

家蚕氨酰-tRNA 合成酶基因分析*

曹广力^{1,2}, 薛仁宇^{1,2}, 朱越雄¹, 魏育红¹, 贡成良^{1,2}

1 苏州大学医学部, 苏州 215123;

2 苏州大学 现代丝绸国家工程实验室, 苏州 215123

摘要: 为了探讨家蚕氨酰-tRNA 合成酶 (BmaaRS) 基因的数目、种类、结构及起源, 利用家蚕基因组数据和 EST 数据进行了 *BmaaRS* 基因的电子克隆, 结果表明家蚕核基因组中含有 2 套不同的 aaRS 核基因, 分别编码线粒体和细胞质 BmaaRS, 但编码线粒体 BmSerRS 的基因有 2 个, 可能缺少编码细胞质的 *BmHisRS* 基因和编码线粒体的 *BmGlnRS*、*BmLysRS*、*BmGlyRS* 和 *BmThrRS* 基因, 这些基因的功能可能由具有相似功能的其他蛋白完成, 或通过某个 *BmaaRS* 基因的可变剪接分别形成不同功能的 BmaaRS。EST 证据表明 *BmaaRS* 基因存在不同形式的可变剪接; BmaaRS 氨基酸序列的相似性及二、三级结构分析表明部分 BmaaRS 存在结构域的扩增, 有些不同的 BmaaRS 具有相同结构域, 相同功能的 BmaaRS 具有相似的三级结构; 进化分析表明 BmaaRS 为 2 套不同来源的 *BmaaRS* 基因编码, 细胞质和线粒体 *BmaaRS* 的起源不同。

关键词: 家蚕; 氨酰-tRNA 合成酶; 核基因; 线粒体; 电子克隆

Analysis of the aminoacyl-tRNA synthetase genes of silkworm(*Bombyx mori*)

CAO Guang-Li^{1,2}, XUE Ren-Yu^{1,2}, ZHU Yue-Xiong¹, WEI Yu-Hong¹, GONG Cheng-Liang^{1,2}

1 Medical college of Soochow University, Suzhou 215123, China;

2 National Engineering Laboratory for Modern Silk Soochow University, Suzhou 215123, China

Abstract: For further research on number, type, composition and origin of *Bombyx mori* aminoacyl-tRNA synthetase (BmaaRS) genes, in silico cloning was performed with *Bombyx mori* genomic and EST databases. There might be two different sets of aaRS nuclear gene in *Bombyx mori* genome, which encode mitochondrial BmaaRS and cytoplasmic BmaaRS, respectively. Among *BmaaRS* genes, there were 2 genes encoding mitochondrial BmSerRS, but no genes encoding cytoplasmic BmHisRS and mitochondrial BmGlnRS, BmLysRS, BmGlyRS, and BmThrRS. The functions of these absent genes could be directly replaced by other proteins with similar functions, or might undergo their distinct BmaaRS functions based on the alternative splice of one certain *BmaaRS* mRNA. Evidence of EST indicated that *BmaaRS* performed different alternative splicing patterns. The homology comparison and advanced structural analysis of BmaaRS demonstrated the existence of extended domains of BmaaRS. This is because some different BmaaRSs contained similar domain. Moreover, BmaaRSs with similar functions possessed the similar tertiary structure. Phylogenetic analysis revealed that BmaaRS encoded by two various sources of *BmaaRS* genes. Mitochondrial and cytoplasmic *BmaaRS* had different origin.

Keywords: *Bombyxmori*; aminoacyl-tRNA synthetase (aaRS); nuclear gene; mitochondrial; in silico cloning

收稿日期: 2009-04-27; 修回日期: 2009-08-12

基金项目: 国家高技术研究发展计划项目 (863 计划) (编号: 2006AA10A118) 和国家自然科学基金项目 (编号: 30671590) 资助

作者简介: 曹广力 (1964—), 男, 副教授, 研究方向: 家蚕分子遗传学。Tel: 0512-65880183; E-mail: guanglicao@hotmail.com

通讯作者: 贡成良 (1965—), 男, 教授, 研究方向: 家蚕病理、生化与分子生物学。Tel: 0512-65880183; E-mail: chlqong@hotmail.com

