

分子伴侣 60 类群的系统发育学分析

徐 毅 周培瑾

(中国科学院微生物研究所 北京 100080)

摘 要 有充分的证据支持线粒体和叶绿体起源的内共生假说, 即: 紫色细菌、 α -类群的成员是线粒体的祖先, 而蓝细菌的成员是叶绿体的祖先。这一证据来自分子伴侣 60 蛋白为基础的系统发育学分析, 并且与 16S rRNA 序列为基础的分析结果是完全一致的。这一结果还说明, 嗜碱芽孢杆菌 C-125 菌株在系统发育地位上更接近枯草芽孢杆菌。

关键词 分子伴侣, 系统发育学, 分子进化, 嗜碱芽孢杆菌 C-125 菌株

真核生物可能是在 14 亿年前由原核生物进化而来。现在已有许多证据支持这一学说。其中有两个被人们普遍接受的假说: 其一, 细胞核、线粒体和叶绿体是由细胞质膜内陷形成的; 其二, 就是更被人们认可的內共生学说, 即: 发酵型细菌与光合细菌相结合, 建立一种永久式的共生关系, 这种光合细菌然后进化成叶绿体, 蓝细菌被认为是最有可能的叶绿体祖先; 线粒体则是由好氧呼吸型细菌与其他细菌间的內共生而产生^[1]。现在, 通过对线粒体细胞色素 C 和 16S rRNA 序列的比较已说明, 线粒体可能是由真细菌中紫色细菌的一个成员进化而来^[2-4], 而对 16S rRNA 序列的分析也表明叶绿体可能是由蓝细菌的一个成员组成^[5-7]。由于线粒体和叶绿体进化的主要证据是来自对核糖体 RNA 基因的序列的比较, 所以还需要其他证据来补充和加强內共生学说, 以及进一步推断內共生生物的系统发育起源。在此方面, 细胞器和真细菌中都存在的序列上高度保守的相应类似物可以提供非常有价值的进化信息。分子伴侣 60 就是这样一种在线粒体、叶绿体以及真细菌中都存在的、在序列和进化上非常保守的蛋白质。因而可以被用来作为研究细胞器起源的一种十分有用的探针, 同时也可以作为生物分类的依据。本文报道, 用作者分离、测序和由此推断出的嗜碱芽孢杆菌 C-125 菌株分子伴侣 60 蛋白序列与其他现在已知的真细菌、线粒体和叶绿体的相应蛋白序列比较, 一方面确定了该嗜碱芽孢杆菌 C-125 菌株的分类地位, 另一方面还揭示了线粒体与真细菌中紫色细菌成员之间的系统发育关系, 以及叶绿体与蓝细菌成员之间的进化关系。

1 材料和方法

1.1 嗜碱芽孢杆菌 C-125 菌株分子伴侣 60 基因的分离和测定

嗜碱芽孢杆菌 C-125 菌株^[8]其总 DNA 被作为靶基因的来源; 枯草芽孢杆菌 BD224 菌株总 DNA 被作为 PCR 反应的模板来扩增其本身的分子伴侣 60 (GroEL) 基因, 然后用 Digoxigenin-DNA labelling Kit (Boehringer GmbH, Mannheim, Germany) 标记该 PCR 产物, 用此标记的 PCR 产物作为探针, 通过 Southern blotting 等方法分离出 C-125 菌株的靶基因, 然后克隆到 *E. coli* 中, 最后用 Dideoxy-chain-termination 方法^[9]测序。

1.2 分子伴娘 60 蛋白的氨基酸序列

所有进行系统发育学分析的分子伴娘 60 蛋白序列是从系统发育上不同的种群中选择的,并且只有那些已确定完整序列的蛋白才被用于此项研究。除了 C-125 菌株的分子伴娘 60 蛋白序列外,其他所有蛋白的氨基酸序列都是从 Swiss-Prot Data 以及 Genbank Data 中获得。C-125 菌株分子伴娘 60 蛋白的氨基酸序列直接由该蛋白基因的核苷酸序列翻译过来。

