

合成宏基因组学：将数字信息转化为生物学功能



美国能源部联合基因组研究所（DoE JGI）的研究人员已经开发出了建立在高通量筛选技术基础上的合成宏基因组学技术。这种技术可以将最新一代测序获得的大量数字信息转化为有形的、可观察到的生物体。

简介

新一代测序技术已经发现了数以百万计的新型基因，包括所有的已知通路和功能活性。这些新型基因里面包括了大量新型生物酶的基因，在能源和环境技术方面具有潜在用途，包括基础研究以及应用生物学技术等。最大的一个挑战就是如何将大量的数字信息（序列，生物信息学，化学信息学）转化为可以观察、具有功能活性的生物（酶活性，细胞死亡，蛋白表达等）。

DoE JGI 已经开发出了合成宏基因组学技术来克服这一挑战。基因和通路以一种不依赖模板的方式被合成，快速经济有效地将数字信息转化为生物化学分子。

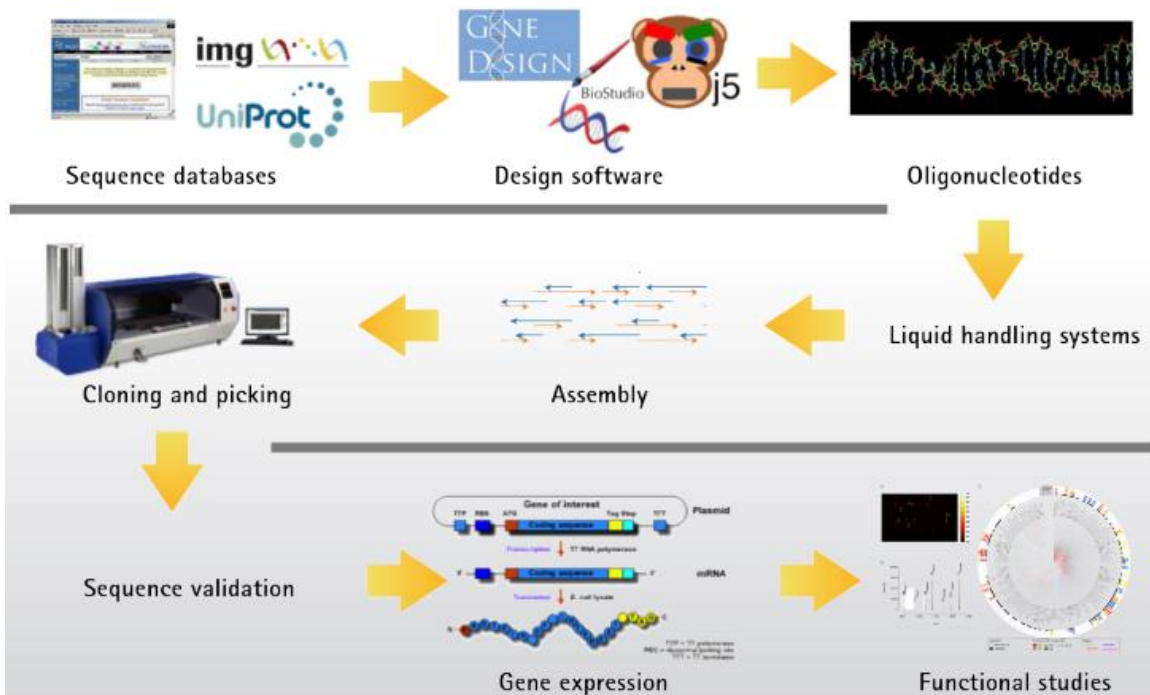
对于商业化生产纤维质生物燃料来说，一个主要的瓶颈就是缺少具有稳定高活性的酶，用于将生物质转化为葡萄糖。离子液体可以被用于前处理溶解纤维素，增加可发酵糖的含量如葡萄糖。为了寻找合适的工业相关的酶，要使用合成宏基因组学方法。能源部联合生物能源研究所（JBEI）设计了特殊工业用途的酶，这些酶要具备下面的条件：（i）70 度的时候仍具有功能性酶活；（ii）pH4.5 条件下仍然稳定；（iii）对离子溶液具有耐受性。这个项目开始的时候，没有酶能够在这些条件下具有足够的活性。

Figure 1. 合成宏基因组学流程：DoE JGI 使用的一个宏基因组学研究流程

宏基因组学研究的标准流程图：在克隆和挑取阶段使用 QPix460 系统可以显著缩短整个流程的时间，增加效率，避免了耗费人力易出错的人工过程。全自动的工作流程一周可以准确处理上万个克隆，而且从涂布到挑取，具有完整的数据记录追踪，大大节省成本。

