

用微生物基因组开发对付感染性疾病的新措施

谢建平^{1,2} 陈永青¹ 王洪海^{1*}

(复旦大学生命科学学院 上海 200433)¹ (西南师范大学生命科学学院 重庆 400715)²

摘要: 综述微生物基因组研究对于深化认识微生物之间关系的作用以及利用微生物基因组开发抗感染性疾病的主要手段和进展, 其中重点是利用生物信息学手段、基因芯片、噬菌体表面展示系统筛选新的抗微生物药物作用靶点。

关键词: 微生物基因组学, 噬菌体表面展示, 核心基因, 分子进化

中图分类号: Q93 **文献标识码:** A **文章编号:** 6253-2654 (2002) 02-0081-05

以 1995 年基因组研究所 (The Institute for Genomic Research-TIGR) 发表流感嗜血杆菌的全基因组序列为标志, 微生物学研究和运用进入了基因组时代和功能基因组时代。目前已完成 56 个微生物的全序列测定 (www.gigr.org/tdb/mdb/mdb.html) (见表 1), 17 个微生物基因组序列正在注释, 110 个左右微生物的基因组序列测定正在进行。现在几乎 2 个月发表一个微生物全基因组序列。

1 微生物基因组研究概况

全基因组序列提供的许多其他途径很难获得的信息。如种独特基因, 保守蛋白, 基因组 GC 含量, 包括整体含量以及变化情况, 重复/插入序列, 构建生理和运输途径,

* 联系作者

收稿日期: 2001-04-22, 修回日期: 2001-11-12

随后分析基因功能。首先完成的具有医学重要意义的微生物的序列测定, 包括人致病菌序列测定。然后是已知最简单、容易生长或在动物内比较容易培养的微生物和模式生物。正在测序的微生物中, 50% 是非致病菌^[1]。它们在基础和应用领域有重要意义: 解决环境问题, 包括废物处理, 再生以及生物修复技术, 人和动物的营养来源, 阐明基础生物学问题如进化等。

表 1 部分完成基因组全序列测定的微生物 (www.gigr.org/tdb/mdb/mdb.html)

Organism	Genome size(Mbp)	ORFs	Unknown function	Unique ORFs
<i>A. permix</i>	1.67	2694	2061(76%)	1538(57%)
<i>A. fulgidus</i>	2.18	2437	1315(54%)	641(26%)
<i>M. thermotautotrophicum</i>	1.75	1855	1010(54%)	496(27%)
<i>M. jannaschii</i>	1.66	1749	1076(62%)	525(30%)
<i>P. horikoshii</i>	1.74	2061	859(42%)	453(22%)
<i>A. aeolicus</i>	1.50	1521	663(44%)	407(27%)
<i>B. subtilis</i>	4.20	4100	1722(42%)	1053(26%)
<i>B. burgdorferi</i>	1.44	1751	1132(65%)	682(39%)
<i>C. trachomatis</i>	1.04	894	290(32%)	255(29%)
<i>D. radiodurans</i>	3.28	3192	1715(54%)	1001(31%)
<i>E. coli</i>	4.60	4288	1632(38%)	1114(26%)
<i>H. influenzae</i>	1.83	1692	592(35%)	237(14%)
<i>H. pylori</i>	1.66	1657	744(45%)	539(33%)
<i>M. tuberculosis</i>	4.41	3924	1521(39%)	606(15%)
<i>M. genitalium</i>	0.58	470	173(39%)	7(2%)
<i>M. pneumoniae</i>	0.81	677	248(37%)	67(10%)
<i>S. ynechocystis</i> sp.	3.57	3168	2384(75%)	1426(45%)
<i>T. maritima</i>	1.86	1877	863(46%)	373(26%)
<i>T. pallidum</i>	1.14	1040	461(44%)	280(27%)
<i>R. prowazkii</i>	1.10	834	48(12%)	207(25%)
<i>C. pneumoniae</i>	1.23	1073	437(40%)	186(17%)
<i>L. lactis</i>	2.35	1495	398(27%)	83(6%)

