

## 湖南的新记录植物(八)

周柳<sup>1</sup>, 田径<sup>1</sup>, 吴尧晶<sup>2</sup>, 李帅杰<sup>1</sup>, 吴玉<sup>1</sup>, 旷仁平<sup>1</sup>, 刘坤<sup>3</sup>, 刘克明<sup>1\*</sup>

(1. 湖南师范大学生命科学学院, 中国湖南长沙 410081; 2. 湖南芷江师范学校, 中国湖南怀化 418000; 3. 湖南省永州市金洞林场, 中国湖南永州 426191)

**摘要:** 报道了湖南种子植物新记录种 6 个, 即: 榕树 *Ficus microcarpa* L.、封怀凤仙花 *Impatiens fenghwaiana* Y. L. Chen、平坝凤仙花 *Impatiens ganpiuana* Hook. f.、裂叶月见草 *Oenothera laciniata* Hill、弯管马铃苣苔 *Oreocharis curvituba* J. J. Wei & W. B. Xu 和象草 *Pennisetum purpureum* Schum. (象草为归化种)。对部分种类的主要特征、生境和地理分布进行了初步讨论。

**关键词:** 种子植物; 植物分布; 新记录; 湖南省

中图分类号: Q949.4

文献标识码: A

文章编号: 1007-7847(2019)01-0035-04

## Newly Recorded Plants from Hunan Province of China (VIII)

ZHOU Liu<sup>1</sup>, TIAN Jing<sup>1</sup>, WU Yao-jing<sup>2</sup>, LI Shuai-jie<sup>1</sup>, WU Yu<sup>1</sup>,  
KUANG Ren-ping<sup>1</sup>, LIU Kun<sup>3</sup>, LIU Ke-ming<sup>1\*</sup>

(1. College of Life Sciences, Hunan Normal University, Changsha 410081, Hunan, China; 2. Hunan Zhijiang Normal School, Huaihua 418000, Hunan, China; 3. Jindong Forest Farm of Yongzhou City, Yongzhou 426191, Hunan, China)

**Abstract:** Six species of seed plants are reported as new records to the flora of Hunan Province. They are *Ficus microcarpa* L., *Impatiens fenghwaiana* Y. L. Chen, *Impatiens ganpiuana* Hook. f., *Oenothera laciniata* Hill, *Oreocharis curvituba* J. J. Wei & W. B. Xu and *Pennisetum purpureum* Schum. (*Pennisetum purpureum* is a naturalized species). The main characteristics, habitats and geographical distribution of some species are discussed preliminarily.

**Key words:** seed plant; plant distribution; new record; Hunan Province

(*Life Science Research*, 2019, 23(1): 035-038)

为编写《湖南植物志》和进行相关研究, 我们对湖南植物区系进行了深入调查, 对中国数字植物标本馆(Chinese Virtual Herbarium, CVH)及湖南师范大学植物标本馆(HNNU)馆藏标本, 特别是近年来采集的标本, 进行了鉴定整理, 在此过程中相继发现了若干湖南种子植物新分布, 此前已作了部分报道<sup>[1-10]</sup>, 本文继续报道湖南的新记录植物, 共 6 种, 分属于 5 科, 5 属。文中引证的标本, 均

收藏在湖南师范大学植物标本馆(HNNU)。

### 1 桑科 Moraceae

#### 1.1 榕树(细叶榕) (图 1A, B)

*Ficus microcarpa* L. f. in suppl. Sp. Pl. 442. 1781; 中国植物志 23(1): 112. 1998; Flora of China 5: 44. 2003.

主要特征: 大乔木, 高达 25 m, 冠幅广展; 老

收稿日期: 2018-10-10; 修回日期: 2018-11-16

基金项目: 国家科技基础性研究专项(2013FY111500); 湖南省林木种质资源调查资助项目[湘财农指(2015) 91 号]; 中国科学院科技服务网络计划资助项目(KFJ-3W-No1); 湖南省外来入侵生物调查资助项目[湘财农指(2014) 91 号]; 湖南省农业野生植物资源调查资助项目(湖南省农业厅); 湖南省古树名木资源普查内业数据技术审查资助项目(湖南省林业厅, 2017); 湖南省生态学重点学科建设项目(0713); 湖南省生物发育工程与新产品研发协同创新中心资助项目(20134486); “作物不育分子机制与资源创新”湖南省重点实验室资助项目(2016TP1011)

作者简介: 周柳(1994-), 女, 湖南常德人, 硕士研究生, 主要从事植物分类学研究, E-mail: 894662083@qq.com; \* 通讯作者: 刘克明(1952-), 男, 湖南沅江人, 湖南师范大学生命科学学院教授, 主要从事植物系统分类学及植物区系学研究, Tel: 0731-88872864, E-mail: lk-ming8@aliyun.com。

树常有锈褐色气生根。叶薄革质, 狭椭圆形, 长 4~8 cm, 宽 3~4 cm, 先端钝尖, 基部楔形, 表面深绿色, 全缘, 侧脉 3~10 对; 叶柄长 5~10 mm。榕果成对腋生或生于已落叶枝叶腋, 成熟时黄或微红色, 扁球形, 直径 6~8 mm。花期 5~6 月。

本种小枝无瘤体, 叶薄革质, 侧脉与中脉成钝角展开, 网脉不显著; 榕果成熟时黄或微红色, 易与近缘种瘤枝榕 *Ficus maclellandi* King 区别。

湖南: 汝城县, 三江口瑶族镇, 仙溪村大坪组 (113°46'54.76"E, 25°21'13.06"N), 海拔 316 m, 村落风水林中, 刘克明、周柳 34238, 2017 年 11 月 19 日; 同地, 周柳、田径 34241, 2018 年 6 月 12 日。照片号: 刘克明 IMG2014-09-9970、P1310713、P13-10714、P1460563。

分布: 广东、海南、广西、福建、台湾、浙江、贵州、云南; 新几内亚; 不丹、印度、马来西亚、缅甸、尼泊尔、斯里兰卡、泰国、越南、澳大利亚<sup>[11-13]</sup>。湖南首次记录。

该榕树生长在湖南省东南部汝城县的三江口瑶族镇, 北与广东省乐昌市的仁化县相邻, 该区属亚热带季风湿润气候, 温暖湿润, 热量丰富, 雨量充沛(汝城县位于湖南省东南部多雨中心的附近, 年均降水量 1 547.1 mm), 光照充足, 优越的自然环境为榕树生长提供了条件。该榕树生长在村落附近的风水林中, 树高 13.5 m, 胸径 168 cm, 冠幅 26 m×28 m, 估测树龄 150 年, 生长繁茂, 冠幅广展, 老树有锈褐色气生根。据当地村民陈述, 该榕树原本是长在一株木荷树(*Schima superba* Gardn. et Champ.)上, 后来木荷树逐渐枯死并被该榕树取代。我们推测, 该榕树很可能是一百多年前由鸟类携带、传播的尚未消化的榕树种子随鸟类排泄物一起落到此处的木荷树上萌发、生长所致, 无疑其属原生个体, 而且可能是我国榕树分布纬度最北的植株。榕树在湖南省的发现, 对研究该种的分布与扩散及当地气候变迁乃至湖南省植物区系等具有一定科学价值。

## 2 凤仙花科 Balsaminaceae

### 2.1 封怀凤仙花(图 1C)

*Impatiens fenghuaiensis* Y. L. Chen in Acta Phytotax. Sin. 27(5): 395. Fig. 1. 1989; 中国植物志 47(2): 194. 2001; Flora of China 12: 106. 2007.

