

张璞, 林育, 游斌杰, 等. 猪的免疫球蛋白基因鉴定和注释 [J]. 中国实验动物学报, 2020, 28(2): 208–215.

Zhang P, Lin Y, You BJ, et al. Identification and annotation of the porcine immunoglobulin genes [J]. Acta Lab Anim Sci Sin, 2020, 28(2): 208–215.

Doi:10.3969/j.issn.1005-4847.2020.02.009

猪的免疫球蛋白基因鉴定和注释

张璞¹, 林育¹, 游斌杰^{2,3,4}, 孙静^{2,3,4,5*}

(1. 四川农业大学畜禽品种资源发掘与利用省重点实验室, 四川 611130; 2. 重庆市畜牧科学院, 重庆 402460; 3. 农业部养猪科学重点实验室, 重庆 402460; 4. 养猪科学重庆市市级重点实验室, 重庆 402460; 5. 重庆市医用动物资源开发与利用工程技术研究中心, 重庆 402460)

【摘要】 目的 本研究旨在鉴定猪的免疫球蛋白基因, 为研究猪对疾病的抗性奠定坚实的基础, 同时对生猪养殖及其作为医学模式动物都有一定意义。**方法** 使用 Blast+, R, DNASTar, WebLogo 等软件通过人免疫球蛋白基因信息对猪的免疫球蛋白基因进行预测分类。**结果** 共鉴定到猪的免疫球蛋白重链(IGH)基因 43 个, 分为 7 个基因家族; 免疫球蛋白 κ 轻链(IGK)基因 27 个, 分为 6 个基因家族; 免疫球蛋白 λ 轻链(IGL)基因 25 个, 分为 9 个基因家族。**结论** 我们发现了 43 个猪免疫球蛋白重链基因, 其中可能存在假基因。同时猪免疫球蛋白两种轻链基因比例接近, 这可能与其免疫特性有关, 但仍需要相关的试验数据来开展研究。

【关键词】 猪; 免疫球蛋白; 基因组

【中图分类号】 Q95-33 **【文献标识码】** A **【文章编号】** 1005-4847(2020) 02-0208-08

Identification and annotation of the porcine immunoglobulin genes

ZHANG Pu¹, LIN Yu¹, YOU Binjie^{2,3,4}, SUN Jing^{2,3,4,5*}

(1. Provincial Key Laboratory for Farm Animal Genetic Resource Exploration and Innovation, Sichuan Agricultural University, Sichuan 611130, China. 2. Chongqing Academy of Animal Sciences, Chongqing 402460. 3. Key Laboratory of Pig Industry Science of Agriculture Ministry, Chongqing 402460. 4. Chongqing Key Laboratory of Pig Industry Science, Chongqing 402460. 5. Technical Engineering Center for the Development and Utilization of Medical Animal Resources, Chongqing 402460)

Corresponding author: SUN Jing. E-mail: sunjing85026@163.com

【Abstract】 Objective The aim of this study was to establish a basis for the identification of porcine immunoglobulin genes in order to facilitate research on swine disease resistance. Further improvements of immunoglobulin gene annotation will facilitate the development of the pig industry and will promote the wider application of pigs as a major biomedical model. **Methods** Porcine immunoglobulin genes were predicted and classified by comparison with human immunoglobulin genes using R, DNASTar, WebLogo, and Blast+. **Results** A total of 43 IGH, 27 IGK, and 25 IGL genes were identified in pigs, which could be divided into 7, 6, and 9 gene families, respectively. **Conclusions** We found 43 immunoglobulin heavy chain genes in pig. Some of them may be pseudogene. Meanwhile, the ratio of light chain genes (IGK:IGL) is approximately 1:1. It may be related to the immunological characteristics of pig, which need related experimental data to do its research.

【Keywords】 pig; immunoglobulin; genome

Conflicts of Interest: The authors declare no conflict of interest.

【基金项目】 重庆市科研院所绩效激励引导专项项目 (cstc2018jxjl0083, cstc2018jxjl0043); 重庆市农发资金 (17407)。
Funded by Performance Incentive Guidance for Scientific Research Institution of Chongqing Science & Technology Commission (cstc2018jxjl0083, cstc2018jxjl0043), and Chongqing Agricultural Development Fund Project (17407).

【作者简介】 张璞 (1996—), 男, 学士, 研究方向: 畜禽品种资源发掘与利用。Email: 964057802@qq.com

【通信作者】 孙静, 副研究员, 博士, 研究方向: 无菌动物的培育与应用。Email: sunjing85026@163.com

家猪(*Sus scrofa domesticus*)属于哺乳纲、真兽亚纲、偶蹄目,是较早被驯化的动物之一^[1],也是最为广泛的肉类来源。在驯化过程中,猪体型的头躯比从原始野猪的 7:3 变成了现代家猪的 3:7 左右,使得现代家猪具有很强的产肉性能^[2]。我国是猪肉生产大国,猪肉是我国城乡居民主要消费的肉类,猪肉消费比重占肉类消费总量的 60% 左右。全世界猪肉的消费比重在总体肉类消费中同样占据重要地位。

