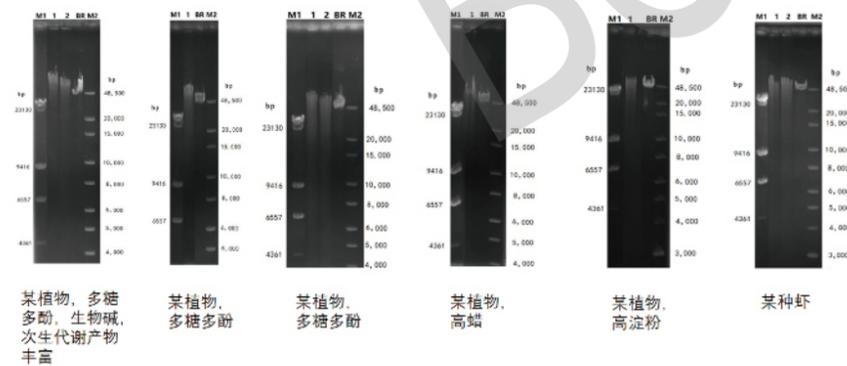


图1 疑难物种成功提取举例



图2 铁皮石斛提取成果

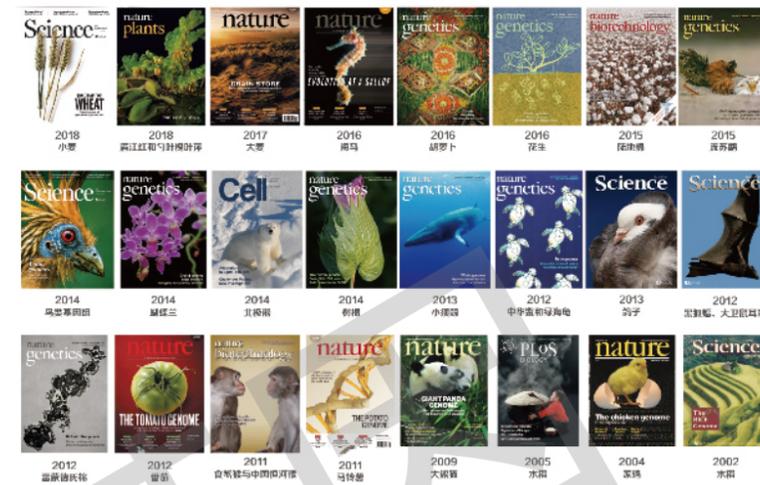


图3 动植物基因组封面文章集锦

▶ 产品应用



▶ 技术流程

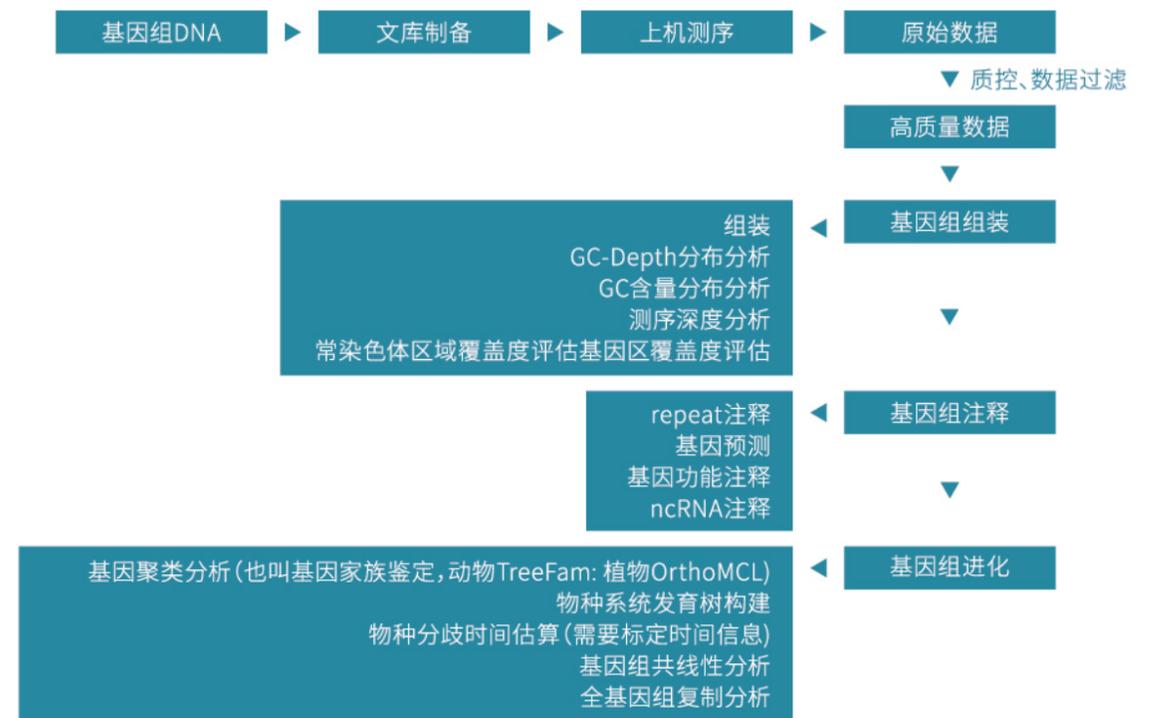


图4 动植物基因组de novo项目流程图

研究内容

基因组Survey分析

1. K-mer分析及基因组大小估算
2. 杂合率估算
3. 初步组装
4. GC-Depth分布分析

基因组组装

1. 组装
2. GC-Depth分布分析
3. GC含量分布分析
4. 测序深度分析
5. 常染色体区域覆盖度评估(需要客户提供BAC或者Fosmid序列)
6. 基因区覆盖度评估(需要客户提供EST或者转录组序列)

基因组注释

1. repeat注释
2. 基因预测
3. 基因功能注释
4. ncRNA 注释

进化分析

1. 基因聚类分析(也叫基因家族鉴定,动物TreeFam;植物OrthoMCL)
2. 物种系统发育树构建
3. 物种分歧时间估算(需要标定时间信息)
4. 基因组共线性分析
5. 全基因组复制分析(动物WGAC;植物WGD)

定制化信息分析

可结合客户的需求,协商确定定制化信息分析内容

样本要求

表1 各个平台DNA送样要求

测序平台	文库类型	样品类型	样品浓度	样本量
Nanopore	20Kb	基因组DNA	20ng/μl	10μg
PacBio	20Kb		20ng/μl	10μg
BGISEQ	350bp		20ng/μl	1.5μg
10X Genomics	—	基因组DNA (主带大于100kb,最小值大于50kb)	20ng/μl	500ng
BioNano	—	gDNA Agarose Plug (250Kb)/ 组织样本	—	4-6μg/Plug
Hi-C	—	完成甲醛交联的DNA/ 新鲜血液/活体组织	—	—

注:大片段文库不建议客户送DNA样本,建议直接送组织;组织样本需求量依据不同物种及组织样本类型而不同,如有需要请咨询当地销售。

测序策略

表2 动植物De novo各个平台的测序策略

测序平台	文库大小	测序读长	推荐测序深度	主要用途
Nanopore	20Kb及以上	10Kb以上	120 X	组装
PacBio	20Kb及以上	10Kb以上	100X	组装
BGISEQ	350bp/2-10Kb	PE150	50-100X	Survey/纠错
BioNano	250Kb	150Kb 以上	150 X	辅助组装/组装结果校正
Hi-C	350-450bp	PE100/PE150	100 X	Hi-C辅助组装定位
10X Genomics	350-450bp	PE100/PE150	60-100X	从头组装/辅助组装

注:上文提及的辅助组装手段可全选,也可根据物种特性及技术特点进行选择。

项目执行周期

标准执行周期 40-120个工作日,与项目的内容和基因组的复杂度有关。

案例分析

案例一:热带玉米基因组及高精度结构变异图谱成功构建,助力玉米遗传改良(华大参与)
Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement

杂志: Nature Genetics 发表时间: 2019.05 测序策略: PacBio + BioNano +10X Genomics+Hiseq

★ 研究结果

1. 热带玉米基因组图谱概貌

该项研究首先以一个热带小粒玉米品种(SK)为材料,应用PacBio测序技术、Bionano Genomics双酶切光学图谱、10X Genomics和Illumina测序数据,组装得到迄今为止质量最好的玉米参考基因组,大小为2.32Gb, contig N50达到15.78Mb,注释获得43,271个基因,为后续通过比较基因组学分析对热带玉米品系遗传多样性进行表征和对重要农艺性状相关基因进行定位提供了高质量的参考基因组和基因集。

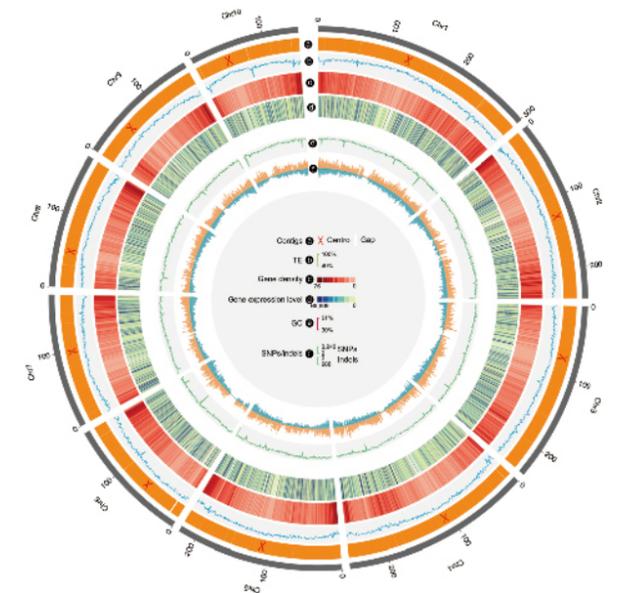


图5 热带玉米SK的参考基因组概貌

2. 结构变异图谱有什么妙用?

该项研究首先以一个热带小粒玉米品种(SK)为材料,应用PacBio测序技术、Bionano Genomics双酶切光学图谱、10X Genomics和Illumina测序数据,组装得到迄今为止质量最好的玉米参考基因组,大小为2.32Gb, contig N50达到15.78Mb,注释获得了43,271个基因,为后续通过比较基因组学分析对热带玉米品系遗传多样性进行表征和对重要农艺性状相关基因进行定位提供了高质量的参考基因组和基因集。

3. 玉米产量相关基因是如何被定位的?

玉米籽粒的重量与产量相关,是玉米驯化和改良过程中的关键选择性状之一。此前虽已针对该性状的控制位点开展了多项研究,但从未用正向遗传学的手段定位过影响该性状的基因。

在该项目中,研究人员利用ZHENG58和SK构建的重组自交系群体在玉米的1号染色体上定位到了一个同时控制粒型和粒重的位点(qHKW1),随后将其精细定位在长度为177Kb的基因组区间内,进一步定位出该位点所在的基因——ZmBAM1d,通过基因表达实验证实该基因正向调控玉米粒重,且在该基因过表达和敲除实验中均未检出对其他农艺性状的影响,表明该基因在玉米品质改良当中有应用前景,可用于提升作物产量。

研究人员进一步将SK和B73的ZmBAM1d基因区域进行比较,发现了与粒重表型直接相关的结构变异,由此说明,结构变异是表型差异的基础,也证明了该项研究中所构建的结构变异组图谱在农艺性状相关基因与位点定位当中的直接作用和未来应用前景。



图6 野生玉米大刍草、SK 以及现代玉米自交系ZHENG58的籽粒形态及百粒重

案例二: 鹅掌楸基因组及进化研究(华大参与)

Liriodendron genome sheds light on angiosperm phylogeny and species-pair differentiation

杂志: *Nature Plants* 发表时间: 2018.12 测序策略: PacBio 87X+ illumina 378X+ BioNano 180X

★ 研究结果

鹅掌楸(*Liriodendron chinense*)俗名“马褂木”,属国家二级珍稀保护树种,树形美观、材质优良、应用广泛,是我国重要的用材树种和园林景观。本研究首次完成了木兰类物种中国鹅掌楸的基因组组装、从全基因组水平解析被子植物的系统演化,确定了以鹅掌楸为代表的木兰类植物在被子植物中的演化地位。基于中国鹅掌楸与北美鹅掌楸的群体分析,解析了鹅掌楸属物种的群体多态性以及遗传结构。该成果对于深入了解珍稀物种鹅掌楸的进化和遗传,开展功能基因鉴定、重要性状的分子调控网络和基于基因组的选择育种等方面的研究,均具有十分重要的理论和应用价值。

木兰类植物是主要被子植物(Mesangiospermae),即核心被子植物中较早演化出来的一支,对于理解被子植物的演化具有非常重要的意义。本研究首次基于全基因组研究,从分子水平揭示了鹅掌楸为代表的木兰类植物形成于单、双子叶植物分化之前。这一研究结果为解决长期困扰学术界关于木兰类植物、双子叶植物与单子叶植物之间的演化关系的争议,提供了新的重要证据。

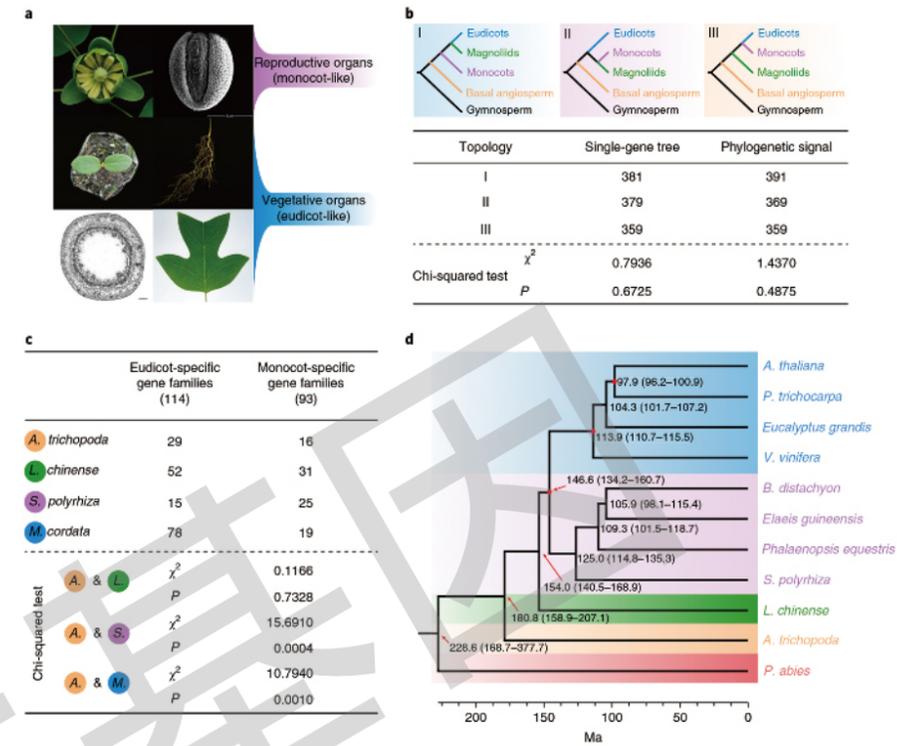


图7 鹅掌楸进化图

本研究还对东亚-东部北美洲间断分布的中国鹅掌楸和北美鹅掌楸的群体演化、种内多样性和种间遗传分化模式等进行了全基因组水平的研究。结果显示,与北美鹅掌楸相比分布于中国的鹅掌楸具有更高的遗传多样性。通过历史种群演化动态回溯研究发现,中国境内东西走向的高山峻岭聚集的复杂地形地貌,可能为第四纪冰川时期分布于中国的鹅掌楸提供了更有利的避难所生境,保存了较高的遗传多样性。本研究成果的发表,对于中国鹅掌楸野生种质资源保护、利用、开发及育种提供了重要依据。

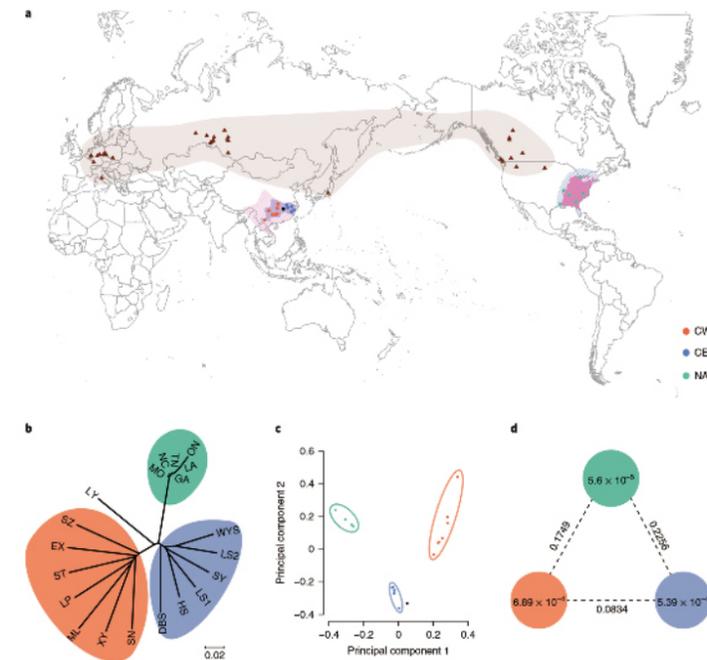


图8 鹅掌楸种群分布图

常见问题

Q: 怎么查询基因组的大小?

A: 查询植物基因组大小的网站: Cvalues database

查询动物基因组大小的网站: Animal genome size database

换算关系: 1pg=978Mb

Q: 如何判断简单基因组和复杂基因组?

A: 通常可以通过已测序的近缘种来判断物种复杂程度,如一般鸟类和哺乳类动物都是简单基因组,如果没有近缘物种供参考,可以先做基因组survey评估。

简单基因组及复杂基因组定义

基因组分类	普通基因组	复杂基因组
基因组大小	≤3.5G	>3.5G
染色体倍性	单倍体或纯合二倍体	杂合二倍体或多倍体
杂合率	<0.5%	≥0.5%
重复序列含量	<50%	≥50%
GC含量	35-65%	<35%或>65%
备注	满足所有条件为普通基因组	符合任一条件即为复杂基因组

Q: 基因组从头测序的组装结果好坏如何判断?

A: 一般用contig N50和scaffold N50来衡量基因组组装结果的好坏。N50是指把组装出的contigs或scaffolds从大到小排列,当其累计长度刚刚超过全部组装序列总长度50%时,最后一个contig或scaffold的大小即为N50的大小,N50对评价组装序列的连续性、完整性有重要意义: N70和N90的计算方法与N50类似,只是百分数变为70%或90%。

Q: PacBio测序的优势是什么?

A: 最大的优势是测序读长长,平均读长在12K以上,且无GC偏向性: 对基因组的组装、大的结构变异检测、转录组全长测序结果均有极大提升。

Q: Bionano 项目对物种有限制吗? 哪些物种可以做呢?

