

# 二倍体和四倍体泥鳅全基因组 DNA 甲基化的比较

张曼曼 冯 兵 罗双双 王卫民 周小云

华中农业大学水产学院/农业动物遗传育种与繁育教育部重点实验室, 武汉 430070

**摘要** 以二倍体和四倍体泥鳅为研究对象, 采用甲基化修饰依赖性内切酶测序技术(MethylRAD-Seq)在全基因组水平上分析泥鳅的 DNA 甲基化特点及倍性间的 DNA 甲基化变异。测序结果共得到 302 111 684 条 Methyl-RAD 序列标签。与参考基因组比对结果显示, 泥鳅的甲基化位点主要分布在基因体区(gene body), 其次为内含子区(intron)和基因间区(intergenic), 而在其他功能元件上的分布较少。四倍体泥鳅的整体甲基化水平比二倍体高, 尤其是在第一外显子区(1<sup>st</sup> Exon)和转录起始位点上游 1 500 bp 至 200 bp 区(TSS1500), 且倍性间差异极显著( $P < 0.01$ )。但在启动子区, 四倍体泥鳅的甲基化水平略低于二倍体。在二倍体和四倍体泥鳅间共筛选到 1 268 个差异甲基化 C<sup>m</sup>CGG 位点和 14 个差异甲基化 C<sup>m</sup>CWGG 位点, 这些位点主要分布于内含子、基因体和基因间区。比较各基因的甲基化水平, 共得到 684 个倍性间差异甲基化基因。KEGG 分析结果显示, 倍性间差异甲基化基因主要富集到与生长发育、免疫及错配修复等相关通路上。

**关键词** 泥鳅; 二倍体; 四倍体; DNA 甲基化; MethylRAD-Seq

**中图分类号** Q 959.46<sup>+</sup>8 **文献标识码** A **文章编号** 1000-2421(2018)05-0095-09

多倍体化是自然界中普遍存在的一种生物学现象, 也是推动生物进化和新物种形成的重要驱动力。多倍化最直接的影响就是造成基因组中大量功能基因的重复, 这些重复基因的演化机制已经成为分子进化领域的研究热点之一。DNA 甲基化是表观遗传修饰的主要途径之一, 在生物体的基因表达、细胞分化以及系统发育过程中发挥重要的调控作用。研究发现, 多倍体的产生往往伴随着广泛的 DNA 甲基化变异<sup>[1-2]</sup>, 这些变异在控制多倍体基因组中冗余基因的表达和保持基因组稳定方面起着重要的调节作用。然而, 目前相关的研究主要集中在拟南芥<sup>[3]</sup>、水稻<sup>[4]</sup>、小麦<sup>[5]</sup>以及火龙果<sup>[6]</sup>等植物中。鱼类是脊椎动物的最大类群, 其多倍体现象也非常普遍, 鱼类多倍化后 DNA 甲基化修饰变化及其对基因表达调控的研究尚未见报道。

泥鳅(*Misgurnus anguillicaudatus*)是我国常见的小型经济鱼类。已有的研究发现, 泥鳅的染色体存在广泛的核型多态性<sup>[7]</sup>。日本的野生泥鳅群体中, 除二倍体( $2n=50$ )外, 还存在部分三倍体( $3n=75$ )<sup>[7]</sup>。我国的泥鳅不仅有二倍体, 在长江流域, 人

们还发现有大量的四倍体( $4n=100$ )<sup>[8]</sup>。通过诱导雌核发育<sup>[9-10]</sup>、雄核发育<sup>[11]</sup>、六倍体<sup>[12]</sup>, 以及对减数分裂染色体行为的观察<sup>[13]</sup>, 研究人员认为这些染色体数为 100 的泥鳅为同源四倍体, 是由二倍体泥鳅经过染色体加倍形成。泥鳅的这种天然四倍体与二倍体共存的现象, 使之成为研究多倍体鱼类表型变异、基因表达变异以及表观遗传修饰变异的理想模式生物。

随着对 DNA 甲基化研究的不断深入, 各种 DNA 甲基化检测方法被开发出来以满足不同类型研究的需求。其中, 甲基化敏感扩增多态性(MSAP)是目前使用最多的一项技术, 具有较高的可靠性和有效性, 但该方法无法做到对位点的定量分析。在模式生物中, 较多的采用全基因组亚硫酸氢盐测序(WGBS)、简化的亚硫酸氢盐测序(RRBS)、甲基免疫共沉淀测序(MeDIP-seq)等方法, 但这些方法只适用于有参考基因组的物种, 对于遗传信息相对匮乏的非模式生物却无能为力。基于此, Wang 等<sup>[14-15]</sup>开发了一种甲基化修饰依赖性内切酶测序技术(methylation-dependent restriction-

收稿日期: 2018-01-22

基金项目: 国家自然科学基金项目(31472267)

张曼曼, 硕士研究生. 研究方向: 鱼类表观遗传学. E-mail: 1527050643@qq.com

通信作者: 周小云, 博士, 副教授. 研究方向: 鱼类表观遗传学、鱼类遗传育种学. E-mail: zhouxy@mail.hzau.edu.cn

site associated DNA sequencing, MethylRAD-Seq),该方法利用一类类似于ⅡB型的甲基化修饰依赖性内切酶(如 *FspEI*、*MspJI* 等)对基因组DNA进行酶切,酶切后产生以  $C^m$ CGG 和  $C^m$ CWGG 为核心甲基化位点的等长标签,对标签文库进行高通量测序就能获得全基因组范围内的甲基化位点序列信息。根据等长标签扩增效率的一致性,甲基化标签的测序深度可以反映位点( $C^m$ CGG/ $C^m$ CWGG)的甲基化水平。因此,在不需要基因组背景信息的条件下可以大规模发掘全基因组范围内的甲基化位点,直接获得甲基化胞嘧啶的序列信息。