1.3 分子伴娘 60 蛋白氨基酸序列的系统发育学分析

将所有被分析的分子伴娘 60 蛋白氨基酸序列放在一个文件名下,输入到“Clustalw”程序包中^[10],通过该程序包进行多序列匹配排列(Multiple Alignments)分析。基本的多序列匹配排列过程主要由三个阶段构成:1. 利用一种慢速、精确的“full dynamic programming”方法^[10]分别匹配排列每一对序列,然后计算出序列间的“distance matrix”;2. 在“distance matrix”基础上,利用 Nei 和 Saitou 的“Neighbor-Joining”方法^[11]构成一个“guide tree”,用来指导以后的多序列匹配排列;3. 按照“guide tree”中的分枝顺序,利用改进的 Feng 和 Doolittle 的“progressive”方法^[12],对所有序列进行阶梯式多序列匹配排列,最后形成一个多序列匹配排列阵,其中形成的缺口(gap)用中性元素填补。将由“Clustalw”程序包计算出的多序列匹配排列阵以一个文件名的形式,输入 Felsenstein 的“PHYLP”v.3.5c 程序包中,进行最终的系统发育分析。

在“PHYLP”程序包中的诸多程序中,作者以“Distance Matrix”程序为主要分析手段,“Parsimony”程序为辅,进行分子伴娘 60 蛋白的系统发育分析。

Neighbor-Joining 和 Bootstrap 程序:输入“Seqboot”程序中一套序列数据,该程序通过“bootstrap”方法^[13],产生 100 套重新取样的序列数据,重新命名后输入到“Protdist”程序,将“Analyze multiple data sets?”选择设为 100,该程序利用 Dayhoff 的 PAM 001 模式^[14]计算 distance matrix,产生一个具有相应于 100 套序列 distance matrix 数据的“outfile”。再利用 Distance Matrix 程序中 Nei 和 Saitou^[11]的 Neighbor-Joining 方法通过连续会聚,形成 100 个 Neighbor-Joining 系统发育树或“treefiles”。将“treefile”重新命名后输入“consense”程序,该程序运行一组“M-sub-L” consensus tree 方法^[15],从 100 套 Neighbor-Joining 的“treefile”数据中产生一个“majority rule consensus tree”。

Distance matrix 的“Fitch”程序:首先将一套序列数据输入“Protdist”程序中,根据 Dayhoff 的 PAM 001 模式^[14]通过序列数据来计算“distance matrix”,然后将这一“distance matrix”输入“Fitch”程序中,把“power”设在 2.0 处,“Global rearrangements”= Yes,通过“Fitch-Margoliash”计算方法^[16],从“distance matrix”数据来估算系统发育树。

Protpars 程序:输入“Protpars”程序中一套序列数据,该程序利用“parsimony”方法^[17,18],估算系统发育树。

2 结果

2.1 嗜碱芽孢杆菌 C-125 菌株分子伴娘 60 基因的克隆和测序

在 Schmitt 等人^[19]工作的基础上,首先以枯草芽孢杆菌的总 DNA 为模板,通过 PCR 技术扩增出枯草芽孢杆菌分子伴娘 60 基因的 PCR 产物。为了检验嗜碱芽孢杆菌

C-125 菌株中是否存在分子伴侣 60 基因,用标记的枯草芽孢杆菌 PCR 产物作为探针与 C-125 菌株限制性内切酶处理过的总 DNA 杂交,杂交结果显示出一条 3kb 左右的正杂交片段,通过克隆测序(图 1),发现存在一个完整的 Open Reading Frame,由于该基因在序列上与其他细菌分子伴侣 60 基因高度同源,所以被认为是 C-125 菌株的分子伴侣 60 基因。

