



浙川乌骨鸡全基因组转座子的鉴定与分析

李东华, 付亚伟, 张晨曦, 曹艳芳, 李文婷, 李转见, 康相涛, 孙桂荣

(河南农业大学牧医工程学院, 郑州 450046)

摘要: 【目的】通过分析浙川乌骨鸡全基因组重测序数据, 获取浙川乌骨鸡转座子的基本信息和特征分布, 探究转座子相关基因参与的通路。不仅对研究浙川乌骨鸡转座子的生物学功能具有重要的意义, 而且为探索基因组扩增、基因组功能及进化研究提供重要基础数据。【方法】对浙川乌骨鸡血液 DNA 进行全基因组重测序, 采用双末端短序列比对基因组, 运用 RepeatModeler、TEclass、RepeatMasker 等使用流程对转座子进行鉴定、构建、校正、分类和注释, 获得浙川乌骨鸡整个基因组所有的转座子, 对转座子的特征、分布及和基因的关系等方面进行分析。并对转座子插入的所有基因进行 GO 和 KEGG 数据库富集, 分别结合 GO 和 KEGG 注释结果对功能进行描述。【结果】经鉴定、分类和注释, 浙川乌骨鸡共鉴定到 370 252 个转座子序列, 分为 19 个超家族, 主要为 CR1、TcMar-Mariner、ERVL、ERV1 等超家族, 进一步说明浙川乌骨鸡转座子类型主要是逆转录转座子; 转座子的数目和染色体长度有关, 染色体越长, 转座子数目越多。在基因组中转座子和基因的密度成反比, 基因密集的区域中转座子密度较低; 基因组内转座子的插入并非随机的, LTR/ERVL、LTR/ERV1、DNA/PIF-Harbinger、DNA/Hat-Charlie、RC/Helitron 等转座子倾向于插入基因外部。GO 富集分析表明, 在生物过程条目中鉴定出的转座子富集相对较多的是细胞过程、单生物过程、代谢过程、生物调节、刺激反应等; 分子功能条目中鉴定出的转座子富集相对较多的是结合、催化活性等; 细胞组分条目中鉴定出的转座子富集相对较多的是细胞部分、细胞器、膜等。KEGG 富集分析表明, 鉴定出的转座子主要在甘油酯代谢、PPAR 信号通路、PI3K-Akt 信号通路、胰岛素抵抗、Jak-STAT 信号通路等通路。着重关注了与浙川乌骨鸡特性相关的通路, 如和色素沉积有关的酪氨酸代谢、和肉质有关的脂代谢、脂肪酸的合成、PI3K-Akt 信号通路等。【结论】转座子含量与基因组大小, 具有一定的正相关性, 而且浙川乌骨鸡转座子在基因组的分布存在一定偏好性。转座子相关基因在色素相关通路中富集, 其可能与浙川乌骨鸡的种质特性有关, 具体的调控机制还有待进一步研究。

关键词: 鸡; 全基因组重测序; 转座子; 特征分析

Genome-Wide Identification and Characterization of Transposable Elements in Xichuan Black-Bone Chicken

LI DongHua, FU YaWei, ZHANG ChenXi, CAO YanFang, LI WenTing, LI ZhuanJian,
KANG XiangTao, SUN GuiRong

(College of Animal Science and Veterinary Medicine, Henan Agricultural University, Zhengzhou 450046)

Abstract: 【Objective】The aim of the study was to analyze the whole genome re-sequencing data of Xichuan black-bone chicken to obtain the identification, classification, distribution of the transposon of Xichuan black-bone chicken, and to explore the

收稿日期: 2019-01-22; 接受日期: 2020-01-13

基金项目: 国家现代农业产业技术体系建设专项资金(CARS-40-K04)、教育部创新团队(IRT16R23)、河南省科技重大专项(151100110800)

联系方式: 李东华, Tel: 15516976656; E-mail: lidonghua6656@126.com. 通信作者康相涛, E-mail: xtkang2001@263.net. 通信作者孙桂荣, E-mail: grsun2000@126.com

pathways involved in transposon-related genes, which not only had important significance for studying the biological function of the Xichuan black-bone chicken transposon elements (TEs), but also provided important basic data for exploring genome amplification, genome function and evolution research. **【Method】**In this study, whole genome resequencing of blood DNA of Xichuan black-bone chicken was performed, and the paired-end mapping methods alignment genome was used. The TEs was identified, constructed, corrected and classified by RepeaterModeler, TEclass, RepeatMasker and other processes. All TEs in the whole genome of Xichuan black-bone chicken were obtained to analyze the characteristics, distribution and relationship of their genes. All genes inserted in the TEs were subjected to GO and KEGG databases enrichment, and the functions were described in combination with GO and KEGG annotation results, respectively. **【Result】** After identification, classification and annotation, 370 252 TEs sequences were identified in Xichuan black-bone chicken, which were divided into 19 superfamilies, mainly CR1, TcMar-Mariner, ERVL, ERV1 and other superfamilies, further indicated that the TEs types of Xichuan black-bone chicken were major TEs. The number of TEs was related to chromosome length, and longer the chromosome was, the more the number of TEs was. Number of TEs was inversely proportional to the density of gene. TEs density was relatively low in gene dense regions. The insertion of TEs in genome was not random, LTR/ERVL, LTR/ERV1, DNA/PIF-Harbinger, DNA/Hat-Charlie, RC/Helitron tend to be inserted outside the gene. GO enrichment analysis indicated some of these genes related to TEs were enriched in the following biological process terms: cell process, single-organism process, metabolic process, biological regulation, response to stimulus. In addition, some of these genes related to TEs were enriched in the following molecular function terms: binding and catalytic activity. Furthermore, some of these genes related to TEs were enriched in the following cell component terms: cell parts, organelles, and membranes. KEGG enrichment analysis indicated that these genes related to TEs mainly focused on Glycerolipid metabolism, PPAR signaling pathway, PI3K-Akt signaling pathway, Insulin resistance, and Jak-STAT signaling pathway. This study mainly focused on the pathways related to the characteristics of Xichuan black-bone chicken, such as tyrosine metabolism related to pigmentation, lipid metabolism related to meat quality, fatty acid synthesis, and PI3K-Akt signaling pathway. **【Conclusion】** There was a positive correlation between TEs content and genome size of Xichuan black-bone chicken. Moreover, the TEs of Xichuan black-bone chicken had a certain preference in the distribution of genome. TEs related genes were enriched in the pigmentation related pathway, which might be related to the germplasm characteristics of Xichuan black-bone chicken. The specific regulatory mechanisms remained to be further studied.

