

普洱茶不同发酵时期微生物群落宏转录组学研究

姜 姝 吕 杰* 李 灏

(北京化工大学 生命科学与技术学院, 北京 100029)

摘 要: 分离提取了普洱茶发酵前期和发酵中期茶叶样本中的微生物菌群全 mRNA, 并通过 Illumina 测序得到普洱茶发酵前期 17376 转录文本结果以及普洱茶发酵中期 14456 个转录文本结果。对测序结果的注释结果表明黑曲霉 (*Aspergillus niger*) 在两个阶段都占有绝对优势; 对两个样本差异表达部分 GO 功能和 KEGG PATHWAY 的对比分析表明, 适应渥堆发酵环境变化的微生物 (如黑曲霉等) 对普洱茶发酵过程起到了决定性的作用。

关键词: 普洱茶; 渥堆发酵; 宏转录组学; 微生物群落

中图分类号: Q789

引 言

普洱茶是以晒干的云南大叶种毛茶为原料, 经一系列复杂工序制成的一种茶。现代普洱茶制造中最为重要的工艺渥堆发酵, 是以微生物活动为中心的微生物固态发酵^[1]。在这一发酵过程中, 微生物对于普洱茶品质的形成起到了极其重要的作用。

从 20 世纪 70 年代后期开始, 人们对于普洱茶发酵过程中的微生物进行了研究, 研究方法主要是传统的平板培养法^[2], 后来又出现了显微观察及显微摄影法加以辅助^[3], 直到近几年才出现使用分子生物学手段对分离微生物进行鉴定, 如利用 PCR-DGGE 对群落结构变化进行分析^[4]。而现有研究仅反映普洱茶发酵过程中的微生物组成和变化, 对实际过程中的代谢反应只能推测。为了原位实时分析不同时期微生物的基因表达, 从而获得其代谢过程中酶的生产, 全面了解不同阶段微生物的作用, 必须引进新的技术。宏转录组学是后基因组时代最具代表性的新技术, 它可在样本较少时获得偏差较小的结果^[5], 目前宏转录组学只在海洋^[6-8]和土壤^[9]等天然环境的微生物菌群研究中有所应用, 而将其运用在工业生产环境中尚未见报道。

本文对普洱茶发酵初期和中期的两个样本进行了宏转录组学对比分析, 通过比较两者的差异来分析微生物群落在普洱茶发酵前中期的群落结构及代谢表达的变化及其在普洱茶品质形成方面的影响。

1 材料与方 法

1.1 样本制备

样本采集于云南省普洱市中国普洱茶研究院普洱茶加工示范车间于 2010-06-16 至 2010-07-12 渥堆发酵过程中的普洱茶, 渥堆总数为 5 t, 发酵过程为: 普洱茶良种场基地的鲜叶 (云南大叶种三、四级) → 杀青 → 揉捻 → 晒干 → 晒青毛茶 → 晒青毛茶拼配 → 增湿渥堆 (2010-06-16) → 翻堆 (2010-06-19) → 翻堆 (2010-06-22) → 翻堆 (2010-06-26) → 翻堆 (2010-07-05) → 通沟 (2010-07-12)。

选取渥堆中距边缘 1 m 以上的 4 个代表性的点, 翻堆时分上、中、下 3 层混合样, 每次取样 100 g 左右, 称出 10 g 茶样放入无菌匀浆袋中, 加入 90 mL 0.8% 吐温, 匀浆拍打 3 min, 最后加入 RNA 保存剂。样品 1 取自 2010-06-17 发酵开始第 2 天 (发酵初期), 样品 2 取自 2010-06-26 发酵开始第 11 天, 第 2 次翻堆 (发酵中期)。

1.2 RNA 提取及 cDNA 制备

将保存于 -70 ℃ 的 5 mL 样本匀浆液融化后离心去上清, 取沉淀 100 mg 进行总 RNA 提取, 再利用 poly A 从提取出的总 RNA 中选择分离 mRNA, 然后将 mRNA 片段化, 用随机引物进行反转录合成第一链 cDNA, 接着进行第二链合成得到双链 cDNA 片段并纯化。