除了可以提供肉制品,猪还是重要的模式动物。尽管猪是非灵长类动物,但是在基因同源度、部分器官解剖学结构、新陈代谢方式等多个方面与人类有着较高相似性^[3-4]。利用其繁殖性能优秀、单胎产子数多等特点,可以在较短时间内提供充足的试验材料。如今猪的克隆技术已较为成熟,可在一定程度上进行相关基因的编辑^[5-6],为疾病研究提供稳定且性状均一的动物模型,现阶段在疾病模型研究中已经有了一定应用^[7-11],同时有望作为器官供体进行器官移植^[12]。

编码免疫球蛋白(immunoglobulin, IG)的基因包括重链(immunoglobulin heavy chain)和轻链(immunoglobulin light chain)基因,重链和轻链的胚系基因(germline gene)由编码可变区(V 区)和恒定区(C 区)的多个基因偏串联形成的基因簇组成。重链可变区基因包括 V、D、J 三类基因片段,轻链可变区基因包括 V、J 两类基因片段^[13]。重链和轻链基因在基因组中通常位于不同染色体上,其胚系基因彼此由内含子隔开。

免疫球蛋白是指具有抗体活性且化学结构与抗体类似的一类球蛋白,具有调节免疫、介导消除抗原等多种功能。免疫球蛋白分子单体都具有由两条重链多肽和两条轻链多肽组成的基本结构,两条轻链分别靠二硫键及非共价作用力与重链结合,两条重链间通过二硫键与非共价键的亲水和疏水作用力结合在一起,形成一个 Y 字型结构^[13]。根据重链和轻链恒定区结构的不同可以将重链分为五类(α 、 δ 、 ϵ 、 γ 、 μ 链),分别组成 IGA、IGD、IGE、IGG、IGM 五类免疫球蛋白;将轻链蛋白分为 κ 型和 λ 型,分别对应 IGL 与 IGL。

基因注释方法主要包括两大类:一是进行序列比对,二是进行从头预测。研究者在对芥菜(*Brassica juncea*)基因组进行基因注释时便采用了这两种注释策略,并结合转录组数据,取得了很好

的注释效果^[14]。彭长军等人对大熊猫(*Ailuropoda melanoleuca*)基因组重复序列采取了基于序列比对的方法注释其转座子元件,选取的数据库为食肉目,仍获得了不错的注释效果^[15]。在对某一种或某一类基因进行预测时,通过序列对比的预测方法较为常用^[16-17]。

最新版猪基因组(*Sus scrofa* 11.1 版本)是 2018 年由英国罗斯林研究所 Alan 团队主导拼装并公布的,其基因组组装指标 contig N50 已经提高到与人接近的水平,是极其优秀的参考基因组。基因组组装得越好,对基因、功能元件、基因组修饰等的注释帮助就越大。当前版本猪基因组组装得足够好,但是猪的免疫球蛋白基因并未获得完全注释。猪与人的基因相似度较高,而人的免疫球蛋白基因研究得足够深入,通过序列对比对猪的免疫球蛋白基因进行注释可能取得不错的效果。研究表达的猪免疫球蛋白基因有助于了解相关免疫球蛋白的组装机制,进一步研究猪对抗原、病原的免疫特性,以更好地发挥其作为农用动物及疾病模型的作用。

1 材料与方法

1.1 资料来源

人的免疫球蛋白基因信息下载自 NCBI 数据库(<https://www.ncbi.nlm.nih.gov/>),猪的最新版基因组数据(*Sscrofa* 11.1)下载自 Ensembl 数据库(http://asia.ensembl.org/Sus_scrofa/Info/Index)。

1.2 数据处理

我们首先利用已知的人免疫球蛋白基因作为模板,使用 Blast+ (<https://www.ncbi.nlm.nih.gov/books/NBK52640/>),将猪基因组数据进行序列比对,筛选出潜在的猪的免疫球蛋白基因,并进行定位与注释;之后结合注释信息使用 R 软件绘制基因图谱;并根据已知的人免疫球蛋白基因信息,与鉴定所得猪的免疫球蛋白基因进行系统发育分析,使用 TreeBest (<http://treesoft.sourceforge.net/treebest.shtml>)基于 Neighbor-joining 计算方法构建系统发育树。

同时我们根据鉴定出的猪的免疫球蛋白重链和轻链基因片段,使用 DNASTAR (DNASTAR Inc., Madison, USA)对其核苷酸序列进行同源性分析,将同源性大于 80% 的序列划分为同一基因家族。此外我们还使用 WebLogo^[18] 分别对猪免疫球蛋白的重链和轻链基因序列绘制序列标识图。

2 结果

2.1 猪免疫球蛋白基因及其基因座结构概述

BLAST 搜索共鉴定出 43 个免疫球蛋白重链基因,其中 IGHE 基因 1 个,IGHG 基因 7 个,IGHV 基因 35 个,这些基因散乱地分布于多个 scaffold 上,难以对其进行具体定位(见表 1)。

同样方法鉴定出 52 个免疫球蛋白轻链基因,其中 IGH 型轻链蛋白基因 27 个,IGL 型轻链蛋白基因 25 个。IGH 基因分布于一个基因组 scaffold 上,所有基因跨越的长度范围为 290,541 bp。其中 V 区基因 26 个,C 区基因 1 个,具体注释信息见图 1。IGL 基因分布于 14 号染色体上,长度范围为 250,644 bp。其中 V 区基因 21 个,C 区基因 4 个,具体注释信息见图 2。

