

上海白猪(上系)基因组遗传变异检测与功能注释分析

盛中华^{1,2}, 陈国宏^{1*}, 潘玉春³, 王起山³, 张哲³

(1. 扬州大学动物科学与技术学院, 扬州 225009; 2. 上海市闵行区动物疫病预防控制中心, 上海 201109;
3. 上海交通大学农业与生物学院, 上海 200240)

摘要: 拟通过全基因组范围内遗传变异的检测和注释分析, 深入了解上海白猪(上系)群体的遗传现状, 以便更好地进行提纯复壮与开发利用。本研究应用基因组简化测序平台—GGRS, 对上海白猪(上系)群体(99头)进行基因组测序, 并综合实验室前期太湖地方猪种和西方引进品种共计447头测序数据进行SNP和InDel等遗传变异检测和功能注释分析。结果显示, 上海白猪基因组测序平均覆盖度为2.87%, 平均测序深度为3.9倍, 共检测到328 586个SNPs、693 220个InDels。进一步分析表明, SNP和InDel在染色体上的分布相对均一, 但在不同染色体上数量和密度的分布存在一定差异。结构注释显示, SNP和InDel对应基因的数量分别为11 496个和13 216个, 按照基因的结构区域分类, SNP和InDel在各类功能基因区间的分布特点一致, 绝大多数的SNP和InDel变异都分布在基因间区, 分别为74.61%和83.38%。内含子中次之, 分别为22.76%和14.98%。基因富集分析结果表明, SNP主要与蛋白代谢过程和高分子代谢过程等相关, InDel多在组织、器官发育和疾病相关的通路中发生富集。此外还发现, 含有高密度SNP和InDel等遗传变异的QTL主要集中在肉质与胴体性状和健康性状。本研究利用先进的简化基因组测序技术对上海白猪(上系)全基因组范围内的遗传变异进行了普查, 并通过对太湖流域地方猪种和西方引进品种的合并比较分析, 阐释了这些遗传变异的染色体分布特征, 基因区间分布规律和功能注释信息, 为后续的开发利用奠定了坚实的基础。

关键词: 上海白猪(上系); 猪基因组; SNP; InDel

中图分类号: S828.2

文献标志码: A

文章编号: 0366-6964(2018)05-0887-10

Identification and Functional Annotation of Genome-wide Genetic Variation in Shanghai White Pigs(Shang)

SHENG Zhong-hua^{1,2}, CHEN Guo-hong^{1*}, PAN Yu-chun³, WANG Qi-shan³, ZHANG Zhe³

(1. College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, China;
2. Minhang District Center for Animal Disease Prevention and Control in Shanghai, Shanghai 201109, China;
3. School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: This study intended to understand the genetic status of Shanghai White pig population through detection and annotation analysis of genome-wide genetic variation in order to better purify, rejuvenate and exploit the resource. In this study, genome sequencing was performed on 99 Shanghai White pigs using the genotype by genome simplifying and sequencing platform-GGRS. A total of 447 sequencing data including Taihu local pig breeds and western introduced breeds were collected for SNP and InDel genetic variation detection and functional annotation analysis.

收稿日期: 2017-09-26

基金项目: 上海市农业基层推广体系项目(沪农科推字(2015)第6-2号); 上海市闵行区自然科学研究课题(2015MHZ018)

作者简介: 盛中华(1982-), 男, 上海人, 畜牧师, 博士, 主要从事畜禽种质资源研究, Tel: 021-54460162, E-mail: 5143035@qq.com

* 通信作者: 陈国宏, 教授, E-mail: ghchen@yzu.edu.cn

The results showed that the average coverage of Shanghai White pig genome was 2.87% and the average sequencing depth was 3.9 times. A total of 328 586 SNPs and 693 220 InDels were detected. Further analysis showed that the distribution of SNP and Indel on chromosomes were relatively uniform, but the distributions of number and density on different chromosomes were different. Structural annotation showed that the corresponding gene number of SNP and InDel were 11 496 and 13 216, respectively. According to the structural region classification of the genes, the distribution characteristics of SNP and InDel were consistent in all kinds of functional gene regions. The most of SNP and InDel variation distributed in the intergenic region, with the ratio of 74.61% and 83.38%, respectively, followed by intron, with 22.76% and 14.98%, respectively. Gene enrichment analysis showed that SNPs were mainly related to protein metabolism and macromolecule metabolism. InDel mostly enriched in tissues, organ development and disease-related pathways. In addition, we also found that the QTLs containing high-density SNP and InDel mainly focused on meat quality, carcass and healthy traits. The whole genome genetic variation of Shanghai White pig was investigated by advanced simplified genomic sequencing. The comparative analysis between Shanghai White pig, Taihu pig breeds and western pig breeds explained the chromosomal distribution characteristics, gene distribution and functional annotation information of these genetic variations, which laid a solid foundation for subsequent exploitation and utilization.

