

Fisher 线性判别函数在基于 COGs 分类的基因组间距离研究中的应用*

刘 蓉¹⁾ 王月兰²⁾ 朱小蓬²⁾ 凌伦奖²⁾** 韩汝珊¹⁾**

(¹⁾北京大学物理系, 北京 100871; ²⁾中国科学院生物物理研究所, 北京 100101)

摘要 利用全基因组信息构建系统发育树. 基于 COGs 类, 对每一个基因组的每一个基因, 都用一个 17 维的向量来描述其编码蛋白隶属于 17 个 COGs 类的程度; 而与一个基因组的所有基因相对应的那些向量就组成一个集合. 接着, 利用 Fisher 线性判别函数, 寻找一组最优化的权重因子; 在此基础上利用 Fisher 线性变换将上述各集合中每一个向量进行线性变换, 使得经 Fisher 线性变换后 17 个 COGs 类对基因组进化的重要程度得到更准确的反映. 最后, 用进行变换后的向量组成的集合间的距离代替基因组之间的距离. 使用这种方法, 分别用 38 个和 43 个基因组做的进化树都支持了 Woese 的三界理论. 该方法克服了其他基于全基因组信息构建系统发育树方法难以对大小相差很大的基因组进行比较的问题, 并能减少基因横向迁移对基因组间距离的扭曲.

关键词 基因组分, 遗传距离, COGs, Fisher 线性判别函数

学科分类号 Q7

在分子进化研究中, 构建系统发育树时常用一种或一组同源分子 (核苷酸、RNA、氨基酸序列等) 作为各基因组的特征分子, 再用这些特征分子之间的某种差异来代表基因组之间的距离^[1]. 然而一种分子或者一组分子都只是整个基因组的很小一部分, 正象一棵或几棵树木很难代表一片森林一样, 这些特征分子也很难代表基因组的整个进化历史. 选取不同的特征分子, 往往得到不同的进化结果, 有时甚至是相互矛盾的结果. 因此越来越多的研究者认为, 为了反映基因组之间的整个进化历史, 应该选取尽可能多的特征分子^[1], 部分研究者甚至建议用整个全基因组的信息来构建系统发育树^[2,3].

我们发展了一种新的基于全基因组信息构建系统发育树的方法. 基于美国国家生物技术信息中心 (NCBI) 的 COGs (clusters of orthologous groups of proteins) 的 17 个功能分类, 对每一个基因组的每一个基因, 都用一个 17 维的向量来描述其相应蛋白质序列隶属于 17 个 COGs 类的程度; 而与一个基因组的所有基因相对应的那些向量就组成一个集合 (17 维向量的定义方法详见 Ling 等发表于《Molecular Phylogenetics and Evolution》).

由于基因组间的大小差别很大, 所含的各种信息的多少和特征对进化的重要程度不一样, 如何将这些信息间的差异准确地转化为基因组间的距离, 对构建接近真实情况的进化树, 无疑是十分重要的. 特别是, 17 个 COGs 类代表不同的生物学功

能, 它们对生物进化的重要程度显然也是不平权的. 因此, 若直接用上述集合间各维平权的欧氏距离来代表基因组间的距离, 必然有一定的偏差. 基于这种考虑, 我们对上述属于各基因组的向量集合进行了 Fisher 线性变换. 而 Fisher 线性判别函数^[4]是一种最优化的判别准则, 它使得变换后集合间样本分得更开, 集合内样本更密集 (集合中任一向量就是这个集合的一个样本). 经过这样的变换, 相当于寻找一种最优化的权重因子, 使得变换后 17 个 COGs 类对基因组进化的重要程度得到更准确的反映. 同时, 由于那些横向迁移来的基因与内源基因明显有很大的不同, 它们通常就是使得集合的方差很大的样本. 经过 Fisher 线性变换后, 集合内样本变得密集, 即集合的方差变小, 意味着有效地降低了基因横向迁移对整个基因组的影响.