表 1 猪的 IGH 基因概述
Table 1 Summaries of IGH gene in pig

所在片段 Scaffold	免疫球蛋白基因 Immunoglobulin genes	分布范围 Distribution range
AEMK02000149. 1	IGHV-g1-Pig	11,133–11,423
AEMK02000566. 1	IGHV-g2-Pig; IGHV-g3-Pig; IGHV-g4-Pig; IGHV-g5-Pig; IGHV-g6-Pig; IGHV-g7-Pig; IGHV-g8-Pig; IGHV-g9-Pig; IGHV-g10-Pig; IGHV-g11-Pig; IGHV-g12-Pig; IGHV-g13-Pig; IGHV-g14-Pig; IGHV-g15-Pig; IGHV-g16-Pig; IGHV-g17-Pig; IGHV-g18-Pig; IGHV-g19-Pig; IGHV-g20-Pig; IGHV-g21-Pig; IGHV-g22-Pig; IGHV-g23-Pig; IGHV-g24-Pig; IGHV-g25-Pig; IGHV-g26-Pig;	15,646–270,287
AEMK02000188. 1	IGHV-g27-Pig; IGHV-g28-Pig; IGHV-g29-Pig; IGHV-g30-Pig; IGHV-g31-Pig; IGHV-g32-Pig; IGHV-g33-Pig	17,872–70,292
AEMK02000452. 1	IGHE-g1-Pig; IGHG-g1-Pig; IGHG-g2-Pig; IGHG-g3-Pig; IGHG-g4-Pig; IGHG-g5-Pig; IGHG-g6-Pig; IGHG-g7-Pig; IGHV-g34-Pig; IGHV-g35-Pig	3,596,046–3,841,144

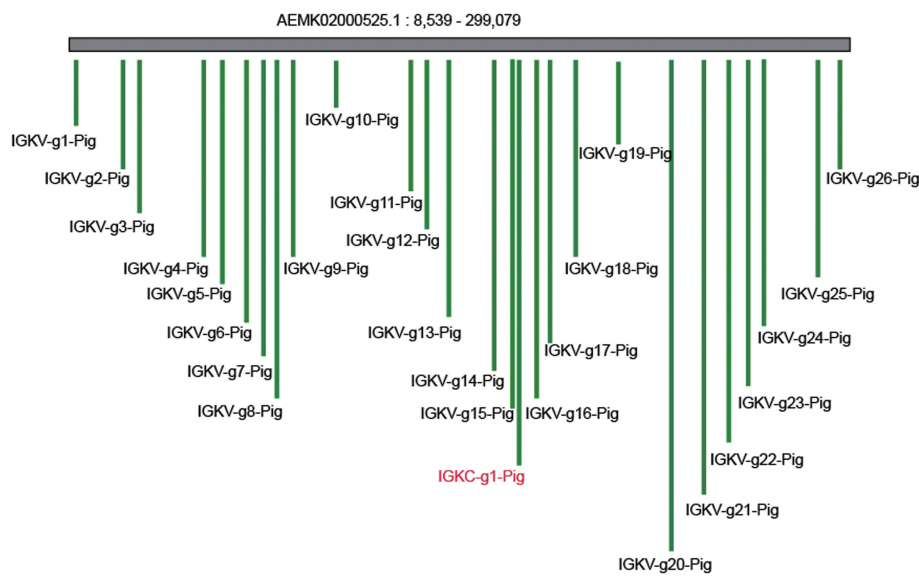


图 1 猪的 IGH 基因物理图谱
Figure 1 Physical map of IGH gene in pig

2.2 猪免疫球蛋白基因的系统发生分析

选取所有鉴定出的猪免疫球蛋白基因片段,对比分析猪与人的免疫球蛋白基因的进化关系。结果如图 3、图 4、图 5 所示,免疫球蛋白重链和轻链基因的 V 区与 C 区的基因被划分成两个分支。对 IGH 与 IGL 基因的分析发现,猪与人的 V_H(重链可变区)和 V_L(L 链可变区)基因聚在一起,表明其具有很高的相似度,而对 IGK 基因的分析发现,一部分猪的 V_K 基因独立聚在一起,跟人的 V_K(K 链可变区)基因同源性很低,不排除其为假基因的可能。

2.3 猪免疫球蛋白基因家族分析

根据鉴定出的猪免疫球蛋白基因序列的相似性,将相似性 80% 及以上的序列划分为一个免疫球蛋白基因家族。免疫球蛋白重链基因可分为七个

家族,每个家族包含的基因数 1 到 18 个不等;轻链 IGK 基因可分为六个基因家族,每个家族基因数 1 到 12 个不等;IGL 基因可分为九个基因家族,每个家族基因数 1 到 6 个不等。同时,根据 IGH、IGK 和 IGL 基因的序列绘制序列标识图(Sequence logo),见图 6、图 7、图 8,提示在同一种免疫球蛋白基因中,大部分位点存在较高的保守性。

3 讨论

此前的研究表明,多种动物的免疫球蛋白基因库远大于其抗体库^[19]。一部分免疫球蛋白基因在复制的过程中发生点突变并成为新的免疫球蛋白基因家族成员,这可能是免疫球蛋白各家族成员庞大且序列相似度较高的原因之一。免疫球蛋白存