Key words: Shanghai White pig(Shang); pig genome; SNP; InDel

上海白猪是通过十几年的选育而培育成的一个地方肉脂兼用型品种,具有肉质好、产仔数较多、胴体瘦肉率较高和耐粗饲等优良特点。早在 1986 年被收录至《中国猪品种志》,2012 年列入《上海市畜禽遗传资源保护名录》。上海白猪(上系)作为上海白猪的一个品系,曾作为供港猪杜长上杂交组合的母本风靡一时,但随着引进的国外猪种严重地冲击国内种猪市场,上海白猪(上系)的群体规模受到了极大的影响,各项生产指标出现一定程度的衰退,因此需要深入了解当前群体的遗传现状,以便更好地进行提纯复壮与开发利用。

遗传变异的检测和功能注释是分析动物群体遗传现状的重要前提和基础,单核苷酸多态性(SNP)和 2 bp~1 kb 之间的小片段插入和缺失(InDel)为猪基因组占比最大的两类分子标记。其中,SNP 因具备充足的信息常用于基因连锁分析和与动植物 DNA 功能性变异密切相关的连锁不平衡分析。如 H. S. Ai 等^[1-2]基于全基因组的 SNP 遗传标记开展了中国部分地方猪的群体结构和连锁不平衡分析等。M. Z. Li 等^[3]利用重测序得到的 SNP 标记,从群体遗传学上进一步阐明了藏猪适应高原环境的遗传机理及杜洛克经人工选择后与藏猪遗传上的差异。InDel 标记具有较好的稳定性和多态性等优

点,频率仅次于 SNP,其中约 1/3 位于已知的基因区域内,还有一些位于决定基因功能的关键性区域如启动子区和外显子区,当前已有多个研究组利用基因芯片开展猪抗病、肉质等相关遗传机制研究^[4-6]。Q. Xiao 等^[7-9]也基于基因组简化测序技术(GGRS)开展了地方猪和引进猪的 SNP 和 InDel 的检测分析。

本研究针对上海白猪核心群体,采用 GGRS 测序技术^[10-11],在全基因组范围内进行 SNP 和 InDel 等遗传变异多态性检测,并进行系统的功能注释分析,以便了解上海白猪当前群体的遗传现状,进而为其保护和利用提供分子生物学依据。

1 材料与方法

1.1 上海白猪及对照群体

试验所用 99 头上海白猪的耳组织样采自上海市闵行区畜禽种场的保种群。同时,为了更准确的开展遗传变异的检测分析和缺失基因型的填补,揭示上海白猪的种质特性,本研究采用与其地理位置邻近的中国地方品种(包括 6 个太湖流域地方品种:梅山猪、二花脸猪、枫泾猪、米猪、沙乌头、嘉兴黑猪)及西方引进品种(包括杜洛克、长白、大白、皮特兰、巴克夏)作为对照品种,共计 447 个个体。

1.2 简化基因组测序

上海白猪采样个体基于 Illumina 公司 HiSeq2000 平台,采用基因组简化测序技术进行双端测序^[10-11]。为了更好的与其它品种进行比较,将上海白猪的测序结果与太湖猪和引进猪种 11 个品种测序结果^[12-14]在一起进行 SNP 和 InDel calling,缺失基因型的填补则利用 Beagle 4.1 来完成^[15-16]。最后,分别在全群和上海白猪群过滤掉最小等位基因频率(MAF)小于 0.05 的 SNP 位点,用于后续分析。

1.3 基因功能区间分布

猪的基因注释数据来自 Ensembl 数据库(Ensembl release 78,ftp://ftp.ensembl.org/pub/release-78/gtf/sus_scrofa/)^[17-18],共包含 25 332 个基因注释信息。根据数据库的信息,本研究主要将基因组区间分为起始或者终止密码子、外显子、内含子、非翻译区(UTR)和基因间区。利用 PERL 脚本编程语言分析 SNP、InDel 是否在上述基因组功能区间。

1.4 基因功能富集分析

位于起始/终止密码子区或者外显子区的变异很有可能改变基因的结构和功能。因此,把位于起始/终止密码子区或者外显子区的变异定义为大效应突变。为了进一步了解各类大效应突变显著参与的通路、生物学功能和参与的分子生物学过程,利用 DAVID v6.7 网络数据库工具及 GO 和 KEGG 数据库^[19-20],对这些与大效应突变相关的基因,按突变类型分别进行功能注释(GO annotation)和通路分析(Pathway analysis)。显著性水平 P 阈值设定为 0.05^[21-22]。针对当前猪的 GO 与 KEGG 注释库还很不全面的问题,利用 PERL 脚本提取猪对应的人同源基因标识用于基因富集分析。