A: Bionano 项目只能针对已有初步组装结果的物种来进行辅助组装,因此每个项目进行之前需要对已有的组装序列进行前期信息分析评估并寻找合适的酶。基因组初步组装结果要求不能太碎,大于100kb的scaffold序列要占大部分,N的含量不能太高,基因组杂合率高的话,辅助组装时也有一定的影响: 酶的选择是利用软件LabelDensityCalculator模拟酶切初步组装结果,计算Label Density,范围在8-15/100kb即可。评估合格的物种可以进行Bionano测序。

华大发表文献

华大基因自1999年成立以来,发表文章总数超过一千篇。截至2019年6月,已经发表动植物de novo文章161篇。

华大与合作伙伴共同发表基因组文章列表(植物)

中文名	拉丁名	发表时间	刊物
玉米	<i>Zea mays</i>	2019.05	Nature Genetics
野生大豆	<i>Glycine soja</i>	2019.03	Nature Communications
芥菜	<i>Brassica nigra</i>	2019.01	Plant Molecular Biology
瑞丽植物园689个植物		2019.01	GigaScience
金鱼草	<i>Antirrhinum majus</i>	2019.01	Nature Plants
鹅掌楸	<i>Liriodendron chinense</i>	2018.12	Nature Plants

中文名	拉丁名	发表时间	刊物
5种孤立作物	班巴拉花生 (<i>Vigna subterranea</i>)、 扁豆 (<i>Lablab purpureus</i>)、 相思树 (<i>Faidherbia albida</i>)、 伯尔硬胡桃 (<i>Sclerocarya birrea</i>)、 辣木 (<i>Moringa oleifera</i>)	2018.12	GigaScience
六倍体小麦-中国春(封面文章)	<i>Triticum aestivum</i>	2018.08	Science
毛竹	<i>Phyllostachys edulis</i>	2018.08	GigaScience
单叶省藤和黄藤	<i>Calamus simplicifolius</i> 、 <i>Daemonorops jenkinsiana</i>	2018.08	GigaScience
满江红、勺叶槐叶萍	<i>Azolla filiculoides</i> 、 <i>Salvinia cucullata</i>	2018.07	Nature Plants
乌拉尔图小麦	<i>Triticum urartu</i>	2018.05	Nature
茶树	<i>Camellia sinensis</i>	2018.04	PNAS
小立碗藓	<i>Physcomitrella patens</i>	2018.02	The Plant Journal
芦笋	<i>Asparagus officinalis</i>	2017.11	Nature Communications
椰子	<i>Cocos nucifera</i>	2017.10	GigaScience
野生油橄榄	<i>Olea europaea</i>	2017.10	PNAS
珍珠粟(御谷)	<i>Cenchrus americanus</i>	2017.09	Nature Biotechnology
杜鹃花	<i>Rhododendron delavayi</i>	2017.08	GigaScience
油菜	<i>Brassica napus</i> cultivar 'ZS11'	2017.08	The Plant Journal
石榴	<i>Punica granatum</i>	2017.06	The Plant Journal
莴苣	<i>Lactuca sativa</i>	2017.04	Nature Communications
大麦(封面文章)	<i>Hordeum vulgare</i>	2017.04	Nature
大麦	<i>Hordeum vulgare</i>	2017.04	Scientific Data
龙眼	<i>Dimocarpus longan</i>	2017.03	GigaScience
土瓶草	<i>Cephalotus follicularis</i>	2017.02	Nature Ecology & Evolution
银杏	<i>Ginkgo biloba</i>	2016.11	GigaScience
窄叶羽扇豆	<i>Lupinus angustifolius</i>	2016.08	Plant Biotechnol J.
胡萝卜	<i>Daucus carota</i>	2016.05	Nature Genetics
矮牵牛	<i>Petunia hybrida</i>	2016.05	Nature Plants
花生(封面文章)	<i>Arachis duranensis</i> <i>Arachis ipaensis</i>	2016.02	Nature Genetics
铁皮石斛	<i>Dendrobium catenatum</i>	2016.01	Scientific Reports
甲藻	<i>Symbiodinium kawagutii</i>	2015.11	Science
小豆	<i>Vigna angularis</i>	2015.10	PNAS
茭白	<i>Zizania latifolia</i>	2015.08	The Plant Journal
铁皮石斛	<i>Dendrobium officinale</i>	2015.06	Molecular Plant
陆地棉(封面文章)	<i>Gossypium hirsutum</i> TM-1	2015.04	Nature Biotechnology
牛耳草	<i>Boea hygrometrica</i>	2015.04	PNAS
麻疯树	<i>Jatropha curcas</i>	2015.03	The Plant Journal
青稞	<i>Lasa Goumang</i>	2015.01	PNAS
蝴蝶兰	<i>Phalaenopsis equestris</i>	2014.11	Nature Genetics
枣	<i>Ziziphus jujuba</i> Mill.	2014.10	Nature Communications
甘蓝型油菜	<i>Brassica napus</i>	2014.08	Science
野生大豆	<i>Glycine max</i>	2014.07	Nature Communications
树棉(亚洲棉,棉花A)	<i>Gossypium arboreum</i>	2014.05	Nature Genetics
甘蓝	<i>Brassica oleracea</i>	2014.05	Nature Communications
辣椒	<i>Capsicum annuum</i> <i>Capsicum annuum</i> var <i>glabriusculum</i>	2014.03	PNAS

中文名	拉丁名	发表时间	刊物
芝麻	<i>Sesamum indicum</i>	2014.03	Genome Biology
胡杨	<i>Populus euphratica</i>	2013.11	Nature Communications
桑树	<i>Morus notabilis</i>	2013.09	Nature Communications
醉蝶花	<i>Tarenaya hassleriana</i>	2013.08	Plant Cell
莲	<i>Nelumbo nucifera</i>	2013.08	Plant Journal
短花药野生稻	<i>Oryza brachyantha</i>	2013.03	Nature Communications
小麦A(野生一粒小麦)	<i>Triticum urartu</i>	2013.03	Nature
小麦D(节节麦、粗山羊草)	<i>Aegilops tauschii</i>	2013.03	Nature
鹰嘴豆	<i>Cicer arietinum</i>	2013.01	Nature biotechnology
梅花	<i>Prunus mume</i>	2012.12	Nature Communications
梨	<i>Pyrus bretschneideri</i>	2012.11	Genome Research
西瓜	<i>Citrullus lanatus</i>	2012.11	Nature Genetics
雷蒙德氏棉/亚洲棉	<i>Gossypium raimondii</i>	2012.08	Nature Genetics
亚麻	<i>Linum usitatissimum</i>	2012.07	Plant Journal
盐芥	<i>Thellungiella salsuginea</i>	2012.07	PNAS
谷子	<i>Setaria italica</i>	2012.05	Nature biotechnology
番茄	<i>Solanum lycopersicum</i>	2012.05	Nature
木豆	<i>Cajanus cajan</i>	2011.11	Nature Biotechnology
白菜	<i>Brassica rapa</i>	2011.08	Nature Genetics
马铃薯	<i>Solanum tuberosum</i>	2011.07	Nature
黄瓜	<i>Cucumis sativus</i>	2009.11	Nature Genetics
水稻(9311)	<i>Oryza sativa</i> L. ssp. <i>indica</i>	2002.04	Science

华大与合作伙伴共同发表基因组文章列表(动物)

中文名	拉丁名	发表时间	刊物
抹香鲸	<i>Physeter macrocephalus</i>	2019.02	Molecular Ecology Resources
泰国斗鱼	<i>Betta splendens</i>	2018.11	GigaScience
赤狐	<i>Vulpes Vulpes</i>	2018.08	Nature Ecology & Evolution
日本鹌鹑	<i>Coturnix japonica</i>	2018.05	GigaScience
紫扇贝	<i>Argopecten purpuratus</i>	2018.04	GigaScience
江豚	<i>Neophocaena asiaeorientalis</i>	2018.04	Nature Communications
德国小蠊、干材白蚁	<i>Blattella germanica</i> , <i>Cryptotermes secundus</i>	2018.02	Nature Ecology & Evolution
普通吸血蝙蝠	<i>Desmodus rotundus</i>	2018.02	Nature Ecology & Evolution
斜纹夜蛾	<i>Spodoptera litura</i>	2017.09	Nature Ecology & Evolution
沙鼠	<i>Psammomys obesus</i>	2017.07	PNAS
欧洲野牛	<i>Bison bonasus</i>	2017.03	GigaScience
甘薯粉虱	<i>Bemisia tabaci</i>	2017.03	GigaScience
乌鳢	<i>Channa argus</i>	2017.03	GigaScience
大银鱼	<i>Protosalanx hyalocranius</i>	2017.02	GigaScience
线纹海马	<i>Hippocampus erectus</i>	2017.02	GigaScience
牙鲆	<i>Paralichthys olivaceus</i>	2016.12	Nature Genetics
海马(封面文章)	<i>Hippocampus comes</i>	2016.12	Nature
五步蛇	<i>Deinagkistrodon acutus</i>	2016.10	Nature Communications
翻车鱼	<i>Mola mola</i>	2016.09	GigaScience
斑点叉尾鮰	<i>Ictalurus punctatus</i>	2016.08	GigaScience

中文名	拉丁名	发表时间	刊物
中华绒螯蟹	<i>Eriocheir sinensis</i>	2016.06	GigaScience
太平洋鲑鱼	<i>Salmo salar</i>	2016.04	Nature
滇池金线鲃、犀角金线鲃、安水金线鲃	<i>Sinocyclocheilus grahami</i> , <i>Sinocyclocheilus rhinoceros</i> , <i>Sinocyclocheilus anshuiensis</i>	2016.01	BMC Biology
非洲猎豹	<i>Acinonyx jubatus</i>	2015.12	Genome Biology
流苏鹬(封面文章)	<i>Philomachus pugnax</i>	2015.11	Nature Genetics
壁虎	<i>Gekko japonicus</i>	2015.11	Nature Communications
亚洲虎蚊	<i>Aedes albopictus</i>	2015.10	PNAS
鹅	<i>Anser cygnoides orientalis</i>	2015.05	Genome Biology
10种蜂类(5种为新测)	<i>Apis mellifera</i> 等	2015.05	Science
大黄鱼	<i>Larimichthys crocea</i>	2015.04	PLOS Genetics
高山倭蛙	<i>Nanorana parkeri</i>	2015.03	PNAS
犬弓蛔虫	<i>Toxocara canis</i>	2015.02	Nature Communications
弹涂鱼	<i>B. pectinirostris</i> , <i>S. histophorus</i> , <i>Periophthalmodon schlosseri</i> , <i>Periophthalmus magnuspinnatus</i>	2014.12	Nature Communications
企鹅	<i>Aptenodytes forsteri</i> , <i>Pygoscelis adellae</i>	2014.12	GigaScience
朱鹮	<i>Nipponia nippon</i>	2014.12	Genome Biology
鳄鱼	<i>Alligator mississippiensis</i> , <i>Crocodylus porosus</i> , <i>Gavialis gangeticus</i>	2014.12	Science
48种鸟类(45种为新测)	<i>Acanthisitta chloris</i> 等	2014.12	Science
稻飞虱	<i>Nilaparvata lugens</i>	2014.12	Genome Biology
双峰驼、单峰驼、羊驼	<i>Camelus bactrianus</i> , <i>Camelus dromedarius</i> , <i>Vicugna pacos</i>	2014.10	Nature Communications
达马拉兰鼯鼠	<i>Fukomys damarensis</i>	2014.08	Cell Reports
泰国肝吸虫	<i>Opisthorchis viverrini</i>	2014.07	Nature Communications
虎皮鹦鹉	<i>Melopsittacus undulatus</i>	2014.07	GigaScience
绵羊	<i>Ovis aries</i>	2014.06	Science
盲鼯形鼠	<i>Spalax galili</i>	2014.06	Nature Communications
鞭虫	<i>Trichuris</i>	2014.06	Nature Genetics
丝绒蜘蛛、白膝头蜘蛛	<i>Stegodyphus mimosarum</i> , <i>Acanthoscurria geniculata</i>	2014.05	Nature Communications
北极熊	<i>Ursus maritimus</i>	2014.05	Cell
湿木白蚁	<i>Zootermopsis nevadensis nuttingi</i>	2014.05	Nature Communications
榕小蜂	<i>Ceratolen solmsi</i>	2014.04	Genome Biology
隧蜂	<i>Lasioglossum albipes</i>	2014.03	Genome Biology
半滑舌鳎	<i>Cynoglossus semilaevis</i>	2014.02	Nature Genetics
行军蚁	<i>Cerapachys biroi</i>	2014.02	Current Biology
飞蝗	<i>Locusta migratoria</i>	2014.01	Nature Communications
小须鲸	<i>Balaenoptera acutorostrata</i>	2013.11	Nature Genetics
白鬻豚	<i>Lipotes vexillifer</i>	2013.10	Nature Communications
东北虎	<i>Panthera tigris</i>	2013.09	Nature Communications
扬子鳄	<i>Alligator sinensis</i>	2013.08	Cell Research
布氏鼠耳蝠	<i>Myotis brandtii</i>	2013.08	Nature Communications

中文名	拉丁名	发表时间	刊物
北京鸭	<i>Anas platyrhynchos</i>	2013.06	Nature Genetics
古马	<i>Equus caballus</i>	2013.06	Nature
藏羚羊	<i>Pantholops hodgsonii</i>	2013.05	Nature Communications
中华鳖和绿海龟	<i>Pelodiscus sinensis ; Chelonia mydas</i>	2013.04	Nature Genetics
游隼、猎隼	<i>Falco peregrinus; Falco cherrug</i>	2013.03	Nature
地山雀	<i>Pseudopodoces humilis</i>	2013.03	Genome Biology
树鼩	<i>Tupaia belangeri</i>	2013.02	Nature Communications
小菜蛾	<i>Plutella xylostella</i>	2013.01	Nature Genetics
鸽子	<i>Columba livia</i>	2013.01	Science
黑狐蝠、大卫鼠耳蝠	<i>Pteropus alecto, Myotis davidii</i>	2012.12	Science
家山羊	<i>Capra hircus</i>	2012.12	Nature Biotechnology
家猪	<i>Sus scrofa</i>	2012.11	Nature
五指山猪	<i>Sus scrofa</i>	2012.11	GigaScience
牡蛎	<i>Crassostrea gigas</i>	2012.09	Nature
勇地雀	<i>Geospiza fortis</i>	2012.08	GigaScience
牦牛	<i>Bos grunniens</i>	2012.07	Nature Genetics
埃及血吸虫	<i>Schistosoma haematobium</i>	2012.01	Nature Genetics
裸鼯鼠	<i>Heterocephalus glaber</i>	2011.10	Nature
食蟹猴和中国恒河猴	<i>Macaca fascicularis, Macaca mulatta lasiota</i>	2011.10	Nature Biotechnology
猪蛔虫	<i>Ascaris suum</i>	2011.10	Nature
顶切叶蚁	<i>Acromyrmex echinatior</i>	2011.06	Genome Research
蚂蚁	<i>Camponotus floridanus , Harpegnathos saltator</i>	2010.08	Science
熊猫	<i>Alluopoda melanoleura</i>	2009.12	Nature
家蚕	<i>Bombyx mori</i>	2004.12	Science
家鸡	<i>Gallus sonneratii</i>	2004.12	Nature



动植物重测序

▶ 产品概述

全基因组重测序是对已知基因组序列的物种进行DNA测序,并在此基础上完成个体或群体分析。全基因组重测序通过序列比对,可以检测到大量变异信息,包括单核苷酸多态性位点(SNP),插入缺失位点(InDel, Insertion/Deletion)、结构变异位点(SV, Structure Variation)位点,拷贝数变异(CNV, Copy Number Variation)位点等,获得同一物种不同个体的遗传变异图谱。基于检测到的变异能进一步的阐述该物种特有的生物学信息。

随着测序成本降低和已知基因组序列物种的增多,全基因组重测序已经成为动植物分子育种、群体进化中最为迅速有效的方法之一。利用全基因组重测序技术有助于快速发现与动植物重要性状相关的遗传变异,应用于分子育种中,缩短育种周期。

▶ 产品优势



数据精准

华大至今完成10万+的动植物重测序样本,严格质量控制流程保证结果准确度。



经验丰富

动植物重测序领域挂名发表文章90余篇,累计影响因子700+,其中一作或通讯作者文章50+,涵盖变异检测、遗传图谱构建&QTL定位、群体进化和GWAS等各研究领域。



项目方案支持

大项目参与方案设计,使项目赢在起跑线。



分析团队实力雄厚

发表影响因子10分以上动植物研究文章的人员20+。



个性化分析

具有丰富个性化分析经验,可根据项目需要选择最适宜的分析软件,只为保障最精准结果。



自主测序平台、成本可控



测序数据duplicates rate低

有效数据大幅提升。



测序数据无Index Hopping担忧

DNBSEQ™技术用滚环扩增(Rolling circle amplification, RCA),只以最初的单链环状文库为模板进行线性扩增,不会有PCR过程中错误的累积,所以由Index hopping带来的错误序列不会累积。



BGISEQ平台PCR-free重测序检测InDel更精准

建库和测序过程“0”PCR,避免引入错误InDel。

研究内容

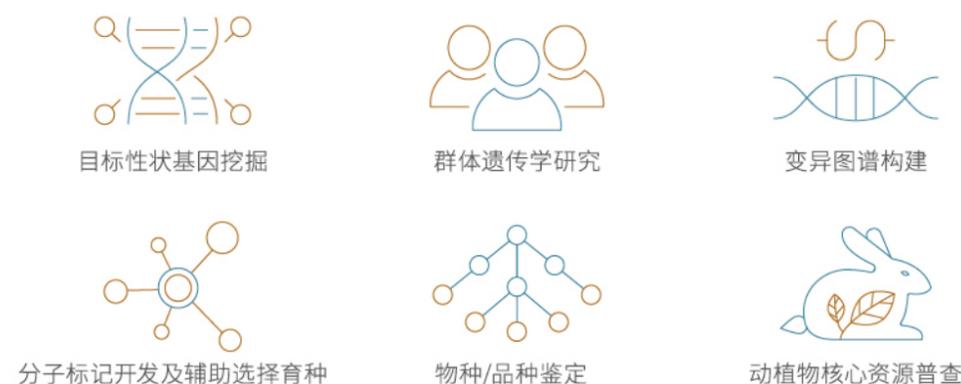
信息分析条款	信息分析内容
标准信息分析	测序数据基本分析 比对 产出数据统计 一致性序列生成 SNP检测、注释及统计
变异检测	InDel检测、注释及统计 SV检测、注释及统计 CNV检测、注释及统计
群体进化	连锁不平衡分析(LD) 群体进化树分析(Phylogeny tree) 群体结构分析(Structure) 群体主成分分析(PCA) 多态性分析
点突变检测	点突变位点检测、注释和统计
BSA分析	子代混合样本SNP-index值计算 筛选两个DNA pool中等位基因频率差异显著的区域 候选基因检测、注释和统计
遗传图谱构建及QTL定位	遗传图谱/Bin Map构建 QTL定位(需提供表型数据) 遗传图谱整合(需提供同一作图群体原图谱数据)
HapMap构建	连锁不平衡分析(LD) 群体进化树分析(Phylogeny tree) 群体结构分析(Structure) 群体主成分分析(PCA) HapMap构建、Tag SNP筛选
全基因组关联分析(GWAS)	连锁不平衡分析(LD) 群体进化树分析(Phylogeny tree) 群体结构分析(Structure) 群体主成分分析(PCA) 基因型缺失数据处理(Imputation) GWAS分析 筛选与性状相关的候选基因 候选基因功能注释
定制化信息分析	可结合客户的需求, 协商确定定制化信息分析内容

高级信息分析

技术流程



产品应用



技术参数

样品要求: 基因组DNA, 无降解或轻微降解;
 样品需求量(单次): $\geq 1 \mu\text{g}$;
 样品浓度: $\geq 12.5 \text{ ng}/\mu\text{L}$;
 推荐数据量: 个体重测序推荐20X以上, 群体重测序推荐5X以上;
 交付周期: 具体项目的交付周期与样本数和信息分析内容有关。

案例二：鹰嘴豆群体进化和GWAS研究(华大参与)

案例一：3K水稻重测序&泛基因组研究(华大参与)

由中国农业科学院作物科学研究所牵头,联合IRRI、上海交大、华大基因、深圳农业基因组研究所、安徽农大等16家单位共同完成“3000份亚洲栽培稻基因组研究”,并于2018年4月发表在《Nature》上。研究针对水稻起源、分类和驯化规律进行了深入探讨,揭示了亚洲栽培稻的起源和群体基因组变异结构,剖析了水稻核心种质资源的基因组遗传多样性。

3000份水稻(来自全球89个国家和地区)代表了全球78万份水稻种质约95%多样性的核心种质。通过全基因组重测序,每个样本平均测序深度14X,利用重测序数据共检测到32M的高质量SNPs和InDels。对亚洲栽培稻群体的结构和分化进行了更为细致和准确的描述和划分,由传统的5个群体增加到9个,分别是东亚(中国)的籼稻、南亚的籼稻、东南亚的籼稻和现代籼稻品种等4个籼稻群体,东南亚的温带粳稻、热带粳稻、亚热带粳稻等3个粳稻群体,以及来自印度和孟加拉的Aus和香稻。研究首次揭示了亚洲栽培稻品种间存在的大量微细结构(>100bp)变异(SVs,包括易位、缺失、倒位和重复)。着重研究453个测序深度>20X的品系的SVs,利用SVs构建的进化树与SNP构建的进化树类似。大量的SVs可能是不同程度杂种不育和XI与GJ杂种衰退的遗传基础。同时构建了亚洲栽培稻的泛基因组,包括12,770个(62.1%)核心(core)基因家族和9,050个(37.9%)分散式(distributed)基因家族。发现了1.2万个全长新基因和数千个不完整的新基因。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。

研究策略

最初测序3024份水稻样本,后来进行质控过滤掉14份,最终保留3010份水稻样本进行深度研究。3K RG测序数据比对到参考基因组日本晴Nipponbare上检测SNPs、indels。合并Nipponbare基因组序列和无冗余的新组装的基因组序列构建泛基因组。利用测序深度>20X,比对深度>15X的453个水稻材料进行SVs和PAVs分析。

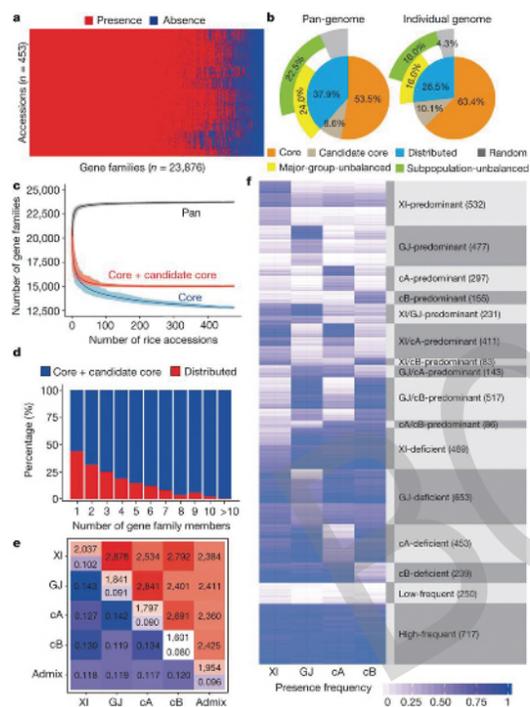


图1 水稻泛基因组
a. 基因家族PAVs
b. 泛基因组和一个单独的基因组的组成成份
c. 基于500个随机筛选的水稻基因组模拟泛基因组和核心基因组
d. 核心和分散式基因家族比例
e. 两个品系间基因家族平均数量差异
f. 5733主要群组不平衡基因家族特性

参考文献

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.

