

文章编号: 1004-0374(2011)04-0390-04

猪的全基因组水平SNP研究现状

陈慧勇¹, 罗赛群¹, 吴珍芳^{2*}

(1 中南大学生物科学与技术学院分子生物学研究中心, 长沙 410078; 2 华南农业大学动物科技系, 广州 510640)

摘要: SNP(single nucleotide polymorphism, 单核苷酸多态)在猪基因组中的分布极其广泛, 平均分布间隔为 300~400 bp, 相关数据库收录已达 55 万条。猪基因组测序已取得实质性进展, 大规模搜索发现基因组及 EST(expressed sequence tag) 序列中的 SNP 已展开, 应用于猪全基因组水平的 SNP 芯片已建立。在此基础上, 基于猪 SNP 标记的遗传图谱绘制、QTL(quantitative trait loci) 定位、遗传多样性检测及全基因组关联分析等也都相继出现。

关键词: 猪; SNP; 全基因组关联分析

中图分类号: Q959.842; Q78 **文献标识码:** A

Progresses of SNP in pig genome-wide studies

CHEN Hui-Yong¹, LUO Sai-Qun¹, WU Zhen-Fang^{2*}

(1 Molecular Biology Research Center, College of Biological Science and Technology, Central South University, Changsha 410078, China; 2 Department of Animal Science, South China Agricultural University, Guangzhou 510640, China)

Abstract: Single nucleotide polymorphisms (SNPs) are abundant and widespread throughout the pig genome, occurring typically on average every 300-400 bps. About 557 k SNPs have been deposited in NCBI's SNP repository dbSNP. The pig whole genome is currently being sequenced with billions of sequences completed. Hundreds of thousands of porcine SNPs are discovered in genome sequence data and expressed sequence tag (EST) collections. High density SNP genotyping assays are designed for porcine genome. Therefore, the extensive applications of porcine SNP are developed, some of which include constructing of genetic linkage map, mapping of quantitative trait loci (QTL), detection of genetic diversity, genome-wide association study.

Key words: pig; SNP; genome wide association study

单核苷酸多态 (single nucleotide polymorphism, SNP), 是基因组水平的单个核苷酸变异所引起的 DNA 序列多态性, 主要指单个碱基的替换, 包括转换和颠换。由于分布的广泛性和稳定性及检测手段的自动化, SNP 逐渐成为基因组连锁作图、精细定位及构建单倍型的最佳遗传标记。利用 SNP 进行的人类复杂疾病全基因组关联分析 (genome-wide association studies) 于 2005 年开始, 至今已有上百万篇的科研论文发表, 这些庞杂的研究结果将给相关疾病的预防、诊断和治疗带来新的期望和挑战^[1]。在此背景下, 各种动物和植物的全基因组关联研究也相继开始。猪在农业和医学领域有着独特而重要

的地位, 它的基因组数据积累已有很大的突破。作为全基因组关联研究的基础, 猪的 SNP 发现和利用研究进展目前尤为迅速。

1 猪SNP数据的积累

据估计, 猪的 SNP 覆盖密度可以达到每

收稿日期: 2010-09-30; 修回日期: 2010-12-02

基金项目: 国家高技术研究发展计划(“863”计划)(2008AA101008); 高等学校博士学科点专项科研基金/新教师基金项目(20090162120026)

*通讯作者: E-mail: wzfemail5000@163.com; Tel: 0766-2986523

300~400 bp 即有 1 个^[2]。截至 2010 年 9 月 17 日, NCBI 的 SNP 数据库 (<http://www.ncbi.nlm.nih.gov/SNP>) 收录的猪 SNP 已近 55.7 万条。大规模的猪 SNP 发现工作始于 2002 年, 由美国农业部发起, 他们通过对可以表达的基因进行体外 PCR 扩增、比较测序等实验手段共发现了 1 650 个 SNP^[3]。近几年, 猪 SNP 数据的积累尤其迅速。

猪的基因组约 2.7 Gb。中国—丹麦合作 (Sino-Danish initiative) 完成的草图覆盖度达 0.66^[4], 可以从 NCBI 网站 (<http://www.ncbi.nlm.nih.gov/sites/entrez>) 下载, 这些数据来自 5 个不同的猪种 (杜洛克、二花脸、汉普夏、长白和大白猪)。Sanger 研究所 (SGSC pig genome sequencing project) 公布的基于克隆测序的基因组数据则处于一种实时更新状态 (http://www.sanger.ac.uk/Projects/S_scrofa/)^[5]。同时, 美国农业部和爱荷华州立大学为首的相关组织也开展了相关的工作, 其第 10 版的猪基因组数据 (Build 10) 也于 2010 年 9 月 22 日公布 (<ftp://ftp.tgac.bbsrc.ac.uk> 或 <http://www.animalgenome.org/blast/>)。另外, 还有法国 INRA (INRA-Genescope) 的数据 (<http://www.piggenome.org>), 包含了伊比利亚猪、长白猪、眉山猪、小型猪、皮特兰、野猪及大白猪等 7 个猪种的信息。