1 材料与方法

本研究中用到的 COGs 数据库来自 NCBI (<http://www.ncbi.nlm.nih.gov/COG>) (2001 年 1 月以前的版本)^[5], 各基因组的基因信息来自 GenBank 数据库中相关的*.gbk 文件. 选取的 38

* 国家自然科学基金资助项目 (39890070, 19890380, 39993420), 中国科学院创新工程项目 (KSCX2-2-07, KJCX1-08) 和北京市科委特别资助项目.

** 通讯联系人.

¹⁾** Tel: 010-62752333, E-mail: rshan@pku.edu.cn

²⁾** Tel: 010-64888544, E-mail: ling@sun5.ibp.ac.cn

收稿日期: 2002-03-06, 接受日期: 2002-04-11

个和 43 个基因组的信息见表 1. 按以下步骤构建进化树:

(1) 依 Ling 等的方法构造每一个基因对应的 17 维矢量:

$$v = (v[1], v[2], \dots, v[17])$$

(2) 计算任两个基因组之间的 Fisher 参数 ω_{ij}^* , ω_{ij}^* 是对方程:

$$S_{\omega_{ij}^*}^{-1} S_{b_{ij}} \omega_{ij}^* = \lambda \omega_{ij}^*$$

的解, 即一般矩阵 $S_{\omega_{ij}^*}^{-1} S_{b_{ij}}$ 的特征根 ω_{ij}^* . 上式中各变量的含义见文献 [4]. ω_{ij}^* 就是使 Fisher 线性判别函数取极大值时最好的投影超平面参数^[4], 即将 i, j 两个集合中的样本分别投影到这个超平面上能使 i, j 两个集合的样本分得越开, 而 i, j 集合内的样本越密集.

(3) 对每两个基因组样本的中心作变换 (因为是线性变换, 所以相当于对每一个样本求变换后再求中心):

$$y_i = \omega_{ij}^{*T} m_i \quad y_j = \omega_{ij}^{*T} m_j$$

其中, m_i 是第 i 个集合的中心, y_i 是该基因组相应的映射空间的中心.

(4) 求任两个基因组之间的距离:

$$d_{ij} = |y_i - y_j|$$

(5) 将 (4) 得到的距离作为两个基因组之间的距离, 用 PHYLIP 软件包的 Neighbor joining^[6] 缺省

参数画进化树, 它实际上是一种基于距离矩阵对真实生命之树的一种最优化的统计推断. 再用 treeview 软件(<http://taxonomy.zoology.gla.ac.uk/rod/rod.html>) 显示.

2 结 果

2.1 38 个基因组得到的结果

38 个基因组的全称和缩写见表 1 中的前 38 个, 构建出的进化树见图 1.

从图 1 可以看出这 38 个基因组明显地分成三大支, 除 Mtub, Halo, Aper 三个例外, 6 个古细菌、23 个真细菌和 6 个真核都分别聚类在三个不同的分支中, 在图 1 中分别标记为 1, 2, 3 三支.

a. 在古细菌这一分支中: 分别聚类成 4, 5 两支, 几个嗜热的生物: 詹氏甲烷球菌 (Mjan), 热自养甲烷杆菌 (Mthe), 嗜酸热原体 (Taci), *Pyrococcus abyssi* (Paby), *Pyrococcus horkoshi* (Phor) 聚类在同一支 5 中, 另一个嗜热的硫代谢古细菌闪烁古生球菌 (Aful) 和一个叫结核分枝杆菌 (Mtab) 的真细菌聚类在另一分支 4 中. 文献 [7] 报道 Mtub 与其他真细菌有明显的不同, 我们的结果支持这种报道; 而文献 [8] 曾报道硫代谢古细菌与其他古细菌分成两支, 我们的结果印证了这一点, 但它不是真核的姊妹分支. 在分支 5 的几个嗜热生物中, Mjan 是从潜水艇热水排放口