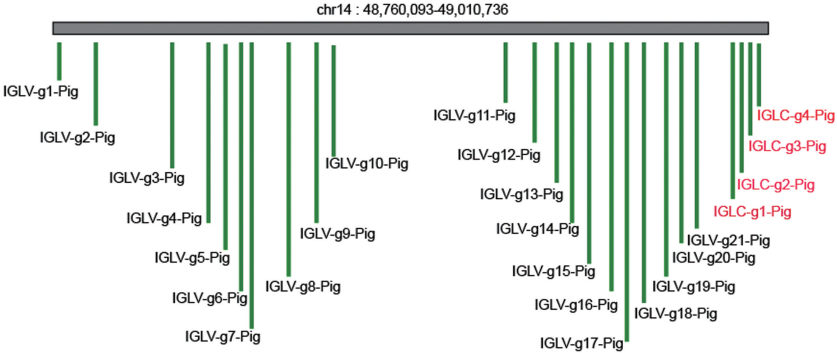


图 2 猪的 IGL 基因物理图谱
Figure 2 Physical map of IGL gene in pig

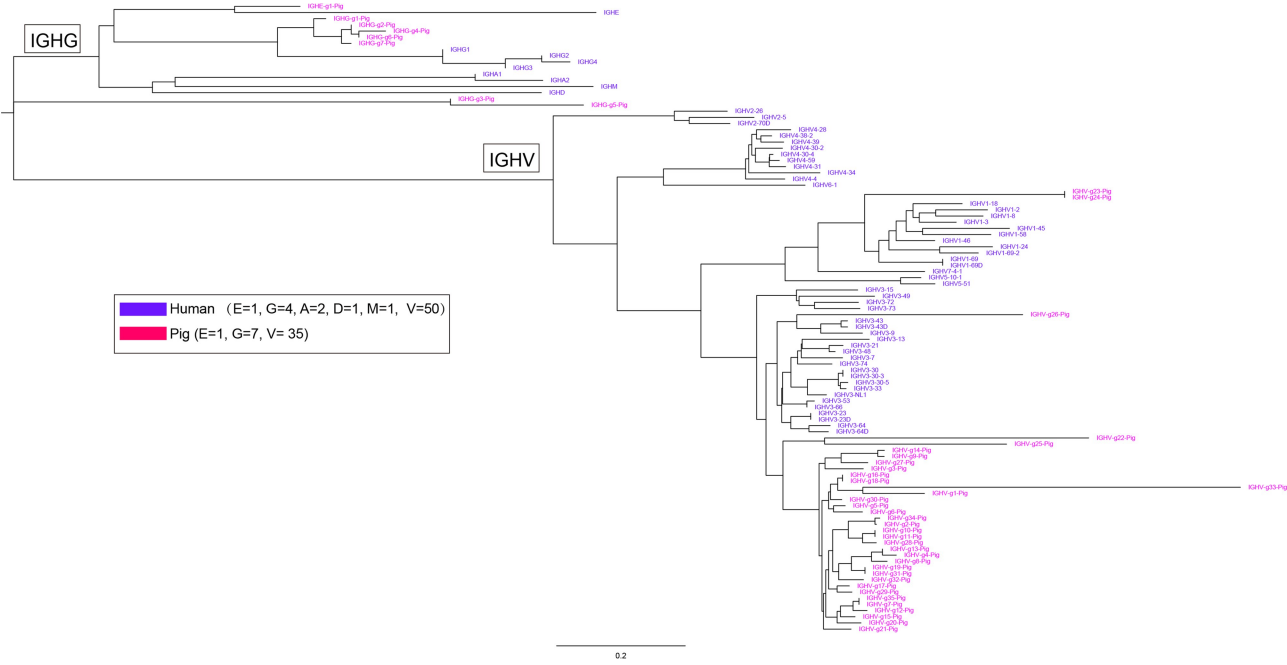


图 3 猪与人的 IGH 基因系统发生分析
Figure 3 Phylogenetic analysis of IGH gene in pig and human