收稿日期: 2012-04-23

基金项目: 云南省科技厅州、市党政“一把手”科技工程 (2008QA028)

第一作者: 女, 1985 年生, 硕士生

* 通讯联系人

E-mail: lvjie@mail.buct.edu.cn

其中各个步骤均使用试剂盒并遵照生产商指导使用步骤完成, RNA 提取用的是 QIAGEN Rneasy Mini Kit 试剂盒, mRNA 分离以及 cDNA 制备用的是 Illumina mRNA-seq sample prep. Kit 试剂盒, 而 cDNA 纯化则用的是 QIAGEN QIAquick PCR Purification Kit 试剂盒。

1.3 cDNA 测序

cDNA 测序采用 Illumina 高通量 DNA 测序技术。首先将目标序列修剪成大小合适的片段(约 800 bp), 在片段两个末端连接上独特的接头后进行 PCR 扩增, 再利用表面连接有一层单链引物的芯片, 将两端分别与其表面的引物碱基互补的单链 DNA 片段“固定”在芯片上, 从而形成桥式结构。再对每个片段进行单分子扩增 1000 倍, 获得单克隆 DNA 簇。

以得到的 DNA 簇为模板, 加入改造过的 DNA 聚合酶和带有 4 种荧光标记的 dNTP, 使目标单链 DNA 片段聚合为双链, 同时使用激光扫描反应板表面, 读取聚合上去的核苷酸种类, 统计收集的荧光信号结果, 可得到模板 DNA 片段的序列。

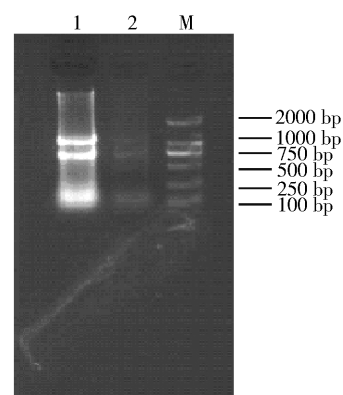
1.4 数据处理

将前面得到的序列结果进行预处理, 去除低质量序列。随机拼接余下序列并利用 NT 数据库进行注释, 对剩余未注释上的序列, 利用 NR 库注释, 获得全部转录本的注释 (blastx NCBI 提供的 NR database)。对注释结果聚类, 得到基因表达量。再利用 R 语言 Deseq 软件包, 对表达量进行统计分析, 将 $P < 0.01$ 显著级的结果认定为显著差异表达。在获得差异表达基因的基础上, 执行 GOfact 功能注释和 KEGG PATHWAY 分析。

2 结果与讨论

2.1 RNA 提取结果及质量分析

将样品 1 取自 2010-06-17 发酵开始第 2 天(发酵初期)的茶叶匀浆液提取到的总 RNA 进行凝胶电泳分析, 如图 1 所示, 结果其质量浓度为 1221.7 ng/ μ L, 体积 40 μ L, 其质量浓度纯度及总量均符合转录组测序要求。对样品 2 取自 2010-06-26 发酵开始第 11 天(第 2 次翻堆, 发酵中期)也执行同样的质量分析, 结果其质量浓度为 203.1 ng/ μ L, 体积 40 μ L, 其质量浓度较低, 但使用最大量进行上样还是可以满足转录组测序要求。



M—RNA 标记物; 1—样本 1; 2—样本 2

图 1 总 RNA 凝胶电泳质量分析结果

Fig. 1 Analysis of the quality of gel electrophoresis from total RNA

2.2 转录文本注释结果及表达量分析

经过测序和拼接共得到 294521 个转录文本, 其中 44258 个序列有注释结果, 去除其中注释结果来源不符合要求的 22544 个序列。两个样本的总表达序列为 9288 个, 总表达量 4094655, 其中样本 1 的表达量为 968312, 样本 2 的表达量为 3126343。样本 1 的特异表达序列为 8088 个, 表达量为 938841; 样本 2 的特异表达序列为 5168 个, 表达量为 301609。所以, 样本 1 的总序列为 17376 个, 总表达量为 1907153; 样本 2 的总注释结果为 14456 个, 总表达量为 3427952。转录文本及表达量分布结果如表 1 所示, 其中 t1 表示样本 1, t2 表示样本 2, A 代表样本 1 的特异表达部分, B 代表样本 2 的特异表达部分, C 代表两个样本的重叠表达部分。