2.2 不同来源分子伴侣 60 蛋白的系统发育学分析

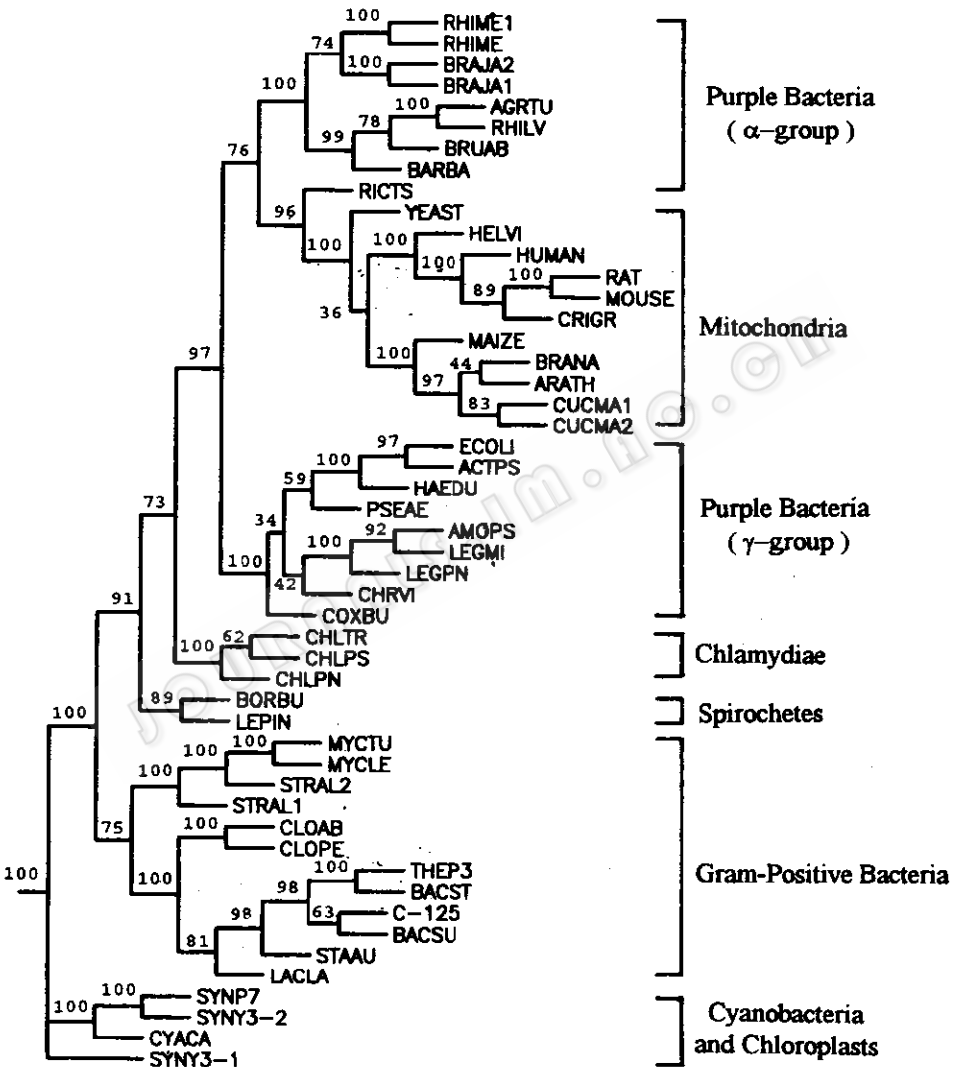


图 2 分子伴侣 60 蛋白家族的“Consensus”系统发育树

该树含有从 100 次 bootstrapping 取样分析的数据,分叉处的数值表明此组蛋白在 100 个由 bootstrapping 所建立的系统发育树中出现的机率。

Fig. 2 Consensus phylogenetic tree of hsp 60 protein family

The tree contains the values from 100 times of bootstrapping, the respective values at the forks show how often the group to the right was found out of 100 trees.

