

DOI: 10.3724/SP.J.1005.2012.00907

原核生物 *eno* 基因在系统进化中应用及水平转移分析

刘庆辉¹, 郭振国¹, 任嘉红^{2,3}

1. 河北大学生命科学学院, 保定 071002;
2. 长治学院生物科学与技术系, 长治 046011;
3. 南京林业大学森林资源与环境学院, 南京 210037

摘要: 多基因系统发育研究方法是系统发育分析中的一个重要手段, 基因树冲突已成为分子系统发育研究中日益突出的问题。烯醇化酶基因(*eno*)及其编码的蛋白广泛存在于五界系统中, 烯醇化酶为糖酵解途径中重要酶类。文章选取原核生物已注释的 *eno* 基因序列进行了系统发育分析。对其中的 138 个模式菌株的 *eno* 基因序列进行系统发育分析和同源性搜索, 发现 19 个模式菌株的 *eno* 基因是通过水平转移而来; 并通过核苷酸组成、密码子偏好性和基因排列等基因特征分析, 进一步验证了水平转移基因的外源性。结果表明: 原核生物 *eno* 序列具有较高保守性, 其大小适中, 是研究原核生物系统发育的良好材料。文章在对基因水平转移的供体和受体菌株生活习性、进化历史以及烯醇化酶的结构和功能的研究过程中提供重要参考价值。

关键词: *eno* 基因; 系统发育; 基因水平转移

Phylogenetic application and analysis of horizontal transfer based on the prokaryote *eno* gene

LIU Qing-Hui¹, GUO Zhen-Guo¹, REN Jia-Hong^{2,3}

1. College of Life Sciences, Hebei University, Baoding 071002, China;
2. Department of Biological Science and Technology, Changzhi College, Changzhi 046011, China;
3. College of Forest Resources and Environment, Nanjing Forestry University, Nanjing 210037, China

Abstract: The phenomenon of conflicting gene trees has become a remarkable and difficult problem. Application of multiple genes has been a widespread practice to reconstruct phylogenies in phylogenetic studies. Enolase is a key glycolytic enzyme, The enzymes from a large variety of organisms, including archaeobacteria, eubacteria and eukaryotes, were studied. We downloaded *eno* sequences from the genomes of bacteria and archaea that have been completely sequenced. The comprehensive homology search and phylogenetic analysis of the *eno* were used, and nineteen horizontally transferred genes were identified. The results of analysis showed lots of differences between the features of horizontal transferred genes and the ones of whole genomic genes, such as nucleotide composition, gene combination, codon usage bias, and selection pressure. These results reconfirmed that the horizontally transferred genes were exogenous. The result revealed that pro-

收稿日期: 2011-11-07; 修回日期: 2012-02-03

基金项目: 国家自然科学基金项目(编号: 31100471)资助

作者简介: 刘庆辉, 硕士研究生, 专业方向: 进化基因组学。E-mail: ffgh36@sina.com

通讯作者: 任嘉红, 博士, 副教授, 研究方向: 微生物分子遗传学。E-mail: renjiahong76@yahoo.com.cn

网络出版时间: 2012-6-28 10:58:58

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20120628.1058.006.html>

karyote *eno* genes were highly conserved, medium-sized, is a good material in the phylogenetic. This paper can provide a reference in study of life habit and evolutionary history of donor and receptor, and enolase structure and function.

Keywords: *Eno*; phylogenetic study; horizontal gene transfer

在生物系统发育研究中, 16S rDNA序列分析是目前的主要应用方法。但 16S rDNA基因序列进化上高度保守, 对属内亲缘密切种间分析时, 保守的 16S rDNA基因序列分辨率不够高^[1]; 另外, 同一菌株中, 16S rDNA可能有 2~3 个拷贝, 研究推测这可能由于基因水平转移(Horizontal gene transfer, HGT)所致^[2]。最近几年, 随着测序技术以及全基因组学的发展, 出现了许多根据其他保守基因序列以及编码蛋白进行系统发育分析的例证^[3-6]。但这些保守序列多数为某一谱系中所特有的一种或多个基因, 无法从整体上反映原核生物间的进化关系。此外, 采用多个基因建树进行系统发育分析时, 不同基因片段可能会展现或多或少不同的分支式样甚至严重的分歧, 即基因树之间发生冲突^[7]。

烯醇化酶(Enolase, 2-phospho-D-glycerate hydrolyase)是Lohman和Mayerhof在研究肌肉提取物中磷酸甘油酸向丙酮酸转换过程中发现的^[8]。它既可催化糖酵解过程中 2-磷酸-D-甘油酸(PGA)向磷酸-烯醇式丙酮酸(PEP)的转化, 同时也在糖异生过程中催化逆向反应, 即作为磷酸丙酮酸水合酶, 使PEP向PGA转化, 因此烯醇化酶在细胞能量代谢过程中起重要的作用^[9]。烯醇化酶由*eno*基因编码。*Eno*基因及其酶蛋白广泛分布于古生菌、真细菌以及真核生物中^[10-18], 真核生物包括真菌^[13,14]、植物^[15]、果蝇^[16]和哺乳动物^[17,18]等。对*eno*序列及其酶蛋白的研究多集中于真核生物以及真核与原核生物互作机制^[19, 20], Hannaert等^[10]研究发现眼虫属(*Euglena*)的*eno*序列可能是从一种螺旋体(Spirochaete)中获得; 而单独对原核生物*eno*序列及其酶蛋白的研究较少。大量研究发现, 生物体中同一蛋白可以有不同的功能, 在细胞内部和外部的功能也有差异, 这些功能之间可能无多大的联系^[21]。基因*eno*编码的烯醇化酶除催化糖酵解过程外, 还在不同的生态系统扮演多种角色, 并发挥不同的功能。例如, 烯醇化酶在酵母中可作为一个热休克蛋白^[22]; 在艰难梭菌(*Clostridium diffi-*

cile)中可作为毒素B^[23]; 在一些病原体中可作为纤溶酶原和层黏连蛋白的受体, 可提高细菌毒力, 增加细菌对组织的侵袭力^[24-26]。此外, 烯醇化酶还具有热激蛋白、RNA降解体^[27]等其他方面的作用。

基因水平转移是指在具有生殖隔离的不同物种之间, 或单个细胞的叶绿体、线粒体等细胞器之间, 以及细胞器和细胞核之间所进行的DNA片段流动^[28]。基因交流可以发生在亲缘关系较近的同属不同种之间, 也可以是亲缘关系较远的, 或没有亲缘关系的生物种之间, 有的甚至跨越了界的限制。*Eno*基因广泛分布于五界系统中, 序列比对发现, *eno*基因在进化过程中具有高度保守性, 进化速度慢^[15,26], 这一特点与糖酵解途径的大多数酶是一致的。

本文从原核生物入手, 共选取 497 属中的 817 个种的 1 105 个 *eno* 基因序列进行系统发育分析。对其中的 140 个模式菌株的 *eno* 基因序列进行系统发育分析和同源性搜索。利用多种软件分析了水平转移基因的 GC%含量、核苷酸组成、密码子偏好性、选择压力及基因排布, 发现它们不同于基因组的普遍特性, 从而进一步确定其外源性。

