

文章编号: 1674-5566(2017)05-0666-08

DOI:10.12024/jsou.20170301971

## 两种拼接方法在瓯江彩鲤转录组研究中的适用性比较

陈晓雯, 王 军, 岳武成, 杜金星, 王成辉

(上海海洋大学 农业部淡水水产种质资源重点开放实验室, 上海 201306)

**摘 要:** 以瓯江彩鲤 (*Cyprinus carpio* var. *color*) 为研究对象, 探讨了目前两种常用的转录组拼接方法在鲤的不同品种/品系转录组研究上的适用性。结果表明, 在转录组拼接和注释方面, 基于序列比对优先策略 (Cufflinks 软件) 拼接的转录组序列平均长度为 1 545 bp, 注释 77 601 个转录本; 基于从头拼接策略 (Trinity 软件) 拼接的转录组序列平均长度为 979 bp, 注释转录本数为 69 406 个。在基因结构和选择性剪切预测方面, Cufflinks 软件所拼接的转录组, 每个基因平均含有 3 个转录本数目, 而对于 Trinity 版本转录组, 每个基因平均含有 4 个转录本 ( $P < 0.001$ )。经比较发现 Cufflinks 版本转录组预测的基因结构较 Trinity 版本更加准确。因此, 与从头拼接策略相比, 序列比对优先策略所拼接的瓯江彩鲤转录本序列长度更长、注释的基因数目更多, 且在基因结构和选择性剪切预测方面更为准确, 适用于后续与功能基因表达相关的功能研究; 相反, 从头拼接策略拼接的转录组更易得到瓯江彩鲤的特异基因序列, 对于后续需要利用瓯江彩鲤特异基因序列进行分子进化相关的研究非常重要。本文的研究结果为根据不同的实验目的合理地选用相应的转录组拼接方法提供了依据, 也为其他鱼类转录组学研究提供了借鉴。

**关键词:** 转录组; 瓯江彩鲤; 拼接方法; 序列比对优先; 从头拼接

**中图分类号:** S 917      **文献标志码:** A

二代高通量测序技术 (Next-generation sequencing) 的产生, 加速了后基因组学的发展, 基因组学、转录组学、蛋白质组学和代谢组学等“组学”研究在这 10 多年间快速发展, 在生物学的各个研究领域都取得了巨大成果<sup>[1]</sup>。其中, 随着高通量测序成本的逐年降低和可视化生物信息学软件的大量开发, 转录组测序已成为进行相关物种功能基因组学研究的最佳选择<sup>[2-11]</sup>。当前, 转录组测序已成为商业化的产品, 各生物测序公司均采取模式化的方式进行测序, 相对而言, 参考转录组的拼接具备一定的灵活性, 需各实验室独立进行。目前转录组的拼接方法主要包含两种策略, 第一种为序列比对优先策略, 相应开发的软件有 Scripture, Cufflinks 等<sup>[12-13]</sup>, 主要是基于有完整参考基因组的物种, 将测序的原始数据比对到参考基因组上, 然后根据比对的重叠区域拼接序列, 获得参考转录组。另一种策略为从头拼接策略, 主要的相关软件有 ABySS<sup>[14]</sup>、

SOAPdenovo<sup>[15]</sup>、Velvet<sup>[16]</sup>、Trinity<sup>[17]</sup> 等, 主要是根据 de Bruijn 图的算法, 利用测序获得的序列进行直接组装, 无需参考基因组序列。

鲤 (*Cyprinus carpio*) 是全世界分布最为广泛的鱼类之一, 也是世界重要的水产养殖对象。鲤在长期的自然选择和人工选择过程中, 形成了许多不同的品种/品系, 不同的鲤鱼品种/品系产生了明显的表型和遗传差异<sup>[18-21]</sup>。如作为食用鱼和观赏鱼在全世界范围内广泛养殖的镜鲤、黄河鲤、建鲤、松浦鲤、兴国红鲤、荷包红鲤、玻璃红鲤和瓯江彩鲤<sup>[18, 22-27]</sup>。鲤的基因组序列已经发表, 相关的基因组序列可以在生物数据库中下载, 为鲤的功能基因组学研究、进化研究奠定了基础, 为不同品种/品系鲤鱼的“组学”研究提供了参考基因组序列。近年来, 在很多不同品种/品系的鲤中广泛开展的转录组研究主要采用的是从头拼接的策略<sup>[8, 10, 28]</sup>。由于不同鲤品种/品系间显著的表型和遗传差异, 基于松浦鲤的鲤参考基因

收稿日期: 2017-03-07      修回日期: 2017-04-30

基金项目: 国家自然科学基金(31372607); 上海市科委项目(15391900800)