本研究以二倍体和四倍体泥鳅为对象,采用 MethylRAD-Seq 技术分析泥鳅的DNA甲基化特点,以及不同倍性泥鳅间的甲基化水平差异及模式变化特征,并对倍性间差异甲基化基因进行GO注释和KEGG Pathway富集分析,探讨DNA甲基化对多倍体泥鳅的基因表达调控作用,以及引起鱼类倍性间表型差异的表现遗传机制。

## 1 材料与方法

### 1.1 实验鱼来源

研究所用二倍体和四倍体泥鳅为笔者所在实验室人工繁育个体。将二倍体和四倍体泥鳅鱼苗分池饲养至2月龄后,分别在背鳍基部注射不同颜色的荧光标记(code wire tag, USA),然后将2种倍性的泥鳅进行混合饲养。养殖至1龄时,随机选取二倍体和四倍体泥鳅各3尾,用于本研究。

### 1.2 DNA的提取和质检

将实验鱼用MS-222(100 mg/L)麻醉后,解剖分离肝脏组织,用QIAamp DNA Mini Kit (250)试剂盒提取基因组DNA,用1%的琼脂糖凝胶电泳检测DNA的完整性,用NanoDrop 2000紫外分光光度计测定DNA的浓度。

### 1.3 MethylRAD-Seq测序文库的构建与测序

MethylRAD-Seq测序文库采用Wang等<sup>[15]</sup>的五标签串联技术构建。具体为:用甲基化修饰依赖性内切酶 *FspEI* 对基因组DNA进行酶切;设计3'端带有4个简并碱基N的接头,与酶切产物连接;PCR扩增富集连接产物,并通过8%聚丙烯酰胺凝胶电泳切取目的条带(100 bp);将同一串5个标签样品的富集产物混合,用 *SapI* 对混合产物进行酶切;将酶切产物加入磁珠中,去除接头序列后与T4 DNA连接酶连接;回收连接产物(约244 bp),通过

PCR为成串的混合产物引入Barcode序列,用1%琼脂糖凝胶电泳检测目的条带(约299 bp),用纯化试剂盒进行纯化,Qubit 2.0质检合格后,使用Illumina HiSeq X Ten平台进行双末端测序(PE150)。

### 1.4 数据处理与分析

将测序获得的原始数据进行质控过滤,去除含有接头、N碱基以及低质量的序列,得到高质量的序列。拼接成对的高质量序列,去除不含酶切识别位点的序列,得到含甲基化位点的高质量测序数据(Methyl-RAD标签)。

由于目前在公共数据库中还没有泥鳅的参考基因组序列,因此,本研究选用斑马鱼的基因组作为参考,从中提取含有 *FspEI* 酶切识别位点( $C^m$ CGG和 $C^m$ CWGG)的序列构建参考序列。然后使用SOAP软件将各样品的Methyl-RAD标签序列与参考序列进行比对,统计甲基化位点数目及位点的深度信息。

根据等长标签扩增效率的一致性,甲基化标签的测序深度可以反映位点( $C^m$ CGG/ $C^m$ CWGG)的甲基化水平,基于此,对各位点进行甲基化水平相对定量,以实现样品间或倍性间各位点甲基化水平的可比性,相对定量方法采用RPM(reads per million)计算。

基于各位点在每个样品中的测序深度信息,统计各基因的甲基化程度,用负二项分布精确检验,计算各位点以及各基因在倍性间的差异 $P$ 值及差异倍数( $\log_2FC$ ), $P < 0.05$ 且 $\log_2FC > 2$ 表明存在显著性差异。对倍性间差异甲基化基因进行GO功能富集分析、KEGG注释及Pathway富集分析。

### 1.5 亚硫酸氢盐测序验证差异甲基化位点

随机选择7个 $C^m$ CGG位点进行亚硫酸氢盐测序(BSP-Seq)验证,取同份泥鳅肝脏的DNA 500 ng按照EZ DNA Methylation-Gold Kit<sup>TM</sup>试剂盒的步骤进行亚硫酸氢盐处理。以处理后的DNA为模板进行PCR扩增,体系为50  $\mu$ L,包括DNA 4  $\mu$ L, TaKaRa *Taq*<sup>TM</sup>HS(5 U/ $\mu$ L) 0.5  $\mu$ L, 1 $\times$ PCR Buffer, dNTPs (2.5 mmol/L) 4  $\mu$ L, Forward/Reverse Primer (10  $\mu$ mol/L) 1  $\mu$ L(表1)。PCR程序为95 $^{\circ}$ C预变性5 min;95 $^{\circ}$ C变性30 s,  $T_m$ 退火40 s, 72 $^{\circ}$ C延伸30 s,进行35个循环;最后72 $^{\circ}$ C延伸8 min。用2%琼脂糖凝胶电泳检测PCR产物,回收目的片段连接到T载体,转化至感受态细胞中,

