BIG DATA

大数据时代的生物技术和农业

李　宏　王拥军◎编著

非外借

科学出版社

大数据时代的生物技术和农业

李　宏　王拥军　编著

科学出版社
北　京

内 容 简 介

本书主要阐述了大数据时代生物技术和农业即将发生的变革。前3章介绍了大数据的基本概念和重要性，人类基因组计划产生的大量数据对医学和药物设计带来的推动力量，生物信息学的发展对于医学研究、临床医疗和药物开发的影响；第4章介绍了大数据引发的农业技术变革，包括作物基因组计划对遗传育种的积极影响，大数据与精准农业等内容；最后一章对大数据的安全性和相关伦理学问题进行了讨论。

本书适合于生物医学、农业生物技术和管理等学科的本科生和研究生阅读，也可作为从事相关学科研究的参考书。

图书在版编目(CIP)数据

大数据时代的生物技术和农业 / 李宏，王拥军编著. — 北京：科学出版社，2019.9

ISBN 978-7-03-062207-5

Ⅰ.①大… Ⅱ.①李… ②王… Ⅲ.①数据处理-应用-生物工程-研究 ②数据处理-应用-农业-研究 Ⅳ.①Q81②S-39

中国版本图书馆CIP数据核字（2019）第188444号

责任编辑：冯 铂 贾 桥 / 责任校对：彭 映

责任印制：罗 科 / 封面设计：墨创文化

科学出版社出版

北京东黄城根北街16号

邮政编码：100717

http://www.sciencep.com

成都锦瑞印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2019年9月第 一 版 开本：B5（720×1000）

2019年9月第一次印刷 印张：11

字数：220 000

定价：99.00元

（如有印装质量问题，我社负责调换）

此为试读，需要完整PDF请访问：www.ertongbook.com

自　序

大数据正在给社会经济发展带来前所未有的机会，世界各国都意识到了这一点。有关大数据方面的书籍，也引起了读者的兴趣。如美国的埃里克·托普(Eric Topol)所著的《颠覆医疗：大数据时代的个人健康革命》《未来医疗：智能时代的个体医疗革命》，德国的埃拉德·约姆-托夫(Elad Yom-Tov)所著的《医疗大数据：大数据如何改变医疗》，日本21世纪医疗论坛所编著的《大数据时代的医疗革命》等书，都是在关注大数据对未来医疗的影响。中国在大数据领域也有所探索。

中国人非常关心医疗问题，原因是中国人口众多，而医疗资源很匮乏，与发达国家相比，我国当前的医疗状况不容乐观。在看病难这种现状没有改变之前，人们总是希望医疗条件有所改善，大数据给人们带来了希望。大数据将颠覆传统医疗，引发医学的一场大变革。人们希望通过这场变革看到医疗现状的改观。农业问题也是中国人非常关注的，中国的人口众多，土地资源有限，如何提高农业生产效率，任务十分艰巨。用有限的资源，养活近14亿人口，中国的农业必须发生转变，传统的种植模式已经无法满足不断增长的人口需要，在这种情况下，精准农业有了市场，而大数据是实现精准农业不可缺少的技术支持。

本书的编写有两个目的，一是让人们了解大数据即将给医疗和农业带来的巨大变革；二是让人们认识大数据的两面性，大数据的个人隐私安全和个人权益，是当前人们担忧的问题。但大数据给社会经济发展带来的好处是肯定的，大数据已经被美国上升为国家战略，被认为是21世纪的“新资源、新石油”，社会经济发展越来越离不开大数据。谁抢占了大数据，谁就抢占了先机。因此，大数据引起了许多企业家的关注，被称为新的财富。

李　宏

2019年2月28日

前　言

大数据时代已经悄然而至，渗透到我们生活的方方面面。大数据正在以破坏性创造的方式，改变传统医学模式和农业的生产模式。与脱氧核糖核酸(deoxyribonucleic acid，DNA)有关联的领域，都因为基因组数据的大量增长而悄然发生改变，但人们还没有完全了解可能发生的改变，以及在今后的几十年会对我们的生活产生何种的影响。

