

DOI: 10.3724/SP.J.1005.2013.00955

非模式生物转录组研究

刘红亮, 郑丽明, 刘青青, 权富生, 张涌

西北农林科技大学动物医学院, 农业部动物生物技术重点实验室, 杨凌 712100

摘要: 转录组代表细胞或组织内全部的 RNA 转录本(RNA transcripts), 反映不同生命阶段、不同组织类型、不同生理状态以及不同环境条件下表达的基因。转录组研究可以从整体水平上反映细胞中基因表达情况及其调控规律。非模式生物(Non-model organism)具有许多模式生物不具备的特征, 其转录组研究对解决基因进化、遗传育种以及生态等诸多方面的问题具有重要意义。但由于非模式生物缺乏参考基因组信息, 传统的转录组研究方法操作复杂, 实验周期长, 花费大, 使得其转录组研究进展缓慢。新一代高通量 RNA 测序技术(RNA-seq)完全改变了转录组学的研究模式, 迅速成为研究非模式生物转录组的先进技术。文章详细论述了近年来利用 RNA-seq 技术进行非模式生物转录组研究的进展情况, 从样品准备、高通量 DNA 测序以及生物信息学分析等方面介绍了利用 RNA-seq 技术研究非模式生物转录组的一般流程及方法, 并对其中有待进一步研究的问题进行了展望。

关键词: 转录组; 非模式生物; RNA-seq

Studies on the transcriptomes of non-model organisms

LIU Hong-Liang, ZHENG Li-Ming, LIU Qing-Qing, QUAN Fu-Sheng, ZHANG Yong

Key Laboratory of Animal Biotechnology of the Ministry of Agriculture, College of Veterinary Medicine, Northwest A&F University, Yangling 712100, China

Abstract: The transcriptome represents the whole complement of RNA transcripts in cells or tissues and reflects the expressed genes at various life stages, tissue types, physiological states, and environmental conditions. Transcriptome analysis provides a comprehensive understanding of gene expression and its regulation. Non-model organism has many interesting traits of which model organisms lack, and the study of its transcriptome has great significance in solving the questions of genetic evolution, genetic breeding, ecology and so on. Because of absence of reference genome information, and traditional transcriptome research methods which are complicated to operate, long experimental period and costly, slow progress has been made in the research of non-model organism transcriptome. Fortunately, RNA sequencing (RNA-seq), the next-generation sequencing technology, has completely changed the way of transcriptome study, becoming an advanced technology on the investigation of non-model organism transcriptome. In this paper, we give a summary of the non-model

收稿日期: 2013-01-27; 修回日期: 2013-03-15

基金项目: 十二五国家科技计划课题(863 项目)(编号: 2011AA100303)资助

作者简介: 刘红亮, 硕士研究生, 专业方向: 动物胚胎工程。E-mail: liuhl355113@sina.com

通讯作者: 权富生, 博士, 副教授, 硕士生导师, 研究方向: 动物胚胎工程与发育生物学。E-mail: quanfusheng@nwsuaf.edu.cn

张涌, 博士, 教授, 博士生导师, 研究方向: 动物胚胎工程与发育生物学。E-mail: zhangyong1956@nwsuaf.edu.cn

网络出版时间: 2013-4-24 16:37:50

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20130424.1637.001.html>

organism transcriptome research using RNA-seq in recent years and briefly describe its general flow and principles from aspects of sample preparation, high throughput DNA sequencing, and bioinformatics analysis. Finally, questions that are still open and awaiting further research are also discussed.

Keywords: transcriptome; non-model organism; RNA-seq

广义的转录组代表细胞或组织内全部的 RNA 转录本, 包括编码蛋白质的 mRNA 和各种非编码 RNA(rRNA、tRNA、microRNA 等)。而狭义的转录组系指所有编码蛋白质的 mRNA 总和。转录组研究能够从整体水平研究基因功能以及基因结构, 揭示特定生物学过程以及疾病发生过程中的分子机理, 已广泛应用于基础研究、临床诊断和药物研发等领域。

随着一系列模式生物(Model organism)基因组测序的完成, 功能基因组学的研究方兴未艾。参照这些模式生物的参考序列(Reference sequence), 研究人员可以很方便的研究该物种的全基因组转录情况、不同个体之间的 SNP 差异、基因拷贝数差异等。而对于非模式生物而言, 情况则不容乐观。虽然具有许多模式生物缺少的有趣特征, 且其转录组研究对解决基因进化、遗传育种以及生态等诸多方面的问题具有重要意义。但是由于大量的人力和财力都投入到了模式生物的基因组研究当中, 导致非模式生物的基因组信息严重缺乏。没有相应物种的参考基因组信息, 使得非模式生物转录组的研究举步维艰。传统的研究方法是建立 cDNA 文库, Sanger 法测序获得基因, 构建基因芯片, 其操作复杂, 实验周期长, 花费大, 让很多研究人员望而却步。

近年来, 随着 DNA 高通量测序技术的发展, 出现了“下一代”大规模平行测序(Massive parallel sequencing, MPS)技术, 如 Roche 公司(454 GS-FLX)、Illumina 公司(Genome Analyzer II)和 ABI(AB SOLiD), 这些大规模平行测序技术的出现已完全改变了转录组研究的方式, 产生了“RNA 测序技术(RNA-seq)”。RNA-seq 也称为转录组测序, 与传统的基因芯片技术^[1]相比, 该技术可以高通量的测定转录组 cDNA 的序列, 揭示特定细胞或组织中表达的全部基因或表达序列标签(Expressed sequence tag, EST)、不同基因的相对表达丰度(表达量)(Abundance)、发现转录水平的 SNP、识别一个基因不同

的转录本和可变剪切位点(不同拼接)、SSR 等遗传多态性和遗传标记、能够检测未知基因、发现新的转录本等。RNA-seq 提供精确的数字化信号, 更高的检测通量以及更广泛的检测范围, 是目前深入研究转录组复杂性的强大工具。

目前, RNA-seq 已经成功用于水稻(*Oryza sativa*)^[2]、玉米(*Zea mays*)^[3,4]、拟南芥(*Arabidopsis thaliana*)^[5,6]的大规模 EST 测序研究, 发现了这些物种更多 EST。但是这些物种都是模式生物, 它们不但具有完整的基因组 DNA 序列信息, 而且具有丰富的转录组序列。这些已知的序列信息为测序序列的基因组定位(Mapping)或组装(Assembly)提供强大的支持。RNA-seq 在用于检测基因表达和转录组研究时, 最显著的优势是无需像基因芯片那样需先解码研究物种的基因信息并设计特异性的探针。因此, RNA-seq 可在没有研究物种基因信息的情况下, 直接对任何物种的转录组进行分析。RNA-seq 的这一特征, 弥补了非模式生物转录组研究中缺乏基因组信息的不足。但是, 与模式生物转录组研究相比, 由于缺乏基因组信息, 非模式生物仍不能进行测序序列的基因组定位和注释, 只能进行从头拼装(*De novo assembly*), 并通过同源比对进行测序序列的注释和分析。这也就要求有强大的生物信息学方法和软件作为基础。不过随着生物信息学方法的不断进步, 近年来 RNA-seq 已成为非模式生物转录组研究的先进技术, 大量非模式生物的转录组得到研究。

本文综述了利用 RNA-seq 进行非模式生物转录组研究的概况, 总结了使用 RNA-seq 技术研究非模式生物转录组的一般流程及方法, 最后对非模式生物转录组研究中有待进一步分析的问题进行展望。

1 非模式生物转录组研究概况

RNA-seq 对非模式生物转录组的研究因无参考基因组(Reference genome)信息, 被称为从头转录组

分析(De novo transcriptome analysis)。2008 年, Vera 等^[7]运用 454 GS-20 测序技术进行了第一例从头转录组分析研究(*M. cinxia*; Lepidoptera: Nymphalidae)。之后已经有大量的非模式生物的转录组通过 RNA-seq 得到研究(图 1)。由图 1 可以看出, 得到研究的非模式生物的数量在逐年增加, 尤其是 2010 年, 有高达 36 个非模式生物物种利用 RNA-seq 技术进行了转录组分析研究。另外, 通过文献查阅, 笔者对近 5 年使用 RNA-seq 技术研究非模式生物的情况进行了汇总(表 1)。

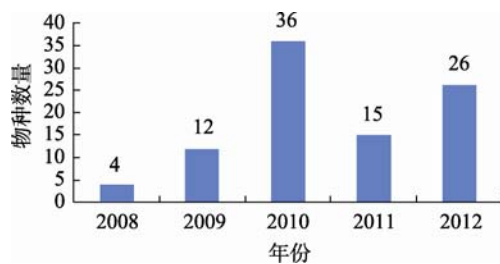


图 1 近 5 年利用 RNA-seq 研究的非模式生物数量统计

由表 1 可以看出, 以上非模式生物的从头转录组研究绝大多数是运用 Roche GS-20 sequencer、Roche 454 GS-FLX、454 GS-FLX Titanium 完成。只有少数是运用 Illumina GAII、GAIIx、HiSeq™ 2000 完成, 包括油菜(*B. napus*)^[19]、烟粉虱(*B. tabaci*)^[23]、甘薯(*I. batatas*)^[50]、弓形虫(*T. gondii*)^[57]、不吉按蚊(*A. funestus*)^[58]、雏豆(*C. arietinum* L.)^[73]、大蒜(*A. sativum*)^[74]、斑海豹(*P. largha*)^[75]、贯叶连翘(*H. perforatum*)^[77]、红花(*C. tinctorius* L.)^[79]、胡黄连(*P. kurroo*)^[80]、盐沼甲(*P. chaldeus*)^[84]、静水椎实螺(*L. stagnalis*)^[86]、桔小实蝇(*B. dorsalis*)^[87]、欧洲笠螺(*P. vulgata*)^[90]、首乌(*P. cuspidatum*)^[94]、紫甘薯(*I. batatas* L.)^[98]。而 ABI 公司的 AB SOLiD 技术则几乎没有运用, 仅仅在甜瓜(*C. melo* L.)^[95]的 SNP 分析上有所运用。究其原因 SOLiD 测序技术读长较短(平均读长 50 bp), 并且测序运行时间较长。由于非模式生物缺乏参考基因组信息, 测序读长越长, 越有利于测序片段的装配。因此, 读长最长的 Roche

