

微生物学报 *Acta Microbiologica Sinica*
54(9):1010–1021; 4 September 2014
ISSN 0001–6209; CN 11–1995/Q
http://journals.im.ac.cn/actamicroen
doi: 10.13343/j.cnki.wsxb.2014.09.006

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

李林^{1,2}, 王钦宏², 杨海麟¹, 王武^{1*}

¹江南大学, 工业生物技术教育部重点实验室, 江苏 无锡 214122

²中国科学院天津工业生物技术研究所, 中国科学院系统微生物工程重点实验室, 天津 300308

摘要:【目的】为了构建和改造产油微藻 (*Chloralla sorokiniana*), 需要从基因组水平解析脂肪酸、三酰甘油、淀粉等生物大分子的合成和降解途径, 并分析和确定其中的重要基因。【方法】本研究选取了初始氮浓度分别为 KNO_3 : 8g/L 和 KNO_3 : 2 g/L 的培养条件培养微藻 *C. sorokiniana*, 培养至 84h 收集藻细胞进行 Illumina HiSeq2000 双端测序, 利用 Trinity 进行 *de novo* 拼接, 得到的转录本通过 Nr 数据库、UniProtKB/Swiss-Prot 数据库、COG 数据库进行功能注释及分类, 通过 KEGG 数据库进行相关代谢途径注释。最后利用 RSEM 计算每个转录本的 RPKM 值, 并对相关代谢途径中的转录本的表达差异进行了初步分析。【结果】 Illumina HiSeq2000 双端测序共获得 49M 100 bp 的 raw reads, *de novo* 拼接得到 49885 个转录本, 其长度范围在 300bp 到 14100bp 之间, N50 为 1941 bp。其中 26479 个转录本成功注释到功能, 2357 个转录本注释到了 EC 编号, 利用这些转录本注释到 207 条代谢途径。在此基础上构建了 *C. sorokiniana* 的脂肪酸、三酰甘油、淀粉等生物大分子的合成及降解途径。并初步确定了代谢途径中基因的上调及下调水平。【结论】通过 RNA-seq 分析实现了对非模式藻株 *C. sorokiniana* 的基因组解析, 在此基础上构建了脂肪酸、三酰甘油、淀粉的生物合成和降解途径及重要基因, 所构建的代谢途径与模式藻株莱氏衣藻是一致的, 并比较了相关途径中的基因的表达差异, 这些信息有助于将来对微藻进行遗传改造提高其产油能力。

关键词: 微藻, *Chlorella sorokiniana*, 三酰甘油, 淀粉, RNA-seq

中图分类号: Q935 **文章编号:** 0001-6209(2014)09-1010-12

由于传统化石燃料的日益枯竭及温室效应的加剧, 生物柴油越来越受到人们的重视。与第一、二代生物柴油相比, 以微藻为原料的生物柴油有明显优势: (1) 微藻相比陆生植物有更高的生长速率和光转化效率^[1]; (2) 微藻能够在盐碱地等荒地生长并

能够利用工业废水生长并积累油脂, 油脂进而通过转酯化反应转化为生物柴油^[2]; (3) 微藻能够固定 CO_2 并将其转化为油脂或淀粉, 从而缓解温室效应。

然而, 目前以微藻为原料生产生物柴油的成本还比较高。微藻中三酰甘油的生物合成途径也是参

基金项目: 国家“973 项目”(微藻能源规模化制备的科学基础, 2011CB200900); 高等学校博士学科点专项科研基金(海洋微藻油脂合成氮源调控机制及信号应答网络研究, 20100093120002)

* 通信作者。Tel/Fax: +86-22-84861951; E-mail: wangwu@jiangnan.edu.cn

作者简介: 李林(1987-), 山东邹平人, 硕士研究生, 研究方向为微藻生物柴油。E-mail: 0820LL@163.com

收稿日期: 2013-12-23; **修回日期:** 2014-03-07

考高等植物和细菌中的代谢途径^[3]; 除模式藻株外, 微藻的基因组信息也知之甚少^[4], 这就大大限制了对微藻进行代谢工程改造。转录组测序技术可以在没有基因组信息的情况下对基因组进行转录水平上的分析, 从而确定编码油脂合成相关酶的基因, 在此基础上解析相关的生物合成途径, 这将有助于对微藻进行代谢工程改造以提高油脂含量。微藻 (*Chloralla sorokiniana*) 是一种单细胞淡水微藻, 在氮胁迫培养条件下, *C. sorokiniana* 的油脂积累量可以达到氮充足培养条件下的 2.5 倍, 因此可以作为生产生物柴油的潜在藻株。

在本研究中, 我们首先测定了 *C. sorokiniana* 在不同初始氮浓度下生长量、油脂积累、葡萄糖的消耗情况; 然后分析了 *C. sorokiniana* 在不同初始氮浓度培养条件下的转录组确定了与脂肪酸、三酰甘油、淀粉的生物合成及降解途径相关的酶的基因, 并在此基础上构建了脂肪酸、三酰甘油、淀粉的生物合成及降解途径。我们基于转录组数据构建的油脂代谢途径与 *Chlamydomonas reinhardtii* 中的油脂代谢途径是一致的。

1 材料和方法

1.1 主要试剂和仪器

酶标仪 (美国 Molecular Devices), 离心机 (德国 Eppendorf), 电子天平 (上海梅特勒), 高效液相色谱 (美国 Agilent Technologies) 恒温振荡器 (美国精骐有限公司)。实验中所用药品购自国药集团化学试剂有限公司。

1.2 *C. sorokiniana* 的培养

C. sorokiniana (UTEX 1602) 来自美国德克萨斯大学藻种培养保藏中心, 在含 1% 葡萄糖的 Kuhl 固体斜面上 4℃ 保存。*C. sorokiniana* 的培养采用优化后的 Kuhl 培养基, 并以 40 g/L Glucose 作为碳源, 8 g/L KNO₃ 或 2g/L KNO₃ 作为氮源进行异养培养, 采用 100 mL/250mL 的装液量, 37℃、220 r/min 培养。

1.3 生物量、油脂含量和葡萄糖的消耗量的测定

C. sorokiniana 分别在不同初始氮浓度条件下培养, 每 12 h 取样, 测定 OD₆₈₀、葡萄糖的消耗量及油脂含量。油脂含量的测定采用优化后的尼罗红染色法^[5], 即将微藻的浓度稀释至 OD₆₈₀ 在 0.1 - 0.3 之

间, 取 1 mL 稀释后的藻液加入 3.33uL 的尼罗红溶液 (浓度为 7.8×10^{-4} mol/L, 溶剂为丙酮), 激发波长 486 nm, 发射波长 570 nm 测定油脂的荧光强度。葡萄糖浓度采用 HPLC 法测定。

1.4 测序样品的制备

C. sorokiniana 培养至 84 h 时, 4℃、2057 × g 离心 10 min 收集藻细胞, 立即放入液氮速冻中并快速转移至 -80℃ 保存。样品用干冰送至上海伯豪生物技术有限公司进行 Illumina HiSeq2000 双端测序。本研究中所测得序列信息已提交至 NCBI Sequence Read Achieve 数据库, Accession number: SRX354139。

1.5 数据预处理及 *De novo* 拼接

测序获得的 raw reads 在进行 *de novo* 拼接之前需进行预处理, 除去低质量的碱基、不确定的碱基 'N' 及不成对的 reads。利用 Python 脚本对原始数据进行预处理, 每个 read 经处理后仅保留质量值 20 以上且不包含 'N' 的最长连续片段并保证片段长度不低于 25bp, 不成对的 reads 也会被去除。从而得到用于 *de novo* 拼接的 clean reads。利用拼接软件 Trinity (r2013-02-25)^[6] 对 clean reads 进行 *de novo* 拼接, 参数选用 Trinity 的省缺参数, Kmer = 25。

