

# 绿色杜氏藻转录组分析

朱帅旗, 龚一富, 杭雨晴, 刘浩, 王何瑜

宁波大学海洋学院, 宁波 315211

**摘要:** 为了深入了解绿色杜氏藻(*Dunaliella viridis*)基因信息及功能、耐盐相关通路(甘油脂代谢)及关键酶, 本文首次通过 Illumina HiSeq™ 2000 高通量测序技术对绿色杜氏藻转录组进行测序, 利用 Trinity 软件将数据组装形成转录本, 对所有转录本进行 COG(Clusters of Orthologous Groups)、GO(Gene Ontology)和 KEGG(Kyoto Encyclopedia of Genes and Genomes)分类和功能注释、Pathway 注释以及蛋白编码区(Opening reading fragment, ORF)的预测, 并对甘油脂代谢通路关键酶基因进行了分析。转录组测序共获得 81 593 个转录本, 其中 ORF 共有 77 117 条, 约占所有转录本的 94.50%。COG 分类结果表明, 16 569 条转录本被分为 24 个类别。GO 分类结果表明, 76 436 条转录本被注释。在所有注释分类中, 生物学过程转录本数量最多, 为 30 678 条, 占总转录本数的 40.14%。KEGG 分析结果表明, 317 个标准途径中包含 26 428 条转录本, 含转录本最多的类别是代谢, 为 9949 条(37.65%)。与代谢有关的途径为 131 条, 占有注释途径的 41.32%。在甘油脂代谢通路中仅发现 1 条关键酶转录本(二羟丙酮激酶), 该酶可能与绿色杜氏藻耐盐胁迫中甘油的合成有较大关系。本研究进一步完善了绿色杜氏藻的基因信息, 为绿色杜氏藻代谢途径研究奠定了坚实的基础。

**关键词:** 绿色杜氏藻; 转录组; 注释; 分类; 甘油

## Transcriptome analysis of *Dunaliella viridis*

Shuaiqi Zhu, Yifu Gong, Yuqing Hang, Hao Liu, Heyu Wang

School of Marine Sciences, Ningbo University, Ningbo 315211, China

**Abstract:** In order to understand the gene information, function, haloduric pathway (glycerolipid metabolism) and related key genes for *Dunaliella viridis*, we used Illumina HiSeq™ 2000 high-throughput sequencing technology to sequence its transcriptome. Trinity soft was used to assemble the data to form transcripts. Based on the Clusters of Orthologous Groups (COG), Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, we carried out functional annotation and classification, pathway annotation, and the opening reading fragment (ORF) sequence prediction of transcripts. The key genes in the glycerolipid metabolism were analyzed. The results suggested that 81 593 transcripts were found, and 77 117 ORF sequences were predicted, accounting for 94.50% of all transcripts. COG classification results showed that 16 569 transcripts were assigned to 24 categories. GO classification annotated 76 436 transcripts. The number of transcripts for biological processes was 30 678, accounting for 40.14%

收稿日期: 2015-01-14; 修回日期: 2015-03-25

基金项目: 浙江省科技厅重点科技创新团队项目(编号: 2012R10029-07, 2010R50029)和宁波科技攻关项目(编号: 2013C10018)资助

作者简介: 朱帅旗, 硕士研究生, 专业方向: 藻类生物化学与分子生物学。E-mail: ljlzj2009@163.com

通讯作者: 龚一富, 博士, 副教授, 研究方向: 植物生化与分子。E-mail: gongyifu@163.com

DOI: 10.16288/j.ycz.15-032

网络出版时间: 2015-5-22 17:15:39

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20150522.1715.001.html>

of all transcripts. KEGG pathway analysis showed that 26 428 transcripts were annotated to 317 pathways, and 131 pathways were related to metabolism, accounting for 41.32% of all annotated pathways. Only one transcript was annotated as coding the key enzyme dihydroxyacetone kinase involved in the glycerolipid pathway. This enzyme could be related to glycerol biosynthesis under salt stress. This study further improved the gene information and laid the foundation of metabolic pathway research for *Dunaliella viridis*.

**Keywords:** *Dunaliella viridis*; transcriptome; annotation; classification; glycerol

绿色杜氏藻(*Dunaliella viridis*)属绿藻门、团藻目、杜氏藻科<sup>[1]</sup>, 是一类存在于海洋和盐湖中的无细胞壁的单细胞真核绿藻, 对外界环境具有极大忍耐力, 可作为耐盐研究的模式生物<sup>[2,3]</sup>。目前, 对杜氏藻耐盐机理研究较多。蔡马等<sup>[4]</sup>研究了不同浓度 NaCl 对杜氏藻中依赖 NAD<sup>+</sup> 的 3-磷酸甘油脱氢酶基因表达的影响, 认为 NaCl 浓度与 3-磷酸甘油脱氢酶基因的表达水平呈负相关, 杜氏藻通过降低或提高甘油含量应对低渗和高渗的外界环境。吕文兵等<sup>[5]</sup>研究了不同盐浓度对杜氏藻叶绿体中编码光系统核心蛋白——D1 蛋白基因(*psbA*)表达的影响, 结果表明, 高浓度盐会导致 *psbA* 基因表达量降低。