作者简介: 陈晓雯(1987—), 女, 博士研究生, 研究方向为水产动物种质资源与种苗工程。E-mail: 396169701@qq.com

通信作者: 王成辉, E-mail: wangch@shou.edu.cn

组序列是否适合于每个鲤品种/品系的转录组拼接尚不得而知。

瓯江彩鲤(*Cyprinus carpio* var. *color*)是浙江省西南部龙泉、青田等县市广泛养殖的地方性鲤群体,因具有全红、大花、麻花、粉玉、粉花等多种体色,色彩绚丽鲜艳,是一种兼具食用和观赏价值的经济鱼类<sup>[29]</sup>。由于长期的自然和人工选择,瓯江彩鲤已产生了明显的适应性进化,在表型和基因型上均与其它鲤群体存在较大差异<sup>[30]</sup>。本实验室在瓯江彩鲤的人工选育和体色进化研究过程中,已积累了大量的转录组数据,为序列的拼接和组装提供了多样化素材。本文以瓯江彩鲤为研究对象,研究两种转录组拼接方法的适用性,同时为其它鱼类的转录组研究提供参考和依据。

## 1 材料与方法

### 1.1 实验材料

本文进行瓯江彩鲤转录组拼接所用到的原始测序数据来自于本实验室前期进行的“全红”(WR)和“粉玉”(WW)两种体色瓯江彩鲤的皮肤转录组测序研究,基于 Illumina HiSeq2500 测序平台,进行  $2 \times 150$  bp 的双末端测序,获得“全红” $9.7 \times 10^9$  bp 和“粉玉” $1.27 \times 10^{10}$  bp 的测序数据(数据未显示),为本文进行转录组拼接的原始数据。

### 1.2 转录组拼接

原始测序数据利用 Trimmomatic 过滤后,获得无污染的、高质量的测序数据,然后将过滤后的“全红”和“粉玉”测序数据合并,用于瓯江彩鲤转录组拼接的输入数据<sup>[31]</sup>。本文采用了两种常规的拼接策略进行瓯江彩鲤的转录组拼接。第一种方法采用了 Tophat-Cufflinks 流程(比对优先策略),基于序列比对的原理进行拼接,利用已发表的鲤全基因组序列为参考序列。其步骤是首先利用 Tophat 软件将过滤后的“全红”和“粉玉”的测序序列分别比对到鲤鱼的基因组上,采用 Tophat 软件默认的最佳参数设置,获得“全红”和“粉玉”转录组测序序列在基因组上的比对信息,储存于 BAM 文件中<sup>[13]</sup>;其次将生成的包含比对

信息的 BAM 文件导入 Cufflinks 软件,基于比对信息中的基因组位置信息进行拼接,分别获得“全红”和“粉玉”的转录组序列<sup>[13]</sup>;最后,利用 Cuffmerge 软件将上一步完成的“全红”和“粉玉”的转录组序列进行合并,获得完整的瓯江彩鲤的参考转录组序列<sup>[13]</sup>。第二种方法采用了基于 *de Bruijn* 图算法的 *de novo* 拼接(从头拼接策略),此方法不需要用到鲤鱼的基因组序列,完全依赖于测序获得的瓯江彩鲤的序列信息。其步骤是首先将“全红”和“粉玉”过滤后的序列全部合并,然后利用 Trinity 软件基于双末端的模式进行拼接,设置的参数如下:16 线程的 CPU,120 Gb 的内存,过滤掉拼接长度小于 300 bp 的转录本序列<sup>[17]</sup>。

### 1.3 转录组的功能注释

采用两种方法拼接获得的转录组后,利用 NCBI Blastx 软件进行拼接转录本的功能注释,将拼接所获得的转录本与蛋白序列的公共数据库 NCBI-NR(非冗余蛋白数据库),Uniprot 蛋白数据(<http://www.uniprot.org/downloads>),斑马鱼蛋白数据([ftp://ftp.ensembl.org/pub/release-87/fasta/danio\\_rerio/pep/](ftp://ftp.ensembl.org/pub/release-87/fasta/danio_rerio/pep/))以及鲤蛋白数据([http://www.carphbase.org/download\\_home.php](http://www.carphbase.org/download_home.php))<sup>[32]</sup>作比对,E 值的阈值设置为  $< 1e - 5$ 。利用 BLAST2GO 软件进行 Gene Ontology 和 KEGG 代谢通路的注释。利用 TransDecoder 软件进行蛋白编码序列(CDS)的预测。