伴随着猪基因组数据库的充实、新的生物信息学分析手段及测序技术的出现, 全基因组范围内快速搜索 SNP 工作几乎同步进行。这方面荷兰瓦格宁根大学做的工作比较多: 2009 年, 他们利用公开的前两个猪基因组数据库进行比对, 发现了近 10 万个 SNP^[6]; 利用 Illumina 及 454 测序技术分析了不同品种猪 (杜洛克、皮特兰、大白、长白及野猪) 的 19 个基因组文库及 2 个 DNA 池样品, 参照猪基因组数据 Build 7 作比对分析, 共发现了 37 万多个 SNP^[7]; 利用 5 头长白与皮特兰的杂交后代构建基因组 DNA 池, 结合 Illumina 测序技术共发现了 17 489 个 SNP^[8]。2008 年, 美国农业部利用焦磷酸测序技术分析了 26 个个体 DNA 池, 共发现了 11 万多个

SNP^[9]。

而伴随 Pig EST database (<http://pigest.geome.iastatae.edu>) 和 Pig EST (<http://pigest.kvl.dk>) 等猪 EST 数据库资源的丰富, 寻找 EST 中 SNP 的工作也在开展。2007 年, 丹麦奥胡斯大学 Panitz 等^[10]在 EST 序列的编码区发现了 7 900 个 SNP。另外, Hayes 等^[11]利用 EST 数据库中的冗余数据找到 676 个经得起反复验证的 SNP。这些编码区的 SNP (coding SNP, cSNP) 功能明确, 分析利用起来相对容易和经济。

2 高通量 SNP 检测技术

SNP 检测技术有很多, 如对 PCR 产物直接测序、质谱或液相色谱分析、对已有的 EST 或基因组数据库进行序列比对、等位基因特异性杂交、等位基因特异性引物延伸以及与酶连接或酶外切反应相结合的技术等, 原理多种多样。表 1 中列了部分 SNP 高通量实验检测技术, 在猪的相关研究中已得到应用。

这些技术有的检测通量突出, 如 Affymetrix 及 Illumina 公司的 SNP 芯片, 在现代全基因组范围内搜索或研究猪的 SNP 有着重要意义。2008 年, Bischoff 等^[12]设计了独特 Affymetrix 寡核苷酸微阵列, 分析比较出眉山猪种中分布独特的 cDNA 序列 SNP。而 Illumina 公司 2009 年的 PorcineSNP60 一共包含了 6 万多个 SNP 位点, 每个包装含有 4~96 张芯片, 检测的样品数从 48 到 1 152 不等^[7]。有的则可以根据实验者的研究内容, 随意确定待测的 SNP 位点, 如 SNPlex 检测的 SNP 可由研究者自身来设计特定探针^[13,14], 而 Affymetrix 芯片检测 SNP 位点则是由芯片事先设计好的。有的一次性可以完成多个样本的多个位点的检测, 如 SNPstream 若借助于 384 孔板, 目前至少可以一次性检测 384 个样品的 12 个 SNP 位点^[15], 而 SNPlex 技术的一个泳道可以完成 48 个 SNP 的检测, 毛细管泳道数越多检测的样本数也就越多, 如 3730 基因分析仪有 384 个

表1 高通量SNP实验检测技术

关键原理	代表技术	一次检测SNP数	一次检测样本数	SNP的选择性
ASO	Affymetrix SNP芯片	上万个	1个	已固定
ASO-酶外切反应	Taqman技术	1个	96或384(受微量滴定板孔数限制)	任意选择
ASO-酶连接反应	SNPlex技术	48个	96(受毛细管泳道数限制)	任意选择
ASPE-酶连接反应	Illumina SNP芯片	上万个	1个	任意选择
SBE-酶连接反应	SNPstream技术	12或48个	384个	任意选择

ASO(allele-specific oligonucleotide hybridization): 等位基因特异性杂交; ASPE(allele-specific primer extension): 等位基因特异性引物延伸; SBE(single-base extension): 单碱基延伸

泳道数,那么就可以一次性完成384个样品的48个位点检测。Taqman技术虽然一个反应只检测一个SNP位点,但借助相关的平台一次可以完成多个样本的检测,而且在检测的SNP选择上比较灵活。对此,研究人员要根据自身的实验内容及检测的通量和成本来选择出最佳的检测方法。

