

低深度全基因组测序加速破译家猪经济性状的遗传结构

华大智造MGISEQ-2000测序平台对猪的重要经济性状进行遗传结构解析

中国农业大学胡晓湘教授与华南农业大学吴珍芳教授团队基于多年研究积累，基于华大智造MGISEQ-2000测序平台，开发了杜洛克猪低深度全基因组测序分析方案，并通过高通量测序技术解析了猪的重要经济性状遗传结构。

该研究成果于2021年以“Accelerated deciphering of the genetic architecture of agricultural economic traits in pigs using a low-coverage whole-genome sequencing strategy”¹为题发表于*GigaScience*。

推荐应用：分子育种

推荐机型：MGISEQ-2000RS

- 数据检测准确性高

DNBSEQ 测序技术可大幅提高检测的准确性，且其低深度重测序(Lc-WGS)基因分型相比其他方法准确性更高¹。

- 实验周期短

华大智造自动化建库方案可大幅缩短实验周期，并减轻手动建库的误差，极大满足分子育种的时效要求。



背景介绍

作为世界最大的猪肉消费国与生猪生产国，我国生猪引进品种高达 90%，面对生猪产业“卡脖子”的严峻现状，促进优良品系育种至关重要。近年来，伴随高通量测序技术的高速发展，物种全基因组测序、功能基因组学以及生物信息学得到迅猛发展，然而，全基因组高成本的特点限制其市场化与产业化转化。低深度全基因组测序方法 (low-coverage whole-genome sequencing, lcWGS) 为畜牧业和食品业的品系选育提供了新的选择²。

低深度全基因组测序方法是指以大约 1× 或更少的测序深度执行低覆盖率全基因组测序^{3, 4}，然后通过填充来恢复丢失的基因型 / 位点，以确保所有个体都拥有一组共享变异的基因类型。这种方法已经用于人类和一些动物物种的全基因组关联研究和基因组选择 / 预测，并被证明是高深度测序的可行替代方案⁵。总体而言，低深度全基因组测序是一种新的低成本的测序标记方法。研究表明，就评估群体参数而言，对大量个体进行低深度全基因组测序这一方案相比在较高测序深度对较少个体采样的方案更为准确⁶。

研究描述

研究团队收集了来自同一育种场的 2869 头杜洛克公猪，随后完成基因组文库构建并利用华大智造 MGISEQ-2000 测序平台进行了平均 0.73× 的低深度测序。采用了 BaseVar-Stitch 流程进行 Reference Panel 构建及基因型填充。并用 3 种不同分型方法 (高精度测序、SNP 芯片、Fluidigm 基因分型) 对低深度数据进行准确性对比，并对不同参数下样本量和测序深度对准确性的影响进行对比。通过低深度测序对 21 个猪重要经济性状进行了全基因组关联分析和遗传结构解析，并系统分析人工选择对于杜洛克猪选育过程中基因组结构变化产生的影响(图 1)。

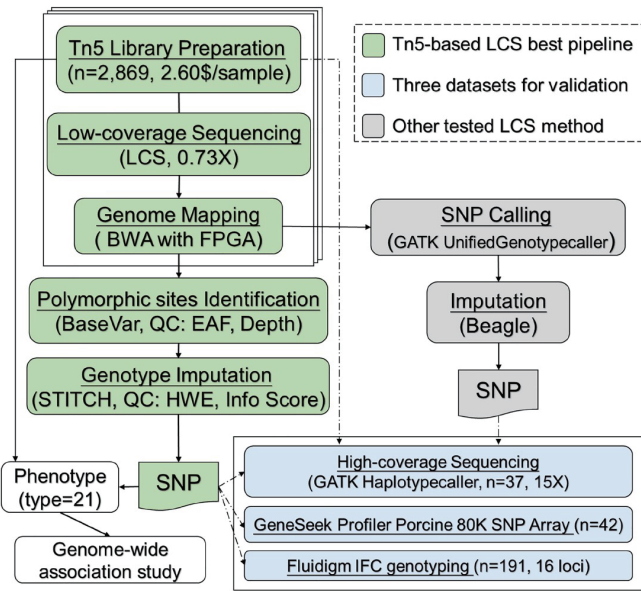


图1.低深度全基因组测序方案设计

实验方法

样本采集

本研究采集了来自同一育种场的2869头杜洛克公猪,使用 DNeasy Blood & Tissue Kit (Qiagen 69506)试剂盒从猪耳部组织中提取基因组DNA。

