

生物信息学在医学微生物领域研究生教学中的应用与思考

蒋琰 伍雪晴 俞云松

[摘要] **目的** 探讨生物信息学在医学微生物领域研究生教学中的应用与思考。**方法** 总结十余年的教学经验,通过整理医学微生物领域实例,建立生物信息学分阶段教学的方法。**结果** 生物信息学主要运用计算机科学、数学的思路和方法并结合生物学工具来阐明和理解大量数据中包括的生物学意义。医学微生物领域近些年在科研上发展很快,十分依赖生物信息学对其大数据的解析和处理工作。在医学微生物背景的研究生教学过程中,生物信息学的学习可以分为三个阶段,一是对概念看得懂、应用懂得用的基本概念阶段;二是会搭建平台、会灵活运用进阶阶段;三是会编程、懂算法的高级阶段。同时,需要鼓励学生在在学习过程中注重实践和应用,并积极阅读原文文献,跟上最新的学科发展步伐。**结论** 通过借助这些方法和经验,在实际教学过程中帮助学生走出生物信息学学习的误区,打好学习的基础,能够将学科用于真正的科研实践当中。

[关键词] 生物信息学; 教学; 医学微生物

Application and thinking of bioinformatics teaching in graduate students who had a medical microbiology background JIANG Yan, WU Xueqing, YU Yunsong. Department of Infectious Diseases, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310016, China.

[Abstract] **Objective** To explore the application and thinking of bioinformatics teaching in graduate students who had a medical microbiology background. **Methods** Based on teaching experience in last decade, multi-stage teaching pattern can be constructed through lots of instance on medical microbiology. **Results** In bioinformatics, many thoughts and methods based on computer science and mathematics combined with biological tools were employed to elucidate and understand the biology interpretation within big data. Research of medical microbiology developed rapidly in recent years and relied on the data analysis by bioinformatics tools. During the bioinformatics teaching in graduate students who had a medical microbiology background, the curriculum can be divided into three stages. Firstly, the fundamental concept stage requires students to understand the biology concept and how to use the web tools. Secondly, the promotion stage requests constructing the data analysis platform or applying the command-line tools. Lastly, the superior stage asks for understanding the computer programming and algorithm. Moreover, practice and application should be encouraged for students and reading original research papers can be a good method to acquire the new scientific information. **Conclusion** These methods and experiences around bioinformatics teaching can help graduate students get rid of the misunderstandings and lay a solid study foundation, utilizing the knowledge for real science and research.

[Key words] bioinformatics; teaching; medical microbiology

生物信息学是一门由计算机科学、数学和生物

DOI: 10.13558/j.cnki.issn1672-3686.2022.006.016

基金项目:浙江省自然科学基金项目(LY22H190001)

作者单位:310016 浙江杭州,浙江大学医学院附属邵逸夫医院感染科

通讯作者:俞云松, Email: yvys119@zju.edu.cn

学综合而成的交叉学科,它包含了对生物学信息数据的获取、加工、存储、分配、分析、解释等方面,运用计算机科学、数学的思路和方法并结合生物学工具来阐明和理解大量数据中包括的生物学意义^[1]。对于生物信息学来说,生物学内容是研究的目标、实体和对象,信息科学是研究的平台、技术和手段。

生物信息学作为一门新兴的交叉学科,是近二十年科学领域发展最迅速的学科之一。一方面得益于生物学领域在研究手段和研究技术上的极大发展,特别是组学技术的不断进步,使获得的生物学数据量呈指数级别的上升,海量的基因组、转录组、蛋白质组数据亟需高效的分析能力支持;另一方面,依托于信息科学特别是计算机科学技术水平的迅速提高,大数据、复杂数据不再成为计算机处理的难题,而机器学习、人工智能等最新的计算机算法使得人们可以借助计算机技术进一步地认识、挖掘生物学数据。

生物信息学的应用几乎覆盖了生物学所有的相关学科,而与人类健康息息相关的生物医学领域,生物信息学更显得极为重要。近些年来,国家医学院校都非常重视生物信息学课程的教学工作,大多数学校在研究生阶段都开展了相应课程^[2,3]。但是,生物信息学学科发展的时间还很短,因此在教学上还面临着很多的困难,一方面是师资力量不足,科班出身的生物信息学教师非常少,很多学校生物信息学授课老师自己也处在边学习边教学的阶段;另一方面,生物信息学学习对计算机基础有较高的要求,医学背景的学生大多缺乏扎实的计算机背景,因此学生普遍会产生畏难心理,大大影响了对生物信息学的学习效果^[4]。本文针对生物信息学教学在医学微生物领域中的现状,利用近十年来研究生教育的经验和思考,探讨生物信息学学科在医学微生物领域研究生教学中的应用与实践,为教师和学生提供教学上和学习上的指导和帮助。