表 1 倍性间差异 C<sup>m</sup>CGG 甲基化位点 BSP-seq 验证所用引物

Table 1 Primers for BSP-seq validation of differential methylated C<sup>m</sup>CGG sites between ploidy

| 甲基化位点<br>Methylated locus | 位点标签<br>Locus tag    | 引物序列(5'-3')<br>Primer sequences(5'-3')                         |
|---------------------------|----------------------|----------------------------------------------------------------|
| Locus 1                   | NC_007114.6-13705823 | F: AAGTGGTTTTAGGTTAGGGTTAATG<br>R: CAAATCCAAAACACTCTAAAACTCTAC |
| Locus 2                   | NC_007114.6-13453513 | F: TTTGAGTTTTGTGAGAATTGGATA<br>R: AACAACAATACCCCTAATTAAACA     |
| Locus 3                   | NC_007125.6-31918368 | F: AGGATGTTTTAAAGGTTTAGGATGA<br>R: ACTTCAATAAACAACAAAAAATTTC   |
| Locus 4                   | NC_007126.6-46389960 | F: GTGGATTTTTTTTGAATAAAGAGTG<br>R: TAAAACCCAAACAAATAAACAAAAAC  |
| Locus 5                   | NC_007135.6-36529469 | F: ATGTTGGATGGATTGAATAATATAT<br>R: ACCTACAACCTTTAACCTCCATTAC   |
| Locus 6                   | NC_007127.6-41782360 | F: TAATGGTATTAGGTTGGTAAGAGAT<br>R: ATAAAACCAACAAAAAACCAAAAC    |
| Locus 7                   | NC_007112.6-36112706 | F: GATGGTATGAAAAAGTTTTTAGAGTTT<br>R: AAAATAAACTTCCAAATCACAACC  |

每个位点至少挑选 30 个单克隆测序。测序获得的目的序列用 QUMA 软件进行甲基化水平分析。

## 2 结果与分析

### 2.1 测序数据统计

通过 MethylRAD-Seq 测序,本研究共得到 530 192 536 条高质量序列,每条序列长 28 bp。将双末端测序的成对序列拼接后,去除不含预期酶切位点的序列,共得到 302 111 684 个 Methyl-RAD 序列标签(表 2)。

将各样品的 Methyl-RAD 标签序列比对到参考标签序列,由于基因组中存在重复序列,位于重复序列的标签能够同时比对到基因组的多个位置,对于这种甲基化标签的实际发生位置难以进行判断。因此,在进行比对时,不考虑重复比对的标签。6 个文库的高质量甲基化标签能比对到基因组唯一位置的比率为 2.19%~2.57%。比对序列数的多少反映了基因组 DNA 甲基化的水平,可以看出,四倍体泥鳅的总体甲基化水平比二倍体高,但这种差异不具有显著性( $P>0.05$ )。

表 2 各样品测序数据量与比对率

Table 2 Sequenced data of the samples and alignment rate to the reference genome

| 样本<br>Sample | 高质量数据<br>Clean data | Methyl-RAD 标签<br>Methyl-RAD tags | 数据量/bp<br>Total clean data | 唯一比对数<br>Unique alignment tags | 比对率/%<br>Unique alignment ratio |
|--------------|---------------------|----------------------------------|----------------------------|--------------------------------|---------------------------------|
| 2N-1         | 90 393 069          | 49 859 198                       | 1 594 534 183              | 1 157 695                      | 2.32                            |
| 2N-2         | 90 393 069          | 50 286 628                       | 1 607 667 740              | 1 216 760                      | 2.42                            |
| 2N-3         | 90 393 069          | 53 160 461                       | 1 699 228 545              | 1 162 006                      | 2.19                            |
| 4N-1         | 90 393 069          | 54 303 885                       | 1 735 243 388              | 1 351 780                      | 2.49                            |
| 4N-2         | 84 310 130          | 46 378 983                       | 1 483 021 296              | 1 192 679                      | 2.57                            |
| 4N-3         | 84 310 130          | 48 122 529                       | 1 537 867 695              | 1 163 158                      | 2.42                            |

### 2.2 甲基化位点在基因组不同功能元件上的分布

统计二倍体和四倍体泥鳅的甲基化位点在基因组不同功能元件上的分布,包括第一外显子(1<sup>st</sup> Exon)、基因体(gene body)、转录起始位点上游 200 bp 内(TSS200)、转录起始位点上游 1 500 bp 至 200 bp

区(TSS1500)、内含子(intron)和基因间区(intergenic)。结果显示,泥鳅的甲基化 C<sup>m</sup>CGG 位点主要分布在基因体区,其次为内含子区和基因间区,而在其他功能元件上(1<sup>st</sup> Exon、TSS1500 和 TSS200)的分布较少(图 1)。甲基化 C<sup>m</sup>CGG 和 C<sup>m</sup>CWGG 位

点的分布情况基本一致(图 1)。

比较甲基化位点在二倍体和四倍体泥鳅间的差异发现,在各功能元件上,四倍体泥鳅的甲基化  $C^mCGG$  位点均比二倍体高( $P < 0.05$ ),且在  $1^{st}$  Exon 和 TSS1500 区,倍性间的甲基化水平存在极显著性差异( $P < 0.01$ )。四倍体泥鳅的基因体、内含子和基因间区的甲基化  $C^mCWGG$  位点比二倍体略高,但  $1^{st}$  Exon、TSS1500 和 TSS200 区比二倍体略低,

且在倍性间的差异均不显著( $P > 0.05$ )。

### 2.3 甲基化位点在不同基因区域的分布

分析甲基化位点在转录起始位点(TSS)上下游各 2 kb 区域、转录终止位点(TTS)上下游各 2 kb 区域、以及基因体的分布,结果显示,二倍体、四倍体泥鳅基因体的甲基化位点数( $C^mCGG/C^mCWGG$ )明显高于 TSS 和 TTS 上下游 2 kb 区域(图 2)。