表 1 通过大规模并行测序技术从头组装转录组的非模式生物

非模式生物	测序平台	完成国家/地区	时间	文献
腐霉菌(<i>Pythium ultimum</i>)	FLX	UK & USA, etc	2008	[8]
巨桉(<i>Eucalyptus grandis</i>)	GS-20 & FLX	USA	2008	[9]
庆网蛱蝶(<i>Melitaea cinxia</i>)	GS-20	USA	2008	[7]
豌豆(<i>Pisum sativum</i>)	GS-20 & FLX	Germany & USA	2008	[10]
珊瑚(<i>Acropora millepora</i>)	FLX	Australia & USA	2009	[11]
绵鳚(<i>Zoarces viviparus</i>)	FLX	Sweden	2009	[12]
墨西哥玉米(<i>Palomero Toluqueño</i>)	GS-20	USA	2009	[13]
鲟鱼(<i>Acipenser fulvescens</i>)	FLX	USA	2009	[14]
甲虫(<i>Chrysomela tremulae</i>)	FLX	UK & France	2009	[15]
黄花蒿(<i>Artemisia annua</i>)	FLX	China	2009	[16]
橄榄(<i>Olea europaea</i>)	FLX	Italy	2009	[17]
美洲栗(<i>Castanea dentata</i>)	GS-20 & FLX	USA	2009	[18]
板栗(<i>Castanea mollissima</i>)	GS-20 & FLX	USA	2009	[18]
油菜(<i>Brassica napus</i>)	GA	UK & China	2009	[19]
六星灯蛾(<i>Zygaena filipendulae</i>)	FLX	Denmark	2009	[20]
麻蝇(<i>Sarcophaga crassipalpis</i>)	FLX	USA	2009	[21]
苹果实蝇(<i>Rhagoletis pomonella</i>)	FLX	USA	2009	[22]
烟粉虱(<i>Bemisia tabaci</i>)	GAII	China	2010	[23]
白鲑(<i>Coregonus clupeaformis</i>)	FLX	Canada & Germany	2010	[24,25]
丹参(<i>Salvia miltiorrhiza</i>)	FLX Titanium	China	2010	[26]
灰飞虱(<i>Laodelphax striatellus</i>)	FLX	China	2010	[27]
吉丁虫(<i>Agrilus planipennis</i>)	FLX	USA	2010	[28]

续表 1

非模式生物	测序平台	完成国家/地区	时间	文献
菌核菌(<i>Sclerotium rolsii</i>)	454 Life Sciences™	Germany	2010	[29]
虹鳉(<i>Oncorhynchus mykiss</i>)	FLX Titanium	USA	2010	[30]
釉蛱蝶(<i>Heliconius melpomene</i>)	FLX	USA & UK, etc	2010	[31]
弄蝶(<i>Erynnis propertius</i>)	FLX Titanium	USA	2010	[32]
凤蝶(<i>Papilio zelicaon</i>)	FLX Titanium	USA	2010	[33]
庆网蛱蝶(<i>Melitaea cinxia</i> L.)	FLX	USA	2010	[33]
黄瓜(<i>Cucumis sativus</i> L.)	FLX	USA & China	2010	[34]
粪金龟(<i>Onthophagus Taurus</i>)	FLX	USA	2010	[35]
螺旋蝇(<i>Cochliomyia hominivorax</i>)	FLX	Brazil	2010	[36]
欧洲鳗鲡(<i>Anguilla anguilla</i>)	FLX Titanium	Italy & Belgium,etc	2010	[37]
多头绒泡菌(<i>Physarum polycephalum</i>)	FLX	Germany	2010	[38]
红吼北蜂鸟(<i>Archilochus colubris</i>)	FLX	USA	2010	[39]
养环鸽(<i>Streptopelia risoria</i>)	FLX	USA	2010	[39]
蓝山雀(<i>Parus caeruleus</i>)	FLX	USA	2010	[39]
虎皮鹦鹉(<i>Melopsittacus undulates</i>)	FLX	USA	2010	[39]
金领娇鹛(<i>Manacus vitellinus</i>)	FLX	USA	2010	[39]
斑姬鹀(<i>Ficedula hypoleuca</i>)	FLX	USA	2010	[39]
鹈鹕(<i>Dromaius novaehollandiae</i>)	FLX	USA	2010	[39]
小嘴乌鸦(<i>Corvus corone</i>)	FLX	USA	2010	[39]
短嘴鸦(<i>Corvus brachyrhynchos</i>)	FLX	USA	2010	[39]
安氏蜂鸟(<i>Calypte anna</i>)	FLX	USA	2010	[39]
西洋参(<i>Panax quinquefolius</i> L.)	FLX Titanium	China	2010	[40]
黑松(<i>Pinus contorta</i>)	GS XLR70 Titanium	USA	2010	[41]
乌鸦(<i>Corvus [corone] corone</i>)	FLX	Germany	2010	[42]
乌鸦(<i>Corvus [corone] cornix</i>)	FLX	Germany	2010	[42]
束带蛇(<i>Thamnophis elegans</i>)	FLX Titanium	USA	2010	[43]
线虫(<i>Trichostrongylus colubriformis</i>)	FLX	USA & Australia	2010	[44]
蚜虫(<i>Aphis glycines</i>)	FLX Titanium	USA	2010	[45]
野生稻(<i>Oryza longistaminata</i>)	FLX	Germany	2010	[46]
高水麻(<i>Amaranthus tuberculatus</i>)	FLX	USA	2010	[47]
黄孢平革菌(<i>Phanerochaete chrysosporium</i>)	GS-20	USA	2010	[48]
真涡虫(<i>Schmidtea mediterranea</i>)	FLX	Spain	2010	[49]
甘薯(<i>Ipomoea batatas</i>)	GAII	China	2010	[50]
甘薯(<i>Ipomoea batatas</i> L.)	FLX Titanium	Peru	2010	[51]
滨螺(<i>Littorina saxatilis</i>)	FLX	UK	2010	[52]
南极帽贝(<i>Laternula elliptica</i>)	FLX	Portugal & UK	2010	[53]
夏栎(<i>Quercus robur</i>)	FLX Titanium	France & Japan	2010	[54]
无梗栎(<i>Quercus petraea</i>)	FLX Titanium	France & Japan	2010	[54]
草苔虫(<i>Bugula neritina</i>)	FLX	China & USA	2010	[55]
淫羊藿(<i>Epimedium sagittatum</i>)	FLX	USA & China	2010	[56]
弓形虫(<i>Toxoplasma gondii</i>)	GA	Japan	2010	[57]
不吉按蚊(<i>Anopheles funestus</i>)	GA IIx	France & USA	2010	[58]
白蜡树(<i>Fraxinus spp.</i>)	FLX Titanium	USA	2011	[59]

续表 1

非模式生物	测序平台	完成国家/地区	时间	文献
臭虫(<i>Cimex lectularius</i>)	FLX Titanium	USA	2011	[60]
白粉虱(<i>Trialeurodes vaporariorum</i>)	FLX	UK & Germany, etc	2011	[61]
紫杉(<i>Taxus cuspidata</i>)	FLX Titanium	China	2011	[62]
蝴蝶兰(<i>Phalaenopsis orchids</i>)	FLX	China	2011	[63]
磷虾(<i>Euphausia superba</i>)	FLX Titanium	UK & France, etc	2011	[64]
螨虫(<i>Phytoseiulus persimilis</i>)	FLX	Egypt	2011	[65]
欧洲蕨(<i>Pteridium aquilinum</i>)	FLX Titanium	USA	2011	[66]
荞麦(<i>Fagopyrum</i>)	FLX	Russia	2011	[67]
响尾蛇(<i>Crotalus adamanteus</i>)	FLX	USA	2011	[68]
鳕鱼(<i>Gadus morhua</i>)	FLX	Norway	2011	[69]
燕麦(<i>Avena barbata</i>)	FLX	USA	2011	[70]
不吉按蚊(<i>Anopheles funestus</i>)	FLX	UK & South Africa	2011	[71]
西葫芦(<i>Cucurbita pepo</i>)	FLX Titanium	Spain	2011	[72]
雏豆(<i>Cicer arietinum</i> L.)	GA	India	2011	[73]
大蒜(<i>Allium sativum</i>)	HiSeq™ 2000	China	2012	[74]
斑海豹(<i>Phoca largha</i>)	HiSeq2000	China	2012	[75]
灯草鹑(<i>Junco hyemalis</i>)	FLX Titanium	USA	2012	[76]
贯叶连翘(<i>Hypericum perforatum</i>)	HiSeq™ 2000	China	2012	[77]
黑琴鸡(<i>Tetrao tetrix</i>)	FLX Titanium	Sweden & USA	2012	[78]
红花(<i>Carthamus tinctorius</i> L.)	HiSeq™ 2000	China	2012	[79]
胡黄连(<i>Picrorhiza kurrooa</i>)	GA IIx	India	2012	[80]
环节动物(<i>Pristina leidy</i>)	FLX Titanium	USA	2012	[81]
沙冬青(<i>Ammopiptanthus mongolicus</i>)	GS XLR70 Titanium	China	2012	[82]
鸡枞菌(<i>Termitomyces albuminosus</i>)	FLX Titanium	China	2012	[83]
盐沼甲(<i>Pogonus chalceus</i>)	HiSeq™ 2000	Belgium	2012	[84]
假山毛榉(<i>Nothofagus nervosa</i>)	FLX	Argentina	2012	[85]
静水椎实螺(<i>Lymnaea stagnalis</i>)	GAIIx	Japan	2012	[86]
桔小实蝇(<i>Bactrocera dorsalis</i>)	GAI	Taiwan, China & USA	2012	[87]
康乃馨(<i>Dianthus caryophyllus</i> L.)	FLX Titanium	Japan	2012	[88]
鲤鱼(<i>Cyprinus carpio</i>)	FLX Titanium	China	2012	[89]
欧洲笠螺(<i>Patella vulgata</i>)	GAI	UK	2012	[90]
旗尾更格芦鼠(<i>Dipodomys spectabilis</i>)	FLX Titanium	USA	2012	[91]
日本沼虾(<i>Macrobrachium nipponense</i>)	FLX	China	2012	[92]
食蟹猴(<i>Macaca fascicularis</i>)	FLX	Korea	2012	[93]
首乌(<i>Polygonum cuspidatum</i>)	HiSeq™ 2000	China	2012	[94]
甜瓜(<i>Cucumis melo</i> L.)	SOLiD	Spain	2012	[95]
豌豆(<i>Pisum sativum</i> L.)	FLX Titanium	Australia	2012	[96]
蚕豆(<i>Vicia faba</i> L.)	FLX Titanium	Australia	2012	[96]
武昌鱼(<i>Megalobrama amblycephala</i>)	FLX Titanium	China	2012	[97]
紫甘薯(<i>Ipomoea batatas</i> L.)	HiSeq™ 2000	USA & China	2012	[98]
大比目鱼(<i>Scophthalmus maximus</i>)	FLX Titanium	Spain	2012	[99]

注: GS-20: Roche 454 GS-20; FLX: Roche 454 GS-FLX; FLX-Titanium: Roche 454 GS-FLX Titanium; GA: Illumina GA; GAI: Illumina GAI; GAIx: Illumina GAIx; HiSeq™ 2000: Illumina HiSeq™ 2000; SOLiD: AB SOLiD™ System; KSA: Kingdom of Saudi Arabia.