## 2 结果

### 2.1 转录组拼接比较

基于 Tophat-Cufflinks 软件和 Trinity 软件拼接的两个瓯江彩鲤转录组分别获得了 98 906 和 153 291 条转录本序列(表 1)。利用 Tophat 软件将近 60% 的“全红”和“粉玉”的转录组序列比对到鲤鱼的基因组上(表 2,图 1)。利用 Trinity 软件拼接获得的转录本的平均长度和 N50 长度较利用 Cufflinks 软件拼接过的转录本长度短(表 1,图 2a)。虽然两种方法拼接的转录本数目不同,但拼接的转录组大小是基本一致的(约  $1.5 \times 10^8$  bp,表 1)。

表1 两种转录组拼接方法的比较  
Tab.1 Comparison of two transcriptome assemblies (Trinity and Cufflinks)

	Trinity	Cufflinks
组装 Assembly		
转录本数目 No. of transcripts	153 291	98 906
N50 转录本数目 No. of N50 transcripts	26 107	23 611
N50 长度 /bp N50 length	1 788	2 114
转录本平均长度/bp Mean transcripts length	979	1 545
最大转录本/bp Largest transcripts	18 397	13 393
预测 CDS 数目 No. of predicted CDS	62 006	80 632
总组装转录本大小/bp Total assembled transcriptome size	$1.5 \times 10^8$	$1.52 \times 10^8$
注释 Annotation		
NCBI-Nr	69 406(45.3%)	77 601(78.5%)
UniProt	62 372(40.6%)	70 728(71.5%)
斑马鱼 Zebra fish	69 216(45.1%)	75 045(75.8%)
鲤 Common carp	76 989(50.2%)	77 416(78.2%)
GO	32 581	36 476
KEGG	26 822	29 854

## 2.2 转录组功能注释比较

利用不同的蛋白数据库对拼接完成的两个转录组进行注释,注释结果表明,在 Trinity 版本的转录组中能注释的转录本数目在 62 372 到 76 989 之间。而 Cufflinks 版本的转录组中能注释的转录本数目在 70 728 到 77 601 之间。以鲤的蛋白数据库为参照,Cufflinks 和 Trinity 版本的转录组分别注释了 77 416 和 76 989 个转录本,分别对应了 27 169 和 22 450 个鲤的蛋白编码基因(表1)。其中 18 886 个基因在两个转录组中共有,Cufflinks 版本转录组中拼接获得了 8 283 个特异的基因序列,而 Trinity 版本转录组中拼接获得了 3 564 个特异的基因序列。利用 GO 数据

库,Cufflinks 和 Trinity 版本转录组中分别注释了 36 476 和 32 581 个转录本序列。基于 KEGG 蛋白数据库,Cufflinks 和 Trinity 版本转录组中 29 854 和 26 822 条转录本序列分别被注释。整体而言,Cufflinks 版本中能注释的转录本数目大于 Trinity 版本中能注释的转录本数目(表1,图2b)。

## 2.3 基因结构预测和比较

基于 18 886 个在两个版本转录组中共同注释的功能基因,统计两个版本转录组中每个基因对应的平均选择性剪切的转录本数目。对于 Trinity 转录组,每个基因平均含有 4 个转录本数目,而对于 Cufflinks 版本转录组,每个基因平均含有 3 个转录本( $P < 0.001$ , t-test)。尽管 Trinity 版本的转录组预测了更多的选择性剪切的转录本,但 Cufflinks 版本转录组预测的基因结构较 Trinity 版本更加准确。例如,*slc9a3r1* (ENSDARG0000000068) 基因在两个版本的转录组中均预测获得了 3 条转录本序列,在 Trinity 版本转录组中 3 条 *slc9a3r1* 转录本序列为 TR183826|c7\_g1\_i1、TR183826|c7\_g1\_i2 和 TR183826|c7\_g1\_i3,基因结构预测表明该 3 条序列为 *slc9a3r1* 基因的 3 条选择性剪切的转录本,属于同一个基因。而在 Cufflinks 版本转录组中 3 条 *slc9a3r1* 转录本序列为 TCONS\_00004298、TCONS\_00004299 和 TCONS\_00004867,基因结构预测的结构表明 3 条序列来自于两条旁系同源的 *slc9a3r1* 基因,TCONS\_00004298 和 TCONS\_00004299 是鲤鱼基因组 scaffold 000000268 上 *slc9a3r1* 基因的 2 条选择性剪切的转录本。而 TCONS\_00004867 与 TR183826|c7\_g1\_i3 序列的相似度很高(>95%),来自于鲤鱼基因组 scaffold 000000306 上的另一个 *slc9a3r1* 旁系同源基因(图3)。

表2 全红(WR)和粉玉(WW)两种体色瓯江彩鲤 RNA 测序数据 Tophat 软件比对信息  
Tab.2 Mapping statistics of RNA-Seq reads from WR and WW color common carp by Tophat software