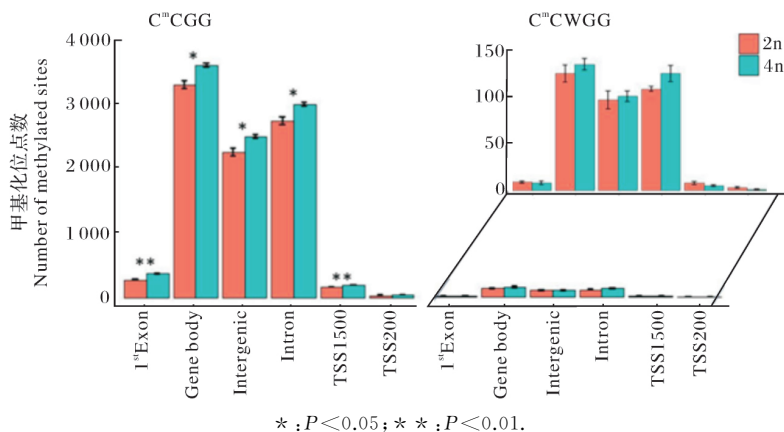


图 1 甲基化位点在基因组各功能元件上的分布

Fig.1 Distribution histogram of methylated sites in each function element of genome

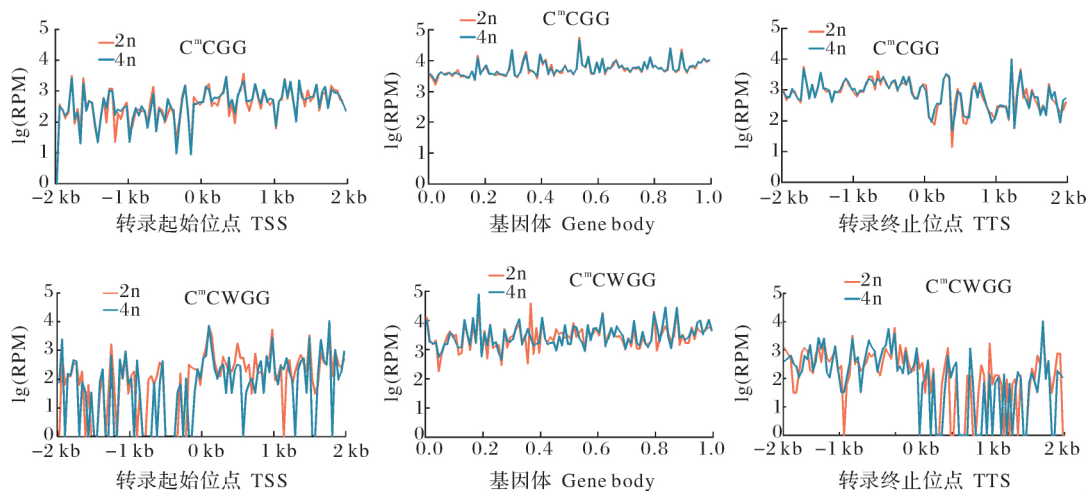


图 2 甲基化位点在基因体(gene body)、转录起始位点(TSS)和转录终止位点(TTS)上下游各 2 kb 区域的分布

Fig.2 Distribution of methylated sites in the gene body, 2 kb of upstream and downstream region of the TSS and TTS

比较二倍体和四倍体泥鳅间甲基化  $C^mCGG$  位点在上述 3 个区域的分布情况,结果显示,在 TSS 上游 2 kb 区域内,四倍体泥鳅的平均甲基化水平比二倍体低 10.19%,而在启动子区(TSS 上游 1.5 kb 至下游 0.5 kb),四倍体泥鳅的平均甲基化水平只比二倍体低 1.61%。在基因体和 TTS 下游 2 kb 区,四倍体泥鳅的甲基化水平分别比二倍体高 0.88% 和 1.74%(表 3)。与甲基化  $C^mCGG$  位点相比,甲基化

$C^mCWGG$  在二倍体、四倍体泥鳅间的变异幅度较大,四倍体泥鳅启动子区的甲基化  $C^mCWGG$  位点数比二倍体少 25.45%,而在启动子、基因体和 TTS 下游 2 kb 区分别比二倍体高 9.09%、16.42% 和 10.38%(表 3)。

### 2.4 泥鳅倍性间差异甲基化位点的分布

根据甲基化  $C^mCGG$  和  $C^mCWGG$  位点在各样品中的测序深度信息,筛选泥鳅倍性间差异甲基化

表 3 不同基因区域上 DNA 甲基化在二倍体、四倍体泥鳅间的差异比较

| Table 3 The difference compares of DNA methylated sites between diploid and tetraploid <i>M. anguillicaudatus</i> in different gene regions |                                 |                                  |
|---------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|----------------------------------|
| 区间<br>Region                                                                                                                                | 4n/2n<br>(C <sup>m</sup> CGG)/% | 4n/2n<br>(C <sup>m</sup> CWGG)/% |
| 启动子区 Promoter                                                                                                                               | -1.61                           | -25.45                           |
| TSS 上游 2 kb<br>2 kb of upstream of the TSS                                                                                                  | -10.19                          | +9.09                            |
| 基因体区<br>Gene body                                                                                                                           | +0.88                           | +16.42                           |
| TTS 下游 2 kb<br>2 kb of downstream of the TTS                                                                                                | +1.74                           | +10.38                           |

注：负号(−)表示与二倍体泥鳅相比，四倍体的甲基化水平降低，而正号(+)表示四倍体的甲基化水平升高。Note: Minus(−) means methylation levels in tetraploids were lower than diploids, while plus(+) means higher than diploids.