1 医学微生物领域对生物信息学学科的需求

医学微生物学又常称为临床微生物学,是一门医学的基础学科,主要研究与医学有关的病原微生物的生物学性状、传染致病的机理、免疫学的基本理论、诊断技术和特异性防治措施等,以达到控制和消灭传染性疾病和与微生物有关的免疫性疾病,保障人类健康的目的。医学微生物学是最早将生物信息学技术应用于学科发展的领域之一,可以说是生物信息学技术革新的前沿阵地。微生物相对真核生物来说,其生物学结构、基因构成、基因组序列等等相对简单,因此生物信息学早期发展的一些数据处理方法都是在微生物上最早完成的,例如人类完成的第一个生命体基因组全序列分析,即是在1995年刊登在《Science》杂志上的流感嗜血杆菌全基因组测序和分析^[5]。

医学微生物涉及到的医学学科很广,包括了临床检验诊断学、传染病学、呼吸病学、重症医学等等,而医学微生物学所研究的主要对象是病原微生物,因此涉及到病原微生物的鉴定、病原菌的耐药性和致病性、病原菌与宿主相互作用等研究方向都需要用到生物信息学的技术手段。

2 生物信息学在医学微生物学领域的应用

微生物具有最简单、最基本的生命形式,因此在生命领域的研究历史当中,一直是作为最早研究的对象。人类完成的第一个生命体基因组测序是流感嗜血杆菌,生物信息很多技术手段,都是在微生物的物种上率先实践。从基因组的角度来看,微生物具有最小、最简单的基因组构成,常见的细菌基因组大小只有1~10 Mbp,多为一条环状闭合DNA,基因组上大部分序列均编码基因,基因长度在1 kbp左右,占整个基因组的大概90%。真菌的基因组大小一般也不超过20 Mbp,而病毒的基因组就更小了,通常只有几kbp到几十kbp。生物信息学技术的发展使其成为医学微生物领域研究不可或缺的重要手段,高通量基因组测序的普及,带来了大量的微生物基因组序列分析、统计分析、大数据挖掘、进化分析需求,这些都离不开生物信息学技术平台。除了基因组学,转录组、蛋白组、代谢组等各种组学也发展迅速。而目前医学微生物领域的研究热点更离不开生物信息技术的支持,例如病原菌分子流行病学调查关注大规模的基于基因组序列的分子进化分析,需要基因组学大数据的联配、同源性和统计分析^[6];病原菌的耐药性、致病性研究,需要基因突变、转录水平、酶动力学等多方面基因组、转录组和蛋白质组关联分析^[7];病原菌与宿主的相互作用研究,则需要蛋白质筛查、转录代谢调控通路的分析和图形展示^[8]。以上这些研究热点,都可以作为医学微生物领域研究生生物信息课题的教学示范以及实践练习内容^[9]。

3 生物信息学分阶段教学

生物信息学普遍被认为是研究生阶段比较难学的课程之一,对于医学背景的研究生来说,计算机的基础和应用能力往往不足,除此之外,很多学生还容易走进学习生物信息学的几个误区:一个是基本概念混杂、概念不清,只会简单应用,不能融会贯通;二是对基本概念涉猎范围广,但仅停留在书本表面,实际应用能力差;三是盲目学习新概念,追求热点,遇难而退从而导致半途而废。

基于对生物信息学研究生阶段教学的十余年经验,根据学生背景和基础情况,进行分阶段教学是走出这些误区比较合理的方式。以医学微生物学领域为例,将生物信息学教学分为三个阶段,一是基本概念阶段,看得懂概念,会懂得应用;二是进阶阶段,会搭建平台,会灵活运用;三是高级阶段,会编程,懂算法。