在人们正在为某些问题疑惑不解的时候，总有一些科学先贤走在了研究的前列。历史事件总有那么多的巧合，并且反复重现。在人们对地球物种多样性的来源疑惑不解的时候，在剑桥大学神学院学习的达尔文因偶然的机会踏上了贝格尔号科考船去了远方，这次远途航行载回了满满的收获——达尔文关于物种进化的思想，他也从一个剑桥大学神学院的学生摇身转变成为影响人类观念的自然科学家，动摇了宗教的神创论。而当人们对遗传物质争议不休的时候，同样是来自剑桥大学的两位年轻学者——詹姆斯·沃森(James Watson)和弗朗西斯·克里克(Francis Crick)提出了 DNA 双螺旋结构模型。与达尔文进化论一样，这一发现具有划时代的意义。生命科学的快车就在双螺旋搭建的轨道上高速前进，跨越了一个半世纪。几个具有重大意义的技术革新，如基因的一代测序、二代测序和各种组学的技术等大大加速了生命科学领域数据的产生速度。基因组测序产生的海量数据正在改变生物医学研究者的研究思路，习惯于从实验入手进行研究的科学家现在面临着一种更优的选择——从数据入手进行研究，这种方法逐渐被人们接受，因为有许多事例证明了从数据入手进行研究是一种成本很低、更省时间和人力的方法，常会出现“弯道超车”的奇迹。

大数据必然会对生物医学和农业带来惊喜，大数据将驱动技术创新，并且助力于生物医学和农业生物技术领域的创业，为创业者提供更好的发展空间和环境。置身于大数据时代的我们将会迎来更多的机遇，借助于大数据，可以即时获得有价值的信息，预测市场变化，采取应对措施。

全书共分为五章，主要阐述了大数据时代生物医学和农业即将发生的变革。前 3 章介绍了大数据的基本概念和重要性，人类基因组计划产生的大量数据对医学和药物设计带来的推动力量，生物信息学的发展对于医学研究、临床医疗和药物开发的影响；第 4 章介绍了大数据引发的农业技术变革，包括作物基因组计划对遗传育种的积极影响，大数据与精准农业等内容；最后一章对大数据的安全性和相关伦理学问题进行了讨论。第 1 章至第 3 章由李宏执笔，第 4 章和第 5 章由

王拥军执笔。

本书的写作过程中得到了单位和学院领导以及同事的大力支持和帮助，科学出版社成都分社的黄桥编辑对本书的出版付出不少心血。在此感谢各位领导、同事和编辑的关心和支持，是他们给了我很大的勇气和动力来完成本书的写作。

大数据正引领着这个时代向新的方向发展，作为一个新兴的事物可能不能很快被大众接受，但大数据给整个社会带来的好处远远超过了它的负面影响，这一点是不可否认的，正如人类历史上的其他新兴事物一样。本书只是让读者对大数据对生物医学和农业带来的变革有所了解，并没有很专业、很深入地切入，希望起到“抛砖引玉”的作用，相信今后这方面的书籍会逐渐增多，大数据也会随着时代的发展出现新的应用。由于编者水平所限，不可能面面俱到，书中的不足之处在所难免，恳请同行和广大读者批评指正！

目　录

第 1 章　大数据的来源

本书所讲的大数据是与生命科学、医学和农业密切相关的数据和信息，当然一般意义上的大数据含义更广，涉及的学科领域和研究范围更大。

1.1　大数据的概念及类型

2012 年是不平凡的一年，“大数据”(big data)一词被各大宣传媒体报道，成了热点词汇。大数据因我们这个时代的信息爆炸而生，用于描述和定义由此产生的海量数据，以及与之相关的技术发展与创新。人类是一个想象力丰富、敢于面对挑战的物种，大数据对人类提出了新的挑战，这也激发了人类提高数据驾驭能力的欲望。正如《纽约时报》2012 年 2 月的一篇专栏中所称，“大数据”时代已经降临，在商业、经济及其他领域中，决策将日益基于数据和分析而做出，而并非基于经验和直觉。