Key words: chicken; whole genome resequencing; transposon elements; characterization analysis

0 引言

【研究意义】转座子（简称 TEs），又称易位子，跳跃基因，指存在于基因组上可以自主复制和位移的一段 DNA 片段^[1]，构成了高等生物基因组的大部分并且其不同生物基因组中的含量有非常大的变化。利用全基因组重测序筛选浙川乌骨鸡有关的转座子，从而发现与浙川乌骨鸡种质特性有关的转座子。**【前人研究进展】**转座子主要分为 Class I 型转座子和 Class II 型转座子两大类^[2]。Class I 型转座子，又称逆转录转座子。以 RNA 为介导进行转座，属于“DNA-RNA-DNA”型，每转座一次相应拷贝数增加，基因组增大；且本身带有调控元件，所以会对宿主基因的表达造成影响。主要包括长末端重复元件（LTRs），内源性反转录病毒（ERV），长散布元件（LINEs）和短散布元件（SINEs），后两者又被统称为非长末端重复元件（Non-LTRs）。Class II 型转座子，也称作 DNA 转座子。以 DNA 为介导进行转座，属“DNA-DNA”

型，造成基因的突变或者重排，一般对基因组大小没有影响。此类转座子包括微型反向重复转座元件（MITEs）和滚筒式转座子（helitron）^[3]。转座子在 20 世纪 50 年代被 McClintock 在玉米中首次发现^[4]，在很长一段时间，一直被当做是基因组中的一类无用的“垃圾序列”^[5]。然而，近年来，随着越来越多物种全基因组测序的应用，关于染色体结构^[6]、基因组大小^[7]、基因组重排^[8]、新基因生成^[9]和基因表达调控^[10]等受转座子影响的研究报道越来越多，在分子生物学研究中被广泛应用于基因治疗、转基因动物制备和物种进化等研究^[11-13]。目前，转座子插入对畜禽的影响也已被广泛报道，CLARK 等^[14]研究发现在狗 silver 基因 SINE 插入进而影响毛色表型。DALL'OLIO 等^[15]研究表明在赛马 MSTN 基因转座子的插入，影响马的赛跑距离。猪 VRTN 基因转座子的插入影响猪脊椎数发育^[16]。**【本研究切入点】**浙川乌骨鸡是 2011 年列入的《中国畜禽遗传资源志-家禽志》中的一个新的遗传资源。具有“乌嘴、乌腿、乌皮、乌骨、

乌肉、绿壳蛋”等特征；遗传性能稳定、觅食力强、适应性广、喜群居、耐寒性强、饲料消耗少、具有药用保健功能等优点。近年来，随着测序技术的发展，越来越多的生物基因组序列被测定，为从基因组水平进行转座子的鉴定、分类、注释等提供了便利，转座子几乎存在于所有生物体基因组中，并在真核生物基因组中占据重要组成部分，尤其是脊椎动物^[17-19]。然而关于鸡转座子方面的研究还鲜有报道。【拟解决的关键问题】本研究利用浙川乌骨鸡全基因组重测序数据，从全基因组范围内进行转座子鉴定、分类和特征分析，为浙川乌骨鸡基因组功能和进化研究等提供重要基础数据。

1 材料与方法

1.1 试验材料

30 周龄的浙川乌骨鸡母鸡 5 个个体来源于河南省淅川县，翅下静脉采血，采集约 1 mL 的血液置于 3 mL 抗凝管（EDTA）中，上下颠倒使抗凝剂和血液混匀后放入-20 冰箱保存，备用。

1.2 血液 DNA 提取

采用氯仿-苯酚法提取鸡血液 DNA，使用 NanoDrop 进行浓度检测。1%的琼脂糖凝胶电泳检测 DNA 的质量和完整性（电压 150V，电泳 30 min）。

1.3 DNA 文库构建和全基因组重测序

使用超声波将浙川乌骨鸡血液 DNA 序列片段化，然后进行纯化、末端修复、3'端加 A、加接头后，磁珠吸附进行富集 400 bp 左右的随机片段，PCR 扩增形成测序文库。文库质检合格后使用 Illumina HiSeq 4000 进行双末端测序（每个个体 30X）。

1.4 全基因组转座子的鉴定和注释

根据 RepeatModeler (<http://www.repeatmasker.org/RepeatModeler/>)，TEclass (<http://www.compgen.uni-muenster.de/teclas>)，RepeatMasker (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>) 使用流程对产生的重复序列的一致性序列进行鉴定、构建、校正、分类和注释。

1.5 转座子家族区分

根据 80-80-80 规则对转座子进行家族划分^[20]。两个元件的序列中的最短序列（>80 bp）满足编码区、两端的重复区域、或者全长的 80%的序列有 80%的相似度中的其中任意一个条件。