1.6 功能注释及代谢途径分析

将拼接得到的转录本利用 BLASTx (2.2.27 +)^[7] 算法分别比对到 NCBI 的 Nr 数据库、UniProtKB/Swiss-Prot 数据库、COG 数据库中, E-value 的阈值设为 10^{-5} , 从而预测转录本的功能并进行功能分类。得到注释的转录本会根据 GI 编号进而得到相应的 GO 编号。拼接得到的转录本同时提交到 KEGG^[8] 的 KAAS 服务器^[9], 选择 SBH 方法进行代谢途径分析。KAAS 将注释每一条提交的转录本序列, 对注释到的转录本分配相应的 EC 编号及所参与的代谢途径。

最后利用 RSEM^[10] 计算每个转录本在不同条件下的 FPKM 值, 利用 FPKM 值对不同培养条件下相关代谢途径中的基因的表达水平进行了比较。

2 结果和讨论

2.1 不同氮浓度培养条件下的生物量及油脂含量

C. sorokiniana 分别在氮充足 (KNO₃: 8g/L) 和氮胁迫条件 (KNO₃: 2g/L) 下进行异养培养。在氮

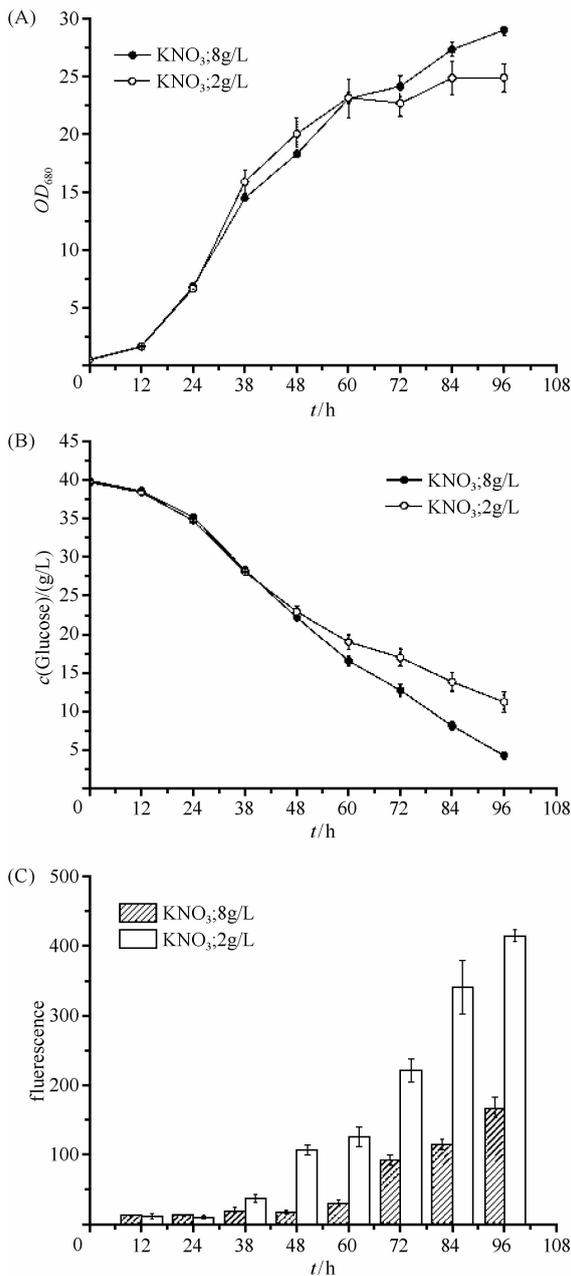


图 1. *C. sorokiniana* 在不同氮浓度培养条件下的生长、油脂含量及葡萄糖的消耗

Figure 1. The cell growth, lipid production and glucose consumption in *C. sorokiniana* at different conditions of nitrogen supply. (A) Cell growth under KNO₃: 8g/L and KNO₃: 2g/L conditions. (B) Lipid production under KNO₃: 8g/L and KNO₃: 2g/L conditions. (C) Glucose consumption under KNO₃: 8g/L and KNO₃: 2g/L conditions.

充足培养条件下, *C. sorokiniana* 的到达稳定生长期时生物量要高于在氮胁迫培养条件下能够到达的生物量(图 1-A), OD₆₈₀ 分别为 28.97 ± 0.45 , $24.87 \pm$

1.19。这与葡萄糖的消耗量是一致的(图 1-C), 培养至 96 h 时, 在氮充足培养条件下的葡萄糖的剩余量为 $4.32 \text{ g/L} \pm 0.51$, 利用率达到 89.2%; 而氮胁迫培养条件下的葡萄糖的剩余量为 $11.22 \text{ g/L} \pm 1.37$, 利用率为 71.7%。表明氮胁迫的培养条件会影响 *C. sorokiniana* 的生长速率。然而, *C. sorokiniana* 在氮胁迫培养条件下却能够积累更多的油脂(图 1-B), 96h 的荧光强度达到了 414.52 ± 8.18 , 而此时在氮充足培养条件下的荧光强度仅为 166.27 ± 13.15 , 二者相差 2.49 倍。表明氮胁迫的培养条件能够促使 *C. sorokiniana* 大量积累油脂。

2.2 转录组测序信息及 de novo 拼接

为了能够了解 *C. sorokiniana* 油脂合成相关酶的基因及相应的代谢途径, 选取了不同培养条件下培养至 84h 的藻液, 4℃ 离心收集藻细胞, 进行 Illumina HiSeq2000 双端测序, 共获得 49,269,563 的 raw reads。经过预处理后获得 44,058,382 用于 de novo 拼接的高质量的 clean reads, 占 raw reads 的 89.42% (表 1)。将所有的 clean reads 利用 de novo 拼接软件 Trinity (r2013-02-25)^[6] 进行 de novo 拼接, Kmer 选用默认值 25, 总共获得 49885 个转录本, 长度范围为 300 - 4100bp, N50 为 1941bp (表 1, 图 2-A)。

表 1. *C. sorokiniana* 的转录组信息

Table 1. Summary for the <i>C. sorokiniana</i> transcriptome		
Contig measurements	Raw reads	Clean reads
KNO ₃ : 2g/L	13,719,355	12,217,431
	13,650,626	12,217,431
KNO ₃ : 8g/L	10,972,576	9,811,760
	10,927,006	9,811,760
Total	49,269,563	44,058,382
total trinity transcripts		49,885
total trinity components		29,189
contig N50		1,941 bp

2.3 功能注释

拼接得到的所有转录本利用 BLASTx 算法在 NCBI 的 Nr 数据库中进行比对, E-value 的阈值选择 10^{-5} , 有 26480 个转录本(占有所有转录本的 53.1%) 能够和 Nr 库中的序列比对上(表 2)。在比对到的序列所属物种中 *Chlorella variabilis* (*C. variabilis*) 占了 81.77%, *Coccomyxa subellipsoidea* (*C. subellipsoidea*) 占了 7.08%, *Volvox carteri* (*V. carteri*) 占了 1.83% (图 2-B)。根据 BLASTx 的比对结果,

17240 个的转录本能够注释到 GO 编号, 同时 13786 个转录本能够与 COG 数据中的序列比对上 (表 2)。根据 GO 注释结果将转录本分成 43 个分类, 分别属于 Cellular component、Molecular function 和 Biological process 3 大类 (图 3)。同样, 根据 COG 数据库的比对结果, 能够比对的转录本按照功能分成 24 个分子家族 (图 4)。

表 2. 所有转录本的注释结果

Table 2. Annotation of all transcripts

Database	Number of annotated transcripts	Percentage of annotated transcripts/%
Nr	26479	53.1
UniProtKB/Swiss-Prot	16748	33.6
GO	17240	34.6
COG	13786	27.6
KEGG	5851	11.7

2.4 代谢途径注释

利用 KAAS 对拼接得到的所有转录本进行了代谢途径注释, 2357 个转录本通过比对分配到了 EC 编号, 并注释到 207 条不同的代谢途径, 包括脂肪酸的生物合成及降解途径、三酰甘油生物合成及降解途径、淀粉生物合成及降解途径等。对于一些重要代谢途径, 编码相关酶的转录本在 *C. sorokiniana* 的转录组数据中都存在 (表 3)。