表 1 转录文本及表达量分布

Table 1 Transcriptomes series and the related expression breakdown

类别	转录文本量		表达量	
	t1	t2	t1	t2
A	8088	0	938841	0
B	0	5168	0	301609
C	9288	9288	968312	3126343
总	17376	14456	1907153	3427952

从总体分布上来说, 虽然样本 1 序列总数多于样本 2 (样本 1 的序列为样本 2 的 120.2%), 但总表达量的情况却正好相反, 样本 1 的总表达量要远少于样本 2 仅为其一半左右 (55.6%)。这说明普洱茶发酵过程微生物的转录表达种类呈下降趋势, 但

是其代谢活动的频繁程度却是呈上升趋势的。

对比两个样本各自的分布组成,样本 1 无论是序列还是表达量,其特异表达部分(A 部分)与重叠表达部分(C 部分)所占比例都几乎相同,分别为 1: 1.15 和 1: 1.03,而样本 2 的情况则完全不同,B 部分所占比例要远小于 C 部分,其序列比为 1: 1.80,而表达量比更是高达 1: 10.4。这说明普洱茶发酵前期微生物转录表达有近一半到中期时停止表达,另外一半的表达活跃程度则大大增强(增加了 2.23 倍),而发酵中期才出现的新转录表达无论是种类还是表达量都少得多。

再对其中的 C 部分进行进一步的分析,将其分为 C1,序列同时存在于样本 1 与样本序 2 中,而存在于样本 1 中的数量多于样本 2,即 $t1 > t2$; C2,序列同时存在于样本 1 与样本 2 中,而数量相等,即 $t1 = t2$; C3,序列同时存在于样本 1 与样本 2 中,而存在于样本 1 中的数量少于样本 2,即 $t1 < t2$,结果如表 2 所示。

表 2 C 部分的转录文本及表达量分布

Table 2 Transcriptomes series and the related expression breakdown of part C

类别	转录文本量	表达量		
		t1	t2	t1 + t2
C1	1113	456234	59304	515538
C2	8	523	523	1046
C3	8167	511555	3066516	3578071
C1 + C2 + C3	9288	968312	3126343	4094655

其中 C2($t1 = t2$)是这 3 个部分里所占比例最小的,而且与其他两部分相比存在着数量级的差异,这是因为代谢表达具有一定的随机性,表达结果完全相同必然是非常少的。而 C1($t1 > t2$)与 C3($t1 < t2$)这两部分无论是序列量还是表达量都存在着非常值得研究探讨的巨大差异。C3 序列量是 C1 的 7.34 倍,而 C3 总表达量则是 C1 的 6.94 倍,表明差异表达部分的大多数的序列表达能力都显著增强。而 C3 表达量占 C 部分总表达量 87.4%,则显示出普洱茶发酵前、中期的微生物代谢差异表达的主要变化是代谢表达活动变得更加频繁。

2.3 转录文本来源分类及群落的构成

根据以往研究,普洱茶发酵过程中耐热真菌起到主要作用^[7],且本研究中使用 poly A 尾吸附技术对 mRNA 进行了分离和纯化,因此得到的转录文

本的结果主要是针对的是普洱茶发酵过程中真菌类微生物。

根据 NCBI 的注释结果,共鉴定 68 个属 145 种微生物,霉菌和酵母占绝大多数,其中种类最多的属为曲霉属(*Aspergillus*),多达 15 种。t1 包含 65 个菌属 124 种微生物,t2 拥有 62 个菌属 110 种微生物。但其中很多微生物仅有几个序列(甚至只有 1 个序列),这很有可能是测序步骤中,随机拼接带来的误差,即便不是误差它们的表达在总表达里所占比例也非常小,对发酵整体的影响可以忽略不计,所以在这里只讨论表达量占总表达量 1% 以上的微生物。