1.6 转座子与基因的关系

基于大于 scaffold N50 以上的所有 scaffold 对转座

子的特征进行分析。对 4 Mb 窗口上转座子和基因的分布进行分析；且对转座子在基因内部、基因外部（上下 2 kb）区域插入偏好性进行统计分析。

1.7 转座子相关基因 GO 功能注释及 KEGG Pathway 分析

为了进一步了解转座子的功能，我们对在基因上下 2 kb 区域以及基因内部有转座子插入的基因使用 Blast 2 GO^[21]进行 Gene Ontology（GO）和 KEGG Pathway 功能注释。

2 结果

2.1 基因组中 TEs 的鉴定

经过过滤分析，我们在浙川乌骨鸡基因组当中鉴定到有 370 252 个转座子序列，基于 80-80-80 规则，分为 19 个超家族（表 1），主要是 CR1、TcMar-Mariner、ERVL、ERV1 等超家族，进一步说明浙川乌骨鸡的转座子类型主要是逆转录转座子。

表 1 基因组转座子统计分析
Table 1 Statistical analysis of genome TEs

分类 Class	类别 Order	超家族 Superfamily	数量 Counts
逆转录转座子 Retrotransposons	LTR	Copia	2 182
		Gypsy	424
		ERV1	12 026
		ERVK	7 502
		ERVL	57 233
		ERV	5 265
	LINE	CR1	243 531
		I	489
		L2	591
		RTE-X	259
DNA 转座子 DNA transposons	SINE	SINE/MIR	208
		CMC-EnSpm	1 808
	RC/Helitron	hAT-Ac	486
		hAT-Charlie	9 076
		MULE-MuDR	462
		P	2 853
		PIF-Harbinger	4 148
		TcMar-Mariner	18 304
		RC/Helitron	3 405

2.2 浙川乌骨鸡转座子的全基因组注释

通过对浙川乌骨鸡全基因组注释分析，比对（Masked）的转座子序列总计占到了浙川乌骨鸡基因组已有测序序列的 18.29%（表 2）。逆转录转座子占总 Masked 序列的 96.88%，其中 LINE 转座子所占比例最高，占逆转录转座子总 Masked 序列的 6.43%。除此之外，L3/CR1 也占到总 Masked 序列的 6.38%。DNA 转座子只占到总 Masked 序列的 1.33%，其中以 DNA 所占 Masked 序列比最高（1.21%）。

表 2 浙川乌骨鸡全基因组转座子注释信息

Table 2 Annotation of TEs superfamilies in whole-genome resequencing assembly

分类	类别	超家族	数量	比对序列	Masked 序列所占比
Class	Order	Superfamily	Counts	Masked (bp)	Percentage of Masked (%)
逆转录转座子 Retrotransposons	LTR	LTR	77112	34594642	2.81
		ERVL	45563	18364082	1.49
		ERV_classI	11039	5027391	0.41
		ERV_classII	3659	2027848	0.16
	SINE	SINE	208	16102	0.01
	LINE	LINE	234961	79044697	6.43
		LINE2	589	336996	0.03
		L3/CR1	233726	78474296	6.38
DNA 转座子	DNA	DNA	36568	14908973	1.21
DNA transposons		hAT-Charlie	8963	1578789	0.13
未分类	Unclassified		60019	25121370	2.04
总计	Total		635295	224900544	18.29