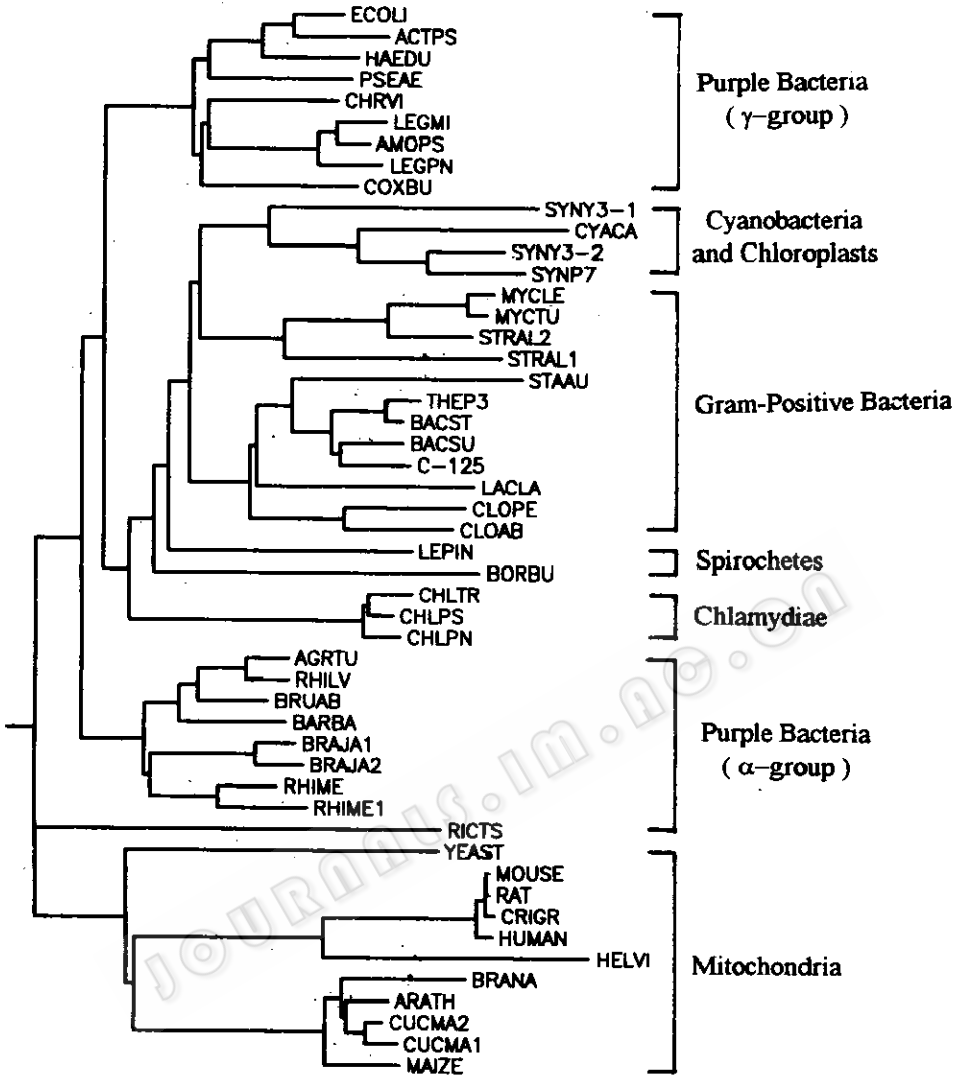


图 3 分子伴娘 60 蛋白家族的“Fitch”系统发育树

图中的标尺单位表示每个位点上氨基酸变异的概率。

Fig. 3 “Fitch” phylogenetic tree of hsp 60 protein family on the basis of “Distance Matrix” method

The bar scale indicates 10% estimated nucleotide substitutions.

选择 50 个不同来源的分子伴娘 60 蛋白序列进行系统发育学分析(表 1)。其中 C-125 菌株的分子伴娘 60 蛋白序列,直接由该基因的核苷酸序列翻译过来,而其他分子伴娘蛋白序列则是从 Swiss-Prot data 获得。在这 50 个分子伴娘 60 蛋白中,首先计算每一对序列的百分相似性(结果略)。在此基础上,用 Clustalw 程序包对这些序列进行逐步地多重序列匹配排列(结果略)。排列阵中的缺口用中性元素填补,这些中性元素代表了氨基酸的缺失或插入。最初排列阵的两端存在着不匹配的独立悬浮片段,为了以后更好地进行序列间一对一的比较和分析,这些悬浮片段应删掉。然后将此修改过的多序列匹配排列阵在 PHYLIP 程序包中进行系统发育学分析。图 2 是以 Neighbor-Joining 方法

为基础的、利用 bootstrap 模式而得到的、更具统计性质的 consensus 系统发育树。在该树分叉上的数值若大于 50, 表明该数值右边的分枝是可靠的, 反之则是可疑的; 图 3 是以 Fitch 方法为基础 Distance matrix 系统发育树; 图 4 是 Parsimony 系统发育进化树, 作为以上进化树的参比。图 2~4 的结果反映了 50 个不同来源(真细菌、线粒体和叶绿体)分子伴娘 60 蛋白之间的系统发育关系。