3.1 基本概念阶段 基本概念阶段要求所有学生都必须掌握,主要是对一些基本概念的理解和对简单应用的实践。在医学微生物领域,主要的研究对象是病原微生物,因此,对微生物的生物学概念要有比较明确的认识,并结合生物数据的分析需求来理解这些生物学概念。在这个阶段要求学生能够掌握利用网页工具解决基本的生物信息数据分析工作。生物信息学发展到现在,实际上很多基本的数据分析工具都已经实现网页化,利用网络页面的工具再结合网站背后的数据库,可以做很多基本的分析工作。例如,在学习DNA序列的同源性比对内容时,要学会利用美国国立生物信息网站(NCBI, <https://www.ncbi.nlm.nih.gov/>)的Blast工具在数据库中进行同源序列的检索,也要对比对结果的概念和意义进行解释。又如,对病原细菌基因组序列的耐药基因识别和序列型分析,可以使用丹麦科技大学开发的Center for Genomic Epidemiology (CGE, <http://www.genomicepidemiology.org/>)网站进行在线分析,只要通过网页递交基因组序列,可以迅速地获得耐药基因和序列型列表,结合对生物学概念的解释,可以高效率地实现教学目标。

3.2 进阶阶段 进阶阶段要求学生掌握一些命令行形式和运用代码的基本技能,生物信息学的一些高级运行程序往往是在Linux环境下编写的开源代码,一般要求学会搭建本地服务器,在服务器上利用命令行形式运用程序。建议授课在专用的计算机教室进行,教师首先搭建好可以本地运行的小型服务器,采用多终端的方式让学生登陆,进行边授课边实践的方式。学生首先需要了解Linux操作系统使用规则,熟悉用命令行形式运行程序的方式。对于有条件的学校,可以购买一些商业软件包,例如包含基因组序列拼接、比对和SNP分析功能的CLC genomics workbench软件包,这些软件包有Linux环境下的运行界面,且数据分析功能强大,方便学生的入门学习。另外,要注意指导学生体会命令行形式和网络页面形式在处理相似分析功能时

的使用体验差异,例如可以用NCBI完整的Blast比对页面开展同源序列检索和比对,同样也可以在Linux系统里利用Blast+命令行实现相似功能,而命令行形式可以更灵活地指定比对序列集合、目标序列库以及序列比对结果的格式,更利于下一步通过编程或对接其它软件开展深层次分析。

进阶阶段可以要求学生掌握运用代码进行数据分析和结果展示的能力,一般来说推荐Python和R两种语言进行代码编写,Python一般用于编写灵活的数据转换、数据提取等用途的脚本,所谓脚本是指依据一定语言格式编写的可执行文件,通常以文本保存,只在被调用时进行解释或者编译。编写脚本一般不会涉及太复杂的编码,运用起来灵活方便,特别适合生物信息分析时的分析文件转换、过渡、提取等操作,如果再利用一些开源的功能包,则可以实现更大更复杂的任务。而R语言是世界上流行的数据分析、统计计算和制图软件,几乎能够完成任何数据处理任务,非常胜任生物信息学数据处理过程中对数据统计和结果图形展示的需求。R语言通常需要编程的代码并不复杂,并且可以借助强大的功能包,实现统计和绘图的灵活要求。

3.3 高级阶段 高级阶段则进入了开发生物信息学算法和数据库的层次,可根据学生的兴趣和接受程度灵活要求,一般要求学生具有非常好的数学和计算机基础,适合基础能力比较强的学生,对普通学习需求的学生可不做过高要求。高级阶段的研究人员一般致力于开发新的算法,编写具有普适功能的软件包,来协助普通生物信息分析人员更好地运用工具开展工作。学生在学习分析工具算法的过程中,如果遇到好的想法和思路,可以在原程序的基础上进行算法的优化和调整,来适应新的用途。高级阶段也可以学习尝试搭建生物信息数据库,利用网络平台建设数据库网站,并开发相应的调用数据库数据的程序和工具。

4 生物信息学教学应注重实践和应用

生物信息学是一门注重实践和应用的学科,在教学中特别注意鼓励学生在实际数据分析实践中学习生物信息的基本概念和思路方法。脱离实践直接去学习方法和概念是枯燥的,在教学过程中,需要引入实际工作中的数据,软件和程序也需要边用边学习,往往是在得到一定的分析结果后,从分析结果出发,体会和理解结果中蕴含的概念和意义。生物信息学数据分析步骤往往具有流水线的