案例二：鹰嘴豆群体进化和GWAS研究(华大参与)

由国际作物干旱热带作物研究所(ICRISAT)与华大基因主导,来自全球范围的21个研究机构组成的科学家团队共同协作,对来自45个国家的429种鹰嘴豆制品系全基因组重测序进行分析,并于2019年4月发表于《Nature Genetics》。研究明确鉴定了该物种耐旱和耐热的功能基因,为培育高产与抗逆的鹰嘴豆提供了宝贵的遗传资源。

研究人员应用429个鹰嘴豆品种4.97M的SNPs构建了进化树,7个野生种作为外群,栽培种形成4个亚群。进化树的拓扑结构与品种生物学地位和种子的2种主要类型均无关,这在作物群体研究当中比较少见,表明鹰嘴豆驯化路径较为复杂。研究通过对来自于不同地理环境的鹰嘴豆样本的群体差异指数和群体多样性进行分析,揭示鹰嘴豆最初起源于地中海东部,经地中海/新月沃地到达中亚,之后可能从中亚同时到达东非(埃塞俄比亚)和南亚(印度)。分析数据也支持埃塞俄比亚是鹰嘴豆的第二起源中心。并且在鹰嘴豆的驯化过程中,有近80%的遗传多样性丢失,利用多样性衰减(ROD)和群体差异指数分析,得到基因组上受选择的候选区间122个,其中包含受选择的基因204个,后经注释发现,受选择的基因主要和胁迫应答、DNA修复、蛋白酶激活、种子发育、发芽、花发育等生理过程相关。研究同时针对272份鹰嘴豆样本,应用GLM/CMLM、FarmCPU和EMMAX模型,对3.65M的SNPs与来自多年多点(1-6个季节,1-6个地点)的20个与抗旱、耐热相关性状进行全基因组关联分析(GWAS),定位得到与13个重要性状相关的262个标记和候选基因,并发现地方种ICC14778具有稳定的耐热、抗旱性状,能用于培育抗性品种。

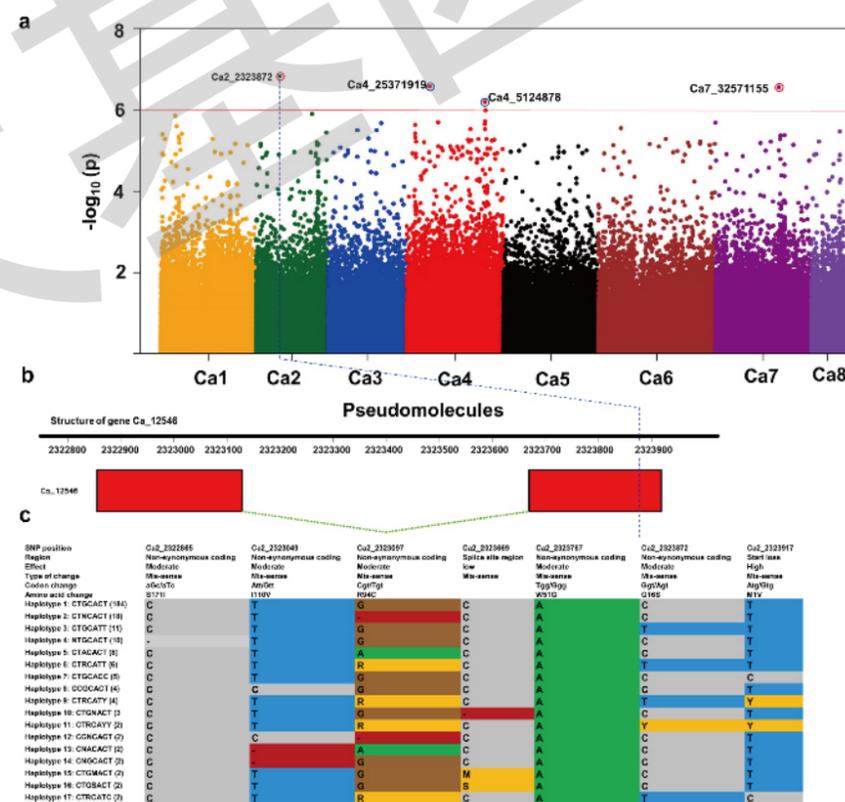


图2 高温胁迫下产量的GWAS分析结果

- SNP位点Ca2_2323872与旱作条件下的产量有关联,候选基因Ca_12546预测和AB抗性相关
- 基因Ca_12546结构信息
- 基因Ca_12546的17个单倍型

参考文献

Varshney R K, Thudi M, Roorkiwal M, et al. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits[J]. Nature genetics, 2019, 51(5): 857.

案例三: QTL定位-超级杂交稻的产量QTL定位(华大参与)

研究人员首次组装成功具有籼稻和热带粳稻血缘的“两优培九”母本品种培矮64S的基因组序列,并基于高密度 Bin Map图谱,不仅更新了“两优培九”父本品种93-11的基因组序列,还结合12个水稻产量性状的考查共检测到43个与水稻产量密切相关的数量性状位点(QTL),其中包括20个尚未报道的新QTL。同时,利用大规模重组自交系群体,结合亲本基因组间大量精确的单核苷酸多态标记(SNP)和插入缺失标记(InDel)对两个控制穗粒数和二次枝梗数的QTL-qSN8和qSPB1分别加以精细定位,最终确定控制抽穗期和穗粒数的DTH8基因和控制穗分支的LAX1基因为候选基因,其中qSN8通过互补实验证实为DTH8基因。这一大规模重组自交系群体与超高分辨率遗传图谱相结合,其一系列新的产量相关基因将会被陆续克隆和开展功能研究,从而为进一步阐明超级稻产量的遗传基础和基因辅助育种搭建了理想平台。

★ 研究策略

利用全基因组重测序进行基因分型,重组自交系亲本93-11(36X)和PA64s(48X)都已经进行了基因组组装,对132个超级杂交稻“两优培九”的核心重组自交系每个样本测序~4X,构建的图谱不仅用于基因组序列的改善还可以进行QTL定位。

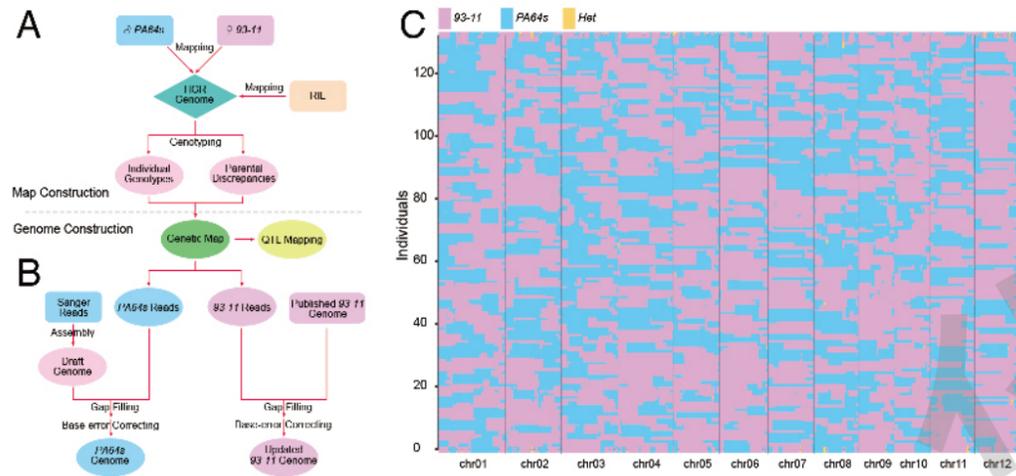


图3 项目研究流程示意图及Bin Map图

★ 参考文献

Gao Z Y, Zhao S C, He W M, et al. Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences [J]. Proceedings of the National Academy of Sciences, 2013, 110(35): 14492-14497.

▶▶ 常见问题

Q: 进行全基因组重测序数据推荐?

A: 每个样本推荐的数据量与样本类型和要做的信息分析内容相关。例如关注个体样本的SNP,对SNP的准确度和覆盖度要求比较高,一般推荐测序深度>30X,对于稀有变异测序深度还要进一步提高;用于研究群体结构的样本,测序深度推荐5X以上,随着测序成本的不断下降,现在越来越多的项目测序深度在10X以上;纯合样本混样检测等位基因频率,推荐平均每个样本的测序深度在1X以上,混合样本测序深度不低于30X;DH和RIL群体构建Bin Map,子代群体测序深度可以测序0.5X/样本。

Q: 样本量选择多大合适?

A: 样本量大小与样本类型和研究目的相关。例如进行群体进化研究推荐30个样本以上,因为从统计学上说30个以上才属于大样本;对于进行基因挖掘的项目来说,无论是利用自然群体进行GWAS分析或是用家系群体进行连锁分析,都是群体越大越好,一般的情况下进行GWAS分析的样本推荐300个样本以上,对于家系群体推荐200个以上。

Q: 连锁图谱构建适用于什么样的群体?

A: 连锁图谱的构建适用于作图群体,它是由性状差异显著的亲本杂交衍生的群体。亲本选择的要求:要考虑亲本间的遗传多态性、目标性状差异、亲本的纯合度和杂交后代的可育性。构建分离群体类型,根据遗传稳定性可将分离群体分成两大类:暂时性分离群体如F₁、F₂、BC等,永久性分离群体如RIL、DH等。

Q: 现在我们重测序数据比对回参考基因组检测SNP用的软件是什么?

A: 现在常用的比对和call SNP的软件包括SOAP+SOAPSnp和BWA+GATK,可以根据需求进行选择。

Q: 重测序reads与参考基因组比对率低,可能的原因是什么?

A: 重测序reads比对率低原因可能是:

1) 因为测序样本与参考基因组亲缘关系比较远。因为动植物品种多样,但是目前已完成基因组组装的往往只是其中的一个品种,同一个物种野生种与驯化种差异还是很大的; 2) 可能因为DNA不纯,存在其他物种的污染; 3) 参考基因组序列组装质量较差,引起比对率低; 4) 比对参数设置严格等。

▶▶ 华大发表文献

华大参与发表文章(一作或通讯作者,2017年至今)

发表日期	期刊	影响因子	物种	研究目的	研究内容
2019.04	Nature Genetics	27.125	鹰嘴豆	群体进化&GWAS	全基因组重测序鹰嘴豆揭示基因组多态性、进化和后进化候选基因
2018.12	Nature Communications	12.124	青稞	进化	青稞的起源与进化
2018.06	Genome Biology	10.81	梨	群体进化,地理起源	亚洲梨和欧洲梨的多样性与独立进化
2018.05	GigaScience	6.871	鹌鹑	基因组+进化	群体基因组数据揭示与鹌鹑重要性状相关的基因
2018.04	Nature Communications	12.124	江豚	基因组+群体进化	鲸类对淡水环境的适应性进化机制
2018.04	Nature	41.577	水稻	群体进化&泛基因组	揭示亚洲栽培稻的起源和群体基因组变异结构,剖析了水稻核心种质资源的基因组遗传多样性
2018.04	Nature Communications	12.124	梅花	群体进化+泛基因组+GWAS	木本植物梅花中花性状的遗传结构
2017.12	Molecular Biology and Evolution	6.202	牛	群体进化	中国牛遗传结构和进化
2017.11	Scientific Reports	4.259	绒山羊	进化	绒山羊羊绒纤维性状进化
2017.06	GigaScience	6.871	古老大型犬科动物	平台比较	BGISEQ-500与HiSeq2500平台比较
2017.02	GigaScience	7.463	小米	Bin Map+QTL 定位	RIL群体构建bin map辅助基因组组装指标提升并定位9个农艺性状QTL

动植物简化基因组测序

产品概述

简化基因组测序 (RRGS, Reduced-Representation Genome Sequencing), 通过对酶切获得的酶切片段进行高通量测序, 能够降低基因组的复杂度, 操作简便, 同时不受参考基因组的限制, 可快速鉴定出高密度的SNPs, 用于遗传图谱构建、群体进化、QTL定位, 辅助scaffold组装到染色体等。

产品优势

- 适用范围广**
有无参考基因组, 二倍体或多倍体均适用。
- 精确度高**
每个酶切tag的平均覆盖深度至少10X, 保证分子标记的高准确性。
- 性价比高**
基于酶切简化的基因组DNA, 数据量低, 特别适合大样本量研究。
- 快速高效**
只需2-3个月即可完成大样本的群体进化研究。

研究内容

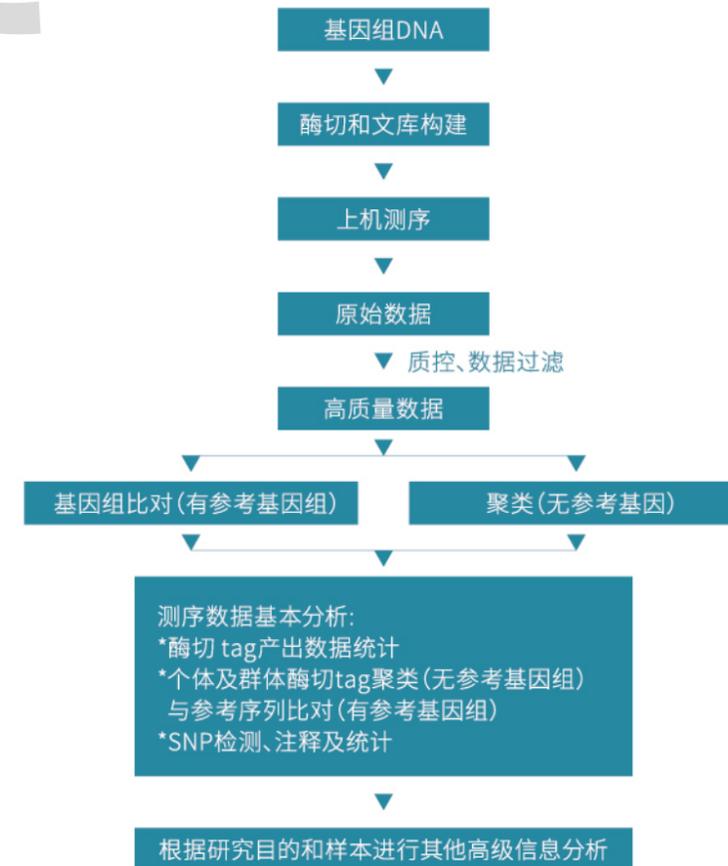
信息分析条款	信息分析内容
标准信息分析	酶切 tag 产出数据统计 个体及群体酶切tag聚类 (无参考基因组) 与参考序列比对 (有参考基因组) SNP检测、注释及统计 SNP分型
高级信息分析	群体进化 群体进化树分析 (Phylogeny tree) 群体结构分析 (Structure) 群体主成分分析 (PCA)
	作图群体极端表型样本BSA分析 子代混合样本 SNP-index 值计算 筛选两个 DNA pool 中等位基因频率差异显著的区域 候选基因检测、注释和统计
	遗传图谱构建及 QTL 定位 遗传图谱/Bin Map 构建 QTL 定位 (需提供表型数据) 遗传图谱整合 (需提供同一作图群体原图谱数据)

高级信息分析	全基因组关联分析 (GWAS)	连锁不平衡分析 (LD) 群体进化树分析 (Phylogeny tree) 群体结构分析 (Structure) 群体主成分分析 (PCA) 基因型缺失数据处理 (Imputation) GWAS分析 筛选与性状相关的候选基因 候选基因功能注释
	定制化信息分析	可结合客户的需求, 协商确定定制化信息分析内容

产品应用



技术流程



▶ 技术参数

样品要求: 基因组DNA, 无降解或轻微降解, 无或微量的RNA污染;
 样品需求量(单次): $\geq 1 \mu\text{g}$;
 样品浓度: $\geq 20 \text{ ng}/\mu\text{L}$;
 推荐数据量: 个体重测序酶切tag 20X以上, 群体重测序酶切tag推荐10X以上;
 交付周期: 具体项目的交付周期与样本数和信息分析内容有关。

▶ 案例分析

案例一: 赤小豆遗传图谱构建及QTL定位(华大参与)

赤小豆是重要的豆科作物($2n=2X=22$), 其低脂肪且富含易于消化的蛋白。中国在1.2万年前就已经开始驯化种植赤小豆。本研究利用简化基因组RAD测序进行基因分型, 利用作图群体定位赤小豆重要农艺性状的QTL。利用栽培种Ass001(红色种皮、淡黄色豆荚、绿茎、直立茎秆、种子大粒)和野生的小豆CWA108(黑色种皮、黑色豆荚、紫茎、缠绕茎、种子小粒)杂交构建 F_2 群体, 群体大小150, 针对此群体及其衍生的 $F_{2,3}$ 、 $F_{3,4}$ 群体对24个数量性状和4个质量性状进行表型鉴定, 利用多年的数据进行QTL定位。利用1571个标记构建了高密度遗传图谱, 共定位了26个QTL, 5个质量性状相关基因。其中包括与开花性状和豆荚成熟性状相关的11个QTL, 百粒重性状相关的6个QTL, 同时豆荚数目、种子厚度、主茎分枝数、最大叶宽、茎节间长度也分别有QTL定位结果。筛选与开花期和豆荚成熟相关的候选基因, 利用拟南芥、赤小豆、大豆、鹰嘴豆等数据库信息进行候选基因的鉴定并进行进化分析。研究发现在赤小豆、菜豆、绿豆、鹰嘴豆、木豆和蔓花生中有四个PHY位点包括两个PHYA、一个PHYB和一个PHYE, 但是在蒺藜苜蓿中只有一个PHYA。这个结果证明了在大豆中发生了两次全基因组复制事件, 导致了大豆中有4个PHYA和两个PHYE拷贝。

★ 研究策略

数量性状受环境影响比较大, 利用多年数据特别是 $F_{2,3}$ 、 $F_{3,4}$ 群体类型的多个样本表型数据计算均值, 代表 F_2 样本表型, 进行QTL定位, 降低表型误差。QTL定位结果与以往近缘物种研究结果相互验证, 不仅能帮助验证定位结果的准确性, 筛选候选基因位点, 同时可以进行例如进化研究等其他有意思的研究内容。

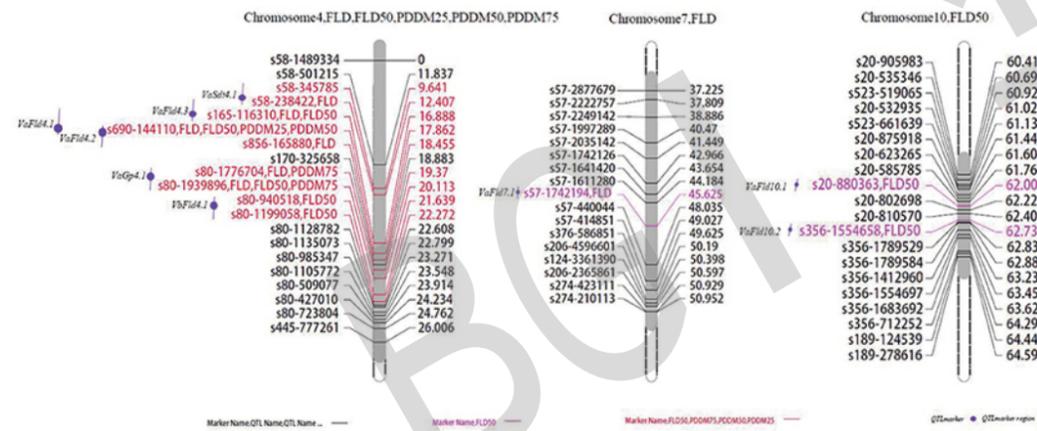


图1 开花和成熟性状QTL定位结果展示
 FLD, 第一次开花时间 FLD50, 50%开花时间 PDDM25, 25% 豆荚成熟
 PDDM50, 50% 豆荚成熟 PDDM75, 75%豆荚成熟

★ 参考文献

Li Y, Yang K, Yang W, et al. Identification of QTL and Qualitative Trait Loci for Agronomic Traits Using SNP Markers in the Adzuki Bean[J]. Frontiers in Plant Science, 2017, 8: 840.

案例二: 武昌鱼遗传图谱构建及QTL定位

武昌鱼主要分布于长江的中下游, 是养殖的重要水产品种。本研究利用SNP高密度标记构建遗传图谱定位重要经济性状相关位点, 促进武昌鱼分子育种进程。本研究利用简化基因组测序技术对两个高杂合武昌鱼亲本和187个 F_1 子代样本进行基因型鉴定, 母本和父本分布获得13.68M和12.62M的RAD tags, 子代测序1.14M-6.62M RAD tags。亲本测序样本组装获得367,640 contigs, 平均长度为374 bp, 双亲间检测多态性SNPs, 子代聚类检测SNP并进行基因分型。最终整合了双亲信息的遗传图谱为24个连锁群, 包含了14648个高可信的SNP, 图谱长度3258.38 cM, 5676个有效位点平均间距0.57 cM。对重要的经济性状进行QTL定位, 针对体长、高度、重量、生殖腺重量、生殖腺发育期和性别检测到8个正向QTL, 但是没有一个QTL能解释 $>10\%$ 的表型变异。将武昌鱼的高密度SNP遗传图谱与斑马鱼基因组序列进行比较研究, 176个作图标记比对到斑马鱼基因组。武昌鱼 $2n=48$ 、斑马鱼 $2n=50$, 武昌鱼24对染色体与斑马鱼具有共线性, 由于染色体数目不对等, 武昌鱼的第四条染色体比对到斑马鱼2条染色体上。

★ 研究策略

武昌鱼 F_1 群体, RAD测序构建高密度连锁图谱。没有参考基因组物种进行简化基因组测序, 可以通过组装再检测SNP, 可以提高标记数量。无参考基因组物种构建遗传图谱与已知近缘物种基因组序列比对, 染色体间存在共线性关系, 那么其他遗传学研究也可以参考已知基因组及其注释信息。

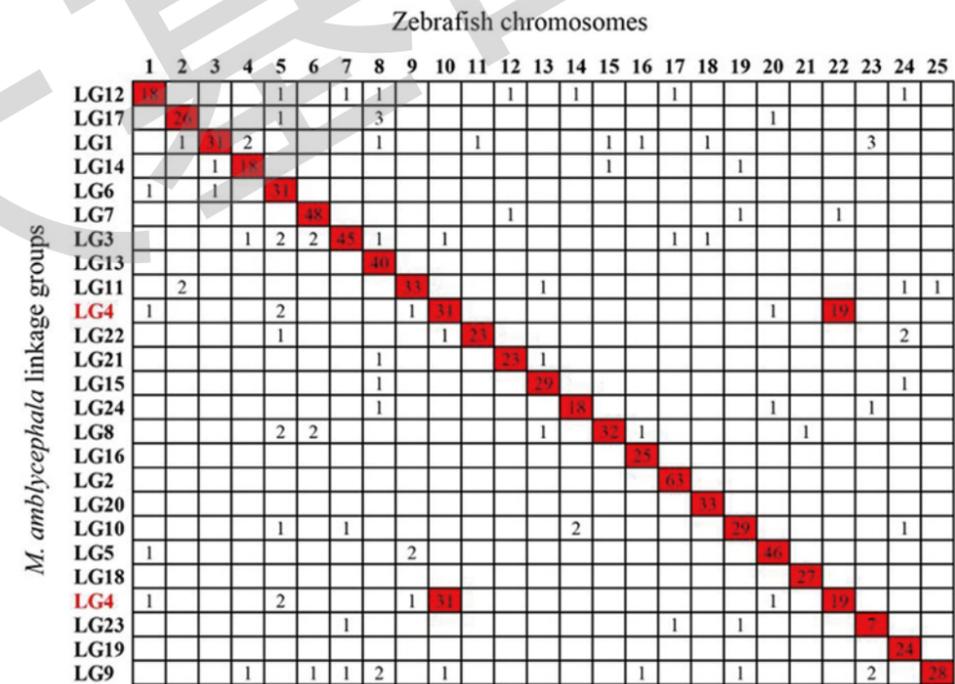


图2 武昌鱼遗传图谱和斑马鱼基因组数据比较结果展示
 纵坐标LGX代表武昌鱼不同的连锁群 横坐标代表斑马鱼不同的染色体

★ 参考文献

Wan S M, Liu H, Zhao B W, et al. Construction of a high-density linkage map and fine mapping of QTLs for growth and gonad related traits in blunt snout bream[J]. Scientific Reports, 2017, 7: 46509.