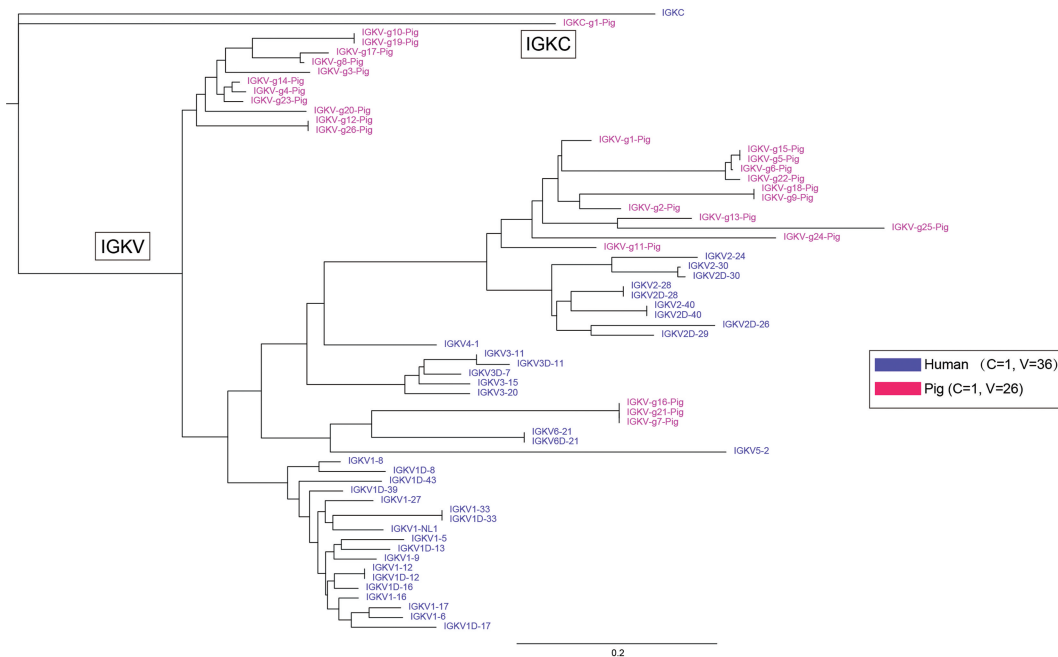


图 4 猪与人的 IGH 基因系统发生分析

Figure 4 Phylogenetic analysis of IGH gene in pig and human

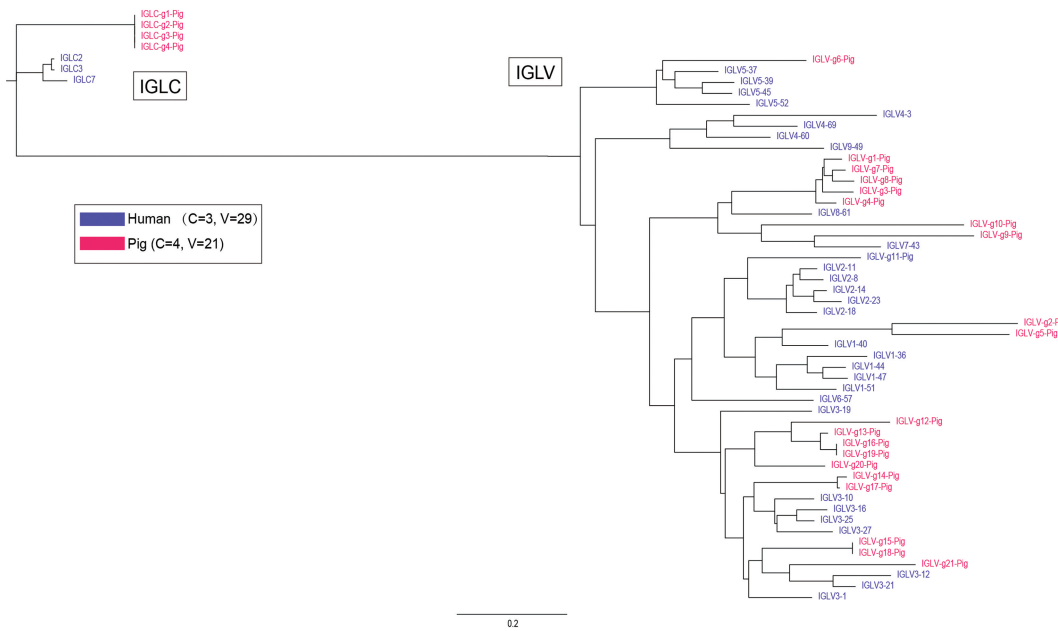


图 5 猪与人的 IGL 基因系统发生分析

Figure 5 Phylogenetic analysis of IGL gene in pig and human

在多种亚型,并非所有编码免疫球蛋白的基因会常时性表达,如敲除 IGD 的小鼠对部分胸腺依赖性和非胸腺依赖性抗原的免疫反应良好^[20],而兔子则完全缺乏 IGD^[21]。一些具有 IGD 基因的哺乳动物也几乎不表达它,但不能忽视这些免疫球蛋白亚型的作用。因此,在基因组层面注释免疫球蛋白基因,了解其基因库特征将有助于揭示其系统发育机制

及表达机制,从而建立更加有效的适应性免疫系统。基于免疫球蛋白基因的保守性,研究者利用脊椎动物免疫球蛋白重链基因克隆了中华鳖 (*Pelodiscus sinensis*) 的免疫球蛋白重链基因,发现中华鳖的三种免疫球蛋白重链亚型 (IGM、IGY 和 IGD) 跟鸡、青蛙、壁虎和人等的相应亚型有 40% ~ 50% 的同源性^[22]。此外,基于序列同源性比对鉴定

