

* 院长论坛 *

20 世纪后期, 生物技术 (BT) 迅猛发展, 无论从数量上还是从质量上都极大地丰富了生物科学的数据资源。截至 2002 年, 来自 NCBI 的基因组数据每 14 个月就翻一番。这种数据资源的急剧膨胀迫使人们寻求一种强有力的工具去组织这些数据, 以利于储存、加工和进一步利用。于是, 一门崭新的、拥有巨大发展潜力的新学科——生物信息学悄然兴起, 并得到极为迅速的发展。

生物信息技术专业人才培养方案

哈尔滨工业大学 廖明宏 王宽全 张宏志 / 文



廖明宏教授

1995 年, 在人类基因组计划 (Human genome project) 第一个五年总结报告中, 科学家们才首次对生物信息学给出了一个较为完整的定义: “它是一门研究包括生物信息的获取、加工、存储、分配、分析、解释在内的所有方面, 并综合运用数学、计算机科学和生物学的各种工具, 来阐明和理解海量生物信息学数据所蕴含的生物学意义的新兴的交叉学科”。生物信息学的发展将会对生命科学带来革命性的变革。

因此, 各国政府和工业界对此

极为重视, 投入了大量资金。欧美各国及日本相继成立了生物信息数据中心, 如美国的国家生物技术信息中心、国家基因组资源中心、英国的欧洲生物信息研究所、日本的国家遗传学研究所等。以西欧各国为主的欧洲分子生物学网络组织 (European Molecular Biology Network, EMBNet) 是目前国际最大的分子生物信息研究、开发和服务的机构, 通过计算机网络使英、德、法、瑞士等国生物信息资源实现共享。我国科技界自完成人类基因组计划 1% 项目以来, 对生物信息

表1 2002年国家教育部备案的一级学科内自主设置的生物信息学专业名单

| | | |
|----------|-----------|------------|
| 数学 | 南开大学 | 生物信息学 |
| 生物学 | 北京大学 | 生物信息学 |
| | 中国农业大学 | 生物信息学 |
| | 南开大学 | 生物信息学 |
| | 南京大学 | 生物信息学 |
| | 浙江大学 | 生物信息学 |
| | 中国科学技术大学 | 生物信息学 |
| | 山东大学 | 基因组学与生物信息学 |
| | 中国科学院研究生院 | 生物信息学 |
| 控制科学与工程 | 清华大学 | 生物信息学 |
| 计算机科学与技术 | 哈尔滨工业大学 | 生物信息技术 |
| | 国防科学技术大学 | 生物信息学 |
| 生物医学工程 | 上海交通大学 | 生物信息学 |
| 基础医学 | 中南大学 | 生物信息物理学 |
| 药学 | 中国药科大学 | 药物生物信息学 |

学研究也给予了极大的重视，南方、北方两大生物信息学中心的相继成立标志着我国生物信息技术领域的重大进展。

然而我国目前想在生物信息学领域进入国际先进行列面临的最大问题是人力资源的极大缺乏，因为生物信息学研究是横跨分子生物学和计算机技术两大专业领域的新兴交叉学科，目前还没有任何一个专业有能力独立培养高水平的研究人才。为了在国内抢占制高点，北京大学已在2001年春天从计算机、生物等系遴选40名优秀学生成立了生物信息学本硕连读班；以清华大学为首的国内各大高校也都在紧锣密鼓地筹建生物信息学本科专业和硕士点。哈尔滨工业大学也于2002年申请并获得批准成立生物信息技术专业。本文重点介绍哈尔滨工业大学生物信息技术

专业人才培养方案。

一、国内外发展现状

随着社会对生物信息学人才需求的不断增加，世界各国大学相继建立了生物信息学本科专业。目前，已经开设生物信息学本科专业的国外大学有：George Washington University, Waterloo, Wright State University, University of California San Diego, RIT, Canisius College, UC Santa Cruz, RPI, Ramapo College, College of the Sciences (Philadelphia), Baylor, College of Staten Island, SUNY Buffalo等等。

不同类型的大学在建设生物信息技术本科专业的侧重和方法也有所不同。其中，RIT和Canisius College两所大学的课程设置最为典型。Canisius College的生物信息学

专业隶属于计算机科学系，而RIT的隶属于生命科学系；Canisius College的相关课程主要是基于现有的计算机专业课程，力求使专业所需的新课程减到最小，而RIT的相关课程设置则显得较为庞大。