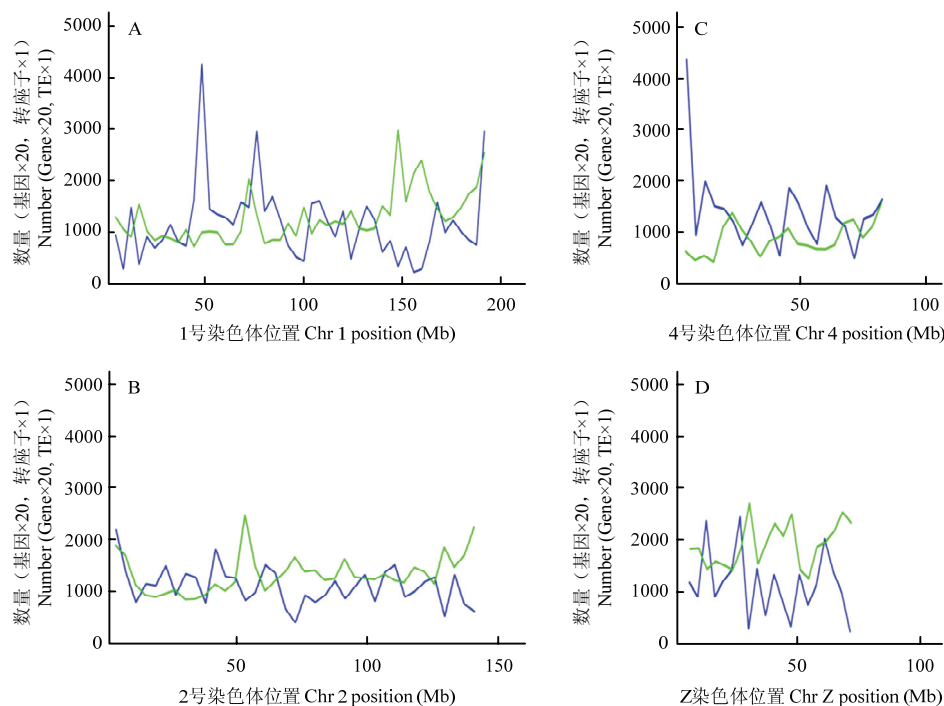
表 3 转座子在染色体上的分布

Table 3 Distribution of transposons on chromosomes

染色体	TEs 数量	所占比例	染色体	TEs 数量	所占比例
Chromosome	TEs Counts	Proportion (%)	Chromosome	TEs Counts	Proportion (%)
Chr1	71257	26.74	Chr15	1767	0.66
Chr2	47150	17.69	Chr17	1584	0.59
Chr3	29426	11.04	Chr18	1732	0.65
Chr4	20621	7.74	Chr19	1511	0.57
Chr5	11618	4.36	Chr20	2043	0.77
Chr6	5725	2.15	Chr21	434	0.16
Chr7	6196	2.32	Chr22	863	0.32
Chr8	5233	1.96	Chr23	901	0.34
Chr9	3829	1.44	Chr24	626	0.23
Chr10	2981	1.12	Chr26	534	0.20
Chr11	2963	1.11	Chr27	1503	0.56
Chr12	2911	1.09	Chr28	753	0.28
Chr13	2670	1.00	ChrZ	33206	12.46
Chr14	1656	0.62	ChrW	4822	0.01

2.3 转座子与基因关系的分析

转座子和基因的关系分析结果见表 3 和图 1 所示，1、2、3、4、Z 号染色体分布的转座子较多（表 3），说明转座子的多少和染色体的长度有关，染色体越长，转座子数目越多；而且转座子比例高，基因的比例则低，反之亦然（图 1）。说明转座子和基因的分布呈现强烈的负相关性。个别转座子在个别区间分布较高的现象说明了转座子在基因组上的分布区间是有一定的偏好性。



A: 1号染色体; B: 2号染色体; C: 4号染色体; D: Z染色体。绿色线: 转座子数量; 蓝色线: 基因数量
A: Chromosome 1; B: Chromosome 2; C: Chromosome 4; D: Chromosome Z. Green line: TEs number; Blue line: Genes number

图1 转座子与基因分布

Fig. 1 Distribution of TEs and genes

2.4 转座子插入基因组位置偏好性

因为转座子的覆盖度和基因密度有负相关关系,我们以超家族分类的转座子为对象,进而分析转座子插入在基因组上的位置与基因在基因组上的位置之间的关联性。通过对转座子插入基因上下游和基因内部的偏好性进行分析发现, LTR/ERV1、LTR/ERV1 等转座子倾向于插入基因外部(图 2-A); DNA 转座子分析发现, DNA/TcMar-Mariner、DNA/PIF-Harbinger、DNA/Hat-Charlie、RC/Helitron 等转座子在基因外部和基因内部都有相对较高的插入,但是更倾向于插入基因外部(图 2-B)

2.5 转座子相关基因 GO 功能注释及分析

为了解这些转座子相关基因功能,我们对 10 845 个基因进行 GO 功能注释。有 7 231 个基因注释在 64 个 GO 条目。对富集转座子数量较多的前 24 个 GO 条目以柱状图方式展示见图 3,由图可知,生物过程条目中富集的转座子相关基因最多(51.8%),富集相对较多条目是单生物过程,代谢过程,生物调节,生物过程的调控,刺激反应等;在细胞组分条目下富集的转座子相关基因占 33%,富集相对较多的是细胞部分,细胞器,膜等;在分子功能中富集的转座子相关基因

占 15.2%,富集相对较多的是结合,催化活性等。

2.6 转座子相关基因功能 Pathway 分析

进一步对 10 845 个与转座子相关基因进行 KEGG 富集分析,共有 2 535 个相关基因注释在 338 个 KEGG 通路中。富集较多的前 20 个通路见表 4,由表可知,主要集中在甘油酯代谢,PPAR 信号通路,PI3K-Akt 信号通路,胰岛素抵抗, Jak-STAT 信号通路等通路。其中着重关注了和浙川乌骨鸡特性相关的通路,如和色素沉积相关的酪氨酸代谢、和肉质品质相关的脂代谢、脂肪酸的合成、PI3K-Akt 信号通路等(图 4),这些通路下的转座子具生物学功能还需进一步研究。

3 讨论

对浙川乌骨鸡全基因组水平转座子的特征及转座子相关基因生物信息学分析等方面进行了研究。结果表明,在浙川乌骨鸡基因组中有大量的转座子存在,且在基因组上的分布不是随机的,具有区域偏好性的,与前人研究一致^[22-23]。有些转座子倾向于插入在基因内部,有些转座子在基因外部。本研究使得我们对浙川乌骨鸡中转座子的特性有了更进一步的了解。