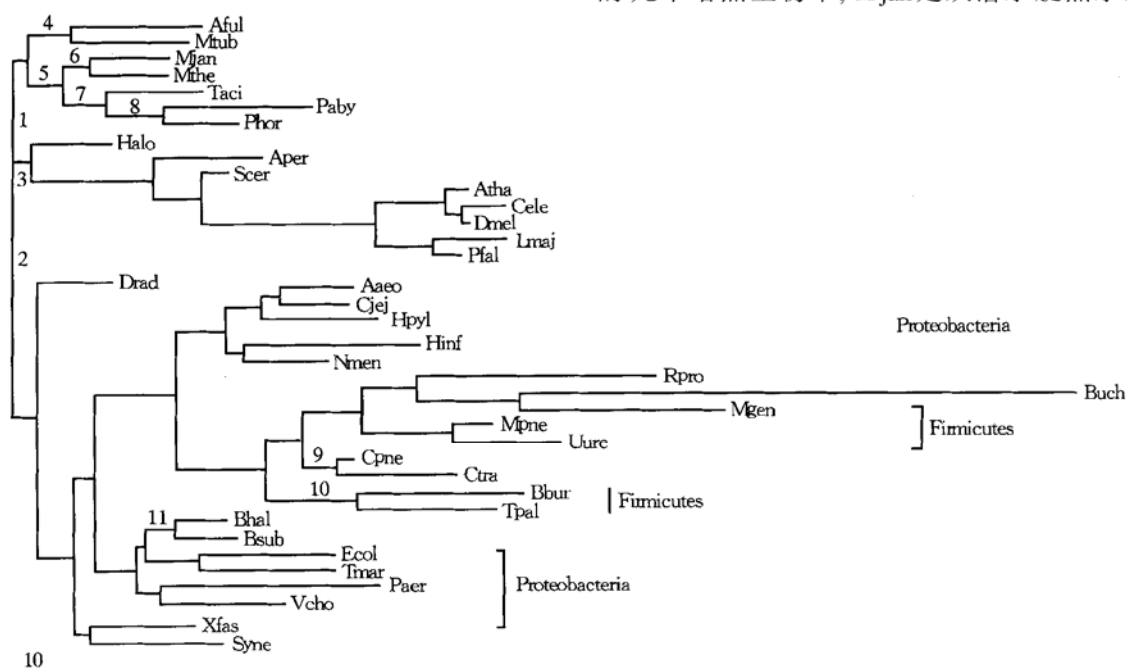


Fig. 1 Phylogenetic tree of 38 genomes

Table 1 Full names and abbreviations of 43 genomes

No	Abbreviaton	Full name	No	Abbreviation	Full name
1	Aful	<i>Archaeoglobus fulgidus</i>	23	Tpal	<i>Treponema pallidum</i>
2	Mtub	<i>Mycobacterium tuberculosis</i>	24	Rpro	<i>Rickettsia prowazekii</i>
3	Mjan	<i>Methanococcus jannaschii</i>	25	Buch	<i>Buchnera</i> sp
4	Mthe	<i>Methanobacterium thermoautotrophicum</i>	26	Mgen	<i>Mycoplasma genitalium</i>
5	Taci	<i>Thermoplasma acidophilum</i>	27	Mpne	<i>Mycoplasma pneumonia</i>
6	Paby	<i>Pyrococcus abyssi</i>	28	Uure	<i>Ureaplasma urealyticum</i>
7	Phor	<i>Pyrococcus horikoshii</i>	29	Cpne	<i>Chlamydia pneumoniae</i>
8	Halo	<i>Halobacterium</i> sp	30	Ctra	<i>Chlamydia trachomatis</i>
9	Aper	<i>Aeropyrum pernix</i>	31	Bhal	<i>Bacillus halodurans</i>
10	Scer	<i>Saccharomyces cerevisiae</i>	32	Bsub	<i>Bacillus subtilis</i>
11	Atha	<i>Arabidopsis thaliana</i>	33	Ecol	<i>Escherichia coli</i>
12	Cele	<i>Caenorhabditis elegans</i>	34	Tmar	<i>Thermotoga maritima</i>
13	Dmel	<i>Drosophila melanogaster</i>	35	Paer	<i>Pseudomonas aeruginosa</i>
14	Lmaj	<i>Leishmania major</i>	36	Vcho	<i>Vibrio cholerae</i>
15	Pfal	<i>Plasmodium falciparum</i>	37	Syne	<i>Synechocystis</i> sp
16	Drad	<i>Deinococcus radiodurans</i>	38	Xfas	<i>Xylella fastidiosa</i>
17	Hpyl	<i>Helicobacter pylori</i>	39	Hp99	<i>Helicobacter pylori</i> J99
18	Aaeo	<i>Aquifex aeolicus</i>	40	NmenA	<i>Neisseria meningitidis</i> Z2491
19	Cjej	<i>Campylobacter jejuni</i>	41	CpneA	<i>Chlamydophila pneumoniae</i>
20	Hinf	<i>Haemophilus influenzae</i>	42	CpneJ	<i>Chlamydophila pneumoniae</i> J138
21	Nmen	<i>Neisseria meningitides</i>	43	CtraM	<i>Chlamydia muridarum</i>
22	Bbur	<i>Borrelia burgdorferi</i>			