两所大学本科专业课程设置中各相关交叉学科课程所占的比例也不尽相同。在Canisius College生物信息学专业的本科课程中，计算机科学类占12%，数学类占9%，化学类占11%，生物学类占10%，生物信息学类占4%，文学艺术类占37%；此外，还提供4门专业选修课（12%）以及2门任选课（4%）供学生选修，让学生根据自身需要进一步加强在计算机或生物不同方向上的能力培养。而在RIT的课程设置中，计算机与IT类占11%，数学类占11%，化学类占10%，其文学艺术类占29%，而生物类则占据了40%。

近年来，国内各大学纷纷建立起生物信息学专业，如表1所示。不难看出，由不同的院校所建立的专业侧重均有所不同，他们分属于不同的一级学科。这些差异表明了国内高校对生物信息学含义的不同理解，并且正在形成各具特色的专业，但这也从一个侧面反映出我国生物信息学相关教育工作还处于起步阶段。

二、专业定位

我们必须认识到，国内在生物信息学方面的研究与国外一些知名研究机构还有很大差距，如果想要

在生物信息学领域前沿占有一席之地,就必须发挥我们特有的优势、突出自己的特色,所以如何选择正确的研究方向和方法是发展生物信息学专业发展的关键。

生物信息技术侧重于采用信息技术研究生物信息学问题。如今,随着人类基因组计划的实施,生物学数据迅速膨胀,有关核酸、蛋白质的序列和结构数据呈指数增长,如何有效地对数据进行有效的搜集、管理、处理、分析和释读已经成为生物信息学发展的关键问题。面对巨大而复杂的生物信息数据,运用计算机管理数据、控制误差、加速分析过程势在必行。同时,哈尔滨工业大学在数据库技术、网络、软件开发和数据分析方面又有着得天独厚的优势,这些都为生物信息技术专业在某些领域的快速发展提供了有效保证。

哈尔滨工业大学生物信息技术专业是由计算机科学与技术学院和生命科学系联合创办的,由于我校在计算机领域的教学和科研方面具有较大的优势,因此,本专业的人才培养定位是以计算机技术为基础,侧重信息技术在生物学领域的应用,使本专业培养的学生具备解决生物信息学关键问题的能力,同时又具有在计算机科学和生物学等相关领域继续发展的潜力。

然而,从学科交叉的角度上讲,本专业培养的学生在具备扎实的计算机应用技术的同时,又必须充分掌握生物信息学所必需的生

物学理论基础,这样才能保证本专业学生真正地做到学以致用。但是,由于交叉学科的特殊性,本专业不能够简单地照抄照搬生物学专业的教学课程,在众多生物技术相关的课程中我们必须有所选择。所以,本专业课程设置也必须以我们的学科定位为依据,在达到学科标准的基础上,突出自己的特色,即计算机相关课程的选择应当以解决生物学关键问题为目的,而对于生物学相关课程的设置要做到求精而不求全。

三、培养方案

这里从计算机技术、生物技术和数学等相关课程的选择和内容的侧重点等方面介绍培养方案的设置。

1. 生物信息学相关计算机技术

(1) 生物信息学数据库

数据库是生物信息学的主要内容,各种数据库几乎覆盖了生命科学的各个领域。各类生物信息学数据库在近年不断出现,其规模呈爆炸趋势增长,同时数据结构日趋复杂。目前生物信息学数据库服务已实现了高度的计算机和网络化。算法和软件的进步、数据库的一体化、服务器——客户模式的建立使之成为生物、医药、农业等学科的强有力工具。由于历史原因,目前的各种不同生物信息学数据源(数据库)的整合性极低,使得各数据源协同应用极为困难。这个问题在处理生物学研究工作中所必须的复杂检索时尤为突出(生物学研究人员为得到所

查询的结果,往往需要在不同数据源之间进行选择、合并和互换,这样就需要大量的时空开销)。

针对生物信息学数据库的特殊性,本专业设置的数据库课程内容与计算机科学与技术专业有所不同。它应更强调大型数据库和网络数据库,并应涉及海量数据存储管理、查询分析,以及分布式异质数据库的查询、访问等技术。

(2) 生物信息学软件工程

在生物信息学领域,面向生物信息学的软件开发还要更多地依赖于开发人员的个人技巧,远未形成类似于计算机软件工程这样的理论体系。一方面,各种各样团体或组织都各自独立地开发自己的工具软件,而这些软件绝大部分出于短期需求考虑,故应用面较窄。另一方面,一些生物信息学家(科学范畴)为了进行自己的研究,只好作为业余程序员来编写一些针对某种特殊用途的专用软件,但同时他们又必须要避免被这些原本属于软件开发人员(工程范畴)的工作占用太多的精力而影响自己的本职工作(科学范畴)。另外,就科学、技术以及商业领域而言,各种技术的快速更新换代又为软件开发标准的制定带来了负面影响,使得组建一个稳定的、跨平台的软件标准变得极为困难,而这又在某种程度上导致软件厂商开发生物信息学软件的成本急剧攀升。

