

中国科学院国家科学图书馆

科学研究动态监测快报

2007年7月1日 第13期（总第22期）

先进工业生物科技专辑

中国科学院国家科学图书馆成都分馆主办

中国科学院国家科学图书馆成都分馆 四川省成都市一环路南二段十六号
邮编：610041 电话：028-85228846 电子邮件：zx@clas.ac.cn

目 录

重点关注

综述：微生物基因组数据资源1

短 讯

科技政策与科研计划

加州大学伯克利分校启动生物燃料前沿研究4

研究与开发

科学家希望利用森林植物资源制备新药5

糖基燃料将更具优势6

垃圾制成生物塑料7

综述：微生物基因组数据资源

引言

根据基因组在线数据库（Genomes OnLine Database）的统计，现在已经公开了 400 多种微生物基因组数据，另有 1000 多个正在进行的计划和一些准备启动的计划。世界各地的学术组织正在开展单个微生物菌的基因组数据和微生物群体的宏基因组学数据的测序，例如基因组研究所、美国能源部联合基因组研究所、Wellcome Trust Sanger 研究所、麻省理工与哈佛的 Broad 研究所，以及圣路易斯的华盛顿大学等。

对于独立的生物体而言，这些组织采用的数据研究策略大体上是相似的。一个微生物基因组的“死”序列可以被拼装到更长的相邻序列中，形成带有基因组草图性质的数据文件。通过迭代过程将相邻序列中的差别减小，即可得到完整的基因组。草图与完整的基因组可以用来预测潜在的基因，并确定被预测的基因带有的功能。

宏基因组序列数据的研究仍处于发展的早期阶段。由于一个微生物群体的宏基因组数据来自一系列的细胞单元，其中有些细胞在基因上关联且与同一种类的不同菌种有着潜在的联系，而另有些细胞在基因上却是截然不同的，所以宏基因组序列数据与单个基因组序列数据相比更加复杂、难以收集完整，因此质量较低。在单个基因组序列数据中常用的传统的基因拼装、基因预测和注释方法在宏基因组序列数据中起不到什么作用。随着时间的推移，如果现有的研究手段能够被更新或更高级的手段所取代，宏基因组数据的研究方法将有望取得更好的发展。

草图与完整的微生物基因组数据与不断增加的宏基因组数据在各种微生物基因组数据资源中被很好地组合起来，这些数据资源的共同目的是提高基因组的初级注释的准确性和完整性，补充基因表达的功能信息。独立的资源经常使用不同的技术和策略来综合和扩展初级注释的结果，例如，基因的功能特性可能来自单个基因的一个特定片段，也可能来自于一个支持多基因比较分析的综合片段。

由于微生物基因组的研究资源和研究方法多种多样，因此有必要理清每一个资源的类型和范围以及不同资源之间的关系、确定潜在数据的关联与转换关系。

微生物基因数据的生命周期

遍布全球的研究机构将它们获取和处理过的微生物基因数据全部公开在不同的数据资源中，并且数据往往事先经过了特定的资源转换。微生物基因数据在不同资源经过的生命周期如图 1 所示。其中的数据资源可参见表 1。