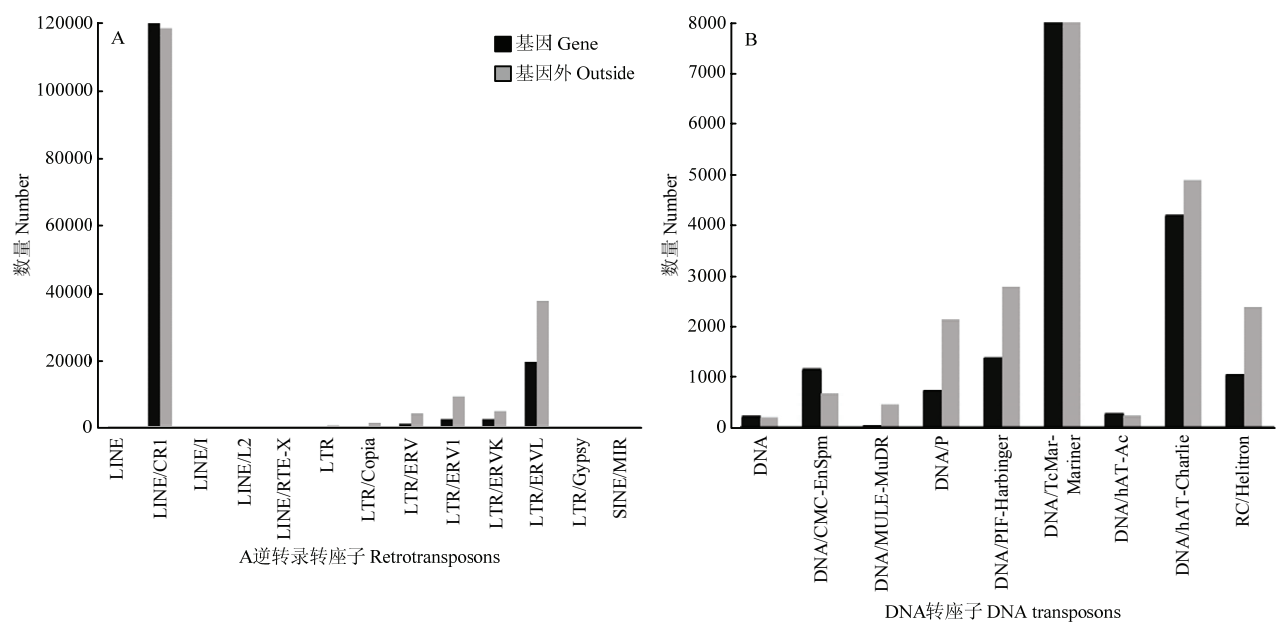


图 2 转座子插入基因组分布偏好性

Fig. 2 Preferential distribution of TEs in genes in genome

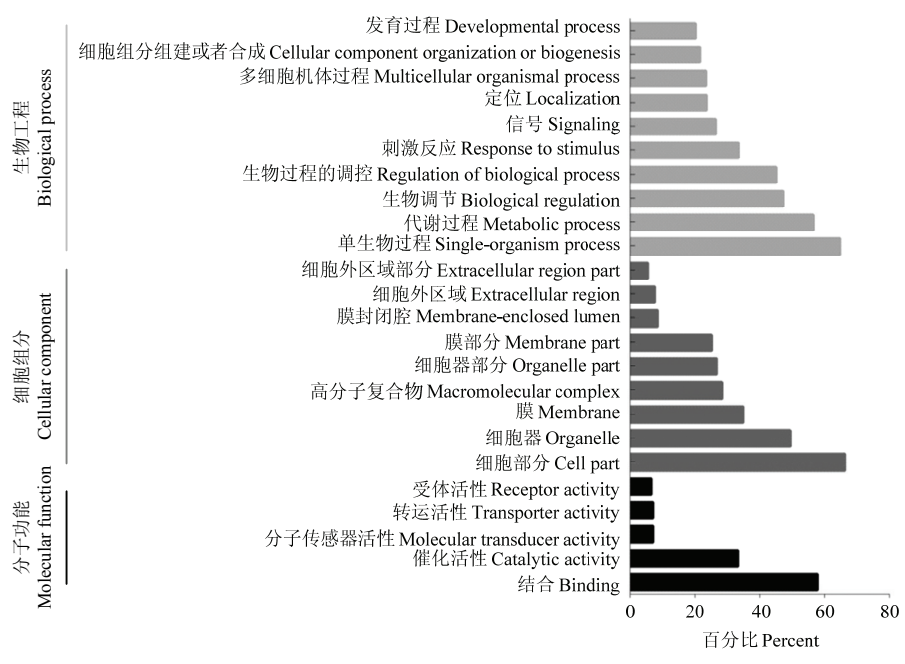


图 3 转座子相关基因的 GO 富集分析

Fig. 3 GO categories enrichment analysis of genes related to TEs

3.1 转座子所占基因组比例

近年来，大规模基因组测序技术为转座子的研究提供了新的契机，通过生物信息学分析发现，在很多的物种中都有转座子的存在，所占基因组的比例等出

现不同的分布趋势，如占拟南芥基因组的 14%^[24]，玉米基因组的 84.2%^[25]，桑树基因组的 37.87%^[26]，人类基因组的 44%^[27]，小鼠基因组的 40%^[28]，斑马鱼基因组的 54.96%^[29]，占浙川乌骨鸡基因组的 18.29%；虽

表 4 转座子相关基因富集较多的 20 个代谢通路

Table 4 Top 20 metabolic pathways involving genes related to TEs

代谢通路	KEGG ID	KEGG 通路名称	Name of KEGG	数量	Numbers
甘油酯代谢	ko00561	Glycerolipid metabolism		392	
PPAR 信号通路	ko03320	PPAR signaling pathway		391	
PI3K-Akt 信号通路	ko04151	PI3K-Akt signaling pathway		211	
胰岛素抵抗	ko04931	Insulin resistance		173	
Jak-STAT 信号通路	ko04630	Jak-STAT signaling pathway		169	
肌动蛋白细胞骨架调节	ko04810	Regulation of actin cytoskeleton		143	
泛素介导的蛋白水解	ko04120	Ubiquitin mediated proteolysis		139	
嘧啶代谢	ko00240	Pyrimidine metabolism		137	
钙信号途径	ko04020	Calcium signaling pathway		136	
癌症中的蛋白多糖	ko05205	Proteoglycans in cancer		129	
神经活性配体-受体相互作用	ko04080	Neuroactive ligand-receptor interaction		123	
粘附斑激酶	ko04510	Focal adhesion		121	
Ras 信号通路	ko04014	Ras signaling pathway		116	
谷氨酸突触	ko04724	Glutamatergic synapse		115	
MAPK 信号通路	ko04010	MAPK signaling pathway		109	
HTLV-I 感染	ko05166	HTLV-I infection		106	
cAMP 信号通路	ko04024	cAMP signaling pathway		98	
磷脂酶 D 信号通路	ko04072	Phospholipase D signaling pathway		98	
表皮生长因子受体酪氨酸激酶抑制剂耐药性	ko01521	EGFR tyrosine kinase inhibitor resistance		92	
胞吞	ko04144	Endocytosis		92	