无论是样本 1 还是样本 2,表达量最大的微生物都是黑曲霉(*Aspergillus niger*),而且都各自占总表达量一半左右,拥有绝对优势。黑曲霉同时也是表达量变化最大的微生物,由 852078 增加至 1778633,说明它是普洱茶发酵过程中的优势微生物;而在总表达量比例上变化最大的则是烟曲霉(*Aspergillus fumigatus*),由 0.76% 上升到了 15.86%,在酵母中,汉逊德巴利酵母(*Debaryomyces hansenii*)所占比例也相对较高,且在发酵过程中表达量有所上升,如图 2~3 所示。

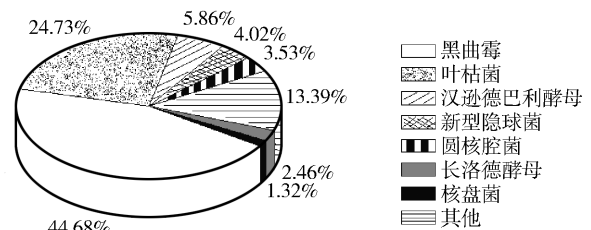


图 2 t1 微生物表达量分布图

Fig. 2 The layout of the microorganism species in t1

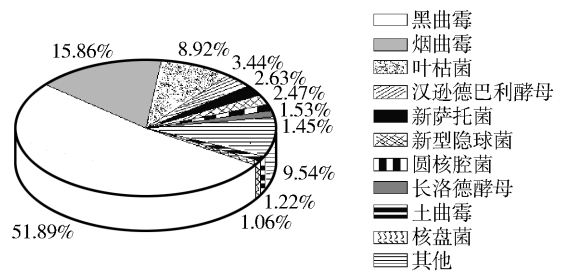


图 3 t2 微生物表达量分布图

Fig. 3 The layout of the microorganism species in t2

整体而言,样本 1 微生物种类多于样本 2,但样本 2 中也存在样本 1 所没有的微生物,这一现象与普洱茶的实际发酵过程相关。普洱茶发酵过程中具有高温湿热环境,部分微生物无法适应这一环境而停止转录表达进入休眠状态甚至死亡,而这一过程

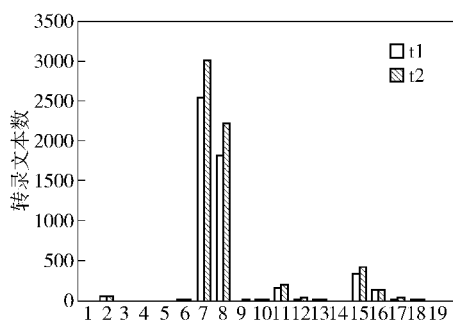
一直处于一个开放的环境中,且还因需要为微生物提供更多氧气等原因而进行翻堆,使得空气、水等外界因素为整个普洱茶发酵过程带来了更多种类的微生物。

2.4 宏转录组学对比分析

2.4.1 GOfact 功能分析注释

根据注释序列的基因 refseq 号在 <http://www.ncbi.nlm.nih.gov/protein> 网站中搜寻 Gene ID 或 SP 号,最终共得到 26144 个 Gene ID,将 Gene ID 整理成列表,并将列表提交至 <http://david.abcc.ncifcrf.gov/home.jsp>,利用得到的相应 GO 号进行 GO 功能分析。根据生物过程 (BP)、细胞组成 (CC) 和分子功能 (MF) 3 个部分分别对普洱茶发酵前、中期菌群转录表达 GOfact 功能进行对比分析。

从总体结构上看,t1 与 t2 GOfact 功能分析的 3 个部分结构差别较小。在 BP 部分中,细胞成分组成或生物转化 (GO:0071841) 和细胞代谢过程 (GO:0044237) 远高于其他类别,同时这两项也是两个样本差别最大之处,t2 均比 t1 增加约 20%。这说明无论在普洱茶发酵初期和发酵中期,微生物生长繁殖所需的基础代谢是微生物的主要代谢活动,而且整体呈上升趋势,如图 4 所示。