1 材料和方法

1.1 材料

从NCBI(<http://www.ncbi.nlm.nih.gov/gene?term=%22eno>)中选取含有 *eno* 序列的原核生物, 下载其 *eno* 序列, 进行比对分析, 同一种的不同菌株间同源性在 99.9%以上的, 选取其中的一个 *eno* 序列进行分析。本文共选取 497 属中的 817 个种, 其 *eno* 序列共计 1 105 个进行系统发育分析。参照 LPSN(<http://www.bacterio.cict.fr/foreword.html>)和《伯杰氏细菌系统学鉴定手册》(第九版), 将模式菌株作为研究对象。下载模式菌株的 *eno* 序列, 同一属中的不同种的模式菌株进行比对分析, 其同源性在 99%以上的, 选取其中的一株进行后续系统发育分析和同源性搜索。

其全基因组信息来源于NCBI (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>), 从基因水平转移数据库HGT-DB (<http://genomes.urv.es/HGT-DB/>)选取标记*eno*为“hgt”的菌株作为参比菌株^[29]。*Eno*序列在细菌中多数为单拷贝序列^[30], 但也有两个拷贝的, 如 *Amycolatopsis mediterranei* U32(NC 014318)。本文共选取 117 属中的 138 个种, 其中放线菌门(Actinobacteria)的 *Catenulispora acidiphila* DSM 44928^T和厚壁菌门(Firmicutes)的 *Desulfitobacterium hafniense* DCB-2^T各含有 2 个*eno*拷贝序列, 共计 140 个*eno*基因序列进行系统发育和相关分析。

1.2 方法

1.2.1 水平转移基因的鉴定

1.2.1.1 系统发育进化分析 首先, 用Galaxy系统^[31](<http://app.biodiscover.com/>)将下载的*eno*序列进行比对, 使用国际上通用的ClustalX, 参数设置使用默认值^[32]。在MEGA5.05软件中对比对后的*eno*序列进行修饰, 使用最小进化法(Minimum evolution)对 1 105 个*eno*序列构建系统发育树。使用邻接法(Neighbor-joining)、最小进化法、UPGMA(Unweighted pair group method with arithmetic mean)和最大似然法(Maximum likelihood)分别对模式菌株的 140 个*eno*基因序列构建系统发育树, 自举检验重复数选择 1 000。对其进行系统发育分析, 利用结果中的物种信息, 通过Taxonomy数据库和《伯杰氏细菌系统学鉴定手册》(第九版)对应成谱系数据。选择未能与同一谱系(门或纲水平)其他菌株很好聚类的菌株, 其*eno*基因作为下一步研究的候选水平转移基因。

1.2.1.2 相似性搜索 对于候选水平转移基因, 在NCBI中用Blastn程序进行相似性搜索, 筛选出E值小于-30, 序列一致性高于 60%(Identity>60%)的结果, 利用比对结果中的物种信息, 将其通过Taxonomy数据库和《伯杰氏细菌系统学鉴定手册》(第九版)对应成谱系数据。再根据以下标准进行水平转移基因的来源物种分析: E 值由低到高进行分析, 非同物种系谱超过 3 个属, 并且这 3 个属为同一谱系(如门或纲水平), 候选的水平转移基因即可能来源于此。对 19 个候选基因和候选来源物种的 *eno* 基因的上下游 1 000 bp 序列进行分析, 比较其相似性和有无转座特征性序列, 综合考虑水平转移基因的

系统发育树拓扑结构及序列比对的同源性, 确定水平转移基因的来源物种、转移方式和转移方向。

1.2.2 水平转移基因的基因特征分析

在MEGA5.05软件中对选取的 140 个 *eno* 序列进行匹配距离计算(Compute pairwise distances), 得到两两之间的遗传距离; 进行序列特征比对(Compute composition distances), 得到其 GC%含量, 并进行相应菌株的 16S rDNA 基因序列的 GC%含量分析, 从 NCBI 基因组文库中得到相应菌株的全基因组 GC%含量数据。

选取 16S rDNA 和腺苷激酶(adenosine kinase, *adk*)、3-磷酸莽草酸 1-羧乙烯基转移酶(3-phosphoshikimate 1-carboxyvinyltransferase, *aroA*)、ATP 合酶(ATP synthase subunit, *atpD*)、伴侣蛋白(co-chaperonin, *groES*)、DNA 聚合酶III(DNA polymerase III, *dnaE*)、谷氨酰胺合成酶(glutamine synthetase, *glnA*)、柠檬酸合酶(citrate synthase, *gltA*)、丝氨酸羟甲基转移酶(serine hydroxymethyltransferase, *glyA*)、苹果酸脱氢酶(malate dehydrogenase, *mdh*)、DNA 错配修复蛋白(DNA mismatch repair protein, *mutS*)、颤动蛋白(twitching motility protein, *pilT*)、DNA 重组蛋白(DNA recombination protein, *recA*)、DNA 指导的 RNA 聚合酶(DNA-directed RNA polymerase beta subunit, *rpoB*)、甲状腺素受体 α (thyroid hormone receptor alpha, *thrA*)、色氨酸合酶(tryptophan synthase beta subunit, *trpB*)和甘油-3-磷酸结合周质蛋白(glycerol-3-phosphate -binding periplasmic protein precursor, *ugpB*)等 16 个看家基因进行基因特征分析, 作为相应菌株全基因组的基因特征。

运用 Codon W 软件计算候选水平转移基因及全基因组的密码子适应指数(Codon adaptation index, CAI)、密码子偏好性指数(Codon bias index, CBI)、有效密码子数(Effective number of codon, Nc)、最优密码子使用频率(Frequency of optimal codons, FOP)和第三密码子核苷酸组成的 GC3s 值等一些其他常用数值和各密码子使用频率。运用 KaKs_Calculator 软件中的所有模型基于进化选择压力的算法, 研究计算部分物种的水平转移基因的非同义替换率(The number of non-synonymous substitutions per non-synonymous site, Ka)和同义替换率(The number of syn-

onymous substitutions per synonymous site, K_s), 选取 Fisher 检验得出的最低 P 值作为该基因的选择压力数据。根据 Microbesonline (<http://www.microbesonline.org/>) 中的基因组数据及位于候选水平转移基因上下游的编码基因的注释信息, 分析候选水平转移基因上下游基因排布情况, 并构建物种树^[33, 34]。根据物种树, 进一步确定基因的外源性。

2 结果与分析

2.1 *eno* 序列分布

截止到 2012 年 05 月, NCBI(<http://www.ncbi.nlm.nih.gov/genome/browse/>) 中收录的已完成全基因组测序的真细菌为 3 156 个; 古生菌为 192 个, 其中 90% 以上都已鉴定含有 *eno* 序列。NCBI 中收录真细菌的 *eno* 序列数为 2 072 个, 隶属于 432 属 724 个种; 古生菌的 *eno* 序列数为 139 个, 隶属于 66 属 107 个种。表 1 列出的是本研究所用谱系的全基因组信息及 *eno* 序列存在的数量。

2.2 水平转移基因的鉴定

2.2.1 系统发育进化分析

通过对 1 105 个 *eno* 序列进行人工修饰, 最终比对序列长度在 1 200 bp 左右。采用最小进化法构建基于 *eno* 序列的系统发育树(附图 1)。附图 1 中, 同