1.5 与 QTL 的映射

猪许多重要性状的 QTL 已相继被定位,并被收集到猪的 QTL 数据库(<http://www.animalgenome.org/>, Release 32, Apr 27, 2017)^[23-24]。截至本研究进行前,该数据库共收集了 12 618 QTLs,基于现有的 QTL 数据库对 SNP 和 InDel 等遗传变异进行功能注释分析。针对有些 QTLs 长度过长,并不能有效地用于后续分析问题,本研究去除了长度超过 1 Mb 的 QTLs,并且把重叠 50% 以上的两个 QTLs 合并成一个新的 QTL,同时设定新 QTL 与其原始性状都相关。此过程利用 QTL 的位置信息与

性状信息,通过 Perl 语言脚本处理完成。

2 结果

2.1 上海白猪(上系)基因组碱基测序质量、覆盖度和深度

经过污染序列的过滤之后,上海白猪共获得大约 430 万条高质量的 reads,两端碱基平均质量得分均在 Q30 以上,即碱基测得的正确率大于 99.9%。其中,最小检测个体和最大检测个体测序 reads 数目分别为 200 万条和 616 万条,个体平均 reads 数目为 409 万条(表 1)。

测序覆盖度是指每个个体测到的基因组总碱基数占该物种基因组总碱基数的百分比,由表 1 可知,每个群体个体的平均覆盖度范围从 1.2%(枫泾)到 3.98%(长白),所有个体的平均测序覆盖度为 2.46%,上海白猪的平均测序覆盖度为 2.87%。测序深度是指基因组上的同一位置被测到的次数。测序深度的计算方法为每个个体测到的高质量 reads 数与唯一位置 reads 数的比值,在本研究中,上海白猪的平均测序深度为 3.90,所有个体的平均测序深度为 4.78 倍,详见表 1。

2.2 上海白猪(上系)基因组变异检测及在染色体上的分布

经 SNP calling 和过滤后,全群共检测到 487 323 个 SNPs、976 235 个 InDels。其中,上海白猪共检测到 328 586 个 SNPs、693 220 个 InDels(表 2)。我们进一步统计了上海白猪各类变异在染色体上的数量和密度分布,发现序列的变异数量一般与染色体的长度相关,染色体长度越长其所含有的变异的数量越多。如表 2 所示,1 号染色上的数量最多,SNPs 和 InDels 分别为 29 857 和 68 997 个,而 Y 染色体上的数量最少,SNPs 和 InDels 分别为 39 和 100 个。SNP 和 InDel 在染色体上的平均密度分别为 12.96/100 kb 和 26.67/100 kb。笔者进一步以 400 kb 为不重叠窗口,图形化展示 SNP 的数量在染色体上分布,可见 SNP 和 InDel 的密度在不同染色体上的分布有较大的差异,同一染色体上,分布相对均匀,具体分布可见图 1 和图 2。

2.3 上海白猪(上系)基因组 SNP 和 InDel 的功能区间分布

当前研究结果显示,上海白猪 SNP 和 InDel 在基因内的数量分别为 11 496 和 13 216 个。基因上各类变异在每个染色体上的数量分布、对应的基因数量见

表 1 测序数据量、覆盖度和深度

Table 1 Distribution of high quality reads number, coverage and depth across population

品种 Breed	样品数 Sample size	平均覆盖度/% Average coverage (percent)	平均测序深度 Average sequencing depth	reads 数量(百万) Number of reads (million)
上海白猪 Shanghai White	99	2.87	3.90	4.30
梅山 Meishan	50	1.87	4.26	3.90
二花脸 Erhualian	31	1.66	7.99	2.83
米猪 Mi	36	2.66	2.80	3.85
枫泾 Fengjing	15	1.20	7.68	4.04
沙乌头 Shawutou	31	1.76	8.37	4.77
嘉兴黑 Jiaying Hei	29	2.17	5.21	3.88
杜洛克 Duroc	48	3.19	3.49	4.18
长白猪 Landrace	37	3.98	3.50	5.10
大白猪 Large White	35	3.47	2.82	3.81
巴克夏 Berkshire	16	2.80	4.74	6.16
皮特兰 Pietrain	20	1.89	2.65	2.00
合计/平均 Total/Mean	447	2.46	4.78	4.09

表 2 SNPs 和 InDels 遗传变异在染色体上的数量及其密度分布

Table 2 The number and density distribution of SNPs and InDels detected on each chromosome