本种叶片侧脉 7~9 对, 基部无腺体; 苞片粉红色, 卵形、心形, 长约 8 mm; 侧生萼片顶部具不

明显的小尖; 翼瓣基部裂片长圆状倒卵形, 与近缘种婺源凤仙花 *Impatiens wuyuanensis* Y. L. Chen 相区别。

湖南: 安仁县豪山乡, 金花村, 金子仙, 海拔 1 248 m, 山坡疏林向阳处, 刘克明、周柳 LXP03-01281, 2016 年 9 月 9 日。照片号: DSCN-6082、DSCN-6107、DSCN-6113。

分布: 江西、湖北<sup>[14-15]</sup>。湖南首次记录。

### 2.2 平坝凤仙花(图 1D, E)

*Impatiens ganpiuana* Hook. f. in Hook. Icon. Pl. t. 2917. 1910; 中国植物志 47(2): 161. 2001; Flora of China 12: 95. 2007.

主要特征: 一年生草本, 高 30~60 cm; 茎、枝上部被褐色绒毛, 基部匍匐。叶椭圆形至长披针形, 长 3~6 cm, 边缘具粗锯齿, 侧脉每边 4~6 条, 叶柄长 1~2 cm。总花梗常具 2 花; 花小, 黄色; 侧生萼片 2; 旗瓣椭圆形; 翼瓣 3 裂, 上部裂片斧形, 先端微凹成 2 枚裂片, 下部裂片阔卵形; 唇瓣漏斗状, 基部下延为细长内弯的长距。蒴果狭纺锤形。花期 7~8 月, 果期 9~10 月。

本种枝上部具开展的褐色绒毛; 总花梗长约 1 cm; 苞片狭披针形, 无齿; 翼瓣上部裂片斧形, 先端微凹成 2 枚裂片, 易与近缘种齿苞凤仙花 *Impatiens martinii* Hook. f. 相区别。

湖南: 城步苗族自治县, 南山牧场, 响水坝, 海拔 1 625 m, 山地沟谷旁潮湿处, 刘坤、周柳 34239, 2018 年 8 月 12 日。照片号: 周柳 DSC-4309、DSC-4313、DSC-4319。

分布: 贵州<sup>[16]</sup>。湖南首次记录。

## 3 柳叶菜科 Onagraceae

### 3.1 裂叶月见草(图 1H)

*Oenothera laciniata* Hill, Veg. Syst. 12, appendix: 64, Pl. 10. 1767; 中国植物志 53(2): 68. 2000; Flora of China 13: 425. 2007.

主要特征: 直立至外倾, 一年生或多年生草本; 茎长 10~50 cm, 被柔毛, 在茎上部常混生腺毛。基部叶线状倒披针形, 边缘羽状深裂; 茎生叶狭倒卵形或狭椭圆形, 下部常羽状裂, 中上部具齿, 上部近全缘。花序穗状, 由少数花组成, 每天日落时每序开一朵花; 花瓣淡黄至黄色, 宽倒卵形, 先端截形至微凹。蒴果圆柱状, 向顶变狭, 长 2.5~5 cm。花期 4~9 月。

本种茎生叶狭倒卵形或狭椭圆形, 宽 0.7~3 cm,

基部楔形,下部羽状深裂,与近缘种待宵草 *Oenothera stricta* Ledeb. et Link 相区别。

湖南:郴州苏仙区,飞天山镇,飞天山,海拔 143 m,逸生于荒地、路边,吴玉、李帅杰 772946,2018 年 5 月 18 日。照片号:吴玉 WUY-8104、WUY-8107、WUY8108。

分布:福建、台湾、广东、浙江<sup>[17-19]</sup>;原产美国东部至中部,在日本、非洲南部、澳大利亚、中美洲、欧洲和南美洲被引种。湖南首次记录。

研究表明,本种适应性广,具有较强的耐寒、耐贫瘠、耐盐碱和耐水湿的能力;在伴生植物拥挤情况下能正常开花结实,大型植株还能通过化学他感能力排挤伴生土著植物,具有较强的侵入能力<sup>[20]</sup>。

#### 4 苦苣苔科 Gesneriaceae

##### 4.1 弯管马铃苣苔(图 1F, G)

*Oreocharis curvituba* J. J. Wei & W. B. Xu in Phytotaxa 280(2): 190. 2016.

主要特征:多年生草本。叶全部基生,莲座状;叶片卵形至椭圆形,顶端钝圆,基部稍偏斜,边缘具小齿,上面有囊泡状小突起,密被柔毛。花白色或淡紫色;花冠细筒状,筒部稍弯曲,喉部具一环紫色星状腺毛;退化雄蕊 1。花期 8~9 月,果期 9~10 月。

本种的花冠筒稍弯曲,喉部具一环紫色星状腺毛,与近缘种 *Oreocharis stenosphon* Mich.Möller & A.Weber 和大齿马铃苣苔 *Oreocharis magnidens* Chun ex K. Y. Pan 相区别。

湖南:桂东县,新坊乡,泥塘村,林下石壁上,海拔 1 172 m,田径、周柳 LXP03-2623 (HNNU),2016 年 9 月 20 日。照片号:IMG-8918、IMG-8937、IMG-8938。

分布:广西、江西<sup>[21-22]</sup>。本种为 2016 年发表的新种,模式标本产地为广西壮族自治区灌阳县千家峒自然保护区。湖南首次记录。

#### 5 禾本科 Gramineae

##### 5.1 象草(图 1I, J)

*Pennisetum purpureum* Schum. in Beaskr. Guin. Pl. 44. 1827; 中国植物志 10(1): 373. 1990; Flora of China 22: 550. 2006.