454 技术(平均读长 400 bp)在非模式生物转录组研究中应用最为广泛, 其次是 Illumina Solexa 技术(平均读长 100 bp)有少数应用。

从表 1 可以看出, 非模式生物的头转录组研究绝大多数是由国外的科研机构完成的, 他们最先将 RNA-seq 技术使用到非模式动物转录组的研究上, 而国内, 则起步稍晚。目前, 国内也已经展开了对中药植物、农业昆虫等物种的头转录组研究, 包括黄花蒿(*A. annua*)^[16]、油菜(*B. napus*)^[19]、烟粉虱(*B. tabaci*)^[23]、丹参(*S. miltiorrhiza*)^[26]、灰飞虱(*L. Striatellus*)^[27]、黄瓜(*C. sativus* L.)^[34]、西洋参(*P. Quinquefolius* L.)^[40]、甘薯(*I. batatas*)^[50]、淫羊藿(*E. Sagittatum*)^[56]、紫杉(*T. cuspidate*)^[62]、蝴蝶兰(*P. orchids*)^[63]、大蒜(*A. sativum*)^[74]、红花(*C. tinctorius* L.)^[79]、首乌(*P. cuspidatum*)^[94]等。这些研究证明 RNA-seq 在发现非模式生物转录组基因及遗传标记中非常有效。例如用新一代高通量测序技术 454 GS-FLX Titanium 对 2 年生丹参根的转录组进行测序, 研究其基因表达谱, 挖掘其功能基因, 获得 46 722 表达序列标签(EST), 序列平均长度 414 bp, 与 Sanger 测序的长度相当; 所得序列与 GenBank 中丹参的 EST 合并拼接, 获得 18 235 条唯一基因(Unigene), 包括 6 620 个序列重叠群(Contig)和 11 615 条单一序列(Singleton), 唯一基因总长 7.89 Mb。在全部唯一基因中, Roche 454 高通量测序发现了 13 980 条新的唯一基因。数据库中的序列同源性比较表明, 其中 73.0% (13 308 条)与其他生物的已知基因具有不同程度的同源性(获得了基因注释)。根据拟南芥蛋白质组数据库注释结果, 被注释序列大约包含 7 800 个转录本; 另有 4 927 条唯一基因 (27.0%)未被注释, 认为是可能的新基因。通过 BLAST 与基因本体论(Gene Ontology)分析获得了可能参与丹参酮合成的序列 27 条(编码 15 个关键酶), 参与丹酚酸合成的序列 29 条(编码 11 个关键酶), 细胞色素 P450 序列 70 条, 转录因子序列 577 条。证明 Roche 454 高通量测序技术作为药用植物功能基因组研究的重要手段可在丹参功能基因的发现中发挥重要作用, 这些基因的发现为丹参酮和丹酚酸类化合物生物合成研究奠定了基础, 同时也为丹参的转录组研究提供了基础数据^[26]。

另外, 将 RNA-seq 和定制基因芯片联合使用, 对于快速研究非模式生物的转录组非常有效^[7,11,12]。应

用第二代测序技术对非模式生物进行转录组测序, 发现基因或 EST 信息, 随后用定制“基因/EST 芯片”对单个个体的基因表达情况进行快速鉴定, 这一研究思路在非模式生物的功能基因组学研究方面越来越流行。例如, 来自美国和芬兰的科学家相互合作, 运用第二代测序技术和定制芯片, 对一种没有参考序列的非模式生物——庆网蛱蝶(*M. cinxia*)进行研究^[7]。研究者首先提取不同家系的庆网蛱蝶幼虫、蛹和成虫混合样本的 RNA, 反转录为 cDNA。用第二代测序技术进行测序。共得到 608 053 个表达序列标签(平均长度 110 bp), 获得 48 354 个重叠群和 59 943 个单拷贝。为了保证实验结果的正确性, 研究者同时取了一部分 RNA 样品, 构建 cDNA 文库, 用传统的 Sanger 法测序, 以此做为第二代测序结果的参照。Sanger 法测序得到 3 888 条序列, 经过拼接后, 得到 813 条可用序列, 其中 749 条(92%)能在第二代测序的结果中找到。这表明第二代测序的结果具有很高的准确率。由于取样时, 研究者取了不同家系的庆网蛱蝶做为样本, 所以依据第二代测序的结果, 研究者还发现了庆网蛱蝶的一些 SNP 位点和可变剪接方式。随后研究者根据测序得到的数据, 用 Agilent 的 e-array 自主设计基因芯片。由于昆虫体内一般都会有一些共生或者寄生的微生物, 因此以成虫为样本时, 就顺带提取了微生物的 RNA, 并进行测序, 针对微生物的探针也包括在基因芯片上。研究者用自主设计的基因芯片, 检测不同家系单个庆网蛱蝶个体的基因表达情况, 结果发现了不同家系的庆网蛱蝶携带的微生物也不同。这一成果对于研究庆网蛱蝶的种群特征具有重要意义。这篇文章提供给我们一种比较新颖的研究思路。当我们想对那些没有参考基因组信息的非模式生物进行基因组研究时, 我们可以先用第二代测序技术对混合样本进行测序, 发现基因, 根据序列数据, 用 Agilent 的 e-array 进行定制基因芯片, 然后用基因芯片对不同种群或者不同个体的基因转录情况进行快速分析。

2 非模式生物转录组研究的一般流程及方法

利用 RNA-seq 技术研究非模式生物转录组的一般流程主要包括: 测序 RNA 样品的准备、高通量 DNA 测序和生物信息学分析(图 2)。由于非模式生物无参考基因组信息, 因此与模式生物转录组研究

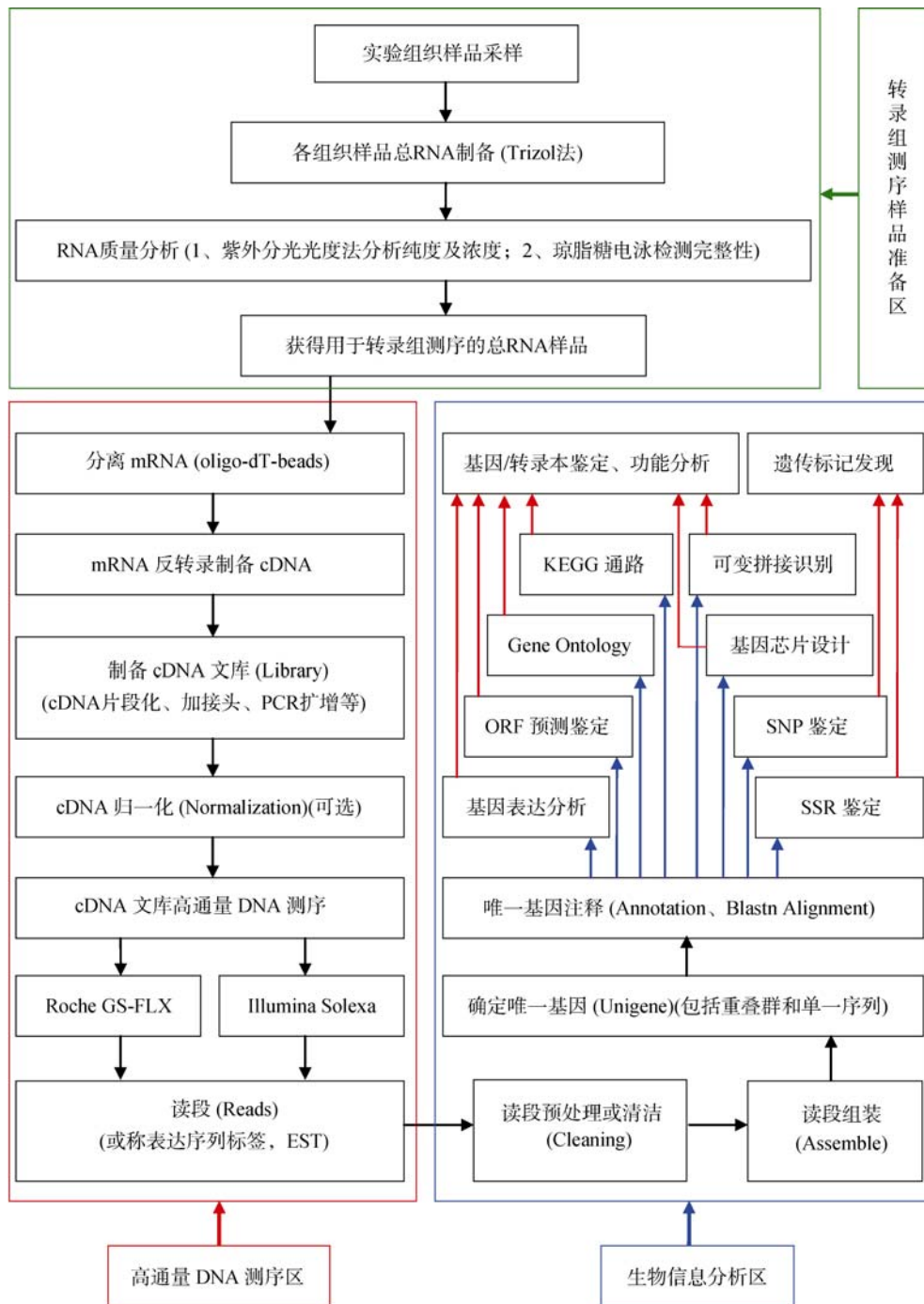


图 2 非模式生物转录组研究流程图

有一定的差异,例如,获得非模式生物的 RNA-seq 数据后,首先需要进行测序读段(Reads)的从头组装(*De novo* assembly),生成唯一基因(包括重叠群和单一序列),才能进行下一步的基因注释和生物信息学分析;而对于模式生物,则可以直接将测序读段定位(Mapping)到参考基因组上进行基因注释及下游