转录组是包括 mRNA 和非编码 RNA 的所有 RNA 总和<sup>[6,7]</sup>, 是将 RNA 反转录成 cDNA, 获得不同基因 mRNA 在某个特定样本中的含量。转录组能够对基因组功能元件、剪切方式<sup>[8,9]</sup>、转录结构、转录后修饰、转录产物分类以及胁迫条件的差异表达进行研究<sup>[10]</sup>。测序是转录组研究的基础, Sanger 双脱氧测序是最早的 DNA 测序技术<sup>[11]</sup>。此后, 新一代测序技术相继产生, 如 454 技术(Roche 公司)、Solexa 技术(Illumina 公司)、SOLiD 技术(ABI 公司)、HeliScope 遗传分析系统(Helicos Biosciences 公司)<sup>[12]</sup>。高通量测序是一种深度测序技术, 先将 RNA 片段化, 再将片段化的 RNA 反转录成 cDNA 并构建文库, 经过质检合格后进行测序。高通量测序分为有参考基因组的组装和无参考基因组的组装(*de novo* assembling)。有参考基因组组装是将测序获得的序列与已知基因组比对, 无参考基因组组装则是通过序列拼接得到转录本进而从头组装。高通量测序具有灵敏度高、无需设计探针、单碱基差异分辨率高以及重复性好等优点<sup>[13~15]</sup>。目前, 植物<sup>[16]</sup>、微生物群落<sup>[17,18]</sup>的转录组研究均有报道。

杜氏藻的基因序列主要通过两种方法获得: 一

种是通过设计简并引物扩增基因核心序列, 再通过 RACE 技术扩增其全长 cDNA 序列; 另一种是构建 cDNA 文库。章丽等<sup>[19]</sup>利用 RACE 技术克隆了杜氏藻玉米黄质生物合成关键酶基因  $\beta$ -胡萝卜素羟化酶基因(*chyb*); 许芳等<sup>[20]</sup>利用 RACE 技术获得杜氏藻 FK506 结合蛋白基因(*fkbp*)cDNA 全长序列; Li 等<sup>[21]</sup>通过构建 cDNA 文库克隆了依靠钠离子的磷酸传递基因, 通过 Northern 杂交技术分析不同 NaCl 浓度对该基因表达的影响。

近年来, 藻类转录组研究主要集中在转录本分类、功能注释及相关通路研究。Rismani-Yazdi 等<sup>[22]</sup>利用第二代焦磷酸测序技术对杜氏藻 *Dunaliella tertiolecta* 转录组进行测序, 获得大量的转录本, 并将转录本进行分类和功能注释, 研究了三酰甘油和淀粉的生物合成与代谢。Yang 等<sup>[23]</sup>通过对绿藻 *Chlamydomonas moewusii* 进行转录组测序, 获得了大量与生物燃料相关的转录本。绿色杜氏藻是研究耐盐的模式微藻, 但对其基因信息及功能、耐盐相关通路(甘油脂代谢)中的关键酶基因还不清楚。本文利用 Illumina HiSeq<sup>TM</sup> 2000 高通量测序技术对绿色杜氏藻进行转录组测序, 获得大量的基因信息, 并对测序组装的转录本进行分类统计、功能注释和 Pathway 分析, 预测转录本 ORF 序列, 通过分析转录组数据研究绿色杜氏藻甘油脂代谢途径关键基因。

## 1 材料和方法

### 1.1 材料

绿色杜氏藻由宁波大学海洋生物工程重点实验室微藻室提供。培养条件参照文献<sup>[24]</sup>。

### 1.2 方法

#### 1.2.1 RNA 的提取和文库构建

取 80 mL 处于平台期末的藻液, 4、5000 r/min

离心 10 min, 弃上清,  $-20^{\circ}\text{C}$  保存备用。总 RNA 提取采用北京奥莱博生物技术有限公司的植物总 RNA 提取试剂盒。提取的总 RNA 经过 Alilent 2100 Bioanalyzer 检测, 符合转录组 RNA 检测标准(RNA 总量  $6\text{ }\mu\text{g}$ ,  $OD_{260/280}$  在 1.8~2.1 范围, rRNA Ratio (28S:18S) 1.5:1, RIN 7(RNA integrity number)), 并通过甲醛变性胶检测总 RNA 的完整性。

提取的总 RNA 用带有 Oligod(T)的 Breads 纯化。Agilent 2100 结果显示的文库大小在 300~400 bp, 由于在建库过程中引入了约 125 bp 的接头序列, 因此插入文库的实际长度约为 200 bp。利用反转录试剂盒(SuperScript II reverse transcriptase, TaKaRa)将 RNA 反转录成 cDNA。末端修复使 5'端磷酸化。在 3'端加 A, 将带有 3'-dTTP 端的双链 DNA 接头连到文库片段。通过 PCR 扩增, 使用 AMPure XP Breads 纯化富集接头产物。利用试剂盒(Qubit DNA brassay kit, Invitrogen)对 PCR 产物精确定量。PCR 产物切胶回收并通过 2100 Bioanalyzer High Sensitivity chip 确定切胶产物片段是否符合测序要求。通过 qPCR 对构建的文库进行摩尔浓度的绝对定量。质检合格的文库进行上机测序。