染色体 Chromosome	SNP 数量 Number of SNP	SNP 密度/100 kb Density of SNP	SNP 基因数 Number of genes for SNP	InDel 数量 Number of InDel	InDel 密度/100 kb Density of InDel	InDel 基因数 Number of genes for InDel
1	29 857	9.47	1 078	68 997	21.88	1 238
2	20 757	12.77	1 040	47 393	29.15	1 203
3	22 463	15.51	772	47 818	33.03	901
4	18 282	12.74	655	37 015	25.80	735
5	14 487	12.99	562	29 170	26.16	686
6	28 165	17.85	1 118	55 253	35.02	1 269
7	19 612	14.55	749	41 818	31.03	821
8	13 498	9.09	418	30 866	20.79	487
9	19 906	12.95	645	38 927	25.33	720
10	13 866	17.53	291	24 005	30.35	324
11	9 679	11.04	213	20 358	23.22	249
12	15 275	24.02	618	27 098	42.61	697
13	20 190	9.23	773	50 301	23.01	933
14	25 460	16.55	816	49 864	32.41	908
15	15 534	9.85	469	34 877	22.12	543
16	10 904	12.55	234	20 133	23.17	257
17	11 499	16.50	374	22 920	32.88	424
18	8 837	14.43	275	18 371	30.01	328
X	10 276	7.12	392	27 936	19.36	489
Y	39	2.38	4	100	6.11	4
合计/平均 Total/Mean	328 586	12.96	11 496	693 220	26.67	13 216