分析。下面简要介绍非模式生物转录组研究的一般流程与方法。

2.1 样品准备

2.1.1 采样

根据所研究的非模式生物及其研究目的,需要

选择性地采集不同的样本。例如, Vera 等^[7]在研究庆网蛱蝶(*M. cinxia*)转录组时, 采集了不同家系庆网蛱蝶的幼虫、蛹和成虫组成混合样本(8 个家系, 80 个个体); Gregory 等^[71]在对不吉按蚊(*A. funestus*)的表达谱重测序(*De novo* expression profiling)的研究中, 为了获得一个覆盖更广泛的转录组数据集, 分别采集了两个品系不吉按蚊不同生长阶段的样本(每个品系取 10 个 4 龄幼虫、10 个蛹, 10 个雌性成虫和 10 个雄性成虫); Meyer 等^[11]在珊瑚(*A. millepora*)转录组重新测序分析研究时, 选择珊瑚的浮浪幼虫作为实验材料, 为了增加表达基因的多样性, 研究者将珊瑚的浮浪幼虫分成两组进行了不同的温度处理(28 和 32 培养 5 d); Jeukens 等^[25]利用 Roche GS-FLX 平台研究白鲑(*C. clupeaformis*)的 SNP 时, 为了增加基因型和表达基因的多样性, 采集了 24 个个体和 3 种不同的组织(白色肌、脑和肝脏); Kaur 等^[96]在发现和验证豌豆和蚕豆的 SSR 遗传标记的研究中, 分别采集了其各个生长阶段的样本, 包括叶子(幼叶和老叶)、茎、花、未成熟的豆荚、成熟的豆荚和未成熟的种子。由此可知, 非模式生物转录组研究的样本采集是一个复杂的过程, 需要研究者根据不同的研究目的, 采集不同的样本, 进行不同的条件处理。

2.1.2 RNA 制备

研究者根据自己的研究目的, 采集到相应的样本之后, 就需要提取样本中的总 RNA。笔者通过文献查阅, 发现非模式生物转录组研究中大部分样本的总 RNA 是通过 TRIzol 试剂(Invitrogen)提取的。当然也有其他方法, 但只有少数运用, 例如, 李滢等^[26]采用通用植物总 RNA 提取试剂盒(百泰克公司)提取丹参根总 RNA; Parchman 等^[41]采用 CTAB 法, 可以简单快速的提取松树总 RNA。

此外, 用于非模式生物转录组测序研究的总 RNA 样品, 要满足 A_{260}/A_{280} 在 1.9~2.1 之间, A_{260}/A_{230} 在 2.0~2.5 之间, 以及 RNA 完整性指数大于 8 才能用于后续的研究。因此在进行下一步 cDNA 合成之前要对 RNA 样品进行 DNase I 的处理(37 °C, 30 min)和进一步纯化(RNeasy MinElute Clean up Kit, Qiagen)。经过 DNase I(RNase Free)处理并检测合格的 RNA 样品即可用于下一步的 cDNA 文库的构建。

非模式生物转录组测序中, 为了尽可能囊括该物种所有基因, 测序的 RNA 样品制备中, 提取各种组织的总 RNA, 再将各种组织的总 RNA 混合成一个样品, 进行高通量测序。

2.2 cDNA 文库制备与测序

由于大部分非模式生物从头转录组研究采用 Roche 454 技术, 以新一代高通量测序平台 Roche 454 GS FLX 为例, 其 cDNA 文库制备与测序流程简述如下: (1)从总 RNA 样品中提取与纯化 mRNA, 这一过程一般使用 mRNA 提取试剂盒完成, 如 Oligotex mRNA Mini kit (Qiagen); (2)以纯化的 mRNA 为模板合成双链 cDNA, 并纯化合成的 cDNA; (3)cDNA 片段化, 即打断成 400~800 bp 的片段; (4)cDNA 片段的末端补平; (5)在两个末端连接上 A 和 B 接头(3'和 5'端具有特异性); (6)具有 A、B 接头的单链 DNA 片段被纯化回收后即组成了样品文库, 至此 cDNA 文库制备完成; (7)将单链 DNA 文库固定在特别设计的 DNA 捕获磁珠上, 使大部分磁珠携带一个独特的单链 DNA 片段, 然后进行乳液 PCR 扩增; (8)经乳液 PCR 扩增后, 乳液混合物被打破, 扩增的片段仍然结合在磁珠上, 携带 DNA 的捕获磁珠随后放入“Pico TiterPlate”(PTP)板中进行后续的测序。

另外, 在 cDNA 文库制备过程中还存在归一化(Normalization)问题。归一化过程一般包括 cDNA 变性与重新组合、DSN(Duplex-specific nuclease)处理以及归一化片段的扩增。归一化会降低样本中高丰度转录本的数量, 有利于新基因的发现, 并且可以使代表不同转录本的读段数量占总测序读段的比例增加, 有利于 SNP 的发现与分析。但是, 也有文献^[14]指出, 归一化对基因发现效率的影响很小, 并且归一化还会消除自然 cDNA 样品中基因表达量的差异, 以及影响等位基因频率(标准化倾向于降低常见等位基因频率)。因此, 研究者在进行非模式生物转录组研究时, 需要根据自己的研究目的选择是否进行 cDNA 文库的归一化。

2.3 数据处理及生物信息学分析

2.3.1 数据过滤

对于非模式生物而言, 获得 RNA-seq 原始数据

后, 首先需要进行序列的从头组装, 这是后续研究与分析的基础。但是, 在序列从头组装之前, 有时还需要根据测序数据情况对其做某些基本的预处理。例如, 数据过滤。

由于 RNA-seq 原始数据中存在大量的测序接头序列、polyA/T 序列、载体序列、低质量的读段(序列中存在大量的 N)、较短的序列(<100 bp, 454 测序数据)以及其他污染序列(如细菌基因序列), 这将严重影响后续重新组装的质量。而数据过滤的目的是为了去除这些冗余和污染序列。一般来说, 测序公司会运用一些相应软件对测序数据进行初步过滤, 如 Roche 公司运用一个信号处理软件自动去除接头序列。而进一步的数据过滤就需要借助于其他软件, 如, SeqClean (DFCI Gene Indices Software Tools [http://compbio.dfci.harvard.edu/tgi/software/]) 可以去除原始数据中的 polyA/T 序列、测序接头序列、低质量序列、载体序列、细菌基因序列等, 因此应用最广泛。另外, 还有 Exonerate (一个通用的序列比对工具 [http://www.ebi.ac.uk/~guy/exonerate/]), 可以在测序中去除接头序列; Lucy (Lucy DNA sequence quality and vector trimming tool [http://lucy.sourceforge.net/]) 可以去除低质量读段。通常数据过滤可以借助几种软件的联合使用, 达到最佳过滤效果。

2.3.2 从头组装

在缺乏参考基因组信息的非模式生物转录组研究中, 获得 RNA-seq 原始数据后, 首先需要将所有测序读段通过从头组装生成重叠群和单一序列, 这是后续处理及生物学功能分析的基础。要完成测序读段的从头组装并不简单, 因此, 为了有效地做好从头组装, 目前生物信息学领域已经开发出多种组装程序^[73,100], 包括: GS Assembler ver. 1.1.02.15 (Roche)、Newbler^[101]、MIRA^[102]、CAP3^[103]、SeqMan^[104]、TGICL^[105]、stackPACK^[106]、Velvet^[107]、AbySS^[108]、ALLPATHS2^[109]、Oases^[110]、SOAPdenovo^[111]、Multiple-k method^[112]、Scaffolding using translation mapping (STM)^[112]、Trinity^[113]、PCAP^[114]。这些从头组装程序的开发和应用, 密切地配合了高通量 DNA 测序技术的实际应用, 使得非模式生物转录组的 RNA-seq 研究得到快速发展。如表 2 所示, 为近 5 年来非模式生物转录组研究中常用的组装程序。可以看到, 应用最为广泛的是 Roche 的 Newbler, 依次是 CAP3、MIRA、SeqMan 和 TGICL, 它们都是用于 Roche 454 测序读段的从头组装。而 SOAPdenovo、Velvet、Trinity 和 PCAP 则是文献中报道的用于更短测序读段(Illumina GA 测序读段)从