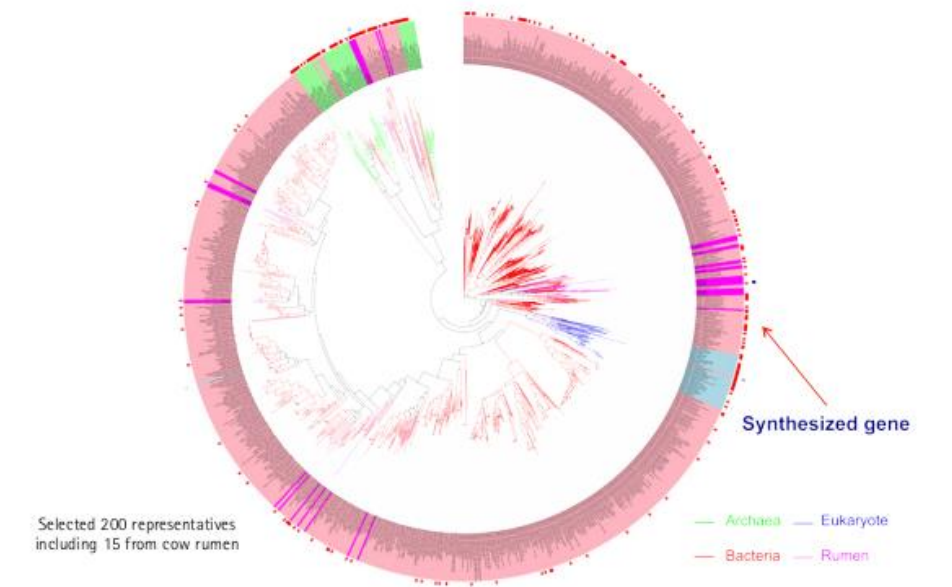
优势

- 同时检测克隆和定量荧光标记--使用表现型筛选克隆
- 从加样、涂布到挑取完全自动化工作流程
- 从样品涂布到挑取、复制和重排记录并追踪样品数据
- 每天准确挑取超过 30000 个筛选到的克隆，效率>98%



第一个研究目标就是已知的可以催化纤维二糖转化为葡萄糖的 GH1s 酶家族。基于系统发育的方法使我们能够从已有的上千序列里面选取 200 个基因，获得最大的功能多样性。200 个基因里面，其中 180 个基因被合成，然后使用 QPix460 系统克隆到表达载体（Figure 2）。

Figure 2. 宏基因组学系统发育树分析，选择有代表性的基因家族

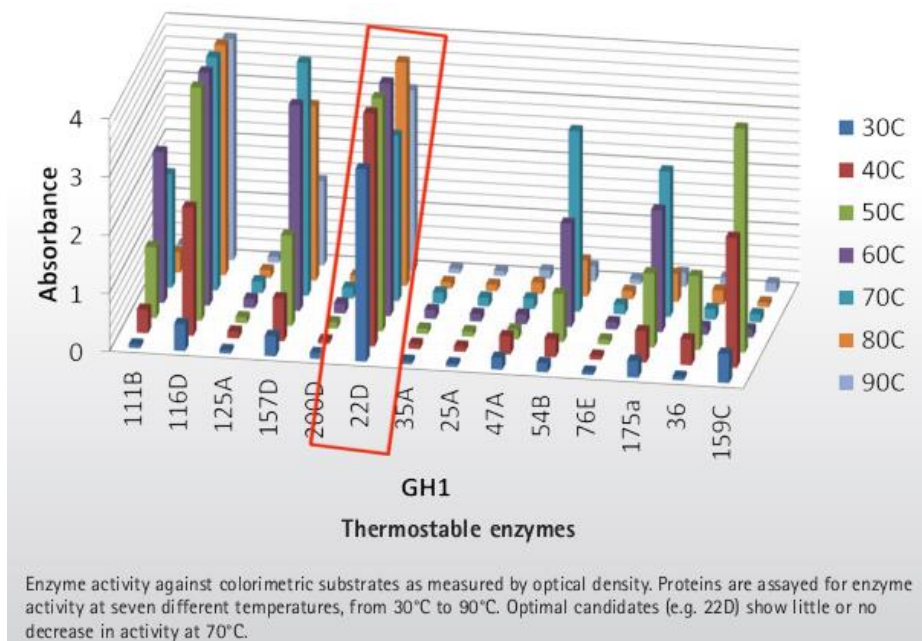


Tree diagram of the metagenome, including bacteria (red), eukaryotes (blue), cow rumen (purple) and archaea/ "extreme prokaryotes" (green). Long branch length indicates increased genetic diversification, whereas short length does not. 200 representative samples from shorter branch areas were selected, including 15 from cow rumen to contribute to phylogeny-driven sample diversity.

DoE JGI 开发的高通量自动化流程包括平行转化、涂布以及从每个基因变异里面挑取多样的克隆，用于测序鉴定。QPix460 系统使这些成为可能，在一个平台上面可以提供高通量的接种、涂布以及克隆挑取。一旦序列验证正确，每个基因的最优克隆被转化到蛋白表达菌株。