注释基因功能微生物基因组测序表明: 每个种都有近一半的功能未知开放阅读框(ORF)和近 25% 的功能独特、与其他已知蛋白质序列无明显相似性的 ORF。这也提示“模式生物”的提法对于微生物可能不甚恰当。多数非致病菌是最近才发现的。现有许多方法可以对基因功能进行注释和预测。(1) 系统演进谱-phylogenetic profiling^[2] 该方法根据特定分布模式在不同种的存在与否进行注释。同一代谢途径中的基因往往具有共同的分布类型, 通过与已知基因的分布类型比较, 可以揭示功能未知基因的功能; (2) 特定途径中丢失某些基因也可以作为预测基因功能的方法^[3] 利用蛋白质功能域将蛋白质连接为不同功能组, 进行功能预测^[4] 由两个二结构域构成的蛋白质发挥作用需要两个结构域都存在。用一个种里存在的二结构域的蛋白质对只含有其中任何一个结构域的蛋白质进行分类, 看它们是否属于同一个代谢途径, 属于同一途径的基因归为一类。该方法也适用于包括多个结构域的蛋白质。因此可以预测多个基因的功能。目前急需建立比较完整的数据库以及不同实验室之间具有可比性的注释方法。

2 微生物基因组研究深化了对微生物的认识

2.1 再次强调微生物的多样性

从进化和环境的观点, 微生物都是生物圈的基础。微生物比动物和植物的出现要早 30 亿年。地球生物量的 60% 是微生物。微生物在遗传、代谢途径和生理方面的多样性比动物和植物要高。微生物能够迅速适应多样性的极端

环境如高温、酸度、盐浓度和缺氧。估计有 $10^6 \sim 10^7$ 种微生物, 目前我们已经能够纯培养的 2×10^5 , 只是总量的 1% ~ 10%^[21] (见表 2)。大多数不能培养的微生物在整个地球生态关系的维持中有重大作用。当然, 对实验室可培养微生物的基因组研究将有助于我们认识基础的生物学过程以及微生物之间的进化关系。

2.2 微生物进化 过去 20 年, 关于分子进化主要是 Woese rRNA 小亚基系统演进树。生物包括 3 个不重叠的界: 细菌, 古生菌和真核类。古生菌类和细菌的细胞结构类似, 古生菌是真核

表 2 微生物的估计种类数目和已经知道的种类数目^[1]

类别	估计数	已知	已知数占估计数的比例
病毒	130, 000	5000	4
古生菌	?	< 500	?
细菌	40, 000	4800	12
真菌	1500, 000	69, 000	125
藻类	60, 000	40, 000	67

类的祖先。但是从全基因组序列来看, 就单个蛋白质而言, 从系统树的根部到各分支, 结论往往截然不同。这提示基因进化不等同于种的进化。原因有: (1) 基因侧向转移如细胞器基因转移到核, 细菌不同种之间交换耐药基因等。这种转移不仅在细菌之间, 在细菌和古生菌之间也可能发生。 *Aquifexaeolicus*, *Thermotaga maritima* 20% ~ 25% 的基因更接近古生菌的而非细菌。同一种包括不同来源的基因, 不同基因有不同的历史, 这对于研究种的进化不利。(2) 其他原因: 基因取代, 基因重复, 在分子水平的趋同。因此, 关于分子进化还需要收集多方面的证据才能得出结论。这也是微生物学基础研究的重要任务。

尽管微生物进化过程中发生过广泛的基因侧向转移, 但各种微生物保留了某些系统演进信号, 即所谓的“进化核心”。通过比较 4 种微生物的全序列发现存在共同的“核心基因^[10]”, 核心基因占各种微生物全部基因组的 31% ~ 35%, 主要编码参与基因组复制和表达的蛋白质。核心基因与真核生物的相应部分比较接近, 而其他只在一两个种中出现的基因更接近细菌的相应部分。这可能是在古生菌的进化中, 细菌基因侧向转移到古生菌, 而与基因组复制和表达有关的核心基因未发生侧向转移。之所以核心基因未发生侧向转移, 原因可能是复制和表达需要的机器比较复杂-由复制体, 转录体和核糖体等多个蛋白构成, 这使得基因侧向转移及被固定下来比较困难。

根据基因含量对全基因组构建平均系统树^[5], 不同基因组共同的、基于多个基因的进化树与根据 rRNA 基因和 RecA 基因得到的进化树十分相似。这是对单基因系统演进的发展。通过分析特定基因存在与否^[7]以及利用高阶分类方法^[7]都得到了类似结果。古生菌基因组中有保守核心^[14], 也有些基因易转移。许多种微生物的基因密度是一致的, 大约每 1000 个碱基一个基因。这些结果提示有一个基本、共同的进化模式, 当然, 个别基因未必遵循该进化模式。通过比较近缘种的基因组还可能发现突变过程、密码子使用等进化非常迅速的特征。

对微生物基础生物学过程的认识有助于我们开发对付微生物感染性疾病的新措施。比如诊断方法、新抗生素和疫苗。

3 微生物基因组研究的应用

3.1 利用微生物基因组和新技术鉴定靶点 除了分子进化研究外, 致病菌全基因组序列为临床提供了新机遇。微生物基因组学面临的巨大挑战是基因组功能分析: 如何大

规模利用新的序列信息将静态的基因组变成微生物细胞中基因表达调控的动态网络,以帮助开发新的抗生素、诊断试剂和疫苗。常用方法有基因芯片、噬菌体表面展示、双向电泳、酵母双杂交和大规模突变等。