案例三: 大豆进化+GWAS

开花期和籽粒大小是驯化相关性状,然而大豆中这些性状相关基因的鉴定工作少见报道。本研究中使用简化基因组 (RAD-seq) 技术对群体大小为286个的大豆自然群体进行基因分型分析,基于检测的SNP标记进行了群体进化和GWAS分析,检测到48个与进化相关的基因位点,其中6个与开花时间性状关联,4个与籽粒大小关联。本研究表明种群分化分析、全基因组关联分析和基因表达分析的联合应用是一种有效地鉴定驯化相关基因的方法。

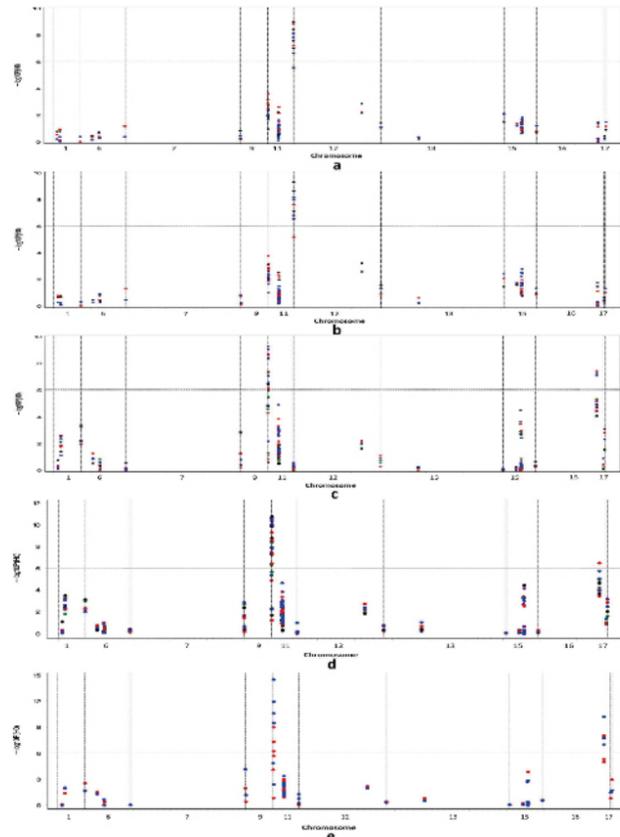


图3 利用驯化相关SNPs进行开花时间、籽粒大小的全基因组关联分析

★ 参考文献

Zhou L, Wang S B, Jian J, et al. Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method[J]. Scientific reports, 2015, 5: 9350.

▶▶ 常见问题

Q: 简化基因组项目常用的内切酶有哪些? 客户可以指定特定的酶来做项目吗?

A: 目前常用的酶是EcoRI、ApeKI。客户也可以指定其它的酶,采用新酶涉及到购买酶、设计接头等,项目执行时间会延长。

Q: 现在GBS和RAD都是简化基因组技术,区别在哪里?

A: GBS和RAD都是基于高通量测序的简化基因组技术,区别主要在于建库流程上,RAD经酶切后还有一次随机打断过程,而GBS建库时不需要物理打段和片段选择,因此建库过程更简便、快速。

Q: 没有参考基因组的RAD测序,检测SNP是否只能检测到纯合的SNP?

A: 没有参考基因组的RAD测序,首先是样品内的测序reads聚类,在每个聚类结果中检测到每个样品的杂合位点信息,并生成Consensus序列,然后再进行样品间的聚类检测SNP。所以也可以检测杂合位点的。

▶▶ 华大发表文献

华大参与发表文章(一作或通讯作者,2017年至今)

发表日期	期刊	影响因子	物种	研究目的	研究内容
2019.04	BMC Genomics		大豆	遗传图谱& QTL定位	大豆叶型相关基因定位
2018.05	Theoretical and Applied Genetics	4.132	大豆	基因克隆及功能验证	酸性磷酸酶基因GmHAD1通过精细定位与大豆中的低磷耐受性相关联
2018.05	Royal Society Open Science	2.243	鲟鱼	遗传图谱	鲟鱼高密度遗传图谱构建
2018.04	Plant Biotechnology Journal	7.443	花生	遗传图谱& QTL定位	使用全基因组重新测序的高密度遗传图谱用于花生中抗病性的精细定位和候选基因发现
2018.03	Science China. Life sciences	2.781	石斑鱼	GWAS	利用RAD-seq基因分型对橙斑石斑鱼 (Epinephelus coioides) 的生长性状进行全基因组关联研究
2017.11	Theoretical and Applied Genetics	4.132	大豆	遗传图谱+ QTL定位	大豆异黄酮含量QTL定位
2017.10	Molecular Breeding	2.465	猕猴桃	遗传图谱+ QTL定位	猕猴桃遗传图谱构建,辅助基因组组装+QTL定位
2017.09	Aquaculture	2.57	金鲳鱼	遗传图谱+ QTL定位	金鲳鱼生长相关性状 QTL定位
2017.05	Frontiers in Plant Science	4.495	赤小豆	遗传图谱+ QTL定位	赤小豆RAD测序构建遗传图谱定位28个性状相关QTL
2017.04	Scientific Reports	4.495	武昌鱼	遗传图谱+ QTL定位	武昌鱼RAD测序构建遗传图谱定位生长和生殖相关性状
2017.04	Ecology and Evolution	2.44	枇杷	进化研究	研究枇杷进化关系
2017.03	G3	2.861	谷子	bin map+ QTL定位	谷子14个农艺性状 QTL定位
2017.02	Theoretical and Applied Genetics	4.132	大豆	BSA+遗传图谱+ QTL定位	大豆抗疫霉根腐病基因 RpsWY精细定位

产品
简介

Product
Profile

转录组学

- 01 | 转录组测序 
- 02 | 全长转录组测序 
- 03 | RNA-seq 
- 04 | Small RNA测序 
- 05 | 长链非编码RNA(lncRNA)测序 

转录组测序

产品概述

转录组测序的研究对象为特定细胞在某一功能状态下所能转录出来的所有RNA的总和,目前该测序技术主要针对mRNA。转录组测序技术在检测基因表达水平变化的同时,还能发现未知转录本和稀有转录本,精确地识别可变剪切位点、基因融合、SNP以及Indel位点,提供最全面的转录组信息;而且不依赖于参考基因组,对于无参考基因组或者参考基因组不完善的物种,可以从头组装转录本。通过高通量测序,能够全面快速地获得某一物种特定组织或器官在某一状态下的几乎所有转录本序列信息,已广泛应用于遗传发育、作物改良和分子育种研究等领域。

产品优势

- 全面丰富的样品类型**
常规建库低至200ng,提供单细胞、FFPE、微量样品的特殊样品服务,微量建库低至pg级。
- 高质量的自主测序平台**
滚环扩增构建DNB测序文库,PCR扩增错误不会累积,无index hopping之忧,低Dup rate无需人为干预,《Nature》、《Cell》、《Immunity》等1000+累积影响因子。
- Dr. Tom系统解决个性化难题**
只需任意一种RNA测序数据,就获得多组学关联信息,多数据库联合分析,多维度数据展示,循环挖掘数据,轻松锁定解释生物学问题的核心基因。
- 极速交付**
样本提取只需2天,从样本检测、建库测序到Dr. Tom交付只需14天,节约您的宝贵时间。

产品应用



技术流程

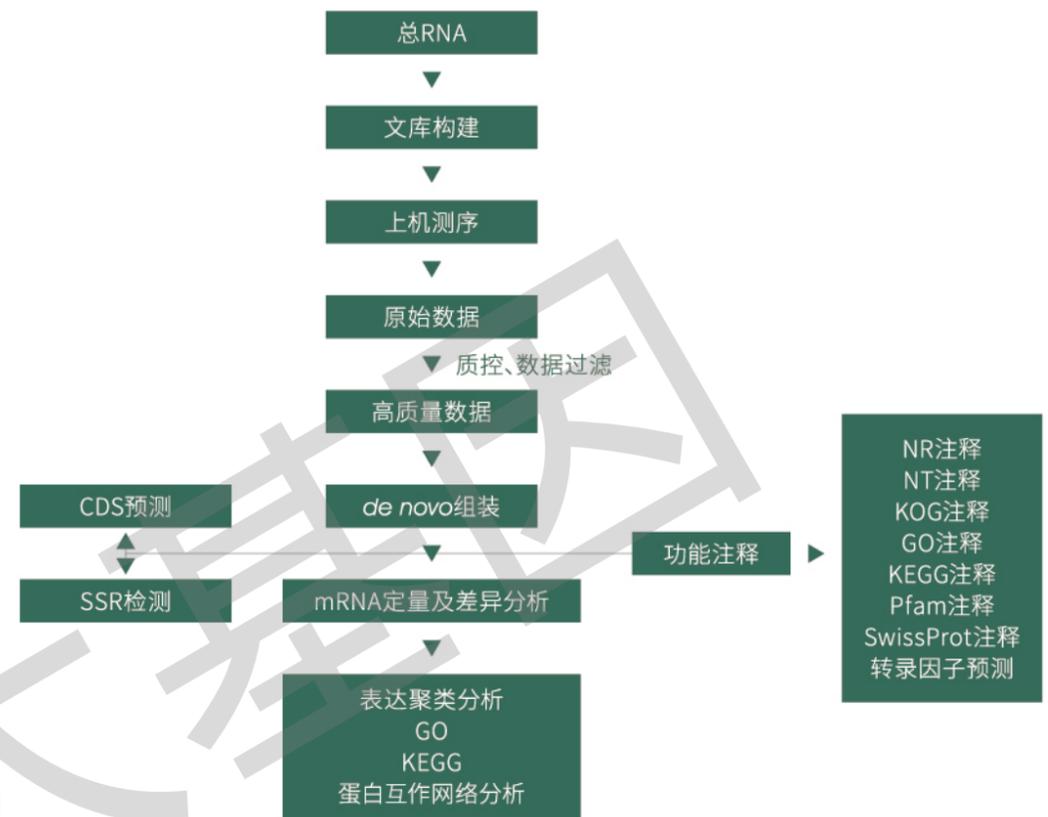


图1 无参考基因组序列的转录组测序技术流程

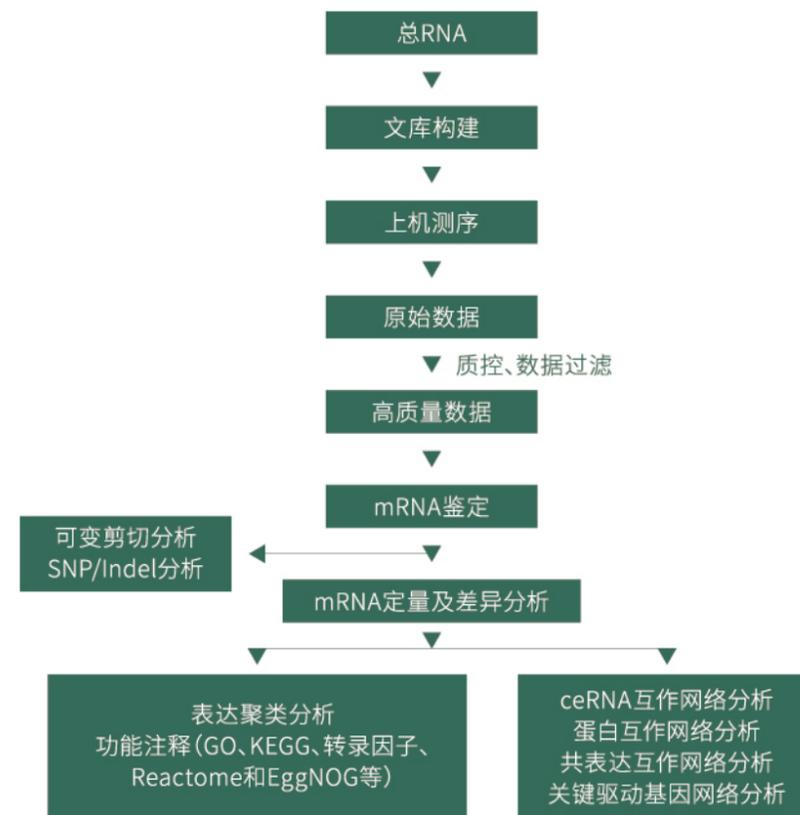


图2 有参考基因组序列的转录组测序技术流程

► 研究内容

无参考基因组序列的转录组标准信息分析

1. 测序数据过滤
2. 转录组 *de novo* 组装
3. Unigene 七大功能数据库注释
4. Unigene 的CDS预测
5. Unigene 的TF编码能力预测
6. Unigene 的SSR检测
7. Unigene 表达量计算 (基因表达水平、PCA分析、基因表达水平箱线图 and 密度图、样品间的表达量韦恩图)
8. 时间序列分析
9. 差异表达基因检测
10. 差异表达基因聚类分析
11. 差异表达基因GO功能分析
12. 差异表达基因Pathway功能分析
13. 差异基因蛋白互作分析
14. 真菌致病基因预测 (真菌样本)
15. 植物抗病基因预测 (植物样本)

有参考基因组序列的转录组标准信息分析

1. 基本数据统计①去除接头序列、低质量序列得到reads信息, ②样品相关性, ③表达量分布, ④RNA分类
2. 参考基因组比对
3. mRNA 鉴定
4. mRNA 定量分析
5. mRNA 差异表达分析 (样本间、组间)
6. mRNA 表达/差异基因聚类
7. mRNA 差异基因GO分类、富集
8. mRNA 差异基因KEGG分类、富集
9. mRNA 结构分析①可变剪切分析, ②SNP/Indel分析

Dr.Tom 信息分析

数据库注释

1. 转录因子注释 (AnimalTFDB/PlantTFDB)
2. Rfam、Pfam、Reactome、COG、EggNOG 和 InterPro 数据库注释

互作网络分析

1. 靶基因分析①miRNA-mRNA 靶向关系分析, ②lncRNA-mRNA 靶向关系分析
2. ceRNA 互作网络分析①lncRNA-mRNA 联合分析, ②circRNA-mRNA 联合分析 (仅限人、小鼠)
3. 蛋白互作网络分析
4. 共表达互作网络分析

特色分析

1. 自定义标签和自有数据上传
2. 外部数据库信息 (TCGA、ARCHS4)
3. 卡方检验
4. 关键驱动基因网络图分析
5. 时间序列分析

► 案例分析

案例一: 转录组测序构建大鲵的参考基因组

★ 案例描述

收集成年中国大鲵的20多个组织样本 (腹部皮肤、背部皮肤、侧部皮肤、肺、心脏、肾、胰腺、小肠、脾脏、胃、脑、脊髓、软骨、眼睛、指尖、长骨、上颌骨、头盖骨、肌肉、卵巢、脂肪、尾部脂肪、血液)。每个样本约为6.5Gb, 通过转录组测序组装出一个完整的参考基因组。组装得到93,366条非冗余转录本, 平均长度1,326bp。将组装得到的序列注释到NR、Swiss-Prot、KEGG、COG、GO数据库进行功能注释, 最后有41,874条序列被注释到。采用BlastX、ESTscan、Transdecoder 软件分别预测CDS序列, 鉴定了26,135个蛋白。

★ 结果展示

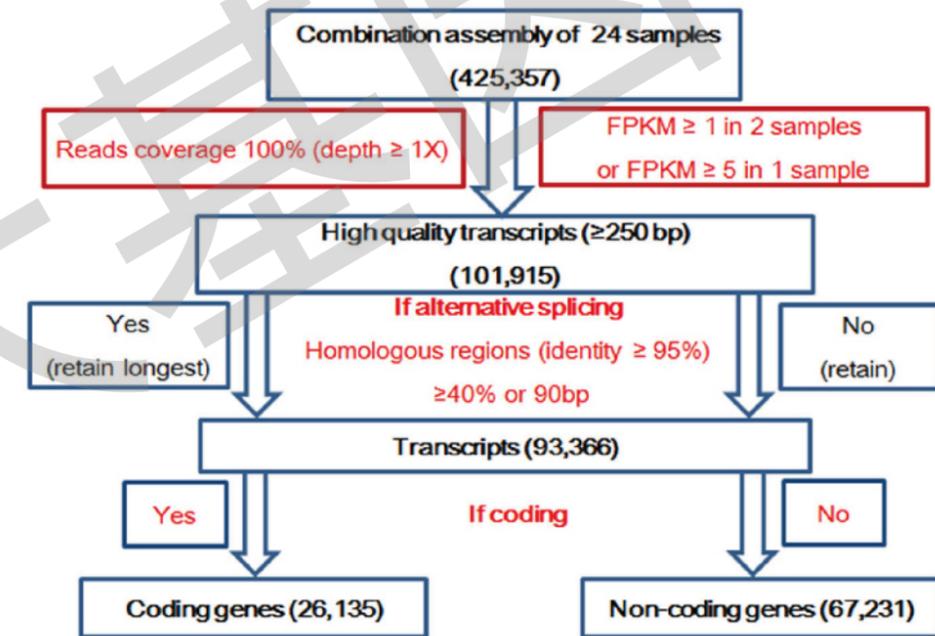


图3 转录组组装流程

★ 参考文献

Geng X, Li W, Shang H, Gou Q, et al. A reference gene set construction using RNA-seq of multiple tissues of Chinese Giant Salamander, *Andrias davidianus*. *GigaScience*. 2017 Feb 15. doi: 10.1093/giga-science/gix006.

案例二: 水稻大规模转录组测序和蛋白质组数据集研究

本研究目的是通过转录组测序和蛋白质组数据, 提高水稻基因组注释。建立了一个全面的转录本数据库, 来自于29个公开的RNA测序数据集, 也重新分析了9个公开的水稻蛋白质组数据集。总之, 鉴定到420K多肽谱。经过严格的筛选, 发现了1584个高置信度的新多肽。这些新的多肽被聚类到692个基因组位点, 这些位点因我们的结果, 使注释得到改进。80%的新肽在至少一种其他植物物种的蛋白质序列中具有同源性。对于在基因间区域聚集的多肽 (可能是潜在的新基因), 共鉴定了101个位点, 其中43个位点有高置信度的蛋白域。这些结果都可以获取, 帮助其他研究者重新注释水稻基因组。

► 技术参数

- 样品类型: 完整且无污染的总RNA;
样品需求量 (单次): 鼠200ng, 其他物种1μg;
样品浓度: 鼠10-1000ng/μL, 其他物种40-2500ng/μL;
样品纯度: 28S/18S ≥ 1.0; 动物RIN ≥ 7.0, 植物和真菌RIN ≥ 6.5; 昆虫无28S/18S及RIN值要求;
测序策略: PE100/PE150;
推荐数据量: 6G或10G;
项目执行周期: 标准执行周期为24个工作日, 极致交付周期为14个工作日 (含样品检测)。

★ 结果展示

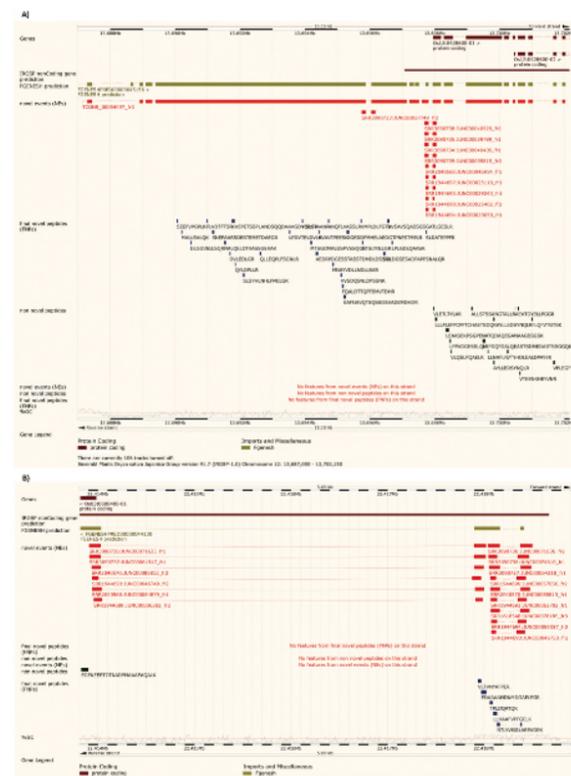


图4 发现的两个典型可变剪切事件 (NEs) 的基因组图谱

★ 参考文献

Z. Ren, D. Qi, N. Pugh, K. et al. Improvements to the rice genome annotation through large-scale analysis of RNA-Seq and proteomics datasets. *Molecular and Cellular Proteomics*, November 2018, mcp.RA118.000832.