主要特征:多年生丛生大型草本,常具地下茎。秆粗壮,直立,高 2~4 m,光滑,在花序基部密

生柔毛。叶舌短小,具长 1.5~5 mm 的纤毛;叶片线形,质较硬,长 20~50 cm,宽 1~3 cm,上面疏生刺毛,近基部有小疣毛,下面无毛,边缘粗糙。圆锥花序长 10~30 cm,主轴密生长柔毛,雄蕊 3,花药顶端具一束短毛。花果期 8~10 月。

本种为多年生丛生大型草本,秆粗壮,直立,高 2~4 m,雄蕊的花药顶端具一束短毛,与国产的同属其他种类易于区别。

湖南:醴陵市,泗汾镇,石湾村,新坊组,海拔 81 m,村边荒草地(可能为逸生),刘克明、周柳 LX-P03-5400,2016 年 9 月 28 日。照片号:周柳、吴尧晶 DSC-5555、DSC-5559、DSC-5630。

分布:原产非洲。我国江西、福建、台湾、江苏、广东、海南、广西、四川、云南引种栽培,偶有逸生,在海南已野化<sup>[23]</sup>。湖南首次记录。

象草是热带和亚热带地区优良饲用植物,具有较高的营养价值,蛋白质含量高,适口性好,嫩

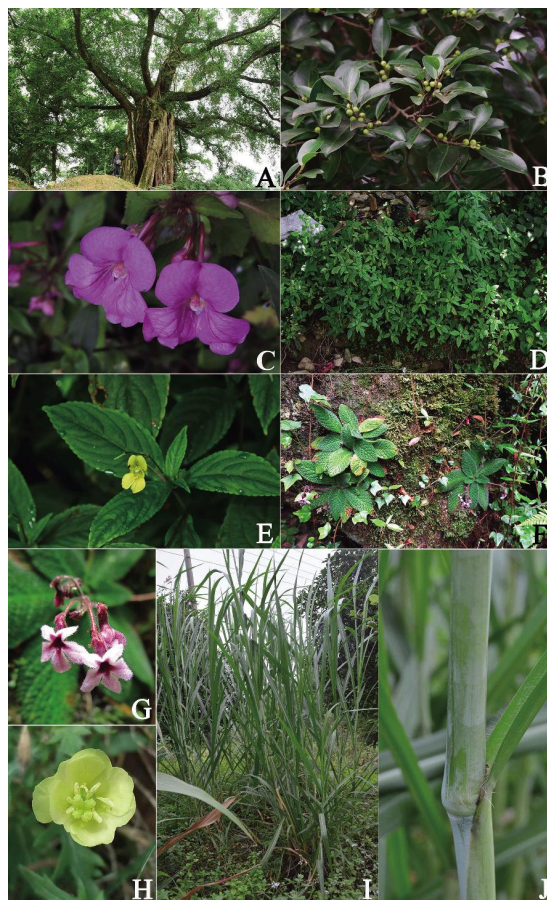


图 1 湖南省 6 种新记录植物的部分照片

Fig.1 Six newly recorded plants from Hunan Province A, B: *Ficus microcarpa* L. f.; C: *Impatiens fenghwaiana* Y. L. Chen; D, E: *Impatiens ganpiuana* Hook. f.; F, G: *Oreocharis curvituba* J. J. Wei & W. B. Xu; H: *Oenothera laciniata* Hill; I, J: *Pennisetum purpureum* Schum..

茎叶柔软多汁,牛、马、羊、兔、鸭、鹅等均喜食,幼嫩期也是养猪、养鱼的好饲料。此外,象草还能替代煤炭、石油发电,具有较高的经济价值。但另一方面,象草为多年生丛生大型草本,高 2~4 m,通常具根状茎,根系十分发达,生长快,适应性及无性繁殖能力均较强,变异性较大,对其引种存在潜在入侵风险。近年来,湖南省除醴陵外,长沙县(北山镇)、涟源(杨市镇)等均有较大面积的种植,且呈半野生状态(我们拍摄的凭证标本植株照片很可能为逸生者)。象草的环境适应性及其繁殖特性表明,其可能是一种具潜在入侵性的外来物种。因此,我们认为,人们在进行象草引种时,应避免在开阔区域种植,其种植地的选择应具有隔离屏障或可控性,同时对其加强管理、观察和监测。如将植株随意丢弃或疏于管理等,很容易造成该种的逃逸与扩散。基于该种在湖南省引种栽培的历史不长(3~4年),对其潜在危害或入侵性尚未被充分认识,对其引种的利、害尚存在不确定性,人们在享受其引种带来好处的同时,对其可能存在的入侵风险应该给予关注。