引入大数据这种新概念的功劳应归功于国际著名期刊《自然》(*Nature*)。早在 2008 年 9 月 4 日，也就是谷歌成立 10 年前际，《自然》期刊推出了一期大数据专辑，包括 8 篇大数据专题文章加上 1 篇编者按。从此，大数据就以一种新概念的形式进入了公众视野。目前对大数据的定义有三种，分别从数据体量、复杂性程度、价值三个角度来界定什么是大数据。国际著名管理咨询公司麦肯锡十分看重大数据的价值，并认为大数据是指那些规模大到传统的数据库软件工具已经无法采集、存储、管理和分析的数据集。麦肯锡对大数据的定义表现出，有效利用大数据需要有超越传统数据分析的技能，否则，大数据只是一堆数字、图像而已。

科技发达的美国在大数据的利用方面也抢得头筹，2012 年 3 月 22 日，奥巴马政府宣布投资 2 亿美元资金用于拉动大数据相关产业发展，并将“大数据”上升为国家战略。奥巴马政府认为大数据是“未来的新石油”，对于国家的经济发展是不可缺少的新资源，一个国家拥有数据的规模、活性及解释运用的能力将成为综合国力的重要组成部分，未来对数据的占有和控制甚至将成为陆权、海权、空权之外的另一种国家核心资产。奥巴马政府对大数据的重视程度已经到了与石油能源同等的水平，美国曾经因争夺石油控制权而对石油富产国发动战争，是因为石油是国家经济的命脉，由此可见大数据的价值是不容低估的。

联合国也在 2012 年发布了大数据政务白皮书，指出大数据对于联合国和各国

此为试读，需要完整PDF请访问：www.ertongbook.com

政府来说是一个历史性机遇，人们如今可以使用极为丰富的数据资源，来对社会经济进行前所未有的实时分析，帮助政府更好地响应社会和经济运行。从政府层面上来讲，推动大数据的应用可以刺激经济的发展，也显示出大数据的重要性。

下面就生物学和医学以及农业大数据作基本的介绍。

1.1.1 生物学和医学大数据

随着现代科学技术的发展，特别是近年来生命科学领域中的基因组测序技术的进步，产生了大量的生物医学数据，其特点是数据量特别庞大。虽然不同来源的数据其格式不同，但在互联网广泛应用的今天，这些已经可以得到解决。不同来源的数据融合、互换、对比以及更新已经成为常态。许多生物医学数据库也建立起来，并通过互联网技术实现数据共享。特别是人类基因组计划产生的庞大数据资源的共享，促进了世界各国的科技合作和研究模式的变革。

大数据提供的信息是多方面的，这对使用者有利。但数据的形式和格式不同，也带来了数据交换的困难和障碍。因此，数据的标准化是十分必要的，如医疗影像行业内通过制定医学数字影像和通信(digital imaging and communication in medicine，DICOM)标准，可以将不同格式的数据转换成标准数据模式。

针对不同格式的生物学数据，需要通过计算机软件对其进行整合集成，并且使用功能强大的查询系统进行数据查询。SRS 和 Entrez 等查询系统就是其中比较典型的例子，用于解决不同格式的生物学数据的利用问题，为生物信息资源的利用提供了极为有效的工具。它们将各种类型的数据库集成在一起，通过统一的界面和查询方法，利用计算机网络，实现了信息共享。SRS 即“Sequence Retrieval System”的英文缩写，由欧洲的 EMBnet 开发，用于查询 EMBnet 收集和整理的许多格式不同的生物数据库。Entrez 查询系统是由美国国家生物技术信息中心(National Center for Biotechnology Information，NCBI)开发的数据库查询系统，使用过 Entrez 查询系统的人会感觉这是一款很不错的查询软件，并且很快就会喜欢上它。Entrez 采用自动列出相关记录的方法，实现同一数据库中不同条目或不同数据库之间的链接。但与 SRS 不同，Entrez 是一个封闭的数据库系统，而 SRS 是一个开放的系统。