氨酰-tRNA合成酶 (aminoacyl-tRNA synthetase, aaRS) 也称氨酰tRNA连接酶或氨基酸活化酶, 是生命进化过程中最早出现的一类蛋白质。aaRS能专一性地辨认氨基酸的侧链和tRNA, 在真核生物体内一些aaRS以多聚复合物的形式行使功能, 而且aaRS参与了转录和翻译水平的调控、编校、tRNA的成熟与转运、RNA剪接、信号传导和免疫应答等功能^[1,2]。

线粒体拥有自己的DNA (mtDNA), 带有少量结构基因, 可以进行转录、翻译和蛋白质合成。人类约98%的线粒体蛋白质由核基因编码, 在胞浆中合成后被转入线粒体^[3], 也就是说, 线粒体的功能是由核基因和线粒体基因编码的蛋白质共同完成^[4]。线粒体蛋白质结构与功能的改变与人类许多疾病相关, 如退行性疾病、心脏病、衰老和癌症等, 线粒体蛋白质的研究日益受到关注^[5]。通常真核生物拥有2套起源不同的aaRS核基因, 分别编码线粒体和细胞质aaRS, 参与线粒体功能的aaRS也是由核基因编码。研究表明, 线粒体aaRS与人类的疾病密切相关^[6]。

根据aaRS包含的与ATP相结合的催化核心结构域, 可以将aaRS分为两类 (I类和II类), 每一种aaRS可以用该种氨基酸的英文三字符号后缀RS表示^[7]。大多数aaRS的tRNA转运识别功能非常专一, 但也有例外, 可以交替识别, 转运不同的tRNA, 例如在有些原核生物中不存在GlnRS, 其功能由GluRS来替代, 大多数古细菌和几种细菌中, AsnRS的功能由AspRS替代^[8,9]。根据内共生理论, 线粒体aaRS可能来源于 α -变形杆菌, 但也可能存在其他进化途径^[10]。

家蚕 (*Bombyx mori*) 是鳞翅目 (Lepidoptera) 昆虫的典型代表, 在家蚕中已有个别氨酰-tRNA合成酶基因 (*BmaaRS*) 的研究报道^[11,12]。家蚕基因组序列的完成, 为鳞翅目昆虫的生物信息学研究提供了良好的平台, 研究aaRS的结构、功能和进化, 对深入理解蛋白质合成机制的演化、酶的催化特异性、以及蛋白质结构与功能关系有实质性的帮助, 也可为人类的相关研究提供相互印证和理论支撑。本文在电子克隆*BmaaRS*的基础上, 对*BmaaRS*进行了生物信息学分析和进化分析, 期望了解*BmaaRS*基因的数目、种类、结构及起源, 为进一步探讨*BmaaRS*基因的结构和功能提供参考。

1 材料和方法

1.1 序列数据

用于比较分析的aaRS序列数据和家蚕基因组数据以及家蚕EST数据均来源于NCBI (National Center for Biotechnology Information) 数据库 (<http://www.ncbi.nlm.nih.gov/>)。

1.2 方法

1.2.1 电子克隆

选取黑腹果蝇 (*Drosophila melanogaster*), 赤拟谷盗 (*Tribolium castaneum*), 冈比亚按蚊 (*Anopheles gambiae*), 埃及斑蚊 (*Aedes aegypti*), 致倦库蚊 (*Culex pipiens quinquefasciatus*), 豌豆蚜虫 (*Acyrtosiphon pisum*), 丽蝇蛹集金小蜂 (*Nasonia vitripennis*) 等昆虫来源的线粒体 aaRS 和细胞质的 aaRS 序列, 特殊情况选取包括古细菌在内的一些物种的 aaRS 序列, 利用 NCBI 在线 (<http://www.ncbi.nlm.nih.gov/>) 的 tblastn 程序搜寻家蚕基因组 wgs (whole – genome shotgun reads) 数据库序列中潜在的 *BmaaRS* 编码区域, 并根据内含子剪接特点 (GT-AG 法则), 基因的 5'与 3'端非编码序列特点 (5'端序列含 TATA 盒和 CAAT 盒, 3'端序列含加尾信号序列 AATAAA) 以及是否有 EST 序列支持等进行电子克隆, 根据电子克隆获得的 *BmaaRS* 基因的 cDNA 序列, 进一步推测其氨基酸序列。

1.2.2 生物信息学分析

信号肽分析利用在线分析软件 (<http://www.cbs.dtu.dk/services/SignalP/>) 进行; 蛋白质二级和三级结构分析应用Swiss-model在线工具 (<http://swissmodel.expasy.org>) 程序^[13]进行, 首先采用自动分析模式, 在无法获得分析结果时, 再采用选择模板进行分析; 结构域分析通过NCBI在线CD-search进行。