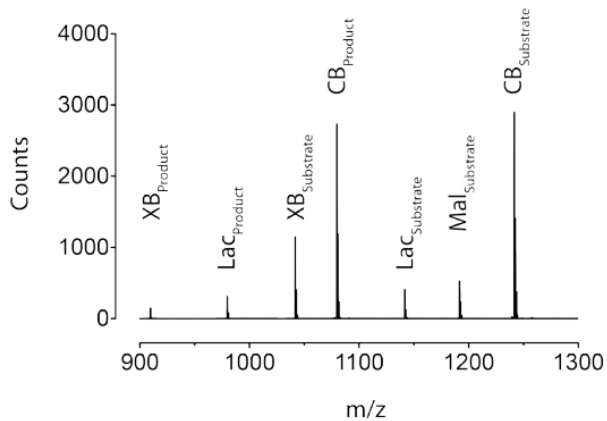
然后分析 GH1 蛋白生产菌株的蛋白表达可溶性。65% 克隆表达的蛋白具有可溶性，然后在上面描述的条件进一步分析。例如，作为早期质量控制，在不同温度条件下分析一系列酶的活性，结果显示，其中一些候选酶具有热稳定性（Figure 3）。

Figure 3. GH1 酶家族活性鉴定，发现具有热稳定性的酶



为了评估这些候选酶在不同条件下的特性，需要约 15000 个反应，筛选这些反应就需要一个超高通量的技术。为此目的，使用了一个基于质谱的高通量技术——纳米结构启动质谱（NIMS）。酶-底物反应物被点样到固相载体上，并通过激光电离。通过检测底物和产物的比值确定酶的活性（Figure 4），选取活性在前 20 的酶进行深入研究，其中有 5 个在上面描述的工业相关条件下具有活性。这表明使用合成宏基因组学可以从公共序列数据库预测的蛋白开始，快速找到具有特殊活性的酶。

Figure 4. 基于 NIMS 的底物和产物的比值



NIMS analysis revealed the top 20 performing enzymes based on the optimal reactivity with substrates as measured by the ratio of substrates to products.

结论

这些工作证明 DoE JGI 开发和使用的技术可以成功地将数字信息转化为生物学相关、具有新功能的酶。来自于 GH1 酶家族的 180 种蛋白，经过 QPix460 系统的高通量、自动化筛选，最后成功找到符合要求的、具有新特性的酶蛋白。这个系统已经成为高效、自动化开发和管理宏基因组库的不可缺少的组成部分。

此外，使用这个高效的自动化系统，成功筛选到 5 个具有新特性的酶。这些酶在 70°C、pH 4.5、高浓度离子溶液中和底物仍然具有较好活性，能够满足工业化的要求。JBEI 目前正在研究这些酶的潜在工业化用途，如能否有效工业化放大，将生物质转化为葡萄糖。

参考资料

1. A nanostructure-initiator mass spectrometry-based enzyme activity assay. Northen TR, Lee JC, Hoang L, Raymond J, Hwang DR, Yannone SM, Wong CH, Siuzdak G. Proc Natl Acad Sci USA. 2008 Mar 11;105(10):3678-83. doi: 10.1073/pnas.0712332105. Epub 2008 Mar 4. PMID:18319341 [PubMed - indexed for MEDLINE]

目前使用 QPix 从事相关应用的国内外用户：

1、US Department of Energy's Joint Genome Institute (DoE JGI) QPix460

2、美国可再生产品制造商 Amyris 购买了 3 台 QPix （开发使用藻类生产生物燃料和将甘蔗转化为乙醇）

3、武汉生物技术研究院 合成生物学实验室 QPix460

美国能源部联合基因组研究所（DOE JGI）位于加州 Walnut Creek，成立于 1997 年，以将能源部在 LBNL、劳伦斯利弗莫尔国家实验室（LLNL）和洛斯阿拉莫斯国家实验室（LANL）基因组中心率先进行 DNA 测序、信息科学和技术发展中的专长和资源整合在一起。1999 年，为加快完成能源部对人类基因组计划的承诺，负责管理 DOE JGI 的美国加州大学，将在加州 Walnut Creek 轻工业园区实验室和办公室租赁出去，以巩固各项活动。这样做取得的重大规模经济使联合基因研究所成为第一个在自然杂志上发表的目标染色体 5、16 和 19 序列分析。继这一成就之后，DOE JGI 通过测序 20 多个微生物物种几种典型的生物体和将这一信息无偿地贡献给公共数据库继续推进基础科学。2004 年，DOE JGI 将自己确定为全国用户设施，而今天有超过 2000 名世界各地的用户。DOE JGI 测序的绝大多数是在生物群落测序项目（CSP）赞助下进行的，测量生物圈以表征与能源部生物能源、全球碳循环和生物地球化学科学任务范围有关的生物体。DOE JGI 的最大客户是美国能源部生物能源研究中心（BRCs），这些中心成立于 2007 年，旨在加速开发下一代纤维素生物燃料的基础研究。该研究所继续得到来自美国能源部科学局生物和环境研究处的大部分资金支持。

目前，DOE JGI 已发展到占地八万平方英尺，聘请由国际公认的遗传学家、医学博士和哲学博士 Eddy Rubin 领导的 250 名员工。DOE JGI 的伙伴实验室包括：LBNL、劳伦斯利弗莫尔国家实验室、洛斯阿拉莫斯国家实验室、橡树岭国家实验室和太平洋西北国家实验室（PNNL）以及哈森阿尔法生物技术研究院（以前与斯坦福人类基因组中心有关）。