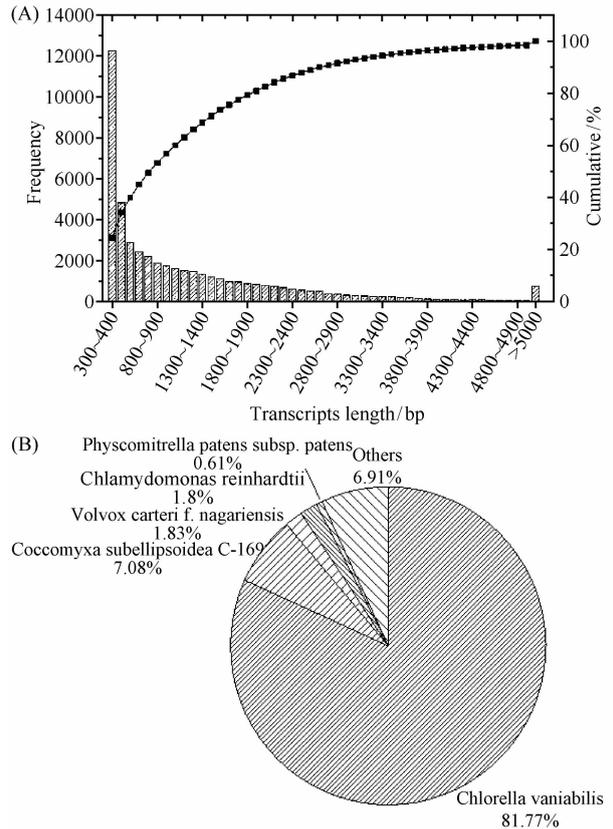


图 2. De novo 拼接及 BLASTx 注释结果

Figure 2. De novo assembly and BLASTx results. (A) Cumulative transcripts length frequency distribution of the *C. sorokiniana* transcriptome assembly; (B) Top-hit species distribution for BLASTx matches for the *C. sorokiniana* transcriptome.

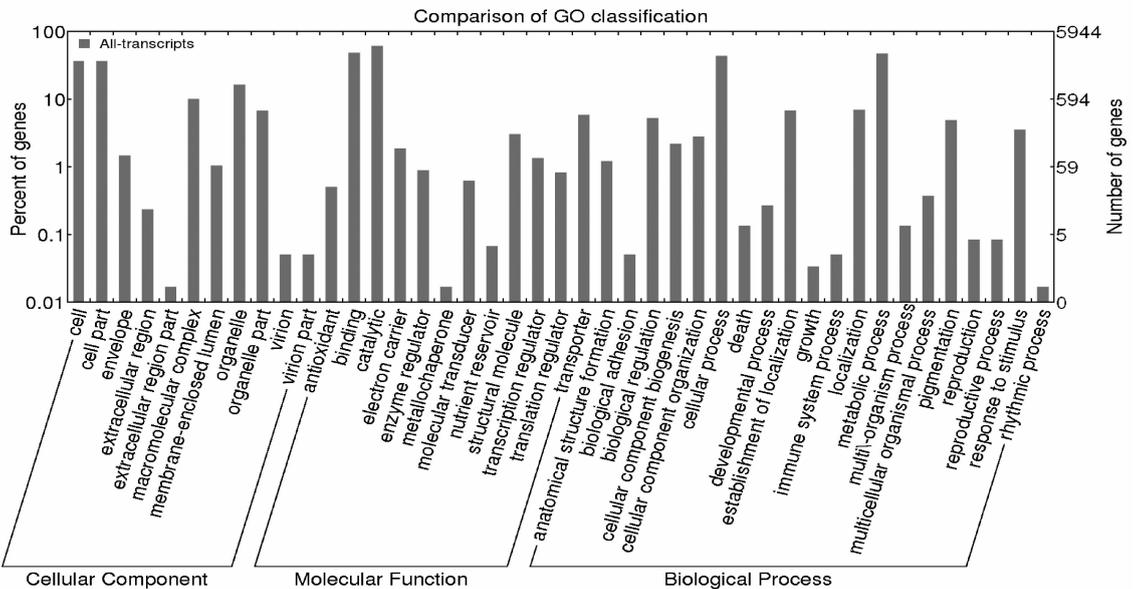


图 3. 转录本的 GO 注释

Figure 3. GO annotation of all transcripts. Best hits were mapped to get GO term, and 17240 transcripts were assigned to at least one GO term. Most transcripts were grouped into three major functional categories, namely cellular component, molecular function and biological process.

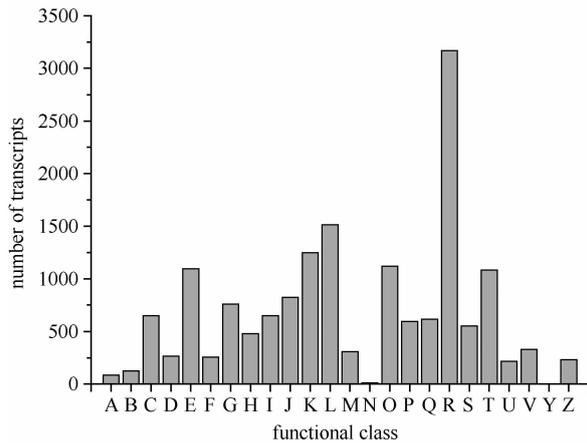


图 4. 转录本的 COG 功能分类

Figure 4. COG annotation of all transcripts. All transcripts were aligned to the COG database using BLASTx algorithm and can be classified functionally into at least 24 groups. (A) RNA processing and modification; (B) Chromatin structure and dynamics; (C) Energy production and conversion; (D) Cell cycle control, cell division, chromosome partitioning; (E) Amino acid transport and metabolism; (F) Nucleotide transport and metabolism; (G) Carbohydrate transport and metabolism; (H) Coenzyme transport and metabolism; (I) Lipid transport and metabolism; (J) Translation, ribosomal structure and biogenesis; (K) Transcription; (L) Replication, recombination and repair; (M) Cell wall/membrane/envelope biogenesis; (N) Cell motility; (O) Posttranslational modification, protein turnover, chaperones; (P) Inorganic ion transport and metabolism; (Q) Secondary metabolites biosynthesis, transport and catabolism; (R) General function prediction only; (S) Function unknown; (T) Signal transduction mechanisms; (U) Intracellular trafficking, secretion, and vesicular transport; (V) Defense mechanisms; (Y) Nuclear structure; (Z) Cytoskeleton.

表 3. *C. sorokiniana* 转录组注释到的重要代谢途径

Table 3. Essential metabolic pathways annotated in the

C. sorokiniana transcriptome

Pathway	Enzyme founded	Known enzymes
Photosynthetic carbon fixation (Calvin cycle)	12 *	13
Glycolysis/Gluconeogenesis	10	10
Pentose phosphate	5	5
Citrate cycle	10	10
Fatty acid biosynthesis	6	6
TAG biosynthesis	4	4
Starch biosynthesis	4	4

* the enzyme Phosphoketolase was not found in the *C. sorokiniana* transcriptome.