▶▶ 常见问题

Q: 与基因芯片相比, 转录组高通量测序有什么优势?

A: 基因芯片是用已知序列的探针和样本mRNA进行杂交来获得mRNA的序列信息, 如果样本mRNA以前从来没有报道过, 那么基因芯片上面就不会有相应的探针序列, 也无法检测到新的mRNA。另外, 核酸杂交的背景噪音很高, 存在交叉杂交现象。转录组是直接对样本mRNA进行测序, 能够发现很多新的mRNA; 转录组测序对基因表达上调或下降的检测范围能够达到几万倍, 远比基因芯片的百倍左右要灵敏。而且在有参考基因组的情况下, 通过转录组测序您还可以分析可变剪切、基因结构变异、全基因组水平基因表达丰度等情况。

Q: 是否需要生物学重复? 重复几次?

A: 是的, 至少需要两次生物学重复, 3次以上的生物学重复更好。2011年7月Hansen发表的文章表明生物学差异是基因自身表达的特性, 与检测技术的选择以及数据处理的方式无关。如果不设生物学重复, 高影响因子的杂志可能会因此而拒稿。

Q: 可否提供BGISEQ转录组标准品测试数据供客户测评?

A: 可以, 我们1个UHRR标准品, 构建了3个文库, 数据都已经上传到EBI。可登录以下网站进行下载: <http://www.ebi.ac.uk/ena/data/view/PRJEB19428>。

▶▶ 华大发表文献

华大参与发表文章(2017年至今)

物种	发表日期	发表期刊	影响因子	文献名称
大鲵	2017.02	GigaScience	4.688	A reference gene set construction using RNA-seq of multiple tissues of Chinese giant salamander, <i>Andrias davidianus</i>
昆虫	2017.08	GigaScience	4.688	Positive and relaxed selection associated with flight evolution and loss in insect transcriptomes
豌豆	2017.08	Science	37.205	Structure and assembly mechanism of plant C2S2M2-type PSII-LHCII supercomplex
酿酒酵母	2017.10	Genome Biology	11.908	Dynamic changes in eIF4F-mRNA interactions revealed by global analyses of environmental stress responses
虎蛾	2018.04	Molecular Phylogenetics and Evolution	4.419	A phylogenomic analysis of lichen-feeding tiger moths uncovers evolutionary origins of host chemical sequestration
拟南芥	2018.08	Plant Methods	4.269	Comparative performance of the BGISEQ-500 and Illumina HiSeq4000 sequencing platforms for transcriptome analysis in plants
水稻	2018.10	Mol Cell Proteomics	5.232	Improvements to the rice genome annotation through large-scale analysis of RNA-Seq and proteomics datasets
鹰嘴豆	2018.10	Plant Biotechnol J.	6.305	Integrated transcriptome, small RNA and degradome sequencing approaches provide insights into <i>Ascochyta</i> blight resistance in chickpea.
雨生红球藻	2019.01	Biotechnol Biofuels	5.497	Comparative transcriptome analysis of a long-time span two-step culture process reveals a potential mechanism for astaxanthin and biomass hyper-accumulation in <i>Haematococcus pluvialis</i> JNU35
苦皮藤	2019.01	BMC Genomics	3.730	De novo leaf and root transcriptome analysis to explore biosynthetic pathway of Celastrolin V in <i>Celastrus angulatus maxim</i>
烟草	2019.01	Front. Plant Sci	3.678	Tolerance of Transplastomic Tobacco Plants Overexpressing a Theta Class Glutathione Transferase to Abiotic and Oxidative Stresses

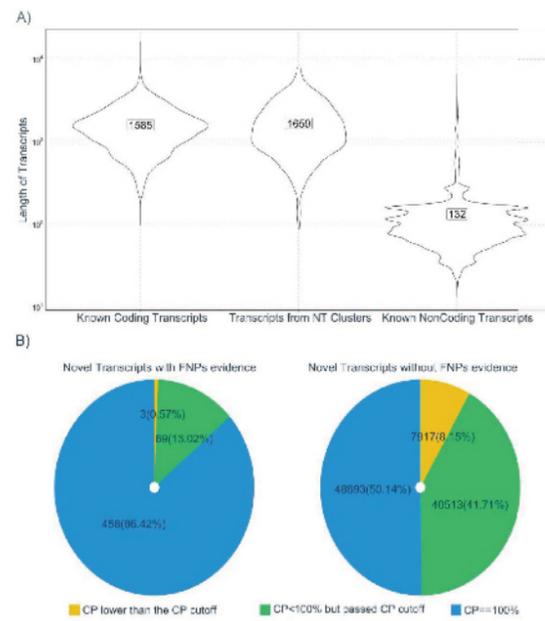


图5 用最终的新多肽 (FNPs) 的证据评估新的可变剪切事件 (NEs)

A. 新转录本长度分布 B. 新转录本编码潜能

华大参与发表文章(2017年至今)

物种	发表日期	发表期刊	影响因子	文献名称
斑马鱼	2019.01	Chemosphere	4.427	Microcystin-LR exposure induced nephrotoxicity by triggering apoptosis in female zebrafish
拟南芥	2019.03	Environmental Science & Technology	6.653	Comparative physiological and transcriptomic analyses reveal the toxic effects of ZnO nanoparticles on plant growth

全长转录组测序 (Iso-Seq)

▶▶ 产品概述

近年来,随着高通量测序技术的发展,转录组测序已经成为研究基因表达调控的主要手段。PacBio的Sequel测序平台,因其长读长优势,可以直接得到全长转录本信息,已被越来越多地用于构建/完善参考物种基因集。华大基因基于PacBio平台,在常规全长转录组的基础上开发了独特的UMI全长转录组产品、5X全长转录组产品以及polyA+全长转录组产品,以更低的成本为大家提供从定性、定量到调控全方位研究方案。

▶▶ 产品应用&产品优势

产品名称	产品应用	产品优势
5X全长转录组	构建参考基因集 基因功能研究 辅助基因组注释 转录本结构变异研究 基因/转录本差异研究	同样的数据量,可得到5倍的有效数 检测到的基因/转录本数目多 基因/转录本定量结果准 其他优势同“常规全长转录组”
PolyA+全长转录组	PolyA相关研究 构建参考基因集 基因功能研究 辅助基因组注释 转录本结构变异研究	可检测PolyA长度分布 可检测PolyA尾中非A碱基类型及出现频率 研究揭示polyA在表达调控中的作用 其他优势同“常规全长转录组”
常规全长转录组	构建参考基因集 基因功能研究 辅助基因组注释 转录本结构变异研	不需要组装,就可以准确的得到全长转录本的序列信息 可以发现新的基因和转录异构体 鉴定可变剪接和基因融合 辅助基因组注释

技术流程

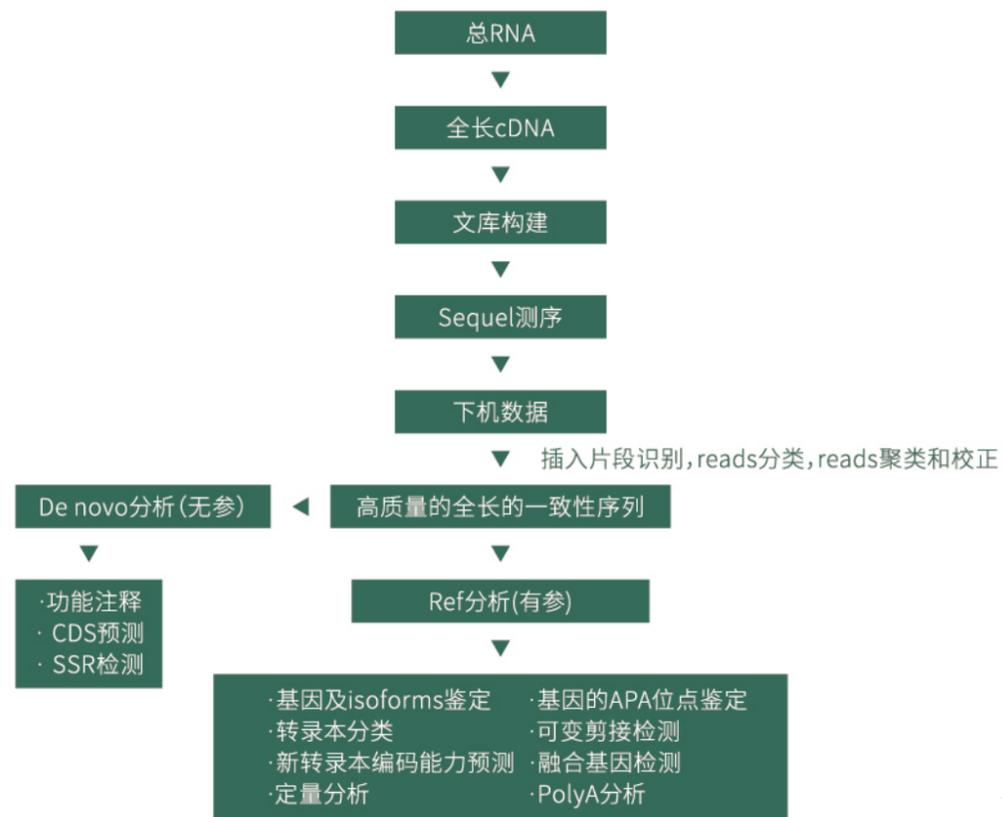


图1 无参考基因组序列的转录组测序技术流程

信息分析内容

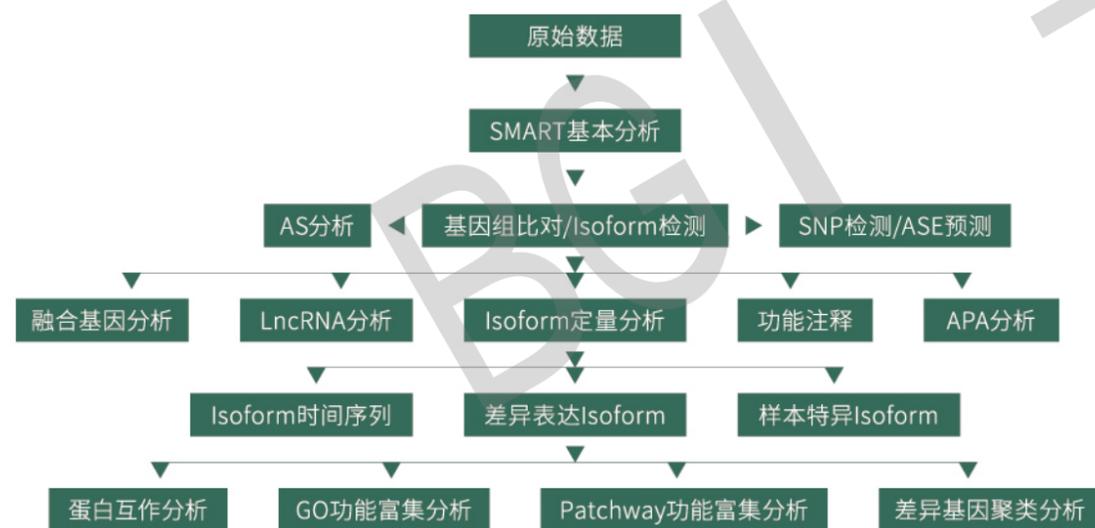


图2 Iso-Seq de novo标准分析条款

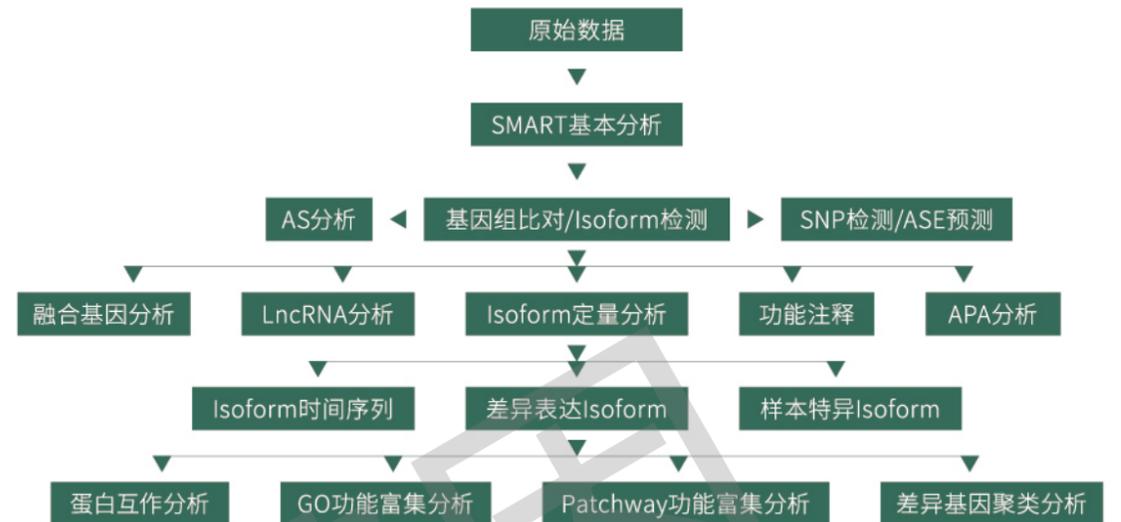


图3 Iso-Seq ref标准分析条款

技术参数

样品类型: 完整且无污染的总RNA;
 样品需求量: $\geq 1\mu\text{g}$;
 样品浓度: $c \geq 285\text{ng}/\mu\text{L}$;
 样品纯度: $\text{RIN} \geq 8$, $28\text{S}/18\text{S} \geq 1.4$;
 推荐数据量: 根据需求选择5X全长转录组文库、polyA+全长转录组文库或常规全长转录组文库,
 测序1个cell或多个cell (sequel平台); 如果用sequel-II平台, 可能需要多个样本pooling;
 项目执行周期: 40个工作日。

案例分析

案例一: 玉米和高粱全长转录组测序

★ 方案设计

1. 材料选择

- ①玉米(B73): 根、茎、叶、幼苗(14日龄), 雌穗(ears, v8期), 雄穗(tassels, v7期), 花粉(r1期)、胚, 胚乳和果皮(授粉后20天从种子获得), 须(silk, r1期), 苞片(最内层)。
- ②高粱(BTx623): 根、茎、叶、幼苗(组织萌芽后14天), 胚, 胚乳和果皮(授粉后20天), 花粉(9-10周阶段), 花序-1(1-5mm) 花序-2(5-10mm), 花序-3(1-2cm)

2. 测序策略

- ①PacBio测序
 建库: Barcode建库; 文库大小: $< 1\text{ kb}$ 、 $1-2\text{ kb}$ 、 $2-3\text{ kb}$ 、 $3-5\text{ kb}$ 、 $> 5\text{ kb}$ 文库
 测序: PacBio RS II: 130 SMRT Cells; PacBio Sequel: 5 SMRT Cells;
- ②Illumina测序
 Illumina HiSeq 2500 PE125

3. 研究内容

- ①Isoform 统计和分类
- ②Isoform 和可变剪接的组织特异性
- ③伴随无义介导的mRNA衰变的AS事件
- ④可变剪接的保守性
- ⑤转录因子Isoform分析
- ⑥多聚腺苷酸化分析(APA)
- ⑦LncRNA分析
- ⑧基因表达聚类分析
- ⑨基因进化分析

★ 主要结论

1. 测序得到总数据量(玉米+高粱): 6,893,280 reads; 长度分布: 256-6643 bp; 玉米有1,570,093(96.7%)条高质量转录本能比到玉米参考基因组, 非冗余序列136,745; 高粱有979,305(89.5%)条高质量转录本能比到参考基因组, 非冗余序列95,380。

2. 玉米的11个组织中有1659个共有isoforms, 高粱的11个组织中有1069个共有isoforms; 在玉米中, 花粉的特异性isoforms占比最高(27.2%), 根的特异性isoforms占比最低(14.5%); 在高粱中, 花序-1的特异性isoforms占比最高(35.8%), 花粉次之(34.1%), 根的特异性isoforms占比最低(20.8%)。

3. 在玉米和高粱中, 分别有18,741(45%) 和13,327(38.5%)个基因存在可变剪接事件。可变剪接类型包括: (ES) Exon skipping; (A5) alternative 5' splice-site; (A3) alternative 3' splice-site; (IR) intron retention; (AF) alternative first exon; (AL) alternative last exon; 其中IR在所有组织中占比最高, AL占比最少。

4. 玉米isoform中有55,080(40.3%)个无义介导的mRNA衰变(NMD); 高粱中有34,322(36%)个NMD; 玉米和高粱中, Non-NMD isoforms均比NMD isoforms表达量高。

5. 玉米和高粱同源基因的可变剪接形式类似; 玉米同源基因的可变剪接数目比高粱多; 玉米和高粱保守isoform中IR可变剪接形式产生NMD的可能性最高; ES产生NMD的可能性最低。

6. 不同组织富集的基因GO功能差异较大; 玉米和高粱同一组织富集的基因GO功能也有差异。

7. 玉米和高粱中检测到的转录因子家族数目和参考基因集一致; 玉米新发现179个转录因子isoforms, 高粱中新发现129个转录因子isoforms。

8. 玉米和高粱的APA motif占比最高的是AATAAA motifs, 排名前三的motifs在两个物种中数量相同。

9. 通过PLEK构建了RNA分类模型, 获得lncRNA序列; 得到1706个高粱新lncRNA, 平均长度1241bp, 长于前期PacBio平台研究结果(平均长度880bp); 高粱lncRNA的平均长度比玉米(535bp)长。

10. 高粱不同组织间基因表达相关性跟组织间的进化关系相关, 进化上越相近的组织, 相关性越高; 玉米不同组织间基因表达相关性跟组织进化无明显的相关性; 玉米和高粱同一组织基因表达相关性比同一物种不同组织相关性强。

基因进化分析发现, 玉米比高粱有更多的“年轻”基因; 玉米的蛋白编码基因中有59.2%是“古老”基因, 3%是“年轻”基因; 高粱的蛋白编码基因中有66%是“古老”基因, 6.7%是“年轻”基因; 和进化“年轻”的基因相比, “古老”基因基因长度和ORF长度更长, Isoform种类更多。生殖组织比营养组织的进化年龄更高; 一般情况, 进化年龄指数(TAI)越高的组织, 转录组多样性指数(TDI)越高; 玉米和高粱同一组织的TAI和TDI指数有细微差别。

★ 参考文献

Wang B, Regulski M, Tseng E, et al. A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing[J]. Genome research, 2018, 28(6): 921-932.

案例二: 全长转录组测序研究玉米转录组的复杂性

★ 方案设计

1. 材料选择

玉米自交系B73不同发育阶段的6个组织(根、花粉、胚芽、胚乳、幼雌穗、幼雄穗), 提取RNA;

2. 测序策略

Illumina测序: 6个组织进行RNA-Seq测序, 每个样品三个重复;

PacBio测序: 每个样本反转录之前加入特异性barcode, 后续进行等量混合, 上机测序47cell;

3. 分析方案

检测玉米可变剪接现象; 转录因子分析; lncRNA分析; 融合基因分析; 甲基化分析。

★ 主要结论

1. 构建5种不同片段大小的文库, 上机测序47cell, 总共产生3,716,604条reads, 过滤低质量的reads后, 总共获得1,553,692条全长的转录本序列(FL)。

2. 和RefGen-V3的isoforms进行长度比较, 发现全长转录本预测出来的转录本整体上比V3基因集的要长。在目前的V3基因集中一个基因平均有2.84个isoforms, 而全长转录组数据显示, 一个基因平均有6.56个isoforms, 是前者的两倍多。Isoforms的组织特异性分析显示: 花粉有更高的组织特异性, 而根的特异性最低。

3. 在玉米的V3参考基因集中, 转录因子数目为2,624, 分为57个家族。全长转录组解决方案将转录因子的数据增加到5,423个, 几乎是两倍。

4. 鉴定出878个lncRNA, 其中11个是以前报道过的, 867个是新的lncRNA。这些lncRNA的平均长度为1.1kb(范围为0.2kb-6.6kb), 比之前的认知的lncRNA要长很多(平均400+kb)。花粉拥有最多的特异的lnc(238个), 穗是最少的lnc(68)个。

5. 从PacBio数据中鉴定出1,430个融合转录本, 其中143个被Illumina数据支持。结果表明, 融合事件多发生在染色体间。

★ 参考文献

Wang B, Tseng E, Regulski M, et al. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing[J]. Nature communications, 2016, 7: 11708.

案例三: 丹参药用成分合成机理研究

★ 方案设计

1. 研究材料

丹参酮一般认为产生于丹参根部周皮部, 研究分别取了根部的周皮(periderm)、韧皮(phloem)、木质(xylem)3种类型的根部组织进行了mRNA测序。

2. 研究方法

3种类型根部样本各设置3个生物学重复, 总共9个样本, 采用HiSeq 2500 PE100进行测序, 每个样本产生~5G raw data。9个样本混合测序, 采用PacBio RSII进行测序, 建<1kb、1-2kb、2-3kb、>3kb四个SMRT bell文库, 总共产生~4.8G raw data;

★ 主要结论

1. 采用HiSeq 2500数据对PacBio RSII平台所产生的subreads进行了校正, 最后得到了16,241个高质量非冗余isoforms。

2. 基于HiSeq 2500产生的mRNA数据的差异表达分析, 发现了在根部周皮部特异表达与者高表达丹参酮合成相关基因, SmCPS1、SmKSL1、GGPS、IPI、CYP等。

3. 最后研究者使用得到的16,241个高质量的Isoforms进行了可变剪接分析, 发现了大约有40%检测基因位点发生了可变剪接现象, 其中有些基因参与了萜类化合物代谢及类异戊二烯代谢。

★ 参考文献

Xu Z, Peters R J, Weirather J, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis[J]. The Plant Journal, 2015, 82(6): 951-961.