1.2.3 进化分析

选取昆虫来源的氨酰-tRNA合成酶序列, 并分别以人种 (*Homo sapiens*) 来源的各 aaRS 和 *E.coli*

K12 株来源的各 aaRS 序列为外群进行进化分析。由于蛛形纲 (Arachnida) 的肩突硬蜱 (*Ixodes scapularis*) 与昆虫的亲缘关系较近而被选择。除家蚕各 aaRS 序列为本文电子克隆推导得到的外, 其他序列数据来源于 NCBI, 由于 NCBI 中一些 aaRS 的序列不完整, 不利于同源性比较而未采用。对同一功能的不同物种来源的 aaRS 序列通过 DNASTar 软件的 Editseq 程序、Bioedit 软件的 Alignment 程序和 ClustalW Multiple Alignment 程序编辑后, 利用 MEGA3 软件的 Neighbor – Joining (NJ), Minimum Evolution, Maximum Parsimony, UPGMA 等不同程序分别进行进化分析。对 ValRS 和 GlnRS/ Glu-ProRS/ GluRS 进化分析首先利用 NCBI 在线 BlastP 初步获得进化树, 然后选择不同物种 (特别是模式生物) 的 aaRS 进行进化分析。

2 结果与分析

2.1 家蚕 aaRS 基因的电子克隆

电子克隆获得的 BmaaRS 的 cDNA 序列及其对应的家蚕基因组序列见表 1 (所列序列不包含两端非编码序列, 且按基因 ORF 的外显子顺序排列, 涉及可变剪接时仅列出最长 cDNA 序列)。虽然部分 BmaaRS 基因在 GenBank 已有公开序列, 如 BmAlaRS-1 基因 (GenBank 登录号: NP_001037452), BmLysRS 基因 (GenBank 登录号: NP_001040293), 但这些序列未能反映基因的外显子数目及是否存在可变剪接, 本文仍然进行了电子克隆。比对分析发现: 可能由于家蚕品系来源不同, 已公开的这些 BmaaRS 基因的氨基酸序列与我们电子克隆后推导的序列存在少量突变, 本文进化分析时仍采用电子克隆序列。由于基因组序列尚未完善, 部分 BmaaRS 的 cDNA 序列通过 EST 的拼接获得, 如 BmAspRS-1, BmGlnRS-2, BmTyrRS-1 等。除 BmProRS-2 基因外, 其他 BmaaRS 基因都有或多或少的 EST 证据支持。同一基因一般对应几个基因组序列数据, 这些冗余序列中只以 1 个表示。电子克隆表明, BmaaRS 可能存在假基因片段, 其上下游缺乏 BmaaRS 的其他区域的同源区, 且无 EST 证据支持, 如与 ValRS 有同源性的 BAAB01143487 (591–460), BABH01035807 (11 853–11 690, 5 814–5 683) 等可能为 BmaaRS 的假基因片段。

2.2 两套 BmaaRS 基因

分析表明, 大多数 BmaaRS 核基因均有 2 个, 其中 1 个为细胞质编码 (BmaaRS-1), 1 个为线粒体编码 (BmaaRS-2), 但存在例外, BmGlnRS、BmLysRS、BmGlyRS、BmHisRS 和 BmThrRS 均只有 1 个基因; BmGluRS-1 与 BmProRS 融合在一起, 以双功能氨酰 tRNA 合成酶 (bifunctional aminoacyl-tRNA synthetase, glutamyl-prolyl-tRNA synthetase) 基因 (BmGlu-ProRS) 形式出现; BmPheRS 和 BmSerRS 基因则有 3 个。

在 BmPheRS 中, 细胞质 BmPheRS-1 由 α 亚基 (BmPheRS-1 α) 和 β 亚基 (BmPheRS-1 β) 构成, 与其他物种相比, BmPheRS-1 β 基因只发现 1 个而不是多个; 与 *E.coli* 相比, 家蚕中无论是细胞质还是线粒体的 BmGlyRS 均无 α 亚基和 β 亚基之分。