2.4.1 脂肪酸的生物合成途径: 基于微藻 *C. sorokiniana* 转录组的功能注释, 发现了 *C.*

sorokiniana 在脂肪酸的合成和降解中的编码相关酶的转录本, 平均每个酶有 2 个转录本。结合 KEGG 的比对结果重新构建了自由脂肪酸的生物合成途径 (图 5)。微藻中, 合成自由脂肪酸的反应首先发生在叶绿体中, 生成 C16:0 (棕榈酸) 和 C18:0-ACP (acyl carrier protein, 酰基转运蛋白), 然后再进一步作为前体物参与中性油脂 (主要是三酰甘油) 的生物合成。*C. sorokiniana* 的分析结果中, 合成自由脂肪酸的第一步是乙酰 CoA 羧化酶 (ACC, EC: 6.4.1.2), 催化乙酰 CoA 和碳酸盐 (HCO_3^-) 生成丙二酰 CoA 的反应, 然后丙二酰 CoA-ACP 酰基转移酶 (MAT, EC: 2.3.1.39) 催化丙二酰 CoA 转化成丙二酰-ACP。在随后的脂肪酸碳链延伸反应中丙二酰-ACP 将作为碳供体与酰基-ACP (或乙酰 CoA) 发生一系列由酮酰基-ACP 合酶 (KAS) 及其同工酶催化完成的缩合反应。第一个缩合反应由 β -酮酰基-ACP 合酶 III (KAS III, EC: 2.3.1.180) 催化并生成四碳产物, 然后 β -酮酰基-ACP 合酶 I (KAS I, EC: 2.3.1.41) 催化生成六碳到十六碳的不同碳链长度的产物, 最后 β -酮酰基-ACP 合酶 II (KAS II, EC: 2.3.1.179) 催化 C16:0-ACP 延长至 C18:0-ACP。在 *Eustigmatos cf. polyphem* (*E. cf. polyphem*)^[11]、*Dunaliella tertiolecta* (*D. tertiolecta*)^[12] 和 *Neochloris oleoabundans* (*N. oleoabundans*)^[13] 的转录组分析中都发现了编码 β -酮酰基-ACP 合酶 I 的转录本, 而在本研究中并未从 *C. sorokiniana* 的转录组数据中发现编码 β -酮酰基-ACP 合酶 I 的转录本。每一次缩合反应的初始产物均是 β -酮酰基-ACP, 然后依次再经过由 β -酮酰基-ACP 还原酶 (KAR, EC: 1.1.1.100) 催化发生还原反应、 β -羟酰脱水酶 (HAD, EC: 4.1.2.59) 催化发生脱水反应及烯酰-ACP 还原酶 (EAR, EC: 1.3.1.9) 催化发生还原反应后, 最终生成饱和脂肪酸并进入下一轮的碳链延伸。每一次循环脂肪酸会延伸 2 个碳原子, 如此进行 6 次或 7 次循环后分别生成 C16:0-ACP 和 C18:0-ACP (图 5)。

脂肪酸碳链的延伸会在以下情况终止: (1) 酰基-ACP 硫酯酶 (OAT, EC: 3.1.2.14) 催化酰基与 ACP 分离; (2) 酰基-ACP 硫酯酶 A (FatA, EC: 3.1.2.14, 3.1.2.-) 水解酰基-ACP 并释放脂肪酸; (3) 酰基转移酶直接将脂肪酸从 ACP 转移至甘油-3-磷酸或单酰甘油-3-磷酸。对于合成不饱和脂肪

酸, 酰基-ACP 去饱和酶 (ADD, EC: 1. 14. 19. 2) 将一个双键引入酰基-ACP 中^[12]。因此这些酶的活性大小决定了 *C. sorokiniana* 中脂肪酸的组成。另外, 在 *C. sorokiniana* 的转录组数据中还发现了其他一些去饱和酶: Δ^{12} (ω^6)-去饱和酶 (Δ^{12} D, EC: 1. 4. 19. -),

催化油酸 (18:1n-9) 形成亚油酸 (LA, 18:2 ω 6); Δ^{15} (ω^3)-去饱和酶 (Δ^{15} D, EC: 1. 4. 19. -), 进一步催化亚油酸形成 α -亚麻酸 (ALA, 18:3 ω 3) (图 5)。油酸和亚油酸是两种人体必需氨基酸, 也是多不饱和脂肪酸生物合成的重要前体物质。

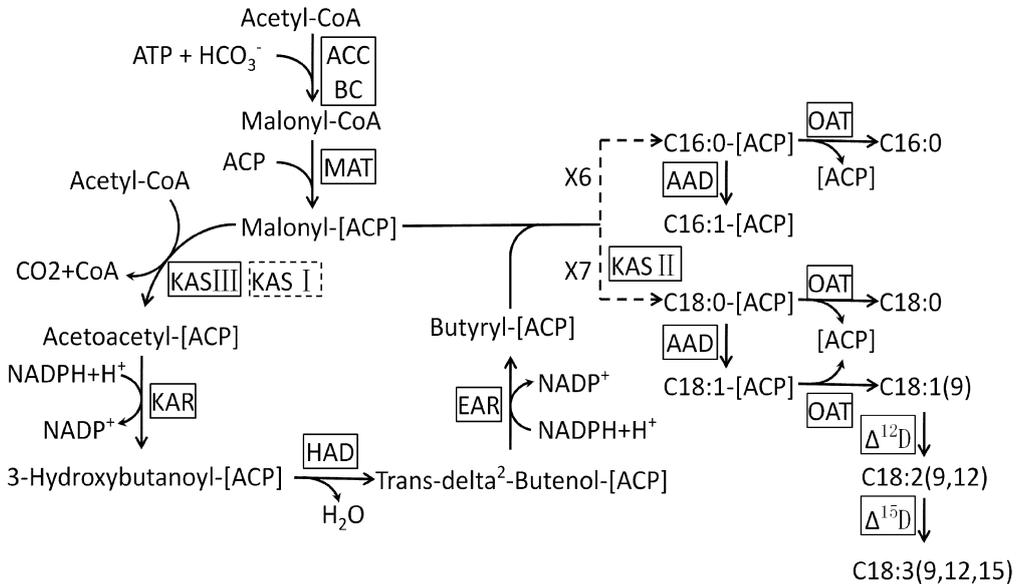


图 5. 基于 *C. sorokiniana* 转录组数据的 *de novo* 拼接及注释结果构建的脂肪酸生物合成途径

Figure 5. Fatty acid biosynthetic pathway reconstructed based on the *de novo* assembly and annotation of *C. sorokiniana* transcriptome. Identified and unidentified enzymes are shown in solid and dashed boxes, respectively, including: ACC, acetyl-CoA carboxylase (EC: 6. 4. 1. 2); BC, biotin carboxylase (EC: 6. 3. 4. 14); MAT, malonyl-CoA ACP transacylase (EC: 2. 3. 1. 39); KAS Beta-ketoacyl-ACP synthase (KAS II, EC: 2. 3. 1. 179; KAS III, EC: 2. 3. 1. 180); KAR, Beta-ketoacyl-ACP reductase (EC: 1. 1. 1. 100); HAD, beta-hydroxyacyl-ACP dehydrase (EC: 4. 2. 1. 59); EAR, Enoyl-ACP reductase (EC: 1. 3. 1. 10); AAD, acyl-ACP desaturase (EC: 1. 14. 19. 2); OAT, oleoyl-ACP thioesterase (EC: 3. 1. 2. 14); $\Delta 12$ D, $\Delta 12$ ($\omega 6$) desaturase (EC: 1. 14. 19. -); $\Delta 15$ D, $\Delta 15$ ($\omega 3$) desaturase (EC: 1. 14. 19. -). The fatty acid biosynthesis in *C. sorokiniana* produces saturated, C16:0 (palmitic acid) and C18:0 (stearic acid), and unsaturated fatty acids C18:1(9) (oleic acid), C18:2(9,12) (linoleic acid), C18:3(9,12,15) (α -linolenic acid).