常见问题

Q: 为什么我测的数据量增加了好几倍,得到全长转录本数基本没什么变化?

A: Sequel平台测序数据量主要受读长和reads数两个因素影响。Reads数的多少主要受限于测序芯片上ZMW孔的多少,sequel一张测序芯片上有1M个孔,一般测序loading率在30-70%之间波动(即每次测序会有300K-700K的有效reads),这个数量不会受测序芯片和试剂的升级影响。在常规建库测序中,测序得到的每一条reads代表一条转录本序列(全长比例大约占50%左右),即一般测序一个cell会得到300K-700K的转录本(其中全长转录本有150K-350K条)。

随着sequel测序试剂的升级,测序读长越来越长,在reads数不变的情况下,测序产量也得到了大幅提升,下表是sequel不同试剂版本对应的产量、读长和reads数,从表中可以看出,3.0试剂相比2.0试剂测序产量提升了8倍,平均reads数只提升了不到50%。所以在全长转录组测序中不能只关注测序产量,得到的转录本数量是一个更重要的关注指标。

	Sequel(2.0试剂)	Sequel(2.1试剂)	Sequel(3.0 LR试剂)	提升率(3.0 LR试剂/2.0试剂)
平均产量	4.5G	10G	40G	8.88
平均读长	13K	23K	81K	6.23
平均Reads数	346k	434k	494K	1.42

多倍通量全长转录组技术是将多条转录本连成一条reads进行建库测序,测得的一条reads代表多条转录本,从而提高了全长转录本的得率。相同的测序数据量,可得到常规文库测序5倍的全长转录本。

Q: 多倍通量全长转录组产品的优势?

A: 相比于常规全长转录组测序,多倍通量全长转录组产品有以下优势:

- ①有效转录本reads数可提升3-5倍
- ②基因/转录本检测数目接近Illumina平台
- ③Isoform准确度高
- ④测序长度分布无偏向性
- ⑤UMI和多倍通量完美结合,基因定量结果更准确
- ⑥采用多倍通量技术,大部分物种测1个sequel cell数据就可满足研究需求

Q: 用多倍通量技术,一个样本需要测多少数据?

A: 采用多倍通量技术,大部分物种测1个sequel cell数据就可满足研究需求

以下是部分已发表物种全长转录组测序数据量,请选择转录本复杂程度相近的物种作参考:

物种	测序Reads数	转录本数目	推荐测序Cell数(sequel)	推荐测序Cell数(多倍通量)
海岛棉	2,542,318	176,849	~6cell	2cell
玉米	3,716,604	111,151	~6cell	2cell
拟南芥	844,003	78,010	~4cell	1cell
菠萝	1,068,545	45,876	~3cell	1cell
毛竹	288,312	42,280	~3cell	1cell
盐生草	433,420	54,835	~3cell	1cell
高粱	1,838,330	27,860	~2cell	1cell
苜蓿	448,454	52,787	~2cell	1cell
黄芪	494,408	27,975	~2cell	1cell
鸡	1,849,786	64,277	~3cell	1cell
真菌	~500K	22,956	~2cell	1cell

Q: PolyA+全长转录组可以做哪些分析?

A: 目前通过PolyA+全长转录组可以研究以下内容:

- PolyA长度分布
- PolyA尾中非A碱基类型及出现频率
- 转录本表达量、结构与功能与polyA的关联。

结合翻译组及RNA半衰期检测,还可以研究PolyA长度及其他碱基修饰与mRNA稳定性或翻译效率的关系

RNA-Seq

产品概述

RNA-Seq是直接对某一物种或特定细胞在某一功能状态下产生的mRNA进行高通量测序,用来研究基因的表达差异情况,已经广泛应用于遗传发育、作物改良和分子育种研究等领域。相比转录组, RNA-Seq产品更加侧重基因定量研究,整体性价比更高。相比表达谱芯片, RNA-Seq产品利用数字化信号,无背景噪音,无交叉杂交;没有物种限制;能发现未知基因和低丰度基因。

产品优势



全面丰富的样品类型
鼠样本常规建库仅需200ng,提供单细胞、FFPE、微量样品的特殊样品服务,微量建库低至pg级。



高质量的自主测序平台
滚环扩增构建DNB测序文库,PCR扩增错误不会累积,无index hopping之忧,低Dup rate无需人为干预,《Nature》、《Cell》、《Immunity》等1000+累积影响因子。



Dr. Tom系统解决个性化难题
只需任意一种RNA测序数据,就获得多组学关联信息,多数据库联合分析,多维度数据展示,循环挖掘数据,轻松锁定解释生物学问题的核心基因。

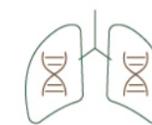


极速交付
样本提取只需2天,从样本检测、建库测序到Dr. Tom交付只需11天,节约您的宝贵时间。

产品应用



不同发育阶段的基因表达模式



各器官的基因表达模式



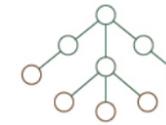
环境适应性机制研究



抗病性机制研究



与微生物互作机制研究



物种进化研究

技术流程

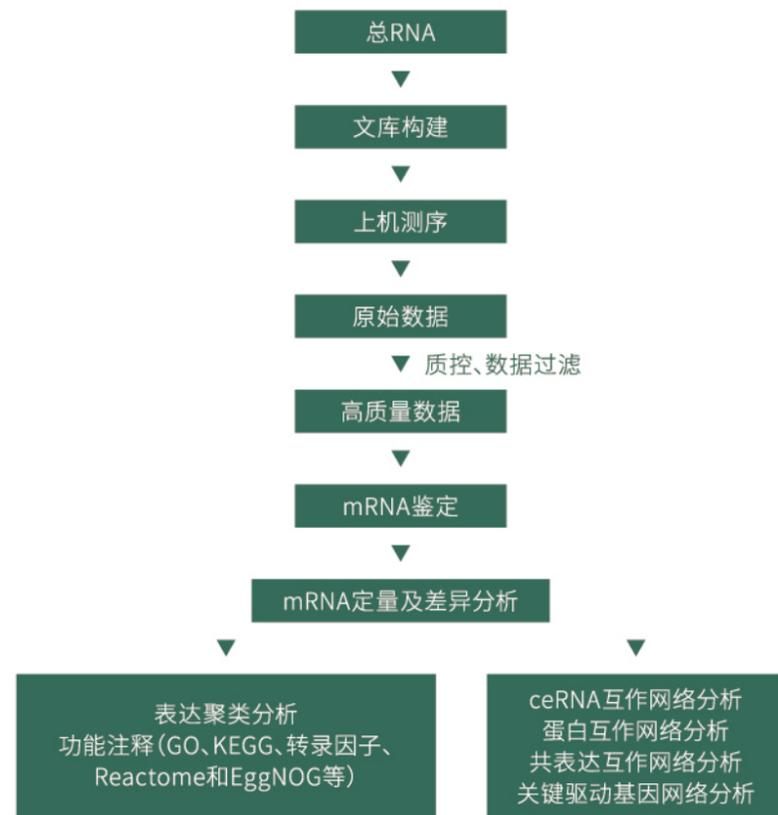


图1 RNA-Seq技术流程

技术参数

样品类型: 完整且无污染的总RNA;
 样品需求量(单次): 鼠200ng, 其他物种1μg;
 样品浓度: 鼠10-1000 ng/μL, 其他物种40-2500 ng/μL;
 样品纯度: 28S/18S≥1.0; 动物RIN≥7.0, 植物和真菌RIN≥6.5; 昆虫无28S/18S及RIN值要求;
 测序策略: SE50;
 推荐数据量: 20M clean reads;
 项目执行周期: 标准执行周期为15个工作日, 极致交付周期为11个工作日(含样品检测)。

案例分析

案例一: RNA-Seq研究拟南芥防御反应机制

★ 案例描述

铜离子是所有生物体所必需的, 研究者之前报道过拟南芥中铜离子能引起防御反应, 这依赖于乙烯信号通路。但是铜离子引起乙烯生物合成的机制仍不清楚。本研究利用 RNA-Seq对6个拟南芥样本, 即铜离子处理3个不同时间点, 2组生物学重复进行测序, 发现铜离子特异性激活拟南芥*AtACS8*基因表达, 来促进乙烯早期生物合成, 进而引起拟南芥防御反应。而且, 铜离子诱导的*AtACS8*表达, 是依赖于*AtACS8*启动子区的铜响应顺式元件(CuRE)。

★ 结果展示

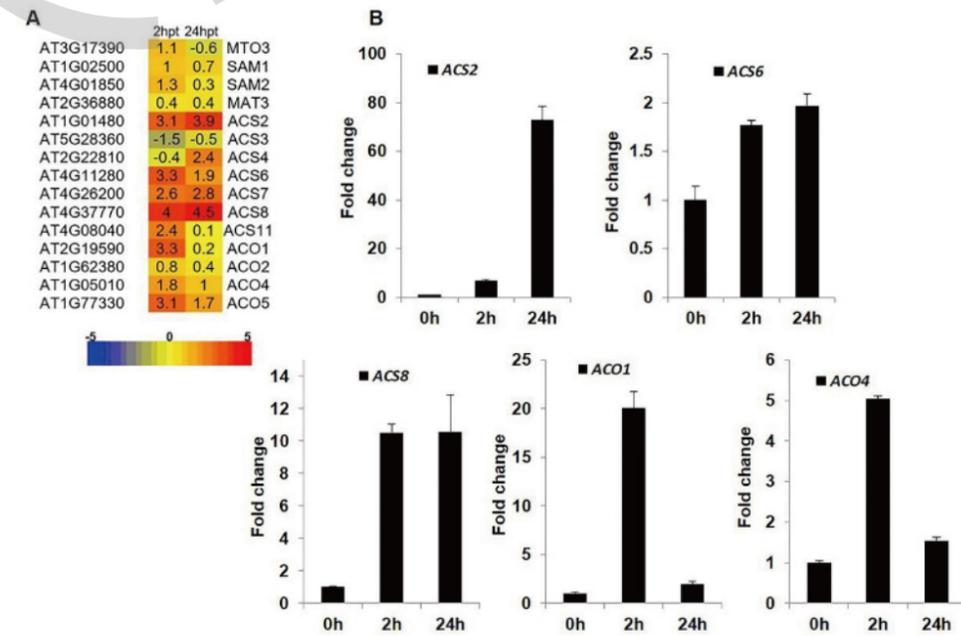


图2 CuSO₄处理后乙烯生物合成通路基因的表达情况

★ 参考文献

Zhang B, Liu H, Ding X, et al. *AtACS8* plays a critical role in the early biosynthesis of ethylene elicited by copper ions in Arabidopsis. *Journal of Cell Science*. 2017 Aug 3. pii: jcs.202424.

研究内容

标准信息分析

1. 基本数据统计
 - ①去除接头序列、低质量序列得到reads信息
 - ②样品相关性
 - ③表达量分布
 - ④RNA分类
2. 参考基因组比对
3. mRNA鉴定
4. mRNA定量分析
5. mRNA差异表达分析(样本间、组间)
6. mRNA表达/差异基因聚类
7. mRNA差异基因GO分类、富集
8. mRNA差异基因KEGG分类、富集

Dr.Tom信息分析

- 数据库注释
1. 转录因子注释(AnimalTFDB/PlantTFDB)
 2. Rfam、Pfam、Reactome、COG、EggNOG和 InterPro数据库注释
- 互作网络分析
1. 靶基因分析
 - ① miRNA-mRNA靶向关系分析
 - ② lncRNA-mRNA靶向关系分析
 2. ceRNA互作网络分析
 - ① lncRNA-mRNA联合分析
 - ② circRNA-mRNA联合分析(仅限人、小鼠)
 3. 蛋白互作网络分析
 4. 共表达互作网络分析
- 特色分析
1. 自定义标签和自有数据上传
 2. 外部数据库信息(TCGA、ARCHS4)
 3. 卡方检验
 4. 关键驱动基因网络图分析
 5. 时间序列分析

案例二: RNA-Seq在作物高效利用氮肥调控机制研究中的应用

★ 案例描述

通过选育半矮化作物,极大地提高了绿色革命品种谷物的产量,但因使用过多无机氮肥而导致环境恶化日益加剧。为加强全球粮食安全及可持续发展,提高作物氮利用效率的需求日益迫切。因此,亟需深入了解作物生长、氮同化、碳固定的共同调节机制。材料为水稻半矮秆品系NJ6-*sd1*、对应野生型品系NJ6(NJ6-*sd1*的氮素利用效率比NJ6低)。通过RNA-Seq和ChIP-Seq两项技术,解析GRF4作用的分子机制。GRF4表达增加改善氮素吸收效率的同时,不影响半矮化性状,能进一步提高作物产量。

★ 结果展示

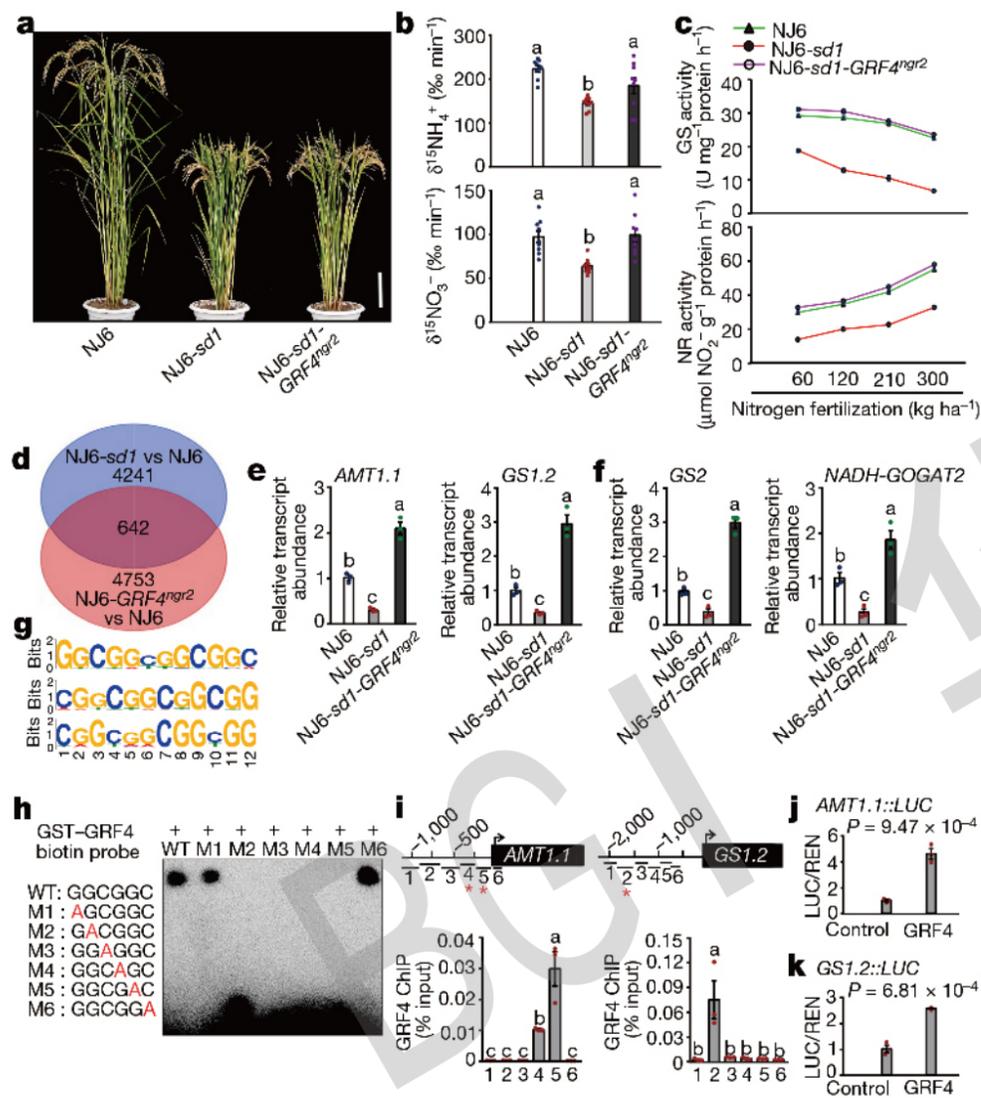


图3 GRF4调控多种氮代谢基因的表达

★ 参考文献

Li S, Tian Y, Wu K, et al. Modulating plant growth-metabolism coordination for sustainable agriculture. Nature. 2018 Aug;560(7720):595-600.

► 常见问题

Q: RNA-Seq必须要有参考基因组序列吗?

A: 不一定要有参考基因组序列,但一定要有参考序列,如unigene序列、mRNA序列、CDS序列等可以作为参考序列。

Q: 若无参考序列,能够进行哪些生物信息分析?

A: 在无参考序列情况下,可通过参考有基因组序列的近缘物种(需预先指定)的基因注释信息来完成相应的标准信息分析。

Q: RNA-Seq推荐测序数据量与基因组大小有关吗?

A: RNA-Seq推荐测序数据量,主要与基因的数量有关,不同物种基因组大小相差比较大,但是编码基因的数量相差并不大,一般物种在3万左右,所以,对于一般物种RNA-Seq推荐10-20 M clean reads数据量。

Q: 测序后有何验证方法?

A: 实验验证的方法最常见的是通过实时荧光定量PCR(qRT-PCR)技术来验证测序结果,还有FISH(原位荧光杂交)、微阵列芯片技术、Northern blot等。功能验证一般是基因敲除、敲低或过表达、转基因等。

Q: BGISEQ RNA-Seq有没有发布demo数据?

A: 已发布一个UHRR人标准品下机数据并共享了数据结果和分析方法。网址:<http://www.bgitechsolutions.com/science/cat-91-588-1464.html>。

► 华大发表文献

华大参与发表文章(2017年至今)

物种	发表日期	发表期刊	影响因子	文献标题
青蒿	2018.01	New Phytologist	7.33	A novel HD-ZIP IV/MIXTA complex promotes glandular trichome initiation and cuticle development in <i>Artemisia annua</i>
拟南芥	2018.03	Nature Communications	12.124	Hydrogen peroxide positively regulates brassinosteroid signaling through oxidation of the BRASSINAZOLE-RESISTANT1 transcription factor
水稻	2018.08	Nature	41.577	Modulating plant growth-metabolism coordination for sustainable agriculture
拟南芥	2018.01	Developmental Cell	9.174	Phytochrome A Negatively Regulates the Shade Avoidance Response by Increasing Auxin/Indole Acetic Acid Protein Stability
秀丽隐杆线虫	2019.01	Cell Rep	8.282	A Persistence Detector for Metabolic Network Rewiring in an Animal
桃树	2019.02	Plant Biotechnology Journal	6.305	A single nucleotide mutation in GID1c disrupts its interaction with DELLA1 and causes a GA-insensitive dwarf phenotype in peach

华大参与发表文章(2017年至今)

物种	发表日期	发表期刊	影响因子	文献标题
番茄	2019.01	New Phytologist	7.33	A tomato B-box protein SIBBX20 modulates carotenoid biosynthesis by directly activating PHYTOENE SYNTHASE 1, and is targeted for 26S proteasome-mediated degradation
拟南芥	2019.03	Environmental Science & Technology	6.198	Comparative physiological and transcriptomic analyses reveal the toxic effects of ZnO nanoparticles on plant growth
啮齿动物	2019.02	Hepatology	13.246	Hepatic IRF6 alleviates liver steatosis and metabolic disorder by transcriptionally suppressing PPAR γ
鱼	2019.03	Environmental Science & Technology	6.198	Hypoxia causes transgenerational impairment of ovarian development and hatching success in fish
鼠	2019.03	Molecular Cancer	6.204	Hypoxic BMSC-derived exosomal miRNAs promote metastasis of lung cancer cells via STAT3-induced EMT
拟南芥	2019.02	Plant Physiology	6.456	The R2R3-MYB transcription factor MYB49 regulates cadmium accumulation
水稻	2019.01	Plant Biotechnology Journal	6.305	The basic helix-loop-helix transcription factor, OsPIL15, regulates grain size via directly targeting a purine permease gene OsPUP7 in rice
斑马鱼	2019.04	Nature Communications	12.124	Primary cilia regulate hematopoietic stem and progenitor cell specification through Notch signaling in zebrafish
斑马鱼	2019.05	Protein & Cell	6.228	Loss-of-function of sox3 causes follicle development retardation and reduces fecundity in zebrafish
拟南芥	2019.05	Molecular Plant	9.326	mTERF5 Acts as a Transcriptional Pausing Factor to Positively Regulate Transcription of Chloroplast psbEFLJ

Small RNA测序

产品概述

Small RNA是生物体内一类重要的功能分子,主要包括miRNA、siRNA和piRNA,通过各种序列特异性的基因沉默作用,包括RNA干扰(RNAi)、翻译抑制、异染色质形成等,诱导基因沉默,调控诸如细胞生长发育、应激反应、沉默转座子等各种各样的细胞生理过程。小RNA测序技术采用胶分离技术,收集样品中18~30nt的RNA片段,利用UMI建库技术、高通量测序技术,一次性获得单碱基分辨率的数百万条小RNA序列信息,依托强大的生物信息分析平台,鉴定小RNA,并预测其靶标基因。