表 2 近 5 年来非模式生物转录组研究中常用的组装程序

组装程序	非模式生物
Newbler	<i>Eucalyptus grandis</i> ^[91] , <i>Coregonus</i> spp. <i>Salmonidae</i> ^[24] , <i>Acropora millepora</i> ^[11] , <i>Castanea dentata</i> and <i>Castanea mollissima</i> ^[112] , <i>Palomero toluqueño</i> ^[113] , <i>Acipenser fulvescens</i> ^[14] , <i>Sarcophaga crassipalpis</i> ^[21] , <i>Rhagoletis pomonella</i> ^[22] , <i>Sclerotium rolfsii</i> ^[29] , <i>Melitaea cinxia</i> L. ^[33] , <i>Onthophagus Taurus</i> ^[35] , <i>Panax quinquefolius</i> ^[40] , <i>Thamnophis elegans</i> ^[43] , <i>Aphis glycines</i> ^[45] , <i>Laternula elliptica</i> ^[53] , <i>Fraxinus</i> spp. ^[59] , <i>Cimex lectularius</i> ^[60] , <i>Phalaenopsis orchids</i> ^[63] , <i>Euphausia superba</i> ^[64] , <i>Phytoseiulus persimilis</i> ^[65] , <i>Junco hyemalis</i> ^[76] , <i>Tetrao tetrix</i> ^[78] , <i>Pristina leidy</i> ^[81] , <i>Nothofagus nervosa</i> ^[85] , <i>Cyprinus carpio</i> ^[89] , <i>Macrobrachium nipponense</i> ^[92] , <i>Macaca fascicularis</i> ^[93]
CAP3	<i>Pisum sativum</i> ^[10] , <i>Artemisia annua</i> ^[16] , <i>Laodelphax striatellus</i> ^[27] , <i>Erynnis propertius</i> , <i>Papilio zelicaon</i> ^[32] , <i>Cucumis sativus</i> L. ^[34] , <i>Oryza longistaminata</i> ^[46] , <i>Amaranthus tuberculatus</i> ^[47] , <i>Phanerochaete chrysosporium</i> ^[48] , <i>Bugula neritina</i> ^[55] , <i>Epimedium sagittatum</i> ^[56] , <i>Pteridium aquilinum</i> ^[66] , <i>Avena barbata</i> ^[70] , <i>Ammopiptanthus mongolicus</i> ^[82] , <i>Termitomyces albuminosus</i> ^[83] , <i>Cyprinus carpio</i> ^[89]
MIRA	<i>Chrysomela tremulae</i> ^[15] , <i>Zygaena filipendulae</i> ^[20] , <i>Heliconius melpomene</i> ^[31] , <i>Cucumis sativus</i> L. ^[34] , <i>Onthophagus Taurus</i> ^[35] , <i>Anguilla anguilla</i> ^[37] , <i>Schmidtea mediterranea</i> ^[49] , <i>Trialeurodes vaporariorum</i> ^[61] , <i>Fagopyrum</i> ^[67] , <i>Anopheles funestus</i> ^[71] , <i>Ammopiptanthus mongolicus</i> ^[82] , <i>Cyprinus carpio</i> ^[89] , <i>Megalobrama amblycephala</i> ^[97] , <i>Scophthalmus maximus</i> ^[99]
SeqMan	<i>Melitaea cinxia</i> ^[7] , <i>Pinus contorta</i> ^[41] , <i>Oncorhynchus mykiss</i> ^[80] , <i>Cochliomyia hominivorax</i> ^[36] , <i>Littorina saxatilis</i> ^[52] , <i>Crotalus adamanteus</i> ^[68]
TGICL	<i>Pythium ultimum</i> ^[8] , <i>Pisum sativum</i> ^[10] , <i>Zoarcetes viviparus</i> ^[112] , <i>Quercus robur</i> , <i>Quercus petraea</i> ^[54]
SOAPdenovo	<i>Bemisia tabaci</i> ^[23] , <i>Ipomoea batatas</i> ^[50] , <i>Phalaenopsis orchids</i> ^[63] , <i>Allium sativum</i> ^[74] , <i>Hypericum perforatum</i> ^[77] , <i>Picrorhiza kurroa</i> ^[80] , <i>Polygonum cuspidatum</i> ^[94] , <i>Ipomoea batatas</i> L. ^[98]
Velvet	<i>Anopheles funestus</i> ^[58] , <i>Cicer arietinum</i> L. ^[73] , <i>Lymnaea stagnalis</i> ^[86] , <i>Bactrocera dorsalis</i> ^[87] , <i>Patella vulgata</i> ^[90]
Trinity	<i>Phoca largha</i> ^[75] , <i>Carthamus tinctorius</i> L. ^[79] , <i>Pogonius chalcus</i> ^[84] , <i>Bactrocera dorsalis</i> ^[87]
PCAP	<i>Dipodomys spectabilis</i> ^[91]

头组装的常用程序。从头组装的目的是生成唯一基因(包括重叠群和单一序列),用于下一步的基因注释和生物信息学分析。

2.3.3 基因注释

测序读段通过从头组装生成重叠群和单一序列后,通常使用 BLAST 程序将这些拼接序列唯一基因(重叠群和单一序列)与核酸、蛋白质序列数据库进行比对(E 值 $<1e^{-5}$),选择匹配最好的一项作为注释信息。常用于比对的核酸数据库为 GenBank 非冗余核酸数据库 Nt;常用的蛋白质数据库则包括 Swiss-Prot/UniProtKB、GenBank 非冗余蛋白数据库 Nr、COG 和 KEGG。基因注释一般包括:使用 BLASTN 与 NCBI 中的非冗余核酸数据库 Nt 进行比对,检索相似序列,对唯一基因进行注释;使用 BLASTX 和 BLAST2GO 与蛋白质数据库(Nr 和 SwissProt)比对,获得 GO 信息注释,并对序列(按分子功能、细胞组分、生物学过程)进行分类;使用 BLASTX 与蛋白质数据库(COG、KEGG)比对,获得 COG 注释和 KEGG 代谢途径注释。例如用新一代高通量测序技术 Illumina HiSeq™ 2000 对红花(*C. tinctorius* L.)转录组进行测序研究中, Huang 等^[79]使用 BLAST 程序($E < 1e^{-5}$)将红花唯一基因与核酸数据库(Nt)和蛋白数据库(Nr、SwissProt、KEGG 和 COG)进行比对,共计注释了 70 342 条唯一基因,大约占总唯一基因的 58%。其中, 21 943 条唯一基因具有 COG 注释信息, 30 203 条唯一基因可以匹配到 121 个 KEGG 代谢通路。此外,还利用 BLAST2GO 程序和 WEGO 软件对红花唯一基因进行了 GO 信息注释,共计 26 332 条唯一基因匹配到 1 754 个 GO-term。

2.3.4 其他生物信息学分析

在非模式生物转录组研究中,利用新一代高通量测序技术产生的大量数据除了可以进行以上基因注释中常见的 GO、KEGG 和 COG 分析外,还可以进行大量的其他生物信息学分析。如基因发现(Gene discovery)^[15,78,82,92,99]、SNP 鉴定^[9,11,14,19,34,52,54,59,64,72,76,84,92,97,115]、SSR 鉴定^[26,41,51,56,60,64,70,72,73,75,92,97]、鉴定差异表达基因^[25,26,67]、鉴定控制特定生物学过程的基因^[67,68]、识别可变拼接^[93]等。例如,在日本沼虾(*M. nipponense*)的转录组

研究中, Ma 等^[92]在 GO、KEGG 和 COG 分析的基础上,利用生物信息学手段发现了许多与性别决定有关的基因,包括 DMRT1、FTZ-F1、FOXL2、FEM1 等。除此之外,还利用 Mreps software (<http://bioinfo.lifl.fr/mreps/>)鉴定出 6 689 个 SSR,利用 VarScan (<http://varscan.sourceforge.net/>)鉴定出 18 107 个高质量 SNP。Logacheva 等^[67]通过比较甜芥(*F. esculentum*)和苦芥(*F. tataricum*)的转录组测序结果,分析了两者的差异表达基因,这些差异基因包括逆转座子

基因以及糖类合成与代谢相关基因。另外,利用其他信息学手段还鉴定出许多与荞麦花序发育相关的基因。Huh 等^[93]在食蟹猴(*M. fascicularis*)转录组研究中,分析鉴定出 4 314 个可变剪接事件。总之,生物信息学分析是我们分析研究非模式生物转录组的重要手段,除以上介绍的几项数据分析外,根据不同的研究目的,我们还可以进行预测新的开放阅读框(ORF)^[67,72],确定基因表达丰度^[26],发现遗传标记^[11,41,85],发现 microRNA^[37]等。

3 结语与展望

非模式生物具有许多模式生物缺少的有趣特征,通过非模式生物转录组的研究可以解答基因进化、遗传育种以及生态方面的许多问题。而新一代测序技术的发展,从根本上延伸了转录组研究的范围,普及到了非模式生物和野生型群体,使之成为了值得期待的研究领域。笔者通过文献查阅,尝试对近 5 年非模式生物转录组研究情况进行了简要汇总,统计了近 5 年来利用 RNA-seq 技术研究非模式生物物种数量,并简要介绍了非模式生物转录组研究的一般流程及方法。由于近年来非模式生物转录组的研究快速发展,新的物种转录组研究层出不穷,希望本文讨论的一些内容可以为正在或即将采用 RNA-seq 技术进行非模式生物转录组研究的学者提供参考。

新一代测序技术的发展为非模式生物转录组的研究提供了新的手段,但目前在实际应用中还面临两个重要的问题。一是复杂的生物信息学分析,即如何解读和分析测序产生的庞大数据量所带来的信息。非模式生物缺少参考基因组信息,无法进行测

序读段的基因组定位及基因注释, 因此需要借助一系列生物信息学手段来解读这些数据。这也为生物信息学带来了挑战, 需要发展更简单有效的生物信息学方法, 使一般生物学实验研究人员不依靠专门的生物信息学专业人员就能够实现测序信息的全程分析, 达到自己的研究目的, 从本专业角度或者研究领域最大限度地挖掘数据中蕴含的有用信息。二是研究成本高昂。目前高通量 DNA 测序仍然依赖于非常昂贵的测序设备, 设备购置的一次性投入巨大, 因此测序仪器无法在普通实验室普及, 仅局限于一些实力雄厚的实验室和生物技术公司, 致使转录组测序的成本很高, 一般科研人员望而却步。例如目前在该领域最常用的 Roche 454 GS FLX 测序平台, 产生不少于 80 万条有效序列的一个完整测序反应的费用约在 13 万元左右。此外, 测序数据的生物信息学分析依赖于性能良好、运算速度快的计算机设备(一般是服务器), 其投入和有效的运作也是生物学实验研究人员面临的难题。随着 DNA 测序技术的快速进步、测序成本的逐渐下降、用户友好型软件包的开发、生物学实验研究人员生物信息学知识的提高, 以及同一实验室中生物学实验研究人员与生物信息学开发研究人员的密切合作和团队协作, 这些问题将逐步得到解决, 促进 RNA-seq 技术的大规模应用。