C. sorokiniana 的转录组中也存在编码脂肪酸分解代谢途径 (β -氧化途径) 的相关酶的转录本。脂肪酸分解代谢途径中包含 4 个关键酶: 酰基 CoA 氧化酶 (AOx, EC: 1. 3. 3. 6)、烯酰 CoA 水合酶 (ECH, EC: 4. 2. 1. 17)、3-羟酰 CoA 脱氢酶 (HADH, EC: 1. 1. 1. 35) 和 乙酰 CoA 转移酶 (ACAT, EC: 2. 3. 1. 16)。脂肪酸在进入 β -氧化途径前先在酰基 CoA 合成酶 (ACSL, EC: 6. 2. 1. 3) 的催化下与 CoA 结合生成酰基 CoA, 然后经过氧化、水合、脱氢后以乙酰 CoA 的形式降解两个碳原子 (图 6)。生成的乙酰 CoA 可以参与柠檬酸循环为细胞生长提供能量或者参与三酰甘油的生物合成等途径。

基于 *C. sorokiniana* 的转录组数据构建的脂肪酸的合成及降解途径与莱氏衣藻的脂肪酸合成与降

解途径是一致的^[3]。可以利用这些转录本的序列信息对 *C. sorokiniana* 进行遗传改造促进脂肪酸的积累。最常见改造手段是过表达乙酰辅酶 A 羧化酶 (ACC, EC: 6. 4. 1. 2)^[14], 另外调控脂肪酸合成途径中的相关缩合酶^[15]及阻断 β -氧化途径^[16]也是促进脂肪酸积累的有效手段。

2.4.2 三酰甘油的生物合成及降解途径: *C. sorokiniana* 在氮胁迫条件下能够大量积累以三酰甘油 (TAG) 为主的油脂。三酰甘油可以作为生产生物柴油、塑料、化妆品及表面活性剂等原料, 用途广泛^[17]。然而由于缺乏微藻中三酰甘油的合成途径的相关信息, 限制了微藻的基因工程改造。本研究基于 *C. sorokiniana* 的转录组信息构建了 *C. sorokiniana* 三酰甘油的合成途径 (图 7)。甘油三磷

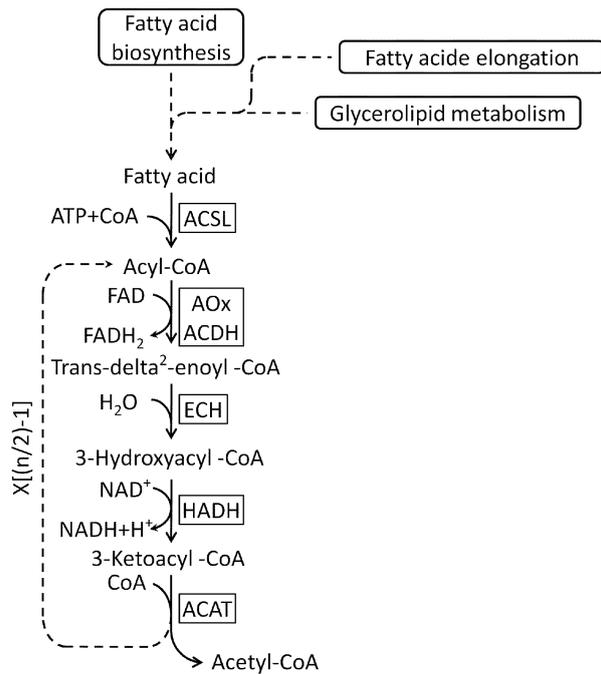


图 6. 基于 *C. sorokiniana* 转录组数据的 *de novo* 拼接及注释结果构建的自由脂肪酸分解代谢途径

Figure 6. Fatty acid catabolic pathways reconstructed based on the *de novo* assembly and annotation of *C. sorokiniana* transcriptome. Identified enzymes are shown in boxes and including: ACSL, acyl-CoA synthetase (EC: 6.2.1.3); AOX, acyl-CoA oxidase (EC: 1.3.3.6); ACDH, acyl-CoA dehydrogenase (EC: 1.3.8.7); ECH, enoyl-CoA hydratase (EC: 4.2.1.17); HADH, 3-hydroxyacyl-CoA dehydrogenase (EC: 1.1.1.35, 1.1.1.211); ACAT, acetyl-CoA C-acyltransferase (EC: 2.3.1.16, 2.3.1.9).

酸 (G-3-P) 和酰基 CoA 是三酰甘油合成的前体物。甘油三磷酸主要来自糖酵解途径或者通过甘油激酶 (GK, EC: 2.7.1.30) 催化甘油磷酸化生成, *C. sorokiniana* 的转录组数据中发现了 3 个编码甘油激酶的转录本。另外, 酰基 CoA 则通过脂肪酸和 CoA 的酯化反应而产生, 该反应可由酰基 CoA 合成酶 (ACSL, EC: 6.2.1.3) 催化完成。三酰甘油合成的前两步反应是酰基 CoA 相继将酰基转移至甘油三磷酸的 1、2 号碳原子上从而形成磷脂酸 (Phosphatidic acid, PA), 催化这两步反应的酶分别是甘油三磷酸酰基转移酶 (GPAT, EC: 2.3.1.15) 和 1-酰基-甘油三磷酸酰基转移酶 (AGPAT, EC: 2.3.1.51)。随即磷酸酯磷酸化酶 (PP, EC: 3.1.3.4) 催化磷酸基团从磷脂酸上移除并形成 1,2-二酰基甘油 (DAG), 1,2-二酰基甘油同时也是合成磷脂酰胆碱 (Phosphatidylcholine,

PC) 和磷脂酰乙醇胺 (phosphatidylethanolamine, PE) 的重要中间代谢物。最后 1,2-二酰基甘油在二酰基甘油酰基转移酶 (DGAT, EC: 2.3.1.20) 催化下在 3 号碳原子的位置上发生酰化反应最终生成三酰甘油。(图 7)

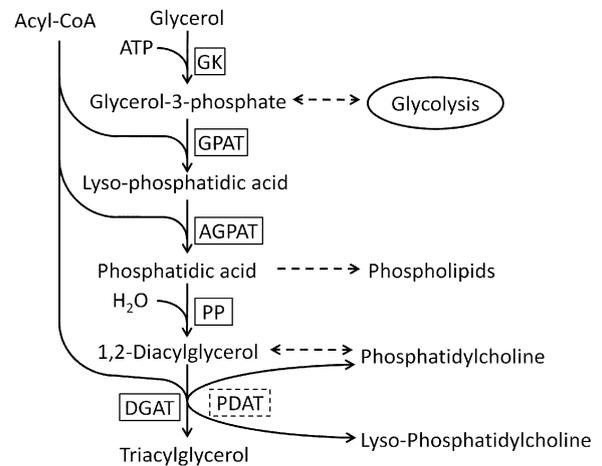


图 7. 基于 *C. sorokiniana* 转录组数据的 *de novo* 拼接及注释结果构建的三酰甘油生物合成途径

Figure 7. Triacylglycerol biosynthetic pathway reconstructed based on the *de novo* assembly and annotation of *C. sorokiniana* transcriptome. Identified and unidentified enzymes are shown in solid and dashed boxes, respectively, including: GK, glycerol kinase (EC: 2.7.1.30); GPAT, glycerol-3-phosphate O-acyltransferase (EC: 2.3.1.15); AGPAT, 1-acyl-sn-glycerol-3-phosphate acyltransferase (EC: 2.3.1.51); PP, phosphatidate phosphatase (EC: 3.1.3.4); DGAT, diacylglycerol O-acyltransferase (EC: 2.3.1.20); PDAT, phospholipid: diacylglycerol acyltransferase (EC: 2.3.1.158). Dashed arrow denote reaction (s) in which the enzymes are not shown.