样本 Sample	总序列数目 Number of sequences	总比对数目(百分比) Total number of mapped reads	多定位的测序序列数目(百分比) Number of multiple mapped reads	单定位的测序序列数目(百分比) Number of single mapped reads
WW-1	33 158 078	21 264 821(64.1%)	4 397 377(13.3%)	16 867 444(50.9%)
WW-2	21 643 228	13 095 967(60.6%)	2 614 112(12.1%)	10 481 855(48.4%)
WW-3	30 239 430	19 355 825(64.0%)	4 080 646(13.5%)	15 275 179(50.5%)
WR-1	17 958 938	11 776 608(65.5%)	2 326 098(13.0%)	9 450 510(52.6%)
WR-2	19 343 932	12 315 237(63.7%)	2 504 592(12.9%)	9 810 645(50.7%)
WR-3	27 285 755	16 882 644(61.9%)	3 495 254(12.8%)	13 387 390(49.1%)

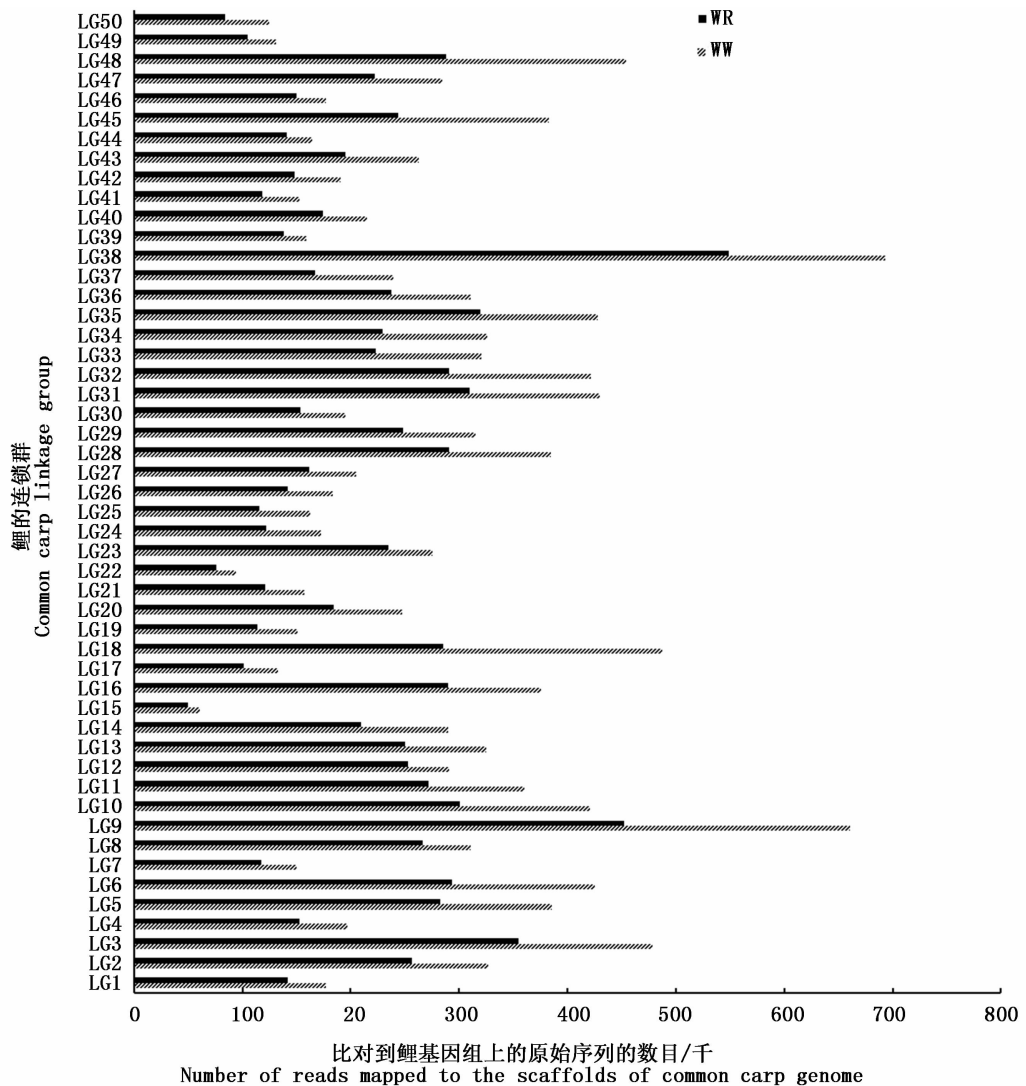


图 1 瓯江彩鲤转录组数据在鲤基因组上的分布情况

Fig. 1 Reads distribution on the scaffold (LG) of common carp genome

### 3 讨论

基于序列比对优先策略的 Cufflinks 软件和基于从头拼接策略的 Trinity 软件都是目前转录组拼接常用的两个软件<sup>[13, 17]</sup>。基于序列比对优先策略开发的 Cufflinks 软件承诺序列比对时最大程度的灵敏性,但是需要利用物种特异的或者近缘的基因组序列<sup>[13]</sup>。