



袁晓龙,张哲,张豪,等.转录组学在动物科学类专业生物信息学教学中的探索及实践[J].黑龙江农业科学,2020(10):117-119.

转录组学在动物科学类专业生物信息学教学中的探索及实践

袁晓龙,张 哲,张 豪,李加琪

(华南农业大学 动物科学学院/广东省农业动物基因组学与分子育种重点实验室,广东 广州 510642)

摘要:生物信息学作为生命科学领域的前沿学科,其教学尚处于起步阶段。转录组测序已有替代传统实时定量 PCR 的趋势,已成为畜牧科学家开展科研工作的常用手段和工具。大规模的转录组数据分析对计算环境的硬件和软件提出了更高的要求,给生物信息学的教学带来了难题。本文结合农业院校动物科学类专业特点和生物信息学转录组课程的教学实践,利用计算机集群和网络公共数据,以锻炼学生的科研能力为导向,开展转录组课程的个性化实践教学,旨在培养学生独立思考和解决问题的能力,提高学生适应新时代动物科学研究的能力。

关键词:转录组;生物信息学;计算机集群;探索与实践

近年来,人类基因组计划的实施,推动了生命科学技术的巨大进步。随着测序成本的大幅下降,已经产生了海量生物基因组数据,亟需强有力的工具挖掘有意义的生物信息,帮助解决基础生命科学问题。这种需求催生了挖掘有用和有意義的生物信息的学科,即生物信息学。生物信息学是指利用数学和统计分析方法,借助高性能计算机,研究生命科学大规模数据的学科,是生命科学、数学、统计学和计算机科学相互交叉的一门学科^[1]。当前,生物信息学已经成为生命科学领域最具活力、最前沿的学科之一,它涉及生物信息的采集、存储、传递、分析和解读,研究内容涵盖了大规模的基因组、转录组、蛋白质组、表观遗传组、微生物组、代谢组等多个组学层次。

转录组数据主要是指围绕着遗传学“中心法则”,利用高通量测序,检测基因组转录后的数据。转录组学数据真实反映了生物体内的生物学信息在转录水平的动态变化,强有力地促进了基础生命科学的发展。与其他学科相比,大规模的转录组数据分析需要借助 Linux 操作系统和 Shell 编程语言,对计算环境的硬件和软件提出了更高的

要求,给生物信息学的教学带来了难题。在畜牧学科中,转录组测序已成为畜牧科学家开展科研工作的常用手段和工具,逐渐有替代传统实时定量 PCR 的趋势。借助转录组测序不仅能挖掘影响动物性状的重要基因、加强动物多样性的保护,还能加快动物育种的选择和选育,具有教学和科研一体化的潜力^[2],是动物科学类专业研究生需要掌握的基本知识和技能。

生物信息学作为一门新兴、前沿的学科,其教学尚处于起步阶段,没有完整的教学模式和教学方法,教学内容繁杂,没有固定的教材。华南农业大学动物科学学院为硕士研究生新生开设了生物信息学课程,但市场上无针对动物科学专业学生的教材,给动物科学专业生物信息学的教学带来了难题。在“双创”背景下,本文结合农业院校动物科学类专业特点和生物信息学转录组课程的教学实践,对生物信息学转录组课程的现状和问题进行仔细分析和思考,以期能为生物信息学课程的教学探索提供一定的帮助。

1 转录组数据的主要分析流程

当前,基于高通量测序技术能够快速获得 mRNAs、miRNAs、lncRNAs 和 circRNAs 等转录水平的组学数据。在严格实验设计和充足的生物学重复下,采集相应的样品,建立测序文库和上机测序,下机数据通过去除低质量数据,把测序数据比对到参考基因组,比对后的数据进行组装和计算相应的表达水平,根据表达水平计算组内和组

收稿日期:2020-05-27

基金项目:广东省普通高校青年科技创新人才项目(2018 KQNCX019)。

第一作者:袁晓龙(1989-),男,博士,副教授,硕导,从事动物遗传育种方向的教学和科研工作。E-mail: yxl@scau.edu.cn。

通信作者:李加琪(1965-),男,博士,教授,博导,从事动物遗传育种方向的教学和科研工作。E-mail:jqli@scau.edu.cn。

间重复性,并计算组内差异表达的基因和转录本,然后对分析结果进行展示^[3]。根据数据分析流程,主要涉及三大类软件:一是数据比对前的处理,主要是利用 FastQC、Cutadapt、FASTX-Toolkit 或 PRINSEQ lite 等软件对测序数据进行质量控制和检测;二是比对-定量-检测,主要是利用 Tophat、STAR、Bowtie2、HISAT 和 Subread 等软件把测序数据比对到参考基因组,利用 Cufflink、StringTie、Feature Counts、RSEM 和 eXpress 等软件将比对后的数据进行组装和定量,利用 Cuffdiff、DESeq2、edgeR 和 Ballgown 等软件筛选差异表达的基因;三是对结果进行分析和图示化,主要是利用 Samtools、Bigwig、R、Heatmap、PCA、GO 和 KEGG 等软件和数据库对数据进行分析和图示化。