家蚕和野桑蚕 (*Bombyx mandarina*) 同属鳞翅目 (Lepidoptera) 蚕蛾科 (Bombycidae) 昆虫, 至今已完成蚕蛾科 mtDNA 基因组全序列测定的包括 7 个株系, 这些线粒体基因组序列中均存在 2 个 tRNA^{Leu} 基因和 2 个 tRNA^{Ser} 基因, 其他 tRNA 基因均只有 1 个。值得注意的是线粒体基因组中虽然有 2 个 tRNA^{Leu} 基因, 但核基因组中只有 1 个 BmLeuRS-2 基因, 表明线粒体中只需要 1 种 BmLeuRS 就可为 2 种不同的 tRNA^{Leu} 分子添加氨基酸。而线粒体基因组中有 2 个 tRNA^{Ser} 基因, 核基因组中也有 2 个编码线粒体 BmSerRS 的基因 (BmSerRS-2a, BmSerRS-2b), 表明线粒体中可能分别需要不同 BmSerRS 蛋白为不同 tRNA^{Ser} 分子添加氨基酸。

2.3 BmaaRS 基因的可变剪接

对家蚕 EST 数据库分析表明, BmaaRS 基因存在可变剪接, 这些可变剪接一方面可能导致一种 BmaaRS 基因分别编码细胞质和线粒体 BmaaRS, 如上述的 BmLysRS 基因和 BmHisRS 基因; 另一方面, BmaaRS 基因的可变剪接可能具有基因表达调控等功能。BmaaRS 基因的可变剪接主要发生在 5' 端, 但也有发生在基因中间区域以及基因的 3' 端。

分析显示: (1) 可变剪接可能导致翻译提前终止。如 BmGluRS-1 基因、BmHisRS 基因和 BmTrpRS-2

基因。具有BmHisRS功能的真核翻译启动因子2 α 激酶基因的可变剪接也导致翻译提前终止；(2) 可变剪接可能导致翻译起始速率的改变。如BmTyrRS-1的基因，其剪接方式为5'端非编码序列的可变剪接，不影响ORF编码序列；(3) 可变剪接可能导致编码蛋白改变。上述BmGluRS-1、BmHisRS、BmTrpRS-2基因的可变剪接可能导致读码框改变，或编码蛋白的序列缺失。又如在BmGluRS-1的基因中，5'端可变剪接导致编码蛋白的N端发生缺失；(4) 可能转录起始不同导致可变剪接。如BmAspRS-1的可变剪接导致第1外显子的改变，由于该基因的起始码位于第2外显子，因而不影响BmAspRS-1的读码框；(5) BmaaRS基因中间区域以及3'端的可变剪接。如BmAsnRS-1的3'端可变剪接导致加尾位点改变，BmGluRS-1基因中间区域的可变剪接导致多个外显子的缺失。

由部分EST序列的分析结果还显示，有些BmaaRS基因可能与其他基因重叠，例如在BmTrpRS-2基因编码区的互补链编码其他基因，其重叠发生在2个cDNA基因的3'端。分析表明BmaaRS基因与其他基因的重叠现象比较罕见，BmaaRS基因可能并不存在反式剪接机制。

2.4 BmaaRS 基因的序列扩增

比较不同昆虫物种的aaRS序列，可以发现BmaaRS基因中存在序列扩增现象。例如BmMetRS-1基因中编码C端的2个外显子是同源扩增形成的(24个ORF外显子中的第22, 23个外显子)，在包含外显子和部分内含子的2段基因组序列(BABH01031386: 6 164–8 538, 7 674–8 048, 分别含ORF的第22个外显子和第23个外显子)均带有末端顺向重复序列cgttctc，两段序列的一致性(Identity)高达96.8% (375 bp中有12 bp不同)，EST证据显示这2个外显子同时出现在成熟mRNA中；该区域的增加将导致编码蛋白的C端结构域从原来的2个HLH (helix-loop-helix) 结构增加为3个。由于同源性比较主要在家蚕和其他昆虫之间，数据相对较少，因此难以发现进化过程中其他结构域的扩增与缺失现象。