对公众开放的基因序列档案数据资源将会通过各种渠道收集微生物基因数据包，这些数据包括基因匹配物、序列名称、基因名称和一些附加的功能注释（例如

相关的蛋白团、蛋白质序列等)。

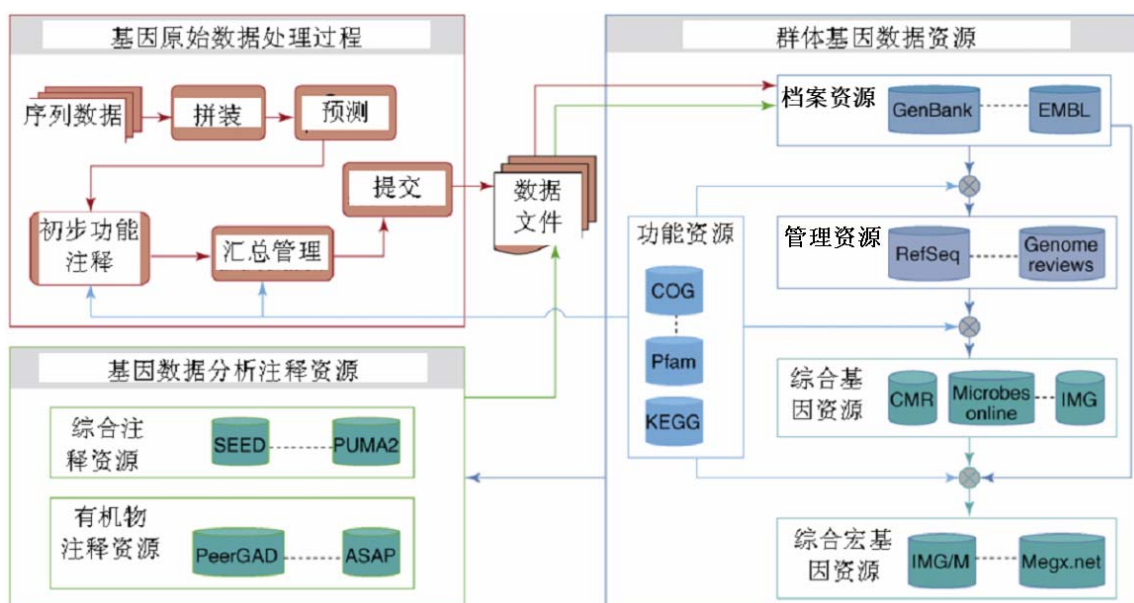


图 1. 微生物基因数据生命周期示意图。从不同的数据处理途径可以得到微生物基因与宏基因序列数据，这些数据被存放在公共的序列数据资源中；经过不同的资源可以逐渐对这些数据进行整合，以提高数据的一致性和基因功能特征的完整性。

表 1 微生物基因数据资源

数据库	URL 链接
ASAP	https://asap.ahabs.wisc.edu/asap/home.php
CMR	http://cmr.tigr.org/
Entrez Gene	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
ERGO	http://www.ergo-light.com/ERGO/
Genome Reviews	http://www.ebi.ac.uk/GenomeReviews/
GOLD	http://www.genomesonline.org/
IMG	http://img.jgi.doe.gov/
IMG/M	http://img.jgi.doe.gov/m
Integr8	http://www.ebi.ac.uk/integr8/
MBGD	http://mbgd.genome.ad.jp/
Megx	http://www.megx.net/
MicrobesOnline	http://www.microbesonline.org/
PUMA2	http://compbio.mcs.anl.gov/puma2/
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/
SEED	http://theseed.uchicago.edu/
UCSC Genome Browser	http://archaea.ucsc.edu/

档案资源为同一基因组保留不同版本的数据包；同时，由于数据包的提供者采用了不同的处理方法、数据管理策略以及确定功能特征的技术，这类资源提供的数据包各自具有不同的精确度和解析度。各种不同的资源会把这些来自基础资源的微生物基因组数据进一步处理，其目的在于：第一，提供微生物基因组序列的最新报

道；第二，逐渐地提高数据的一致性以及基因功能特征注释的完整性。以美国国家生物技术信息中心(NCBI)的 RefSeq 资源和欧洲生物信息学研究所(EBI)的 Genome Reviews 资源为代表的管理型公共资源就是围绕上述目的开展研究的，其研究手段包括将来自基础资源的基因组序列数据进行合成、检查与管理。基因组序列的注释方法还扩展到了包括手动管理、计算分析以及与其它资源相结合等。