## 1.2.2 测序数据分析

1.2.2.1 原始数据的处理 测序结果的原始数据(Raw reads)中有些是低质量的, 因此需要对原始数据进行筛选。本研究使用 Q30(碱基准确率达到 99.9%)的质控标准。

1.2.2.2 序列拼接 利用 Trinity 和 Velvet 软件对

过滤后的 Clean reads 进行从头组装(以 N50 作为评估组装的参数)<sup>[25]</sup>。利用 Bowtie (<http://bowtie-bio.sourceforge.net/index.shtml>)将 100 bp 能重叠的 Reads 拼接<sup>[26]</sup>, 拼接形成的片段即为 Contig 组装片段。利用 paired-end reads 确定同一转录本的不同 Contig 及其距离, 将这些 Contig 连接, 获得两端不能延长的序列, 该序列即为转录本。

1.2.2.3 功能注释和 ORF 预测 将转录本与 NCBI 中非冗余蛋白数据库(Nr)(<http://www.nlm.nih.gov/>)、去冗余蛋白数据库(SwissProt)、蛋白质直系同源簇<sup>[27,28]</sup>、KEGG<sup>[29]</sup>比对。利用 BlastN 将转录本与 NCBI 非冗余核酸数据库(Nt)比对( $E\text{-value cutoffs } 1.00\text{E-}05$ ), 得到相似度最高的蛋白信息对其进行注释。利用 InterProScan 软件对蛋白功能进行注释, Blast2GO 将注释的蛋白映射到相应的 GO(基因本体)中<sup>[30]</sup>。利用 WEGO 软件将 GO 功能进行分类<sup>[31]</sup>。转录本与蛋白数据库比对排名最高的编码区序列即为其 ORF。

## 2 结果与分析

### 2.1 数据统计

利用 Trinity 和 Velvet 软件将短 Reads 进行从头组装形成转录本。结果表明(表 1), 通过 Trinity 软件组装的 N50 数值最大, 为 2256 bp。因此选择了 Trinity 的组装结果。本文共发现 81 593 条转录本, 转录本总长达到 95 822 901 bp, 平均长度为 1174.4 bp, 其中最长的转录本长度为 18 972 bp, 最短的转录本长度为 201 bp。

表 1 组装统计

软件	Kmer(bp)	数量	总长度(bp)	平均长度(bp)	最长(bp)	最短(bp)	N50(bp)
Velvet	31	58 385	72 594 854	1243.38	18 560	100	2153
	37	53 423	66 199 631	1239.16	18 560	100	2066
	41	51 367	62 426 468	1215.30	16 916	100	2012
	47	48 502	55 894 030	1152.41	16 683	100	1895
	51	47 330	51 737 854	1093.13	16 604	100	1794
	57	45 193	46 550 962	1030.05	16 662	101	1714
	61	43 583	42 395 668	972.76	15 831	100	1610
	67	40 310	36 321 440	901.05	15 831	100	1462
	71	37 801	32 255 027	853.29	15 831	102	1358
Trinity		81 593	95 822 901	1174.4	18 972	201	2256

注: Kmer(bp)表示组装时的参数; N50(bp)表示转录本 N50 的大小。

## 2.2 转录本长度分布

对绿色杜氏藻转录本长度进行统计(表 2),结果表明,长度在 200~300 bp 范围的转录本数目最多,为 20 886 条,占有所有转录本的 25.6%,其次是大于 2000 bp 和 300~400 bp 的转录本数目,分别为 15 344 条(占 18.8%)和 10 421 条(占 12.8%)。

## 2.3 转录本的功能注释

### 2.3.1 注释结果统计

分别将各转录本注释到 Nt、Nr、SwissProt、COG、KEGG、Interpro、GO 数据库中,对注释到每个库中的转录本数目进行统计,结果见表 3。Nr 数据库注释的转录本最多,为 47 552 条,占有所有转录本的 58.28%;KEGG 数据库注释的转录本数量为 38 247 条,占 46.88%。SwissProt 数据库注释的转录本为 28 999 条,占 35.54%。

### 2.3.2 转录本 COG 分类

将转录本进行 COG 分类(图 1),结果表明,

表 2 转录本组装长度统计

转录本长度范围(bp)	转录本数量	占总转录本比例(%)
200~300	20 886	25.6
300~400	10 421	12.8
400~500	6057	7.4
500~600	4013	4.9
600~700	3139	3.9
700~800	2597	3.1
800~900	2322	2.9
900~1000	2119	2.6
1000~1500	8701	10.6
1500~2000	5994	7.4
2000	15 344	18.8