2.5 BmaaRS 的三级结构分析

应用Swiss-model在线工具对BmaaRS蛋白质进行三级结构分析，其中BmAsnRS-1、BmAspRS-2、BmGluRS-2、BmHisRS-1、BmHisRS-2、BmLeuRS-2、BmThrRS-2、BmTrpRS-1、BmTrpRS-2、BmTyrRS-1、BmTyrRS-2、BmValRS-1采用选择模板进行分析。BmaaRS与模板比较获得的部分区域三级结构如图1所示。同一功能的BmaaRS的三级结构相似性很高，如细胞质蛋白BmSerRS-1与线粒体蛋白BmSerRS-2a和BmSerRS-2b，但其他兼具BmaaRS功能的蛋白结构与对应的BmaaRS的结构相差较大，如DNA polymerase accessory subunit (BmGlyRS-2) 与对应的BmGlyRS-1，线粒体核糖体蛋白L39 (BmThrRS-2) 与对应的BmThrRS-1。BmHisRS-2由于缺乏相应的结构模板，只能得到区域323-501 aa的三级结构图，同样，BmThrRS-2由于缺乏相应的结构模板而只能得到79-132 aa (以1wwtA为模板) 和209-266 aa (以1nyqB为模板) 区域的三级结构图。BmGlu-ProRS分别显示的是与GluRS和ProRS 2个模板对应的三级结构。

2.6 BmaaRS 的结构域分析

利用在线CD-search对BmaaRS的结构域进行分析，各BmaaRS的主要结构域如表2所示。总体上，与其他物种aaRS的结构域相同，属于I类的BmaaRS具有由“HIGH”和“KMSKS”两个标签序列构成的Rossmann折叠的催化核心结构域，II类BmaaRS具有被 α 螺旋所环绕的反平行的 β 折叠结构^[14]。属II类的各BmaaRS均存在多个不同类型的结构域，而I类除了BmGlu-ProRS外，各BmaaRS的结构域相对较少。各BmaaRS还存在各自的特殊结构域，如BmHisRS-2中，有类似OB-折叠类型的核酸结合结构域(pfam01336)，G、H、P、S和T结构域(pfam00587)，D、K和N结构域(cd04320)等等，部分BmaaRS间还具有相同或相似的结构域。

2.7 BmaaRS 的进化分析

对同一功能的不同物种来源的aaRS序列进行进化分析。由于BmGluRS和BmGlu-ProRS以及BmProRS和BmGlu-ProRS的功能关系，我们将BmGlu-ProRS分别与BmGluRS和BmProRS进行聚类分析。结果(图略)显示，不同算法得到的进化树具有相同或非常相似的拓扑结构，细胞质和线粒体的aaRS分别聚为1组(可能由单一基因编码的细胞质和线粒体LysRS例外)。(1) AlaRS-2的进化树

表现较为特殊,这可能与AlaRS的进化模式相关,研究认为细胞质中的AlaRS是由真核生物线粒体中AlaRS基因进化而来的^[15]; (2) 细胞质aaRS和线粒体aaRS均分在不同组,表明其起源不同,但也存在极少数序列不完全对应,如人体虱(*Pediculus humanus corporis*)的AlaRS-2,致倦库蚊的GluRS-2,肩突硬蜱的ValRS-2等,并未聚类在线粒体aaRS组; (3) 以人类或*E.coli*的aaRS蛋白序列作外群并不理想,只有在GlyRS和ValRS的聚类中*E.coli*的aaRS表现为外群。由于*E.coli*的aaRS不在线粒体aaRS蛋白,但聚类分析表明,*E.coli*的不同aaRS可以或者与细胞质aaRS蛋白聚类,或者与线粒体aaRS蛋白序列聚类,这也可能反映了细胞质aaRS蛋白和线粒体aaRS蛋白的起源; (4) 在PheRS的进化分析中,细胞质PheRS蛋白的 α 亚基与线粒体PheRS蛋白的亲缘关系较近,可能表明了线粒体PheRS蛋白的起源; (5) 在SerRS的进化分析中,线粒体SerRS确实存在2类不同蛋白,且这2类蛋白与细胞质SerRS蛋白相比也反映了起源的不同; (6) 以aaRS了解家蚕的起源并不理想,家蚕与其他昆虫的亲缘关系随不同aaRS的比较而发生变化。