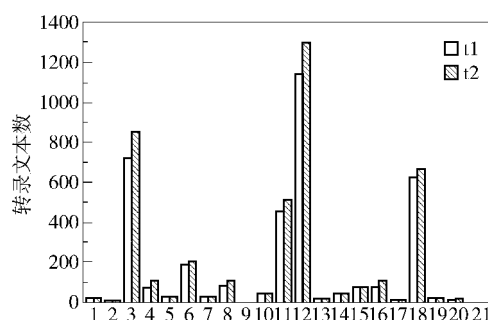


1—细胞凋亡; 2—细胞周期生化过程; 3—染色体分离; 4—细胞粘着; 5—细胞增殖; 6—膜的组成和生物转化; 7—细胞成分组成或生物转化; 8—细胞代谢过程; 9—细胞分裂; 10—细胞定位; 11—生物管理; 12—细胞信息传递; 13—稳定细胞内环境; 14—细胞发育过程; 15—规定的生物过程; 16—细胞应激反应; 17—调控生物质量; 18—细胞壁组织和生物转化; 19—细胞增长

图 4 转录本序列 BP 分析图

Fig. 4 Histogram for GOfact biological processes of transcriptomes in t1 and t2

而 CC 部分中,所占比例最大的是细胞核 (GO:0005730),其次是细胞质 (GO:0005737)、核糖体 (GO:0005840) 和线粒体 (GO:0005739),两个样本差别比较小,在于两个样本的绝对优势微生物都是黑曲霉,如图 5 所示。

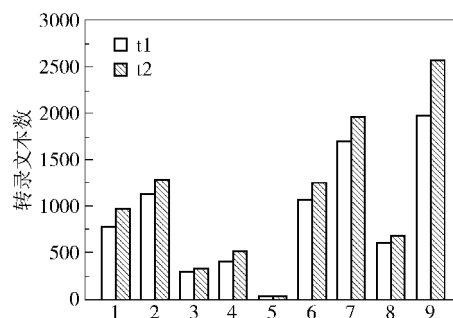


1—细胞壁; 2—细胞芽; 3—细胞质; 4—细胞骨架; 5—内膜系统; 6—内质网; 7—胞外物质; 8—高尔基体; 9—溶酶体; 10—微体 (包括过氧化物酶体和乙醛酸循环体); 11—线粒体; 12—细胞核; 13—细胞间隙; 14—液泡; 15—质膜; 16—蛋白酶体; 17—囊泡转运复合体; 18—核糖体; 19—信号识别颗粒; 20—纺锤体; 21—孢子

图 5 转录本序列 CC 分析图

Fig. 5 Histograms for CCfact biological processes of transcriptomes in t1 and t2

MF 部分中虽然两个样本的总体情况也趋于一致,但这一项分析中几乎所有分类 t2 都是高于 t1 的,无论是与 DNA 或者 RNA 复制相关的核苷结合 (GO:0000166)、核酸结合 (GO:0003676) 及 DNA 或 RNA 结合 (GO:0003677、GO:0003723),还是与能量代谢相关的水解酶活动 (GO:0016787)。结合其他两项 GOfact 分析,可以看出两个样本整体构成基本一致,t2 在各个方面都稍呈上升趋势,如图 6 所示。



1—蛋白质结合; 2—DNA 或 RNA 结合; 3—转录因子活动; 4—激酶活动; 5—受体结合或活动; 6—水解酶活动; 7—核酸结合; 8—结构分子活动; 9—核苷结合

图 6 转录本序列 MF 分析图

Fig. 6 Histogram for GOfact molecular functions of transcriptomes in t1 and t2

2.4.2 KEGG PASSWAY 分析

将结果分为 t1 特异表达、t2 特异表达及重叠表达 3 个部分进行 PASSWAY 分析,根据注释结果将表达量 500 作为特异表达的认定阈值 (即非特异表达量为 0,表达量大于等于 500 作为特异表达),而