位点，结果共得到 1 268 个差异甲基化 C<sup>m</sup>CGG 位点和 14 个差异甲基化 C<sup>m</sup>CWGG 位点。分析差异甲基化位点在基因组各功能元件上的分布，结果显示，差异甲基化 C<sup>m</sup>CGG 位点主要分布在内含子(35.74%)、基因体(28.48%)和基因间区(24.74%)，而在 1<sup>st</sup> Exon、TSS1500 和 TSS200 区的分布较少。差异甲基化 C<sup>m</sup>CWGG 位点的分布趋势与差异甲基化 C<sup>m</sup>CGG 位点类似(图 3)。

2.5 二倍体和四倍体泥鳅间基因差异甲基化分析

根据各甲基化位点的测序深度信息，统计各基因的甲基化水平，比较二倍体、四倍体泥鳅间的差异

甲基化基因，结果共得到 671 个 C<sup>m</sup>CGG 位点所在的差异甲基化基因和 12 个 C<sup>m</sup>CWGG 位点所在的差异甲基化基因。与二倍体相比，四倍体泥鳅中有 404 个 C<sup>m</sup>CGG 位点所在的基因甲基化水平升高，267 个基因的甲基化水平降低；8 个 C<sup>m</sup>CWGG 位点所在的基因的甲基化水平升高，4 个甲基化水平降低。

2.6 倍性间差异甲基化基因的 GO 功能富集

倍性间差异甲基化基因 GO 注释显示，有 419 个 C<sup>m</sup>CGG 位点所在的基因与 GO 数据库中的基因具有相似性，占总数的 62.44%，且较多的单个倍性间差异甲基化基因与多种基因相对应，共建立了 1 962 条对应关系。根据富集的显著性(P 值)，分别筛选分子功能、生物学过程和细胞组件中位于前 10 位的通路如图 4 所示。C<sup>m</sup>CGG 差异甲基化基因分析结果显示，在生物学过程中，与精氨酸-tRNA 连接酶活性(arginine-tRNA ligase activity)相关的基因富集最为显著；在生物学过程中，归入精氨酰-tRNA 氨酰化(arginyl-tRNA aminoacylation)和动作电位中膜去极化(membrane depolarization during action potential)的基因富集最为显著；而在细胞组件中，位于 t-RNA(m1A)甲基转移酶复合体(tRNA(m1A) methyltransferase complex)上的基因富集最为显著。C<sup>m</sup>CWGG 的差异甲基化基因中，在神经连接蛋白家族(neuroligin family protein binding)、突触后密度组装中的桥尾蛋白聚类(gephyrin clustering involved in postsynaptic density assembly)等 GO Term 中显著富集。

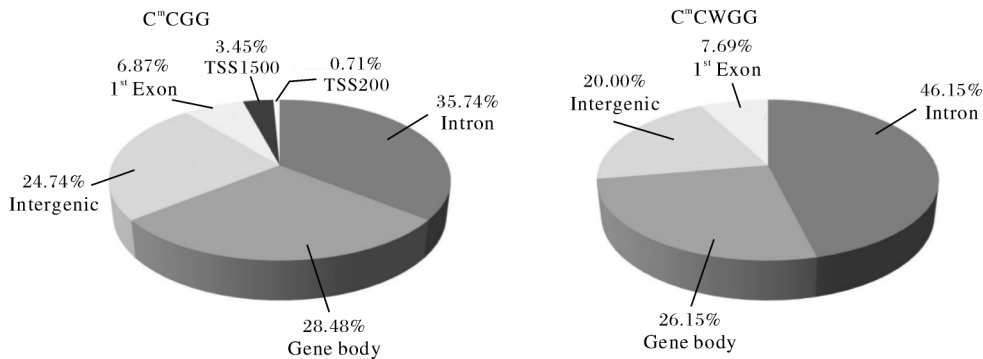


图 3 差异甲基化位点在基因组各功能元件上的分布

Fig.3 Distribution of differential methylated sites in each function element of the genome

2.7 倍性间差异甲基化基因的 KEGG Pathway 富集分析

为了解差异甲基化基因的生物学功能及作用，本研究对筛选得到的差异甲基化基因进行了 Path-

way 分析。结果显示， $P < 0.05$  的基因显著富集通路有 29 个，较多的基因富集在轴突导向(axon guidance)、MAPK(MAPK signaling pathway)、肌动蛋白细胞骨架调节(regulation of actin cytoskeleton)

和 rap1(rap1 signaling pathway)等信号通路上。其中,MAPK、肌动蛋白细胞骨架调控(regulation of actin cytoskeleton)、细胞粘连(focal adhesion)、ECM受体互作(ECM-receptor interaction)等7个通路跟个体的生长发育相关,而Toll和Imd信号通路(Toll and Imd signaling pathway)、血小板活化(platelet activation)等8个通路跟免疫相关,此外,与遗传信息过程相关的错配修复通路(mismatch

repair)也有显著富集(图5)。

## 2.8 BSP-seq 法验证 MethylRAD-Seq 测序数据

为验证 MethylRAD-Seq 测序结果,本研究随机选择了7个C<sup>m</sup>CGG位点进行BSP-Seq实验,结果显示,所有7个位点的甲基化水平与 MethylRAD-Seq 测序结果一致(图6),2种方法对位点甲基化水平定量的相关系数为0.9945,表明本研究利用 MethylRAD-Seq 测序获得的数据较为准确。

