



青鱼survey基因组测序与发育相关基因复制的研究



夏慧敏 罗立鸣 陆颖*

上海海洋大学 水产种质资源发掘与利用教育部重点实验室

引言 Introduction

青鱼 (black carp; *Mylopharyngodon piceus*) 属于鲤科，起源于东亚地区，自然分布于从珠江向北延伸到阿穆尔河，是中国传统的“四大家鱼”之一。目前对青鱼的研究主要集中在人工养殖、营养与饲料、疾病防治等领域，而基因组基础数据很少。

在本研究中，报告了青鱼的survey基因组及全部基因注释结果，并对与生长发育有关基因和基因家族的进化过程进行了分析。

方法 Methods

1. 样品采集和基因组DNA测序

成年雄性青鱼尾鳍样品取自中国江苏省苏州市吴江区四大家鱼养殖场。采用 Illumina HiSeq × Ten 测序平台进行全基因组随机测序。

2. 基因组组装和功能预测

使用 SOAP-denovo 进行基因组的组装，利用 RepeatMasker 检测基因组重复序列，用 BUSCO 对组装的基因组的完整性进行评估。以 GenBank 斑马鱼蛋白序列作为同源参考，使用 Augustus 对 *De novo* 测序结果进行基因预测，建立青鱼基因模型。

3. 同源性分配和系统发育分析

通过 OrthoMCL 对 11 种脊椎动物的蛋白质基因家族进行聚类，用 PhyML 创建系统发育树。根据系统发育树使用 CAFE 鉴定基因家族的扩张和收缩，扩张家族成员用 KEGG 通路数据库进行注释。

结果 Results

1. 基因组组装与注释 (GenBank 登录号 PRJNA720528)

Total length (bp)	956,826,496
N50 length of initial contigs (bp)	651
N50 length of scaffolds (bp)	9,583
Quantity of scaffolds (>1K) (bp)	103,140
Largest scaffold (bp)	1,194,444
Largest contig (bp)	106,276
GC content	37.5%
Content of repetitive sequences	34.8%
Heterozygous rate	1.5×10^{-3}
Completeness of the genome estimated by BUSCO test	90.8%
Quantity of predicted protein-coding genes	27,541
Average length of the encoding sequences (bp)	1,132