在古老的线粒体基因组中,编码aaRS的基因可能经历了基因丢失过程,少数aaRS基因转移到了核内,而大多数则发生了丢失,其功能被古细菌起源的aaRS基因所取代^[16]。例如,细胞核编码的缬氨酸tRNA合成酶(ValRS)被认为是具有线粒体起源的酶类,而编码线粒体中的ValRS基因很可能就是被古细菌的ValRS基因所取代^[17]。据此,我们选择不同物种(特别是模式生物)的细胞质和线粒体ValRS进行进化分析,其中昆虫只选择了家蚕和黑腹果蝇的ValRS。NJ分析结果如图2所示,秀丽隐杆线虫(*Caenorhabditis elegans*)的vrs-1编码蛋白(GenBank登录号: NP_505070)和vrs-2基因编码蛋白(GenBank登录号: NP_493377)较为特殊,其他物种细胞质和线粒体ValRS不能聚为2组,而是交叉聚类,如线粒体蛋白中,斑马鱼(*Danio rerio*, GenBank登录号: XP_686760)、小鼠(*Mus musculus*, GenBank登录号: NP_780346)和人类(GenBank登录号: NP_065175)聚为1个亚组,而与黑腹果蝇(GenBank登录号: NP_648268)和家蚕BmValRS-2分开聚类,酿酒酵母(*Saccharomyces cerevisiae*)中ValRS蛋白(GenBank登录号: NP_011608)兼具细胞质和线粒体ValRS蛋白功能,但表现较为原始,与 α -变形杆菌纲的甲烷氧化菌(*Methylobacterium nodulans*, GenBank登录号: ACL61266)和 γ -变形杆菌的大肠埃希菌(*Escherichia coli*, GenBank登录号: NP_418679)亲缘关系较近。这一聚类分析不能反映出ValRS是否起源于线粒体基因组,以及线粒体中的ValRS与古细菌的ValRS关系,与文献[10]结果一致,但根据BmValRS与其他物种ValRS的关系,表明家蚕中aaRS为2套不同起源的基因。

通过GlnRS和GluRS的电子克隆,我们发现家蚕中只存在细胞质的GlnRS(BmGlnRS-1)和双功能aaRS的BmGlu-ProRS以及线粒体的GluRS(BmGluRS-2),BmGluRS-2似乎兼具了线粒体的BmGlnRS和BmGluRS功能,支持内共生假说^[18]。由此而选择不同物种(特别是模式生物)的细胞质GluRS(Glu-ProRS)与细菌的GlnRS进行进化分析,以了解家蚕中细胞质的BmGlnRS和BmGlu-ProRS与细菌的GlnRS的进化关系,NJ分析结果如图3。由于链孢霉中某种蛋白(GenBank登录号: XP_956785)为假设蛋白,注释其具有GluRS功能,进化分析表明该蛋白可能不是专一的GluRS。图中可见,包括 α -变形菌纲(Alphaproteobacteria)、 β -变形菌纲(Betaproteobacteria)、 γ -变形菌纲(Gammaproteobacteria)、 δ -变形菌纲(Deltaproteobacteria)、 ϵ -变形菌纲(Epsilonproteobacteria)在内的不同细菌的GlnRS和GluRS分别聚类在一起,并分别与真核生物细胞质GlnRS和线粒体GluRS聚类为2组,真核生物Glu-ProRS单独聚为1组,反映了原核生物的GlnRS和真核生物的GlnRS起源相同,GluRS的起源也是如此。

3 讨论

由于aaRS参与多种生物学功能,而线粒体氨酰-tRNA合成酶与人类的疾病密切相关,因此对aaRS的研究在理论上和应用上都有重要意义。从昆虫到人类等的某些高等真核生物的aaRS都具有一个多酶复合体,电子克隆表明,家蚕除了具有各种细胞质和线粒体的BmaaRS外,也具有形成BmaaRS多酶复合体的辅助蛋白成员的基因,且均有EST证据支持其表达(本文未列出)。与人类LysRS基因相似,BmLysRS基因可以形成2种不同cDNA,分别产生细胞质和线粒体的BmLysRS。因此家蚕aaRS

可能与人类等高等生物一样, 在各方面具有相似性。家蚕作为生物学研究材料具有独特的优越性, 并且随着家蚕后基因组时代的到来, 对家蚕分子生物学的深入研究也必将为人类相关研究提供理论依据和相互印证的可能。