获得的一种极端嗜热的产烷生物^[9], Mthe 是一种嗜热的甲烷杆菌^[10], 它们都是与烷相关的生物, 因此又聚类在一个共同的分支 6 中; Phor、Paby 和 Taci 聚类在另一个分支 7 中, 其中 Taci 是一种从煤堆中分离出的嗜热、嗜酸的支原菌^[11], 而 Phor 和 Paby 是从海洋中热水喷发口分离出的两种极端嗜热的古细菌^[12, 13], 因此 Phor 和 Paby 又聚类在一个共同的分支 8 中, Taci 离它们的距离相对较远。

b. 在真细菌这一分支中: 23 个真细菌全部聚类在这一分支中。具体细节比较合理, 比如: 两个属于螺旋体科的细菌: 布氏疏螺旋体 (Bbur) 和苍白密螺旋体 (Tpal) 聚类在一个共同的分支 10 中; 属于衣原体的两个细菌: 肺炎衣原体 (Cpne) 和沙眼衣原体 (Ctra) 聚类在同一分支 9 中。紫色细菌门在图 1 中分成两大块: 一块主要由 γ 亚门 (Ecol, Paer, Vcho, Xfas) 组成, 另一块主要由 ϵ 、 α 等亚门组成 (Cjej, Hpyl 属 ϵ 亚门, Nmen 属

β 亚门, Rpro 属 α 亚门, Hinf, Buch 属 γ 亚门)。硬壁菌门在图 1 中分成两支: 一支为 Mollicutes 的三个细菌: 生殖道枝原体 (Mgen)、肺炎枝原体 (Mpne) 和解脲枝原体 (Uure); 另一支 (标记为 11) 是两个芽孢杆菌属的耐盐芽孢杆菌 (Bhal) 和枯草芽孢杆菌 (Bsub)。这种同一门分成几支的现象与文献 [14] 类似, 但又不完全相同。

c. 在真核这一分支中: 6 个真核均聚类在这一分支中, 并且将动物和植物分开了, 如拟南芥 (Athe) 与线虫 (Cele) 的距离比线虫与果蝇 (Dmel) 的距离更远。Plasmodium falciparum (Pfal) 和 Leishmania major (Lmaj) 之间的距离相对较近, 它们离线虫、拟南芥和果蝇的距离都较远。酵母 (Scer) 离 Atha, Cele, Dmel, Pfal, Lmaj 的距离都较远。这里, Scer, Lmaj, Pfal 三者之间离基部的相对位置可能不准确, 这是因为目前还没有 Lmaj 和 Pfal 的完整的全基因组数据, 但使用我们的方法仍然将它们聚类在了真核这一分支中, 说

明我们的方法具有一定的鲁棒性。

有两个例外 (Halo 和 Aper), 它们分别靠近基部. 文献 [5] 报道 GenBank 中关于 Aper 有很多功能预测有误, 因此 Aper 的错分有可能是 GenBank 注释数据本身不准确造成的. 文献 [15] 报道嗜盐古细菌 Halo 在很多方面与复杂的真核生物有很相似的特征, 说明我们的结果将 Halo 划分在真核这一分支中靠近基部的地方似乎有其合理之处. 同时, 我们的方法更强调全基因组信息和 COGs 功能分类.