#### 参考文献(References):

- [1] 刘克明,刘林翰,李丙贵. 湖南的新记录植物(一)[J]. 广西植物(LIU Ke-ming, LIU Lin-han, LI Bin-gui. Some new recorded taxa from Hunan Province[J]. Guihaia), 1995, 15(2): 116-119.
- [2] 刘克明,李丙贵. 湖南植物区系增补[J]. 植物研究(LIU Ke-ming, LI Bin-gui. Additions to flora of Hunan[J]. Bulletin of Botanical Research), 1996, 16(3): 281-285.
- [3] 刘克明,胡光万. 湖南的新记录植物(二)[J]. 生命科学研究(LIU Ke-ming, HU Guang-wan. Some new recorded plants from Hunan (II)[J]. Life Science Research), 1998, 2(1): 37-39.
- [4] 林祁,刘克明. 湖南植物区系新记录(三)[J]. 生命科学研究(LIN Qi, LIU Ke-ming. New records of flora of Hunan (III)[J]. Life Science Research), 1999, 3(3): 246-249.
- [5] 胡光万,雷立公,刘克明. 湖南的新记录植物(四)[J]. 植物研究(HU Guang-wan, LEI Li-gong, LIU Ke-ming. Some newly recorded plants from Hunan Province of China (IV)[J]. Bulletin of Botanical Research), 2002, 22(2): 136-138.
- [6] 胡光万,刘克明,刘林翰. 湖南的新记录植物(五)[J]. 植物研究(HU Guang-wan, LIU Ke-ming, LIU Lin-han. Some newly recorded plants from Hunan Province of China (V)[J]. Bulletin of Botanical Research), 2003, 23(4): 399-401.
- [7] 黄文新,刘克明,蔡秀珍,等. 湖南的新记录植物(六)[J]. 植物研究(HUANG Wen-xin, LIU Ke-ming, CAI Xiu-zhen, et al. Some newly recorded plants from Hunan Province of China (VI)[J]. Bulletin of Botanical Research), 2004, 24(4): 396-399.
- [8] 喻勋林,刘克明,黄孔泽,等. 湖南省新记录植物[J]. 中南林业学院学报(YU Xun-lin, LIU Ke-ming, HUANG Kong-ze, et al. Newly recorded plants from Hunan Province of China[J]. Journal of Central South Forestry University), 2006, 26(2): 143-144.
- [9] 喻勋林,刘克明,谷志容. 湖南省新记录植物(II)[J]. 中南林业科技大学学报(YU Xun-lin, LIU Ke-ming, GU Zhi-rong. Newly recorded plants from Hunan Province of China (II)[J]. Journal of Central South University of Forestry & Technology), 2007, 27(3): 66-69.
- [10] 彭令,刘雷,肖顺勇,等. 湖南的新记录植物(七)[J]. 湖南师范大学自然科学学报(PENG ling, LIU Lei, XIAO Shun-yong, et al. Some newly recorded plants from Hunan Province of China (VII)[J]. Journal of Natural Science of Hunan Normal University), 2016, 39(6): 26-31.
- [11] ZHOU Z K, GILBERT M G. Moraceae[M]//WU C Y, RAVEN P H, HONG D Y. Flora of China (Vol. 5). Beijing: Science Press & St. Louis: Missouri Botanical Garden Press, 2003: 44.
- [12] 杨大荣,徐磊,彭艳琼,等. 云南省榕小蜂和榕树的物种组成及多样性[J]. 生物多样性(YANG Da-rong, XU Lei, PENG Yan-qiong, et al. Species composition and diversity of fig wasps and figs in Yunnan[J]. Biodiversity Science), 2004, 12(4): 611-617.
- [13] 林淑玲,赵南先,陈贻竹,等. 榕树(*Ficus*)在中国的分布及其在协同进化研究上的意义[J]. 生态学报(LIN Shu-ling, ZHAO Nan-xian, CHEN Yi-zhu, et al. Distribution of figs (*Ficus*) in China and its significance in the issues for interspecific co-evolution[J]. Acta Ecologica Sinica), 2007, 27(10): 4278-4288.
- [14] 刘虹,祝贺,田焕焕,等. 湖北凤仙花属植物分布新纪录[J]. 中南民族大学学报(自然科学版)(LIU Hong, ZHU He, TIAN Huan-huan, et al. New records of *Impatiens* spp. in Hubei Province[J]. Journal of South-Central University for Nationalities (Natural Science Edition)), 2017, 36(3): 35-38.
- [15] 陈艺林. 中国凤仙花属四新种[J]. 中国科学院大学学报(CHEN Yi-lin. Four new species of *Impatiens* (Balsaminaceae) from China[J]. Acta Phytotaxonomica Sinica), 1989, 27(5): 395-404.
- [16] CHEN Y L, AKIYAMA S, OHBA H. Balsaminaceae[M]//WU C Y, RAVEN P H, HONG D Y. Flora of China (Vol. 12). Beijing: Science Press & St. Louis: Missouri Botanical Garden Press, 2007: 95.
- [17] CHEN J R, HOCH P C, WAGNER W L. *Oenothera* L.[M]//WU C Y, RAVEN P H, HONG D Y. Flora of China (Vol. 13). Beijing: Science Press & St. Louis: Missouri Botanical Garden Press, 2007: 425.
- [18] 曾宪锋,邱贺媛,唐光大,等. 广东省 5 种新记录植物[J]. 华南农业大学学报(ZENG Xian-feng, QIU He-yuan, TANG Guang-da, et al. Five newly recorded plants from Guangdong Province[J]. Journal of South China Agricultural University), 2008, 29(3): 59-60.
- [19] 赖秀雅,吴庆玲,李想,等. 浙江归化植物新资料[J]. 温州大学学报(自然科学版)(LAI Xiu-ya, WU Qing-ling, LI Xiang, et al. New records of naturalized plants in Zhejiang Province[J]. Journal of Wenzhou University (Natural Sciences)), 2008, 29(5): 13-16.
- [20] 蒋明,曹家树,丁炳扬,等. 新外来杂草裂叶月见草的生物学特性及防控对策[J]. 生物学通报(JIANG Ming, CAO Jia-shu, DING Bing-yang, et al. New exotic weeds biological characteristics and control strategies of evening primrose[J]. Bulletin of Biology), 2004, 39(9): 20-21.
- [21] WEI J J, XIONG G C, ZOU C Y, et al. *Oreocharis curviflora*, a new species of Gesneriaceae from northeastern Guangxi, China[J]. Phytotaxa, 2016, 280(2): 190.
- [22] 冯璐,王浩威,肖敏,等. 江西省齐云山地区种子植物新资料[J]. 亚热带植物科学(FENG Lu, WANG Hao-wei, XIAO Min, et al. New records of seed plants from Qiyunshan Nature Reserve, Jiangxi Province[J]. Subtropical Plant Science), 2018, 47(1): 72-76.
- [23] CHEN S L, PHILLIPS S M. *Pennisetum* Richard[M]//WU C Y, RAVEN P H, HONG D Y. Flora of China (Vol. 22). Beijing: Science Press & St. Louis: Missouri Botanical Garden Press, 2006: 550.

# 一种高效自动化的转录组差异性表达分析方法

赵 芳, 高 静\*, 刘振羽, 王永军, 邬学敏

(内蒙古农业大学 计算机与信息工程学院 内蒙古自治区农牧业大数据研究与应用重点实验室, 中国内蒙古  
呼和浩特 010018)

**摘 要:** 差异性表达分析是转录组研究的核心目标之一, 对于揭示基因功能和调控规律具有重要意义。但该分析属于多步迭代且耗时较长的计算密集型过程, 软件之间具有复杂的数据依赖关系, 输入输出格式不尽相同。传统方式下, 软件安装与使用复杂、繁琐的手动操作、分析环境的不易迁移以及无法按需集成都是有待解决的关键问题。针对上述问题, 文中首次将云开源项目 Docker 容器技术应用于生物信息领域, 提出一种高效自动化的转录组差异性表达分析方法。首先将最佳实践流程在 Docker 容器中内置与集成, 其次多脚本联动与 Web 服务相结合, 最后形成一个轻量级、易迁移、高效自动化的转录组差异性表达分析“黑匣子”。实验结果表明, 与传统方式相比, 该方法的分析时间缩短约 72%, 效率提升 2 倍多, 为研究人员提供了更高效的技术支持。

**关键词:** Docker 容器; 转录组; 差异性表达分析; 最佳实践流程; 脚本

中图分类号: Q-31

文献标识码: A

文章编号: 1007-7847(2019)01-0039-07

## A Highly Efficient and Automated Method for Differential Expression Analysis of Transcriptome

ZHAO Fang, GAO Jing\*, LIU Zhen-yu, WANG Yong-jun, WU Xue-min

(Inner Mongolia Autonomous Region Key Laboratory of Big Data Research and Application of Agriculture and Animal Husbandry, College of Computer and Information Engineering, Inner Mongolia Agricultural University, Hohhot 010018, Inner Mongolia, China)

**Abstract:** The differential expression analysis is one of the core objectives of transcriptome research, which is of great significance to reveal the gene function and regulation rules. However, the analysis is a multi-step iteration with a time-consuming and computationally intensive process. There are complex data dependencies between software, and the input and output formats are not the same. In the traditional way, the complex installation and use of software, tedious manual operation and difficult migration of analysis environment are the key problems to be solved. In view of the above problems, an efficient and automated method for transcriptome differential expression analysis was proposed. For the first time, Docker container technology, an open source cloud project, was applied in the field of biological information. First of all, the best practice process was embedded and integrated in Docker container. Then the multi-script interaction and web services were combined, and eventually, a “black box” of lightweight, portable, and highly automated transcriptome differential expression analysis was formed. Experimental results showed that, compared with the traditional method, this method reduced the analysis time by about 72%, and improved the efficiency more than twice. Therefore, the method may provide more efficient technical support for researchers.