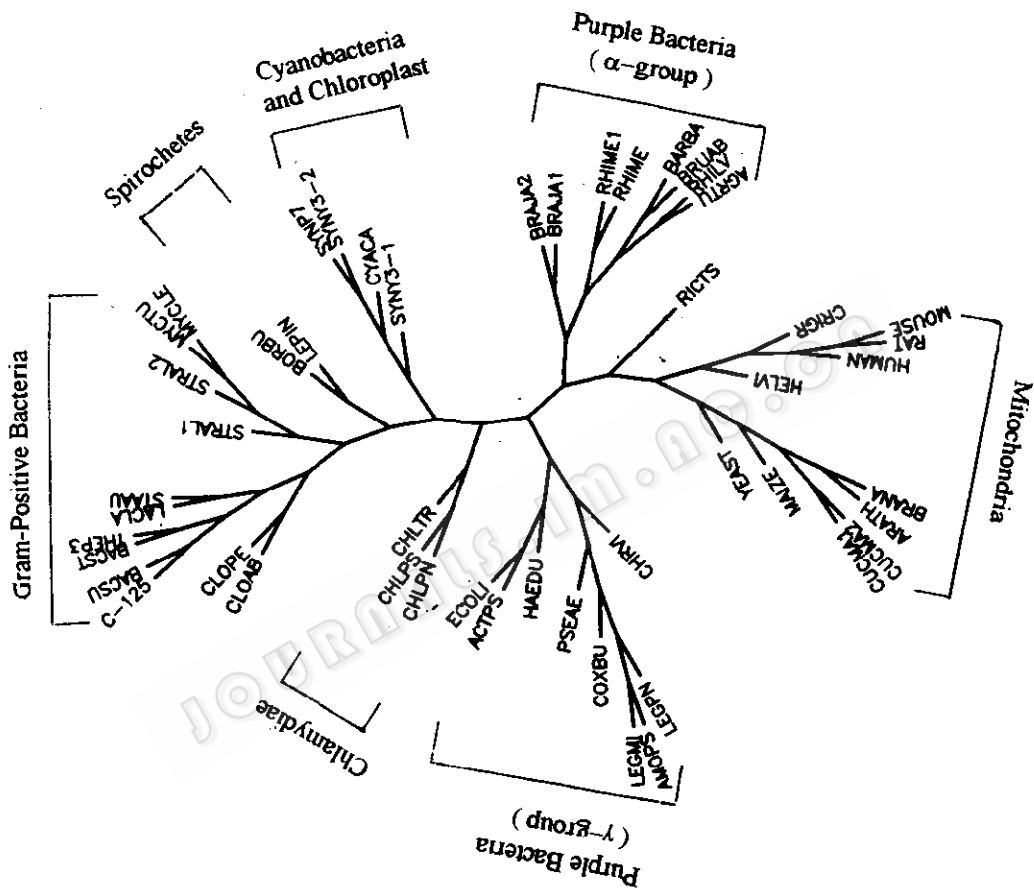


图 4 分子伴娘 60 蛋白家族的“Parsimony”系统发育树
Fig. 4 “Parsimony” phylogenetic tree of hsp 60 protein family

表 1 分子伴娘 60 蛋白的来源
Table 1 Chaperonin 60-protein sources

代 码 Code*	俗 名 Common name	学 名 Scientific name	数据库的进入号 Accession no.**
CHRV1	Purple bacteria	<i>Chromatium vinosum</i>	P31293
PSEAE		<i>pseudomonas aeruginosa</i>	P30718
COXBU		<i>Coxiella burnetii</i>	P19421
AMOPS	symbiotic bacterium	Not Yet classified	P26004
LEGM1		<i>Legionella micdadei</i>	P26194