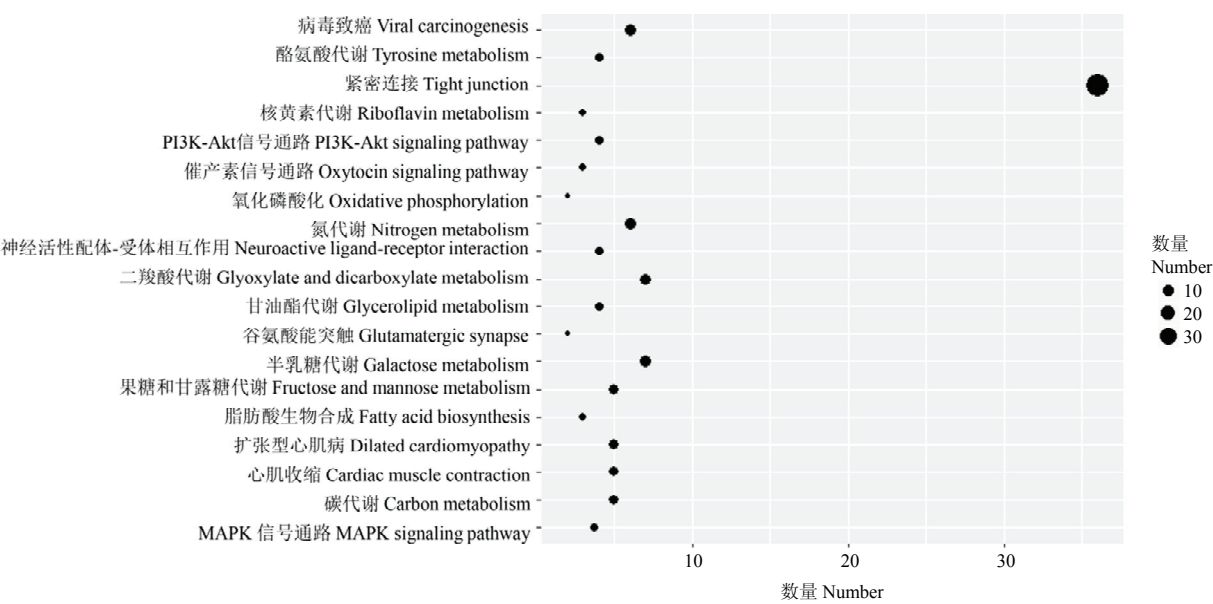


图 4 转座子相关基因的 Pathway 富集分析

Fig. 4 Pathway enrichment analysis of TEs related genes

然各个物种的转座子含量不同,然而其转座子含量与基因组大小密不可分,具有一定的正相关关系。通常认为基因组大小与 DNA 含量增加和删减丢失共同作用有关,其中转座子拷贝的增多是导致基因组增大的一个重要因素^[30-31]。

3.2 转座子和基因的关系

虽然转座子占浙川乌骨鸡基因组的 18.29%,但是这些转座子在基因组上的分布却并非是随机的,呈现相反趋势,基因分布集中的位置,转座子的分布会减少,在拟南芥和人等其他一些测序物种的转座子的比较中也是相同的发现^[32-34],这种现象可能是基因组自身的一种保护方法,避免转座子的插入导致功能基因的失活。通过对转座子和基因之间的相对位置进行分析,以确定是否某些转座子在基因组上存在一定的偏好性,进而影响某个基因的功能。研究发现,一些超家族的转座子在基因组上位置分布确实存在一定偏好性。可能和浙川乌骨鸡的进化历程有密切关系,某些转座子的激活受不同的环境、遗传等因素的影响,大量扩增,从而导致进一步富集到基因组的某些区域。

3.3 Pathway 富集分析

通过进一步对转座子相关基因功能注释,这些转座子相关基因参与了多个生物学过程,物种的表型、生产性能等受转座子的调节。已有研究表明,通过分子重构分别获得的 Sleeping Beauty、Frog Prince 和 Hsmar1 复活的活性 Tc1/Mariner 转座子,在鱼类、青蛙、鼠等动物中被广泛研究^[35-39]; TAKEUCHI 等^[40]和 ASAKAWA 等^[41]利用 Tol2 转座子介导 Gal4-UAS 系统,影响斑马鱼神经元和行为相关基因功能。LINE-1 转座子通过影响基因的表达或调控,进而导致基因疾病^[42-43]。在家禽中, *ev-21* 转座子的插入导致了慢羽表型的突变^[44-45], *ev-10* 或者 *ev-19* 不仅能够降低白来航鸡 8%—9% 的产蛋率,还能降低蛋重和蛋比重^[46]。*SLCO1B3* 基因的 5'非编码区有转座子的反向插入,从而启动了 *SLCO1B3* 基因在卵壳腺中的特异性表达,形成绿壳蛋^[47]。在研究中, *SLCO1B3* 基因也存在转座子的插入。转座子相关基因在酪氨酸代谢通路中富集,其可能与浙川乌骨鸡“乌嘴、乌腿、乌皮、乌骨、乌肉”等典型的种质特性有关,具体的调控机制还有待进一步研究。