**Key words:** Docker container; transcriptome; the differential expression analysis; the best practice process;

收稿日期: 2018-09-14; 修回日期: 2018-10-12

基金项目: 国家自然科学基金资助项目(61462070); 内蒙古农业大学博士科研启动基金资助项目(BJ09-44)

作者简介: 赵芳(1993-), 女, 内蒙古清水河人, 硕士研究生, 主要从事生物大数据相关研究; \*通讯作者: 高静(1970-), 女, 满族, 辽宁沈阳人, 博士, 内蒙古农业大学教授, 博士生导师, 主要从事大数据与云计算及软件工程研究, Tel: 0471-4317420, E-mail: Gaojing@imau.edu.cn.

scripts

转录组差异性表达分析是 RNA 测序数据分析的核心内容,也是转录组研究的基本目标之一。其通过分析不同条件下的转录组数据,进而识别差异基因或异构体,对于揭示基因功能和调控规律具有重要意义<sup>[1]</sup>。但该过程是一个多步迭代的复杂流程,涉及多软件、多步骤,软件之间还具有复杂的数据依赖关系。传统方式下,研究人员手动执行各环节,前一任务的输出即为后一任务的输入,只有前一任务执行完毕后才可串行执行后续依赖任务<sup>[2]</sup>。而且,各软件的输入输出格式不尽相同,均需研究人员进行手动的格式转换。该过程中,复杂的命令行操作以及大量的参数配置很大程度上增加了转录组差异性表达分析的难度,导致效率低、出错率高、可重复性差。此外,环境的不易迁移性也进一步导致数据分析问题成为生物信息学的研究瓶颈。

近年来,针对转录组差异性表达分析,人们开发了一些高性能的自动化分析框架,如 system-PipeR<sup>[3]</sup>、SAMSA<sup>[4]</sup>、miRPursuit<sup>[5]</sup>。这些框架实现了转录组差异性分析的自动化,一定程度上克服了传统方式需大量手动操作的弊端。但在框架部署与初始配置时,仅提供命令行或特定的编程语言接口,仍需用户进行较多的手动操作,不利于缺乏编程基础的生物信息人员使用。此外,移植性较差,需用户在不同的环境下进行多次重复性部署,且用户无法按需集成新的工具,无法满足研究人员的实际分析需求。

因此,文中针对以上低效重复的手动操作、环境的不可迁移以及无法按需集成等问题,提出了一种基于 Docker 容器技术的转录组差异性表达分析方法,形成一套可重复、易移植、高效自动化的轻量级分析框架。

## 1 概念定义与原理

### 1.1 Docker 容器技术

Docker 是基于 Go 语言实现的云开源项目,其主要目标是通过应用组件的封装、分发、部署、运行等生命周期的管理,使用户的 App (Web 应用或数据库应用等)及其运行环境能够做到“一次封装,到处运行”。Docker 具有部署速度快、开发测试敏捷、提高系统利用率、降低资源成本、跨

(Life Science Research, 2019, 23(1): 039~045)

环境可移植性好等优点。

因此,本研究首次将云开源项目 Docker 虚拟化技术应用生物信息学领域,用于解决复杂分析的不可重复性、不易迁移性以及无法按需集成等核心难题,使经验较少或非生物信息专业的研究人员执行复杂的流程分析时就像运行一个简单的生物信息学工具一样。

### 1.2 最佳实践流程

转录组差异性表达分析流程较为复杂,涉及多步骤的处理,如序列比对、转录本组装、转录本合并等。各处理过程需借助不同的工具软件来实现,但目前各步骤的处理软件层出不穷,软件之间复杂的数据依赖关系使得研究人员往往存在着软件选择难的问题,面临着各软件相互衔接难的困境<sup>[6]</sup>。

2016 年, Ghosh 和 Cean 针对转录组差异性表达分析提出了一套最佳实践流程<sup>[7]</sup>,该流程在提出的几年里被生物信息学领域专家所认可并延续至今仍被广泛使用。因此,文中将该最佳实践流程 Bowtie2<sup>[8]</sup>-TopHat2<sup>[9]</sup>-Cufflinks<sup>[10]</sup>-cummeRbund<sup>[11]</sup>与数据预处理环节相整合,在已构建好参考基因组的前提下,将转录组差异性表达分析主要分为 8 个处理步骤,形成图 1 所示完整流程。

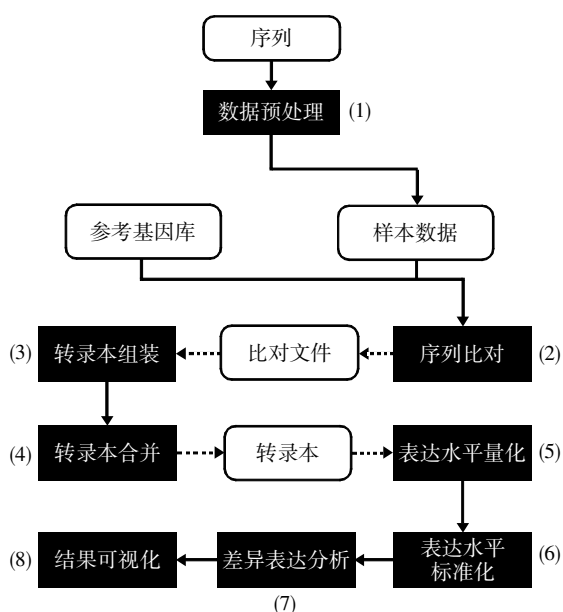


图 1 转录组差异性表达分析处理流程

Fig.1 The transcriptome differential expression analysis process

1) 数据预处理: 首先, 用 SRA-Toolkit (<https://hpc.nih.gov/apps/sratoolkit.html>) 将 sra 格式的原始测序数据转换为 fastq 格式; 其次, 通过 FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) 对测序数据进行质量检测; 最后, 通过 Trimmomatic 质量检测<sup>[12]</sup>将质量低或带有接头(a-dapters)的测序数据进行修剪和过滤, 进而得到高质量测序数据。

2) 序列比对: TopHat2 (依赖 Bowtie2, 比对过程中会调用 Bowtie2)将测序数据与参考基因组进行比对, 输出结果文件为 accepted\_hits.bam。

3) 转录本组装: Cufflinks 将 TopHat2/Bowtie2 比对结果 accepted\_hits.bam 及参考基因组作为输入文件构建转录本, 并计算出各转录本 FPKM 值。

4) 转录本合并: Cuffmerge 将多个转录本合并成一套转录本集合, 将 Cufflinks 生成的 gtf 文件融合为一个更加全面的注释文件 merged.gtf。

5) 表达水平量化: Cuffquant 对单个 bam 文件表达水平进行定量, 用于后续 Cuffdiff 分析。

6) 表达水平标准化: 将 Cuffquant 的输出结果作为 Cuffnorm 的输入文件, 对转录组标准化后的表达水平进行计算, 可得到一系列可比较的基因、转录组、CDS 组和 TSS 组的表达值。