总之, 我们的结果支持了 Woese 将生物分成古细菌、真细菌和真核生物三界的理论.

2.2 43 个基因组得到的结果

43 个基因组的全称和缩写见表 1, 构建出的进化树见图 2.

为了测试方法的稳定性和有效性, 我们在上面 38 个基因组的基础上, 又增加了一些同一种中不同菌株的数据. 由于它们属于同一个种, 它们之间的距离应该相当接近, 在进化树上也应该位于同一分支. 它们之间的距离应该小于不同种之间的距离. 我们的结果正好验证了这一点. 从图 2 中可以看到, 以下 4 组数据分别聚类在 4 个不同的分支上, 同一组数据中的距离都比较近, 均远远小于与其他基因组之间的距离. a. CpneA, Cpne, CpneJ; b. CtraM, Ctra; c. Hp99, Hpyl; d. NmenA, Nmen.

从图 1 和图 2 还可以看到, 38 个基因组和 43 个基因组所做的进化树, 在拓扑结构上具有很好的稳定性. 43 个基因组得到的结果也支持了 Woese 的三界理论.

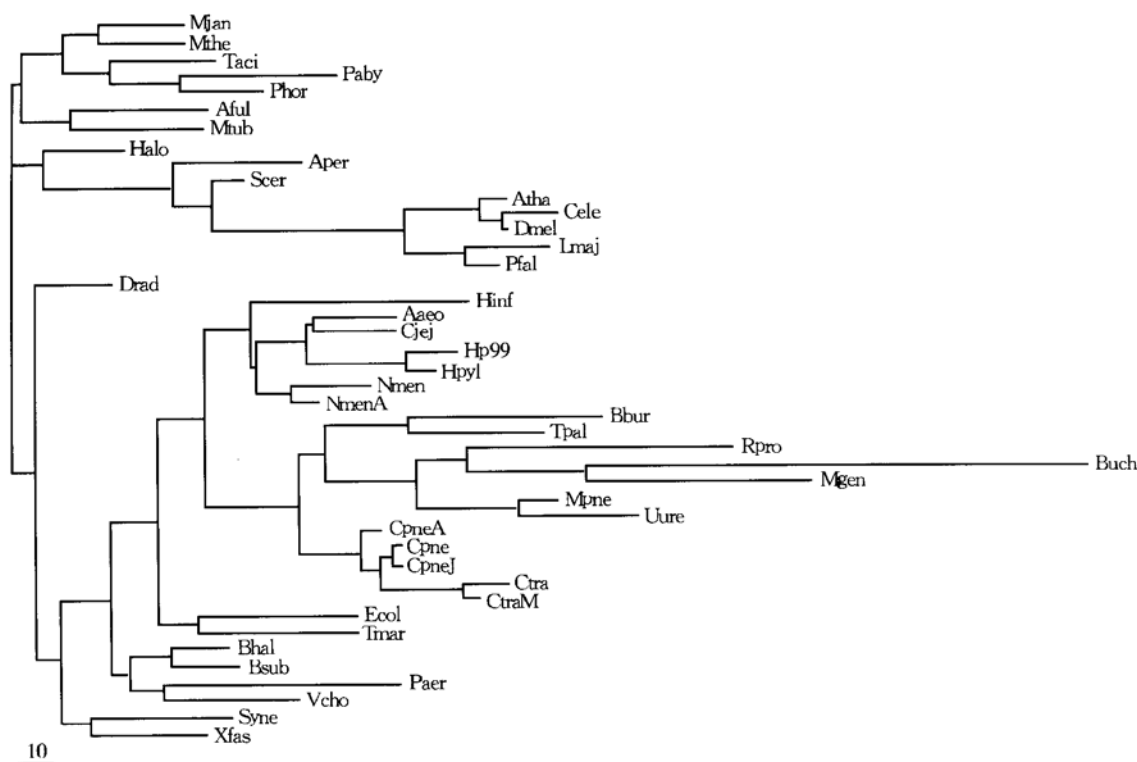


Fig. 2 Phylogenetic tree of 43 genomes

3 讨 论

在数学上, 求集合间距离时, 常用集合间样本的中心距离或集合间样本的最短距离、最长距离等来表示, 而在求样本间的距离时又常用欧氏距离. 对于集合中样本为多维矢量的情况, 由于欧氏距离相当于多维到一维的一个各维平权的投影. 因此直接用集合间样本的最短、最长和中心的欧氏距离,

各维的影响均是相同的. 我们做过试验, 如果先求两个基因组的所有基因之间的欧氏距离, 然后直接用它们之间的最短距离、最长距离来代表基因组之间的距离, 效果较差. 其原因是: 最短距离或最长距离都相当于用其中的一个基因代表整个基因组, 显然不合理. 中心距离比最短距离和最长距离更为合理一些, 因为它考虑了一个基因组的所有基因对这个基因组的贡献. 所以结果稍微合理一些, 但仍