产品优势



全面的样品类型
真核、原核生物、血清血浆、FFPE样品、外泌体、RIP富集RNA等多种类型。



更低的样品总量
小RNA样本实现最低1 ng起始量,低起始量建库成功率达95%。



更准确的测序结果
建库引入UMI技术,实现无偏向的精准定量。



更丰富的分析内容
采用Dr.Tom多组学数据挖掘系统,10大数据库注释,多维度结果图片展示,数据图表循环挖掘。

产品应用



致病调控机理研究



动植物组织器官
分化和发育的研究

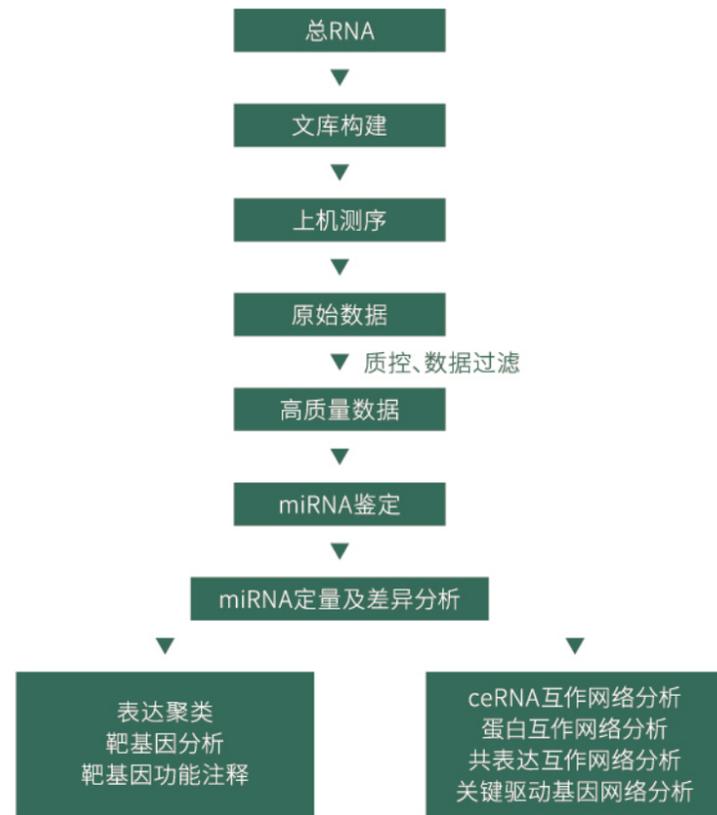


动植物真菌共生研究



原核细菌特性研究

技术流程



技术参数

样品类型: 动植物总RNA样品、微生物总RNA样品、200nt以内的小片段RNA样品、经切胶回收的 small RNA样品; 组织培养与细胞系样品、血清/血浆样品、组织液样品, FFPE样品、RIP样品等;

样品最低需求量(单次): $\geq 1\text{ng}$;

测序策略: SE50;

推荐数据量: 20M clean reads;

项目执行周期: 标准流程的执行周期为30个工作日(样品数小于10个)。

案例分析

案例一: miRNA机制研究——寄生昆虫衍生的miRNA调节宿主发育

★ 案例描述

寄生蜂会产生毒液、多DNA病毒(PDV)、畸形细胞等,改变宿主的生理状态有益于后代存活,但改变的潜在分子机制仍不清楚。本研究发现寄生蜂(*C. vestalis*)的畸形细胞、多DNA病毒(嵌入到寄生蜂基因组中特有的共生病毒)CvBV可以产生miRNA并通过不同方式将其传递到宿主中。寄生宿主小菜蛾(*P. xylostella*)中某些miRNA主要由畸形细胞产生,而寄生宿主中由CvBV编码的miRNA的表达水平比非寄生宿主高100倍。一种畸形细胞产生的miRNA(Cve-miR-281-3p)和一种CvBV产生的miRNA(Cve-miR-novel22-5p-1)通过调节宿主蜕皮激素受体(EcR)的表达来阻止宿主生长。研究结果显示miRNA在动物寄生中可以进行跨物种调控,及寄生期间宿主生理改变中miRNA的功能。

★ 结果展示

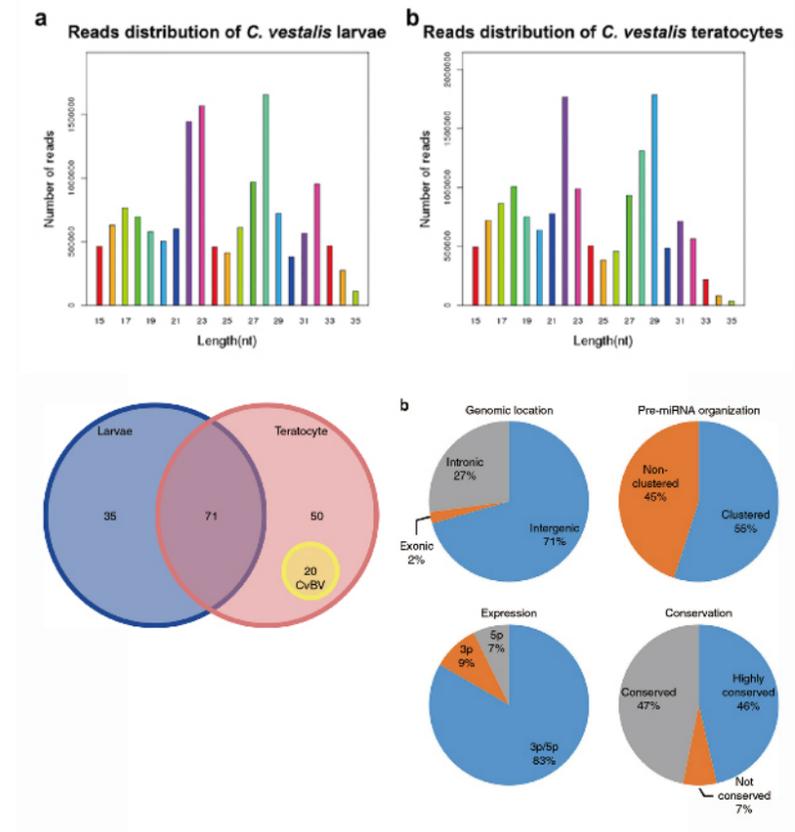


图1 寄生蜂幼虫和畸形细胞中的miRNA特征

研究内容

标准信息分析

1. 数据产出统计,对原始信息采集数据去接头污染,去低质量reads
2. small RNA信息采集结果的长度分布
3. small RNA在选定的参考基因组上的分布
4. small RNA分类统计
5. miRNA定量分析
6. miRNA差异表达分析
7. miRNA表达/差异表达聚类分析
8. miRNA靶基因分析
9. 差异miRNA靶基因GO注释和KEGG通路分析

高级信息分析

- 数据库注释
1. 转录因子注释(AnimalTFDB/PlantTFDB)
 2. Rfam、Pfam、Reactome、COG、EggNOG和InterPro数据库注释

互作网络分析

1. 靶基因分析
 - ① miRNA-mRNA靶向关系分析
 - ② lncRNA-mRNA靶向关系分析
2. ceRNA互作网络分析
 - ① lncRNA-mRNA联合分析
3. 蛋白互作网络分析
4. 共表达互作网络分析

特色分析

1. 外部数据库关联分析(TCGA、ARCHS4)
2. 关键驱动基因网络图分析
3. 时间序列分析

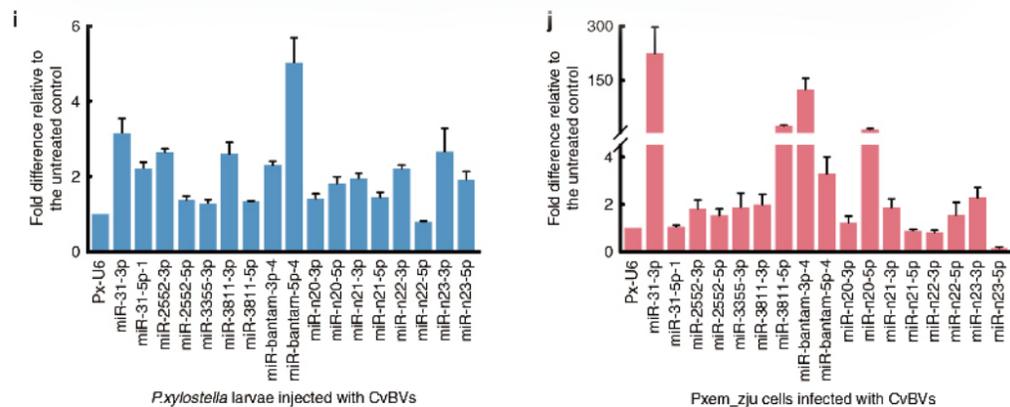


图2 CvBV编码的miRNA在CvBV感染的宿主小菜蛾中的表达

★ 参考文献

Wang Z, Ye X, Shi M, et al. Parasitic insect-derived miRNAs modulate host development[J]. Nature communications, 2018, 9(1): 2205.

▶▶ 常见问题

Q: 华大小RNA可以操作什么类型的样品建库测序?

A: 动植物总RNA样品、微生物总RNA样品、200nt以内的小片段RNA样品、经切胶回收的small RNA样品、组织培养与细胞系样品、血清/血浆样品、组织液样品, FFPE样品、RIP样品等, 只要满足送样要求, 均可用于小RNA建库测序。

Q: 进行small RNA测序对组织样品提取总RNA有什么特殊要求?

A: 提取总RNA时建议不要使用Qiagen等公司的过柱试剂盒, 也不要使用LiCl沉淀, 以免丢失小片段RNA。如果直接提供small RNA样品, 可以使用small RNA提取专用试剂盒来进行提取。

Q: miRNA验证方法?

A: 茎环实时定量PCR-Stem-loop quantitative real-time PCR (qRT-PCR)、Quantitative PCR (qPCR)、qRT-PCR等。

Q: Small RNA分析需要合作伙伴提供参考序列么?

A: 需要, 最好是全基因组序列, 如果没有, 也可以提供近源物种的全基因组序列或者是本物种的EST等序列作为参考序列, 另外还需要老师提供相关的exon/intro/ repeat信息以及基因编码序列。

Q: 原核样本小RNA建库起始量要求多少?

A: 最少1μg。

▶▶ 华大发表文献

物种	发表日期	发表期刊	影响因子	文献标题
鹰嘴豆	2018.10	Plant Biotechnol J.	6.305	Integrated transcriptome, small RNA and degradome sequencing approaches provide insights into Ascochyta blight resistance in chickpea.

物种	发表日期	发表期刊	影响因子	文献标题
小菜蛾	2018.06	Nature Communications	12.124	Parasitic insect-derived miRNAs modulate host development
小菜蛾	2018.02	Frontiers in Immunology	6.4	Genome-Wide Identification of Destruxin A-Responsive Immunity-Related MicroRNAs in Diamondback Moth, <i>Plutella xylostella</i>
荔枝	2017.12	New phytologist	7.3	Coupling of microRNA-directed phased small interfering RNA generation from long noncoding genes with alternative splicing and alternative polyadenylation in small RNA-mediated gene silencing
蜜蜂	2017.08	Plos Genetics	6.1	Plant microRNAs in larval food regulate honeybee caste development
棉花	2017.07	The Plant Journal	5.9	microRNAs involved in auxin signalling modulate male sterility under high temperature stress in cotton (<i>Gossypium hirsutum</i>)
大麦	2017.06	Molecular Biology and Evolution.	6.2	Small RNA activity in archaeological barley shows novel germination inhibition in response to environment
小麦	2017.03	New Phytologist	7.3	<i>Puccinia striiformis</i> f. sp. <i>tritici</i> microRNA-like RNA 1(Pst-milR1), an important pathogenicity factor of Pst, impairs wheat resistance to Pst by suppressing the wheat pathogenesis-related 2 gene
拟南芥	2017.03	The Plant Journal	5.9	microRNA/microRNA* complementarity is important for the regulation pattern of NFYA5 by miR169 under dehydration shock in <i>Arabidopsis</i>
小鼠	2017.02	Cell Death and Disease	6.0	MicroRNA-144 is regulated by CP2 and decreases COX-2 expression and PGE2 production in mouse ovarian granulosa cells
大豆	2015.03	The Plant Cel	9.6	Evolutionary patterns and coevolutionary consequences of MIRNA genes and microRNA targets triggered by multiple mechanisms of genomic duplications in soybean.
植物	2015.03	Mol Plant	6.6	To bloom or not to bloom: role of microRNAs in plant flowering.

物种	发表日期	发表期刊	影响因子	文献标题
鞭毛虫	2014.09	PNAS	9.4	Both endo-siRNAs and tRNA-derived small RNAs are involved in the differentiation of primitive eukaryote Giardia lamblia
小鼠	2014.08	Journal of Autoimmunity.	8.4	DNA methylation and mRNA and microRNA expression of SLE CD4+ T cells correlate with disease phenotype
文昌鱼	2009.07	Genome Biology	10.8	Identification and characterization of novel amphioxus microRNAs by Solexa sequencing
飞蝗	2009.01	Genome Biology	10.8	Characterization and comparative profiling of the small RNA transcriptomes in two phases of locust

长链非编码RNA(lncRNA)测序

▶ 产品概述

长链非编码RNAs (long non-coding RNAs, lncRNAs) 是一类长度大于200nt且不编码蛋白质的RNA, 参与了X染色体沉默、基因组印记以及染色质修饰、转录激活、转录干扰、核内运输等多种重要的调控过程, 涉及到表观遗传调控、转录调控以及转录后调控等多个层面, 是非编码RNA(non-coding RNAs, ncRNAs) 的重要组成部分。lncRNA测序是指使用特定方法对样本中大于200nt的非编码RNA进行测序研究, 从而快速、全面、准确地获得与特定生物学过程(例如发育、疾病等)相关lncRNA信息的研究方法。

▶ 产品优势



技术稳定
同一样本重复建库相关性Pearson值大于0.993。



信息全面
基于高通量测序技术, 一次性获得样本中几乎全部的lncRNA信息。



性价比高
一次测序可以得到mRNA、lncRNA、circRNA信息。



经验丰富
已完成万例lncRNA项目, 涉及人鼠及动植物物种。

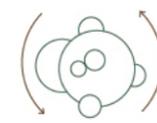


分析交互
采用Dr.Tom多组学数据挖掘系统, 10大数据库注释, 多维度结果图片展示, 数据图表循环挖掘。

▶ 产品应用



细胞分化和发育的研究



调控机理的研究



疾病标志物的寻找

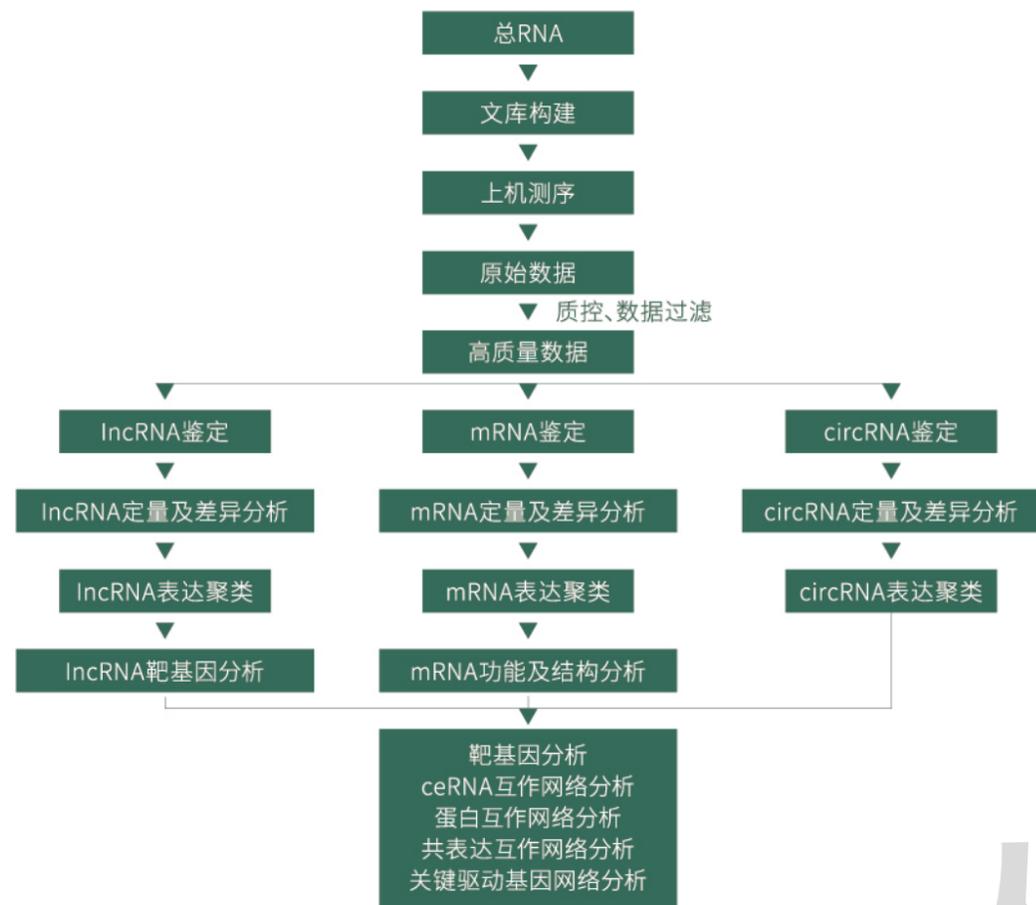


疾病的分子诊断



基因药物的研发

技术流程



技术参数

样品类型: 人、动植物总RNA样品; 组织培养与细胞系样品、FFPE样品等;
 样品最低需求量(单次): $\geq 200\text{ng}$;
 测序策略: PE100;
 推荐数据量: 10G clean data;
 项目执行周期: 标准流程的执行周期为40个工作日(小于10个样本)。

案例分析

案例一: 棉花表型调控lncRNA——研究棉花衣分率和绒毛纤维

★ 案例描述

棉纤维是纺织工业重要的天然纤维。纤维起始数决定棉花衣分率,其是棉纤维产量的重要指标。虽然已通过转录组方法对棉纤维发育进行了分析,但长非编码RNA(lncRNA)在棉绒和绒毛纤维起始调控方面的分子机制仍不甚清楚。本研究将Xu142与无纤维突变体Xu142 fl杂交,获得了三个具有不同衣分率的品系。研究人员从开花后(DPA)0天和5天的Xu142、无纤维突变体Xu142 fl以及三个不同衣分率品系的胚珠(附着纤维)中收集表皮细胞,以进行深度转录组测序。结果共鉴定了2641个新基因、35802个lncRNA和2262个环状RNA(circRNA),其中645个lncRNA优先在无纤维突变体Xu142 fl中表达,651个lncRNA优先在纤维附着品系中表达。研究人员通过病毒诱导的基因沉默(VIGS)实验证明三种lncRNA参与棉纤维发育过程。本研究结果表明,沉默X142 fl中的XLOC_545639与XLOC_039050增加了胚珠的纤维起始数,但沉默Xu142中的XLOC_079089则导致纤维变短。

★ 结果展示

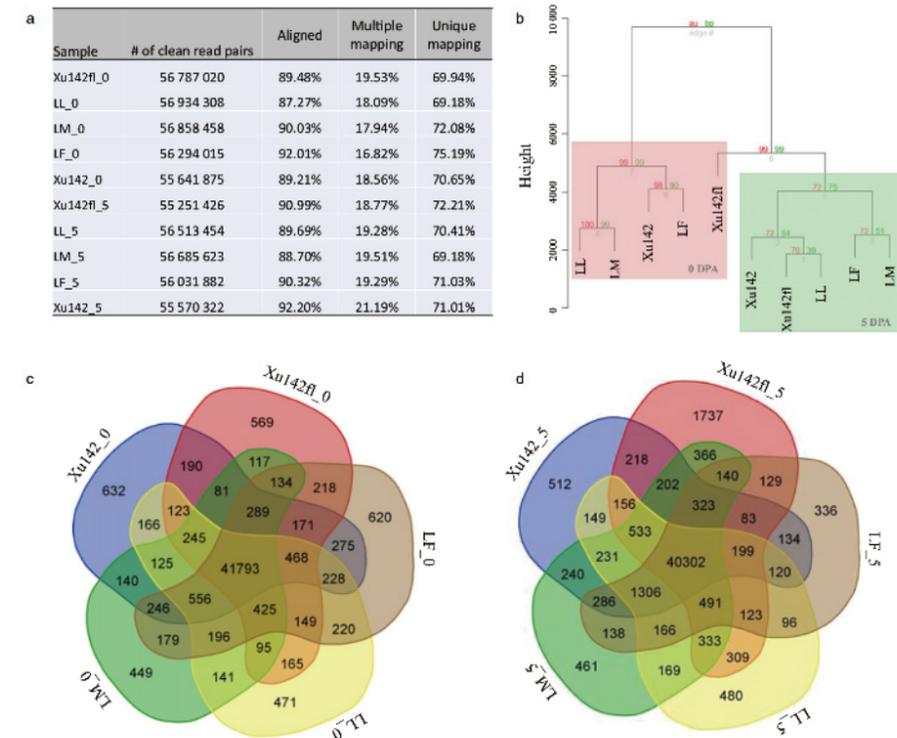


图1 测序数据分析

a. 10个样本的clean reads数及基因组比对率 b. 10个样本聚类分析
 c-d. 0 DPA和5 DPA时期父母本及子代表达基因维恩图

信息分析内容

标准信息分析

1. 基本数据统计
 - ① 去除接头序列、低质量序列得到reads信息
 - ② 样品相关性
 - ③ 表达量分布
 - ④ RNA分类
2. 参考基因组比对
3. lncRNA、mRNA、circRNA鉴定
4. lncRNA、mRNA、circRNA定量分析
5. lncRNA、mRNA、circRNA差异表达分析(样本间、组间)
6. lncRNA、mRNA、circRNA表达/差异基因聚类
7. mRNA差异基因GO分类、富集
8. mRNA差异基因KEGG分类、富集
9. mRNA结构分析
 - ① 可变剪切分析
 - ② SNP/Indel分析