三酰甘油的合成除了上述途径外, 在一些植物和酵母中还存在另外一条三酰甘油的合成途径, 即利用磷脂作为酰基供体合成三酰甘油^[18], 该反应由磷脂:二酰基甘油酰基转移酶 (PDAT, EC: 2.3.1.158) 催化完成, 在 *E. cf. polyphem*^[11] 和 *D. tertiolecta*^[12] 的转录组中都发现了编码磷脂:二酰基甘油酰基转移酶的转录本, 表明在微藻中可能存在这条合成途径, 而在 *C. sorokiniana* 的转录组中没有发现编码磷脂:二酰基甘油酰基转移酶的转录本, *C. sorokiniana* 中是否存在这条途径有待深入研究。

另外, 在 *C. sorokiniana* 的转录组数据中也发现了编码三酰甘油分解代谢相关酶的转录本。三

酰甘油的降解过程分为两步, 首先是酯键的水解反应使脂肪酸从甘油碳骨架上脱落, 催化这一步反应的是三酰甘油脂肪酶 (TAGL, EC: 3. 1. 1. 3), 在转录组数据中有四个转录本能够编码这个酶。随后, 水解得到的脂肪酸进一步通过 β -氧化途径被降解 (图 6) 或者进入其他代谢途径。利用基因工程手段抑制三酰甘油降解相关的酶可以作为提高三酰甘油积累量的一种手段^[19], 调控相关酰基转移酶的活性也可以提高油脂含量及组成^[19-21]。*C. sorokiniana* 转录组中确定的编码相关酶的转录本能够为 *C. sorokiniana* 的基因工程改造提供明确序列信息, 因此可以尝试不同的手段提高油脂含量及组成^[22-23]。

2. 4. 3 淀粉的生物合成及降解途径: 基于 KEGG 代谢途径的注释结果, 我们在 *C. sorokiniana* 的转录组中发现了大量编码淀粉合成及分解代谢途径中酶的转录本。并在此基础上构建了 *C. sorokiniana* 的淀粉生物合成途径 (图 8), 参与该途径中的酶包括: 磷酸葡萄糖变位酶 (PGM, EC: 5. 4. 2. 2), 催化葡萄糖-6-磷酸 (Glc-6-P) 转化为葡萄糖-1-磷酸 (Glu-1-P); ADP-葡萄糖焦磷酸化酶 (AGPase, EC: 2. 7. 7. 27), 催化葡萄糖-1-磷酸和 ATP 产生 ADP-葡萄糖并生成无机磷酸; 淀粉合酶 (SS, EC: 2. 4. 1. 21), 以 ADP-葡萄糖为底物, 在葡萄糖残基之间形成 α -1, 4-糖苷键并生成直连淀粉; α -1, 4-葡聚糖分支酶 (BE, EC: 2. 4. 1. 18), 在葡聚糖中形成 α -1, 6-糖苷键并最终生成淀粉。该淀粉合成途径与 *C. reinhardtii* 中的淀粉合成途径也是一致的。在此基础上可以对淀粉合成途径中的关键酶进行调控, 目前对 ADP-葡萄糖焦磷酸化酶的研究较为深入, 提高 ADP-葡萄糖焦磷酸化酶的活性可以提高淀粉的合成速率^[25]。

淀粉的降解有两条途径: 水解途径和磷酸解途径, 参与这两条途径的酶的转录本在 *C. sorokiniana* 转录组数据中都存在。参与水解途径的酶有 α -淀粉酶 (α -AMY, EC: 3. 2. 1. 1), 催化淀粉水解成低聚糖; 寡聚-1, 6-葡糖苷酶 (O1, 6G, EC: 3. 2. 1. 10), 进一步将低聚糖水解成 α -D-葡萄糖。水解得到的葡萄糖可以参与糖酵解途径或者经己糖激酶 (HXK, EC: 2. 7. 1. 1) 磷酸化后重新进入淀粉合成途径; 淀粉也可以经过 β -淀粉酶 (β -AMY, EC: 3. 2. 1. 2) 水解生成麦芽糖。磷酸解途径的重要酶是淀粉磷酸化

酶 (SPase, EC: 2. 4. 1. 1), 催化淀粉的 α -1, 4 糖苷键的降解得到葡萄糖并催化其与无机磷酸反应生成葡萄糖-1-磷酸及相应的葡聚糖 (图 8)。

2. 4. 4 不同代谢途径间的联系: 细胞中脂肪酸、三酰甘油、淀粉等生物大分子的合成及降解途径是紧密联系在一起的。淀粉分解代谢后可以通过糖酵解途径为其他物质的生物合成提供原料 (图 8)。*C. sorokiniana* 的转录组中有编码从丙酮酸到乙醇代谢途径的相关酶的转录本: 丙酮酸脱羧酶 (PDC, EC: 4. 1. 1. 1), 催化丙酮酸转化成乙醛并释放 CO_2 ; 乙醇脱氢酶 (ADH, EC: 1. 1. 1. 1), 催化乙醛和 $\text{NADH} + \text{H}^+$ 生成乙醇。在 *D. tertiolecta*^[12] 和 *N. oleoabundans*^[13] 的转录组数据中也有类似的结果。同样, 淀粉的降解得到的葡萄糖也可以作为脂肪酸合成的原料, 在 *C. sorokiniana* 转录组中有 6 个编码丙酮酸脱氢酶复合物 (PDHC, EC: 1. 2. 4. 1, 2. 3. 1. 2, 1. 8. 1. 4) 的转录本, 该复合物催化丙酮酸脱羧形成乙酰 CoA, 后者可以参与脂肪酸的生物合成。另外, 由于淀粉的合成和脂肪酸的合成是相对竞争的关系, 在 *C. reinhardtii* 中阻断淀粉合成途径使得油脂的积累量增加了 2 倍^[26]。除碳源外, *C. sorokiniana* 在合成油脂过程中还需要不断补充乙酰 CoA 和 NADPH^[27]。乙酰 CoA 主要由柠檬酸裂解反应产生, 催化该反应的酶是 ATP: 柠檬酸裂解酶 (ACL, EC: 2. 3. 3. 8), *C. sorokiniana* 的转录组中存在 2 个编码该酶的转录本。NADPH 则主要来自苹果酸的脱羧反应, 该反应由苹果酸脱羧酶 (MDH, EC: 1. 1. 1. 40) 以苹果酸为底物生成丙酮酸和 CO_2 , 同时将 NADP^+ 还原为 NADPH^[27], 有 7 个编码苹果酸脱羧酶的转录本。可以通过基因工程手段过表达 ATP: 柠檬酸裂解酶和苹果酸脱羧酶基因为脂肪酸合成提供充足的原料。

2. 5 油脂代谢相关途径中基因表达差异的初步分析

根据 FPKM 值我们比较了氮胁迫 (KNO_3 : 2g/L) 和氮充足培养条件 (KNO_3 : 8g/L) 下的油脂相关代谢途径中的基因的表达水平, 发现在参与脂肪酸、三酰甘油及淀粉的合成与降解途径的总共 37 个基因中上调 2 倍以上的有 10 个, 下调 2 倍以上的有 2 个, 剩余 25 个基因的变化在 2 倍以内, 表 4 中列出了油脂代谢相关途径中变化 2 倍以上的基因。表达量上调 2 倍以上的基因主要存在于油脂合成途径