微生物基因组学可以通过使用资源所提供的各种分析工具来进行研究，以进一步丰富公众性的档案资源和管理资源中的数据。例如，NCBI 的 Entrez Gene 支持对 RefSeq 资源中附有功能注释的单个微生物基因组数据进行基因组中心研究 (gene-centric exploration)。EBI 的 Integr8 也可以对来自 Genome Reviews 资源的附有功能注释的微生物基因组数据进行类似的处理，并能够利用数学统计方法提供特定基因的综合分析 (genome-specific summaries)。UCSC Genome Browser 资源能够支持结合计算分析（如操纵子）和实验研究（如生物芯片）获得的数据与档案型基因组学数据，进行基因组中心研究。

一些资源还具备多重基因组对比分析的能力，例如 CMR、ERGO、Microbes Online、PuMA2、IMG 和 MGD 等资源。其中，大多资源能够进一步修正微生物基因组的注释信息，例如，CMR 资源可以利用 Genome Properties 系统对微生物基因组数据进行功能特征的确定，该系统主要根据代谢途径、细胞活性与细胞结构来研究基因功能的注释。还有一些资源能够支持群体基因的注释，这一类的注释通常是不断积累起来的，即这类注释可以补充入已有的注释，而不是代替已有的注释。

注释的积累模型可以使科学家更好地研究基因组的特征功能，但也可能会产生多个注释之间不一致或相互矛盾的结果。例如，PeerGAD、PseudoCAP 和 ASAP 资源都采用了这样的模型，它们都注明了这种情况，并根据所采用的注释的不同复杂程度采取了相应的机制。基因语义 (gene ontology, GO) 等本体论和控制语句已被用于功能角色的统一描述。同类注释机制已经开始用于描述特定的生物体，例如假单胞菌 *Pseudomonas syringae* pv. *tomato* 和 *Pseudomonas aeruginosa* pv.

不同于特定生物的注释系统，SEED 资源采用的子系统可以同时多种生物体的基因进行注释，子系统带有一系列功能角色，能够执行特定的生物过程。通过采用不断扩展的 SEED 特定控制语句，子系统的特定专家群体能够维护每一个子系统的基因注释集。为新基因组数据开发的子系统带来了特定基因组子系统差异。在这里，主要使用整合机制而不是同类注释机制来解决注释不一致和相互矛盾的问题。

经管理的（已检查、校正和扩充的）微生物基因组注释通常也提交给档案型基因组序列数据库，作为基因组数据包的初始版本的修正，并最终与其它资源整合在一起。

与单个微生物基因组数据的获取方法相类似，微生物群落的宏基因数据也是首

先提交到档案型数据资源——例如在 GeneBank 资源中，数据包以分开的形式提供，只支持 BLAST 式搜索（参见 <http://www.ncbi.nlm.nih.gov/BLAST/Genome/EnvirSamplesBlast.html>）。同时，对于微生物群落中的海量微生物而言，基因拼装和注释的质量与单个微生物基因组草图相接近，单个的数据包的记录方式与单个基因组数据的记录方式也大体相同。

尽管目前在获取和处理方面还存在一些局限，宏基因组学数据还是可以通过一些有价值的分析工作得到改善，人们最近已经对此进行了新的研究和阐述。研究宏基因组数据需参照单个基因组数据的处理方法，或参照生物材料特性研究过程中获得的宏基因组序列的样本。例如，以 IMG/M 资源和 Megx.net 资源为代表的新兴宏基因组数据资源不仅具备单个基因组数据的对比分析功能，还提供宏基因组数据的对比分析功能。

陈方 综述自 Current Opinion in Biotechnology 2007, 18:267-272

检索日期：2007年6月26日

短 讯

科技政策与科研计划

加州大学伯克利分校启动生物燃料前沿研究

美国政府 2007 年 6 月 26 日宣布，美国劳伦斯伯克利国家实验室将投资 1.25 亿美元创建一个生物燃料研究中心。该研究中心作为美国能源部资助的三个中心之一，将建立在东海湾附近地区，该中心会是生物技术工业和硅谷的核心，将会刺激一个绿色产业的兴起并帮助开发新一代生物燃料。恰恰在五个月以前，加州大学伯克利分校获得了石油巨头 BP 公司的史无前例的 5 亿美元资助，以成立一个单独的生物燃料研究所。