随着 DNA 测序技术的快速发展, 在解码许多模式生物的完整基因组 DNA 序列信息后, 越来越多的非模式生物的基因组也在逐渐测序中。大量物种的完整基因组序列信息的测定, 将进一步推动同一物种内不同生物转录组的测定和准确信息的解码, 简化目前非模式生物转录组测序中生物信息学分析的复杂性。此外, 将 RNA-seq 技术与传统转录组研究方法联合使用, 可以更有效的研究和分析非模式生物转录组。例如, Vera 等^[7]应用第二代测序技术对非模式生物进行转录组测序, 发现基因或 EST 信息, 随后用定制“基因/EST 芯片”对单个个体的基因表达情况进行快速鉴定, 这一研究为非模式生物转录组的研究以及基因功能的挖掘提供了一个美好的前景。

参考文献(References):

- [1] Li X, Chen GH, Zhang WY, Zhang XS. Genome-wide transcriptional analysis of maize endosperm in response to *ae wx* double mutations. *J Genet Genom*, 2010, 37(11): 749–762. DOI
- [2] 李湘龙, 柏斌, 吴俊, 邓启云, 周波. 第二代测序技术用于水稻和稻瘟菌互作早期转录组的分析. *遗传*, 2012, 34(1): 102–112. DOI
- [3] Emrich SJ, Barbazuk WB, Li L, Schnable PS. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res*, 2007, 17(1): 69–73. DOI
- [4] Ohtsu K, Smith MB, Emrich SJ, Borsuk LA, Zhou RL, Chen TL, Zhang XL, Timmermans MC, Beck J, Buckner B, Janick-Buckner D, Nettleton D, Scanlon MJ, Schnable PS. Global gene expression analysis of the shoot apical meristem of maize (*Zea mays* L.). *Plant J*, 2007, 52(3): 391–404. DOI
- [5] Jones-Rhoades MW, Borevitz JO, Preuss D. Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. *PLoS Genet*, 2007, 3(10): 1848–1861. DOI
- [6] Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB. Sampling the arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiol*, 2007, 144(1): 32–42. DOI
- [7] Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol*, 2008, 17(7): 1636–1647. DOI
- [8] Cheung F, Win J, Lang JM, Hamilton J, Vuong H, Leach JE, Kamoun S, Levesque CA, Tisserat N, Buell CR. Analysis of the *Pythium ultimum* transcriptome using Sanger and pyrosequencing approaches. *BMC Genomics*, 2008, 9(1): 542. DOI
- [9] Novaes E, Drost DR, Farmerie WG, Pappas GJ, Grattapaglia D, Sederoff RR, Kirst M. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics*, 2008, 9(1): 312. DOI
- [10] Bräutigam A, Shrestha RP, Whitten D, Wilkerson CG, Carr KM, Froehlich JE, Weber APM. Low-coverage massively parallel pyrosequencing of cDNAs enables proteomics in non-model species: Comparison of a species-specific database generated by pyrosequencing with databases from related species for proteome analysis of pea chloroplast envelopes. *J Biotechnol*, 2008, 136(1–2): 44–53. DOI
- [11] Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, Colbourne JK, Willis BL, Matz MV. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFLX. *BMC Genomics*, 2009, 10(1): 219. DOI
- [12] Kristiansson E, Asker N, Förlin L, Larsson DGJ. Characterization of the *Zoarces viviparus* liver transcriptome

- using massively parallel pyrosequencing. *BMC Genomics*, 2009, 10(1): 345. [DOI](#)
- [13] Vega-Arreguin J, Ibarra-Laclette E, Jiménez-Moraila B, Martínez O, Vielle-Calzada J, Herrera-Estrella L, Herrera-Estrella A. Deep sampling of the *Palomero* maize transcriptome by a high throughput strategy of pyrosequencing. *BMC Genomics*, 2009, 10(1): 299. [DOI](#)
- [14] Hale MC, McCormick CR, Jackson JR, DeWoody JA. Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. *BMC Genomics*, 2009, 10(1): 203. [DOI](#)
- [15] Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, Ffrench-Constant RH. Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera. *Insect Biochem Mol Biol*, 2009, 39(5–6): 403–413. [DOI](#)
- [16] Wang W, Wang YJ, Zhang Q, Qi Y, Guo DJ. Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics*, 2009, 10(1): 465. [DOI](#)
- [17] Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, Giuliano G, Chiusano ML, Baldoni L, Perrotta G. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics*, 2009, 10(1): 399. [DOI](#)
- [18] Barakat A, DiLoreto DS, Zhang Y, Smith C, Baier K, Powell WA, Wheeler N, Sederoff R, Carlson JE. Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *BMC Plant Biol*, 2009, 9(1): 51. [DOI](#)
- [19] Trick M, Long Y, Meng JL, Bancroft I. Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnol J*, 2009, 7(4): 334–346. [DOI](#)
- [20] Zagrobelny M, Scheibye-Alsing K, Jensen NB, Moller BL, Gorodkin J, Bak S. 454 pyrosequencing based transcriptome analysis of *Zygaena filipendulae* with focus on genes involved in biosynthesis of cyanogenic glucosides. *BMC Genomics*, 2009, 10(1): 574. [DOI](#)
- [21] Hahn DA, Ragland GJ, Shoemaker DD, Denlinger DL. Gene discovery using massively parallel pyrosequencing to develop ESTs for the flesh fly *Sarcophaga crassipalpis*. *BMC Genomics*, 2009, 10(1): 234. [DOI](#)
- [22] Schwarz D, Robertson HM, Feder JL, Varala K, Hudson ME, Ragland GJ, Hahn DA, Berlocher SH. Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*. *BMC Genomics*, 2009, 10(1): 633. [DOI](#)
- [23] Wang XW, Luan JB, Li JM, Bao YY, Zhang CX, Liu SS. *De novo* characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*, 2010, 11(1): 400. [DOI](#)
- [24] Renaut S, Nolte AW, Bernatchez L. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol Ecol*, 2010, 19(S1): 115–131. [DOI](#)
- [25] Jeukens J, Renaut S, St-Cyr J, Nolte AW, Bernatchez L. The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Mol Ecol*, 2010, 19(24): 5389–5403. [DOI](#)
- [26] Li Y, Sun C, Luo HM, Li XW, Niu YY, Chen SL. Transcriptome characterization for *Salvia miltiorrhiza* using 454 GS FLX. *Acta Pharmaceutica Sin*, 2010, 45(4): 524–529. [DOI](#)
- [27] Zhang FJ, Guo HY, Zheng HJ, Zhou T, Zhou YJ, Wang SY, Fang RX, Qian W, Chen XY. Massively parallel pyrosequencing-based transcriptome analyses of small brown planthopper (*Laodelphax striatellus*), a vector insect transmitting rice stripe virus (RSV). *BMC Genomics*, 2010, 11(1): 303. [DOI](#)
- [28] Mittapalli O, Bai XD, Mamidala P, Rajarapu SP, Bonello P, Herms DA. Tissue-specific transcriptomics of the exotic invasive insect pest emerald ash borer (*Agrilus planipennis*). *PLoS ONE*, 2010, 5(10): e13708. [DOI](#)
- [29] Schmid J, Muller-Hagen D, Bekel T, Funk L, Stahl U, Sieber V, Meyer V. Transcriptome sequencing and comparative transcriptome analysis of the scleroglucan producer *Sclerotium rolfsii*. *BMC Genomics*, 2010, 11: 329. [DOI](#)
- [30] Salem M, Rexroad CE 3rd, Wang JN, Thorgaard GH, Yao JB. Characterization of the rainbow trout transcriptome using Sanger and 454-pyrosequencing approaches. *BMC Genomics*, 2010, 11(1): 564. [DOI](#)
- [31] Ferguson L, Lee SF, Chamberlain N, Nadeau N, Joron M, Baxter S, Wilkinson P, Papanicolaou A, Kumar S, Kee TJ, Clark R, Davidson C, Glithero R, Beasley H, Vogel H, Ffrench-Constant R, Jiggins C. Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. *Mol Ecol*, 2010, 19(Suppl. 1): 240–254. [DOI](#)
- [32] O'Neil ST, Dzurisin JD, Carmichael RD, Lobo NF, Emrich SJ, Hellmann JJ. Population-level transcriptome sequencing of nonmodel organisms *Erynnis propertius* and *Papilio zelicaon*. *BMC Genomics*, 2010, 11(1): 310. [DOI](#)
- [33] Wheat CW. Rapidly developing functional genomics in