2. 系统发育的构建

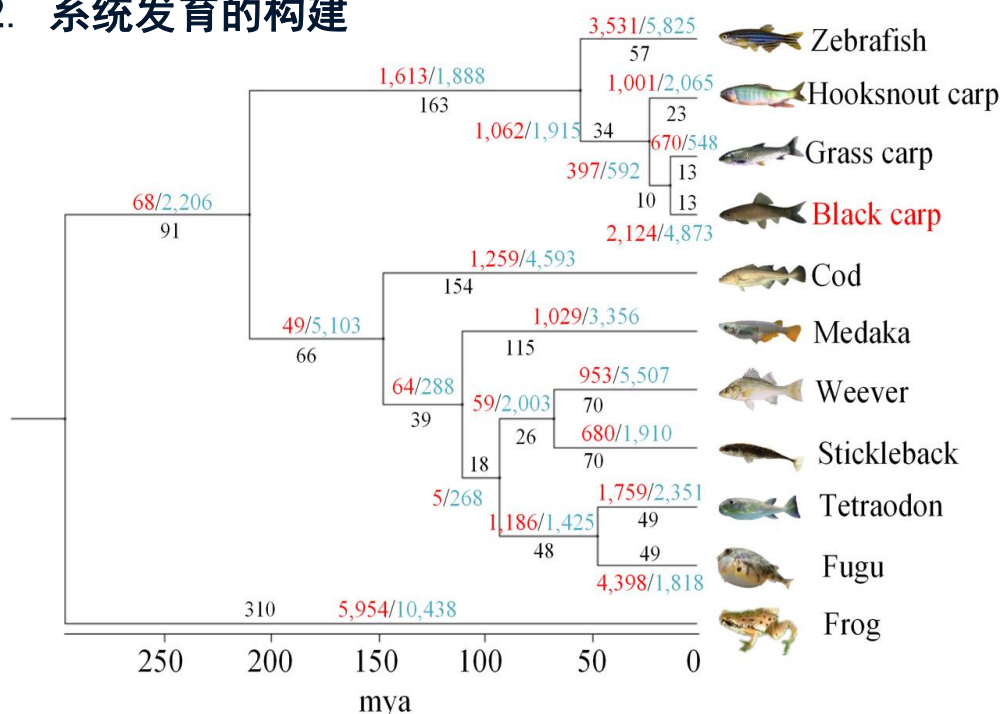


图1. 11种脊椎动物基因组的系统发育树

利用563个单拷贝基因构建系统发育树，结果显示青鱼与草鱼亲缘关系最近；计算估计青鱼与草鱼的分化时间为1300万年前，与斑马鱼的分化发生在约5700万年前。

3. 快速扩张基因家族的功能预测

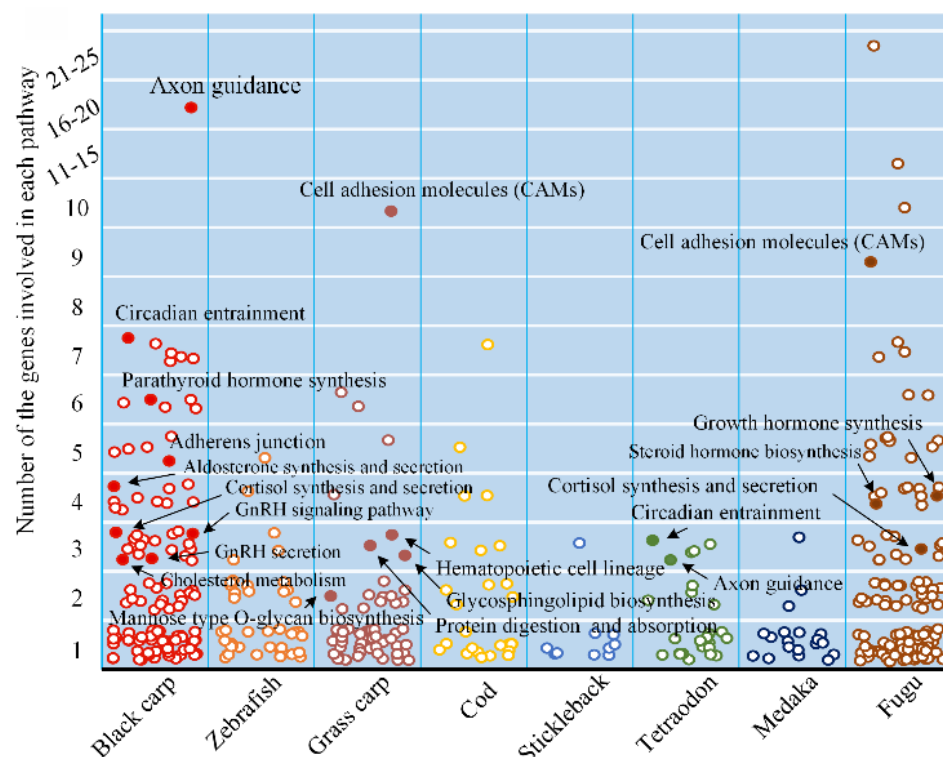


图2. 快速扩张基因家族的富集通路

比较8个物种，快速扩张基因家族参与的代谢通路，发现，在青鱼、草鱼等大型鱼类中有更多的与发育相关通路(骨骼生长、激素调节、节律控制和细胞分化等)的基因发生了明显复制。在河豚中，虽然也有较多的、与体型发育相关的基因家族发生扩张，但这些基因参与的多是抑制生长的代谢途径(生长激素抑制素等)。

结论 Conclusion

青鱼和草鱼等鱼类的体型增加可能是由于在进化过程中，大量相关功能基因的复制造成的。