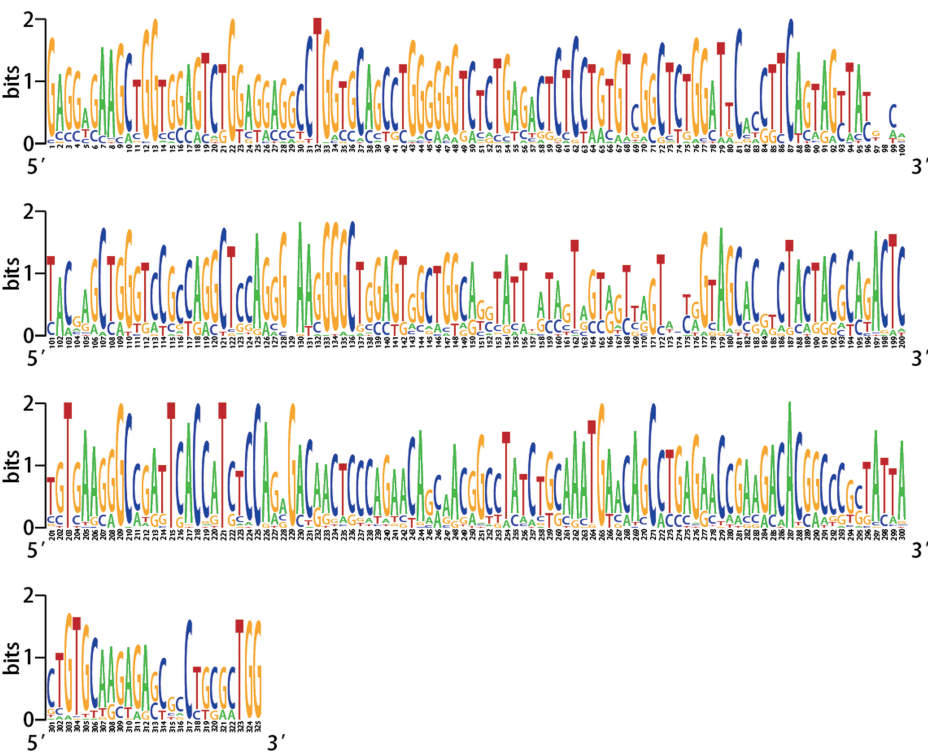


图 6 猪 IGH 基因的序列标识图
Figure 6 Sequence logo of IGH gene in pig

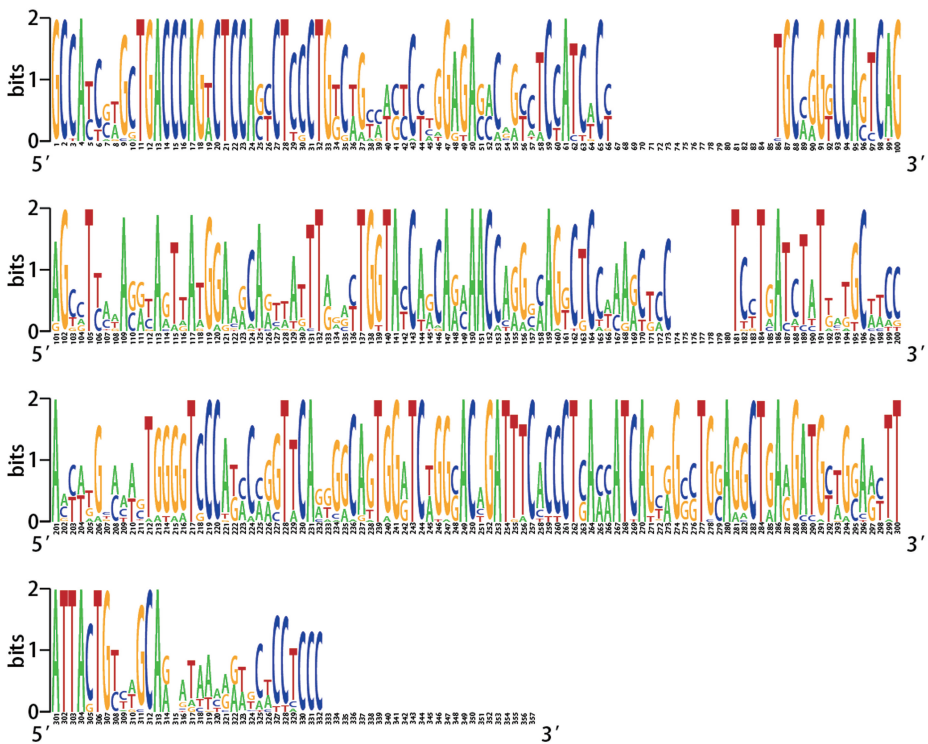


图 7 猪 IGK 基因的序列标识图
Figure 7 Sequence logo of IGK gene in pig

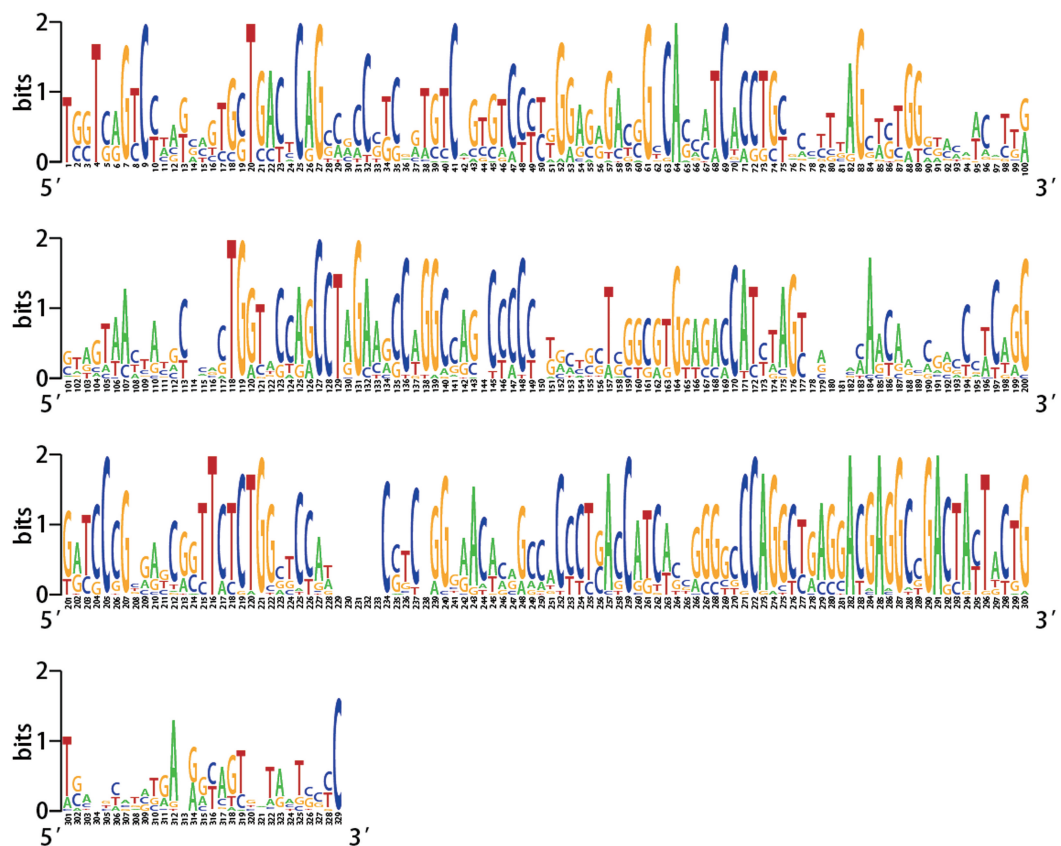


图 8 猪 IGL 基因的序列标识图

Figure 8 Sequence logo of IGL gene in pig

出爬行动物变色蜥 (*Anolins carolinensis*) 的免疫球蛋白重链基因座^[23], 预示基于序列比对、利用生物信息学分析手段预测免疫球蛋白基因是可行的。

由于猪与人在基因同源度、部分器官解剖学结构等多个方面的高度相似性, 从生物信息的角度, 通过比对人的免疫球蛋白基因与猪基因组可成功筛选鉴定出相关的猪免疫球蛋白基因。此前对猪的免疫球蛋白可变区的研究共发现约 33 个 V_H 基因。由于轻链基因的研究较少, 共发现 14–15 个 V_L 基因, 而且 V_K 基因数目存在较大波动^[19]。本研究共鉴定猪的 IGH 基因 43 个, 可分为 7 个基因家族; IGH 基因 27 个, 可分为 6 个基因家族; IGL 基因 25 个, 可分为 9 个基因家族。这将有助于对猪免疫特性等研究, 以期更好地发挥其作为疾病模型的作用。结合更多免疫球蛋白的研究, 可深入探索猪免疫球蛋白重排特性、连接特性等。