大数据时代的来临对实验科学产生了重大影响，这种影响不仅是在研究方法上，而且也改变了研究者的视角。以往，研究者通过实验研究来观察现象，实验目的是获得结论或者是提出一种新假设。而现在，生物医药领域的科学研究已经转变成了数据驱动过程，通过对海量数据的研究来探索其中的规律，可以直接提出假设或得出可靠的结论。数据挖掘已经成为生物医学研究的必备手段，但数据挖掘需要具备特定的分析技能，因此必须有相应学科的人员参与，这就需要进行团队合作。利用同行研究数据进行 Meta 分析，在生物医学领域已经流行起来。

推动生物医学研究的测序技术发展很快，从早期的第一代 DNA 测序技术，发展到高通量测序技术，到现在的第三代 DNA 测序技术，其技术本身发生了质的飞跃。此外，医学影像数据也不断增多，健康档案和文献数据也丰富了大数据资源。广泛存在的数据共享正带来新的挑战，如分析和移动大量数据集存在的计算和输送上的困难，以及如何保护研究参与者的隐私问题。对强大可靠的计算平台的需求，正带来生物医学研究中云计算使用的快速增长。

1. 新的 DNA 测序技术与序列数据

第二代测序技术(next generation sequencing)也叫高通量测序技术，可以一次对几十万到几百万条 DNA 分子进行序列测定，使得对一个物种的转录组和基因组进行细致全貌的分析，以及在极短时间内对人类转录组和基因组进行细致研究成为可能。第二代测序的核心思想是边合成边测序(sequencing by synthesis，SBS)，即通过捕捉新合成的末端的标记来确定 DNA 的序列。与传统的桑格(Sanger)测序技术相比，新一代测序平台最大的变化是无须克隆这一烦琐的过程，而是使用接头进行高通量的并行聚合酶链反应(polymerase chain reaction，PCR)直接测序，并结合微流体技术，利用高性能的计算机对大规模的测序数据进行拼接和分析。新一代测序平台所产生的数据量巨大。使用第一代 ABI 3730XL 毛细管电泳测序仪进行基因分析，每年至多能完成 6000 万碱基的测序量，这样的速度要对人类基因组 300 亿碱基序列测序，唯一可行的就是多国合作。DNA 测序技术跟随着人类基因组计划的步伐，实现了令人惊喜的跨越。2005 年罗氏公司和 454 生命科学公司联合开发出了新一代测序技术，使用焦磷酸测序仪进行基因分析的速度已经远远超越了第一代的 ABI 仪器的速度。如今，新一代测序平台 SOLiD 单次运行，便可以分析 6Gbp(10 亿碱基对)的碱基序列，每周能够产生大约 100×10^9 个 DNA 碱基序列；Solexa 能够对最长 150 个碱基的 DNA 片段进行测序，每周能够产生大约 200×10^9 个 DNA 碱基序列；而 Illumina Genome Analyzer(GAII)测序系统仅在两个小时的运行时间里，就得到 10TB 的信息。这些由数据机器组成的“数据工厂”每天通过流水线可产生数量庞大的数据。在飞速增长的数据量面前，科研人员感受到了巨大的压力，在数据存储、数据分类、数据处理等方面面临种种困难和考验。

另外还有一个非常现实的问题，目前的 DNA 测序仪生产商仅仅提供用于某些特定基因信息分析的软件，如靶标重测序、基因表达分析、染色质免疫沉淀反应或基因组从头测序等，而并未提供任何其他类型的下游生物学信息分析软件用于第二代测序数据分析，从而使相对便宜的第一代测序受到很多科研机构的青睐。要发挥第二代测序的优势，必须解决第二代测序数据分析问题。只有开发出经济实惠的分析软件以及数据管理系统，第二代测序才能真正实现大范围普及。