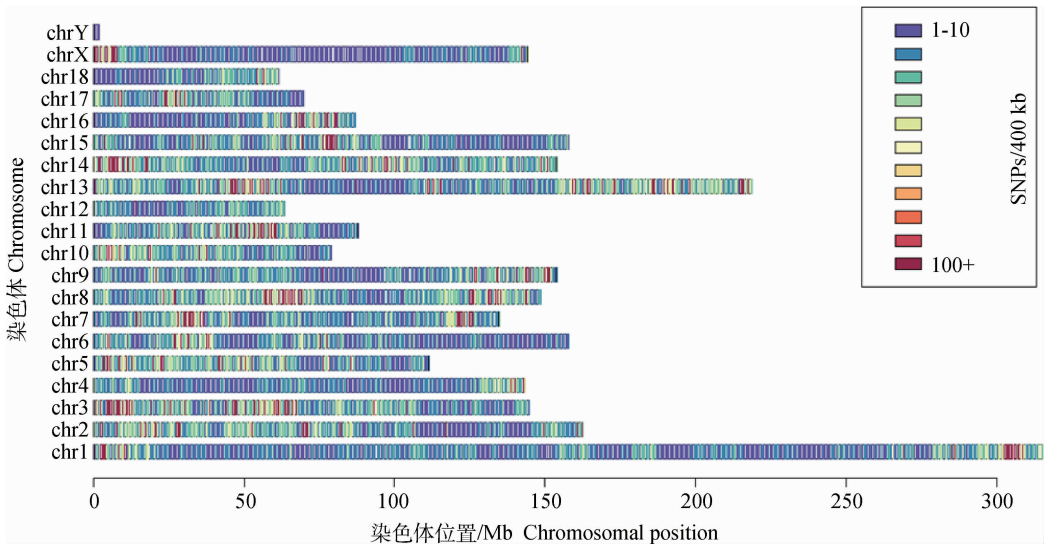


图 1 SNPs 在染色体上的分布

Fig. 1 The location distribution of SNPs across genomes

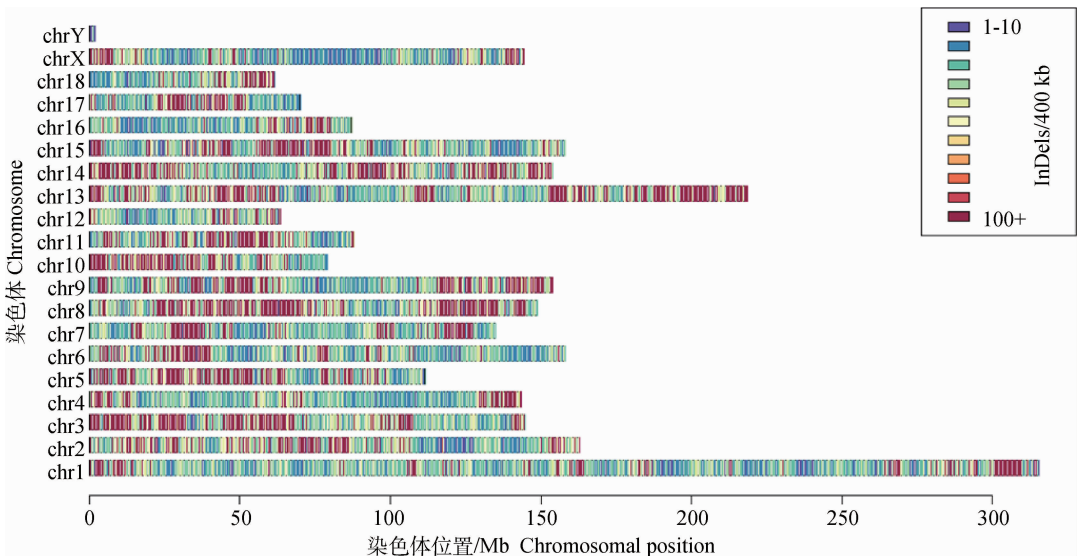


图 2 InDels 在染色体上的物理位置分布

Fig. 2 The location distribution of InDels across genome

表 2。其次,按照基因的结构区域分类,我们进一步对各类变异做了统计分析。结果显示,SNP 和 InDel 在各类功能基因区间的分布特点基本一致。它们在不同的基因区间的数量分布详见表 3。由表 3 可知,绝大多数的 SNP 变异分布在基因间区,约占 74.61%;内含子中比例次之,约为 22.76%;外显子区域中仅占 1.67%。和 SNP 的分布规律一致, InDel 在基因间区的比例也最高,约为 83.38%,内含子中比例次之,约为 14.98%。

2.4 上海白猪(上系)基因组 SNP 和 InDel 的功能注释及通路分析

通过功能注释及通路分析,分别鉴定出了上海白猪各类大效应突变对应的基因显著参与的通路和生物学过程,结果见表 4。SNP、InDel 显著参与的通路/生物过程数目分别为 13 和 40 个。大效应突变的基因富集分析结果显示,不同类型变异富集的通路存在差异。SNP 显著参与的通路大部分是与蛋白高分子合成与调控有关。比如,蛋白代谢过程

(GO:0019538)、高分子代谢过程(GO:0043170)等。而 InDel 多在组织、器官发育和疾病相关的通路中发生了富集,比如单细胞生物过程(GO:0044707)和

致心律失常性右心室心肌病(ssc05412)等。综上,本研究结果初步表明,不同类型的变异可能影响不同的基因网络调控及生物功能。

表 3 SNPs 和 InDels 在不同的基因区间的数量分布

Table 3 The number distribution of SNPs and InDels across genome region

分类 Classification	SNP 数量(百分比/%) Number of SNP (percent)	InDel 数量(百分比/%) Number of InDel (percent)
基因间区 Intergenic region	245 169 (74.61)	578 027 (83.38)
3'-UTR	1 756 (0.53)	2 739 (0.39)
5'-UTR	592 (0.18)	977 (0.14)
外显子 Exon	5 492 (1.67)	6 592 (0.95)
内含子 Intron	74 783 (22.76)	103 869 (14.98)
剪接区 Splice area	84 (0.03)	152 (0.02)
外显子 ncRNA Exon ncRNA	173 (0.05)	189 (0.03)
内含子 ncRNA Intron ncRNA	537 (0.16)	675 (0.97)
合计 Total	328 586	693 220

表 4 各类大效应变异显著富集的 GO 和 Pathway

Table 4 Significant GO and Pathway enriched by the large-effect variants

变异类型 Variant type	功能注释 Functional annotation	基因数 Number of genes	P 值 P-value
SNP	GO:0019538;protein metabolic process	42	2.46E-3
	GO:0043170;macromolecule metabolic process	64	2.63E-03
	GO:0044238;primary metabolic process	68	1.21E-02
	GO:0071704;organic substance metabolic process	70	2.33E-02
	GO:1901576;organic substance biosynthetic process	42	2.34E-02
	GO:0009059;macromolecule biosynthetic process	36	2.38E-02
	GO:0044249;cellular biosynthetic process	41	2.81E-02
	GO:0034645;cellular macromolecule biosynthetic process	35	2.83E-02
	GO:0080090;regulation of primary metabolic process	39	2.87E-02
	GO:0060255;regulation of macromolecule metabolic process	39	3.00E-02
	GO:0009058;biosynthetic process	42	3.37E-02
	GO:0008152;metabolic process	71	4.62E-02
	GO:0019222;regulation of metabolic process	40	4.91E-02
InDel	GO:0044707;single-multicellular organism process	729	3.40E-13
	GO:0044767;single-organism developmental process	689	4.12E-13
	GO:0048856;anatomical structure development	688	1.17E-12
	GO:0032502;developmental process	701	2.63E-12
	GO:0007275;multicellular organism development	601	5.30E-12
	GO:0048518;positive regulation of biological process	665	7.34E-12

(转下页 Carried forward)

(续表 4 Continued)

