学习报告

尊敬的王导：

展信佳！感谢您接纳我作为您的学生，并且分享浩如烟海的珍贵资料。现向您汇报本周学习情况及规划。

一、*Bioinformatics Data Skills*阅读

1.通读

在这三天，我首先通览了目录及前言，认识到这是一本面向有一定生物学或计算机基础的生物信息学初学者的入门读物，主要介绍：

* 良好的分析习惯（第一章）
* 操作系统(linux)、编程语言(shell，R)和常用工具(awk, grep, sed)等（第二章到第八章、第十二章到第十三章）
* 常见数据类型（gff, fasta, sam等）的处理（第九章到第十一章）

2.前言

在前言中，我认为最有意思的地方在于作者不断强调数据分析的稳健性（robust）和可重复性（reproducible）。作者称*robust*是指工作能够抵御“无声错误”、混杂因素、软件错误以及杂乱的数据；而*reproducible*是指工作可以被其他研究人员重复进行，且可以得出相同的结果。这可能是生物信息学最重要的精神理念，否则作者不会再三列出这两个单词。

但同时，作者指出了读者需要具备的前置能力，譬如：懂一门脚本语言、会用一种文本编辑器、懂基本的Unix命令行操作、掌握基础生物学知识、对正则表达式有基本的了解等等。遗憾的是，上述能力我不认为自己完全具备，甚至感到陌生，这也对应了文中“The result is that sections of this book are quite advanced and will be difficult for some readers. Don’t feel discouraged! Like most of science, this material is hard, and may take a few reads before it fully sinks in.”但愿我能够在学习本书时有毅力同时学习这些本领，逐步适应。

3.Part I (Chapter 1)

第一部分包含一个章节，“如何学习生物信息学”。

开篇，作者指出当下数据数量随着检测成本降低而指数级增长，从而引入研究生物信息学需要具备的重要品质：robust和reproducible，即前言中的两个重点单词。随后进行细化分析：重复通常不仅依赖数据和代码，还依赖软件版本、数据下载时间和版本等详细信息，这些元数据或数据是确保再现性的关键细节，故而生信学家务必reproducible；生物信息学工作以数据为主，其错误比湿实验更难发现，而生物信息学项目，甚至是经过大型用户群测试的工具，也可能无法在特定有机体上很好地工作，故而生信学家务必robust。

接下来，作者给予了一些具体的建议：代码应具有可读性，数据应便于计算机读取，善于测试代码，使用现成函数库等等。

总览第一章，作者主要介绍研究生物信息学应具备的良好分析习惯，为后续正式的学习内容做铺垫。

二、针对后续学习的规划

1.*Bioinformatics Data Skills*

本书正文共469页（不含前言），现已读至17页，距离下学期开学还有约80天，计划每周阅读40页左右。寒假期间无课业压力，可能进度会比上学期间更快。

2.Python

本想寒假期间再着手学习，但目前看来对python的需求是燃眉之急。故决定阅读主书目同时进行自学，不求精通，但求入门，学习进度每周汇报。

这是大学期间的第一篇学习报告，其中格式必然有不足之处，对书中内容的理解也可能存在错误，学习规划或许也有不合理之处，恳请您不吝赐教，让我的生物信息学学习有长足进步，更快融入您的小组，非常感谢！

丁加

2024年11月24日