一谱系(门或纲水平)的菌株聚类明显, 其中包括真细菌域的 α -变形菌纲、 β -变形菌纲、 γ -变形菌纲、放线菌纲、拟杆菌纲、鞘脂杆菌纲、芽孢杆菌纲和梭菌纲等聚群, 以及古生菌域的纳古菌门、泉古菌门和广古菌门的甲烷微菌纲、甲烷杆菌纲、热原体纲和古丸菌纲等聚群。各聚群相互间的同源性较低, 具有明显差异。

通过对 140 个模式菌株的 *eno* 序列比对并进行人工修饰, 最终比对序列长度在 1 300 bp 左右, 采用不同方法构建系统发育树(图 1 和附图 2~4)。从图 1 中可以看出, 以广古菌门(Euryarchaeota)的 4 个菌株作为外群, 其他菌株能够很好的聚在一起, 除了厚壁菌门(Firmicutes)形成两大遗传距离较远的聚类群之外, 其他同一谱系的菌株相互聚群, 各个谱系能够很好的区分开来; 同一聚群内部, 菌株间的相似性较高, 不同聚群间差异明显。但也有个别的 *eno* 序列未与来源同一谱系的其他菌株的 *eno* 序列聚群, 包括放线菌门的 *Ca.acidiphila* DSM 44928^T、 γ -变形菌纲的 *Francisella tularensis* FSC 147^T、厚壁菌门的 *De.hafniense* DCB-2^T 以及广古菌门的 *Methanosaeta thermophila* DSM 6194^T 等菌株。

表 1 真细菌和古生菌的部分全基因组信息

分类	已全基因组测序的菌株数	<i>eno</i> 数量	基因组大小(Mb)	首次完成全基因组测序菌株
Alphaproteobacteria	221	179	0.85~8.84	<i>Rickettsia conorii</i> Malish7 (NC_003103)
Betaproteobacteria	128	169	1.56~9.04	<i>Neisseria meningitidis</i> Z2491 (NC_003116)
Gammaproteobacteria	452	589	0.16~7.15	<i>Haemophilus influenzae</i> Rd KW20 (NC_000907)
Deltaproteobacteria	78	48	1.70~13.03	<i>Geobacter sulfurreducens</i> PCA (NC_002939)
Epsilonproteobacteria	70	66	1.58~2.56	<i>Helicobacter pylori</i> 26695 (NC_000915)
Actinobacteria	232	218	0.93~10.47	<i>Streptomyces coelicolor</i> A3(2) (NC_003888)
Firmicutes	524	415	0.58~7.20	<i>Mycoplasma genitalium</i> G37 (NC_000908)
Chlamydiae	35	21	1.00~2.41	<i>Chlamydia trachomatis</i> D/UW-3/CX (NC_000117)
Cyanobacteria	45	47	1.40~9.04	<i>Nostoc</i> sp. PCC 7120 (NC_003272)
Bacteroidetes	80	79	0.64~9.13	<i>Bacteroides thetaiotaomicron</i> VPI-5482 (NC_004663)
Thermotogae	14	13	1.80~2.30	<i>Thermotoga maritima</i> MSB 8 (NC_000853)
Spirochaetes	47	53	0.92~4.69	<i>Borrelia burgdorferi</i> B 31 (NC_001318)
Planctomycetes	6	5	5.44~7.15	<i>Rhodopirellula baltica</i> SH1 (NC_005027)
Aquificae	11	8	1.50~1.98	<i>Aquifex aeolicus</i> VF5 (NC_000918)
Chlorobi	11	12	1.96~3.30	<i>Chlorobium tepidum</i> ATCC 49652 (NC_002932)

Deferribacteres	4	4	2.51~3.20	<i>Denitrovibrio acetiphilus</i> DSM12809 (NC_013943)
Euryarchaeota	81	97	1.20~7.84	<i>Methanocaldococcus jannaschii</i> DSM2661 (NC_000909)

2.2.2 相似性搜索

对不同方法构建的系统发育树比对分析, 最终得到 19 个 *eno* 序列与来源同一谱系其他菌株的 *eno* 序列具有明显差异, 初步鉴定为候选的水平转移基因。其中 4 个 *eno* 序列为基因水平转移数据库 HGT-DB 中已标记为“hgt”的序列(图 1), HGT-DB 数据库从 GC% 含量和密码子使用偏好性等方面鉴定其是否为水平转移基因, 鉴定过程存在一定的局限性^[29], 本文选取这 4 个 *eno* 序列进行后续深入的研究。所

选 19 个 *eno* 序列的菌株遗传信息和 GenBank 登录号见表 2。对这 19 个候选水平转移基因进行相似性搜索, 在比对结果中, 除同一谱系外, 这些基因序列与其他谱系的菌株有较高的相似性, 具有明显的水平转移的迹象, 初步确定其来源物种(表 2)。