抗生素开发比较漫长而艰巨。首先往往需要鉴定靶点,靶点一般是致病菌感染过程中生存所必需的基因及其产物,确定基因是否为必需的金标准是克隆该基因,研究突变或敲除该基因的效应;

生物信息学从基因组序列直接发现一些重要线索:AA-tRNA 合成是细胞存活必不可少,参与氨酰 AA-tRNA 合成的酶具有高度的种特异性。许多致病菌的 AA-tRNA 合成途径与其哺乳动物宿主的相应途径没有任何关系。AARS 酶在过去 25 年中一直是开发抗生素的理想靶点。例如^[11]酪氨酰 tRNA 的抑制剂 indolmycin, 异亮氨酸 tRNA 的抑制剂 pseudomonic acid, 苏氨酸 tRNA 的 borrelidin, 苯丙氨酸 tRNA 的 ochratoxin。通过比较基因组发现:结核分枝杆菌的基因组中缺乏标准的天冬氨酰 tRNA 合成酶和谷氨酰氨 tRNA 合成酶^[11]。因此,结核分枝杆菌等致病菌特异性的 AA-tRNA 合成途径及其相关的酶可望成为新的治疗药物和诊断的新靶点。

基因芯片发现基因组中可能的药物作用靶点和验证先导物^[9],在结核分枝杆菌中的研究很好地体现了这两个用途。结核病是全世界重新关注的焦点。日益严重的耐药性使得开发新的抗结核病药物合成为全球各大制药公司投资的重点之一。异烟肼 (INH) 在结核病人临床上使用最广泛,细菌对它的耐药性也最频繁。因此,基因芯片在微生物功能基因组中的运用首先是研究 INA 诱导的结核分枝杆菌基因表达差异,期望发现新的抗结核病药物作用靶点。研究发现 INA 诱导几种基因的表达发生改变^[8];进而选择性抑制枝菌酸生物合成。在被诱导的基因中有些被证明是 INH 作用靶点,这些基因所在途径中被诱导改变的其他基因也可能作为药物开发的新靶点。诱导或抑制的特定表达谱类型也可能作为筛选具有类似效应新化合物的指标。

噬菌体表面展示筛选功能未知的基因及其产物作为靶点,噬菌体表面展示分离得到的肽可以作为用小分子化合物库筛选功能未知靶点的竞争性检测中的替代配体,利用放射性标记,荧光标记可以将上述筛选实验修改为许多种高通量的筛选模式。配体取代筛选在细胞因子受体,核激素受体, G-蛋白偶联受体中应用比较普遍。相应的检测方法也比较多。与靶蛋白关键功能位点结合的肽可以作为高通量筛选中的替代配体。肽的结合位点有:酶的活性位点,活性位点以外的结合部位如别构调节位点,蛋白质-蛋白质相互作用的关键位点。

确证功能未知的肽是否适合作为药高通量筛选配体的方法有:(1)在体内表达细胞内表达可以确证肽是否与功能未知靶点的关键位点结合,如果靶点是细菌生长必要的,肽与功能位点结合将抑制靶点的功能和细菌生长。有多种途径可以使肽在细胞内进行调控表达。例如将肽与葡萄球菌核酶融合在酵母中表达,寻找抑制细胞内激素信号传导,转录沉默,纺锤体检查点的化合物。筛选与功能已知的必需靶点结合的肽。然后,将其与 GST 融合,由严紧调控型启动子控制。诱导表达可以调控细菌的生长。(2)检测功能未知靶点的其他方法包括化合物与靶蛋白直接结合。化合物与蛋白质结合后,蛋白质的热溶解温度发生改变,这可以通过微热量计或与其他荧光探针结合进行测量。

利用基因组筛选抗生素的例子^[15]是：从 *P. falciparum* 基因组中筛选到两个编码合成异戊烯类如胆固醇所需途径 1-脱氧-D-木糖-5 磷酸 (DOXP) 中关键酶的基因。DOXP 在细菌、藻类和高等植物中产生异戊烯二磷酸。疟原虫的该酶位于一个来自藻类的特殊分化的细胞器内，当疟原虫在红细胞中生长时，该酶才表达。先前发现抑制该途径中 DOP 还原异构酶的化合物可以抑制一些细菌的生长，抑制剂 FR900098 和可以抑制体外生长的疟原虫，也可以治疗由其他疟原虫感染的小鼠，而且，这些化合物的毒性低，稳定性好，制造成本低，有可能成为开发新抗疟疾药物的基础。