7) 差异表达分析: 通过 Cuffdiff 计算不同条件下转录本表达水平的显著性差异, 进而寻找差异基因。

8) 结果可视化: 采用 cummeRbund 软件对 Cuffdiff 输出结果以热图形式进行可视化呈现, 以更直观的方式描述目标差异基因。

转录组差异性表达分析流程中所需软件如表 1 所示。

表 1 最佳实践流程软件列表  
Table 1 List of best practice softwares

步骤 Steps	软件 Software	网址 URL
数据预处理	SRA-Toolkit	<a href="https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software">https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software</a>
	FastQC	<a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc/">http://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
	Trimmomatic	<a href="http://www.usadellab.org/cms/index.php?page=trimmomatic">http://www.usadellab.org/cms/index.php?page=trimmomatic</a>
序列比对	Bowtie2	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
	TopHat2	<a href="http://ccb.jhu.edu/software/tophat/faq.shtml">http://ccb.jhu.edu/software/tophat/faq.shtml</a>
转录本组装	Cufflinks	<a href="http://cole-trapnell-lab.github.io/cufflinks/">http://cole-trapnell-lab.github.io/cufflinks/</a>
转录本合并	Cuffmerge	<a href="http://cole-trapnell-lab.github.io/cufflinks/">http://cole-trapnell-lab.github.io/cufflinks/</a>
表达水平量化	Cuffquant	<a href="http://cole-trapnell-lab.github.io/cufflinks/">http://cole-trapnell-lab.github.io/cufflinks/</a>
表达水平标准化	Cuffnorm	<a href="http://cole-trapnell-lab.github.io/cufflinks/">http://cole-trapnell-lab.github.io/cufflinks/</a>
差异表达分析	Cuffdiff	<a href="http://cole-trapnell-lab.github.io/cufflinks/">http://cole-trapnell-lab.github.io/cufflinks/</a>
结果可视化	cummeRbund	<a href="http://www.bioconductor.org/packages/release/bioc/html/cummeRbund.html">http://www.bioconductor.org/packages/release/bioc/html/cummeRbund.html</a>

## 2 基于 Docker 的转录组差异性表达分析方法

### 2.1 层次构建

基于 Docker 容器技术的转录组差异性表达分析方法形成的整体框架层次主要分为基础层、核心层、调度层以及应用层(图 2)。基础层除硬件设施与操作系统外, 还需为上层提供运行环境的 Docker 引擎; 核心层为转录组差异性表达分析容器, 属于核心功能组件, 可按需集成不同的复杂分析流程; 调度层即通过多脚本嵌套触发对服务、资源的协调和控制; 应用层即为用户接口层。该框架各层次之间协同配合, 进而实现高效自动化的转录组差异性表达分析。

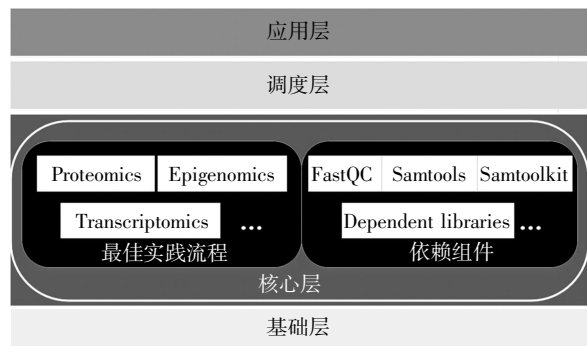


图 2 整体框架层次图

Fig.2 Framework hierarchy diagram

### 2.2 核心层

鉴于 Docker 容器可重复、配置部署快速简单、易于迁移、易于扩展的轻量级特性, 文中将 1.2 节转录组最佳实践流程涉及的软件及其依赖环境封装预配置到 Docker 容器中, 形成一个轻量级自动化的转录组差异性表达分析容器, 并将其命名为

DEA-container (differential expression analysis container)。容器内部主要通过多脚本联动的方式触发启动 DEA-container, 控制和协调各组件的执行 (核心脚本信息如表 2 所示), 从而实现转录组差异性表达的自动化分析。

表 2 核心脚本列表  
Table 2 The list of core script

脚本名称 ScriptName	功能 Function
get_input.sh	获取用户任务数据并生成用户任务表单
init_env.sh	进行初始化环境检测
set_parameters.sh	读取用户表单并设置分析参数
pipeline.sh	执行转录组差异性表达最佳实践流程
pull_output.sh	将处理结果文件提取并反馈给用户

此外, DEA-container 还为用户提供了友好的 Web 图形化界面。我们将全过程封装整合, 仅对外提供单一的数据输入和输出端点。用户仅需将目标分析序列提交。全过程无需人工干预, 经后台服务器分析完毕后会自动将分析结果文件通过 Web 界面反馈至用户, 供用户下载。

2.3 多脚本联动过程

根据整体的框架构建, 各层次之间的协调配合需要通过多脚本联动方式对转录组差异性表达的自动化进行实现, 具体实现分为以下处理流程, 其执行过程如图 3 所示。

1) 数据获取: 将用户基于 Web 图形界面提交的任务及分析参数形成任务表单, 采用 get\_input.sh 脚本读取用户数据及相关分析参数并将数据存储于指定位置。

2) 初始化检测: 通过 init\_env.sh 脚本对宿主主机 Docker 服务进行状态监测, 触发启动 DEA-

container 并对环境进行初始化, 主要完成 DEA-container 配置参数、最佳实践流程组件以及任务数据状态的检测。

3) 设置分析参数: 对用户提交的任务表单进行分析参数的读取, 即分析数据类型, 包括分析物种以及物种的名称。通过 set\_parameters.sh 脚本对目标参考基因组、任务数据的存放位置、输入输出参数进行指定。

4) 差异性表达分析: 参数设置完毕后, 通过 pipeline.sh 脚本触发最佳实践分析流程, 分别进行数据预处理、序列比对、转录本组装、转录本合并、差异表达分析等核心处理步骤, 最终实现转录组差异性表达的自动化分析, 以更直观、专业的热图方式描述差异基因。

5) 结果反馈: 将结果文件及其热图通过 Web 图形化界面呈现给用户。同时, 为保证研究人员可进行后续研究与分析, 通过 pull\_output.sh 脚本对核心结果文件进行提取并以下载列表的形式反馈至用户, 方便下载和进行二次处理分析。

以上各环节均存在检错、纠错机制, 当某一功能环节检测报错后会选取最近点任务进行断点恢复, 这种实时的状态检测与纠错可有效避免任务的中断和失败。在处理过程中, 各脚本均会对上一步骤执行状态进行检测, 并根据不同的返回值进行相应的操作。若返回值为 1, 表明上一步骤执行异常, 存在错误, 程序自动进入纠错环节, 将上一步骤产生的结果文件删除后重新执行该步骤, 运行完毕会再次检测, 检测通过后才可执行下一处理步骤。若返回值为 0, 则表示上一步骤执行成功, 可以顺序执行下一处理步骤。