在以上两个项目的研究中，其中一个目标是从植物纤维生产可再生的公路运输用燃料，以缓解全球变暖和空气污染，降低美国对外国石油进口的依赖，开辟新的国际能源市场。一些专家认为，如果研究成功，为发动机提供可替换燃料仍需 5—10 年的时间。

该项目在未来五年内每年都会得到 2500 万美元的资助，用于研究高技术化学，以从植物纤维生产清洁燃料。该中心命名为联合生物能源研究所（Joint BioEnergy Institute），最终将会雇用 130 名科学家、研究人员和支撑人员，合作单位还包括加州大学戴维斯分校、斯坦福大学、美国劳伦斯利物莫国家实验室（Lawrence Livermore National Laboratory）以及 Sandia 国家实验室。

能源部资助的另外两个生物能源中心位于美国橡树岭国家实验室和威斯康星大学。三个中心联合着 18 所大学、7 家联邦实验室和至少 1 家非盈利性组织，将从事基础的生物燃料研究。

如果生产成本能显著降低，纤维素乙醇与基于玉米谷物的乙醇比较更具价格低廉、能效高等潜力，而这要依赖于植物木质和纤维部分碳的分解。

该新建的研究中心包括四个研究分部：

- ◆ 第一个分部研究新植物，从农场作物中获取可再生碳，以代替不可再生的石油和煤炭；
- ◆ 第二个分部研究用酶来打扫环境，从而把植物纤维转化成糖；
- ◆ 第三个分部研发生物燃料，包括从纤维来源的乙醇，以及其他富能燃料，如丁醇、煤油和航空燃料等；
- ◆ 第四个分部支持以上三个研究链和生物燃料工业。

纤维素乙醇的发展可以为小农场的发展带来新的契机。

陈云伟 译自 <http://www.sfgate.com/cgi-bin/article.cgi?file=/c/a/2007/06/27/BIOFUELS.TMP>

检索日期：2007 年 6 月 27 日

研究与开发

科学家希望利用森林植物资源制备新药

——北安大略形成潜在生物制品经济中心

北方针叶林可能会成为新的亚马逊雨林，这就是北美五大湖森林研究中心（Great Lakes Forestry Centre, 缩写 GLFC）和北安大略医学院（Northern Ontario School of Medicine, 缩写 NOSM）所研究的目标。

长期以来这里的针叶林一直是 paclitaxel 这类抗癌成分的原产地，从现在起又有人开始在这里发掘其他的有益生物化学成分。过去人们一般去亚马逊和南美洲研究新药，但近来，北美五大湖森林研究中心的科学家在北方针叶林就可以寻找到有用资源。

北美五大湖森林研究中心正与北安大略医学院合作启动一项北部生物探测计划“the Boreal Bioprospecting Initiative”。该探测计划将研究利用有营养或医疗价值的经济林产品如木材、菌类、药草和灌木的方法，对其具有经济价值的分子、有机或基因材料进行研究。该探测计划的核心是一个大型材料库，包涵大约 1000 种天然产物粗植物提取物和 800 种纯化合物，在这其中有许多是有着抗氧化性质的新天然产物，它们可以防治和中断经触发我们机体的过激内在防卫反应而产生的疾病。这个库对寻找癌症、中风、帕金森等相关疾病的新治疗方法可能会作出关键贡献。

研究人员已经发现了北部的植物与他们南部邻居的区别非常大。由于南美洲有着较北部林区更多的昆虫或植食性动物，那里的植物已经进化出一种“低密度多成分”的性质。而针叶林植物，由于它们的生长期相对较短，则集中产生高浓度的某些成分。

