



三种鲟鱼全长转录组测序及分析

刘洁 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

鲟鱼(Sturgeon)现存27个种,有约2亿年的化石记录,是最古老的鱼类之一。现有鲟鱼的染色体都是多倍体,经历了1-3次不等的全基因组复制。它们起源于已经灭绝的二倍体祖先。缓慢的进化速率使鲟鱼成为研究脊椎动物进化的重要模型。然而,由于鲟鱼基因组的结构复杂,且存在大量微染色体,因而基因组研究还相对滞后,缺少可靠的基础数据。

本研究利用第三代PacBio测序技术,对达氏鳇(*Huso dauricus*)、史氏鲟(*Acipenser schrenckii*)和闪光鲟(*Acipenser stellatus*)的血液mRNA进行全长度转录组测序,共获得了超过33000条可靠的转录本数据。通过分析,发现了已发表鲟鱼基因组中存在的基因注释的明显错误,推测了鲟鱼全基因组复制(WGD)的发生时间。

方法 Methods

1. 鲟鱼血液RNA提取、PacBio测序;
2. 序列组装,去冗余,获得转录组序列(CD-hit-est);
3. 基因功能注释和系统进化分析;
4. 通过鉴定SNP和计算基因非同义替换率推测WGD发生的时间。

结果 Results

1. 转录组组装与注释

	达氏鳇	史氏鲟	闪光鲟
测序 read 总长(Gb)	16.3	13.5	13.9
组装后的转录本数量	12696	12622	8089
具有完整 ORF 的转录本数量	6560	7056	3997
CDS 平均长度 (bp)	887	916	889
5'-UTR 平均长度 (bp)	428	326	395
3'-UTR 平均长度 (bp)	1009	982	981
与数据库比对上的转录本数量	9813	10684	6143
Interpro	9799	10671	6133
KEGG	3804	5645	2882
Swissprot	7393	8757	4572
预测的转录因子数量	529	449	358
检测到的 LncRNA 数量	1566	1182	1210
GenBank 登录号	PRJNA 692330	PRJNA 692324	PRJNA 692082

2. 全长转录本与已发表的小体鲟(*A. ruthenus*)基因组注释数据的比较

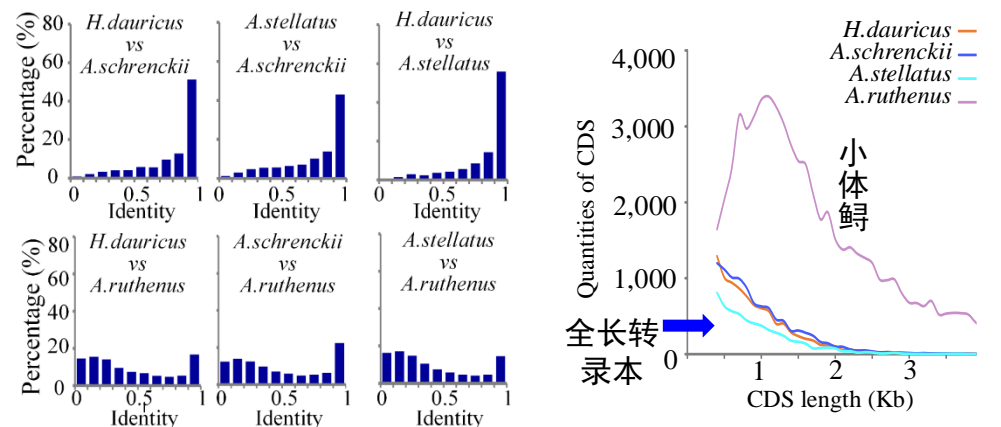


图1. 基因蛋白序列相似性比较

图2. CDS长度比较

3种全长转录本的数据分析显示,鲟鱼基因的CDS在序列和长度上均很保守。小体鲟的基因组注释结果与之相比,不仅相似度低,而且长度大大超过了全长转录本的长度分布。说明,对于鲟鱼这样基因组结构复杂的物种来说,缺少全长转录本数据支持的基因组注释结果误差很大,并可能导致后续进化分析的大量误差。

3. 主要鲟鱼物种系统发育树的构建

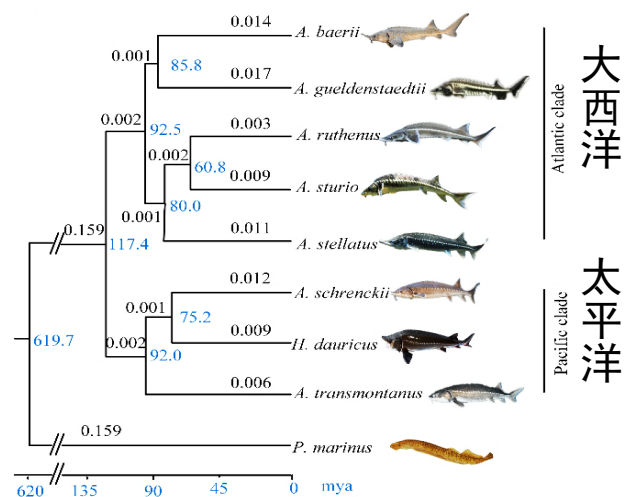


图3. 八种鲟鱼的系统发育树

系统进化树将8种鲟鱼分为大西洋分支和太平洋分支,与它们的地理分布一致。鲟和鳇的亲缘相近,很可能是单一起源。

3. 采用两种方法计算的 WGD 发生时间

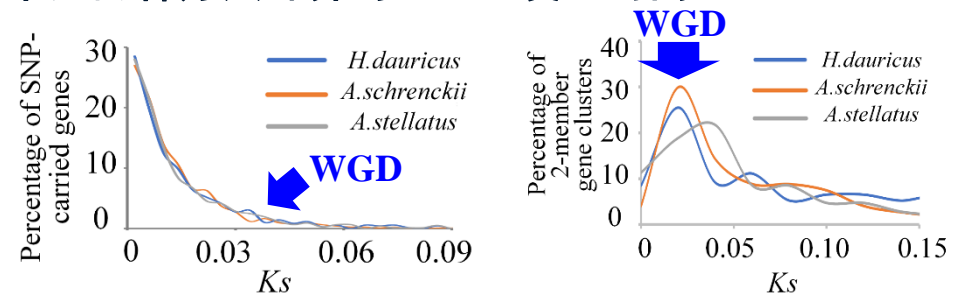


图4. 基因中非同义SNP累积的时间曲线

图5. 基因复制的时间曲线

双拷贝基因的分化时间和基因中SNP累积时间的两种计算结果基本吻合,确定了鲟鱼WGD发生于约1-1.5亿年前。

结论 Conclusion

- (1) 鳇和鲟有共同起源;
- (2) 通过计算基因中SNP的累积可以推测WGD时间;
- (3) 全长转录组序列是准确的基因组基础数据。

原文链接



扫一扫