2.3 水平转移基因的特征分析

将 *eno*、16S rDNA 和全基因组序列的 GC% 含量进行比对, 并作出比值分布图(图 2)。对比值进行 *F*-

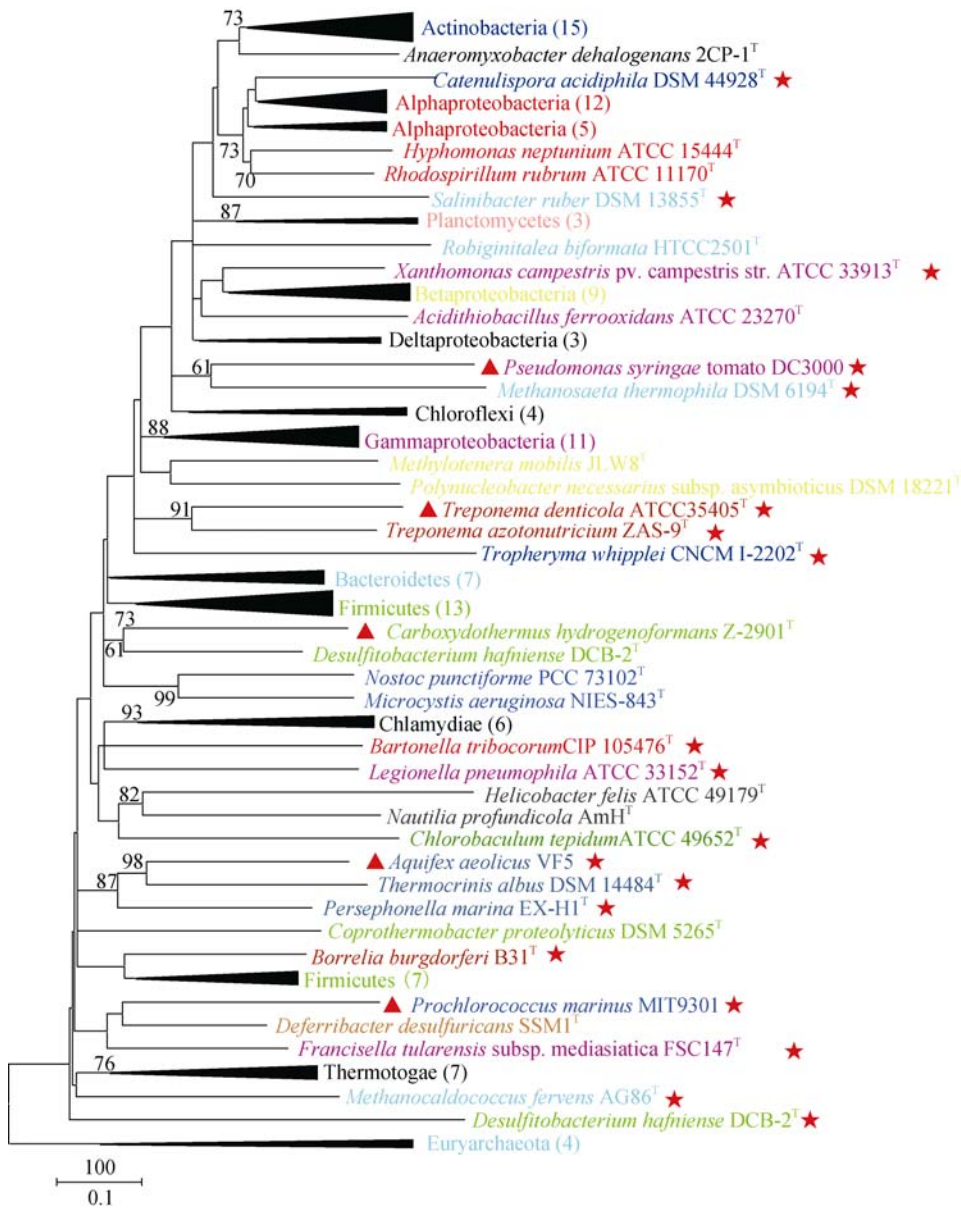


表 2 候选菌株的遗传信息和转移情况

候选菌株	位置(bp)	上下游反向重复序列	来源物种	E 值	转移方式	转移时间
<i>Chlorobaculum tepidum</i> ATCC49652 ^T	147061-148374	上: TGGCTTG-TGA 下: TTGAGACA-A	<i>Listeria welshimeri</i> serovar 6b SLCC5334 ^T	4.00E-46	转座	绿菌门分化前
<i>Pseudomonas syringae</i> DC3000	5211342-5212625	下: TTG-CATCCGCA	<i>Edwardsiella tarda</i> EIB202	2.00E-45	随机插入	属分化后
<i>Prochlorococcus marinus</i> MIT9301	212228-213520	上: (1)TGATTATTAAT (2)ATTTAAATA	<i>Edwardsiella tarda</i> EIB202	2.00E-45	转座	属分化后
<i>Treponema denticola</i> ATCC35405 ^T	972150-973451	下: TGTATCAACAAC	<i>Bacillus amyloliquefaciens</i> FZB42 ^T	3.00E-48	随机插入	科分化后
<i>Aquifex aeolicus</i> VFS ^T	326762-328042	上: GTGGGAGGGA	<i>Caldicellulosiruptor bescii</i> DSM 6725 ^T	5.00E-58	随机插入	门分化前
<i>Tropheryma whippelii</i> str. Twist ^T	899968-901308	上: TCTCTCAGTA 下: GCGTTACTGT	<i>Oligotropha carboxidovorans</i> OM5 ^T	1.00E-34	转座	属分化后
<i>Catenulispora acidiphila</i> DSM44928 ^T	5746040-5747326	上: GCTGCGCAGC 下: TCGGCCITGGC	<i>Methylobacterium radotolerans</i> JCM 2831	0	转座	目分化后
<i>Salinibacter rubber</i> DSM13855 ^T	998974-1000263	无	<i>Nocardopsis dassonvillei</i> subsp. dassonvillei DSM 43111 ^T	3.00E-156	随机插入	科分化后
<i>Treponema azotonutricium</i> ZAS-9 ^T	3312753-3314054	上: GTCCTATCA 下: AAGAAAAGCTGCC	<i>Geobacter sulfurreducens</i> ATCC 51573 ^T	6.00E-114	转座	科分化后
<i>Borrelia burgdorferi</i> B31 ^T	345063-346364	上: CAAAAA 下: AAATCCCT	<i>Clostridium acetobutylicum</i> ATCC 824 ^T	2.00E-152	转座	属分化前
<i>Bartonella tribocorum</i> CIP105476 ^T	913849-915117	上: TATTTTCT 下: TGTAAAAAT	<i>Thermoanaerobacter</i> <i>pseudethanolicus</i> ATCC 33223 ^T	6.00E-70	转座	属或科分化后
<i>Xanthomonas campestris</i> ATCC33913 ^T	197783-1979075	下: GCGGTACTGC	<i>Cupriavidus taiwanensis</i> LMG 19424 ^T	0	随机插入	科分化前
<i>Legionella pneumophila</i> ATCC33152 ^T	2281241-2282509	上: TTTACAAGCG	<i>Listeria welshimeri</i> serovar 6b SLCC5334 ^T	1.00E-33	随机插入	该属分化后
<i>Francisella tularensis</i> FSC147 ^T	1477673-1479043	上: TTATCATCA 下: TATTACTAA	<i>Clostridium perfringens</i> ATCC 13124 ^T	2.00E-113	转座	该属分化前
<i>Desulfotobacterium hafniense</i> DCB-2 ^T	2235600-2236913	下: GAAAATCCGTG	—	>e-10	随机插入	属或科分化前
<i>Persephonella marina</i> EX-H1 ^T	1583661-1584947	下: (1)GTATATTCAG (2)ATCTTTTAT	<i>Caldicellulosiruptor bescii</i> DSM 6725 ^T	9.00E-106	转座	门分化前
<i>Thermocrinis albus</i> DSM14484 ^T	549297-550583	上: (1)TCCTCCAC (2)GTGAGAAAAGG	<i>Desulfotomaculum carboxydovorans</i> CO-1-SRB ^T	3.00E-41	转座	门分化前
<i>Methanosaepta thermophila</i> DSM16194 ^T	895433-896695	上: GGAGAGATCAT 下: ATCTGAAGGAGA	<i>Thermaerobacter marianensis</i> DSM 12885 ^T	1.00E-46	转座	属分化后
<i>Methanocaldococcus fervens</i> AG86 ^T	426388-427659	上: AATATTTTGTAGTAT 下: TAAGCAATG	<i>Caldicellulosiruptor bescii</i> DSM 6725 ^T	5.00E-71	转座	目分化后

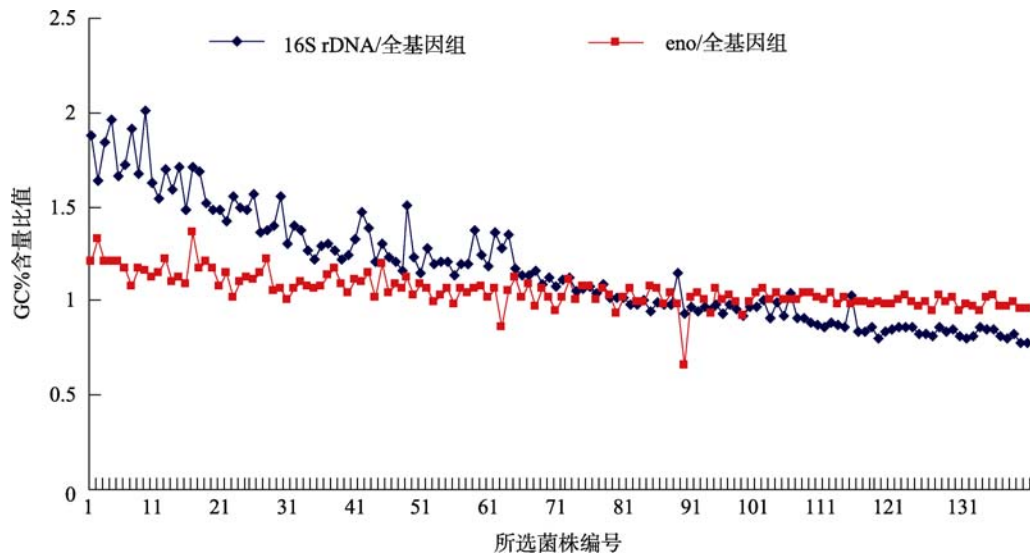


图 2 所选菌株的 *eno*、16S rDNA 与全基因组的 GC%含量比值(以全基因组的 GC%含量为标准)