4 结论

通过对浙川乌骨鸡全基因组重测序数据的转座子

分析,基于 80-80-80 规则,转座子分为 19 个超家族,在各条染色体的位置分布研究发现,它们在染色体上的分布不均匀,与基因的密度成反比,在基因密集的区域中转座子密度较低,转座子在基因组内的插入非随机的。GO 富集分析表明,转座子相关基因主要在细胞过程、单生物过程、结合、催化活细胞部分、细胞器、膜等条目下富集较多。KEGG 富集分析表明,鉴定出的转座子相关基因主要在甘油酯代谢、PPAR 信号通路、PI3K-Akt 信号通路、胰岛素抵抗、Jak-STAT 信号通路等通路。

随着高通量测序技术的不断提高,使得更好的分析转座子成为可能,进而为揭示转座子的转座机制提供重要的依据,同时也将为研究基因功能,探索物种内部的进化提供一种全新的思路。

References

- [1] OKADA N. Transfer RNA-like structure of the human Alu family: implications of its generation mechanism and possible functions. *Journal of Molecular Evolution*, 1990, 31(6): 500-510.
- [2] FINNEGAN D J. Eukaryotic transposable elements and genome evolution. *Trends in Genetics*, 1989, 5(4): 103-107.
- [3] LISCH D. How important are transposons for plant evolution? *Nature Reviews Genetics*, 2013, 14(1): 49.
- [4] MCCLINTOCK B. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America*, 1950, 36(6): 344.
- [5] DOOLITTLE W F, SAPIENZA C. Selfish genes, the phenotype paradigm and genome evolution. *Nature*, 1980, 284(5757): 601-603.
- [6] HOSKINS R A, CARLSON J W, KENNEDY C, ACEVEDO D, EVANSHOLM M, FRISE E, WAN K H, PARK S, MENDEZLAGO M, ROSSI F. Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science*, 2007, 316(5831): 1625-1628.
- [7] NENE V, WORTMAN J R, LAWSON D, HAAS B, KODIRA C, TU Z J, LOFTUS B, XI Z, MEGY K, GRABHERR M. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*, 2007, 316(5832): 1718-1723.
- [8] GALLAGHER L A, SHENDURE J, MANOIL C. Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. *mBio*, 2011, 2(1): 00315-00310.
- [9] REARDEN A, MAGNET A, KUDO S, FUKUDA M. Glycophorin B and glycophorin E genes arose from the glycophorin A ancestral gene via two duplications during primate evolution. *Journal of Biological Chemistry*. 1993, 268(3): 2260-2267.

- [10] MANDAOKAR A, KUMAR V D, AMWAY M, BROWSE J. Microarray and differential display identify genes involved in jasmonate-dependent anther development. *Plant Molecular Biology*, 2003, 52(4): 775.
- [11] MACIA A, BLANCO-JIMENEZ E, GARC A-P, REZ J L. Retrotransposons in pluripotent cells: Impact and new roles in cellular plasticity. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 2015, 1849(4): 417-426.
- [12] SLOTKIN R K, MARTIENSSEN R. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, 2007, 8(4): 272-285.
- [13] TIAN Z, ZHAO M, SHE M, DU J, CANNON S B, LIU X, XU X, QI X, LI M W, LAM H M. Genome-wide characterization of nonreference transposons reveals evolutionary propensities of transposons in soybean. *Plant Cell*, 2012, 24(11): 4422-4436.
- [14] CLARK L A, WAHL J M, REES C A, MURPHY K E. Retrotransposon insertion in *SILV* is responsible for merle patterning of the domestic dog. *Proceedings of the National Academy of Sciences of the United States of America*, 2006;103(5):1376-1381.
- [15] STEFANIA DO, EMILIO S, LUCA F, MARCO T. Analysis of the 227 bp short interspersed nuclear element (SINE) insertion of the promoter of the myostatin (MSTN) gene in different horse breeds. *Veterinaria Italiana*, 2014;50(3):193-197.
- [16] MIKAWA S, SATO S, NII M, MOROZUMI T, GOU Y, IMAEDA N, YAMAGUCHI T, HAYASHI T, AWATA T. Identification of a second gene associated with variation in vertebral number in domestic pigs. *BMC Genetics*, 2011, 12(1):5-5.
- [17] IVICS Z, GARRELS W, MÁTÉS L, YAU T Y, BASHIR S, ZIDEK V, LANDA V, GEURTS A, PRAVENEK M, RÜLCKE T, KUES W A, IZSVÁK Z. Germline transgenesis in pigs by cytoplasmic microinjection of *Sleeping Beauty* transposons. *Nature Protocols*, 2014, 9(4): 810-827.
- [18] BAILLIE J K, BARNETT M W, UPTON K R, GERHARDT D J, RICHMOND T A, DE S F, BRENNAN P M, RIZZU P, SMITH S, FELL M. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*, 2011, 479(7374): 534-537.
- [19] GOODIER J L. Retrotransposition in tumors and brains. *Mobile DNA*, 2014, 5(1): 11-17.
- [20] WICKER T, SABOT F, HUAVAN A, BENNETZEN J L, CAPY P, CHALHOUB B, FLAVELL A, LEROY P, MORGANTE M, PANAUD O. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 2007(8): 973-982.
- [21] CONESA A, GTZ S, GARC A G, MEZ J M, TEROL J, TAL N M, ROBLES M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 2005, 21(18): 3674-3676.
- [22] ASAF L, SCHRAGA S, GIL A. Large-scale discovery of insertion hotspots and preferential integration sites of human transposed elements. *Nucleic Acids Research*, 2009, 38(5): 1515-1530.
- [23] LINHEIRO R S, BERGMAN C M. Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. *PLoS ONE*, 2012, 7(2): e30008.
- [24] INITIATIVE AG. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 2000, 408(6814): 796-815.
- [25] SCHNABLE P S, WARE D, FULTON R S, STEIN J C, WEI F, PASTERNAK S, LIANG C, ZHANG J, FULTON L, GRAVES T A. The B73 maize genome: complexity, diversity, and dynamics. *Science*, 2009, 326(5956): 1112-1115.
- [26] MA B, XIN Y, KUANG L, HOU F, HE N. Identification and characterization of reverse transcriptase fragments of long interspersed nuclear elements (LINEs) in the *Morus notabilis* genome. *American Journal of Molecular Biology*, 2017, 7(3): 138-152.
- [27] MILLS R E, BENNETT E A, ISKOW R C, DEVINE S E. Which transposable elements are active in the human genome? *Trends in Genetics*, 2007, 23(4): 183-191.
- [28] SMIT A F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Current Opinion in Genetics & Development*, 1999, 9(6): 657-663.
- [29] 高波, 王伟, 钱跃, 陈才, 钟继汉, 沈丹, 陈伟, 宋成义. 斑马鱼转座子时空表达特性. *生物信息学*, 2017, 15(4):201-206.
- GAO B, WANG W, QIAN Y, CHEN C, ZHONG J H, SHEN D, CHEN W, SONG C Y. Temporal and spatial expression characteristics of transposons in zebrafish. *Chinese Journal of Bioinformatics*, 2017, 15(4): 201-206. (in Chinese)
- [30] HAWKINS J S, GROVER C E, WENDEL J F. Repeated big bangs and the expanding universe: Directionality in plant genome size evolution. *Plant Science*, 2008, 174(6): 557-562.
- [31] SANMIGUEL P, BENNETZEN J L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany*, 1998, 82(suppl_1): 37-44.
- [32] PEREIRA V. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biology*, 2004, 5(10): R79.
- [33] WRIGHT S I, AGRAWAL N, BUREAU T E. Effects of recombination rate and gene density on transposable element distributions in