通常情况下,对于模式动植物或具有完整基因组序列的物种,转录组拼接往往采用的是 Tophat-Cufflinks 的拼接流程。然而在本实验中,“全红”和“粉玉”的转录组序列中仅有约 60% 左右的序列比对到基因组上(表 2),低于 Tophat 软件要求的 70% 以上,一般情况下,

若序列比对的比例低于 70%,往往会造成一定程度的转录组组装的错误<sup>[13]</sup>。对本实验而言,低比例的序列比对率可能由以下原因造成:(1)目前拼接的鲤鱼的基因组序列仍不完善,拼接的基因组框架序列存在一定的缺口<sup>[32]</sup>;(2)由于瓯江彩鲤是我国主要分布于浙江瓯江流域的特色鲤鱼品系,长期的自然选择和适应性进化造成了其与普通鲤鱼产生了一定的遗传变异,进而造成低比例的序列比对率<sup>[30, 33-34]</sup>;(3)鲤鱼一直以来被认为是古老的异源四倍体鱼类,基因组复制,多拷贝基因的现象也会造成序列比对率下降,进而影响转录组的拼接<sup>[32, 35]</sup>。而基于从头拼接策略的 Trinity 软件在转录组拼接过程中不需要依靠鲤鱼

的基因组序列,完全依赖于瓯江彩鲤自身的测序序列,拼接的转录组序列将是瓯江彩鲤特异的基因序列,而且可能会拼接出普通鲤鱼中不存在的新基因序列<sup>[17]</sup>。但基于 Trinity 软件拼接的转录

组序列由于没有基因组序列作为参考,往往会出现序列长度短,片段化严重,基因序列不完整现象,在多个物种中都有类似的报道<sup>[36]</sup>,本文的研究结果也能很好地证明这一点。

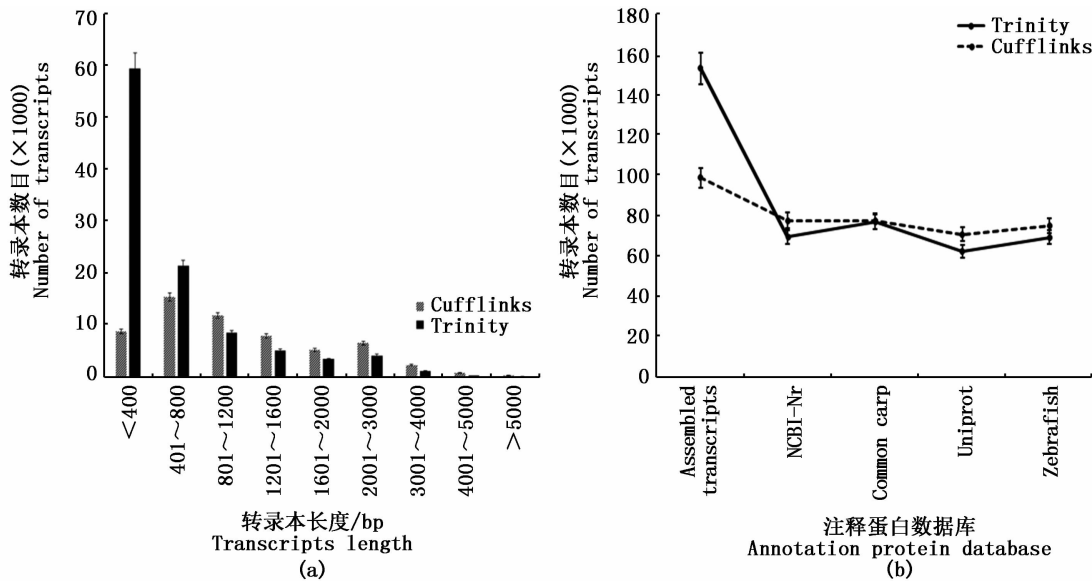


图2 两种转录组拼接方法所拼接转录本的长度 (a) 和注释信息 (b)

Fig. 2 The length of two transcriptome assemblies (Trinity and Cufflinks) (a) and their annotation (b)