不理想. 将一个基因组的全部基因所对应的向量相加, 得到一个基因组一个总的向量, 然后对该总向量进行归一化处理, 再用两个基因组之间归一化后的总向量之间的欧氏距离作为基因组之间的距离. 结果比前三种方法要好, 但它们均未考虑 COGs 类之间不平权的影响. 因此都比用 Fisher 线性变换后求距离的效果差.

在我们的方法中, 由于使用的是中心之间的某种距离, 求中心时相当于对整个基因组的某种信息做了一种平均. 因此基因组大小相差很悬殊并不成为问题. 如在我们所讨论的 43 个基因组中, 真细菌 Mgen 的基因组只有 0.58 M; 果蝇的基因组有 120 Mbp. 基因组间的大小相差很悬殊, 我们的结果仍然与用实验做的结果很一致, 说明我们的方法能够应用到基因组大小相差较大的基因组中去.

最近的研究表明, 估计在一个基因组中有 1.5% ~ 14.5% 的基因与横向迁移有关^[6], 如此大量的横向迁移, 在我们的方法中, 它们对构建进化树的影响也不明显.

综上所述, 我们小组发展的这种新的基于全基因组信息构建系统发育树的方法, 比较有效地降低了基因横向迁移对全基因组的影响, 同时也能应用到大小相差较大的基因组中去. 结果有力地支持了 Woese 的三界理论.

参 考 文 献

- 1 Masatoshi N, Sudhir K. Molecular Evolution and Phylogenetics. New York: Oxford University Press. 2000. 81
- 2 Ekaia F, Lazcano A, Dujon B. The genomic tree as revealed from whole proteome comparisons. Genome Res, 1999, 9 (6): 550~557
- 3 Lin J, Gerstein M. Whole-genome trees based on the occurrence of folds and orthologs: implications for comparing genomes on different levels. Genome Res, 2000, 10 (6): 808~817