文库制备与测序

投入 50 ng gDNA 随后对 gDNA 进行打断 。使用 KAPA HiFi HotStart Ready Mix (Roche)进行PCR 扩增;扩增引物对基于华大智造测序仪进行设计。其中,反向引物包含96种不同的Barcode序列,用以区分各个文库。将文库制备成DNA纳米球 (DNB) 后,在MGISEQ-2000或BGISEQ-500基因测序仪上进行双端100碱基读长测序 (PE100)。

数据分析

使用 GTX-ALIGN 将来自低覆盖度样本的测序数据比对到 Sscrofa11.1 参考基因组上,应用 GATK 中的插入序列重排和碱基质量重校准模块对插入序列候选基因座周围的数据进行重排,并对碱基质量进行重新校正。使用 vcfTools 程序、BioMart 工具进行群体遗传学分析。使用混合线性模型方法进行基于标记 SNPs 的全基因组关联分析。

样本采集	文库制备和测序	生信分析	结果分析
来自同一育种场的 2869头杜洛克公猪	<div></div> WGS文库构建 MGISEQ-2000基因测序仪	GTX-ALIGN vcfTools BioMart	不同平台基因分型效果对比 基于Lc-WGS的分析流程性能评估 解析重要经济性状遗传结构

结果

基因分型效果对比

华大智造测序平台在基因分型中表现出了更低的 duplication 率，高数据产出率，低数据损失率的优势（表 1）。因此，后续研究采用 MGISEQ-2000 进行正式的方案设计。

	illumina	MGI
platform	X-ten	MGISEQ-2000
Samples/lane	96	84
Bases number	~136G	~151G
Reads number	~455M	~755M
Good index bases	~109G	~138G
Total coverage	98.51%	98.55%
Ave coverage	29.03%	42.56%
Ave depth	0.40	0.63
Mapping rate	99.28%	97.86%
PCR duplication	12.64%	4.27%

表1.Illumina平台和华大智造测序平台基因分型效果对比

基于 Lc-WGS 的 BaseVar -STITCH 分析流程性能评估

基于低深度测序数据的 BaseVar STITCH 分析流程相较于 15x 全基因组测序，可得到高准确性基因型 ($R^2=0.919$, $GC=0.970$)，准确性高于基于 GATK-Beagle 分析流程结果 ($R^2=0.484$, $GC=0.709$) (图 2 A)。BaseVar STITCH 结果相较于 GGP-80 数据 (SNP 芯片) 显示出更高的 GC 一致性和 R 值 ($R^2=0.997$, $GC=0.990$) (图 2 B)。基于 Fluidigm 基因分型 (16 个位点, 191 个个体)，平均 GC 为 0.991, BaseVar STITCH 数据与其数据基本接近。

得出结论 BaseVar STITCH 流程是一种新的分析方法,适用于 LcWGS 策略的变异检测和基因填充。当样本量 >500 时，采用 0.5×WGS 数据进行 STITCH 分析时对结果几乎没有影响。当样本量增加至 1985 时，测序深度为 0.1×WGS 时的数据与 0.5×WGS 测序结果一致 (图 2C 和 D)。总体而言，随着测序深度或样本量的增加，结果的可信度将会大幅提升，而当单个位点的总测序深度 >200× 时就可保证研究中基因型的高度可信。