由此可见,本专业开设的软件工程课程必须面向生物信息学,具

体现在：针对目前国内外各种生物学新技术的数量迅猛增加，开发生物信息学软件必须要考虑新技术（模块）在现有软件平台上的“即插即用”，因此我们必须在课程中加入基于构件的软件开发的新技术；针对生物信息学软件开发标准不统一、兼容性不好等问题，本专业的软件工程课程将更多地强调软件体系结构的建立等内容。

（3）程序设计语言

为了保证学生具有进行生物信息学软件开发的基本能力，生物信息学专业的学生必须能够熟练使用些传统的编程语言（C、C++、Java等）。同时，为了解决海量生物信息数据结构化的表述方法，我们还要加强对一些新型程序设计语言的培养，例如HTML、XML等，新近发展起来的GEML、BioML及OpenBSA生物信息数据表述方式都采用了XML标准，将各种生物信息数据表述成结构化形式，同时又不失灵活性。诸如这样的教学内容可以达到激发学生在将所学得的计算机技术应用于生物学研究领域的过程中的创新意识。

（4）其他计算机相关课程

对生物信息专业而言，并非所有计算机课程均有必要开设。根据本专业的定位，我们着重保留了软件方面的课程，有意识地删减了一些硬件课程（例如汇编语言等），而对于计算机组成原理等基础类课程，我们仍然加以保留，目的是为使

将来向多领域横向发展留有余地（例如生物芯片、生物计算等生物信息相关领域都需要有较深厚的计算机理论基础）。

2. 生物学相关课程

生物学研究的主要内容重点体现在基因组学（Genomics）和蛋白质组学（Proteomics）两方面，具体说就是从核酸和蛋白质序列出发，分析序列中表达的结构功能的生物信息。可见，对于生物信息技术专业，一些作为基础的生物技术知识必不可少，为此我们开设了生物学、分子生物学以及遗传学等生命科学专业课程。通过这些课程的学习，学生能够掌握专业所需的生物技术知识，例如序列比对和序列分析、多序列比对、系统发育和简图分析、蛋白质结构分析、蛋白质结构预测、基因组学和蛋白质组学、遗传密码的起源以及基于结构的药物设计等其他领域的相关知识。

根据本专业生物技术相关课程设置“求精不求全”的原则，我们考虑将现代基础化学、有机化学、生物化学以及物理化学等课程合并，针对本专业的特殊背景取其精华，合并为一到两门课程，在不影响学生对一些基础理论的把握的情况下，减轻学生的负担。

3. 生物特征识别技术

哈尔滨工业大学计算机学院在人体生物特征识别领域已经有了较长时间的技术积累，并在其与人体生物特征识别的身份鉴别和医学诊

断领域取得了一些令人鼓舞的成果。另一方面，就生物信息学而言，来自生物（人体）的宏观生物信息也不容忽视，为传统生物信息学引入宏观信息，从学科发展的角度上讲也不失为一个新的思路。例如，将二者结合起来，研究宏观病理特征在基因和蛋白质层面上对应的表现，有可能得到新的理论突破，同时也可以拓展本专业学生的研究领域。因此，我们拟开设人体生物信息识别技术和诊断信息学两门课程。

4. 相关数学课程

随着生物信息学的快速发展，愈来愈多的数学工具被广泛应用于生物信息学研究之中，主要有统计学，包括多元统计学，它是生物信息学的数学基础之一；概率论与随机过程理论，如近年来兴起的隐马尔科夫链模型（HMM），在生物信息学中有重要应用；运筹学，如动态规划法是序列比对的基本工具，最优化理论与算法，在蛋白质空间结构预测和分子对接研究中有重要应用；拓扑学，这里指几何拓扑，在DNA超螺旋研究中是重要工具，在多肽链折叠研究中有应用；函数论，如傅里叶变换和小波变换等都是生物信息学中的常规工具；信息论，在分子进化、蛋白质结构预测、序列比对中有重要应用；而人工神经网络方法则用途极为广泛；计算数学，如常微分方程数值解法是分子动力学的基本工具；群论，在研究遗传密码和DNA序列的对称性方面有重要应用；组合数学，在分子进化和基因组