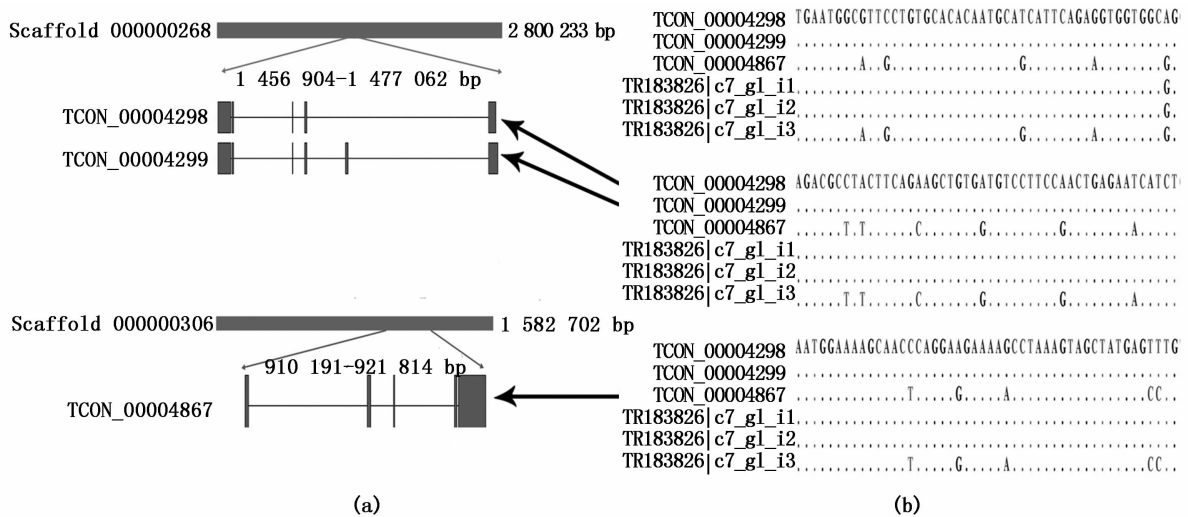


图3 *slc9a3r1* 基因结构预测 (a) 和两种拼接方法的序列比对 (b)

Fig. 3 Gene structure of *slc9a3r1* predicted by Cufflinks (a) and alignment of *slc9a3r1* sequences assembled by Trinity and Cufflinks (b)

由于鲤一直以来被认为是异源四倍体鱼类,基因复制的现象在鲤中非常常见,因而会造成大量的旁系同源基因的出现,这对于准确地预测基因的结构是个挑战<sup>[32]</sup>。本研究利用两种方法拼

接的转录组结果表明利用 Cufflinks 软件拼接的转录组能够更加准确地预测瓯江彩鲤基因结构和选择性剪切情况<sup>[37]</sup>。本研究中,利用 Cufflinks 拼接的转录组序列长度更长,注释地基因数目更

多,基因结构和选择性剪切预测得更加准确,更加适用于进行后续的瓯江彩鲤基因表达量和差异基因表达的研究。相反,基于 Trinity 软件拼接的转录组序列为瓯江彩鲤特异的转录本序列,对瓯江彩鲤的研究也具有重要的意义,对后续需要利用特异的瓯江彩鲤基因序列的分子进化相关研究是非常有必要的。

本研究以瓯江彩鲤为研究对象,探讨了两种常见的转录组拼接方法在瓯江彩鲤转录组学研究上的适用性。就目前而言,两种拼接方法均对瓯江彩鲤的转录组学研究具有重要意义,需要根据不同的研究目的,选用合理的拼接方法。同时本研究也为其他鱼类的转录组学研究提供了参考和方向。

