生物信息学札记（第 4 版）

樊龙江

浙江大学作物科学研究所
浙江大学生物信息学研究所
浙江大学 IBM 生物计算实验室

2017 年 9 月

本材料已由浙江大学出版社出版：《生物信息学》，樊龙江主编，2017

部分内容可通过下列网址获得：
http://ibi.zju.edu.cn/bioinplant/

札记前言

第一版

这份材料是我学习和讲授《生物信息学》课程时的备课笔记，材料大多是根据当时收集的一些外文资料翻译编辑而成。学生在学习过程中经常要求我给他们提供一些中文的讲义或材料，这促使我把我的这份笔记整理并放到网上，供大家参考。要提醒使用者的是，这份材料仅是根据我对生物信息学的一些浮浅的认识整理而成，其中的错误和偏颇只能请读者自鉴了。

2001 年 6 月

第二版

自 1999 年开始接触生物信息学以来，一晃已近六年，而本札记也近四岁了。2001 和 2002 年中国科学院理论物理所的郝柏林院士在浙江大学首次开设生物信息学研究生课程，我作为他的助教系统地学习了生物信息学；同时，借着我国水稻基因组测序计划的机遇，在他的带领下从 2001 年开始从事水稻基因组分析，从此自己便完全投入到这一崭新、引人入胜的领域中来。

不断有来信向我索要本札记的电子版文件，同时在不少网站上看到推荐该札记的内容。生物信息学、基因组学等发展很快，现在再回头审看该札记，有些部分已惨不忍读，这促使我下决心更新它。但因时间和学识问题，还是有不少部分自己不甚满意，就只有待日后再努力了。欢迎告诉我札记中的 BUG，我的信箱 fanlj@zju.edu.cn 或 bioinplant@zju.edu.cn。

2005 年 3 月 30 日

第三版

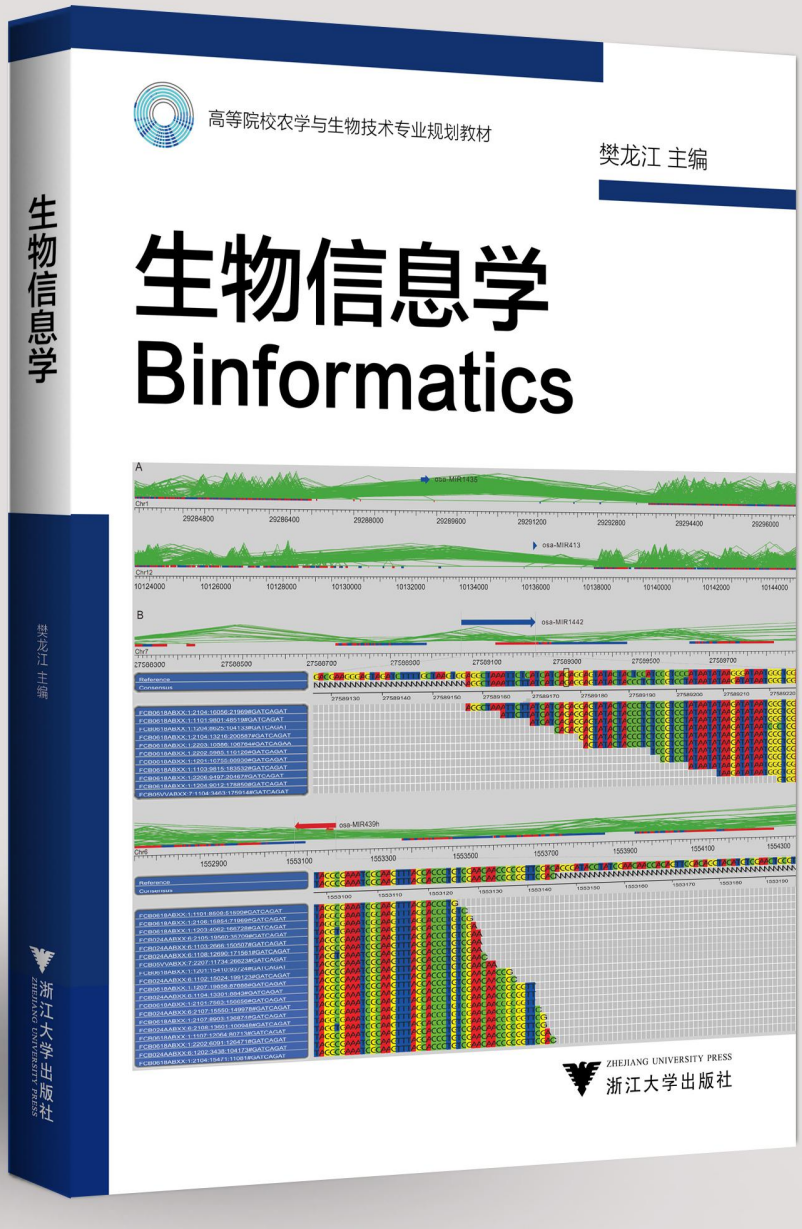
近年来高通量测序技术产生的序列数据大量出现（如小 RNA 和大规模群体 SNP 数据），本次更新根据这一进展增加了两章内容，分别是第七章有关小 RNA 的分析和第八章遗传多态性及正向选择检测。两章内容由我的博士生王煜为主编写，李泽峰和刘云参与了文献整理。另外还更新了第四章有关水稻基因组分析一节。