3 猪SNP的应用现状

SNP由于分布的广泛性和检测手段的便捷,被认为是第三代遗传标记,它正被应用于猪遗传图谱的绘制、QTL的精细定位、性状的关联分析,乃至物种进化、遗传多样性或个体身份的分析 and 鉴定等。

猪最早的遗传图谱绘制于1995年^[16,17],主要借助于微卫星或RFLP标记。网络版的遗传图谱中又加入了少量的AFLP和SNP标记(<http://www.thearkdb.org>)。2009年,丹麦奥胡斯大学成功地利用462个SNP构建了猪18个常染色体新一代的遗传图谱,密度是每3.94 cM一个标记^[18]。这些SNP来源于440个不同的基因,位于基因的编码区或5'和3'非编码区,与相关的基因呈现一种高度的连锁不平衡状态,它们的应用将会极大地促进QTL的定位研究,乃至关键基因的揭示。所以,丹麦奥胡斯大学很快便进行了肉质性状的QTL定位工作,找到相关连锁的SNP标记,为下一步的标记辅助选择做准备^[19]。

SNP与复杂性状的关联分析揭示相关的遗传机理是其最常见的应用,这方面的报道很多,尤其是研究特定基因在猪的生长、繁殖、胴体、肉质及抗病或发病性状中的决定意义,此外还有全基因组水平的关联分析在近两年也开始了。2010年,荷兰猪遗传所利用Illumina公司的PorcineSNP60进行了公猪脂肪组织雄烯酮水平的关联分析,结果揭示了两个重要的染色体控制区域^[20]。由SNP组成的单倍型用于复杂性状的关联分析往往包含更多的信息。一个由34个SNP组成的27 kb染色体区域的单倍型变化就被用于猪在家养与人工选择下的遗传变异情况研究^[21]。Meidtnier等^[22]利用PPARD基因内部的SNP单倍型进行背膘厚的遗传控制机理分析,发现了一个关键的单碱基控制位点,并进行了表达水平的验证。国内,上海交通大学王起山博士^[23]对中国眉山猪保种群群的ESR基因三个SNP位点进行单倍型分析,发现了对产仔数影响较大的一个单倍型。另外,于2010年9月完成验收的中国农业大学等单位主持的“973”项目“猪、鸡重要

经济性状遗传的分子机制”及广东温氏食品集团等单位主持的“863”项目“猪分子细胞工程育种技术创新与优势性状新品系培育”也涉及了部分内容。随着猪SNP数据的积累、高通量检测技术的成熟和高密度SNP基因组图谱的建立,猪全基因组范围内的关联研究也势必形成影响。

同时,SNP也是遗传多样性或进化分析及品种或亲缘关系鉴定的重要基础。SNP标记用于猪个体身份鉴定的对应计算方法已建立^[24]。Rohrer等^[25]曾成功地利用60个SNP进行商业化猪群中个体身份的鉴定,它的敏感度高于10个微卫星标记的联合使用。全世界最大的一个猪遗传多样调查计划(European Union pig diodiversity project II, Pig-BioDiv II)调查了100多个中国和欧洲猪种,以期对相关遗传资源的保护提供参考,目前已进行了十多年,其中使用的标记除了微卫星外还包括大量的SNP标记^[26]。

目前,猪全基因组范围内的SNP积累已达到一定的水平,相关的实验技术和统计算法也部分建立,它的大范围应用即将到来。但我们也要看到猪的相关研究积累与人的还存在很大的差距,比如人的SNP数据收集量相对要庞大许多,更新也异常及时,SNP的染色体定位信息明确,对应的SNP检测芯片也有各种系列以满足不同疾病不同检测规模研究的需要。猪的SNP芯片产品目前比较少,其包含的SNP许多还没有进行染色体定位,所以利用目前的SNP芯片技术还不足以进行猪数量性状的QTL定位研究。另外,也必须考虑这样的现实:基因在发挥作用时,它是多个层次的,有基因组、转录及翻译等各个水平的调控,每个水平的调控都与环境存在或多或少的联系,且基因与基因间还存在复杂的网络联系;基因的变异形式也是多样的,而不仅仅是单个核苷酸的改变,同时也是动态的,有的基因在同一个个体的不同组织或不同阶段可能会有多种结构。所以,在挖掘和利用这些丰富的SNP标记进行猪的复杂性状关联分析时,我们必须谨慎,不能一味地跟踪研究热潮。

[参 考 文 献]