对重叠表达则使用 R 语言 Deseq 软件包,对表达量进行统计分析, $P < 0.01$ 就认为是显著差异表达。

t1 特异表达结果无论的种类及数量均显著多于 t2,如图 7~8 所示,这表明发酵初期微生物处于一种相对天然的环境中,转录表达种类较多,而到发酵中期很多表达停止。而发酵中期新出现的表达种类则非常少,且集中在生物合成方面。 β -丙氨酸是合成泛酸的前体,而萜类化合物是构成某些植物香精、树脂、色素等的主要成分。显著差异表达中,核糖体占据最大比例,而嘌呤、嘧啶的代谢则表明微生物繁殖及转录表达发生了微妙变化。此外,多种多聚糖和不饱和脂肪酸的合成代谢可能与普洱茶特殊品质及功效成分形成相关,这与 t2 特异表达里的萜类化合物生物合成代谢都说明微生物在普洱茶发酵过程中占据重要地位,如图 9 所示。

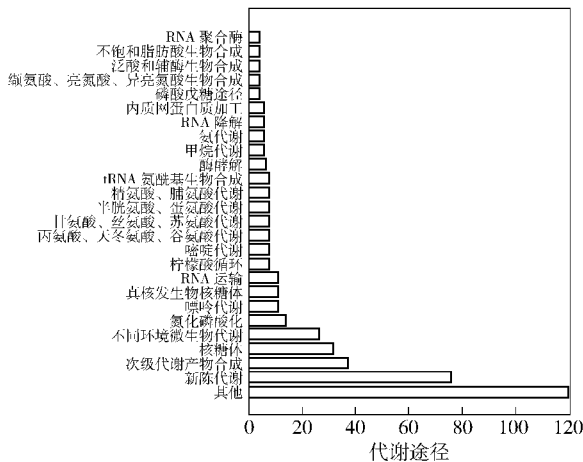


图 7 t1 特异表达 PASSWAY 分析图

Fig. 7 The KEGG pathways in the transcriptomes in t1

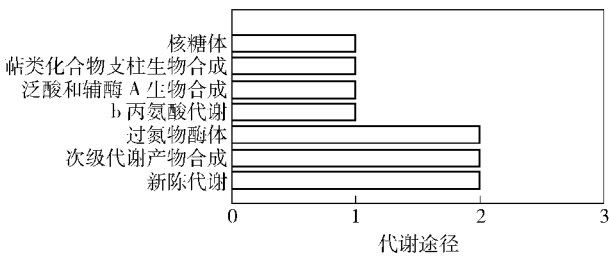


图 8 t2 特异表达 PASSWAY 分析图

Fig. 8 The KEGG pathways in the transcriptomes in t2

PASSWAY 分析结果表明,很多发酵初期的转录表达达到发酵中期均停止,随着环境温度(堆温)升高,很多微生物由于不适应环境的变化休眠或者死亡,相对应的代谢活动也停止。与前面的微生物分析及 GO 注释结果结合分析可发现,这些不适应环境变化的微生物代谢活动的停止并没有对整个发酵

过程产生重大影响,而剩下的微生物不但适应环境变化继续繁殖生长,还比之前略有上升趋势,它们才是对发酵过程起决定性作用的微生物。

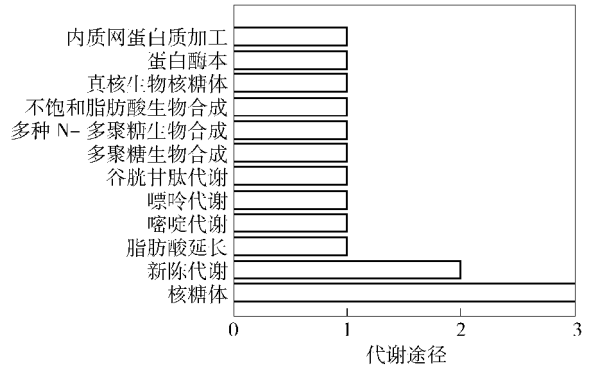


图 9 显著差异 PASSWAY 表达分析图

Fig. 9 Comparison between the KEGG pathways in the transcriptomes t1 and t2

3 结论

(1)宏转录组学对比分析表明,普洱茶发酵前期和中期优势微生物均为黑曲霉。一些不适应发酵环境变化的微生物停止转录表达,但并没有对发酵进程造成重大的影响,而其余的微生物不但适应环境变化继续生长繁殖代谢活力还有上升趋势,它们才对渥堆发酵过程起了决定性作用。