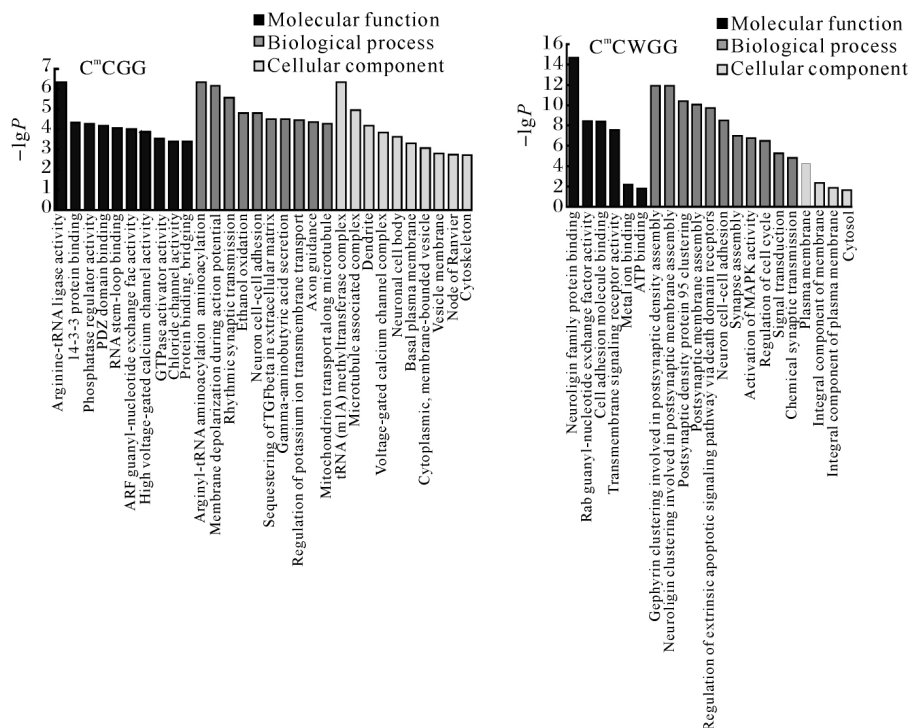


图4 二倍体、四倍体泥鳅间差异甲基化基因的GO富集分析

Fig.4 Go enrichment analyses of differential methylated genes between diploid and tetraploid *M. anguillicaudatus*

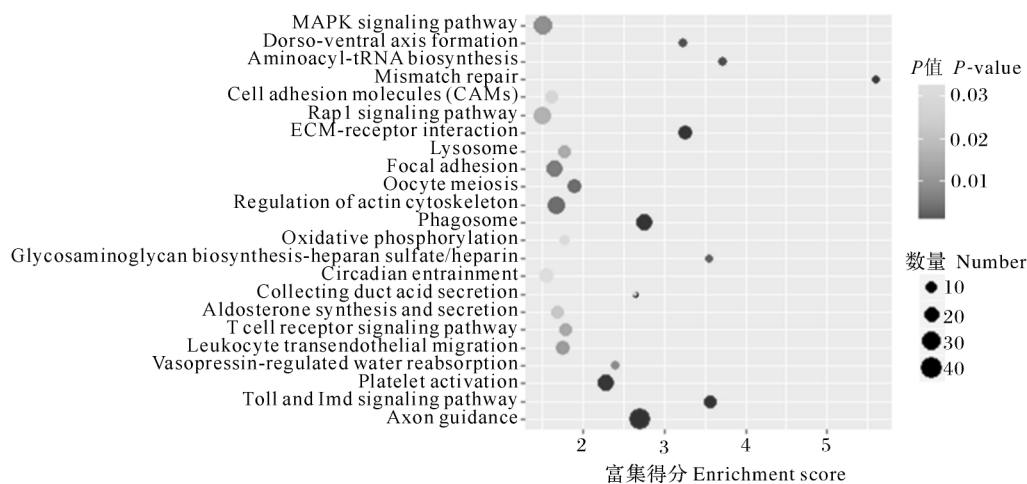


图5 二、四倍体泥鳅间差异甲基化基因KEGG Pathway富集气泡图

Fig.5 Enriched bubble chart of KEGG gathering for the KEGG functional classification for the genes with differential methylation genes between-ploidy

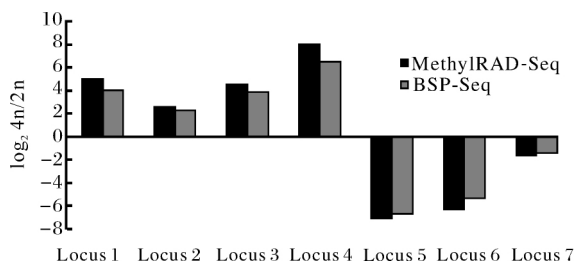


图 6 亚硫酸氢盐测序验证甲基化组测序数据

Fig.6 Validation of the MethyIRAD-Seq data by BSP-Seq for selected methylated sites

### 3 讨论

本研究采用 MethyIRAD-Seq 技术分析了二、四倍体泥鳅全基因组范围内 C<sup>m</sup>CGG 和 C<sup>m</sup>CWGG 位点的甲基化水平和模式信息。结果显示,在泥鳅中,基因体尤其是外显子区的 DNA 甲基化远比其他区域高,这与在其他物种包括植物<sup>[16]</sup>、脊椎动物<sup>[17]</sup>和无脊椎动物<sup>[18]</sup>中的研究结果一致,比如,在海鞘基因组中,约有 60% 的基因存在外显子区甲基化,占其全基因组甲基化的大部分<sup>[18]</sup>。蜜蜂基因组整体甲基化水平远低于脊椎动物,已检测到的甲基化均发生于外显子区,而基因启动子、重复元件和转座子并不发生甲基化<sup>[19]</sup>。Bonasio 等<sup>[20]</sup>认为,外显子 DNA 甲基化,尤其是外显子-内含子边界甲基化程度的急剧转变可能与 RNA 的选择性剪切有关。这种推测在对蜜蜂的研究中得到证实,GB18602 基因在蜂后和工蜂中由于甲基化模式的不同而产生一长一短 2 种不同的转录物,其较长的转录物中缺失了 1 个包含终止密码子的外显子<sup>[21]</sup>。本研究中,四倍体泥鳅基因体的甲基化水平比二倍体高,推测多倍体泥鳅在形成的过程中,可能是通过 DNA 甲基化增加转录变异体 (transcript variants) 的数量来提高表型的灵活性,从而增加四倍体泥鳅对环境的适应能力。