- [1] Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med*, 2010, 363(2): 166-76
- [2] Jungerius BJ, Gu J, Crooijmans RP, et al. Estimation of the extent of linkage disequilibrium in seven regions of the porcine genome. *Anim Biotechnol*, 2005, 16(1): 41-54

- [3] Fahrenkrug SC, Freking BA, Smith TP, et al. Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. *Anim Genet*, 2002, 33(3): 186-95
- [4] Wernersson R, Schierup MH, Jørgensen FG, et al. Pigs in sequence space: a 0.66X coverage pig genome survey based on shotgun sequencing. *BMC Genomics*, 2005, 6(1): 70
- [5] Humphray SJ, Scott CE, Clark R, et al. A high utility integrated map of the pig genome. *Genome Biol*, 2007, 8(7): R139
- [6] Kerstens HH, Kollers S, Kommadath A, et al. Mining for single nucleotide polymorphisms in pig genome sequence data. *BMC Genomics*, 2009, 10: 4
- [7] Ramos AM, Crooijmans RP, Affara NA, et al. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One*, 2009, 4(8): e6524
- [8] Amaral AJ, Megens HJ, Kerstens HH, et al. Application of massive parallel sequencing to whole genome SNP discovery in the porcine genome. *BMC Genomics*, 2009, 10: 374
- [9] Wiedmann RT, Smith TP, Nonneman DJ. SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC Genet*, 2008, 9: 81
- [10] Panitz F, Stengaard H, Hornsh H, et al. SNP mining porcine ESTs with MAVIANT, a novel tool for SNP evaluation and annotation. *Bioinformatics*, 2007, 23(13): 1387-91
- [11] Hayes BJ, Nilsen K, Berg PR, et al. SNP detection exploiting multiple sources of redundancy in large EST collections improves validation rates. *Bioinformatics*, 2007, 23(13): 1692-3
- [12] Bischoff SR, Tsai S, Hardison NE, et al. Identification of SNPs and INDELS in swine transcribed sequences using short oligonucleotide microarrays. *BMC Genomics*, 2008, 9: 252
- [13] Pindo M, Vezzulli S, Coppola G, et al. SNP high-throughput screening in grapevine using the SNPlex genotyping system. *BMC Plant Biol*, 2008, 8: 12
- [14] Tobler AR, Short S, Andersen MR, et al. The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *J Biomol Tech*, 2005, 16(4): 398-406
- [15] Bell PA, Chaturvedi S, Gelfand CA, et al. SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques*, 2002, Suppl: 70-2, 74, 76-7
- [16] Archibald AL, Haley CS, Brown JF, et al. The PiGMap consortium linkage map of the pig (*Sus scrofa*). *Mamm Genome*, 1995, 6(3): 157-75
- [17] Rohrer GA, Alexander LJ, Hu Z, et al. A comprehensive map of the porcine genome. *Genome Res*, 1996, 6: 371-91
- [18] Vingborg RK, Gregersen VR, Zhan B, et al. A robust linkage map of the porcine autosomes based on gene-associated SNPs. *BMC Genomics*, 2009, 10: 134
- [19] Li HD, Lund MS, Christensen OF, et al. Quantitative trait loci analysis of swine meat quality traits. *J Anim Sci*, 2010, 88(9): 2904-12
- [20] Duijvesteijn N, Knol EF, Merks JW, et al. A genome-wide association study on androstrenone levels in pigs reveals a cluster of candidate genes on chromosome 6. *BMC Genet*, 2010, 11: 42
- [21] Ojeda A, Huang LS, Ren J, et al. Selection in the making: a worldwide survey of haplotypic diversity around a causative mutation in porcine IGF2. *Genetics*, 2008, 178(3): 1639-52
- [22] Meidtner K, Schwarzenbacher H, Scharfe M, et al. Haplotypes of the porcine peroxisome proliferator-activated receptor δ gene are associated with backfat thickness. *BMC Genet*, 2009, 10: 76
- [23] 王起山. 动物数量性状候选基因的单体型分析方法研究[D]. 上海交通大学, 2007: 44-53
- [24] Hill WG, Salisbury BA, Webb AJ. Parentage identification using single nucleotide polymorphism genotypes: application to product tracing. *J Anim Sci*, 2008, 86(10): 2508-17
- [25] Rohrer GA, Freking BA, Nonneman D. Single nucleotide polymorphisms for pig identification and parentage exclusion. *Anim Genet*, 2007, 38: 253-8
- [26] Chen K, Baxter T, Muir WM, et al. Genetic resources, genome mapping and evolutionary genomics of the pig (*Sus scrofa*). *Int J Biol Sci*, 2007, 3(3): 153-65