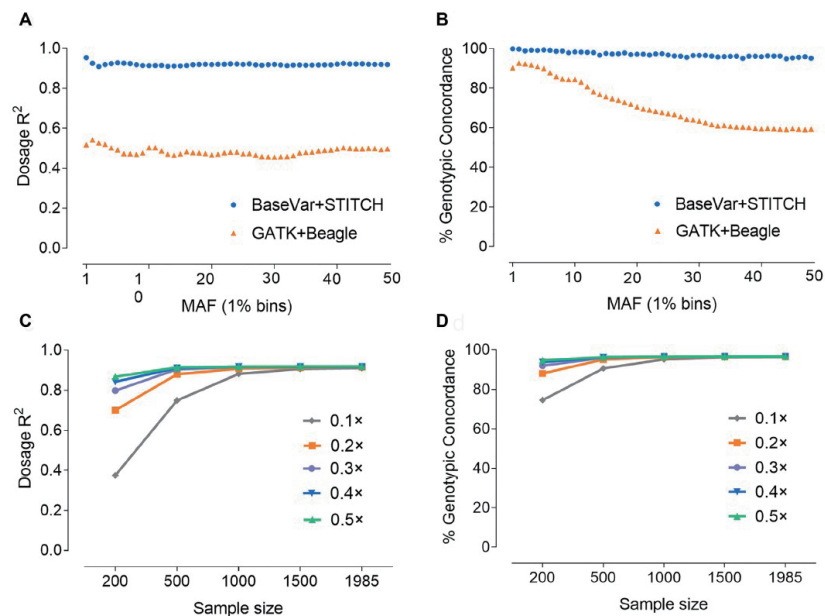


图2.BaseVar-STITCH在不同次等位基因频率(MAFs)和样本量中的分析性能。(注:计算基因型与估算剂量之间的相关性(R^2)和基因型一致性(GC),以评价基因型的准确性)

解析猪重要经济性状遗传结构

通过对 21 个猪重要经济性状进行了 GWAS 和遗传结构解析 (图 3)，并系统解析了人工选择在育种过程中对猪的基因组结构产生的影响。对鉴定到的主效 QTL 位点，利用进一步的高密度标记法鉴定到与猪乳头数相关的主效候选基因 ABCD4、与背膘厚性状相关的主效候选基因 HMGA1；对于遗传力高但是有多个微效基因控制的性状。鉴定得

到了大量重要神经通路中影响采食行为的微效基因，并进一步验证了数量性状经典的“无穷小模型”，为下一代分子育种提供了理论基础。此外，本研究还发现由于该群体经受长期人工选择，生长类性状相关的性状表现出 QTL 固定、遗传力丢失变化，说明该父系群体在前期杂交育种中生长性状的选育取得的显著成效。

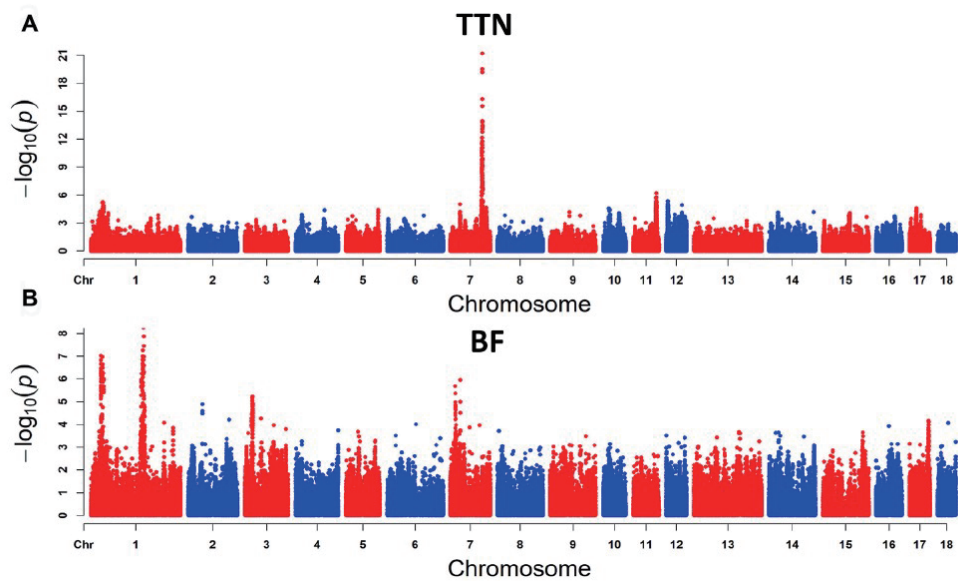


图3.总乳头数 (TTN) 和背膘厚 (BF) 的曼哈顿图。A和B描述TTN和BF在全基因组上的关联信号。

结论

该研究首次发明了采用低深度测序的 BaseVar-Stitch 基因分型流程, 获得了目前杜洛克猪最大群体 (2869 头) 的高密度 SNP 标记集 (11.7M), 并且成本极低。采用三种不同方法分型准确性测评均超过 99%, 证明了本研究创制的大样本低深度测序策略对传统基于小样本高深度的方案具有优越性。如结合华大智造高通量自动化建库工作站 - 硬件加速 - 算法优化等策略, 大幅度降低从生物样本到基因型的实验周期。