与基因体甲基化调控基因的选择性剪切不同,启动子区的甲基化往往会抑制基因的表达。比如在海兔中,CREB2 基因启动区 CpG 岛的甲基化水平与基因表达呈负相关,并且会影响到记忆相关突触的可塑性调控<sup>[22]</sup>;拟南芥 FW4 基因的启动子区经甲基化修饰后,基因的表达被抑制<sup>[23]</sup>。Schübeler<sup>[24]</sup>指出,许多转录因子,如 AP-2、c-Myc/Myn、CREB 等,对其结合位点的 DNA 甲基化非常

敏感,即使是在 GC 含量相对较少的启动子、增强子等基因调控区,甲基化也会影响转录因子与之结合,从而抑制基因的表达,因此,在拟南芥中,95% 表达基因的启动子区呈低或无甲基化状态<sup>[23,25]</sup>。我们的研究发现,与其他区域相比,泥鳅启动子区 (TSS1500、TSS200 和 1<sup>st</sup> Exon) 也呈现较低的甲基化状态。此外,与二倍体相比,四倍体泥鳅启动子区的甲基化水平略有降低,我们推测,在泥鳅基因组多倍化的过程中,启动子区的甲基化模式发生了部分的解除或重新编排,从而诱发某些基因的激活与沉默,使新形成的四倍体泥鳅能够更好地适应环境的需要。

在二倍体、四倍体泥鳅间共筛选到 683 个差异甲基化基因,与二倍体相比,四倍体泥鳅中有 412 个基因的甲基化水平升高,271 个基因的甲基化水平降低。Pathway 富集分析显示,差异甲基化基因主要富集到生长发育、免疫和错配修复等信号通路上。Cui 等<sup>[26]</sup>采用全基因表达芯片法比较不同生长速度鸡的胸脯肉,发现与细胞连接相关通路(包括紧密连接、肌动蛋白细胞骨架调控、细胞粘连和 ECM 受体互作)在肌内脂肪沉积中起着重要作用。本研究中,二倍体、四倍体泥鳅间差异甲基化基因在这些通路中也有显著富集,Feng 等<sup>[27]</sup>研究发现,四倍体泥鳅的肥满度显著高于二倍体,由于 DNA 甲基化与基因表达之间存在密切相关性,倍性间差异甲基化基因在这些通路上的富集表明,DNA 甲基化的变异可能是影响该通路上基因表达,进而影响倍性间肥满度差异的原因之一。与遗传信息相关的错配修复通路是真核生物维持遗传信息稳定的重要调控机制,已有的研究表明,该通路中的基因与 DNA 的重新甲基化<sup>[28]</sup>、G/T 错配的切除<sup>[29]</sup>等有偶联关系,泥鳅基因组多倍化过程中甲基化模式的变异必然经历去甲基化、从头甲基化等一系列过程,错配修复通路中的差异甲基化基因可能参与了多倍体泥鳅 DNA 甲基化模式的调整过程,进而在维持多倍体泥鳅基因组稳定方面起到了重要的作用。本研究结果不仅可为进一步开展泥鳅多倍化中基因组稳定的内在机制研究奠定基础,同时也为探索鱼类倍性间表型差异的分子机制提供可靠的参考依据。