- ecological model systems via 454 transcriptome sequencing. *Genetica*, 2010, 138(4): 433–451. [DOI](#)
- [34] Guo SG, Zheng Y, Joung JG, Liu SQ, Zhang ZH, Crasta OR, Sobral BW, Xu Y, Huang SW, Fei ZJ. Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genomics*, 2010, 11(1): 384. [DOI](#)
- [35] Choi JH, Kijimoto T, Snell-Rood E, Tae H, Yang Y, Moczek AP, Andrews J. Gene discovery in the horned beetle *Onthophagus taurus*. *BMC Genomics*, 2010, 11(1): 703. [DOI](#)
- [36] Carvalho RA, Azeredo-Espin AM, Torres TT. Deep sequencing of New World screw-worm transcripts to discover genes involved in insecticide resistance. *BMC Genomics*, 2010, 11(1): 695. [DOI](#)
- [37] Coppe A, Pujolar JM, Maes GE, Larsen PF, Hansen MM, Bernatchez L, Zane L, Bortoluzzi S. Sequencing, de novo annotation and analysis of the first *Anguilla anguilla* transcriptome: EelBase opens new perspectives for the study of the critically endangered European eel. *BMC Genomics*, 2010, 11(1): 635. [DOI](#)
- [38] Barrantes I, Glockner G, Meyer S, Marwan W. Transcriptomic changes arising during light-induced sporulation in *Physarum polycephalum*. *BMC Genomics*, 2010, 11(1): 115. [DOI](#)
- [39] Kunstner A, Wolf JB, Backstrom N, Whitney O, Balakrishnan CN, Day L, Edwards SV, Janes DE, Schlinger BA, Wilson RK, Jarvis ED, Warren WC, Ellegren H. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Mol Ecol*, 2010, 19(Suppl. 1): 266–276. [DOI](#)
- [40] Sun C, Li Y, Wu Q, Luo HM, Sun YZ, Song JY, Lui EM, Chen SL. *De novo* sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics*, 2010, 11(1): 262. [DOI](#)
- [41] Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics*, 2010, 11(1): 180. [DOI](#)
- [42] Wolf JB, Bayer T, Haubold B, Schilhabel M, Rosenstiel P, Tautz D. Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Mol Ecol*, 2010, 19(Suppl. 1): 162–175. [DOI](#)
- [43] Schwartz TS, Tae H, Yang Y, Mockaitis K, Van Hemert JL, Proulx SR, Choi JH, Bronikowski AM. A garter snake transcriptome: pyrosequencing, *de novo* assembly, and sex-specific differences. *BMC Genomics*, 2010, 11(1): 694. [DOI](#)
- [44] Cantacessi C, Mitreva M, Campbell BE, Hall RS, Young ND, Jex AR, Ranganathan S, Gasser RB. First transcriptomic analysis of the economically important parasitic nematode, *Trichostrongylus colubriformis*, using a next-generation sequencing approach. *Infect Genet Evol*, 2010, 10(8): 1199–1207. [DOI](#)
- [45] Bai XD, Zhang W, Orantes L, Jun TH, Mittapalli O, Mian MAR, Michel AP. Combining next-generation sequencing strategies for rapid molecular resource development from an invasive aphid species, *Aphis glycines*. *PLoS ONE*, 2010, 5(6): e11370. [DOI](#)
- [46] Yang HY, Hu LW, Hurek T, Reinhold-Hurek B. Global characterization of the root transcriptome of a wild species of rice, *Oryza longistaminata*, by deep sequencing. *BMC Genomics*, 2010, 11(1): 705. [DOI](#)
- [47] Riggins CW, Peng YH, Stewart CN Jr, Tranel PJ. Characterization of *de novo* transcriptome for waterhemp (*Amaranthus tuberculatus*) using GS-FLX 454 pyrosequencing and its application for studies of herbicide target-site genes. *Pest Manag Sci*, 2010, 66(10): 1042–1052. [DOI](#)
- [48] Sato S, Feltus FA, Iyer P, Tien M. The first genome-level transcriptome of the wood-degrading fungus *Phanerochaete chrysosporium* grown on red oak. *Curr Genet*, 2009, 55(3): 273–286. [DOI](#)
- [49] Abril J, Cebrià F, Rodríguez-Esteban G, Horn T, Fraguas S, Calvo B, Bartscherer K, Saló E. Smed454 dataset: unravelling the transcriptome of *Schmidtea mediterranea*. *BMC Genomics*, 2010, 11(1): 731. [DOI](#)
- [50] Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, Chen X, Li Y. *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genomics*, 2010, 11(1): 726. [DOI](#)
- [51] Schafleitner R, Tincopa LR, Palomino O, Rossel G, Robles RF, Alagon R, Rivera C, Quispe C, Rojas L, Pacheco JA, Solis J, Cerna D, Kim JY, Hou J, Simon R. A sweetpotato gene index established by *de novo* assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. *BMC Genomics*, 2010, 11(1): 604. [DOI](#)
- [52] Galindo J, Grahame JW, Butlin RK. An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *J Evol Biol*, 2010, 23(9): 2004–2016. [DOI](#)
- [53] Clark M, Thorne M, Vieira F, Cardoso J, Power D, Peck L. Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing. *BMC Genomics*, 2010,

- 11(1): 362. [DOI](#)
- [54] Ueno S, Le Provost G, Léger V, Klopp C, Noirot C, Frigerio JM, Salin F, Salse J, Abrouk M, Murat F, Brendel O, Derory J, Abadie P, Léger P, Cabane C, Barré A, de Daruvar A, Couloux A, Wincker P, Revirion MP, Kremer A, Plomion C. Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics*, 2010, 11(1): 650. [DOI](#)
- [55] Wang H, Zhang HM, Wong YH, Voolstra C, Ravasi T, Vladimir BB, Qian PY. Rapid transcriptome and proteome profiling of a non-model marine invertebrate, *Bugula neritina*. *Proteomics*, 2010, 10(16): 2972–2981. [DOI](#)
- [56] Zeng SH, Xiao G, Guo J, Fei ZJ, Xu YQ, Roe BA, Wang Y. Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. *BMC Genomics*, 2010, 11(1): 94. [DOI](#)
- [57] Yamagishi J, Wakaguri H, Ueno A, Goo YK, Tolba M, Igarashi M, Nishikawa Y, Sugimoto C, Sugano S, Suzuki Y, Watanabe J, Xuan XN. High-resolution characterization of *Toxoplasma gondii* transcriptome with a massive parallel sequencing method. *DNA Res*, 2010, 17(4): 233–243. [DOI](#)
- [58] Crawford JE, Guelbeogo WM, Sanou A, Traoré A, Vernick KD, Sagnon N, Lazzaro BP. *De novo* transcriptome sequencing in *Anopheles funestus* using Illumina RNA-seq technology. *PLoS ONE*, 2010, 5(12): e14202. [DOI](#)
- [59] Bai XD, Rivera-Vega L, Mamidala P, Bonello P, Herms DA, Mittapalli O. Transcriptomic signatures of ash (*Fraxinus* spp.) phloem. *PLoS ONE*, 2011, 6(1): e16368. [DOI](#)
- [60] Bai XD, Mamidala P, Rajarapu SP, Jones SC, Mittapalli O. Transcriptomics of the bed bug (*Cimex lectularius*). *PLoS ONE*, 2011, 6(1): e16336. [DOI](#)
- [61] Karatolos N, Pauchet Y, Wilkinson P, Chauhan R, Denholm I, Gorman K, Nelson DR, Bass C, French-Constant RH, Williamson MS. Pyrosequencing the transcriptome of the greenhouse whitefly, *Trialeurodes vaporariorum* reveals multiple transcripts encoding insecticide targets and detoxifying enzymes. *BMC Genomics*, 2011, 12(1): 56. [DOI](#)
- [62] Wu Q, Sun C, Luo HM, Li Y, Niu YY, Sun YZ, Lu AP, Chen SL. Transcriptome analysis of *Taxus cuspidata* needles based on 454 pyrosequencing. *Planta Med*, 2011, 77(4): 394–400. [DOI](#)
- [63] Fu CH, Chen YW, Hsiao YY, Pan ZJ, Liu ZJ, Huang YM, Tsai WC, Chen HH. OrchidBase: a collection of sequences of the transcriptome derived from orchids. *Plant Cell Physiol*, 2011, 52(2): 238–243. [DOI](#)
- [64] Clark MS, Thorne MAS, Toullec JY, Meng Y, Peck LS, Moore S. Antarctic krill 454 pyrosequencing reveals chaperone and stress transcriptome. *PLoS ONE*, 2011, 6(1): e15919. [DOI](#)
- [65] Cabrera AR, Donohue KV, Khalil SMS, Scholl E, Opperman C, Sonenshine DE, Roe RM. New approach for the study of mite reproduction: The first transcriptome analysis of a mite, *Phytoseiulus persimilis* (Acari: Phytoseiidae). *J Insect Physiol*, 2011, 57(1): 52–61. [DOI](#)
- [66] Der JP, Barker MS, Wickett NJ, Depamphilis CW, Wolf PG. *De novo* characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics*, 2011, 12(1): 99. [DOI](#)
- [67] Logacheva MD, Kasianov AS, Vinogradov DV, Samigullin TH, Gelfand MS, Makeev VJ, Penin AA. *De novo* sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*). *BMC Genomics*, 2011, 12(1): 30. [DOI](#)
- [68] Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB. A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. *Toxicon*, 2011, 57(5): 657–671. [DOI](#)
- [69] Johansen SD, Karlsen BO, Furmanek T, Andreassen M, Jørgensen TE, Bizuayehu TT, Breines R, Emblem Å, Kettunen P, Luukko K. RNA deep sequencing of the Atlantic cod transcriptome. *Comp Biochem Physiol Part D Genomics Proteomics*, 2011, 6(1): 18–22. [DOI](#)
- [70] Swarbreck SM, Lindquist EA, Ackerly DD, Andersen GL. Analysis of leaf and root transcriptomes of soil-grown *Avena barbata* plants. *Plant Cell Physiol*, 2011, 52(2): 317–332. [DOI](#)
- [71] Gregory R, Darby AC, Irving H, Coulibaly MB, Hughes M, Koekemoer LL, Coetzee M, Ranson H, Hemingway J, Hall N, Wondji CS. A *de novo* expression profiling of *Anopheles funestus*, malaria vector in Africa, using 454 pyrosequencing. *PLoS ONE*, 2011, 6(2): e17418. [DOI](#)
- [72] Blanca J, Canizares J, Roig C, Ziarsolo P, Nuez F, Pico B. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics*, 2011, 12(1): 104. [DOI](#)
- [73] Garg R, Patel RK, Tyagi AK, Jain M. *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res*, 2011, 18(1): 53–63. [DOI](#)
- [74] Sun XD, Zhou SM, Meng FL, Liu SQ. *De novo* assembly and characterization of the garlic (*Allium sativum*) bud transcriptome by Illumina sequencing. *Plant Cell Rep*, 2012, 31(10): 1823–1828. [DOI](#)