第三代测序是一种单分子测序方法，如 Helicos Biosciences 公司的 tSMSTM

技术平台、SMRT 技术以及 Oxford Nanopore Technologies 公司研发的纳米孔测序技术。单分子测序的分辨率具有第二代测序方法不可比拟的优势，由于没有 PCR 扩增等步骤，不存在 PCR 扩增引起的碱基错配，因此单分子测序的精准度很高，特别适用于特定序列的单核苷酸多态性(single nucleotide polymorphism，SNP)检测，稀有突变及其频率测定。例如在医学研究中，对于 *FLT3* 基因是否是急性髓细胞白血病(acute myeloid leukemia，AML)的有效治疗靶标一直存在质疑。研究人员用单分子测序分析耐药性患者 *FLT3* 基因，发现该基因下游的稀有突变与耐药性有关，证明了 *FLT3* 基因是急性髓细胞白血病的有效治疗靶标，消除了一直以来对于这一基因靶标的疑惑。删除了 PCR 步骤的单分子测序能够在更短时间内得到序列数据，减少了客户等待时间，为该技术的临床应用奠定了基础。

2. 医学影像

很多大型医院的医生在给患者开处方之前习惯让患者去做 CT 成像、磁共振成像、超声成像、核医学成像等检查，以便对疾病进行诊断。医生根据医学影像来诊断疾病，这些看似符合逻辑的行为，也助涨了医院在医学检查方面的行为取向，不但增加了患者就医成本，还使医学影像数据海量增长。目前，各大医院的医学影像数据已激增至数十乃至数百万亿字节。伴随医学影像数据激增的是过度的医学检查，这不但不利于改善健康状况，反而会增加发生疾病的风险。如 X 线胸部透视，很多人在一般的体检项目中都做过，但真正查出疾病的只占少数，对于健康的人而言，辐射显然是有害的。还有用于心血管系统成像分析的造影剂，对人体的影响比 X 线透视更严重。特别是有肾衰的人，最好少做造影成像分析。

在临床诊断和医学研究上，医学图像确实是不可缺少的重要参考资料，能够帮助医生判断病情，提出有针对性的治疗方案。这就要求医学图像的分辨率和准确性都很高，但是对于不同的疾病需要采用的成像技术有差异，而且同一种疾病也会采用几种成像技术进行分析，由此产生的图像数据差异极大，异构明显，增加了医学图像数据分析的复杂性。

3. 健康档案

用于记录居民健康状况的健康档案，可从社区、家庭和个人等不同层面反映对疾病的易感性，用于对流行性疾病的防治和家族性遗传疾病以及个人健康的医疗诊断。健康档案可以分为个人健康档案、社区电子健康档案和家庭健康档案三类。个人健康档案包含了个人一生中所有的健康信息；社区电子健康档案是以个人健康档案为基础进行汇总，对于区域疾病防治、建立区域医疗体系非常重要。家庭健康档案一般用于呈家族性遗传特征的人类疾病的分析和治疗。随着医学科学的发展和社会的进步，除了传统的医学图像数据、药物敏感性数据和各种检测数据外，今后基因检测数据或个人基因组数据也会逐渐纳入健康档案中来。

其实健康档案很早就有，在历史上的一些皇家疾病就是通过家族对遗传性疾病的详细记载被发现的，如 19 世纪英国王室家庭的血友病、德意志封建统治家族哈布斯堡家族的下颌突出等。只不过那时候的健康档案局限于王室和贵族，属于家庭和个人健康档案。对于一些流行性疾病，如天花、黑死病、麻疹、梅毒等，在历史上也有记载，这类似于社区健康档案，但它的范围更广，面向的是整个疾病流行的地区。如对雅典瘟疫的记载，反映了在当时这个城市的疾病流行情况。但由于科学技术水平的限制，历史上对疾病的记载都是以文字档案的形式存在，不仅成本较高，长期保存有困难，还容易因为火灾或腐蚀而失传，而电子档案可以克服这些缺陷。