表 3 转录本注释统计

数据库	注释转录本数目	注释转录本比例(%)	与数据库比对设置的 E 值	数据库版本
Nt	15 328	18.79	1.00E-05	201301
Nr	47 552	58.28	1.00E-05	201301
SwissProt	28 999	35.54	1.00E-10	201301
COG	21 958	26.91	1.00E-10	No version
KEGG	38 247	46.88	1.00E-10	Release58
Interpro	24 786	30.38	Interproscan 4.8	v36
GO	20 609	25.26		

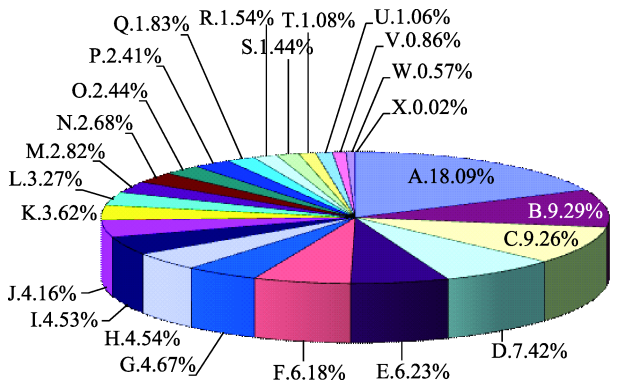


图 1 绿色杜氏藻转录本的 COG 分类

A: 一般功能预测; B: 翻译后修饰、蛋白翻转、分子伴侣; C: 翻译、核糖体结构和生物合成; D: 复制、重组和修复; E: 碳水化合物转运和代谢; F: 氨基酸转运和代谢; G: 功能未知; H: 能量产生和转化; I: 转录; J: 脂类转运和代谢; K: 辅酶转运和代谢; L: 无机离子转运和代谢; M: 细胞壁/膜/包膜生物合成; N: 信号转导机制; O: 次生代谢物生物合成、转运和代谢; P: 核苷酸转运和代谢; Q: 细胞骨架; R: 细胞周期控制、细胞分裂和染色体分区; S: 细胞内运输、分泌和膜泡运输; T: 染色质结构和动力学; U: 防御系统; V: 细胞迁移; W: RNA 加工和修饰 RNA; X: 核结构。

注释到 COG 数据库中 21 958 条转录本中有 16 569 条绿色杜氏藻转录本被分为 24 个大类。转录本最多的是一般功能预测类,有 2997 条序列(18.09%);其余依次为翻译后修饰、蛋白翻转、分子伴侣(1539, 9.29%)、翻译、核糖体结构和生物合成(1534, 9.26%)、复制、重组和修复(1229, 7.42%)。

### 2.3.3 转录本 GO 分类

绿色杜氏藻转录本的 GO 分类结果见图 2,共有 76 436 条转录本被注释分类。其中,生物学过程注释的转录本数量最多,为 30 678 条,占总数的 40.14%。而在生物学过程中又以代谢过程的转录本数量最多,为 11 136 条,占生物学过程的 36.30%,占有所有分类转录本的 14.57%。

分子功能中结合作用和催化活性所占的转录本数量最多,分别为 11 955 条和 11 835 条,占分子功能转录本的 15.69%和 15.48%。细胞组分是含转录本数量最少的,只有 18 977 条。在细胞组分中,细胞和细胞器拥有的转录本数最多,均为 6097 条。因此,与结合作用有关的转录本数量最多。