### 参 考 文 献

- [1] ZENG Z X, ZHANG T, LI G R, et al. Phenotypic and epigenet-

- ic changes occurred during the autopolyploidization of *Aegilops tauschii* [J]. Cereal Res Commun, 2012, 40(4): 476-485.
- [2] XIAO J, SONG C, LIU S J, et al. DNA methylation analysis of allotetraploid hybrids of red crucian carp (*Carassius auratus* var.) and common carp (*Cyprinus carpio* L.) [J]. PLoS ONE, 2013, 8(2): e56409.
- [3] MADLUNG A, MASUELLI R W, WATSON B, et al. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids [J]. Plant Physiol, 2002, 129(2): 733-746.
- [4] 吴绍华, 薛晶晶, 张红宇, 等. 双胚苗水稻单倍体及其杂交后代基因组 DNA 甲基化特异位点的分析及功能探讨[J]. 中国水稻科学, 2011, 25(3): 249-255.
- [5] 彭海, 张静, 吴先军. 植物基因表达中的倍性效应: 研究进展、问题与展望[J]. 中国科学(C 辑), 2008, 38(1): 1-7.
- [6] 刘鹏飞, 乔光, 文晓鹏. 火龙果组培苗 DNA 甲基化变化及应答赤霉素效应[J]. 华中农业大学学报, 2016, 35(5): 18-26.
- [7] ARAI K. Genetics of the loach, *Misgurnus anguillicaudatus*, recent progress and perspective [J]. Folia Biol (Kraków), 2003, 51(suppl): 107-117.
- [8] 印杰, 赵振山, 陈小奇, 等. 二倍体和四倍体泥鳅染色体组型比较[J]. 水生生物学报, 2005, 29(4): 469-472.
- [9] ARAI K, MATSUBARA K, SUZUKI R. Chromosomes and developmental potential of progeny of spontaneous tetraploid loach *Misgurnus anguillicaudatus* [J]. Nippon Suisan Gakk, 1991, 57(12): 2173-2178.
- [10] ARAI K, MATSUBARA K, SUZUKI R. Production of polyploids and viable gynogens using spontaneously occurring tetraploid loach, *Misgurnus anguillicaudatus* [J]. Aquaculture, 1993, 117(3/4): 227-235.
- [11] ARAI K, IKENO M, SUZUKI R. Production of androgenetic diploid loach *Misgurnus anguillicaudatus* using spermatozoa of natural tetraploids [J]. Aquaculture, 1995, 137: 131-138.
- [12] ARAI K, TANIURA K, ZHANG Q. Production of second generation progeny of hexaploid loach [J]. Fisheries Sci, 1999, 65(2): 186-192.
- [13] LI Y J, YU Z, ZHANG M Z, et al. The origin of natural tetraploid loach *Misgurnus anguillicaudatus* (Teleostei: Cobitidae) inferred from meiotic chromosome configurations [J]. Genetica, 2011, 139(6): 805-811.
- [14] WANG S, LV J, ZHANG L, et al. MethylRAD: a simple and scalable method for genome-wide DNA methylation profiling using methylation-dependent restriction enzymes [J]. Open Biol, 2015, 5(11): 150130. <http://dx.doi.org/10.1098/rsob.150130>.
- [15] WANG S, LIU P, LV J, et al. Serial sequencing of isolength RAD tags for cost-efficient genome-wide profiling of genetic and epigenetic variations [J]. Nat Protoc, 2016, 11: 2189-2200.
- [16] TAKUNO S, GAUT B S. Gene body methylation is conserved between plant orthologs and is of evolutionary consequence [J]. Proc Natl Acad Sci USA, 2013, 110(5): 1797-1802.
- [17] JJINGO D, CONLEY A B, YI S V, et al. On the presence and role of human gene-body DNA methylation [J]. Oncotarget, 2012, 3(4): 462-474.
- [18] SUZUKI M M, KERR A R W, SOUSA D D, et al. CpG methylation is targeted to transcription units in an invertebrate genome [J]. Genome Res, 2007, 17(5): 625-631.
- [19] WANG Y, JORDA M, JONES P L, et al. Functional CpG methylation system in a social insect [J]. Science, 2006, 314(5799): 645-647.
- [20] BONASIO R, LI Q, LIAN J, et al. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator* [J]. Curr Biol, 2012, 22(19): 1755-1764.
- [21] LYKO F, FORET S, KUCHARSKI R, et al. The honey bee epigenomes: differential methylation of brain DNA in queens and workers [J]. PLoS Biol, 2010, 8(11): e1000506.
- [22] RAJASETHUPATHY P, ANTONOV I, SHERIDAN R, et al. A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity [J]. Cell, 2012, 149(3): 693-707.
- [23] ZHANG X, YAZAKI J, SUNDARESAN A, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis* [J]. Cell, 2006, 126(6): 1189-1201.
- [24] SCHÜBELER D. Epigenetic islands in a genetic ocean [J]. Science, 2012, 338(6108): 756-757.
- [25] ZILBERMAN D, GEHRING M, TRAN R K, et al. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription [J]. Nat Genet, 2007, 39(1): 61-66.
- [26] CUI H X, LIU R R, ZHAO G P, et al. Identification of differentially expressed genes and pathways for intramuscular fat deposition in pectoralis major tissues of fast-and slow-growing chickens [J]. BMC Genomics, 2012, 13(1): 213.
- [27] FENG B, YI S V, LI R, et al. Comparison of age and growth performance of diploid and tetraploid loach *Misgurnus anguillicaudatus* in the Yangtze River Basin, China [J]. Environ Biol Fish, 2017, 100(Suppl1): 1-14.
- [28] MORTUSEWICZ O, SCHERMELLEH L, WALTER J, et al. Recruitment of DNA methyltransferase I to DNA repair sites [J]. Proc Natl Acad Sci USA, 2005, 102(25): 8905-8909.
- [29] DIBRA H K. Determination of an interaction between the DNA repair proteins MLH1 and SMD4 and aspirin regulation of DNA repair gene and protein expression in colorectal cancer [D]. Wolverhampton: University of Wolverhampton, 2010.

## Comparative studies on DNA methylation variation between diploid and tetraploid *Misgurnus anguillicaudatus*

ZHANG Manman FENG Bing LUO Shuangshuang WANG Weimin ZHOU Xiaoyun

*College of Fisheries, Huazhong Agricultural University/Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education, Wuhan 430070, China*

**Abstract** Using MethylRAD-Seq method, we analyzed DNA methylation distribution in the genome of *Misgurnus anguillicaudatus* and compared DNA methylation variation between the diploids and tetraploids. In total, 302 111 684 MethylRAD sequence tags were obtained and mapped to the reference genome. The results showed that, the DNA methylated sites distributed mainly in gene body, followed by intron and intergenic, but fewer methylated sites were found in other functional elements. Total DNA methylation level in the tetraploids was higher than that in the diploids, especially significant differences ( $P < 0.01$ ) of DNA methylation level existed in 1<sup>st</sup> Exon and upstream 200–1 500 bp region of the transcription start site (TSS 1500). However, in the promoter region, DNA methylation level in the tetraploids was lower than that in the diploids. For DNA methylation level of each sites, 1 268 C<sup>m</sup>CGG and 14 C<sup>m</sup>CWGG sites showed significant difference between the diploids and tetraploids, which mainly distributed in intron, gene body and intergenic. A total of 684 genes exhibited significant methylation variations between ploidy. KEGG Pathway enrichment analysis showed that, the genes with DNA methylation variation between ploidy enriched mainly in the growth development, immunity and mismatch repair related signaling pathways.

**Keywords** *Misgurnus anguillicaudatus*; diploid; tetraploid; DNA methylation; MethylRAD-Seq

(责任编辑:边书京)