变异类型 Variant type	功能注释 Functional annotation	基因数 Number of genes	P 值 P-value
	GO:0048731;system development	550	2.43E-11
	GO:0009653;anatomical structure morphogenesis	376	2.86E-11
	GO:0009888;tissue development	263	4.87E-11
	GO:0030154;cell differentiation	481	3.30E-10
	GO:0048869;cellular developmental process	522	4.95E-10
	GO:0016043;cellular component organization	752	2.77E-09
	GO:0007399;nervous system development	280	5.98E-09
	GO:0048522;positive regulation of cellular process	587	1.12E-08
	GO:0048513;animal organ development	407	2.16E-08
	GO:0071840;cellular component organization or biogenesis	771	2.44E-08
	GO:0048468;cell development	281	7.12E-08
	GO:0051128;regulation of cellular component organization	318	1.43E-07
	GO:0022008;neurogenesis	205	2.23E-07
	GO:0048583;regulation of response to stimulus	440	4.13E-07
	ssc04512;ECM-receptor interaction	30	4.78E-06
	ssc04360;Axon guidance	36	7.10E-05
	ssc04510;Focal adhesion	50	1.86E-04
	ssc04974;Protein digestion and absorption	25	6.87E-04
	ssc04151;PI3K-Akt signaling pathway	70	2.87E-03
	ssc04910;Insulin signaling pathway	33	4.43E-03
	ssc05222;Small cell lung cancer	24	4.60E-03
	ssc05100;Bacterial invasion of epithelial cells	21	9.55E-03
	ssc04921;Oxytocin signaling pathway	37	9.64E-03
	ssc04024;cAMP signaling pathway	44	1.06E-02
	ssc05205;Proteoglycans in cancer	41	2.19E-02
	ssc05412;Arrhythmogenic right ventricular cardiomyopathy (ARVC)	18	2.29E-02
	ssc04810;Regulation of actin cytoskeleton	42	2.45E-02
	ssc05414;Dilated cardiomyopathy	21	2.49E-02
	ssc05410;Hypertrophic cardiomyopathy (HCM)	20	2.54E-02
	ssc04015;Rap1 signaling pathway	43	3.16E-02
	ssc04931;Insulin resistance	26	3.26E-02
	ssc04919;Thyroid hormone signaling pathway	26	3.60E-02
	ssc04014;Ras signaling pathway	45	3.79E-02
	ssc05166;HTLV-I infection	50	4.11E-02

2.5 上海白猪(上系)基因组 SNPs 和 InDels 位于 QTL 的映射

经过过滤和合并之后,上海白猪 SNPs 和 InDels在五大类 QTLs 注释性状中的数量分布详见表 5。坐落在不同性状中的各类变异数量统计分析结果显示,SNP 变异主要分布在肉质与胴体性状

(256 573)和健康性状(246 230),外貌性状中最少(205 204);InDel 变异在各个性状中的分布规律与 SNP 基本一致,肉质与胴体性状占比最多(283 936),外貌性状中最少(233 299)。

按照变异是否位于基因上,笔者又进一步开展了相关统计分析。结果显示,SNP 和 InDel 位于基

因上的数量与变异对应的基因落在 QTL 区间上的比例基本一致,即主要分布在肉质与酮体性状和健康性状相关的基因上,而外貌性状中最少。这表明,

与肉质和健康相关的基因可能具有较高多态性,而与外貌性状相关的基因可能相对比较保守,存在的变异较低。

表 5 同时位于 QTL 与基因上的各类变异数量分布

Table 5 The number distribution of variants located in both QTLs and gene regions

性状 Trait	SNP			InDel		
	数量 ¹ Number ¹	数量 ² Number ²	基因数 ³ Number of genes ³	数量 ¹ Number ¹	数量 ² Number ²	基因数 ³ Number of genes ³
生产 Production	234 436	72 860	10 162	260 889	83 262	11 701
繁殖 Reproduction	222 973	67 919	9 456	248 295	77 291	10 885
健康 Health	246 230	76 051	10 428	273 659	86 455	12 003
外貌 Exterior	205 204	63 315	9 120	233 299	73 741	10 554
肉质和胴体 Meat & Carcass	256 573	79 563	10 898	283 936	90 234	12 505

¹. 位于 QTLs 中的变异数量; ². 位于基因上的变异数量; ³. 与变异对应的基因数量

¹. The number of variants located in QTLs; ². The number of variants located in the genes; ³. The number of genes corresponding to the variants

3 讨论

SNP 是一种重要的遗传分子标记,也是当下分子遗传学研究应用最为广泛的分子标记。在本研究中,检测到的 328 586 个 SNPs 中,有接近 6% 为新发现的 SNPs,这表明在上海白猪中有着自己特异性,其基因组上仍然有大量的 SNPs 有待于被发现。通过测序技术发现,特色培育品种的遗传变异对研究中国地方猪种和培育品种都非常必要。同时,对于丰富 SNP 数据库也有着重要的意义。