检验,结果显示 eno 与全基因组GC%含量的比值在0.655~1.368之间,平均值为1.051,方差为0.007;16S rDNA与全基因组GC%含量的比值在0.779~2.011之间,平均值为1.164,方差为0.088。与前者相比,16S rDNA的GC%含量与全基因组GC%存在明显差异。 eno 序列与全基因组序列的GC%含量具有很高的一致性,但有3个明显的差异点,分别为 *Fervidobacterium nodosum* Rt17-B1^T、*Persephonella marina* EX-H1^T和 *Ch.tepidum* ATCC 49652^T。其中 *Pe.marina* EX-H1^T和 *Ch.tepidum* ATCC 49652^T的 eno 基因为本研究中的候选水平转移基因, *Fe.nodosum* Rt17-B1^T在系统发育树中与同一谱系的其他菌株聚类,其 eno 基因是否为水平转移基因还有待进一步研究确定。从图中也可看出在菌株的全基因组GC%含量较低时,其 eno 基因的GC%含量高于全基因组的GC%含量;在全基因组的GC%含量较高的菌株中,其 eno 基因的GC%含量与全基因组的GC%含量持平或低于全基因组的GC%含量。

对19个候选菌株的 eno 水平转移基因的密码子指数数据与全基因组的比值进行了统计(图3)。发现候选的19个菌株的 eno 基因的GC、CAI、Fop、Nc及GC3s这5个密码子指数与全基因组的比值在1附近,而候选水平转移基因的CBI与全基因组的CBI比值变化较大。其中菌株 *Ch. tepidum* ATCC 49652^T、*Me.fervens* AG86^T、*Ps.syringae* DC 3000、

Fr.tularensis FSC 147^T、*Ba.tribocorum* CIP 105476^T、*Tr.azotonutricium* ZAS-9^T、*Me.thermophila* DSM 6194^T、*Tr.denticola* ATCC 35405^T、*Ca.acidiphila* DSM 44928^T和 *Bo.burgdorferi* B31^T的 eno 基因的GC、CAI、CBI、Fop、Nc及GC3s这6个密码子指数与全基因组的密码子指数具有差异性,可能是因为这些菌株的 eno 基因水平转移进入基因组的时间较短,密码子偏好性还存在。其他几个菌株的 eno 基因的密码子指数与全基因组的密码子指数比较接近,这有可能是因为这些菌株的 eno 基因转移进入基因组的时间较长,在进化过程中,密码子的使用已经逐渐偏好使用寄主基因的密码子^[35]。

统计19个候选菌株的 eno 基因和全基因组的密码子使用情况,对候选的19个菌株的 eno 基因密码子使用频率与全基因密码子使用频率进行比较。一般认为比值若在0.5~2.0,表明两者对密码子偏好性比较接近,反之则差异较大。从附表1可以得出:*Tr.whipplei* Twist^T的候选基因与全基因组的密码子使用频率比值大于2小于0.5的密码子数量最少为9个, *Ch.tepidum* ATCC 49652^T的候选基因与全基因组的密码子使用频率比值大于2小于0.5的密码子数量最多为50个,结果显示候选的19个菌株的 eno 基因与全基因组相比具有显著的密码子偏好性,也进一步支持了候选基因是通过水平转移而来的结论。

对19个候选菌株的 eno 基因和来源物种的 eno

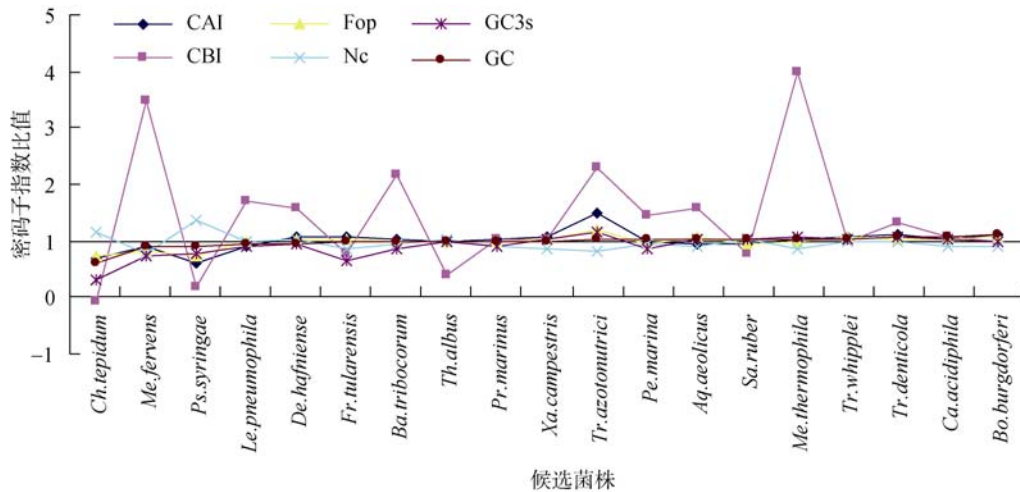


图 3 候选基因与全基因的密码子指数比值(以 GC 比值的升序排列)

基因序列进行了选择压力分析(附表 2)。分析结果显示不同物种的 *eno* 基因间的 Ka/Ks 相近, 都是 $Ka < Ks$ 的负向选择。说明氨基酸在负选择压力下非常保守的突变; 也进一步证明了 *eno* 基因进化上非常保守。

在 MicrobesOnline 网站中, 分析 19 个候选菌株的 *eno* 基因上下游基因排布, 其中 *Tr.azotonutricium* ZAS-9^T 和 *Ch.tepidum* ATCC 49652^T 未被 Microbesonline 数据库收录, 我们分析了其余 17 个候选菌株的 *eno* 基因上下游基因排布情况(图 4 和附图 5)。图 4A 物种树包括了 3 个候选物种, 并且都隶属于产水菌门, 其中产水菌门的 9 株菌株聚为一类, 与其他菌株相互分开, 但彼此之间也有一些差异, 推测水平转移时间发生在产水菌门分化之前; 图 4B、C、D 中 *Bo.burgdorferi* B31^T、*Tr.whipplei* Twist^T、*Fr.tularensis* FSC147^T 与同属的其他菌株聚类明显, 而与其他属的菌株差异明显; 附图 1 中, *Bo.burgdorferi* B31^T、*Tr.whipplei* Twist^T、*Fr.tularensis* FSC147^T 与同属的其他菌株聚群, 但不与同一科水平上的其他菌株聚群, 推测它们的水平转移事件发生在相应属分化之后。根据附图 5 中候选物种与其他物种在基因排布上的差异, 结合附图 1 中对于原核生物的 1 105 个 *eno* 序列构建的系统发育树信息, 进而推测发生水平转移的时间(表 2)。