续表 1

代 码 Code	俗 名 Common name	学 名 Scientific name	数据库的进入号 Accession no. **
LEGPN		<i>Legionella pneumophila</i>	P26878
ACYPS	symbiotic bacterium	Not Yet classified	P25750
ECOLI		<i>Escherichia coli</i>	P06139
HAEDU		<i>Haemophilus ducreyi</i>	P31294
AGRTU		<i>Agrobacterium tumefaciens</i>	P30779
BRUAB		<i>Brucella abortus</i>	P25967
RHILV		<i>Rhizobium leguminosarum</i>	P34939
BARBA		<i>Bartonella bacilliformis</i>	P35635
RHIME		<i>Rhizobium meliloti</i>	P35469
RHIME1		<i>Rhizobium meliloti</i>	P35471
BRAJA1		<i>Bradyrhizobium japonicum</i>	P35861
BRAJA2		<i>Bradyrhizobium japonicum</i>	P35862
RICTS		<i>Rickettsia tsutsugamushi</i>	P16625
CHLPN		<i>Chlamydia pneumoniae</i>	P31681
CHLPS		<i>Chlamydia psittaci</i>	P15599
CHLTR		<i>Chlamydia trachomatis</i>	P17203
CLOAB		<i>Clostridium acetobutylicum</i>	P30717
CLOPE		<i>Clostridium perfringens</i>	P26821
BACSU		<i>Bacillus subtilis</i>	P28598
C-125	Alkalophiles	Alkalophilic <i>Bacillus</i> sp. C-125	
BACST		<i>Bacillus stearothermophilus</i>	L10132
THEP3	Thermophiles	Not Yet classified	P26209
MYCLE		<i>Mycobacterium leprae</i>	P09239
MYCTU		<i>Mycobacterium tuberculosis</i>	P06806
STRAL1		<i>Streptomyces albus</i>	Q00767
STRAL2		<i>Streptomyces albus</i>	Q00767
STAAU		<i>Staphylococcus aureus</i>	Q08854
LACLA		<i>Lactococcus lactis</i>	P37282
BORBU	spirochete	<i>Borrelia burgdorferi</i>	P27575
LEPIN		<i>Leptospira interrogans</i>	P35468
CYACA-C	Red algae	<i>Cyanidium caldarium</i>	P28256
SYNP7	Blue-Green algae	<i>Synechococcus</i> sp. PCC7942	P22879
SYNY3-1	Blue-Green algae	<i>Synechocystis</i> sp. PCC6803	P22034
SYNY3-2	Blue-Green algae	<i>Synechocystis</i> sp. PCC6803	Q05972
YEAST-M	Baker's yeast	<i>Saccharomyces cerevisiae</i>	P19882
ARATH-M	Mouse-ear cress	<i>Arabidopsis thaliana</i>	P29197
MAIZE-M	Maize	<i>Zea mays</i>	P29185
BRANA	Rape	<i>Brassica napus</i>	P35480
CUCMA1	Pumpkin	<i>Cucurbita maxima</i>	Q05045
CUCMA2	Pumpkin	<i>Cucurbita maxima</i>	Q05046
MOUSE-M	Mouse	<i>Mus musculus</i>	P19226
RAT-M	Rat	<i>Rattus norvegicus</i>	P19227
CRIGR-M	Chinese hamster	<i>Cricetulus griseus</i>	P18687
HUMAN-M	Human	<i>Homo sapiens</i>	P10809
HELVI		<i>Heliothis virescens</i>	P25420

* C.叶绿体; M.线粒体; ** 'P or Q' 后面的数字表示 SwissProt 数据库的进入号; 其他的是 Genbank 的进入号。

C. chloroplast; M. mitochondrion; ** Numbers preceded by a 'P or Q' are SwissProt database entries;

The other is Genbank entry.

3 讨论

由于大量不同来源的分子伴娘 60 蛋白序列的发表以及该蛋白序列的保守性,已促使作者利用这些信息来分析生物间的系统发育和进化关系。图 1 结果可明显地看出,以分子伴娘 60 蛋白序列为基础建立起来的系统发育树,与 Woese^[20]用比较 16S rRNA 序列建立起来的系统发育树,在许多方面都是一致的。例如,在真细菌门和其亚类群分类地位、线粒体和叶绿体生物起源的问题方面,两者基本一致。即真细菌门是由紫色细菌、蓝细菌、螺旋体、革兰氏阳性菌、衣原体等类群所组成,其中紫色细菌类群又分成 α -、 β -、 γ -等亚类群;线粒体起源于紫色细菌类群、 α -亚类群的成员;而叶绿体则是由蓝细菌类群的成员进化而成的。