除了进行 SNP 的检测外,本研究还对上海白猪 InDel 变异进行了检测,并发现这些变异广泛存在于上海白猪的基因组中。这些结果将有利于人们对猪基因组信息更加全面和深入的认识。通过与已有研究比较^[4,7,12,25],本研究所检测的 InDels 中,26.3% 也被其它研究检测到(若已报道的 InDel 与笔者检测的重叠区域超过任一方长度的 80%,则认为是同一个)。这些研究也从另一个侧面证实了本研究检测的可靠性。

在揭示上海白猪基因组遗传分子的特征上,本研究主要对遗传变异分子在染色体上的数量、密度、物理位置等分子特征进行了分析。研究结果表明,上海白猪的遗传分子具有下列特征:第一,SNP 在不同染色体上的密度分布存在差异。值得关注的是 12 号染色体,其变异的密度明显高于其它染色体。笔者查看了 Ensembl 的基因数据库和 QTLs 数据

库,同样发现 12 号染色体具有最高密度的基因分布,而且其 QTL 密度也在前 3 位。由此可以看出,12 号染色体可能在基因网络和性状表现上发挥着非常重要的作用,值得研究者对此染色体重点关注。第二,除少量位置出现过密或过疏,变异在同一条染色体上的物理分布相对均一。因此,本研究检测的 SNP 可以作为全基因组范围内的遗传标记,用于遗传多样性、群体结构、信号选择、全基因组关联分析等研究中。

此外,研究发现,很多基因受到 InDel 改变阅读框的影响。其中,一些基因与繁殖性状相关,例如 *CXCL10* 基因,S. Dall'olio 等^[26]报道,*CXCL10* 基因可能参与胚胎的发育和着床,该基因中存在的多态变异可以作为意大利大白猪产仔数关联研究的标记。另外,一些基因是与免疫相关,如 *IRF7* 基因,*IRF7* 基因属于干扰素调控因子家族成员,在调节 I 型干扰素抵抗病原体的感染中发挥着非常关键的作用^[27-28]。而 *IFIT1* 基因同时作为传递和受体分子来抵抗几类病毒家族病原体^[29]。值得注意的是,基因富集分析结果也表明,InDel 相关基因在多个与疾病相关的通路中发生了富集(如“*ssc05414*: Dilated cardiomyopathy”和“*ssc05410*: Hypertrophic cardiomyopathy (HCM)”)。因此推测,上海白猪的 InDel 极有可能在繁殖性状与免疫性状中发挥着重要作用。

QTL 数据库包含了猪经济性状的重要遗传信

息,因此对于动物遗传研究学者是一个巨大的宝藏。为了更清楚地了解各类变异可能影响的性状,建立遗传变异与性状之间可能存在的联系,本研究开展了基于 QTL 数据库的映射分析。同时又对各类变异在不同性状中的分布特点进行了统计分析。结果显示,各类变异在 QTL 上的比例分布与基因上的比例分布相近(除个别区间),这表明各类变异在不同性状 QTL 中基因上的分布相对均一。另外,本研究还统计了各类变异在新定义 QTL 中的密度,整体而言具有较高变异密度的 QTL 主要是与肉质与胴体及健康性状相关,这些遗传标记将能进一步用于解析为何上海白猪在肉质和抗病力上有较好表现。这些研究结果加深了人们对上海白猪基因组遗传变异分布特点的了解,为进一步深入的研究其功能和加快分子育种进程奠定了基础,也为其后续的保护和利用提供了分子生物学依据。

4 结 论

本研究利用简化基因组测序技术对肉质好、胴体瘦肉率较高和耐粗饲的上海白猪(上系)全基因组范围内的遗传变异进行了检测,并通过对比太湖流域地方猪种和西方引进品种的合并比较分析,阐释了这些遗传变异的染色体分布、基因区间和功能注释特征。本研究深入分析了上海白猪(上系)群体的遗传现状,为更好地进行提纯复壮、分子设计育种与开发利用奠定基础。

参考文献(References):

[1] AI H S, HUANG L S, REN J. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and western pigs revealed by genome-wide SNP markers[J]. *PLoS One*, 2013, 8(2): e56001.

[2] AI H S, FANG X D, YANG B, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing[J]. *Nat Genet*, 2015, 47(3): 217-225.

[3] LI M Z, TIAN S L, JIN L, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars[J]. *Nat Genet*, 2013, 45(12): 1431-1438.

[4] HAN S H, SHIN K Y, LEE S S, et al. SINE indel polymorphism of AGL gene and association with growth and carcass traits in Landrace \times Jeju Black pig F₂ population[J]. *Mol Biol Rep*, 2010, 37(1): 467-

471.

[5] SCHIAVO G, RIBANI A, UTZERI V J, et al. A genomic landscape of mitochondrial DNA insertions in the nuclear pig genome [J]. *J Anim Sci*, 2016, 94(S1): 181.