此外, 在 MicrobesOnline 数据库中, 通过构建不同基因的物种树, 可发现潜在的水平转移现象^[33, 34], 图 4C 中, 棒杆菌科(Corynebacteriaceae)的菌株与其他放线菌门的菌株分开, 推测棒杆菌亚目的 *eno* 基因通

过水平转移而来。附图 1 中选取棒杆菌科的 *Corynebacterium* 1 属的 11 个 *eno* 序列进行了系统发育分析, 其与放线菌门的其他菌株聚成一类。该属的各菌株间同源性较高, 与其他属菌株具有明显差异, 其 *eno* 序列是否为水平转移而来还有待进一步验证。

3 讨论

3.1 *eno* 序列在系统分类学上应用

原核生物系统分类学研究中, rDNA 序列分析法是主要的鉴定指标。其中 16S rDNA 基因序列因其细胞所共有、功能同源且最为古老、既含有保守序列又含可变序列、分子大小适合操作等特点, 成为最常用的系统发育标记分子。但由于 16S rDNA 基因的高度保守性, 使得对亲缘关系较近的物种分辨率不够高。随着测序技术以及全基因组学的发展, 人们逐渐将目光转移到其他保守基因序列以及编码蛋白上^[3-6]。当用多个基因建树进行系统发育分析时, 由于随机误差、系统误差以及生物学因素等原因, 不同基因片段可能会展现或多或少的基因树冲突^[7]。为尽量真实的反映原核生物的系统发育进化关系, 需要结合多个保守基因, 综合全面地进行分析。这些基因序列应具有保守性强、大小适中易于操作、原核生物普遍具有及保守区和可变区并存等特点。

本文以隶属于真细菌和古生菌的 497 属中的 817 个种为研究对象, 构建了基于 1 105 个 *eno* 序列的系统发育树(附图 1)。以隶属于古生菌的部分菌株

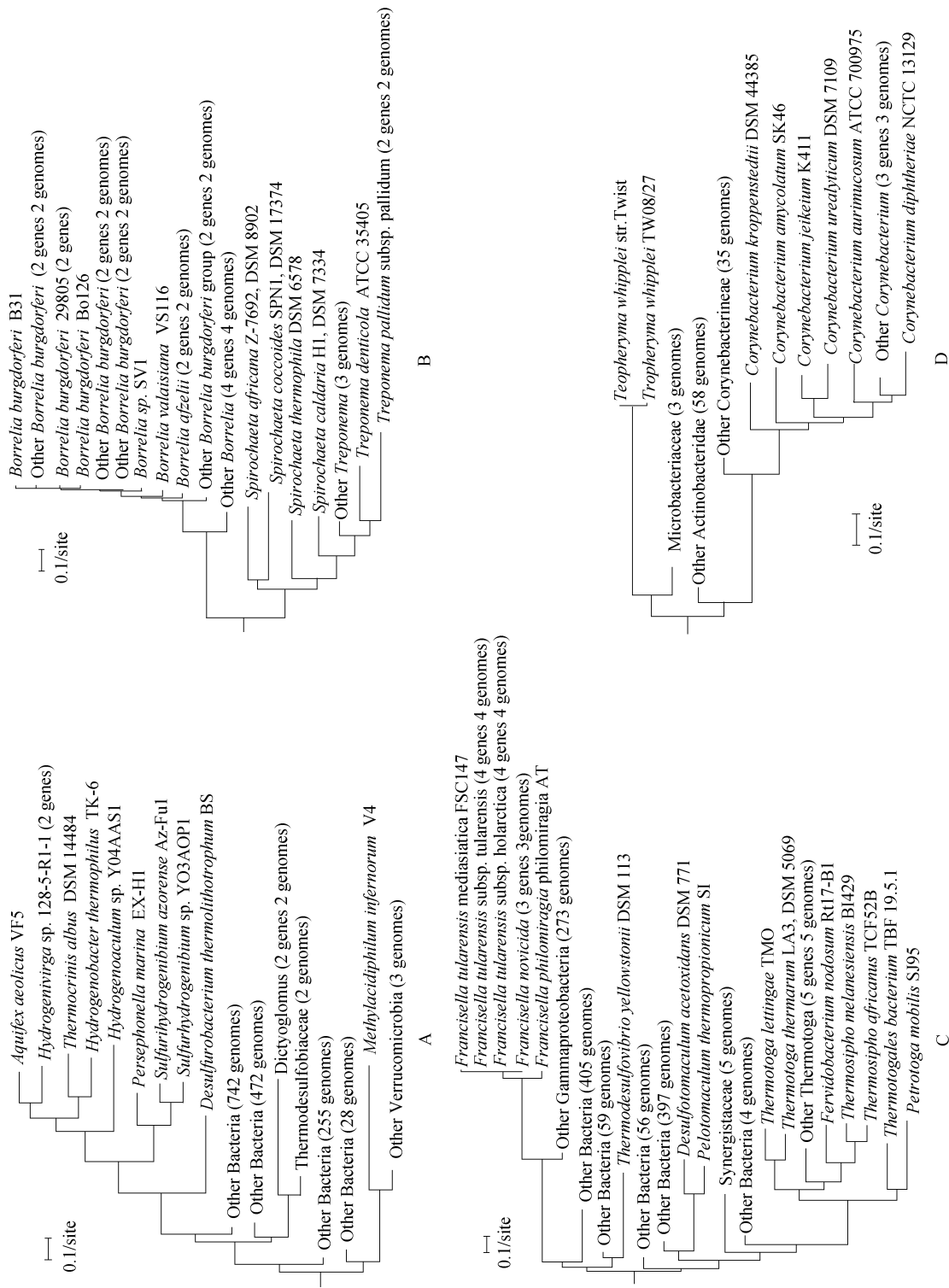


图 4 部分候选基因排列分布物种树

A 为 *Aq. aeolicus* VF5^T、*Pe. marina* EX-H1^T 和 *Th. albus* DSM 14484^T; B 为 *Bo. burgdorferi* B31^T; C 为 *Fr. tularensis* FSC147^T; D 为 *Tr. whipplei* Twist^T。

的*eno*序列作为根类群, 其他菌株能够很好的聚为一群, 并形成 α -变形菌纲、放线菌纲、芽孢杆菌纲和拟杆菌纲等谱系的小聚群。同一谱系的多数菌株形成单一聚群, 不同分类单元的聚群相互分开, 差异明显。此外, 同一聚群的内各菌株在种甚至菌株水平, 可以很好的区分开来(附图 6)。对基因序列的GC%含量的研究文献较多, 证实不同基因的GC%含量具有差异性^[36-37]。16S rDNA具有高度保守性, 其GC%含量在 50%与 60%之间, 不同分类单元间的GC%含量差异较小, 不能通过GC%含量将其分类。本研究的 140 个*eno*基因的GC%含量与其全基因组的GC%含量具有很高的相似度(图 2), 菌株*eno*的GC%含量可以间接反应全基因组的GC%含量。Pancholi^[8]对包括大肠杆菌、疟原虫、拟南芥、果蝇和人类在内的 39 个代表物种的*eno*编码蛋白序列进行了比对分析, 烯醇化酶蛋白序列具有很高的保守性, 两个物种间的同源性在 40%~90%之间, 序列的保守区和可变区间隔排列, 在比对的中部 280 bp、C端和N端附近含有长片段的高保守性区域, 便于设计引物, 进行全序列或可变区序列的测定。本文通过对部分原核生物的烯醇化酶蛋白序列进行比对分析, 也发现序列的保守区和可变区间隔排列这一特征(附图 7)。