(2)宏转录组学技术可用于研究普洱茶发酵过程乃至其他复杂的食品工业发酵过程。而宏转录组学研究比起以往其他的研究方法,更关注于整个微生物群落代谢变化和应答反应,使其对发酵结果即目标产物积累变化更具有针对性的指导作用。

参考文献:

[1] 黄振兴,赵明星,阮文权,等.普洱茶渥堆过程中微生物对其品质形成的影响及其研究进展[J].安徽农业科学,2008,36(28):12496-12498.
Huang Z X, Zhao M X, Ruan W Q, et al. Effects of microorganisms on the quality of Puer tea during piling and its research progress [J]. Journal of Anhui Agri Sci, 2008, 36(28): 12496-12498. (in Chinese)

[2] 方祥,陈栋,李晶晶,等.普洱茶不同贮藏时期微生物种群的鉴定[J].现代食品科技,2008,24(2):105-108.
Fang X, Chen D, Li J J, et al. Identification of microbial species in Pu-erh tea with different storage time [J]. Modern Food Science and Technology, 2008, 24(2): 105-108. (in Chinese)

- [3] Abe M, Takaoka N, Idemoto Y, et al. Characteristic fungi observed in the fermentation process for puer tea [J]. *International Journal of Food Microbiology*, 2008, 124(2): 199–203.
- [4] 杨瑞娟, 吕杰, 严亮, 等. 普洱茶渥堆发酵中嗜热真菌的分离和鉴定[J]. *茶叶科学*, 2011, 31(4): 371–378.
Yang R J, Lu J, Yan L, et al. Isolation and identification of thermophilic fungi during the fermentation of puer tea [J]. *Journal of Tea Science*, 2011, 31(4): 371–378. (in Chinese)
- [5] Poretsky R S, Bano N, Buchan A, et al. Analysis of microbial gene transcripts in environmental samples [J]. *Applied and Environmental Microbiology*, 2005, 71(7): 4121–4126.
- [6] Poretsky R S, Hewson I, Sun S L, et al. Comparative day/night metatranscriptomic analysis of microbial communities in the north pacific subtropical gyre [J]. *Environmental Microbiology*, 2009, 11(6): 1358–1375.
- [7] Frias-Lopez J, Shi Y M, Tyson G W, et al. Microbial community gene expression in ocean surface waters [J]. *Proceedings of the National Academy of Sciences*, 2008, 105(10): 3805–3810.
- [8] Hewson I, Poretsky R S, Dyhrman S T, et al. Microbial community gene expression within colonies of the diazotroph, *Trichodesmium*, from the southwest pacific ocean [J]. *The ISME Journal*, 2009, 3: 1286–1300.
- [9] Bailly J, Fraissinet-Tachet L, Verner M C, et al. Soil eukaryotic functional diversity, a metatranscriptomic approach [J]. *The ISME Journal*, 2007, 1: 632–642.

Comparative metatranscriptomic analysis of microbial communities at two different stages during the pile-fermentation of Puer tea

JIANG Shu LU Jie LI Hao

(College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: mRNAs of microbial communities during different stages of the pile-fermentation of Puer tea have been extracted for Illumina sequencing, which produced 17376 putative mRNA reads from an earlier stage and 14456 from a mid-stage of fermentation. Taxonomic binning of annotated mRNAs indicated that *Aspergillus niger* was the most dominant contributor of transcripts of microbial communities in both stages. GOfact functional annotation and KEGG pathway analysis suggested that adaption of *Aspergillus niger* to changes in the environment played a decisive role during the pile-fermentation of Puer tea.

Key words: Puer tea; pile-fermentation; metatranscriptomic; microbial communities