- 4 边肇祺, 张学工. 模式识别. 北京: 清华大学出版社. 1999. 87~90
- 5 Bian Z Q, Zhang X G. Pattern Recognition. Beijing: Qing Hua University Press. 1999. 87~90
- 6 Tatusov R L, Natale D A, Garkavtsev I V, et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Research, 2001, 29 (1): 22~28
- 7 Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 1987, 4 (4): 406~425
- 8 Cole S T, Brosch R, Parkhill J, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature, 1998, 393 (6685): 537~544
- 9 Lake J A. Prokaryotes and archaeobacteria are not monophyletic: rate invariant analysis of rRNA genes indicates that eukaryotes and eocytes form a monophyletic taxon. New York: Cold Spring Harbor Sym on Quantitative Biology, 1987, 52: 839~846
- 10 Ones W J, Leigh J A, Mayer F, et al. Methanococcus jannaschii sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. Arch Microbiol, 1983, 136: 254~261
- 11 Touzel J P, Conway De Macario E, Nolling J, et al. DNA relatedness among some thermophilic members of the genus methanobacterium: emendation of the species Methanobacterium thermoautotrophicum and rejection of Methanobacterium thermoformicum as a synonym of Methanobacterium thermoautotrophicum. Int. J Syst Bacteriol, 1992, 42: 408~411
- 12 Darland G, Brock T D, Samsonoff W, et al. A thermophilic, acidophilic mycoplasma isolated from a coal refuse pile. Science, 1970, 170: 1416~1418
- 13 Gonzalez J M, Masuchi Y, Robb F T, et al. Pyrococcus horikoshii sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. Extremophiles, 1998, 2: 123~130
- 14 Erauso G, Reysenbach A L, Godfroy A, et al. Pyrococcus abyssi sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. Arch Microbiol, 1993, 160: 338~349
- 15 Baldauf S L, Palmer J D, Doolittle W T. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. Proc Natl Acad Sci USA, 1996, 93 (15): 7749~7754
- 16 Ng W V, Kennedy S P, Mahairas G G, et al. Genome sequence of Halobacterium species NRC-1. Proc Natl Acad Sci USA, 2000, 97 (22): 12176~12181
- 17 Garcia Vallvé S, Romeu A, Palau J. Horizontal gene transfer in bacterial and archaeal complete genomes. Genome Res, 2000, 10 (11): 1719~1725

The Application of Fisher Linear Discriminant to Distance Between Genomes Which Based on COGs*

LIU Rong¹⁾, WANG Yue-Lan²⁾, ZHU Xiao-Peng²⁾, LING Lun-Jiang^{2)**}, HAN Ru-Shan^{1)**}

⁽¹⁾ Department of Physics, Peking University, Beijing 100871, China;

⁽²⁾ Institute of Biophysics, The Chinese Academy of Sciences, Beijing 100101, China)

Abstract A new method to construct a phylogeny tree based on whole genome information is introduced. Each gene of an organism is represented by a 17 dimensional vector, each dimension of which relates to one of the 17 COGs (clusters of orthologous groups of proteins) classes. All the vectors of a genome constitute a set. Then Fisher linear discriminant was used to find a set of optimal weights which reflect more accurately the different contribution of the 17 COGs classes to the genome's evolution. That is, under the Fisher criteria, each vector of

a genome is linear mapped. After that, the distance between two genomes was represented by the distance between the related two sets constituted by mapped vectors. At last, the distance matrix was used to construct a phylogenetic tree by PHILP software package. Phylogeny trees of 38 and 43 genomes constructed by this method respectively well support the “three primary kingdom” theory of Woese. This method rectifies the shortcoming of other methods which are difficult to compare genomes differing remarkably in genome size. In addition, the method diminishes the distortion on the distances between genomes brought by lateral gene transfer.

Key words gene composition, phylogenetic distance, COGs, Fisher linear discriminant

* This work was supported by grants from The National Natural Sciences Foundation of China (39890070, 19890380, 39993420), Knowledge Innovation Project of The Chinese Academy of Sciences (KSCX2-2-07 and KSCX1-08) and a Special Grant Science and Technology Committee of Beijing.

** Corresponding author. ¹⁾** Tel: 86-10-62752333, E-mail: rshan@pku.edu.cn ²⁾** Tel: 86-10-64888544, E-mail: ling@sun5.ibp.ac.cn

Received: March 6, 2002 Accepted: April 11, 2002

神经介素 U (neuromedin U) 研究概况

郑旭煦 胡应和 欧阳克清 李新平 蔡绍哲

(重庆大学生物工程学院, 教育部生物力学和组织工程重点实验室, 重庆 400044)