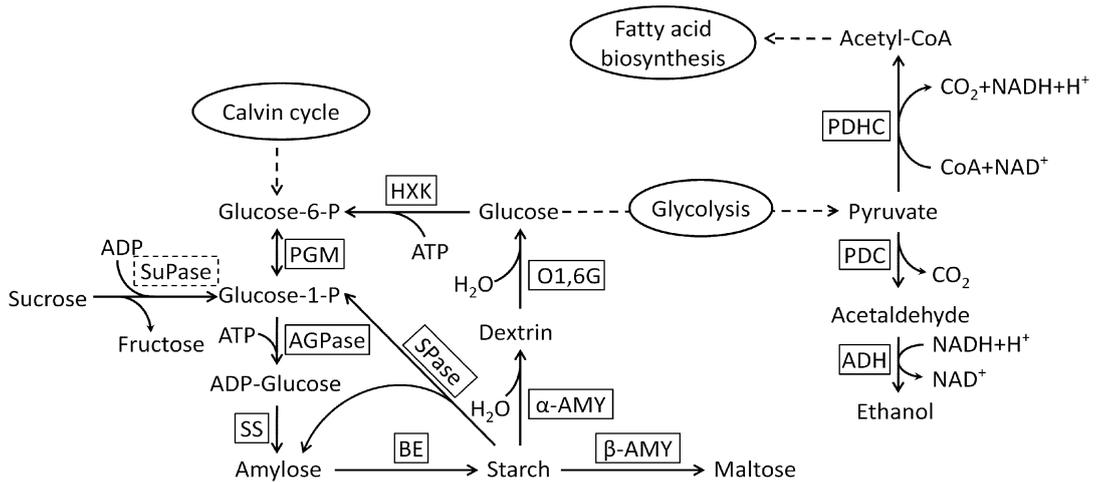


图 8. 基于 *C. sorokiniana* 转录组数据的 *de novo* 拼接及注释结果构建的淀粉生物合成、代谢及乙醇发酵途径

Figure 8. Starch biosynthesis and catabolism, and ethanol fermentation pathways reconstructed based on the *de novo* assembly and annotation of *C. sorokiniana* transcriptome. Identified and unidentified enzymes are shown in solid and dashed boxes, respectively. Enzymes involved in starch biosynthesis including: PGM, phosphoglucomutase (EC: 5. 4. 2. 2); AGPase, ADP-glucose pyrophosphorylase (EC: 2. 7. 7. 27); SS, starch synthase (EC: 2. 4. 1. 21); BE, 1,4- α -glucan branching enzyme (EC: 2. 4. 1. 18); and HXK, hexokinase (EC: 2. 7. 1. 1). Enzymes involved in starch catabolism include: β -AMY, β -amylase (EC: 3. 2. 1. 2); α -AMY, α -amylase (EC: 3. 2. 1. 1); O1,6G, oligo-1,6-glucosidase (EC: 3. 2. 1. 10); and SPase, starch phosphorylase (EC: 2. 4. 1. 1). Enzymes involved in ethanol fermentation via pyruvate include: PDC, pyruvate decarboxylase (EC: 4. 1. 1. 1); ADH, alcohol dehydrogenase (EC: 1. 1. 1. 1). Enzyme PDHC Pyruvate dehydrogenase complex (EC: 1. 2. 4. 1, 2. 3. 1. 12, 1. 8. 1. 4), transforms pyruvate into acetyl-CoA which may be then used in the lipid biosynthesis pathway.

中,包括生物素羧化酶(BC, EC: 6. 3. 4. 14)、 β -酮酰基-ACP 还原酶(KAR, EC: 1. 1. 1. 100)、 β -羟酰脱水酶(HAD, EC: 4. 1. 2. 59)、甘油三磷酸酰基转移酶(GPAT, EC: 2. 3. 1. 15)、1-酰基-甘油三磷酸酰基转移酶(AGPAT, EC: 2. 3. 1. 51)(表 4)。而在氮胁迫条件下乙酰 CoA 羧化酶(ACC, EC: 6. 4. 1. 2)的表达量下调(FPKM 值分别是 2. 79 和 4. 02),该现象在 *N. oleoabundans* 的转录组分析^[13]中也存在,但催化该步反应的生物素羧化酶(BC, EC: 6. 3. 4. 14)表达量上调了 2. 4 倍。 $\Delta^{12}(\omega^6)$ -去饱和酶($\Delta^{12}D$, EC: 1. 4. 19. -)(FPKM 值分别是 967. 05 和 1744. 16)和 $\Delta^{15}(\omega^3)$ -去饱和酶($\Delta^{15}D$, EC: 1. 4. 19. -)(FPKM 值分别是 135. 2 和 246. 62)的表达量下调预示着在氮胁迫培养条件下 *C. sorokiniana* 更倾向于积累饱和脂肪酸。参与脂肪酸的 β -氧化途径中的 6 个基因中有 4 个在氮胁迫条件下表达量下调,其中乙酰 CoA 转移酶(ACAT, EC: 2. 3. 1. 16)和 3-羟酰 CoA 脱氢酶(HADH, EC: 1. 1. 1. 35)下调 2 倍以上(表 4),酰基 CoA 合成酶(ACSL, EC: 6. 2. 1. 3)表达量下调了 1. 3 倍(FPKM 值分别是 53. 12 和 68. 96),酰基 CoA 脱氢酶(ACDH, EC: 1. 3. 8. 7)表达量下调了

1. 1 倍(FPKM 值分别是 0. 66 和 0. 73),烯酰 CoA 水合酶(ECH, EC: 4. 2. 1. 17)的表达量基本没有变化(FPKM 值分别是 19. 39 和 19. 09)表明氮胁迫条件下脂肪酸的降解途径受到抑制。

淀粉合成代谢途径中的大部分基因在氮胁迫条件下表达量上调,除己糖激酶(HXK, EC: 2. 7. 1. 1)表达量上调 2 倍以上(表 4)外其余基因的变化都在两倍以内,其中磷酸葡萄糖异构酶(PGM, EC: 5. 4. 2. 2)表达量上调 1. 2 倍(FPKM 值分别是 5. 84 和 4. 79),淀粉合成酶(SS, EC: 2. 4. 1. 21)表达量上调 1. 1 倍(FPKM 值分别是 72. 61 和 64. 69), α -1,4-葡聚糖分支酶(BE, EC: 2. 4. 1. 18)表达量上调 1. 1 倍(FPKM 值分别是 91. 98 和 80. 32),同时在淀粉降解途径中, β -淀粉酶(β -AMY, EC: 3. 2. 1. 2)和 α -淀粉酶(α -AMY, EC: 3. 2. 1. 2)的表达量没有明显变化,寡聚-1,6-葡糖苷酶(O1,6G, EC: 3. 2. 1. 10)表达量下调了 1. 6 倍(FPKM 值分别是 0. 38 和 0. 61)。在氮胁迫条件下丙酮酸脱氢酶复合物(PDHC, EC: 1. 2. 4. 1, 2. 3. 1. 2, 1. 8. 1. 4)的表达量上调 2 倍以上(表 4)表明有更多的丙酮酸转化为乙酰 CoA 为脂肪酸的合成提供充足原料。

表 4. 油脂相关代谢途径中的基因表达差异

Table 4. Different gene expression involving in lipid metabolism

Gene products	FPKM (2 g/L)	FPKM (8 g/L)	Different expression
BC, biotin carboxylase (EC:6.6.4.14)	24.81	10.17	+2.4*
KAR, Beta-ketoacyl-ACP reductase (EC:1.1.1.100)	4.92	0.45	+10.9
HAD, beta-hydroxyacyl-ACP dehydrase (EC:4.2.1.59)	39.65	14.72	+2.7
GPAT, glycerol-3-phosphate O-acyltransferase (EC:2.3.1.15)	8.07	2.53	+3.2
AGPAT, 1-acyl-sn-glycerol-3-phosphate acyltransferase (EC:2.3.1.51)	7.61	2.41	+3.2
HADH, 3-hydroxyacyl-CoA dehydrogenase (EC:1.1.1.35, 1.1.1.211)	24.54	55.69	-2.3
ACAT, acetyl-CoA C-acyltransferase (EC:2.3.1.16)	0.47	1.58	-3.4
HXK, hexokinase (EC:2.7.1.1)	4.04	1.1	+3.7
PDC, pyruvate decarboxylase (EC:4.1.1.1)	99.15	47.69	+2.1
ADH, alcohol dehydrogenase (EC:1.1.1.1)	59.25	22.8	+2.3
PDHC Pyruvate dehydrogenase complex (EC:2.3.1.12)	76.49	35.95	+2.1
PDHC Pyruvate dehydrogenase complex (EC:1.8.1.4)	92.11	36.31	+2.5

* + : upregulation; - : downregulation.

3 结论

在本研究中,我们采用 RNA-seq 对非模式藻株 *C. sorokiniana* 进行了转录组测序,通过对转录组数据的分析及挖掘,发现了许多编码与脂肪酸、三酰甘油及淀粉合成、降解相关的酶,并在此基础上构建了 *C. sorokiniana* 脂肪酸、三酰甘油、淀粉等生物大分子的合成与降解途径。根据转录组信息构建的代谢途径与模式藻株 *C. reinhardtii* 中的相关代谢途径是一致的。通过不同培养条件下转录组的初步分析确定了油脂相关代谢途径中基因的表达差异。由转录组测序构建的代谢途径及相关基因的差异表达信息有助于对 *C. sorokiniana* 及其他产油微藻进行遗传改造,提高油脂的积累量。

参考文献

[1] Melis A. Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Science*, 2009, 177(4): 272-280.