本研究向用户提供两种服务界面: 1) 用户图

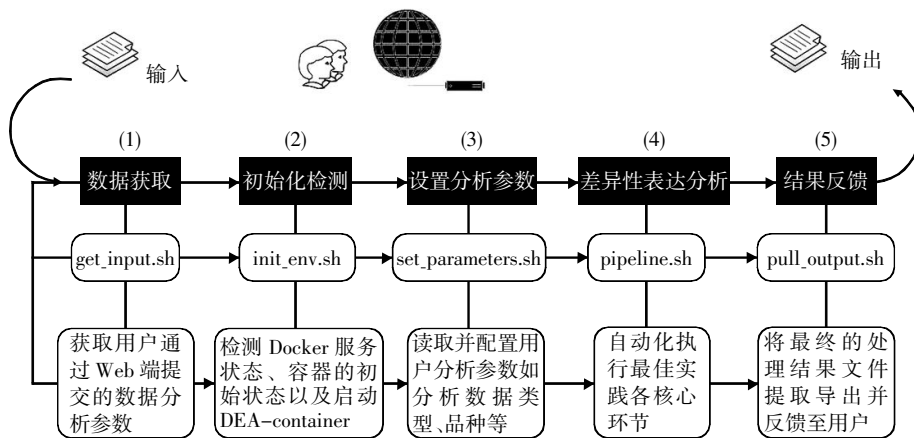


图 3 多脚本联动过程

Fig.3 The process of scripts linkage

形界面: 该方式仅需用户将 sra 格式的原始序列通过 Web 用户图形界面进行提交, 所需的计算资源由该平台的后台服务器提供, 全过程无需用户干预, 处理完毕后分析结果会通过 Web 界面反馈至用户并供用户下载; 2) 命令行界面: 本研究发布了 DEA-container 以及相关的一键化部署脚本, 用户将其下载至本地并通过一键化部署脚本可在任意物理主机基于命令行方式进行特定的数据分析。二者相比, 前者更适合于一些经验较少的生物研究人员或非生物信息学研究者进行通用常规的分析, 后者更适合于具备丰富的数据分析经验且需进行特定参数分析的研究人员。

### 3 实验结果与性能分析

为多方面验证 DEA-container 性能, 本研究采用蒙古羊亚系小尾寒羊的高通量测序数据, 探讨 *FSHR* 基因在不同繁殖力的小尾寒羊中的差异性表达。我们分别选取 SRR5786000、SRR5786001、SRR5786002、SRR5786003 双端数据(共计 59.4 GB)作为样本数据集进行分析。样本数据集的详细信息如表 3 所示。

表 3 样本数据集  
Table 3 List of test samples

序列 Runs	长度/bp Length/bp	大小/GB Size/GB	类型 Type	下载 Download
SRR5786000	21 742 494	14.7	PAIRED	NCBI
SRR5786001	22 071 999	14.7	PAIRED	NCBI
SRR5786002	21 904 607	15.3	PAIRED	NCBI
SRR5786003	21 365 214	14.7	PAIRED	NCBI

#### 3.1 准确性验证

准确性是评估一切方法框架最基本的前提。为验证 DEA-container 框架分析结果的准确性, 首先采用两种方式对同一原始数据进行处理分析, 即传统人工方式和 DEA-container 框架全自动方式; 随后将两种方式的结果数据进行收集与对比, 通过验证结果文件的一致性与完整性来确保数据的准确性。文中主要选取的核心输出文件有 TopHat2 结果文件: accepted\_hits.bam、Cuffmerge

结果文件: merged.gtf、Cuffnorm 结果文件: genes.fpkm\_table、Cuffdiff 结果文件: isoforms.fpkm\_tracking。验证方法如图 4 所示, 结果表明不同处理方式所产生的结果文件内容完全一致, 证明 DEA-container 框架具有准确性。

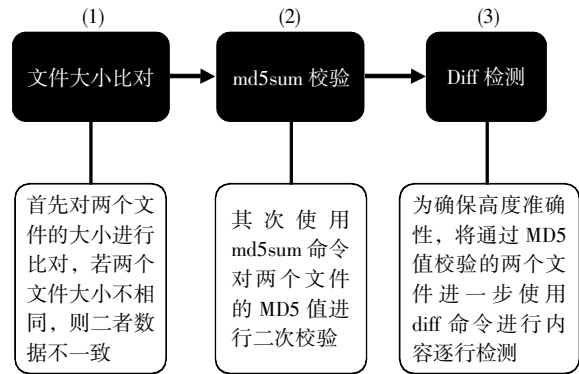


图 4 准确性验证方法  
Fig.4 The method for verifying accuracy

#### 3.2 高效性验证

为测试 DEA-container 的处理效率, 本研究一方面在实验室的惠普工作站(16 GB RAM, 4 CPU Intel Xeon)实验平台对小尾寒羊中目标基因的差异性表达分析任务通过 Web 端进行 6 次提交和执行; 另一方面, 采用传统人工方式对相同数据集进行人工的分析处理。我们分别记录了不同处理方式下每个完整分析周期的总运行时间并求取平均值, 结果如表 4、图 5 所示。

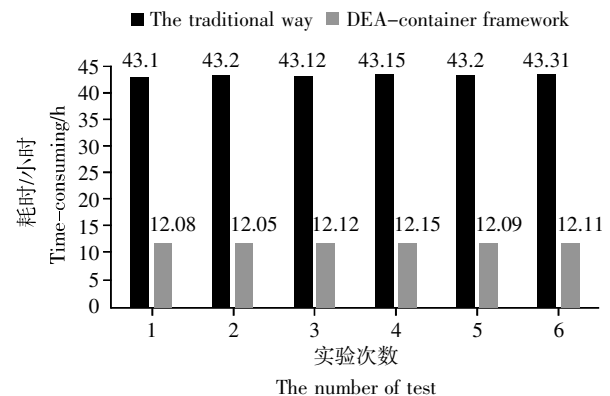


图 5 高效性对比验证  
Fig.5 The comparison of the efficiency

表 4 处理时间统计表  
Table 4 Statistics of the processing time

方式 Method	配置信息 Configuration	样本 Sample	平均时间 Average time
传统人工方式	16 GB RAM, 4 CPU	59.4 GB	43:18:00
DEA-container 框架	16 GB RAM, 4 CPU	59.4 GB	12:06:00

在传统人工方式处理过程中, TopHat2 耗时约 26 h, Cufflinks 耗时约 9.9 h, Cuffnorm、Cuffquant、Cuffdiff 总耗时约 7.4 h, 一个完整周期约耗时 43 h 18 min。其处理效率差, 耗时长, 需研究人员手动部署和执行各个处理步骤, 工作繁琐且难度大易出错, 在大数据时代下该方式不可取<sup>[13]</sup>。文中采用 DEA-container 框架全自动化方式进行的 6 次重复测试均实现了转录组差异性表达分析全流程的自动化执行, 对 6 次执行效率进行计算后取平均值, 得到一次完整的分析周期约为 12 h 6 min。从以上分析可知, DEA-container 的处理效率与传统人工方式相比提升了 2 倍多, 分析时间缩短约 72%, 很大程度提升了分析效率, 在有效降低成本的同时保证了数据分析的高效性。

### 3.3 部署时间对比验证

DEA-container 的突出优势为一次封装、随处运行、可跨平台、不依赖任何编程语言。为了验证这些优势, 本研究采用两种方式, 即 DEA-container 一键化脚本部署方式和传统人工方式, 将转录组差异性表达分析最佳实践流程部署至 11 台主机上。但由于传统人工方式基于 Linux 的命令行, 使大量的参数配置以及复杂的命令行操作均受到有无 Linux 基础的限制, 导致部署时间受到较大的影响, 因此我们分别选取 Linux 技能熟练程度不同的人员进行操作, 对两种方式的部署时间进行记录, 最后求取平均值, 结果如图 6 所示。