神经介素 (neuromedin) 是一组平滑肌刺激性多肽, 通常分成四类: 蛙皮素类、蛙褶肽类、神经紧张素类、神经介素 U (neuromedin U, NMU)。在这组多肽中, 神经介素 U 最后被发现, 并因其子宫收缩活性而得名。神经介素 U 是 Minamino 等在 1985 年首次从猪脊髓中发现并分离提纯到的多肽 (包括 NMU-8 和 NMU-25), 随后又相继从大鼠、几内亚小猪、狗、家兔、鸡和澳大利亚树蛙等物种中分离得到。NMU 产生于一个由 174 个氨基酸组成的前体。该前体包含一个疏水信号肽和几个二元蛋白酶加工位点, 这些二元加工位点靠近 C 端, 释放 NMU 和其他分泌型多肽。NMU 氨基酸序列在自始至终的进化过程中呈现显著的保守性, 其核心多肽 (F-L-F-R-P-R-N-NH₂) 在所有哺乳动物物种中完全保存, 且与平滑肌收缩活性密切相关。

NMU 广泛分布于哺乳动物胃肠道的神经元中。NMU 的第一个生物学活性是刺激平滑肌收缩, 它能刺激人回肠纵向肌和膀胱、鸡嗦嚙、大鼠胃环状肌等离体肌肉组织的收缩, 但不能刺激猪空肠和豚鼠小肠收缩, 提示 NMU 对肠道的调节作用存在物种差异; NMU 具有增加动脉血压、调节肠系膜血流和改变肠道离子传输的活性。NMU 也存在于垂体的甲状腺和内分泌细胞中。在大鼠皮下注射 NMU 将导致促肾上腺皮质激素水平短暂增加和血清皮质酮水平持续增加, 大鼠垂体前叶和人脑垂体促肾上腺皮质激素细胞有高水平 NMU 免疫反应性, 提示 NMU 具有调节垂体肾上腺轴的作用。NMU 还存在于不同的脑区域, 几个涉及感觉的结构 (包括脊髓、三叉神经感觉核、前庭神经核等) 含有丰富的 NMU 纤维; 在下丘脑弓状核内鉴定

出含有 NMU 免疫反应性的神经元细胞体, 提示 NMU 具有调节食物摄取和神经内分泌的作用; 在神经核的横卧和黑质中发现高水平 NMU 免疫反应性, 暗示 NMU 具有调节多巴胺能作用的功能; 在海马回、杏仁体和大脑边缘体系中也发现 NMU 免疫反应性, 提示 NMU 和 NMU 受体与神经精神病症相关。

多个研究组最近分别证明了视紫红质 GPCR 家族的两个“孤儿 GPCR 受体”——FM-3/GPR66 和 FM-4/TGR-1 是 NMU 的特异受体 (分别指定为 NMU1R 和 NMU2R)。人 NMU1R 和 NMU2R 的基因图谱分别为 SHGC-3353 和 SHGC-8848, 定位于染色体 2q34~q37 和 5q31.1~q31.3; 其开放阅读框分别编码由 403 和 412 个氨基酸组成的蛋白质, 两者 51% 同源, 人鼠同源分别为 73% 和 75%。NMU1R 在人的组织中广泛表达, 最高表达出现在外周器官, 特别是胃肠道和泌尿生殖系统, 但在中枢神经系统中表达很低。NMU2R 在中枢神经系统, 特别是下丘脑和黑质中高效表达, 表明 NMU2R 具有介导中枢神经系统中 NMU 的特殊作用。在转染了受体的 HEK293 细胞中, NMU 能动员其细胞内的钙, 信号响应呈量效相关, 但此钙信号对百日咳毒素不敏感, 表明 NMU1R、NMU2R 与 G 蛋白 Gq 家族偶联。

总之, NMU 是一种广泛分布于胃肠、泌尿生殖和中枢神经系统的一种神经多肽, 虽然许多生物学活性已知, 但其生理学功能了解甚少。继续研究 NMU 及其受体的分布、作用机理和表达调控, 将有助于对 NMU 及其受体生理功能的了解, 发挥它们在高通量药物筛选和制药等行业的重要作用。