2010 年 1 月

第四版

2014 年浙江大学开展本科生教材建设工作，我当时作为系主任要带头，就承诺编写我主讲的《生物信息学》教材。编写教材的确不是一件容易的事，经过几番挣扎和多方努力，总算完成了编写，算是了却了一桩心思。该教材内容比较完整，也跟踪了生物信息学领域的最新进展。我就权且把该教材内容作为札记的第四版，也算给该札记一个完美的结尾。

2017 年 9 月


高等院校农学与生物技术专业规划教材
樊龙江 主编
生物信息学
Binformatics
生物信息学
樊龙江 主编
浙江大学出版社
ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

《生物信息学》

前言

自开始接触生物信息学以来，一晃已近二十年了。我是在攻读博士期间开始注意并学习生物信息学的。我的博士生导师胡秉民为应用数学专业教授，主要从事生态系统模型模拟研究。虽然已具备一定数量统计和数量遗传学基础，但当时对于生物信息学，我还是非常陌生的，通过自学才开始一点点了解这门新兴学科。2001-2003 年间，中国科学院理论物理所郝柏林院士在浙江大学首次开设“生物信息学”研究生课程，我作为他的助教，系统地学习了生物信息学；同时，在他的带领下从事水稻基因组分析。自那时起，浙江大学生物信息学学科和相应研究机构也逐步建立起来。2004 年郝院士离开杭州加入复旦大学，生物信息学研究生课程就由我和朱军教授承担下来。现在该课程作为浙江大学全校性研究生公共课程，已成为一门重点建设课程，每年选课人数都在 150 人左右。

上个世纪末，我国生物信息学还处于起步阶段，学习资料很少。学生时常索要学习材料，于是我整理了备课笔记，取名《生物信息学札记》，于 2001 年 6 月挂到实验室主页上供学生参考。随着生物信息学发展，分别于 2005 年 3 月和 2010 年 1 月更新札记两次。由于网络传播的作用，许多生物信息学初学者都读过该札记，在国内形成一定的影响。本书是在该札记框架基础上，补充大量新材料编写而成。

生物信息学学科内容涵盖广且发展很快。基于国内外生物信息学相关教材，以及自身对生物信息学的粗浅理解，我把生物信息学大致分为四部分（篇）内容：第一部分即基础篇，为生物信息学的基础知识。这部分内容总体变化不大（与 10-15 年前比较），它是生物信息学的核心知识，生物信息学教学最重要部分，为应为必讲内容；第二部分高通量测序数据分析篇，最近十年才出现的生物信息学新内容。2005 年高通量测序技术突破后，针对该技术产生的序列数据，出现大量生物信息学新算法和新工具；第三部分生物信息学外延与交叉，重点介绍与生物信息学密切相关的其他生物学学科。生物信息学引入了这些学科的部分核心技术（或反过来被引入），如数量遗传学、群体遗传学和新兴学科合成生物学；第四部分为生物信息学资源与实践篇。生物信息学数据库和软件工具对生物学学科至关重要，所以这部分也是生物信息学重要组成部分。同时，该篇中以实践为目的的生物信息学教学资源是课堂教学的一个很好补充。

我重点编写了本书第一部分基础篇。我的学生参与撰写了有关章节，同时也邀请了相应领域研究者参与部分章节撰写（徐海明：数量遗传学；阮松林：蛋白质组学），最后由我统稿。我们尽可能完整地列出参考书目，标注材料来源，但一定还会有所遗漏。本书受浙江大学本科专业核心课程教材建设专项经费资助出版。

每次拿起书稿总是能发现一些错误或不准确的地方，但由于出版计划一再拖延，只好交稿付印了。如果你发现书中问题，望赐教指正（fanlj@zju.edu.cn），以便我们再版时更正。

樊龙江

2016 年 9 月

《生物信息学》简要目录及 PDF 下载（二校稿，以出版为准）

详细目录