## 2.5 KEGG pathways 分析

利用 KEGG 数据库及 Pathway 对绿色杜氏藻转录本进行分类和分析,结果见表 4。26 428 条转录



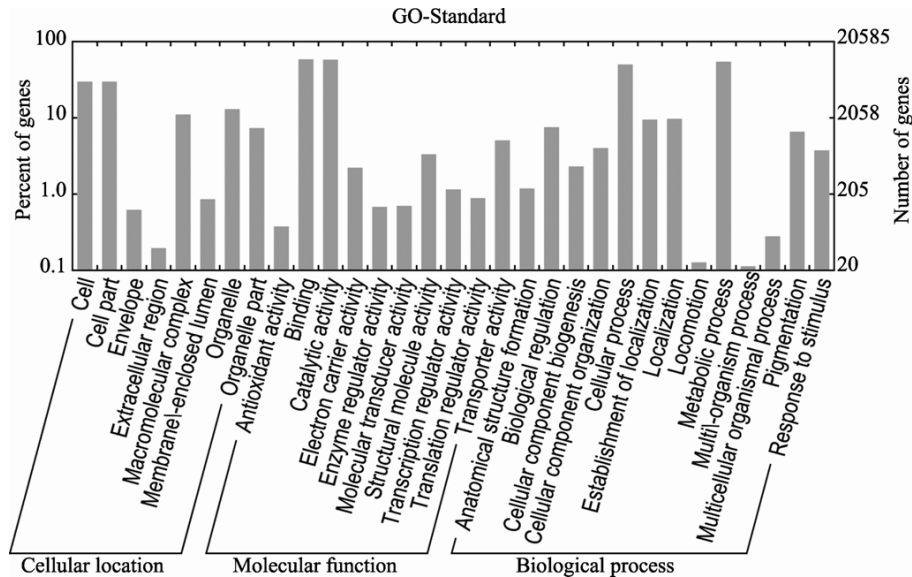


图 2 非冗余转录本的 GO 分类

本能够注释到 317 个 KEGG 标准 Pathway。转录本最多的是代谢和遗传信息程序, 分别为 9949 条和 9699 条, 分别占有所有注释转录本的 37.65% 和 36.70%。在代谢中又以能量代谢转录本最多, 为 2197 条, 占有所有注释转录本的 8.31%。而在遗传信息程序中, 折叠、筛选和降解以及复制和修复所含转录本最多, 分别为 2488 条(9.41%)和 2729 条(10.33%)。对 Pathway 的数据分析表明, 与代谢有关的 Pathway 最多, 为 131 条, 占有所有注释 Pathway 的 41.32%。在该类别中与辅因子和维生素代谢有关的 Pathway 数量最多, 为 28 条, 占注释 Pathway 的 8.83%。遗传信息程序和细胞程序所注释的 Pathway 数量较其他分类少, 分别为 25 条(占有所有注释 Pathway 7.89%)和 21 条(占有所有注释 Pathway 6.62%)。

## 2.6 转录本 ORF 预测

将转录本序列与数据库进行 BlastX 比对, 再附加 ESTscan 软件功能, 得到预测的 ORF 序列。结果表明(图 3), 共有 77 117 条预测到了 ORF 序列, 占有所有转录本数量的 94.50%, ORF 平均长度为 640.9 bp, 长度小于 300 bp 所占的比例最高。

## 2.7 绿色杜氏藻甘油脂代谢途径关键基因分析

杜氏藻是研究盐胁迫的模式生物。研究表明, 细胞膜通过改变细胞体积和形状调节渗透压, 该调

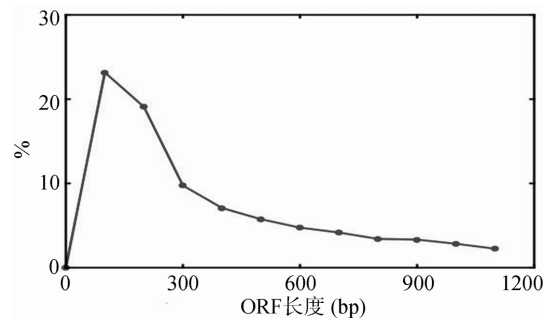


图 3 ORF 长度统计

节通过甘油的合成和降解来完成<sup>[32]</sup>。因此, 甘油是一种盐胁迫的主要诱导物。甘油脂代谢途径是合成甘油的重要通路。通过对该通路关键基因的分析将对杜氏藻盐胁迫机理的研究具有重要作用<sup>[33]</sup>。本研究对绿色杜氏藻转录组数据进行 Pathway 富集性分析, 结果表明(表 5), 共有 72 条转录本归类于甘油脂代谢途径, 占脂代谢转录本总数的 8.4%。其中, 三酰甘油脂肪酶(EC 3.1.1.3)、甘油激酶(EC 2.7.1.30)、半乳糖苷酶(EC 3.2.1.22)所占的转录本数最多, 都为 8 条。三酰甘油脂肪酶是合成三酰甘油的关键酶, 甘油激酶是甘油转变成甘油三磷酸的关键酶, 而半乳糖苷酶则是甘油脂代谢中合成半乳糖苷的关键酶。杜氏藻甘油的调节通过胞内淀粉合成与降解来实现<sup>[33]</sup>。淀粉水解产生的葡萄糖通过磷酸化, 并且在醛缩酶的催化下, 生成 3-磷酸甘油醛以及二羟丙酮磷酸。在

NAD<sup>+</sup>依赖性 3-磷酸甘油脱氢酶的作用下,将磷酸二羟丙酮还原为 3-磷酸甘油<sup>[34]</sup>。在甘油代谢过程中,二羟丙酮激酶是关键酶。本文通过分析发现,在该代谢通路中,存在二羟丙酮激酶,且仅含 1 条转录本。因此,绿色杜氏藻在耐盐胁迫中甘油的合成可能与二羟丙酮激酶有较大的关系。

3 讨 论

随着高通量测序技术的不断发展,对基因信息缺乏的物种进行转录组测序会越来越广泛。本研究首次采用 Illumina HiSeqTM 2000 高通量测序技术对绿色杜氏藻转录组进行测序,经拼接和组装共获得