2 转录组在动物科学类专业生物信息学教学中面临的主要难题

一是动物科学类专业的学生在本科阶段与生物信息学有关的课程较少。本科阶段与生物信息学相关的课程只有《动物遗传学》《生物统计学》和《家畜育种学》,刚入学的硕士研究生,对数据分析过程的抽象内容难以理解。转录组学数据量较大,只能借助高性能计算机处理数据,而学生对计算机编程语言、Linux 操作系统、计算软件接触较少,难以实现对转录组数据的存储、检索、运行和计算分析。二是转录组数据分析依赖高性能的硬件和软件。转录组在分析的过程中产生大量的中间数据,数据的处理对运行环境的存储和运行空间都提出了更高的要求。普通的计算机很难完成转录组数据的处理,需要借助高存储和运行空间的计算机。三是生物信息学相关数据库和软件资源丰富、更新频率高。学生对各种生物学数据库不了解,难以从数据库里提取有效的信息加以利用。例如,美国国立生物技术信息中心 NCBI 数据库、核酸序列数据库 EMBL 和 DDBJ 数据库及 UCSC 数据库,这些数据库包含数以亿计的生物信息,并且实时更新。初次接触这些数据库,眼花缭乱,不能准确地进行检索^[4-5]。四是生物信息学的实训课较多。为保证教学效果,教师课前需对实训项目进行设计和模拟^[6],这样增加了教师的工作量。另外,用于分析转录组数据的软件和参考的文献资料多为英文环境,学生在学习生物信息学过程中感觉吃力、寸步难行,使学生的学习兴趣下降。

3 基于计算机集群的转录组分析教学实践

3.1 计算机集群的配置和登录

现华南农业大学动物科学学院建有生物信息学教学课堂,搭配 65 台 PC 计算机、一套多媒体教学设备和一个计算机集群,计算机集群上搭建了各种用于转录组数据分析的软件。使用多媒体技术,直观和生动地阐述抽象的知识点,能提高学生的兴趣。学生可以通过客户端远程登陆,在集群上建立自己的文件夹,利用搭建好的流程,从 NCBI、UCSC 或者 EMBL 下载目标动物的全基因组序列,根据教学内容,在 Linux 环境下一步一步地进行操作和实践,对候选的基因进行 KEGG 和 GO 的信号通路和基因功能富集分析,结合自己的实验设计和科学问题对结果进行解读和总结。借助生物信息学课堂平台,学生可以在校园内,不受时间和地点的限制登录计算机集群,完成课后实训作业,提高了学习效率^[6-7]。

3.2 科研导向的项目式教学

围绕着畜牧学中的常见科学问题,例如筛选影响性状和表型的候选基因、研究某条件刺激下的响应机制及影响某发育过程的候选基因,教师通过网络公共数据库,下载相关的文献和原始数据。通过板书和 PPT 讲解的方式,让学生充分理解、知晓文献数据的来源,包括实验设计、测序流程、数据质控、上游数据分析。把 3~5 个学生划分为一个团队,给予每个团队不同的科学问题及相应的数据^[8],学生借助生物信息学课堂的 PC 计算机或自己的计算机,通过客户端登陆计算机集群,参照授课内容的数据分析流程和参数设置,课后分析数据和结果^[9]。带着科学问题进行数据分析操作,大大提高了学习的积极性^[6]。学生自己对数据分析流程和结果进行总结,通过 PPT 方式在课堂上展示和讨论。教师带着学生对每个团队的结果进行分析和主导讨论,并与文献中的结果相比较。这种“参与式”教学的方式^[10],从实验设计、数据分析、结果总结到回答科学问题,每个学生都有参与,学习效果良好。不仅让学生更深层次地了解转录组数据分析原理和意义,锻炼科学和严谨的实验设计思维,还能提高学生对结果进行分析和总结的能力。