目前研究人员已发现了一些新的研究方向，如枫蜜表现出的抗氧化性。项目主管 Abou-Zaid 认为，他们筛选了成千上万的用于治疗某类疾病的化合物，其工作是真正的高基准、高风险性的，但一旦成功，就会拥有自主知识产权。

目前，通过与渥太华 Ensyn 科技公司的合作，Abou-Zaid 已经申请获得了一项国际专利，该专利方法可更加有效地从东部红豆杉的针叶和枝条中提取 paclitaxel。

加拿大现正生产相当大量的紫杉醇，但还需要达到在技术上优于他人的更高目标，同时面临与全世界的竞争。

郑颖 译自 <http://www.saultstar.com/webapp/sitepages/content.asp?contentid=574418&catname=Local%20News&classif=>，检索日期：2007年6月26日

糖基燃料将更具优势

威斯康星大学的研究人员发现了一种将大量可再生生物质资源（包括有生命的或刚刚死亡的生物材料，用于燃料或工业生产）转化为液体燃料的新路径。

研究证明，糖基燃料能有效驱动电动机，并且减少对石油等化石燃料的需求，从而减少碳的排放和温室气体的生成。

威斯康星大学的 James Dumesic 与其合作者发现了一种将果糖（或一般的糖或单糖）转化为 2,5-二甲基呋喃（2,5-dimethylfuran）的方法，一种缩写为 DMF 的合成燃料。DMF 可以有效地被分配、储藏和被发动机燃烧，且其效率比玉米基燃料（乙醇）更高。

美国的研究人员认为两年内即可使他们的两步生产方法完全成熟，DMF 由糖生成，与乙醇相似，但是能源密度比乙醇高出 40%，这种密集的材料能生产更多的汽油。该两步方法包括：

首先，向糖和水混合液中添加盐和酸催化剂，然后将该溶液与碳氢化合物混合，此步产生一种称为 5-羟甲基糖醛（5-hydroxymethylfurfural, HMF）的中间化合物；

第二步，将 HMF 与氢及铜-钌基催化剂混合，此步 HMF 转化成 DMF。

该方法很容易实现量产，然而，还需要更多的研究以保证不会产生有毒的副产物。

陈云伟 译自 <http://www.itwire.com.au/content/view/13137/1066/>

检索日期：2007年6月27日

垃圾制成生物塑料

二氧化碳、橘皮、鸡毛、橄榄油、土豆皮、大肠杆菌，化学家就像是收垃圾的一样在搜寻这些东西，来制造生物可分解的、可持续利用的、可再生的塑料。大多数生物塑料来源于植物如玉米、大豆、甘蔗以及柳枝稷，但是最近科学家们转向了垃圾，致力于制造被称作是绿色高分子的产品，本质上就是利用垃圾制造的塑料。

Geoff Coates 是康奈尔大学的一名化学家，也是创造绿色高分子的领头人之一。Coates 的实验室大概占了康奈尔 Spencer T. Olin 实验室的整个 5 楼，而附近的一家煤电厂是 Coates 的灵感源泉。在这个实验室，Coates 发现了能把二氧化碳转化为聚合物的催化剂。

对生物复合材料的探求源自 1913 年，当时，一名法国科学家和一名英国科学家都提出了大豆基塑料专利的申请。

早期生物塑料的研究多数由 Henry Ford——一个坚信大豆潜能的人资助。一张著名的 1941 年的照片上，Henry Ford 正挥动斧头去砍一辆小汽车的后部，旨在表明大豆制作的生物复合体的强度甚至可以用于制造汽车的车身。但是，很快大豆输给了石化基塑料，因为在那时石油供给充足，而且不像大豆一样有生长季节的限制；另外一个重要问题是渗透性：大豆塑料不防水。石油是生物和化学相对惰性的，而绝大多数生物系统需要水。