序列研究中十分有用。

由于学时的限制,相关的数学课程只能选择核心课程加以介绍,其他课程可在研究生阶段进一步深化。

5. 其他课程

本专业除开设上述课程外,还将按教育部规定开设思想政治理论及人文类课程,以丰富学生的学习和生活、培养面向社会的新型人才。

6. 实践教学

本专业学生需要在计算机科学与技术专业和生物技术专业进行相关课程的实践学习,掌握两个专业的基本技能。在毕业设计阶段,学生可选择与生物信息学相关的理论

或应用课题进行毕业设计,从而掌握利用所学知识开展生物信息技术研究和应用的基本技能。

四、课程拓扑结构

课程拓扑结构由基础理论系列、计算机技术系列、生物技术系列、生物信息技术系列和全校公共课系列组成(限于篇幅,此处省去课程拓扑图,读者如需阅读,请参看本刊网站上的转载)。专业主干课程包括:高等多元分析、数据结构与算法、操作系统、数据库系统、生物信息学软件工程、计算机组成原理、计算机网络、模式识别、生物学、生物化学、分子生物

学、遗传学、生物信息学、统计遗传学、分子进化分析、基因组信息学等。总学时2700学时,总学分为184学分。

结束语

生物信息技术专业的人才培养在我国刚刚起步,在国际上也没有现成的培养方案。我们在制定生物信息技术专业人才培养方案时充分考虑专业自身特点,并结合我校的学科优势,能够较好地体现我校的特色。由于还没有毕业生,该培养方案的实际效果如何还有待进一步考验。这里,仅作为一项有益的探索提供给大家借鉴。Edu

瑞星反病毒 2005 年 4 月播报

| | | | | | | |
|---------------------------|---|------|--|--|------|------|
| 本月热门病毒 | MSN 木马变种 P/Q/R/S/T/U (Worm.MSN.Bropia.p/q/r/s/t/u): 警惕程度★★★★, 蠕虫病毒, 通过 MSN 传播, 依赖系统: WIN9X/NT/2000/XP。中毒电脑向 MSN 好友自动发送病毒文件, 其他用户运行文件后就会中毒, 其中部分中毒电脑可能被黑客控制。病毒还会造成杀毒软件以及一些常用工具软件出现异常无法使用。该病毒目前属于活跃期, 用户一定要小心防范。 | | | | | |
| | 梅隆变种 D (I-Worm.Win32.Mellon.d): 警惕程度★★★★, 蠕虫病毒, 通过电子邮件及 IRC 软件进行传播, 依赖系统: WIN9X/NT/2000/XP。病毒会搜索本机 Outlook 地址簿, 向找到的电子邮件地址发送病毒邮件。该病毒还通过 IRC 软件进行传播, 消耗系统资源, 并可能造成网络堵塞。 | | | | | |
| 本月发作病毒 | 病毒名称 | 发作时间 | 病毒类型 | 发作现象 | 传播途径 | 危害级别 |
| | SCO 炸弹变种 AR (Worm.Novarg.ar) | 随机 | 蠕虫病毒 | 病毒会盗取“淘宝网”、“工商银行”、“招商银行”、“农业银行”、“建设银行”等网上拍卖网站、网上银行及其他一些网上支付系统的账号和密码, 给用户的财产带来损失。 | 电子邮件 | ★★★★ |
| | 好友盗号器变种 R (Trojan.QQSW.EasyGet.r) | 随机 | 木马病毒 | 病毒运行后会自动用IE浏览器显示一张美女图片。病毒会不断记录用户的键盘操作, 从而窃取用户的QQ号、密码等信息, 并发送到黑客指定的信箱中。 | 网络 | ★★★☆ |
| 外星木马 (Backdoor.Ethorse.a) | 随机 | 后门程序 | 此病毒是一个国产后门程序, 表面上看是个图片, 用户点击后会提示“无法打开”, 其实这时候就已经被感染, 黑客可以远程窃取用户的QQ密码等信息。 | 网络 | ★★★ | |
| 本月升级版本 | 本月的MSN木马变种P/Q/R/S/T/U (Worm.MSN.Bropia.p/q/r/s/t/u)和梅隆变种D (I-Worm.Win32.Mellon.d)值得注意。“MSN木马变种P/Q/R/S/T/U”会向MSN好友自动发送病毒文件, 病毒还会造成杀毒软件以及一些常用工具软件出现异常无法使用。而“梅隆变种D”则会通过电子邮件及IRC软件进行传播, 危害较大。可以上网的用户推荐采用在线杀毒方式快速查杀这些病毒, 也可以使用单机杀毒软件2005版, 企事业单位用户最好能使用网络版来抵御这些病毒。《瑞星在线杀毒》无需升级、《瑞星杀毒软件2005版》和《瑞星杀毒软件2005下载版》每工作日正常升级, 遇紧急病毒第一时间提供解决方案, 每周升级的新病毒总数不少于400个。截止到4月底瑞星杀毒软件将升级至17.24版本, 望广大用户及时升级。 如遇病毒, 请拨打反病毒急救电话: 010-82678800 或 使用瑞星在线杀毒: http://online.rising.com.cn | | | | | |