[2] Chisti Y. Biodiesel from microalgae. *Biotechnology Advances*, 2007, 25(3): 294-306.

[3] Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *The Plant Journal*, 2008, 54(4): 621-639.

[4] Guarnieri MT, Nag A, Smolinski SL, Darzins A, Seibert M, Pienkos PT. Examination of triacylglycerol biosynthetic pathways via *de novo* transcriptomic and

proteomic analyses in an unsequenced microalga. *PLoS One*, 2011, 6(10): e25851.

[5] Elsey D, Jameson D, Raleigh B, Cooney MJ. Fluorescent measurement of microalgal neutral lipids. *Journal of Microbiological Methods*, 2007, 68(3): 639-642.

[6] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Muceli 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. *Nature Biotechnology*, 2011, 29(7): 644-652.

[7] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden T. BLAST+: architecture and applications. *BMC Bioinformatics*, 2009, 10(1): 421.

[8] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 2000, 28(1): 27-30.

[9] Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, 2007, 35(suppl 2): W182-W185.

[10] Bo L, Colin N. RSEM: accurate transcripts quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 2011, 12:323.

[11] Wan L, Han J, Sang M, Li A, Wu H, Yin S, Zhang C. *De novo* transcriptomic analysis of an oleaginous microalga: pathway description and gene discovery for production of next-generation biofuels. *PLoS One*, 2012, 7(4): e35142.

- [12] Rismani-Yazdi H, Haznedaroglu B, 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(1): 148.
- [13] Rismani-Yazdi H, Haznedaroglu BZ, Hsin C, Peccia J. Transcriptomic analysis of the oleaginous microalga *Neochloris oleoabundans* reveals metabolic insights into triacylglyceride accumulation. *Biotechnology for Biofuels*, 2012, 5(1): 74.
- [14] Roessler PG, Bleibaum JL, Thompson GA, Ohlrogge JB. Characteristics of the gene that encodes acetyl-CoA carboxylase in the diatom *Cyclotella cryptica*. *Annals of the New York Academy of Sciences*, 1994, 721(1): 250-256.
- [15] Verwoert IGS, Linden K, Walsh M, Nijkamp HJ, Stuitje A. Modification of Brassica napus seed oil by expression of the *Escherichia coli* fabH gene, encoding 3-ketoacyl-acyl carrier protein synthase III. *Plant Molecular Biology*, 1995, 27(5): 875-886.
- [16] Radakovits R, Jinkerson RE, Darzins A, Posewitz MC. Genetic engineering of algae for enhanced biofuel production. *Eukaryotic Cell*, 2010, 9(4): 486-501.
- [17] Scott SA, Davey MP, Dennis JS, Horst I, Howe CJ, Lea-Smith DJ, Smith AG. Biodiesel from algae: challenges and prospects. *Current Opinion in Biotechnology*, 2010, 21(3): 277-286.
- [18] Dahlqvist A, Ståhl U, Lenman M, Banas A, Lee M, Sandager L, Ronne H, Stymne S. Phospholipid: diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proceedings of the National Academy of Sciences*, 2000, 97(12): 6487-6492.
- [19] Padham AK, Hopkins MT, Wang T-W, McNamara LM, Lo M, Richardson LGL, Smith MD, Taylor CA, Thompson JE. Characterization of a plastid triacylglycerol lipase from arabidopsis. *Plant Physiology*, 2007, 143(3): 1372-1384.
- [20] Kroon JTM, Wei W, Simon WJ, Slabas AR. Identification and functional expression of a type 2 acyl-CoA: diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals. *Phytochemistry*, 2006, 67(23): 2541-2549.
- [21] Wendel AA, Lewin TM, Coleman RA. Glycerol-3-phosphate acyltransferases: Rate limiting enzymes of triacylglycerol biosynthesis. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 2009, 1791(6): 501-506.
- [22] Zou J, Katavic V, Giblin EM, Barton DL, MacKenzie SL, Keller WA, Hu X, Taylor DC. Modification of seed oil content and acyl composition in the brassicaceae by expression of a yeast sn-2 acyltransferase gene. *The Plant Cell Online*, 1997, 9(6): 909-923.
- [23] Courchesne NMD, Parisien A, Wang B, Lan CQ. Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. *Journal of Biotechnology*, 2009, 141(1-2): 31-41.
- [24] Radakovits R, Jinkerson RE, Darzins A, Posewitz MC. Genetic engineering of algae for enhanced biofuel production. *Eukaryotic Cell*, 2010, 9(4): 486-501.
- [25] Smith AM. Prospects for increasing starch and sucrose yields for bioethanol production. *The Plant Journal*, 2008, 54(4): 546-558.
- [26] Wang ZT, Ullrich N, Joo S, Waffenschmidt S, Goodenough U. Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii*. *Eukaryotic Cell*, 2009, 8(12): 1856-1868.
- [27] Ratledge C. Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production. *Biochimie*, 2004, 86(11): 807-815.

De novo transcriptomic analysis of *Chlorella sorokiniana*: Pathway description and gene discovery for lipid production

Lin Li^{1,2}, Qinhong Wang², Hailin Yang¹, Wu Wang^{1*}

¹Key Laboratory of Industrial Biotechnology, Ministry of Education, Jiangnan University, Wuxi 214122, Jiangsu Province, China

²Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

Abstract: [**Objective**] The paucity of genomic information limits the metabolic engineering of non-model microalgae *Chlorella sorokiniana*. Our study aimed to elucidate the fatty acid, triacylglycerol and starch biosynthetic pathways in the microalgae *C. sorokiniana* based on *de novo* transcriptomic analysis. [**Methods**] We cultured *C. sorokiniana* with different nitrogen concentrations (KNO₃: 8g/L and 2g/L), then sequenced the transcriptome using Illumina HiSeq2000 platform. We used Trinity to *de novo* assemble the reads so as to obtain transcripts, aligned all the transcripts with Nr database, UniProtKB/Swiss-Prot database and COG database to annotate the function and classify using BLASTx algorithm, and assigned the transcript with metabolic pathway by aligning with KEGG database. Then we used RSEM to calculate FPKM value, and used it for preliminary analysis of different gene expression in the related pathways. [**Results**] Over 49M high quality raw reads were produced with the length of 100bp, We used Trinity to assemble these reads into 49885 transcripts with an N50 of 1941bp, ranging from 300bp to 14100bp. 26479 transcripts were annotated through BLASTx similarity search, 2357 transcripts were assigned with EC number, and 207 metabolic pathways were assigned in total. Based on these analyses, we reconstructed the fatty acids, triacylglycerol and starch biosynthetic pathways in *C. sorokiniana*. We also identified preliminarily different gene expression in the pathways. [**Conclusion**] Using RNA-seq technology, we reconstructed the metabolic pathways involving in the fatty acid, triacylglycerol and starch biosynthesis in non-model microalgae *C. sorokiniana* without genomic data, which is consistent with those in model microalgae *Chlamydomonas reinhardtii*, and compared the gene expression level under different conditions. These information is very useful for the metabolic engineering of *C. sorokiniana* and other microalgae to enhance the production of lipids.

Keywords: microalgae, *Chlorella sorokiniana*, triacylglycerol, starch, RNA-seq

(本文责编:王晋芳)

Supported by the Key Project of China National Programs for Fundamental Research and Development (2011CB200900) and by the Ph. D. Programs Foundation of Ministry of Education of China (20100093120002)

* Corresponding author. Tel/Fax: +86-22-84861951; E-mail: wangwu@jiangnan.edu.cn

Received: 23 December 2013/ Revised: 7 March 2014