健康档案具有以下三个特点：一是具有持续、大量增长的特点；二是数据格式复杂，不容易整合；三是数据模式会根据时间的推移不断变化、演进。此外，在收集日常健康数据进入健康档案时，如何保证数据的准确性、有效性也是建立健康档案时必须面临的问题和挑战。在计算机广泛普及的时代，临床医生在与患者接触的十分钟里，有七八分钟花在输入数据等事项中，真正与患者交谈的时间仅有二三分钟。因此，难免会出现对病情了解不够准确的问题。

个人健康档案系统在科研、医疗、公共卫生等领域的应用十分广泛，重点包括个人健康状况相关因素分析，疾病地域分布、年龄分布、个人生活史、遗传史等流行病学分析，疾病转归相关因素分析，各种疾病多种治疗手段疗效及费用对比分析，各医疗机构三日确诊率、各种诊断符合率、切口感染率、床位周转率、医疗事故与差错等各种医疗质量和医疗效率指标统计。因此，健全个人健康档案系统就需要对这些因素进行全面的规范和管理。

建立健全医疗健康档案，还需要解决如下几个问题：一是应完善健康档案的存储体系及备份方案；二是建立数据交换标准与方法，通过医疗机构间信息交换提高基层医疗水平；三是建立健康档案的安全机制，对于涉及居民隐私信息的，要保障其信息不得对外泄漏，确实具有医学研究价值的信息，在取得居民同意后才可公开其信息用于科学研究。

4. 医学文献

在现代生物与医学科技快速发展的时代，传统的“生物医学模式”正在向“生物－心理－社会”模式转化。医学涉及学科众多，学科的交叉渗透促进了医学研究的发展，在各类刊物上发表的医学文献数量剧增。医学文献不仅成为重要的资源，而且成为医学界知识更新的主要来源和重要工具。很多国家建立了专门的医学文献数据库，如国外的 PubMed、BMC、BMJ 以及国内的 SinoMed、CBMdisc。在互联网信息资源中，有 30%以上都是医学信息资源。医学文献正在以每年 7%的速度快速增长。例如：国际著名生物医学数据库 PubMed 的数据量达到近 2000 万条，每年还递增 60 万～70 万条；生物医学与药理学文献数据库 Embase 的数据量

达 1100 余万条，每年还新增 50 万条。临床医生平均每天必须阅读 19 篇专业文献，才能跟上当今医学的发展，才能算是一个合格的临床医生。现代医学正进入“信息爆炸”的时代，增强临床医生的综合素质，提高医学信息检索与知识更新能力，才能应对医学信息复杂化的挑战。

在生物医学大数据时代来临之际，我们应尽快构建一个实时、便捷、全方位的医药领域研究与应用系统。在生物医学信息和文献的收集与管理方面，美国国家生物技术信息中心、欧洲生物信息学研究所(European Bioinformatics Institute，EBI)以及日本 DNA 数据库(DNA Data Bank of Japan，DDBI)做出了典范，值得我国借鉴。虽然面临的困难很大，但生物信息资源的建设意义重大。目前我国还主要处在对医疗流程的信息化管理、质量控制等初级阶段，尚未开展面对“大数据”的系统研究与挖掘。但这种研究与挖掘必将成为生物医学发展的趋势，未来的赢家必然是掌握了以大数据为核心的技术。大数据的到来，既对临床医生、医院、研究人员、医疗监管机构等都提出了巨大挑战，也为生物医学研究带来了前所未有的机遇。如何有效地利用这些信息并最大限度地减少伦理相关问题对个人和公众的困扰，是亟待解决的重要课题。