越来越多的证据显示HGT为原核生物进化的重要动力, Santiago等^[38]对 17 个真细菌和 7 个古生菌的全基因组序列分析发现: 通过水平转移得到的基因可以占到全基因数的 1.5%~14.5%。建立在单基因基础上的系统发育未考虑远源物种之间的基因转移, 选取通过水平转移而来的单基因, 势必会影响系统发育分析的结果。在确定种的分类地位和系统发育地位时需要选取多个保守基因, 结合多项指标综合考虑, 才能得出较准确的系统发育信息。*eno*基因具有高保守性、长度适中易于测定、保守区与可变区交替排列和同属间差异性较大等特点, 其编码的烯醇化酶为糖酵解途径中重要的酶类, 在各类生物体中分布广泛, 作为一项精细的分子指标, 可以补充因rDNA基因序列同源性高, 而不能界定系统发育位置关系十分密切种(菌株)的缺点。*eno*基因结合其他保守基因在种群分类、鉴定和系统发育学研究中具有很好的应用价值^[19, 20], 其的应用将会使原核生物的系统分类更科学、更全面。

3.2 *eno* 序列及其酶蛋白的水平转移分析

在五界系统的基因组相互比对中, 发现存在大量基因水平转移现象^[39-47], 其广泛存在于细菌与细菌^[42]、细菌与古菌^[43, 44]、细菌与动物^[45]、细菌与植物^[46]以及高等生物^[47]之间。目前, 基因水平转移数据库 HGT-DB(<http://usuaris.tinet.cat/debb/HGT/welcome.OLD.html>)已收录了 94 种原核生物(78 种真细菌和 16 种古生菌)发生基因水平转移的数据和相关证据。

原核生物间的水平转移方式主要有转化、接合和转导。真核生物中, 转化、接合和转导等现象很少发生, 其基因水平转移的方式更加复杂。最易于被人们接受的真核基因水平转移的方式是: 宿主生物和与其寄生的生物通过相互接触实现基因水平转移, 或者借助病毒、细菌、昆虫等载体来实现基因的水平转移。对基因水平转移的转移方式研究, 除对基因序列进行分析外, 还需要结合供体和受体的生活习性、地理分布、进化历史等。研究发现, 与亲缘关系和地理位置相比, 相同环境生态位更能决定一个转移基因是否能融入新的基因组中^[48]。通过本文研究发现*eno*基因在进化过程中受到基因水平转移的影响^[20], 在细菌与细菌、细菌与古菌间发生过水平转移事件。本研究不仅为菌株间的系统发育分析提供参考价值, 还可为发生水平基因转移时期, 各种生物的氧气耐受力等生活习性以及周围环境的研究提供重要研究数据。

*eno*基因编码的烯醇化酶在不同的生物体中, 除在糖酵解过程中起催化作用外, 还具有其他特殊的功能。这些功能包括热激休克蛋白、毒素、RNA降解体以及抗体识别和肿瘤转移等^[22-27]。本研究中 *Bo.burgdorferi* B31^T的*eno*序列为水平转移而来, 其烯醇化酶除催化糖酵解途径外, 还是细胞表面纤溶酶结合蛋白^[26]; 其*eno*序列的来源物种为 *Cl.acetobutylicum* ATCC 824^T, 该种的烯醇化酶是否具有纤溶酶结合蛋白活性还有待进一步验证。本研究表明, 在进行不同菌株间烯醇化酶的特殊结构和功能研究时, 通过同源性分析, 可为不同菌株间烯醇化酶的结构、功能研究提供参考价值。

附录: 附图 1-7 和附表 1-2 见文章电子版(www.Chinagene.cn)。

参考文献(References):

- [1] Stackebrandt E, Goebel BM. Taxonomic Note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol*, 1994, 44(4): 846-849. [DOI](#)
- [2] Yap WH, Zhang ZS, Wang Y. Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora Chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J Bacteriol*, 1999, 181(17): 5201-5209. [DOI](#)
- [3] Gaunt MW, Turner SL, Rigottier-Gois L, Lloyd-Macgilp SA, Young JP. Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. *Int J Syst Evol Microbiol*, 2001, 51(6): 2037-2048. [DOI](#)
- [4] Matthew AP. rRNA and *dnaK* relationships of *Bradyrhizobium* sp. nodule bacteria from four Papilionoid legume trees in Costa Rica. *Syst Appl Microbiol*, 2004, 27(3): 334-342. [DOI](#)
- [5] Andam CP, Mondo SJ, Parker MA. Monophyly of *nodA* and *nifH* genes across Texan and Costa Rican populations of *Cupriavidus* nodule symbionts. *Appl Environ Microbiol*, 2007, 73(14): 4686-4690. [DOI](#)
- [6] Sánchez B, Zúniga M, González-Candelas F, de los Reyes-Gavilán CG, Margolles A. Bacterial and eukaryotic phosphotolases: phylogeny, distribution and evolution. *J Mol Microbiol Biotechnol*, 2010, 18(1): 37-51. [DOI](#)
- [7] 邹新慧,葛颂. 基因树冲突与系统发育基因组学研究. *植物分类学报*, 2008, 46(6): 795-807. [DOI](#)
- [8] Pancholi V. Multifunctional α -enolase: its role in diseases. *Cell Mol Life Sci*, 2001, 58(7): 902-920. [DOI](#)
- [9] Subramanian A, Miller DM. Structural analysis of α -enolase, mapping the functional domains involved in down-regulation of the c-myc protooncogene. *J Biol Chem*, 2000, 275(8): 5958-5965. [DOI](#)
- [10] Hannaert V, Brinkmann H, Nowitzki U, Lee JA, Albert MA, Sensen CW, Gaasterland T, Muller M, Michels P, Martin W. Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoeba balamuthi*, and from the chloroplast and cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. *Mol Biol Evol*, 2000, 17(7): 989-1000. [DOI](#)
- [11] Piast M, Kustrzeba-Wójcicka I, Matusiewicz M, Banaś T. Molecular evolution of enolase. *Acta Biochim Pol*, 2005, 52(2): 507-513. [DOI](#)
- [12] Gerlt JA, Babbitt PC, Jacobson MP, Almo SC. Divergent evolution in enolase superfamily: strategies for assigning functions. *J Biol Chem*, 2012, 287(1): 29-34. [DOI](#)
- [13] Brewer JM. Yeast enolase: Mechanism of activation by metal ion. *CRC Crit Rev Biochem*, 1981, 11(3): 209-254. [DOI](#)
- [14] Gandhi NS, Young K, Warmington JR, Mancera RL. Characterization of sequence and structural features of the *Candida krusei* enolase. *In Silico Biol*, 2008, 8(5-6): 449-460. [DOI](#)
- [15] Van Der Straeten D, Rodrigues-Pousada RA, Goodman HM, van Montagu M. Plant enolase: Gene structure, expression, and evolution. *Plant Cell*, 1991, 3(7): 719-735. [DOI](#)
- [16] Bishop JG, Corces VG. The nucleotide sequence of a *Drosophila melanogaster* enolase gene. *Nucleic Acids Res*, 1990, 18(1):191. [DOI](#)
- [17] Sakimura K, Kushiya E, Obinata M, Takahashi Y. Molecular cloning and the nucleotide sequence of cDNA to mRNA for don-nearonal enolase ($\alpha\alpha$ enolase) of rat brain and liver. *Nucleic Acids Res*, 1985, 13(12): 4365-4378. [DOI](#)
- [18] Giallongo A, Oliva D, Cali L, Barba G, Barbieri G, Feo S. Structure of the human gene for α -enolase. *Eur J Biochem*, 1990, 190(3): 567-573. [DOI](#)
- [19] Baptiste E, Philippe H. The potential value of indels as phylogenetic markers: Position of *Trichomonads* as a case study. *Mol Biol Evol*, 2002, 19(6): 972-977. [DOI](#)
- [20] Keeling PJ, Palmer JD. Lateral transfer at the gene and subgenomic levels in the evolution of eukaryotic enolase. *Proc Natl Acad Sci USA*, 2001, 98(19): 10745-10750. [DOI](#)
- [21] Jeffery CJ. Moonlighting proteins. *Trends Biochem Sci*, 1999, 24(1): 8-11. [DOI](#)
- [22] Sundstrom P, Aliaga GR. Molecular cloning of cDNA and analysis of protein secondary structure of *Candida albicans* enolase, an abundant, immunodominant glycolytic enzyme. *J Bacteriol*, 1992, 174(21): 6789-6799. [DOI](#)
- [23] Bissert F, Keith G, Rihn B, Amiri I, Werneburg B, Girardot R, Baldacini O, Green G, Nguyen VK, Monteil H. *Clostridium difficile* toxin B: Characterization and sequence of three peptides. *J Chromatogr*, 1989, 490(1): 91-100. [DOI](#)
- [24] Ehinger S, Schubert WD, Bergmann S, Hammerschmidt S, Heinz DW. Plasmin (ogen)-binding α -enolase from streptococcus pneumoniae: crystal structure and evaluation of plasmin (ogen)-binding sites. *J Mol Biol*, 2004, 343(4): 997-1005. [DOI](#)
- [25] Seweryn E, Pietkiewicz J, Szamborska A, Gamian A. Enolase on the surface of prokaryotic and eukaryotic cells is a receptor for human plasminogen. *Postepy Hig Med Dosw*, 2007, 15(61): 672-682. [DOI](#)