- [75] Gao XG, Han JB, Lu ZC, Li YF, He CB. Characterization of the spotted seal *Phoca largha* transcriptome using Illumina paired-end sequencing and development of SSR markers. *Comp Biochem Physiol Part D Genomics Proteomics*, 2012, 7(3): 277–284. [DOI](#)
- [76] Peterson MP, Whittaker DJ, Ambreth S, Sureshchandra S, Buechlein A, Podicheti R, Choi JH, Lai Z, Mockatis K, Colbourne J. *De novo* transcriptome sequencing in a songbird, the dark-eyed junco (*Junco hyemalis*): genomic tools for an ecological model system. *BMC Genomics*, 2012, 13(1): 305. [DOI](#)
- [77] He M, Wang Y, Hua WP, Zhang YZ, Wang ZZ. *De novo* sequencing of *Hypericum perforatum* transcriptome to identify potential genes involved in the biosynthesis of active metabolites. *PLoS ONE*, 2012, 7(7): e42081. [DOI](#)
- [78] Wang B, Ekblom R, Castoe TA, Jones EP, Kozma R, Bongcam-Rudloff E, Pollock DD, Höglund J. Transcriptome sequencing of black grouse (*Tetrao tetrix*) for immune gene discovery and microsatellite development. *Open Biol*, 2012, 2(4): 120054. [DOI](#)
- [79] Huang LL, Yang X, Sun P, Tong W, Hu S Q. The first Illumina-based *de novo* transcriptome sequencing and analysis of safflower flowers. *PLoS ONE*, 2012, 7(6): e38653. [DOI](#)
- [80] Gahlan P, Singh HR, Shankar R, Sharma N, Kumari A, Chawla V, Ahuja PS, Kumar S. *De novo* sequencing and characterization of *Picrorhiza kurroa* transcriptome at two temperatures showed major transcriptome adjustments. *BMC Genomics*, 2012, 13(1): 126. [DOI](#)
- [81] Nyberg KG, Conte MA, Kostyun JL, Forde A, Bely AE. Transcriptome characterization via 454 pyrosequencing of the annelid *Pristina leidyi*, an emerging model for studying the evolution of regeneration. *BMC Genomics*, 2012, 13(1): 287. [DOI](#)
- [82] Zhou YJ, Gao F, Liu R, Feng JC, Li HJ. *De novo* sequencing and analysis of root transcriptome using 454 pyrosequencing to discover putative genes associated with drought tolerance in *Ammopiptanthus mongolicus*. *BMC Genomics*, 2012, 13(1): 266. [DOI](#)
- [83] Yang F, Xu B, Zhao SJ, Li JJ, Yang YJ, Tang XH, Wang F, Peng MZ, Huang ZX. *De novo* sequencing and analysis of the termite mushroom (*Termitomyces albuminosus*) transcriptome to discover putative genes involved in bioactive component biosynthesis. *J Biosci Bioeng*, 2012, 114(2): 228–231. [DOI](#)
- [84] Van Belleghem SM, Roelofs D, Van Houdt J, Hendrickx F. *De novo* Transcriptome assembly and SNP discovery in the wing polymorphic salt marsh beetle *Pogonus chalceus* (Coleoptera, Carabidae). *PLoS ONE*, 2012, 7(8): e42605. [DOI](#)
- [85] Torales SL, Rivarola M, Pomponio MF, Fernandez P, Acuna CV, Marchelli P, Gonzalez S, Azpilicueta MM, Hopp HE, Gallo LA, Paniego NB, Marcucci Poltri SN. Transcriptome survey of Patagonian southern beech *Nothofagus nervosa* (= *N. alpina*): assembly, annotation and molecular marker discovery. *BMC Genomics*, 2012, 13(1): 291. [DOI](#)
- [86] Sadamoto H, Takahashi H, Okada T, Kenmoku H, Toyota M, Asakawa Y. *De novo* sequencing and transcriptome analysis of the central nervous system of mollusc *Lymnaea stagnalis* by deep RNA sequencing. *PLoS ONE*, 2012, 7(8): e42546. [DOI](#)
- [87] Hsu JC, Chien TY, Hu CC, Chen MJ, Wu WJ, Feng HT, Haymer DS, Chen CY. Discovery of genes related to insecticide resistance in *Bactrocera dorsalis* by functional genomic analysis of a *de novo* assembled transcriptome. *PLoS ONE*, 2012, 7(8): e40950. [DOI](#)
- [88] Tanase K, Nishitani C, Hirakawa H, Isobe S, Tabata S, Ohmiya A, Onozaki T. Transcriptome analysis of carnation (*Dianthus caryophyllus* L.) based on next-generation sequencing technology. *BMC Genomics*, 2012, 13(1): 292. [DOI](#)
- [89] Ji PF, Liu GM, Xu J, Wang XM, Li JT, Zhao ZX, Zhang XF, Zhang Y, Xu P, Sun XW. Characterization of common carp transcriptome: Sequencing, *de novo* assembly, annotation and comparative genomics. *PLoS ONE*, 2012, 7(4): e35152. [DOI](#)
- [90] Werner GD, Gemmell P, Grosser S, Hamer R, Shimeld SM. Analysis of a deep transcriptome from the mantle tissue of *Patella vulgata* Linnaeus (Mollusca: Gastropoda: Patellidae) reveals candidate biomineralising genes. *Mar Biotechnol (NY)*, 2013; 15(2): 230–243. [DOI](#)
- [91] Marra NJ, Eo SH, Hale MC, Waser PM, Dewoody JA. A priori and a posteriori approaches for finding genes of evolutionary interest in non-model species: Osmoregulatory genes in the kidney transcriptome of the desert rodent *Dipodomys spectabilis* (banner-tailed kangaroo rat). *Comp Biochem Physiol Part D Genomics Proteomics*, 2012, 7(4): 328–339. [DOI](#)
- [92] Ma K, Qiu GF, Feng JB, Li JL. Transcriptome analysis of the oriental river prawn, *Macrobrachium nipponense* using 454 pyrosequencing for discovery of genes and markers. *PLoS ONE*, 2012, 7(6): e39727. [DOI](#)
- [93] Huh JW, Kim YH, Park SJ, Kim DS, Lee SR, Kim KM, Jeong KJ, Kim JS, Song BS, Sim BW, Kim SU, Kim SH, Chang KT. Large-scale transcriptome sequencing and gene analyses in the crab-eating macaque (*Macaca fascicularis*) for biomedical research. *BMC Genomics*, 2012, 13(1): 163. [DOI](#)

- [94] Hao DC, Ma P, Mu J, Chen SL, Xiao PG, Peng Y, Huo L, Xu LJ, Sun C. *De novo* characterization of the root transcriptome of a traditional Chinese medicinal plant *Polygonum cuspidatum*. *Sci China Life Sci*, 2012, 55(5): 452–466. [DOI](#)
- [95] Blanca J, Esteras C, Ziarsoolo P, Perez D, Fernandez V, Collado C, Rodriguez R, Ballester A, Roig C, Canizares J, Pico B. Transcriptome sequencing for SNP discovery across *Cucumis melo*. *BMC Genomics*, 2012, 13(1): 280. [DOI](#)
- [96] Kaur S, Pembleton LW, Cogan NO, Savin KW, Leonforte T, Paull J, Materne M, Forster JW. Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics*, 2012, 13(1): 104. [DOI](#)
- [97] Gao ZX, Luo W, Liu H, Zeng C, Liu XL, Yi SK, Wang WM. Transcriptome analysis and SSR/SNP markers information of the blunt snout bream (*Megalobrama amblycephala*). *PLoS ONE*, 2012, 7(8): e42637. [DOI](#)
- [98] Xie FL, Burklew CE, Yang YF, Liu M, Xiao P, Zhang BH, Qiu DY. *De novo* sequencing and a comprehensive analysis of purple sweet potato (*Ipomoea batatas* L.) transcriptome. *Planta*, 2012, 236(1): 101–113. [DOI](#)
- [99] Pereiro P, Balseiro P, Romero A, Dios S, Forn-Cuni G, Fuste B, Planas JV, Beltran S, Novoa B, Figueras A. High-throughput sequence analysis of turbot (*Scophthalmus maximus*) transcriptome using 454-pyrosequencing for the discovery of antiviral immune genes. *PLoS ONE*, 2012, 7(5): e35369. [DOI](#)
- [100] Kumar S, Blaxter ML. Comparing *de novo* assemblers for 454 transcriptome data. *BMC Genomics*, 2010, 11(1): 571. [DOI](#)
- [101] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen ZT, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu PG, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005, 437(7057): 376–380. [DOI](#)
- [102] Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. In: Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB). 1999. [DOI](#)
- [103] Huang XQ, Madan A. CAP3: A DNA sequence assembly program. *Genome Res*, 1999, 9(9): 868–877. [DOI](#)
- [104] Swindell SR, Plasterer TN. SEQMAN. Contig assembly. *Methods Mol Biol*, 1997, 70: 75–89. [DOI](#)
- [105] Pertea G, Huang XQ, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, 2003, 19(5): 651–652. [DOI](#)
- [106] Miller RT, Christoffels AG, Gopalakrishnan C, Burke J, Ptitsyn AA, Broveak TR, Hide WA. A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. *Genome Res*, 1999, 9(11): 1143–1155. [DOI](#)
- [107] Zerbino DR, Birney E. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res*, 2008, 18(5): 821–829. [DOI](#)
- [108] Jackman SD, Birol I. Assembling genomes using short-read sequencing technology. *Genome Biol*, 2010, 11(1): 202. [DOI](#)
- [109] MacCallum I, Przybylski D, Gnerre S, Burton J, Shlyakhter I, Gnirke A, Malek J, McKernan K, Ranade S, Shea TP, Williams L, Young S, Nusbaum C, Jaffe DB. ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. *Genome Biol*, 2009, 10(10): R103. [DOI](#)
- [110] Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: Robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, 2012, 28(8): 1086–1092. [DOI](#)
- [111] Li RQ, Li YR, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics*, 2008, 24(5): 713–714. [DOI](#)
- [112] Surget-Groba Y, Montoya-Burgos JI. Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Res*, 2010, 20(10): 1432–1440. [DOI](#)
- [113] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*, 2011, 29(7): 644–652. [DOI](#)
- [114] Huang XQ, Wang JM, Aluru S, Yang SP, Hillier L. PCAP: A whole-genome assembly program. *Genome Res*, 2003, 13(9): 2164–2170. [DOI](#)
- [115] Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS. SNP discovery via 454 transcriptome sequencing. *Plant J*, 2007, 51(5): 910–918. [DOI](#)