高级信息分析:

1. 数据库注释
 1. 转录因子注释(AnimalTFDB/PlantTFDB)
 2. Rfam、Pfam、Reactome、COG、EggNOG和InterPro数据库注释
 2. 互作网络分析
 1. 靶基因分析
 - ① miRNA-mRNA靶向关系分析
 - ② lncRNA-mRNA靶向关系分析
 2. ceRNA互作网络分析
 - ① lncRNA-mRNA联合分析
 - ② circRNA-mRNA联合分析(仅限人、小鼠)
 3. 蛋白互作网络分析
 4. 共表达互作网络分析
- 特色分析
1. 外部数据库关联分析(TCGA、ARCHS4)
 2. 关键驱动基因网络图分析
 3. 时间序列分析

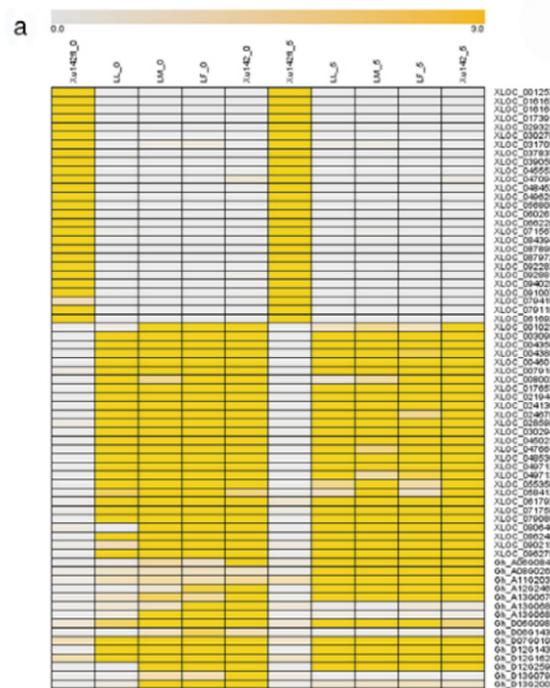


图2 与纤维合成调控相关lincRNA进行表达量热图分析

► 华大发表文献

物种	发表日期	发表期刊	影响因子	文献标题
葡萄	2019.04	Scientific Reports	4.122	Identification and functional prediction of cold-related long non-coding RNA (lincRNA) in grapevine
秀丽隐杆线虫	2019.01	Genome Biology	13.214	Systematic evaluation of <i>C. elegans</i> lincRNAs with CRISPR knockout mutants
蓖麻	2018.07	The Plant Journal	5.775	Differential expression networks and inheritance patterns of long non-codingRNAs in castor bean seeds
棉纤维	2018.05	Plant Biotechnol J.	6.305	Transcriptomic repertoires depict the initiation of lint and fuzz fibres in cotton (<i>Gossypium hirsutum</i> L.)
草莓性状	2018.03	Plant Cell Physiol	4.76	Comparative Transcriptome Profiling Analysis of Red- and White-Fleshed Strawberry (<i>Fragaria</i> × <i>Ananassa</i>) Provides New Insight into the Regulation of Anthocyanins Pathway
山羊繁育	2017.09	PLoS One	2.766	Identification and analysis of differentially expressed long non-coding RNAs between multiparous and uniparous goat (<i>Capra hircus</i>) ovaries
枳	2017.02	Scientific Reports	5.228	Genome-wide screening and characterization of long noncoding RNAs involved in flowering development of trifoliolate orange (<i>Poncirus trifoliata</i> L. Raf.)
猕猴桃生长发育	2016.03	Front. Plant Sci.	4.497	Comprehensive Transcriptome Profiling Reveals Long Noncoding RNA Expression and Alternative Splicing Regulation during Fruit Development and Ripening in Kiwifruit (<i>Actinidia chinensis</i>).
酵母	2016.09	PLoS One	2.766	Identification of Non - Coding RNAs in the <i>Candida parapsilosis</i> Species Group
毛果杨	2015.06	Journal of Experimental Botany	5.677	Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in <i>Populus trichocarpa</i>

★ 参考文献

Hu H, Wang M, et al. Transcriptomic repertoires depict the initiation of lint and fuzz fibres in cotton (*Gossypium hirsutum* L.)[J]. *Plant Biotechnol J.* 2018 May;16(5):1002-1012. doi: 10.1111/pbi.12844.

► 常见问题

Q: 什么是长链非编码RNA? 以及它们的作用?

A: 哺乳动物基因组序列的约5%~10%被稳定转录,蛋白质编码基因仅约占1%,其余4%~9%的序列都转录为非编码RNA。而非编码RNA(non-coding RNA) 是指不能翻译为蛋白的功能性RNA分子。长链非编码RNA为这4%~9%中长度大于200nt的非编码RNA。

它们的作用主要体现在四个方面:

- 第一,影响周边基因的表达;
- 第二,调控蛋白质活动及定位;
- 第三,产生小分子RNA;
- 第四,对其他RNA的调控作用。

Q: 长非编码RNA的建库方案是什么?

A: 长非编码RNA建库主要采用去除rRNA,建链特异性文库。

Q: lincRNA测序的数据中是否也含有mRNA、circRNA?

A: 是的。所以我们推荐在做lincRNA测序时,也可利用同一套测序数据进行mRNA、circRNA的分析。

Q: 怎么用分子生物学实验方法来验证分析结果?

A: 因为转录本组装的复杂性,我们推荐使用传统的克隆或PCR的方法验证分析结果,差异表达的lincRNA也可以用RT-PCR的方式进行验证。

产品
简介

Product
Profile

表观组学

01 | 全基因组甲基化测序 (WGBS) 

02 | ChIP-seq 



全基因组甲基化测序 (WGBS)

产品概述

DNA甲基化是重要的表观遗传学标记信息,获得全基因组范围内所有C位点的甲基化水平数据,对于表观遗传学的时空特异性研究具有重要意义。全基因组甲基化测序将bisulfite处理与高通量测序技术相结合,能够高效准确地绘制全基因组DNA甲基化图谱,实现高精度甲基化修饰模式的分析,在表观基因组学研究中具有里程碑式的意义。该方法能够广泛应用于细胞分化、组织发育等基础机制研究,以及动植物育种、人类健康与疾病等应用性研究。

产品优势



检测精度高
单碱基分辨率,精确分析每一个C碱基的甲基化状态。



性价比高
相对于传统的PCR+sanger测序法费用少。



效率高、耗时少
借助新一代高通量测序平台对全基因组5mC信息进行高效扫描。



双链环化建库方式,GC偏向性更低,全基因组覆盖度更平均,甲基化率更真实。

产品应用



个体发育与细胞分化



动植物育种



畜牧养殖与营养

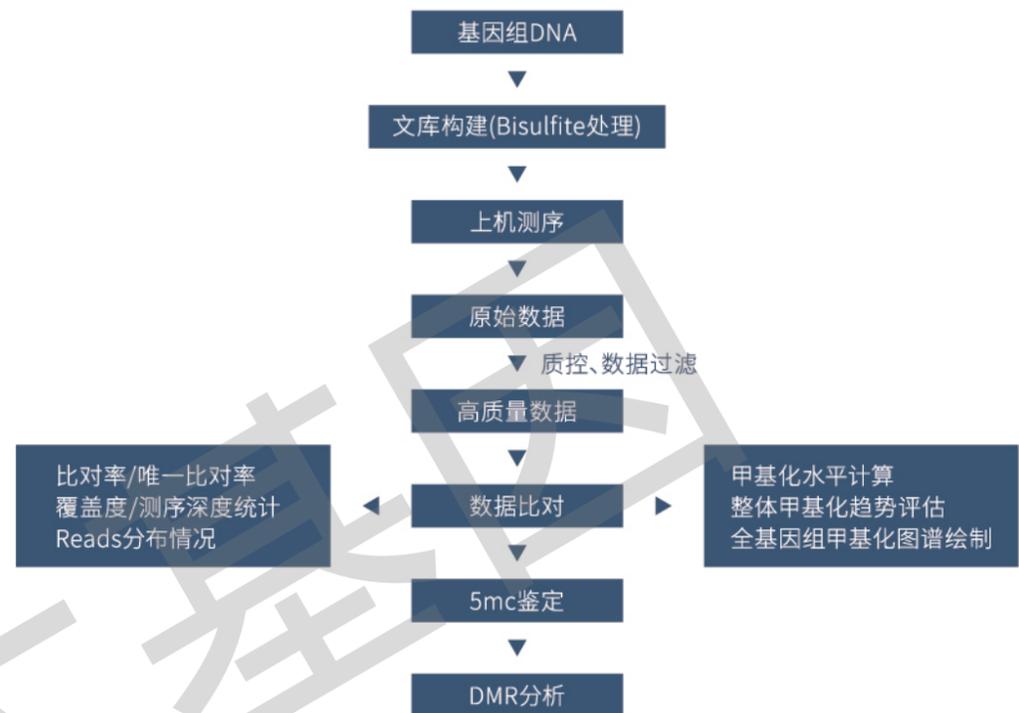


物种及种群进化

技术参数

- 样品总量: ≥ 1 ug;
- 样品纯度: OD260/280 = 1.8~2.0, A260/A230 ≥ 1.6 ; 没有蛋白、多糖和RNA污染;
- 样品完整性: DNA 无明显降解,需提供凝胶电泳检测胶图;
- 样品溶剂: 建议样品用高质量试剂盒提取,溶于ddH2O。确保溶剂内不含有影响酶活的酒精,苯酚,氯仿或其它有机溶剂
- 测序策略: PE101;
- 推荐数据量: $\geq 30\times$;
- 项目执行周期: 40个工作日。

技术流程



研究内容

- 标准信息分析
1. 数据基本处理与质控
 2. 全基因组甲基化水平分析
 3. 甲基化C碱基中CG、CHG与CHH的分布比例
 4. 甲基化CG、CHG和CHH的甲基化水平分布
 5. 甲基化的CG、CHG、CHH附近碱基的序列特征分析
 6. 染色体水平的甲基化C碱基密度分布
 7. 基因组不同转录元件中的DNA平均甲基化水平
 8. 全基因组差异甲基化区域DMR的检测
 9. DMR相关基因的GO和Pathway分析
 10. 其他定制化信息分析(可结合客户的需求,协商确定定制化信息分析内容。)

案例分析

案例一: 鱼类性别改变过程中的甲基化修饰

★ 案例描述

性别决定机制大致可以分为两种,一种是遗传决定性别,一种是环境决定性别。半滑舌鲷的性染色体类型为ZW型,属于遗传决定性别的物种,但是有一部分ZW型的半滑舌鲷个体却表现为雄性。

本研究利用半滑舌鲷作为研究对象,通过对雄性、雌性、假雄性以及其后代进行了甲基化测序,深入阐述了DNA甲基化在半滑舌鲷性别改变过程中的机理,并发现DNA甲基化对与半滑舌鲷的性别决定因素是可以稳定遗传给后代的。

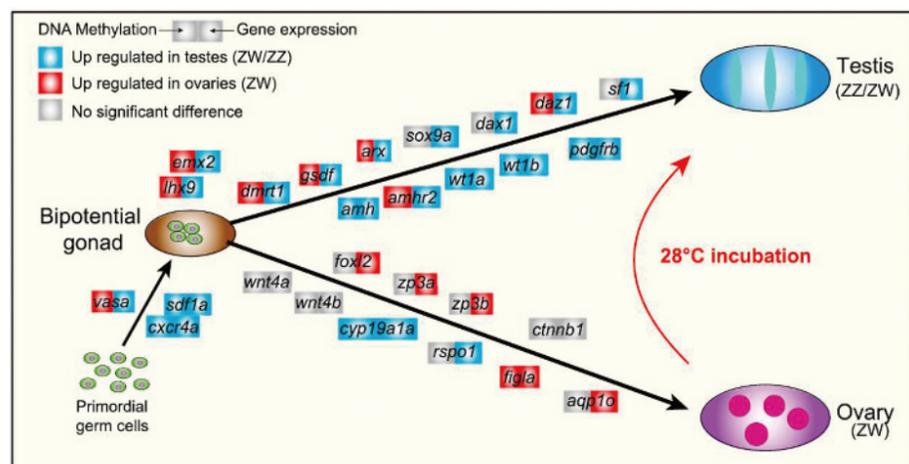


图1 候选基因中DNA甲基化与基因调控情况

▶ 华大发表文献

物种	文章标题	发表期刊
半滑舌鳎	Epigenetic modification and inheritance in sexual reversal of fish.	Genome Res. 2014
蚂蚁	Genome-wide and caste-specific DNA methylomes of the ants <i>Camponotus floridanus</i> and <i>Harpegnathos saltator</i>	Curr Biol. 2012
线虫	Differential DNA methylation in discrete developmental stages of the parasitic	Genome Biol. 2012
黄曲霉	Bisulfite sequencing reveals that <i>Aspergillus flavus</i> holds a hollow in DNA methylation.	PLoS One. 2012
水稻	Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression	BMC Genomics. 2012
家蚕	Single base-resolution methylome of the silkworm reveals a sparse epigenomic map.	Nat Biotechnol. 2010

★ 参考文献

Shao C, Li Q, Epigenetic modification and inheritance in sexual reversal of fish. *Genome Res.* 2014 Apr;24(4): 604-15. doi: 10.1101/gr.162172.113. Epub 2014 Feb 2.

▶ 常见问题

Q: Bisulfite-Seq在项目开始之前需要考虑哪些因素?

A: 是否为低甲基化率的物种:

该物种的基因组完成情况如何(影响BS-SEQ的比对):

基因组是否存在复杂因素: GC含量偏高、杂合度偏高、转座子、重复区域等。

Q: 可以对无参考基因组的物种进行Bisulfite研究吗?

A: Bisulfite-Seq强烈依赖基因组的完成程度,基因质量的好坏直接影响后续的分析结果,因此更适合有完整基因组信息的物种。

Q: Bisulfite的转化率是多少?

A: Bisulfite转化率达到99%以上: 如果样品的DNA没有不发生甲基化的DNA作为对照,都会在样品中混入control DNA来验证Bisulfite的转化率。

ChIP-seq

产品概述

染色质免疫共沉淀(ChIP)是在体内环境中研究蛋白质与DNA相互作用的经典实验方法,广泛应用于组蛋白修饰、特定转录因子的基因调控作用等相关领域。随着新一代测序技术的发展和成熟,染色质免疫沉淀实验与高通量测序的整合——Chromatin Immunoprecipitation Sequencing (简称ChIP-Seq),可在全基因组范围对蛋白结合位点进行高效而准确的筛选与鉴定,同时也为进一步的深入研究打下基础。

ChIP-Seq 技术采用特异性抗体对目的蛋白进行免疫沉淀,分离出与蛋白结合的基因组DNA片段,通过高通量测序与数据分析,在全基因组范围内寻找与目的蛋白结合的DNA位点,并可基于多个样品进行差异比较。

产品优势



数据利用率高
SE50最优读长,没有数据浪费。



性价比高
基于抗体富集目标区域,有效降低测序数据量。



检测范围广
全基因组范围内的DNA与蛋白作用区域进行测序分析。



微量样品
BGISEQ-500平台最低起始量仅需5ng,5ng以下亦可做定制化。

产品应用



个体发育与细胞分化



动植物育种



畜牧养殖与营养



物种及种群进化

技术参数

样品类型: ChIPed DNA 样品(未经PCR 扩增);

样品总量: $\geq 10\text{ng}$;

样品浓度: $\geq 1\text{ng}/\mu\text{L}$;

样品纯度: $\text{OD}_{260}/\text{OD}_{280} = 1.8\sim 2.0$;

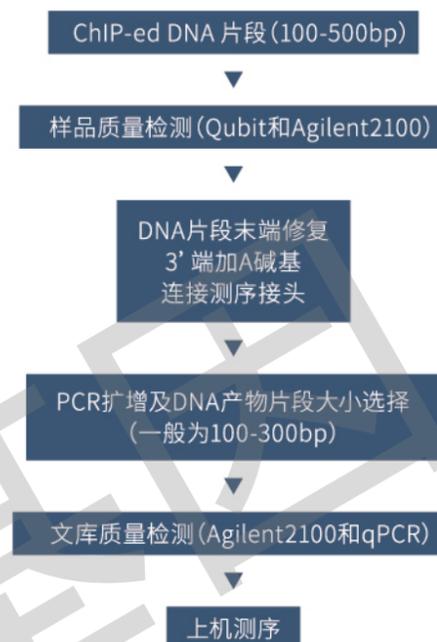
DNA 片段大小: 分布在100~500bp范围,且主带明显。请提供DNA打断后的检测胶图以确定DNA片段大小是否符合要求; 并请附加一份详细的样品信息单,及提供ChIP后的q-PCR验证结果;

测序策略: SE50;

推荐数据量: 20M或40M高质量reads;

项目执行周期: 标准流程的执行周期为30个工作日。

技术流程



研究内容

标准信息分析

1. 数据过滤质控
 - a) 数据产出统计
 - b) 质量、碱基比例分布图
2. ChIP 测序序列与参考基因组序列的比对(基于1)
 - a) 比对结果统计
 - b) 基因组测序深度累积分布
 - c) 基因组测序深度分布
3. Peak 分析(基于2)
 - a) Peak扫描
 - b) Peak长度分布
 - c) Peak深度分布
4. Peak注释(基于3)
 - a) Peak在基因功能元件上的分布
 - b) Peak相关基因分析
 - c) Peak相关基因的GO功能显著性富集分析
 - d) Peak相关基因的Pathway功能显著性富集分析
5. 鉴定样品间差异Peak(基于2,3)
对两个样品进行差异分析,确定存在样品间差异修饰的区间。
6. 样品间差异Peak注释(基于5)
 - a) 差异Peak的基因功能元件分布
 - b) 差异Peak相关基因分析
 - c) 差异Peak相关基因的GO功能显著性富集分析
 - d) 差异Peak相关基因的Pathway功能显著性富集分析
7. 免费提供UCSC genome browser 使用说明(基于2,3)

高级信息分析

1. Motif 分析(基于3)

▶▶ 案例分析

案例一: 玉米基因表达的表现调控机理研究

玉米是一种十分重要的粮食作物,其基因组十分复杂,并且在表观遗传学以及进化上有着极其悠久的历史。本文以玉米品种B73为对象,研究了玉米根与幼苗组织的表观修饰(组蛋白甲基化、组蛋白乙酰化、DNA甲基化等)、mRNA和SmRNA的表达情况,并深入研究了DNA甲基化、组蛋白修饰等表观修饰对基因表达的调控作用。

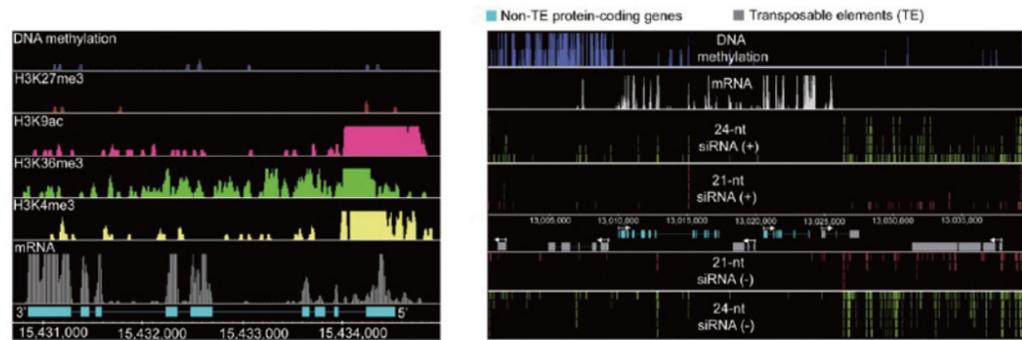


图2 玉米表观组、转录组及小RNA图谱

▶▶ 华大发表文献

物种	文章标题	发表期刊
玉米	Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize.	Plant Cell, 2009
方法学	Single-tube linear DNA amplification (LinDA) for robust ChIP-seq	Nat Methods, 2011

★ 研究结果

1. 编码蛋白基因的转录活性与 H3K4me3 /H3K4me2的比率大小正相关;
2. 另外发现了两类新的siRNA;
3. 三种组蛋白H3K4me3, H3K9ac, H3K36me3在编码蛋白的功能基因上丰度很高,而在转座子相关基因上很少,并且对基因表达起到正调控的作用。

★ 参考文献

Wang X, Elling AA, Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell*. 2009 Apr;21(4): 1053-69. doi: 10.1105/tpc.109.065714. Epub 2009 Apr 17.

▶▶ 常见问题

Q: N-ChIP和X-ChIP的区别是什么?

A: N-ChIP基于内切酶micrococcal nuclease (MNase) 酶切,切割核小体,适用于组蛋白的研究; X-ChIP基于化学交联,适用于大多数研究情况。

Q: ChIP-Seq是否需要做阴性对照测序?

A: 一般情况下,建议选择Input作为对照进行测序。

Q: 样品制备过程中是否需要进行PCR扩增? PCR扩增后是否会影响最终结果?

A: 由于ChIP下来的DNA样品量通常非常少,所以在样品建库制备过程中需要经过一步PCR扩增,主要是为了获得足够上机反应的DNA量。如果提供的DNA样品足量,则可减少PCR循环数或不进行PCR扩增。PCR扩增可能会增加结果的偏向性。

Q: 影响ChIP-Seq结果的因素有哪些?

A: 抗体的质量与特异性、需要富集的目标区域在基因组上的比例、ChIP的实验操作、DNA片段长度范围等都会影响ChIP-Seq的结果。