3.2 寻找新的诊断试剂和诊断标记 在感染不同阶段特异表达，且能够被宿主识别的基因产物可作为诊断标记。PCR 比较常用，缺点是不能检测结核分枝杆菌等潜伏期长的致病菌，也不能区分不同感染阶段。利用不同感染阶段特征性抗原进行免疫学检测可能发现潜伏的致病菌，检测疾病以及发病阶段，且不需要表达有功能的多肽，没有活性，甚至部分片段就足以检测病人血清中的抗体。

通过鸟枪法构建随机基因组库，表达基因组的每个 ORF 可以筛选到适合上述检测目的的抗原基因。该方法的缺点是：(1) 不适合有毒性的基因或由于各种原因产物不能积累的基因；(2) 鸟枪法检测的往往是强表达抗原。反应性弱，表达弱或难以克隆的基因没法检测。在异源宿主中逐个基因表达可以解决上述问题。

3.3 开发新疫苗 细菌基因组中编码细胞表面蛋白的基因存在产生抗原变异的机制。包括：*H. influenzae*, *Helicobacter pylori*, *M. tuberculosis* 基因编码区和基因间区域 5-端在 DNA 重复序列内进行链误配的链剪切机制；*Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Treponema pallidum* 编码外膜蛋白的同源基因间重组；*Plasmodium falciparum*, *Borrelia burgdorferi* 表面表达蛋白随菌落而变化。临床菌株的相应表面蛋白也发生表型变异。这些现象提示致菌可能在每次细胞分裂时使抗原性蛋白发生改变，从而逃避免疫系统的作用。这样，常规的疫苗开发方法可能需要改进。

利用基因组数据开发疫苗的例子^[10]有：(*Neisseria meningitidis*) 测序 (3-12 月)——从基因组中选择表面蛋白或分泌蛋白 (几小时)——在宿主中异源表达蛋白质，逐个进行小鼠等动物免疫 (几个月)——免疫血清筛选杀菌活力或与表面蛋白结合的能力——根据滴度高低选择疫苗候选分子进行下一步研究。

4 展望

微生物学在新世纪之初已经取得了好的进展。利用功能基因组的手段，发掘微生物基因组中的信息，将改善我们对生物学基础过程的认识，更好的认识微生物的多样性，保护和修复环境，充分利用微生物基因组信息开发新的抗生素、诊断试剂和疫苗，提高人类的生活质量。

致谢 本文承蒙陆德如先生的热忱指导，特此谢忱！

参考文献

- [1] Nelson K E, Paulsen I T, Heidelberg J F, Fraser C M, Nature biotech 2000, 18: 1049 ~ 1054.
- [2] Pellegrine M, Marcotte E M, Thompson M J, et al. Proc Natl Acad. Sci. USA. 1999, 96: 4285 ~ 4288.
- [3] Eisen J A, Hanawalt P C. Mutat Res. 1999, 435: 171 ~ 213.

- [4] Enright A J, Lliopoulos I, Kyripides N C. *et al* Nature. 1999, **402**: 86 ~ 90.
- [5] Snel B, Bork P, Huynen M A. Nat Genet. 1999, **21**: 108 ~ 110.
- [6] Fiz-Gibbon S T, House C H. Nucleic acids Res. 1999, **27**: 4218 ~ 4222.
- [7] Tekaa F, Lazcano A, Dujon B. Genome Res. 1999, **9**: 550 ~ 557.
- [8] Wilson M, DeRisi J D, Kristensen H H, *et al*. Proc Natl Acad. Sci. USA. 1999, **96**: 12833 ~ 12838.
- [9] Nierman W C, Eisen J A, Fleischmann R D, *et al*. Curr. opin. struct. biol. 2000, **10**: 343 ~ 348.
- [10] Makarova K S, Aravind L, Galperin M Y, *et al*. Genome Res. 1999, **9**: 608 ~ 628.
- [11] Raczniak G, Ibbas M, Soll D. Toxicology 2001, **160**: 181 ~ 189.
- [12] Christensen D J, Gottlin E B, Benson R E. *et al*. Drug discover. today. 2001, **6**: 721 ~ 727.
- [13] Fraser C M, Eisen J A, Salzberg S L. Nature. 2000, **406**: 799 ~ 803.
- [14] Graham D E, Overbeek R, Olsen G. Woese CR. Proc Natl Acad. Sci. USA 2000, **97**: 3304 ~ 3308.
- [15] Jomaa H. Science, 1999, **285**: 1573 ~ 1576.