- Arabidopsis thaliana*. *Genome Research*, 2003, 13(8): 1897.
- [34] LOWE C B, BEJERANO G, HAUSSLER D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(19): 8005-8010.
- [35] IZSV K Z, IVICS Z, PLASTERK R H. Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *Journal of Molecular Biology*, 2000, 302(1): 93-102.
- [36] HORIE K, KUROIWA A, IKAWA M, OKABE M, KONDOH G, MATSUDA Y, TAKEDA J. Efficient chromosomal transposition of a Tc1/mariner-like transposon Sleeping Beauty in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, 98(16): 9191-9196.
- [37] CARLSON C M, DUPUY A J, FRITZ S, ROBERGPÉREZ K J, FLETCHER C F, LARGAESPADA D A. Transposon mutagenesis of the mouse germline. *Genetics*, 2003, 165(1): 243.
- [38] MISKEY C, IZSV K Z, PLASTERK R H, IVICS Z. The Frog Prince: A reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. *Nucleic Acids Research*, 2003, 31(23): 6873-6881.
- [39] MISKEY C, PAPP B, MÁTÉS L, SINZELLE L, KELLER H, IZSV K Z, IVICS Z. The ancient mariner sails again: Transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. *Molecular & Cellular Biology*, 2007, 27(12): 4589-4600.
- [40] TAKEUCHI M, MATSUDA K, YAMAGUCHI S, ASAKAWA K, MIYASAKA N, LAL P, YOSHIHARA Y, KOGA A, KAWAKAMI K, SHIMIZU T. Establishment of Gal4 transgenic zebrafish lines for analysis of development of cerebellar neural circuitry. *Developmental Biology*, 2015, 397(1): 1-17.
- [41] ASAKAWA K, SUSTER M L, MIZUSAWA K, NAGAYOSHI S, KOTANI T, URASAKI A, KISHIMOTO Y, HIBI M, KAWAKAMI K. Genetic dissection of neural circuits by Tol2 transposon-mediated Gal4 gene and enhancer trapping in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(4): 1255.
- [42] ZHONGJIE Y, QIPENG L, SHAN C, XIAOYU L. The function of LINE-1-encoded reverse transcriptase in tumorigenesis. *Hereditas*, 2017, 39(5): 368-376.
- [43] QIAN L, JINHUI W, XIAOYU L, SHAN C. The connection between LINE-1 retrotransposition and human tumorigenesis. *Hereditas*, 2016: 93-102.
- [44] BOULLIOU A, LE P J, HUBERT G, DONAL R, SMILEY M. The endogenous retroviral ev21 locus in commercial chicken lines and its relationship with the slow-feathering phenotype (K). *Poultry Science*, 1992, 71(1): 38.
- [45] LU X Q, HAN J R, LIU X F, LIN T H, LI Y L. The LTR of endogenous retrovirus ev21 retains promoter activity and exhibits tissue specific transcription in chicken. *Chinese Science Bulletin*, 2009, 54(24): 4664-4670.
- [46] GAVORA J S, KUHNLEIN U, CRITTENDEN L B, SPENCER J L, SABOUR M P. Endogenous viral genes: Association with reduced egg production rate and egg size in White Leghorns. *Poult Sciences*, 1991, 70(3): 618-623.
- [47] WANG Z, QU L, YAO J, YANG X, LI G, ZHANG Y, LI J, WANG X, BAI J, XU G. An EAV-HP Insertion in 5' flanking region of *SLCO1B3* causes Blue Eggshell in the chicken. *Plos Genetics*, 2013, 9(1): e1003183.

(责任编辑 林鉴非)