化石燃料很快占领了塑料市场，农业基塑料还远远落后，随着 Coates 和其他人开发的催化剂的发展，一种全新的聚合物系列将商业化。

选择二氧化碳作为高分子的原料不是一个显而易见的选择。混合二氧化碳和某种环氧化合物，是不会形成聚合物的，除非有正确的催化剂。Coates 的催化剂—— β -二米基醋酸锌盐，或者用 Coates 的话来说是“锌基仙子”，能够把一个地质级的反应时间加速到实验室级反应时间。

绿色高分子商业似乎在各处兴旺起来。一家德国公司——Rodenburg 生物聚合物公司采用土豆废渣来制备塑料。而在 Boston 的 Metabolix 公司则采用基因工程修饰的大肠杆菌来生产天然的聚醚类高分子。

田雅娟 译自 <http://www.iht.com/articles/2007/06/19/healthscience/snbioplastic.php>

检索时间 2007 年 6 月 27 日

版权及合理使用声明

中科院国家科学图书馆《科学研究监测动态快报》（简称《快报》）遵守国家知识产权法的规定，保护知识产权，保障著作权人的合法权益，并要求参阅人员及研究人员认真遵守中国版权法的有关规定，严禁将《快报》用于任何商业或其他营利性用途。未经中科院国家科学图书馆同意，用于读者个人学习、研究目的的单篇信息报道稿件的使用，应注明版权信息和信息来源。未经中科院国家科学图书馆允许，院内外各单位不能以任何方式整期转载、链接或发布相关专题《快报》。任何单位要链接、整期发布或转载相关专题《快报》内容，应向国家科学图书馆发送正式的需求函，说明其用途，征得同意，并与国家科学图书馆签订协议。中科院国家科学图书馆总馆网站发布所有专题的《快报》，国家科学图书馆各分馆网站上发布各相关专题的《快报》。其它单位如需链接、整期发布或转载相关专题的《快报》，请与国家科学图书馆联系。

欢迎对中科院国家科学图书馆《科学研究监测动态快报》提出意见与建议。

中国科学院国家科学图书馆

National Science Library of Chinese Academy of Sciences

《科学研究动态监测快报》(简称《快报》)是由中国科学院国家科学图书馆编辑出版、由中国科学院规划战略局等中科院的职能局和专业局支持指导的半月信息报道类刊物,于2004年12月正式启动。每月1日或15日出版。2006年10月,国家科学图书馆按照统一规划、系统布局、分工负责的思路,对应院1+10科技创新基地,重新规划和部署了系列化的《快报》。系列《快报》的重点服务对象首先是院领导、院专业局职能局领导和相关管理人员;其次是包括研究所领导在内的科学家;三是院外相关科技部委的决策者和管理人员以及相关重点科学家。系列《快报》内容将恰当地兼顾好决策管理者与战略科学家的需求,报道各科学领域的国际科技战略与规划、科技计划与预算、科技前沿与热点、重大研发与应用、科技政策与管理等方面的最新进展与发展动态。

系列《快报》共分12个专辑,分别为由中国科学院国家科学图书馆总馆承担的交叉与重大前沿专辑、现代农业科技专辑、大装置与空间科技专辑、科技战略与政策专辑;由兰州分馆承担的资源环境科学专辑、地球科学专辑;由成都分馆承担的先进工业生物科技专辑、信息科技专辑;由武汉分馆承担的先进能源科技专辑、生物安全专辑、先进制造与新材料科技专辑;由上海生命科学信息中心承担的生命科学专辑。

编辑出版:中国科学院国家科学图书馆

联系地址:北京市海淀区北四环西路33号(100080)

联系人:冷伏海 朱相丽

电话:(010)62538705、62539101

电子邮件:lengfh@mail.las.ac.cn; zhuxl@mail.las.ac.cn:

先进工业生物科技专辑

联系人:邓勇 房俊民

电话:(028)85228846、85223853

电子邮件:dengy@clas.ac.cn; fjm@clas.ac.cn