#### 参考文献:

- [1] MARGULIES M, EGHOLM M, ALTMAN W E, et al. Genome sequencing in microfabricated high-density picolitre reactors [J]. *Nature*, 2005, 437(7057): 376-380.
- [2] WANG J T, LI J T, ZHANG X F, et al. Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (*Cyprinus carpio*) [J]. *BMC Genomics*, 2012, 13(1): 96.
- [3] BOUGAS B, NORMANDEAU E, AUDET C, et al. Linking transcriptomic and genomic variation to growth in brook charr hybrids (*Salvelinus fontinalis*, Mitchell) [J]. *Heredity*, 2013, 110(5): 492-500.
- [4] COROMINAS J, RAMAYO C Y, PUIG-OLIVERAS A, et al. Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition [J]. *BMC Genomics*, 2013, 14: 843.
- [5] JANSEN L, HOLLUNDER J, ROBERTS I, et al. Comparative transcriptomics as a tool for the identification of root branching genes in maize [J]. *Plant Biotechnology Journal*, 2013, 11(9): 1092-1102.
- [6] JIMENEZ-GURI E, HUERTA-CEPAS J, COZZUTO L, et al. Comparative transcriptomics of early dipteran development [J]. *BMC Genomics*, 2013, 14: 123.
- [7] LIAO X L, CHENG L, XU P, et al. Transcriptome analysis of Crucian Carp (*Carassius auratus*), an important aquaculture and hypoxia-tolerant species [J]. *PLoS One*, 2013, 8(4): e62308.
- [8] WANG C H, WACHHOLTZ M, WANG J, et al. Analysis of the skin transcriptome in two oujiang color varieties of common carp [J]. *PLoS One*, 2014, 9(3): e90074.
- [9] HUANG S, WANG J, YUE W C, et al. Transcriptomic variation of hepatopancreas reveals the energy metabolism and biological processes associated with molting in Chinese mitten crab, *Eriocheir sinensis* [J]. *Scientific Reports*, 2015, 5: 14015.
- [10] LI G X, ZHAO Y L, LIU Z H, et al. De novo assembly and characterization of the spleen transcriptome of common carp (*Cyprinus carpio*) using Illumina paired-end sequencing [J]. *Fish Shellfish Immunol*, 2015, 44(2): 420-429.
- [11] WANG J, LAMER J T, GAUGHAN S, et al. Transcriptomic comparison of invasive bigheaded carps (*Hypophthalmichthys nobilis* and *Hypophthalmichthys molitrix*) and their hybrids [J]. *Ecology and Evolution*, 2016, 6(23): 8452-8459.
- [12] GUTTMAN M, GARBER M, LEVIN J Z, et al. *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs [J]. *Nature Biotechnology*, 2010, 28(5): 503-510.
- [13] TRAPNELL C, ROBERTS A, GOFF L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks [J]. *Nature Protocols*, 2012, 7(3): 562-578.
- [14] BIROL I, JACKMAN S D, NIELSEN C B, et al. *De novo* transcriptome assembly with ABySS [J]. *Bioinformatics*, 2009, 25(21): 2872-2877.
- [15] LUO R B, LIU B H, XIE Y L, et al. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler [J]. *Gigascience*, 2012, 1: 18.
- [16] ZERBINO D R, BIRNEY E. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs [J]. *Genome Research*, 2008, 18(5): 821-829.
- [17] GRABHERR M G, HAAS B J, YASSOUR M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. *Nature Biotechnology*, 2011, 29(7): 644-652.
- [18] 常玉梅, 孙效文, 梁利群. 中国鲤几个代表种群基因组 DNA 遗传多样性分析 [J]. *水产学报*, 2004, 28(5): 481-486.
- CHANG Y M, SUN X W, LIANG L Q. Genetic diversity analysis of genomic DNAs of several representative populations of common carp in China [J]. *Journal of Fisheries of China*, 2004, 28(5): 481-486.
- [19] 闵霞. 几种鲤鱼遗传多样性分析和松浦镜鲤分子辅助育种标记筛选 [D]. 上海: 上海海洋大学, 2011.
- MIN X. The genetic diversity of several common carp (*Cyprinus carpio* L.) and screening molecular marker assisting breeding for Songpu mirror carp [D]. Shanghai: Shanghai Ocean University, 2011.
- [20] WANG C H, LI S F, NAGY Z T, et al. Molecular genetic structure and relationship of Chinese and Hungarian common carp (*Cyprinus carpio* L.) strains based on mitochondrial sequence [J]. *Aquaculture Research*, 2010, 41(9): 1339-1347.
- [21] WANG C H, LIU H, LIU Z Z, et al. Mitochondrial genetic diversity and gene flow of common carp from main river