特点,从上游到下游需经过一系列的处理,因此挑选合适的实例,可以很方便地指导学生体会数据分析的全过程,加强对整体性的理解。例如,可以从病原细菌全基因组测序的原始序列出发,利用服务器上搭建好的软件平台,从上游的原始序列的质量控制、序列拼接,到基因组的完成图、基因注释,再到目标识别如耐药基因、毒力基因、移动元件等的识别,以及更下游的全基因组同源性分析、绘制基于基因组的遗传进化树等等。一系列的数据分析,便于学生理解和消化,学生既掌握了分析的方法和工具,又会对每个分析步骤中所涉及到的生物学概念和意义有了更深刻的理解。

5 如何适应生物信息学学科迅猛发展的教学要求

回顾生物信息学学科发展的历史,可以说它是近些年来发展最迅速的学科之一,从第一个生命体基因组被解析到现在不到三十年的时间,但学科的认识水平发生了很大的变化。二十多年前DNA序列的测定还处在通量极小、测序极其缓慢的年代,那个时候要完成哪怕一株细菌的全基因组测序可能都要花掉上百万元的费用和超过一整年的时间,而随着近些年来第二代、第三代高通量测序技术平台的成熟,完成一个细菌基因组只需要几天时间和几百元的费用,同时,随着生物学概念的不断延伸,相应的生物信息分析技术也发生了翻天覆地的变化,过去的基因组序列分析方法到现在可能已经完全不适用了。因此,在生物信息学学科教学中,要指导学生学习教育分析方法的思维共性和方式,同时鼓励学生多阅读最新的原文文献,掌握学科发展的前沿,对于最新的数据分析工具或者算法,要勇于尝试。

6 生物信息学教学成效的思考

在实际教学过程中,很多学生反映生物信息学难学,为什么会觉得难学?对非计算机科学背景的学生来说,复杂的代码往往让人有种天然的恐惧感。要帮助学生解决这样的心理问题,可以从两个方面来着手。一是指导学生认识学习生物信息学的几个阶段,即本文前述的基本概念阶段、进阶阶段和高级阶段,让他们充分体会到生物信息学习是分成多个台阶的,要评估好个人目前处在哪一个层次,先把这个层次的基础打好,再努力走上下一个台阶,切不可眼高手低,比如刚入门是就直接跨入

代码编写阶段,这样很容易遭受挫折失去了学习的兴趣。二是要积极通过实践,获得学习的成就感,保持对学科学习的兴趣。如果没有动手亲自处理数据,直接学习那些枯燥的生物信息学名词,容易出现乏味和疲惫,而从实际的数据出发,利用一些成熟的软件往往能画出漂亮的图形,学习的过程很容易就产生成就感,至于怎么理解图形的概念和意义,可以放在其后慢慢钻研。

综上所述,医学微生物领域研究生在生物信息学学习过程中,应该在客观评估自身基础的前提下,充分认识到这门学科的多层次性和应用实践性,而生物信息学教师要努力为学生提供教学上的实用平台,鼓励实践和创新,指导学生在正确的道路上努力学习。

参考文献

- 1 胡杨.《生物信息学》课程教学模式探讨[J].生物信息学, 2018,16(2):72-75.
- 2 李坤平,潘天玲.大数据时代医药类硕士研究生生物信息学通识教育探索与实践[J].卫生职业教育,2018,36(14):5-6.
- 3 杨俊丽,韩竞,叶鸿瑾,等.医学院校生物信息学课程建设与实践[J].基础医学教育,2019,21(2):113-115.
- 4 李曦,颜晗.浅谈医学研究生的生物信息学教学[J].高等教育,2015,10(28):111-112.
- 5 Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd[J]. Science, 1995, 269(5223):496-512.
- 6 David S, Reuter S, Harris SR, et al. Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread[J]. Nat Microbiol, 2019, 4(11):1919-1929.
- 7 Allen JP, Ozer EA, Minasov G, et al. A comparative genomics approach identifies contact-dependent growth inhibition as a virulence determinant[J]. Proc Natl Acad Sci U S A, 2020, 117(12):6811-6821.
- 8 Murray GL, Tsyganov K, Kostoulias XP, et al. Global gene expression profile of *Acinetobacter baumannii* during bacteremia[J]. J Infect Dis, 2017, 215(suppl_1):S52-S57.
- 9 陈艳炯,杨娥,寻萌,等.生物信息学资源及其在医学微生物学教学中应用的体会[J].医学教育研究与实践,2020,28(3):482-486.

(收稿日期 2022-04-07)

(本文编辑 蔡华波)