1.1.2 农业大数据

农业大数据是大数据理念、技术和方法在农业上的实践。农业大数据与耕地、播种、施肥、杀虫、收割、存储、育种等各环节有密切关系，涉及范围十分广泛，具有跨行业、跨专业、跨业务等特征，对数据进行分析、挖掘以及可视化，最终服务于农业生产。农业一直是技术相对落后的产业，农村的基础设施相对落后，这就给农业大数据的应用设置了障碍。互联网和计算机的普及使农业大数据搭上了信息高速公路的快车，通过进一步完善农业基础设施，农业大数据的应用前景看好。

目前，大力发展农业大数据，加快推进信息化发展，促进信息化和现代化融合，已经成为各国发展农业的重要趋势。2013 年，英国正式启动“农业技术战略”，提出高度重视利用“大数据”和信息技术提升农业生产效率。美国、法国等通过政府推进农业大数据技术也取得了较好的成效。

中国是世界上人口最多的国家，其粮食需求量巨大，以世界 7%的耕地、6%的淡水资源养活全世界约 1/5 的人口，我国农业发展任重而道远。近 50 年来，作物科学的发展有力推动了“中国绿色革命”，中国的粮食总产和单产均提高了 5 倍多。“中国绿色革命”对于世界粮食增产所做的贡献非常巨大，可以说中国应该是世界绿色革命的起源地与代表国。中国是世界上最早培育与推广杂交水稻的国家，这对于绿色革命之后中国粮食产量的大幅度提高产生了重要的作用。长期以来，中国是一个以农业为基础的国家，农业为今天国民经济的高速发展做出了巨大的贡献。

1. 作物基因组

迅速发展的基因组测序技术在越来越多的农作物基因组研究上取得了突出成果。继水稻基因组测序之后，世界各国又完成了 64 种作物基因组测序，其中包括主要粮食作物(小麦、玉米、高粱、谷子等)、经济作物(棉花、大豆等)、园艺作物(主要蔬菜、果树等)等。值得提出的是，这些项目中有 25 种是由中国独立或参与完成的，这标志着我国作物科学已经进入基因组学时代。基因组测序不仅在全基因组水平揭示了物种的组成，而且为种质资源变异组学研究、育种基因组与栽培基因组研究奠定了基础。

基因组测序与分析发现，不同作物虽然因重复序列比例不同而呈现出基因组大小的巨大差异，但其二倍体基因组中的基因数量相似，均为 3 万～4 万个；在这些功能基因之中，有许多决定作物数量性状位点(quantitative trait loci，QTL)，在作物抗旱性、抗寒性等特异性状方面，这些数量性状位点所起的作用较大。例如，高粱的抗旱性与其携带较多的抗旱基因有关，小麦的抗寒性与其携带较多的抗寒基因有关。如果要培育抗旱型作物品种，就需要多聚集类似高粱基因组中的抗旱基因，同样，培育抗寒型的作物品种，就要把小麦基因组中的抗寒基因借用过来。如果育种工作者发挥到极致，充分利用这些优良性状的遗传基因，那么，全世界的农业生产将得到很大的提高。我国北方地区冬季天气寒冷，而且年平均降雨量较小，属于缺水地区，如果培育出既抗寒冷又抗干旱的作物品种，就可以利用北方广阔的土地种植这种作物，将荒野变成粮仓。