在图 2~4 中,应该注意的几个方面,即真细菌门、紫色细菌类群、 α -亚类群的一个成员 *Richettsia tsutsugamushi* 与真核生物线粒体在分子伴娘 60 蛋白序列上的特殊亲缘关系以及真细菌门、蓝细菌类群的一个成员 *Synechocystis* sp. PCC6803 菌株与植物叶绿体在分子伴娘 60 蛋白序列上的密切亲缘关系。这种一致性从另一方面进一步支持和加强了建立在 16S rRNA 序列分析基础上的线粒体和叶绿体起源的内共生学说,即在远古时代,一种好氧呼吸型细菌侵入另一种细菌中,建立起一种永久的共生关系,通过此后长期的进化就形成了现今的线粒体。同理,一种光合细菌侵入另一种细菌内,建立起共生关系,最后进化形成现今的叶绿体^[1]。图 2~4 中,C-125 菌株与枯草芽孢杆菌的亲缘关系在分子伴娘 60 蛋白序列的基础上比其他成员更密切,也就是说,C-125 菌株的分类地位更靠近枯草芽孢杆菌。总之,分子伴娘 60 蛋白的广泛存在及其在序列上的保守性,使其成为系统发育分析上一种很有实用价值的“分子进化计时器”。

参 考 文 献

- [1] Prescott L M, Harley J P, Klein D A. Microbiology. Oxford: Wm C Brown Publishers, 1993. 416.
- [2] Dickerson R E. *Nature*, 1980, **283**: 210~212.
- [3] Woese C R, Gibson J, Fox G E. *Nature*, 1980, **283**: 212~213.
- [4] Yang D, Oyaizu Y, Oyaizu H *et al.* *Proc Natl Acad Sci USA*, 1985, **82**: 4443~4447.
- [5] Gray M W, Doolittle W F. *Microbiol Rev*, 1982, **46**: 1~42.
- [6] Wallace D C. *Microbiol Rev*, 1982, **46**: 208~240.
- [7] Gray M W, Sankoff D, Cedergren R J. *Nucleic Acids Res*, 1984, **12**: 5837~5852.
- [8] Honda H, Kudo U, Ikura Y *et al.* *Can J Microbiol*, 1985, **31**: 538~542.
- [9] Sanger F, Nicklen S, Coulson A R. *Proc Natl Acad Sci USA*, 1977, **74**: 5463~5467.
- [10] Thompson J D, Higgins D G, Gibson T J. *Nucleic Acids Res*, 1994, **22**: 4673~4680.
- [11] Saitou N, Nei M. *Mol Biol Evol*, 1987, **4**: 406~425.
- [12] Feng D F, Doolittle R F. *J Mol Evol*, 1987, **25**: 351~360.
- [13] Felsenstein J. *Evolution*, 1985, **39**: 783~791.
- [14] Dayhoff M O. Atlas of Protein Sequence and Structure, Vol 5, Supplement 3, 1978. Washington DC: National Biomedical Research Foundation, 1979.
- [15] Margush T, McMorris F R. *Bulletin of Mathematical Biology*, 1981, **43**: 239~244.
- [16] Fitch W M, Margoliash E. *Science*, 1967, **155**: 279~284.
- [17] Eck R V, Dayhoff M O. Atlas of Protein Sequence and Structure Maryland: National Biomedical Research Foundation, Silver Spring, 1966.

- [18] Fitch W M. *Systematic Zoology*, 1971, **20**: 406~416.
[19] Schmidt A, Schiesswohl M, Voelker U *et al.* *J Bacteriol*, 1992, **174**: 3993~3999.
[20] Woese C R. *Microbiol Rev*, 1987, **51**: 221~271.

PHYLOGENY OF MOLECULAR CHAPERONE 60 PROTEINS

Xu Yi Zhou Peijin

(Institute of Microbiology, Academia Sinica, Beijing 100080)

Abstract We present considerable evidence supporting the endosymbiotic hypothesis for the origin of mitochondria and chloroplasts that the member of α -group, purple bacteria, is the ancestor of mitochondria and the member of cyanobacteria is the progenitor of plant chloroplasts. This evidence is based on the relationships among the chaperone 60 proteins, and this result is in very high agreement with that from 16S rRNA sequence analysis. In addition, this result also shows that the alkalophilic *Bacillus* sp. C-125 is closer to *Bacillus subtilis* than the other members in genus *Bacillus* in phylogenetic level.

Key words Molecular chaperone 60, Phylogeny, Molecular evolution, Alkalophilic *Bacillus* sp. C-125