4 教学方法和模式讨论

首先,以进一步提升学生的科研潜力为导向,

对教学内容进行改革,加强实践环节。注重培养学生独立思考、独立解决问题、独立操作的能力,拓宽学生的知识面;其次,课程内容设置为基本理论课和实践课,避免灌输转录组测序的理论和方法,避免过度关注转录组数据的结果,忽视了对科学问题、实验设计、数据来源及分析过程的讲解,导致学生对生物信息学的学习产生了厌倦心理;再次,鼓励学生参加学校各类创新和创业大赛^[11],围绕着感兴趣的科学问题,自行设计科学实验,独立进行操作、分析和解决数据,帮助他们根据实验结果发表高水平的国际论文,提高学生的积极性^[12];然后,授课教师结合实际项目和科学问题进行授课,能充分调动学生的积极性,全员行动和全员参与,在项目开展的同时,能让学生感受到获得知识的幸福感;最后,检索每年国际著名杂志发表的转录组学相关文献,补充到理论课中,让学生了解生物信息学最新的发展方向^[13]。

5 结语

在大数据时代,是否拥有数据分析能力直接影响着学生就业。授课教师借助网络资源,开展项目式教学,以科学问题为导向,以锻炼学生的科研思维和能力为目标,培养学生分析和解决问题的能力,提高学生的学习积极性、学习能力和动手能力,鼓励学生利用生物信息学的手段解决实际科研问题,为畜牧业的发展贡献力量。生物信息学具有极强的专业性和应用前景,但大多数教学内容抽象、知识点更新频繁、学生基础薄弱,在今后的教学工作中,需要改进教学内容、教学方法和教学手段,不断总结经验、摸索新方法。

参考文献:

- [1] 陈志雄,王少奎,李亚娟.科研导向的农学类专业生物信息学教学内容改革[J].安徽农业科学,2018,46(32):240-242.
- [2] 朱明库.教学科研一体化模式在生物信息学课程教学中的应用探究[J].中国多媒体与网络教学学报(上旬刊),2018(8):58-59.
- [3] Conesa A, Madrigal P, Tarazona S, et al. A survey of best practices for RNA-seq data analysis[J]. Genome Biology, 2016,17(1):181.
- [4] 张渝洁,邢晋祎.生物信息学实验教学中的网络资源及其利用[J].安徽农业科学,2019(11):276-278.
- [5] 文可佳.生物信息学实验教学中如何巧用数据库[J].中小学数字化教学,2018(2):10-13.
- [6] 马磊,张婷婷.以实践为主的生物信息学教学改革及成效[J].教育现代化,2019(7):32-33,36.
- [7] 刘艳,吕爱平,唐凯临,等. SPOC 模式生物信息实验教学云平台设计与实践[J].实验室科学,2018,21(1):99-102.
- [8] 徐鑫.导师定制的个性化生物信息学课程教学研究[J].安徽农业科学,2019,47(15):274-275,277.
- [9] 马立飞,张立娟,吴婷婷,等.基于翻转课堂的项目式教学模式在生物信息学课程中的教学探索[J].河北北方学院学报(社会科学版),2019(2):114-116.
- [10] 何勇涛,金晓琳,倪青山,等.参与式微生物组在生物信息学教学中的实践研究[J].重庆医学,2019,48(7):1245-1247.
- [11] 范丙友,贾小平,郭丽丽,等.“双创”教改背景下生物信息学课程考核方式改革与实践[J].大学教育,2018(11):101-103.
- [12] 徐后喜,傅淑平.中医药院校本科生生物信息学教学改革研究与实践[J].中国农村教育,2019(20):35-37.
- [13] 马光.网络学习在生物信息学教学改革中的重要性探析[J].现代农村科技,2019(8):67-68.

Exploration and Practice of Transcriptomics in Bioinformatics Teaching of Animal Science

YUAN Xiao-long, ZHANG Zhe, ZHANG Hao, LI Jia-qi

(School of Animal Science, South China Agricultural University, Key Laboratory of Agricultural Animal Genomics and Molecular Breeding of Guangdong Province, Guangzhou 510642, China)

Abstract: Bioinformatics, as a frontier subject in the field of life science, is still in its infancy. Transcriptome sequencing has replaced the traditional real-time quantitative PCR, and has become a common means and tool for animal scientists to carry out scientific research. The large-scale transcriptome data analysis puts forward higher requirements for the hardware and software of computing environment, which brings difficulties to the teaching of bioinformatics. In this paper, combined with the characteristics of animal science specialty in agricultural colleges and the teaching practice of bioinformatics transcriptome course, using computer cluster and network public data, aiming at training students' scientific research ability, personalized practical teaching of transcriptome course was carried out, aiming to cultivate students' ability of independent thinking and problem-solving, and improved their ability to adapt to the new era of animal science research.

Keywords: transcriptome; bioinformatics; computer cluster; exploration and practice