表 4 KEGG 注释结果统计表

类别	途径	注释基因(26 428)	占总基因数的比例(%)	各类别占总基因数及比例	通路数(317)	占总通路数的比例(%)	各类别占总通路数及比例
代谢	氨基酸代谢	1537	5.82	9949, 37.65%	14	4.42	131, 41.32%
	其他次级代谢的生物合成	179	0.68		11	3.47	
	碳水化合物代谢	1919	7.26		15	4.73	
	能量代谢	2197	8.31		12	3.79	
	多糖生物合成和代谢	825	3.12		18	5.68	
	脂质代谢	850	3.22		16	5.05	
	辅因子和维生素代谢	1288	4.87		28	8.83	
	核苷酸代谢	782	2.96		2	0.63	
	外源性物质生物降解与代谢	372	1.41		15	4.73	
遗传信息程序	折叠,筛选和降解	2488	9.41	9699, 36.7%	11	3.47	25, 7.89%
	RNA 家族-非编码 RNAs	1	0		0	0	
	复制和修复	2729	10.33		9	2.84	
	转录	1695	6.41		5	1.58	
环境信息程序	翻译	2786	10.54	1200, 4.54%	8	2.52	44, 13.88%
	膜运输	289	1.09		5	1.58	
	信号转导	626	2.37		16	5.05	
	信号分子及相互作用	285	1.08		15	4.73	
细胞程序	细胞通讯	145	0.55	2221, 8.4%	4	1.26	21, 6.62%
	细胞增殖和衰亡	924	3.5		7	2.21	
	细胞运动	416	1.57		5	1.58	
	运输和分解代谢	736	2.78		5	1.58	
有机体系统	循环系统	55	0.21	1502, 5.68%	2	0.63	50, 15.77%
	发育	39	0.15		3	0.95	
	消化系统	220	0.83		9	2.84	
	内分泌系统	422	1.6		7	2.21	
	环境适应性	123	0.47		3	0.95	
	排泄系统	153	0.58		5	1.58	
	免疫系统	307	1.16		14	4.42	
	神经系统	121	0.46		4	1.26	
	感觉系统	62	0.23		3	0.95	
人类疾病	癌症	422	1.6	1857, 7.03%	15	4.73	46, 14.51%
	心血管疾病	60	0.23		4	1.26	
	免疫系统疾病	97	0.37		4	1.26	
	传染性疾病	504	1.91		16	5.05	
	新陈代谢疾病	104	0.39		2	0.63	
	神经退行性疾病	670	2.54		5	1.58	

表 5 绿色杜氏藻甘油脂代谢途径基因

KEGG 编号	酶号	酶基因	蛋白质定义	转录本数
k00901	2.7.1.107	<i>dgk</i>	Diacylglycerol kinase	4
k00655	2.3.1.51	<i>plsC</i>	1-acyl-sn-glycerol-3-phosphate acyltransferase	3
k19002	2.4.1.-	<i>sqd2</i>	Sulfoquinovosyl transferase	4
k00635	2.3.1.20	<i>dgat2</i>	Diacylglycerol O-acyltransferase 2	1
k00629	2.3.1.15	<i>gpat3/4</i>	Glycerol-3-phosphate O-acyltransferase 3/4	5
k00679	2.3.1.158	<i>dat</i>	Diacylglycerol acyltransferase	2
k01046	3.1.1.3	<i>tgl4</i>	Triacylglycerol lipase	8
k00005	1.1.1.6	<i>gldA</i>	Glycerol dehydrogenase	4
k00864	2.7.1.30	<i>glpK</i>	Glycerol kinase	8
k01189	3.2.1.22	<i>galA</i>	Alpha-galactosidase	8
k03715	2.4.1.46	<i>mgd</i>	1,2-diacylglycerol 3-beta-galactosyltransferase	6
k00863	2.7.1.29	<i>dak1/dak2</i>	Dihydroxyacetone kinase	1
k13519	2.3.1.-	<i>lpt1</i>	Lysophospholipid acyltransferase	4
k00631	2.3.1.15	<i>plsB</i>	Glycerol-3-phosphate O-acyltransferase	4
k00128	1.2.1.3	<i>aldh</i>	Aldehyde dehydrogenase (NAD+)	3
k06118	3.13.1.1	<i>sqd1</i>	UDP-sulfoquinovose synthase	2
k00630	2.3.1.15	<i>ats1</i>	Glycerol-3-phosphate O-acyltransferase	3
k00635	2.3.1.20	<i>dgat1</i>	Diacylglycerol O-acyltransferase 1	2

81 593 个转录本，平均长度为 1174.4 bp，对所有的转录本进行开放阅读框预测，共发现了 77 117 条 ORF，约占所有转录本的 94.50%，N50 达到了 2256 bp。李林等<sup>[35]</sup>通过对微藻 *Chloralla sorokinia* 进行转录组测序及分析共得到 49 885 条转录本，N50 为 1941 bp。绿色杜氏藻的 N50 比该微藻的 N50 高，说明本次转录组测序的质量很高，完全符合 Illumina 测序的预期结果。另一方面，绿色杜氏藻的转录本较该藻多 31 708 条，说明绿色杜氏藻可能在生命活动中功能更加复杂，进化更加完善。

通过与 Nt、Nr、SwissProt、COG、KEGG、Interpro、GO 数据库比对，发现 Nr 数据库所注释的转录本数量最多，占有所有转录本的 58.28%。通过 COG 分类发现，16 569 条转录本分布于 24 个大类。转录本最多的是一般功能预测，共 2997 条序列(18.09%)。将转录本进行 GO 分类，76 436 条转录本被注释分类。在所有注释分类中生物学过程注释的转录本数量最多，达到 30 678 条，占总数的 40.14%。其中，参与代谢有关的转录本数量最多。通过 KEGG Pathway 分析发现 26 428 条转录本注释到了 317 个标准 Pathway。其中，所含转录本最多的是代谢和遗传信