由图 6 可知, 传统人工方式的部署时间受 Linux 技能熟练程度影响较大, 其完整部署时间平均约为 26.09 h, 大量时间耗费在手动的环境配置与软件的部署。与传统人工方式不同, DEA-container 一键化脚本部署与研究人员 Linux 技能熟练程度无关, 该方式将手动操作的全过程封装且自动化, 完整的脚本执行时间平均约 26.7 min, 高效而自动化, 完全解决了软件的环境配置难、安装难等问题, 成功地避免了研究人员重复繁琐

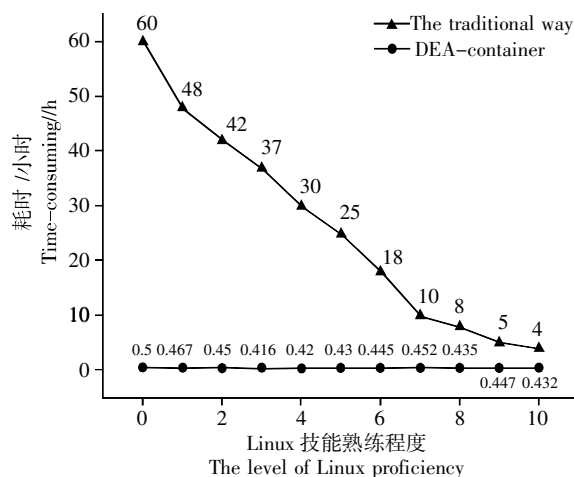


图 6 部署时间对比验证

Fig.6 The time comparison for deployment

的手动操作, 很大程度上减少了用户工作量、降低了用户工作难度, 有效提升了分析效率。同时, DEA-container 一键化脚本部署方式降低了对生物研究人员计算机专业技能的要求, 更为简单易用。

### 3.4 实验小结

两种处理方式的比对结果如表 5 所示, 与传统人工方式相比, DEA-container 方式具有如下优势: 1) 分析效率显著提升, 是传统人工方式的 2 倍多, 分析时间缩短约 72%; 2) 具备较好的移植性, 可跨平台随处运行; 3) 环境部署快速便捷, 仅需执行一键化部署脚本即可; 4) 易扩展, 可按需集成多种分析流程与工具软件; 5) 将复杂流程全面封装, 极大简化了用户操作。

## 4 结语

本文针对转录组差异性表达分析复杂流程形成了一个简单易用的框架, 用户只需要通过 Web 界面对目标分析序列进行提交, 无需做其他额外操作, 而且后台处理分析完毕后会直接给出分析结果供用户下载, 极大地简化了操作流程, 提高了分析效率, 使得研究人员可以更好地进行数据

表 5 全局对比  
Table 5 The global contrast

方式 Method	样本数据 Sample	准确性 Accuracy	分析周期 Cycle	耗时 Time-consuming	可移植性 Portability	按需集成 On-demand integration
传统人工方式	小尾寒羊 59.4GB	100%	43:18:00	26.09 h	不可移植, 仅对当前配置环境有效	可按需集成, 但较为耗时, 需具备计算机基础的研究人员进行多次繁琐的手动配置操作
DEA-container 方式	小尾寒羊 59.4GB	100%	12:06:00	26.7 min	移植性较好, 仅需配置一次, 封装后随处运行	可按需集成, 仅需一次封装, 安装部署快速便捷

分析工作,从海量数据中得到更可靠并具有生物学意义的结果。

目前,我们已将转录组差异性表达分析容器及相关内容公布发放至 GitHub (<https://github.com/fullblossom/the-Differential-expression-analysis-container>),可供用户下载使用。在后续的工作中我们将按需求形成多种分析容器,并将其公布至平台供用户下载和使用,以期为生物学研究者提供简单易用且高效的分析工具。

#### 参考文献(References):

- [1] PERTEA M, KIM D, PERTEA G M, *et al.* Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown[J]. *Nature Protocols*, 2016, 11(9): 1650-1667.
- [2] ZHANG L, LIU X J. A comprehensive review on RNA-seq data analysis[J]. *Transactions of Nanjing University of Aeronautics and Astronautics*, 2016, 33(3): 339-361.
- [3] H BACKMAN T W, GIRKE T. systemPipeR: NGS workflow and report generation environment[J]. *BioMed Central Bioinformatics*, 2016, 17: 388.
- [4] WESTREICH S T, KORF I, MILLS D A, *et al.* SAMSA: a comprehensive metatranscriptome analysis pipeline[J]. *BioMed Central Bioinformatics*, 2016, 17: 399.
- [5] CHAVES I, COSTA B V, RODRIGUES A S, *et al.* miRPursuit: a pipeline for automated analyses of small RNAs in model and nonmodel plants[J]. *FEBS Letters*, 2017, 591(15): 2261-2268.
- [6] MA T L, ZHANG A D. Omics informatics: from scattered individual software tools to integrated workflow management systems[J]. *IEEE/ACM Transactions on Computational Biology & Bioinformatics*, 2017, 14(4): 926-946.
- [7] GHOSH S, CHAN C K K. Analysis of RNA-seq data using TopHat and Cufflinks[J]. *Plant Bioinformatics*, 2016, 1374: 339-361.
- [8] LANGMEAD B, SALZBERG S L. Fast gapped-read alignment with Bowtie2[J]. *Nature Methods*, 2012, 9(4): 357-359.
- [9] KIM D, PERTEA G, TRAPNELL C, *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions[J]. *Genome Biology*, 2013, 14(4): R36.
- [10] TRAPNELL C, ROBERTS A, GOFF L, *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks[J]. *Nature Protocols*, 2012, 7: 562-578.
- [11] GOFF L, TRAPNELL C, KELLEY D. cummeRbund: visualization and exploration of Cufflinks high-throughput sequencing data [EB/OL]. <http://www.bioconductor.org/packages/release/bioc/html/cummeRbund.html>, 2012.
- [12] BOLGER A M, LOHSE M, USADEL B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. *Bioinformatics*, 2014, 30(15): 2114-2120.
- [13] 王俊,郭丽,吴建盛,等. 大数据背景下的生物信息学研究现状[J]. *南京邮电大学学报(自然科学版)* (WANG Jun, GUO Li, WU Jian-sheng, *et al.* The present situation of bioinformatics research under the background of big data[J]. *Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition)*), 2017, 37(4): 62-67.

## 撤稿声明

### 条件性敲除 *APC* 基因小鼠肠道腺瘤模型的构建

廖超男<sup>1a</sup>, 杨治平<sup>1b</sup>, 林良武<sup>1c</sup>, 杨智英<sup>2</sup>, 蔡金杏<sup>1a</sup>, 熊璐<sup>1a</sup>, 黄河<sup>1a\*</sup>

(1. 中南大学 a. 基础医学院, b. 湘雅三医院, c. 粉末冶金研究院, 中国湖南长沙 410013; 2. 长沙卫生职业学院, 中国湖南长沙 410100)

原始文章发表于《生命科学研究》杂志 2016, 20(5): 424-428。

DOI: 10.16605/j.cnki.1007-7847.2016.05.009

经调查,该文在严谨性、规范性方面存在缺陷,文章撰写人提出撤稿。根据本刊的撤稿规定,认为作者的申请符合相关条款,同意撤稿。

特此告知,并诚挚地向读者致歉。

《生命科学研究》编辑部

2019年2月20日