- [26] Floden AM, Watt JA, Brissette CA. *Borrelia burgdorferi* enolase is a surface exposed plasminogen binding protein. *PLoS One*, 2011, 6(11): e27502. [DOI](#)
- [27] Kaberdin VR, Lin-Chao S. Unraveling new roles for minor components of the *E. coli* RNA degradosome. *RNA Biol*, 2009, 6(4): 402–405. [DOI](#)
- [28] 李志江, 海权, 刁现民. 基因水平转移的评判方法和转移方式研究进展. *遗传*, 2008, 30(9): 1108–1114. [DOI](#)
- [29] Garcia-Vallve S, Guzman E, Montero MA, Romeu A. HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Res*, 2003, 31(1): 187–189. [DOI](#)
- [30] Lal SK, Johnson S, Conway T, Kelley PM. Characterization of a maize cDNA that complements an enolase-deficient mutant of *Escherichia coli*. *Plant Mol Biol*, 1991, 16(5): 787–795. [DOI](#)
- [31] Afgan E, Baker D, Coraor N, Goto H, Paul IM, Makova KD, Nekrutenko A, Taylor J. Harnessing cloud computing with Galaxy Cloud. *Nat Biotechnol*, 2011, 29(11): 972–974. [DOI](#)
- [32] Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, 1997, 25(24): 4876–4882. [DOI](#)
- [33] Alm EJ, Huang KH, Price MN, Koche RP, Keller K, Dubchak IL, Arkin AP. The MicrobesOnline Web site for comparative genomics. *Genome Res*, 2005, 15(7): 1015–1022. [DOI](#)
- [34] Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res*, 2010, 38(Suppl. 1): D396–D400. [DOI](#)
- [35] 段海蓉, 丘德彬, 贡成良, 黄茉莉. 家蚕核型多角体病毒水平转移基因分析. *遗传*, 2011, 33(6): 636–647. [DOI](#)
- [36] Mizi A, Zouros E, Moschonas N, Rodakis GC. The complete maternal and paternal mitochondrial genomes of the mediterranean mussel *Mytilus galloprovincialis*: implications for the doubly uniparental inheritance mode of mtDNA. *Mol Biol Evol*, 2005, 22(4): 952–967. [DOI](#)
- [37] Yukawa H, Omumasaba CA, Nonaka H, Ko's P, Okai N, Suzuki N, Suda M, Tsuge Y, Watanabe J, Ikeda Y, Verte's AA, Inui M. Comparative analysis of the *Corynebacterium glutamicum* group and complete genome sequence of strain R. *Microbiology*, 2007, 153(4): 1042–1058. [DOI](#)
- [38] Garcia-Vallve S, Romeu A, Palau J. Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res*, 2000, 10(11): 1719–1725. [DOI](#)
- [39] Beiko RG, Harlow TJ, Ragan MA. Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci USA*, 2005, 102(40): 14332–14337. [DOI](#)
- [40] Lerat E, Daubin V, Ochman H, Moran NA. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol*, 2005, 3(5): e130. [DOI](#)
- [41] Whitchurch CB, Tolker-Nielsen T, Ragas PC, Mattick JS. Extracellular DNA required for bacterial biofilm formation. *Science*, 2002, 295(5559): 1487–1487. [DOI](#)
- [42] 杨成运, 李友国, 魏力, 程国军, 周俊初. 华癸中生根瘤菌 2020 三个内源质粒的功能及其与豌豆根瘤菌共生质粒之间的相互关系. *中国科学·辑: 生命科学*, 2008, 38(3): 250–258. [DOI](#)
- [43] Koonin EV, Wolf YI. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res*, 2008, 36(21): 6688–6719. [DOI](#)
- [44] Lake JA, Rivera MC. Deriving the genomic tree of life in the presence of horizontal gene transfer: conditioned reconstruction. *Mol Biol Evol*, 2004, 21(4): 681–690. [DOI](#)
- [45] 程廷才, 夏庆友, 刘春, 赵萍, 查幸福, 徐汉福, 向仲怀. 家蚕 *chi*、*gluE* 和 *fruA* 基因与微生物相应基因的同源性及其基因水平转移初探. *遗传学报*, 2004, 31(10): 1082–1088. [DOI](#)
- [46] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409(6822): 860–921. [DOI](#)
- [47] Diao X, Freeling M, Lisch D. Horizontal transfer of a plant transposon. *PLoS Biol*, 2006, 4(1): e5. [DOI](#)
- [48] Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*, 2011, 480(7376): 241–244. [DOI](#)