息程序这两个分类，转录本分别为 9949 条(37.65%)和 9699 条(36.70%)。Pathway 分析表明，与代谢有关的 Pathway 数量为 131 条，占有所有注释 Pathway 的 41.32%。GO 分类和 KEGG Pathway 分析结果一致，进一步说明绿色杜氏藻在生长过程中代谢活动非常旺盛。该结果与实验中所选用平台期末的藻体进行转录组测序可能有直接关系。因为在平台期末藻体达到最大生物量，其代谢活动增强，有利于获得大量的基因信息。Alkayal 等<sup>[36]</sup>通过构建杜氏藻 *D. salina* cDNA 文库和 EST 文库并与蛋白数据库比对得到了功能注释的转录本 1901 条，其中与氨基酸代谢有关的转录本占 35.7%，与能量代谢有关的转录本占 21.4%。不难看出，杜氏藻 *D. salina* 通过构建 EST 文库得到的转录本主要与代谢相关，而本研究通过 RNA-seq 测序分析也发现绿色杜氏藻中与代谢相关的转录本数量最高，不同的是在功能注释的转录本中，杜氏藻 *D. salina* 以蛋白质代谢为主，而绿色杜氏藻则以能量代谢为主。因此，推测绿色杜氏藻的能量代谢活动可能比 *D. salina* 强。

转录组测序不仅能够得到大量的基因信息，而且还能够对生物合成途径中的关键基因进行分析。陈

昊等<sup>[37]</sup>对油桐种子的 3 个不同时期的 RNA 进行转录组测序, 并分析了油脂的代谢合成途径。将代谢途径的转录本与 KEGG 数据库进行比对, 获得了关键酶同源蛋白。通过分析还发现了一系列重要的生物合成途径。李铁柱等<sup>[38]</sup>利用转录组测序及分析, 将杜仲的果实和叶片的转录本根据功能分为 25 类, 通过与 KEGG 数据库的比对, 将转录组数据分为 125 类, 其中包括 DNA 剪切、植物激素生物合成及脂类代谢, 萜类化合物、类固醇类化合物、苯丙氨酸的生物合成等。本研究获得了绿色杜氏藻的转录组数据, 数据分析获得了绿色杜氏藻耐盐作用相关的基因和途径(甘油酯代谢合成途径), 以期为绿色杜氏藻关键酶基因表达的分子机理研究提供帮助。

## 参考文献

- [1] 张俊杰, 郭连成, 崔志强, 隋丽英. 盐藻工厂化养殖技术与产品开发(一). 盐业与化工, 2009, 38(6): 35–38. [\[DOI\]](#)
- [2] Zhang N, Wang F, Meng X, Luo S, Li Q, Dong H, Xu Z, Song R. Molecular cloning and characterization of a trehalose-6-phosphate synthase/phosphatase from *Dunaliella viridis*. *Mol Bio Rep*, 2011, 38(4): 2241–2248. [\[DOI\]](#)
- [3] 陈琴, 张福, 姜润林. 3 种营养物质对极端嗜盐绿色杜氏藻生长的影响. 苏盐科技, 2006, (4): 21–23. [\[DOI\]](#)
- [4] 蔡马, 姜建国, 余士元. NaCl 浓度对杜氏盐藻中 1 种 NAD<sup>+</sup> 依赖的 3-磷酸甘油脱氢酶基因表达的影响. 华中农业大学学报, 2013, 32(5): 77–83. [\[DOI\]](#)
- [5] 吕文兵, 刘红涛, 薛乐勋. 不同盐浓度和光照强度对杜氏盐藻 *psbA* 基因表达的影响. 华中师范大学学报(自然科学版), 2010, 44(2): 288–292. [\[DOI\]](#)
- [6] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 2009, 10(1): 57–63. [\[DOI\]](#)
- [7] Costa V, Angelini C, De Feis I, Ciccociola A. Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnol*, 2010, 2010: Article ID 853916. [\[DOI\]](#)
- [8] Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*, 2011, 12(2): 87–98. [\[DOI\]](#)
- [9] Au KF, Jiang H, Lin L, Xing Y, Wong WH. Detection of splice junctions from paired-end RNA-Seq data by SpliceMap. *Nucl Acids Res*, 2010, 38(14): 4570–4578. [\[DOI\]](#)
- [10] 祁云霞, 刘永斌, 荣威恒. 转录组研究新技术: RNA-Seq 及其应用. 遗传, 2011, 33(11): 1191–1202. [\[DOI\]](#)
- [11] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*, 1977, 74(12): 5463–5467. [\[DOI\]](#)
- [12] 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, Weiner MP, Yu PG, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005, 437(7057): 376–380. [\[DOI\]](#)
- [13] Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, 2008, 453(7199): 1239–1243. [\[DOI\]](#)
- [14] Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, McKernan KJ, Grimmond SM. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods*, 2008, 5(7): 613–619. [\[DOI\]](#)
- [15] Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 2008, 320(5881): 1344–1349. [\[DOI\]](#)
- [16] 梁烨, 陈双燕, 刘公社. 新一代测序技术在植物转录组研究中的应用. 遗传, 2011, 33(12): 1317–1326. [\[DOI\]](#)
- [17] 刘万飞, 王西亮, 赵宇慧, 曾滢瑶, 耿佳宁, 胡松年. 基于第二代测序技术的细菌基因组与转录组研究策略简介. 微生物学通报, 2011, 38(11): 1705–1714. [\[DOI\]](#)
- [18] 金迪, 王加启, 赵圣国, 卜登攀, 孙鹏, 周凌云. 元转录组学在微生物群落研究中的应用. 微生物学通报, 2012, 39(2): 237–245. [\[DOI\]](#)
- [19] 章丽, 龚一富, 刘晓丹, 张文青. 盐生杜氏藻  $\beta$ -胡萝卜素羟化酶基因(*chyb*)的克隆及表达分析. 农业生物技术学报, 2013, 21(8): 920–930. [\[DOI\]](#)
- [20] 许芳, 石科, 李靓, 张楠楠, 薛乐勋. 杜氏盐藻 FKBP cDNA 的克隆及其在鞭毛解组装中的功能. 郑州大学学报(医学版), 2011, 46(6): 825–828. [\[DOI\]](#)
- [21] Li Q, Gao XS, Sun Y, Zhang QQ, Song RT, Xu ZK. Isolation and characterization of a sodium-dependent phosphate transporter gene in *Dunaliella viridis*. *Biochem Biophys Res Commun*, 2006, 340(1): 95–104. [\[DOI\]](#)
- [22] Rismani-Yazdi H, Haznedaroglu BZ, Bibby K, Peccia J.