在小鼠和人等多种动物上的研究表明, 免疫球蛋白基因可变区的基因数目往往远大于恒定区的基因数目, 可变区的 V 基因片段通常具有较高的相似性, 通常按照 80% 的相似性进行基因家族的划

分。小鼠的 V_H 基因十分丰富, 总数上百个^[24], 至少可以分为九个家族。研究表明小鼠免疫球蛋白重链基因受到增强子的调控^[25]。人的 V_H 基因按照相似性至少可以分为七个家族, 并且各个家族的基因片段并不是相互独立, 可与其他家族混杂分布于染色体上。从猪的免疫球蛋白重链和轻链基因序列标识图(图 6–8)还可以看出其 IGH、IGH 和 IGL 基因序列上具有一定的保守性, 其 V_H 、 V_K 和 V_L 基因片段的序列也具有较高的相似性。

对免疫球蛋白轻链的研究在多种哺乳动物中都发现了 κ 型和 λ 型两种轻链类型, 但是在鸟类上仅发现 λ 型轻链^[26]。在哺乳动物中, κ 型和 λ 型轻链蛋白所占比例有所不同。人和小鼠以 κ 型轻链为主, 牛、马、猫、狗等动物主要的轻链蛋白却是 λ 型, 由浆细胞产生的猪免疫球蛋白的 κ 型和 λ 型轻链蛋白比例近乎相等^[27]。本研究发现, 猪的 IGH 和 IGL 型免疫球蛋白基因数目较为接近, 虽然免疫球蛋白的含量并不能仅靠基因评定, 但可通过其可变区 V、J 基因片段去探索猪的 VJ 重组多样性, 为进一步研究猪的 κ 型和 λ 型免疫球蛋白奠定基础。

参 考 文 献(References)