同时, 利用 GWAS 深入开展了关于猪生长、胴体及采食三类性状的遗传参数和遗传结构的研究。对主效控制猪乳头数 QTL 的性状, 利用高密度标记一步法鉴定到候选主效基因 ABCD4、对主效控制背膘厚 QTL 的性状鉴定到候选主效基因 HMGA1; 对于遗传力高但由微效基因控制的性状, 鉴定到了大量影响采食行为的重要神经网络中的微效基因, 并进一步证实了数量遗传经典的“无穷小模型”, 为分子育种提供了理论基础; 该群体表现出生长类性状相关的性状表现出 QTL 固定、遗传力丢失变化, 说明了该父系群体在前期杂交育种中生长性状的选育取得的显著成效。

本研究采用低深度测序技术的方式, 填补了在没有优质参考单倍型数据库的畜禽物种中的高效高质填充 SNP 方法, 对育种中重要的“卡脖子”问题提出解决方案, 展示了其在功能基因定位和重要经济性状遗传结构解析潜在的巨大价值, 为推动我国十四五规划种业振兴行动方案提供了完全自主知识产权的新型遗传分析方法和育种理论依据。



基因测序仪 MGISEQ-2000RS

基于华大智造自主研发的 MGISEQ-2000 高通量测序平台能够极大地降低测序成本, 得益于华大智造的 DNBSEQ 非线性扩增技术, MGISEQ-2000 可以保持约 2%-3% 的低冗余度, 为您的科学研究提供理想工具。

参考文献

1. Yang, R. *et al.* Accelerated deciphering of the genetic architecture of agricultural economic traits in pigs using a low-coverage whole-genome sequencing strategy. *GigaScience* 10, doi:10.1093/gigascience/giab048 (2021).
2. Casellas, J., Martín de Hijas-Villalba, M., Vázquez-Gómez, M. & Id-Lahoucine, S. Low-coverage whole-genome sequencing in livestock species for individual traceability and parentage testing. *Livestock Science* 251, 104629, doi:https://doi.org/10.1016/j.livsci.2021.104629 (2021).
3. Nosková, A. *et al.* Characterization of a haplotype-reference panel for genotyping by low-pass sequencing in Swiss Large White pigs. *BMC genomics* 22, 290, doi:10.1186/s12864-021-07610-5 (2021).
4. Buckley, R. M. *et al.* Best practices for analyzing imputed genotypes from low-pass sequencing in dogs. *Mammalian genome : official journal of the International Mammalian Genome Society* 33, 213-229, doi:10.1007/s00335-021-09914-z (2022).
5. Zhao, C. *et al.* Towards a Cost-Effective Implementation of Genomic Prediction Based on Low Coverage Whole Genome Sequencing in Dezhou Donkey. *Frontiers in genetics* 12, 728764, doi:10.3389/fgene.2021.728764 (2021).
6. Lou, R. N., Jacobs, A., Wilder, A. P. & Therikildsen, N. O. A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular ecology* 30, 5966-5993, doi:10.1111/mec.16077 (2021).

推荐订购信息

产品类型	产品名称	产品货号
仪器	基因测序仪 MGISEQ-2000RS	900-000035-00
	MGISP-100RS 自动化样本制备系统	900-000070-00
	MGISP-960RS 自动化样本制备系统	900-000100-00
软件	MegaBOLT 生信分析加速器(工作站式服务器)	970-000085-00
建库试剂	MGI Easy 通用 DNA 文库制备试剂套装(16RXN)	1000006985
测序试剂	MGISEQ-2000RS 高通量测序试剂套装(FCL PE100)	1000012554

深圳华大智造科技股份有限公司

深圳市盐田区北山工业区综合楼11栋



4000-688-114



www.mgi-tech.com



MGI-service@mgi-tech.com

股票简称：华大智造

股票代码：688114



仅供研究使用

版权声明：本手册版权属于深圳华大智造科技股份有限公司所有，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识均属于深圳华大智造科技股份有限公司及其提供者所有。

版本：2023年8月版

撰稿：李艳萍 隋佳原

责任编辑：王其伟

审稿：江遥