- drainages in China[J]. *Freshwater Biology*, 2010, 55(9): 1905-1915.
- [22] 闻秀荣. 优良鲤鱼品种——德国镜鲤[J]. *新农业*, 1990(8): 44.  
WEN X R. Good carp variety-Germany mirror carp[J]. *New Agriculture*, 1990(8): 44.
- [23] 徐桂珍. 鲤鱼优良新品种——建鲤[J]. *农林科学实验*, 1994(4): 26-27.  
XU G Z. Good new variety of carp-Jian carp[J]. *Experiment of agriculture and Forestry Science*, 1994(4): 26-27.
- [24] 石连玉. 鲤鱼新品种——松浦镜鲤[J]. *农家参谋*, 2011(9): 7.  
SHI L Y. New variety of carp-Songpu carp[J]. *The Farmers Consultant*, 2011(9): 7.
- [25] 惠同胜. 几种适合北方高寒地区养殖的鲤鱼品种[J]. *黑龙江水产*, 2016(1): 16-18.  
HUI T S. Several kinds of carp species suitable for breeding in northern Alpine Area[J]. *Heilongjiang Fisheries*, 2016(1): 16-18.
- [26] WANG C H, WANG J, XIANG S P, et al. Parental genetic effects evaluation of growth-related traits of red common carp in China[J]. *Fisheries Science*, 2009, 75(5): 1301-1305.
- [27] 王成辉. 中国红鲤遗传多样性研究[D]. 上海: 上海水产大学, 2002.  
WANG C H. Study on genetic diversity of red common carps in China [D]. Shanghai: Shanghai Fisheries University, 2011.
- [28] JIANG Y L, ZHANG S H, XU J, et al. Comparative transcriptome analysis reveals the genetic basis of skin color variation in common carp [J]. *PLoS One*, 2014, 9(9): e108200.
- [29] 朱丽艳, 马玉清, 项松平, 等. 不同体色瓯江彩鲤生长动态的观察与分析[J]. *上海海洋大学学报*, 2013, 22(3): 341-348.  
ZHU L Y, MA Y Q, XIANG S P, et al. Observation and analysis of growth dynamics in Oujiang color common carp with different pigmentation types [J]. *Journal of Shanghai Ocean University*, 2013, 22(3): 341-348.
- [30] WANG C H, LI S F. Phylogenetic relationships of ornamental (koi) carp, Oujiang color carp and Long-fin carp revealed by mitochondrial DNA COII gene sequences and RAPD analysis [J]. *Aquaculture*, 2004, 231(1/4): 83-91.
- [31] BOLGER A M, LOHSE M, USADEL B. Trimmomatic: a flexible trimmer for Illumina Sequence Data [J]. *Bioinformatics*, 2014, 30(15): 2114-2120.
- [32] XU P, ZHANG X F, WANG X M, et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio* [J]. *Nature Genetics*, 2014, 46(11): 1212-1219.
- [33] WANG C H, LI S F, XIANG S P, et al. Genetic parameter estimates for growth-related traits in Oujiang color common carp (*Cyprinus carpio* var. *color*) [J]. *Aquaculture*, 2006, 259(1/4): 103-107.
- [34] WANG C H, LI S F, XIANG S P, et al. Additive, dominance genetic effects for growth-related traits in common carp, *Cyprinus carpio* L. [J]. *Aquaculture Research*, 2006, 37(14): 1481-1486.
- [35] DAVID L, ROTHBARD S, RUBINSTEIN I, et al. Aspects of red and black color inheritance in the Japanese ornamental (Koi) carp (*Cyprinus carpio* L.) [J]. *Aquaculture*, 2004, 233(1/4): 129-147.
- [36] LU B X, ZENG Z B, SHI T L. Comparative study of *de novo* assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq [J]. *Science China Life Sciences*, 2013, 56(2): 143-155.
- [37] 卢戎. 基于第二代测序的转录组组装软件比较研究[D]. 兰州: 兰州大学, 2013.  
LU X. Comparison of transcriptome assembly software for next-generation sequencing technologies [D]. Lanzhou: Lanzhou University, 2013.

## Comparison of two transcriptome assembly strategies on Oujiang color common carp (*Cyprinus carpio* var. *color*)

CHEN Xiaowen, WANG Jun, YUE Wucheng, DU Jinxing, WANG Chenghui

(Key Laboratory of Freshwater Fishery Germplasm, Ministry of Agriculture, Shanghai Ocean University, Shanghai 201306, China)

**Abstract:** Two transcriptome assembly strategies (mapping-first and de novo assembly-first) were used in this study to evaluate the applicability on transcriptome study for common carp varieties, using Oujiang color common carp as studied species. Regarding the transcriptome assembly and annotation, our study showed the average length of assembled transcripts using mapping-first strategy (Cufflinks software) was 1 545 bp, the number of annotated transcripts was 77 601. The average length of transcript assembled using *de novo* assembly-first strategy (Trinity software) was 979 bp, the number of annotated transcripts was 69 406. Considering the gene structure and alternative splicing prediction, our results presented three and four alternative splicing transcripts for every gene for Cufflinks and Trinity assembly, respectively ( $P < 0.001$ ). Meanwhile, the gene structure prediction was more accurate in Cufflinks assembly than Trinity assembly. Our results indicated the longer transcripts, the more annotated transcripts, and transcripts with more accurate gene structure were identified in the transcriptome assembled using mapping-first strategy (Cufflinks software), which is more suitable for the following gene expression analysis. On the contrary, the transcriptome assembled using de novo assembly-first strategy (Trinity software) provided more accurate specific gene sequences, which is vital for evolutionary studies of Oujiang color common carp. Our study demonstrated that both assembly strategies were essential for the Oujiang color common carp transcriptome, and provided guidance for the choice of transcriptome assembly strategies on other varieties of common carp.

**Key words:** transcriptome; Oujiang color common carp; assembly; mapping-first; *de novo* assembly