- [1] Groenen MAM, Archibald AL, Uenishi H, et al. Analyses of pig genomes provide insight into porcine demography and evolution [J]. *Nature*, 2012, 491(7424): 393–398.
- [2] 龙科任. 利用基因组重测序鉴定巴克夏猪的人工选择痕迹 [D]. 雅安: 四川农业大学; 2015.
Long KR. Whole-genome resequencing provides insight into its artificial selection footprint [D]. Ya'an; Sichuan Agricultural University; 2015.
- [3] Spurlock ME, Gabler NK. The development of porcine models of obesity and the metabolic syndrome [J]. *J Nutr*, 2008, 138(2): 397–402.
- [4] Lunney JK. Advances in swine biomedical model genomics [J]. *Int J Biol Sci*, 2007, 3(3): 179–184.
- [5] Denner J. Advances in organ transplant from pigs Inactivation of PERVs in the pig genome increases the safety of xenotransplantation [J]. *Science*, 2017, 357(6357): 1238–1239.
- [6] Ruan DG, Peng JY, Wang XS, et al. XIST derepression in active X chromosome hinders pig somatic cell nuclear transfer [J]. *Stem Cell Rep*, 2018, 10(2): 494–508.
- [7] Perleberg C, Kind A, and Schnieke A. Genetically engineered pigs as models for human disease [J]. *Dis Model Mech*, 2018, 11(1): pii: dmm030783.
- [8] Prather RS, Lorson M, Ross JW, et al. Genetically engineered pig models for human diseases [J]. *Annu Rev Anim Biosci*, 2013, 1: 203–219.
- [9] Miranda FJB, de Souza DB, Frazao-Teixeira E, et al. Experimental infection with the *Toxoplasma gondii* ME-49 strain in the Brazilian BR-1 mini pig is a suitable animal model for human toxoplasmosis [J]. *Mem Inst Oswaldo Cruz*, 2015, 110(1): 95–100.
- [10] Koopmans SJ, Schuurman T. Considerations on pig models for appetite, metabolic syndrome and obese type 2 diabetes: From food intake to metabolic disease [J]. *Eur J Pharmacol*, 2015, 759: 231–239.
- [11] Yan S, Tu ZC, Liu ZM, et al. A huntingtin knockin pig model recapitulates features of selective neurodegeneration in Huntington's Disease [J]. *Cell*, 2018, 173(4): 989–1002.
- [12] Pan DK, Liu T, Lei TT, et al. Progress in multiple genetically modified minipigs for xenotransplantation in China [J]. *Xenotransplantation*, 2019, 26(1): e12492.
- [13] 蒋磊. 多种动物 IgG(Y)Fc 段基因分析及遗传变异研究 [D]. 泰安: 山东农业大学; 2014.
Jiang L. The genetic analysis and genetic variation of kinds of animal's IgG(Y)Fc Fragment [D]. Taian; Shandong Agricultural University; 2014.
- [14] Yang J, Liu D, Wang X, et al. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection [J]. *Nat Genet*, 2016, 48(10): 1225–1232.
- [15] 彭长军, 牛李丽, 邓家波, 等. 两种方法对大熊猫基因组重复序列的注释比较 [J]. *四川动物*, 2017, 36(2): 121–130.
Peng CJ, Niu LL, Deng JB, et al. Annotations of the repeat elements in *Ailuropoda melanoleura* genome based on two strategies [J]. *Sichuan J Zool*, 2017, 36(2): 121–130.
- [16] Hong W, Zhao H. Vampire bats exhibit evolutionary reduction of bitter taste receptor genes common to other bats [J]. *Proc Biol Sci*, 2014, 281(1788): 20141079.
- [17] Nasipak BT, Kelley DB. The genome of the diploid anuran *Xenopus tropicalis* contains a novel array of sarcoplasmic myosin heavy chain genes expressed in larval muscle and larynx [J]. *Dev Genes Evol*, 2008, 218(7): 389–397.
- [18] Crooks GE, Hon G, Chandonia JM, et al. WebLogo: a sequence logo generator [J]. *Genome Res*, 2004, 14(6): 1188–1190.
- [19] Butler JE, Wertz N. The porcine antibody repertoire: variations on the textbook theme [J]. *Front Immunol*, 2012, 3: 153.
- [20] Nitschke L, Kosco MH, Köhler G, et al. Immunoglobulin D-deficient mice can mount normal immune responses to thymus-independent and -dependent antigens [J]. *Proc Natl Acad Sci U S A*, 1993, 90(5): 1887–1891.
- [21] Lanning DK, Shi-Kang Z, and Knight KL. Analysis of the 3' Cmu region of the rabbit Ig heavy chain locus [J]. *Gene*, 2003, 309(2): 135–144.
- [22] Xu Z, Wang GL, Nie P, et al. IgM, IgD and IgY and their expression pattern in the Chinese soft-shelled turtle *Pelodiscus sinensis* [J]. *Mol Immunol*, 2009, 46(10): 2124–2132.
- [23] Deza FG, Espinel CS, Mompó SM. The immunoglobulin heavy chain locus in the reptile *Anolis carolinensis* [J]. *Mol Immunol*, 2009, 46(8–9): 1679–1687.
- [24] Brodeur PH, Riblet R. The immunoglobulin heavy chain variable region (IgH-V) locus in the mouse. I. One hundred IgH-V genes comprise 7 families of homologous genes [J]. *Eur J Immunol*, 1984, 14(10): 922–930.
- [25] Banerji J, Olson L, Schaffner W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes [J]. *Cell*, 1983, 33(3): 729–740.
- [26] Bencina M, Cizelj I, Bercic RL, et al. Shared epitopes of avian immunoglobulin light chains [J]. *Vet Immunol Immunopathol*, 2014, 158(3–4): 175–181.
- [27] Arun SS, Breuer W, Hermanns W. Immunohistochemical examination of light-chain expression (lambda/kappa ratio) in canine, feline, equine, bovine and porcine plasma cells [J]. *Zentralbl Veterinarmed A*, 1996, 43(1–10): 573–576.

[收稿日期] 2019-07-31