在农业上长期得以运用的常规育种技术已经为农业生产做出了巨大贡献，但传统育种方法的缺点是周期较长、效率较低。虽然近年来发展起来的基因工程技术加快了农业育种的步伐，但由于人们对转基因安全性的担忧，使得这种新技术受到了很多的限制。而近年来快速发展的基因组编辑技术，特别是被喻为“遗传手术刀”的 CRISPR/Cas9 技术，能对目标基因进行精准的编辑，通常可以实现只有少数碱基的取代或者删除，这与基因组上时常发生的自然或诱导突变本质是完全一样的。通过基因组编辑，针对作物某个性状的目标基因，育种工作者可以参照其他优良种质或者近缘物种的同源基因来进行编辑，这正是常规育种希望达到的目的。但是在效率和可控制性等方面，基因组编辑技术明显优于常规育种手段。如果把常规的基因工程比作是大的外科手术，基因组编辑则相当于微创外科手术。美国农业部已经宣称基因组编辑作物不属于转基因生物范畴。虽然欧盟还未表态，但最近德国和瑞典的相关主管部门已经宣称，一些基因组编辑的作物和常规育种作物本质上是完全等同的。一向对转基因作物有强烈排斥意识的欧洲，在看待基因组编辑作物方面的态度明显趋向于平和，这对基因组编辑技术未来在农业上的应用是一件好事，至少不会遇到转基因作物那样尴尬的局面。

要对一个作物品种的遗传种质进行改良，在进行育种时还需要有多个参照基

因组，其中特别重要的是杂合基因组。通过对作物的遗传改良和育种，提高了作物的产量，同时确保了粮食生产安全。对一个物种而言，完整的高质量基因组序列是其广义研究中不可估量的宝贵资源，并且是基因组学、基因功能、分子和进化研究的坚实基础，基因组参考序列的质量在一定程度上也体现了该物种的研究进展和水平。水稻是重要的粮食作物，在遗传学、分子生物学及基因组学研究中具有重要地位，同时也是第一个全基因组测序的禾谷类作物。粳稻品种日本晴和籼稻品种 93-11 分别通过逐个克隆法(clone-by-clone)和全基因组鸟枪法(whole genome shotgun，WGS)进行了全基因组测序，其中日本晴参考序列被公认为现有作物基因组序列中质量最高的，但其中仍然存在着组装错误和空缺的问题。

近十年中，在植物基因组学研究中的诸多科学里程碑式的成果使得人们能够在分子水平对等位基因进行更精确的鉴定。这些里程碑式的成果包括拟南芥、水稻和白杨基因组的测序，表达序列标签(expressed sequence tags，EST)数据库的建立，芯片技术的产生以及拟南芥、水稻、玉米和其他农作物的广泛的突变体的收集、分子标记和大量的重组近交系资源。现在有很多使用基因组学的研究方法用以补充标准的正向或反向遗传学(reverse genetics)研究途径。EcoTILLING、基因芯片定位和结合映射均为能够辅助鉴定可导致优良性状的基因和等位基因的方法。然后，通过使用分子标记、连锁作图和关联分析对优良性状的基因进行准确定位，结合图位克隆技术锁定目标基因位点，可加速所期望的等位基因向优良种质的快速渗透。

获得一个物种的全基因组信息，对深入认识某些特异性状的遗传机制从而进行深度开发利用具有重要意义。然而，在地球上现有的 30 多万种高等植物中，目前已知全基因组序列的仅有一种双子叶植物拟南芥和一种单子叶植物水稻，只是众多物种世界中 1 到 2 个代表而已，物种全基因组信息是弥足珍贵的稀缺资源。但是，对于今后生物科学和农业生产来讲，仅有的一两个物种全基因组信息是远远不够的，还需要了解更多物种的全基因组序列。基于此目的，科学家针对一些重要的农作物，启动了基因组测序计划，其中包括玉米、大豆、番茄和马铃薯等。

2. 气象数据

气象与农业生产关系密切。我国长江中下游地区，有一条流传许久的农谚——“寸麦不怕水，尺麦怕寸水”。这是对小麦生产与雨水关系的总结。如何有效地利用气象数据来搞好生产，是现代农业发展的一个重要方面。为了发展现代农业和提高农业发展效益，解决现有农业生产中存在的各种供求矛盾，2014 年我国提出了“智慧农业”这一新概念。

智慧农业就是将物联网技术运用到传统农业中去，运用传感器和软件通过移动平台或者电脑平台对农业生产进行控制，使传统农业更具有“智慧”。可以根

此为试读，需要完整PDF请访问：www.ertongbook.com