[6] ZANG L, WANG Y D, SUN B X, et al. Identification of a 13 bp indel polymorphism in the 3'-UTR of DGAT2 gene associated with backfat thickness and lean percentage in pigs [J]. *Gene*, 2016, 576 (2): 729-733.

[7] XIAO Q, ZHANG Z, SUN H, et al. Genetic variation and genetic structure of five Chinese indigenous pig populations in Jiangsu Province revealed by sequencing data[J]. *Anim Genet*, 2017, 48(5): 596-599.

[8] XIAO Q, ZHANG Z, SUN H, et al. Pudong White pig: a unique genetic resource disclosed by sequencing data[J]. *Animal*, 2017, 11(7): 1117-1124.

[9] ZHANG Z, WANG Z, YANG Y, et al. Identification of pleiotropic genes and gene sets underlying growth and immunity traits: a case study on Meishan pigs[J]. *Animal*, 2016, 10(4): 550-557.

[10] CHEN Q, MA Y F, YANG Y M, et al. Genotyping by genome reducing and sequencing for outbred animals [J]. *PLoS One*, 2013, 8(7): e67500.

[11] YANG Y M, WANG Q S, CHEN Q, et al. A new genotype imputation method with tolerance to high missing rate and rare variants[J]. *PLoS One*, 2014, 9(6): e101025.

[12] WANG Z, CHEN Q, LIAO R, et al. Genome-wide genetic variation discovery in Chinese Taihu pig breeds using next generation sequencing [J]. *Anim Genet*, 2017, 48(1): 38-47.

[13] WANG Z, CHEN Q, YANG Y, et al. Genetic diversity and population structure of six Chinese indigenous pig breeds in the Taihu Lake region revealed by sequencing data [J]. *Anim Genet*, 2015, 46 (6): 697-701.

[14] WANG Z, CHEN Q, YANG Y M, et al. A genome-wide scan for selection signatures in Yorkshire and Landrace pigs based on sequencing data [J]. *Anim Genet*, 2014, 45(6): 808-816.

[15] BROWNING B L, BROWNING S R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals[J]. *Am J Human Genet*, 2009, 84(2): 210-223.

[16] HOWIE B N, DONNELLY P, MARCHINI J. A flexi-

- ble and accurate genotype imputation method for the next generation of genome-wide association studies [J]. *PLoS Genet*, 2009, 5(6): e1000529.
- [17] FLICEK P, AHMED I, AMODE M R, et al. Ensembl 2013 [J]. *Nucleic Acids Res*, 2013, 41(D1): D48-D55.
- [18] ZERBINO D R, WILDER S P, JOHNSON N, et al. The ensembl regulatory build [J]. *Genome Biol*, 2015, 16: 56.
- [19] ASHBURNER M, BALL C A, BLAKE J A, et al. Gene ontology: tool for the unification of biology [J]. *Nat Genet*, 2000, 25(1): 25-29.
- [20] KANEHISA M, GOTO S. KEGG: kyoto encyclopedia of genes and genomes [J]. *Nucleic Acids Res*, 2000, 28(1): 27-30.
- [21] HUANG D W, SHERMAN B T, LEMPICKI R A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources [J]. *Nat Protoc*, 2009, 4(1): 44-57.
- [22] HUANG D W, SHERMAN B T, TAN Q, et al. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists [J]. *Nucleic Acids Res*, 2007, 35(W1): W169-W175.
- [23] HU Z L, PARK C A, WU X L, et al. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era [J]. *Nucleic Acids Res*, 2013, 41(D1): D871-D879.
- [24] HU Z L, PARK C A, REECY J M. Developmental progress and current status of the Animal QTLdb [J]. *Nucleic Acids Res*, 2016, 44(D1): D827-D833.
- [25] YANG R, FANG S, WANG J, et al. Genome-wide analysis of structural variants reveals genetic differences in Chinese pigs [J]. *PLoS One*, 2017, 12(10): e0186721.
- [26] DALL'OLIO S, FONTANESI L, TOGNAZZI L, et al. Genetic structure of candidate genes for litter size in Italian Large White pigs [J]. *Vet Res Commun*, 2010, 34(S1): S203-S206.
- [27] LIANG Q M, DENG H Y, SUN C W, et al. Negative regulation of IRF7 activation by activating transcription factor 4 suggests a cross-regulation between the IFN responses and the cellular integrated stress responses [J]. *J Immunol*, 2011, 186(2): 1001-1010.
- [28] NING S, PAGANO J S, BARBER G N. IRF7: activation, regulation, modification and function [J]. *Genes Immun*, 2011, 12(6): 399-414.
- [29] DIAMOND M S. IFIT1: A dual sensor and effector molecule that detects non-2'-O methylated viral RNA and inhibits its translation [J]. *Cytokine Growth Factor Rev*, 2014, 25(5): 543-550.

(编辑 郭云雁)