- Transcriptome sequencing and annotation of the microalgae *Dunaliella tertiolecta*: pathway description and gene discovery for production of next-generation biofuels. *BMC Genomics*, 2011, 12: 148. [DOI]
- [23] Yang SH, Guarnieri MT, Smolinski S, Ghirardi M, Pienkos PT. De novo transcriptomic analysis of hydrogen production in the green alga *Chlamydomonas moewusii* through RNA-Seq. *Biotechnol Biofuels*, 2013, 6(1): 118. [DOI]
- [24] 王冬梅, 王玮蔚, 孙雪, 徐年军. 一株杜氏藻的分子鉴定与耐盐特性. *植物生理学报*, 2014, 50(3): 315–323. [DOI]
- [25] 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]
- [26] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 2009, 10(3): R25. [DOI]
- [27] Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science*, 1997, 278(5338): 631–637. [DOI]
- [28] Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov S, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 2003, 4: 41. [DOI]
- [29] Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in KEGG. *Nucl Acids Res*, 2006, 34(Suppl.1): D354–D357. [DOI]
- [30] Zdobnov EM, Apweiler R. InterProScan—an integration platform for the signature recognition methods in InterPro. *Bioinformatics*, 2001, 17(9): 847–848. [DOI]
- [31] Ye J, Fang L, Zheng HK, Zhang Y, Chen J, Zhang ZJ, Wang J, Li ST, Li RQ, Bolund L, Wang J. WEGO: a web tool for plotting GO annotations. *Nucl Acids Res*, 2006, 34(Suppl. 2): W293–W297. [DOI]
- [32] Sadka A, Lers A, Zamir A, Avron M. A critical examination of the role of *de novo* protein synthesis in the osmotic adaptation of the halotolerant alga *Dunaliella*. *FEBS Lett*, 1989, 244(1): 93–98. [DOI]
- [33] 李红萍, 焦新之. 杜氏盐藻渗透调节过程中的甘油代谢途径. *植物生理学报*, 1994, 20(1): 91–99. [DOI]
- [34] 魏娟, 吕芳芳, 唐欣昀, 张海生, 曹媛媛, 赵良侠. 葡萄糖对杜氏盐藻 3-磷酸甘油脱氢酶活性的影响. *广西植物*, 2009, 29(4): 556–560. [DOI]
- [35] 李林, 王钦宏, 杨海麟, 王武. 微藻(*Chlorella sorokiniana*)的转录组分析: 油脂生物合成相关的途径解析和基因挖掘. *微生物学报*, 2014, 54(9): 1010–1021. [DOI]
- [36] Alkayal F, Albion RL, Tillett RL, Hathwaik LT, Lemos MS, Cushman JC. Expressed sequence tag (EST) profiling in hyper saline shocked *Dunaliella salina* reveals high expression of protein synthetic apparatus components. *Plant Sci*, 2010, 179(5): 437–449. [DOI]
- [37] 陈昊, 蒋桂雄, 龙洪旭, 谭晓风. 基于油桐种子 3 个不同发育时期转录组的油脂合成代谢途径分析. *遗传*, 2013, 35(12): 1403–1414. [DOI]
- [38] 李铁柱, 杜红岩, 刘慧敏, 乌云塔娜, 王淋, 叶生晶. 杜仲果实和叶片转录组数据组装及基因功能注释. *中南林业科技大学学报*, 2012, 32(11): 122–130. [DOI]

(责任编辑: 赵方庆)