对电子克隆只发现单个 *BmaaRS* 基因可解释为: (1) 线粒体的 *BmGlnRS* 和 *BmGluRS* 可能只需 1 个基因完成其功能, 因为在某些原核生物中, *GlnRS* 的功能则由 *GluRS* 来替代^[18]; (2) 1 个基因的可变剪接形成 2 种不同剪接体, 翻译后形成 2 种 *BmaaRS*。如 EST 证据表明 *BmLysRS* 基因的 5' 端存在可变剪接, 因此可能与人类的 *LysRS* 基因相似^[19]。又如 EST 证据显示 *BmHisRS* 基因可以发生可变剪接, 可能如酿酒酵母一样^[20], 1 个 *BmHisRS* 基因分别为细胞质和线粒体共用; (3) 由其他功能蛋白来行使 *BmaaRS* 的功能。如 DNA polymerase accessory subunit 基因, 具有 *BmGlyRS-2* 功能; (4) 可变剪接与其他功能蛋白相互协同行使 *BmaaRS* 的功能。如 EST 证据显示 *BmHisRS* 基因可以发生可变剪接, 可能如酿酒酵母一样^[20], 1 个 *BmHisRS* 基因分别为细胞质和线粒体共用, 同时核基因组中还存在一个具有 *HisRS* 功能的真核翻译启动因子 2 α 激酶基因, 具有 *BmHisRS-1* 功能。又如在 *BmThrRS* 基因中存在可变剪接, 不能排除 *BmThrRS* 基因编码具有两种多肽的功能, 而在家蚕核基因组中还存在有 *BmThrRS* 功能的线粒体核糖体蛋白 L39 基因。

某些原核生物中 *GlnRS* 的功能由 *GluRS* 来替代, 这类 *GluRS* 被称为“无辨别力”的谷氨酰-tRNA 合成酶 (Non-discriminating glutamyl-tRNA synthetase, *GluRSnd*), 在原核生物中 *Gln-tRNA^{Gln}* 不能直接合成, 而是通过“误氨酰化”催化谷氨酸与 tRNA^{Gln} 连接生成 *Glu-tRNA^{Gln}*, 然后在 *Glu-tRNA^{Gln}* 酰氨转移酶的催化下, 以谷氨酰氨为供体, 最后生成 *Gln-tRNA^{Gln}*。在家蚕中只有编码线粒体的 *BmGluRS-2*, 而无编码线粒体的 *BmGlnRS-2*。NCBI 在线 blastn 分析表明, 与其他线粒体的 *GluRS* 一样, *BmGluRS-2* 既具有辨别 *GluRS* 核心催化结构域, 又具有“无辨别力” *GluRS* 核心催化结构域, 因此而推测 *BmGluRS-2* 具有 *GlnRS* 的功能, 但线粒体中 *BmGluRS-2* 是否真正具有 *GlnRS* 的功能尚不清楚。对 *GlnRS* 和 *GluRS* 进行的进化分析中, 所分析的所有细菌的 *GlnRS* 中均含有“无辨别力” *GluRS* 核心催化结构域, 而在所分析的所有细菌的 *GluRS* 中均含有辨别 *GluRS* 核心催化结构域, 具有辨别 *GluRS* 的核心催化结构域只存在于某系细菌和细胞器中。基因复制导致了细菌和真核生物中具有不同催化专一性的 aaRS 的形成, 如 *AsnRS* 可能是由真核生物中的 *AspRS* 通过基因水平转移至某些细菌和古细菌中^[21], 而某些细菌中的 *GlnRS* 则被认为是由真核生物中的 *GluRS* 通过基因复制和基因水平转移进化而来, 并由真核生物通过基因水平转移到这类细菌之中^[22,23]。本文对 *GlnRS* 和 *GluRS* 进行的进化分析中, 由于分析的细菌类型不多, 未能反映基因的水平转移, 但结合 *ValRS* 的进化分析进一步表明家蚕中 aaRS 确实为 2 套不同起源的基因。

aaRS 被认为是理想的药物设计的靶点^[24], 作用于 *LeuRS* 的编校结构域的含硼化合物已用于抑制真菌的生长, 成为外用抗菌素^[25]。真核生物的 aaRS 较原核生物的 aaRS 具有明显的高度专一性, 而家蚕为鳞翅目昆虫的模式生物, 与哺乳动物细胞一样, 家蚕细胞中也存在着两套蛋白质翻译系统, 分别存在于细胞质中和线粒体中, 两套翻译系统均有自己相对独立的一套 aaRS 及 tRNA。从进化的角度而言, 线粒体 aaRS 与原核生物 aaRS 相似度较高, 而相对细胞质 aaRS 不保守, 这一现象与线粒体的内共生起源假说也是相符的。这种翻译系统的二重性使得很多以原核生物 aaRS 为靶点设计的药物具有线粒体毒性, 从而给这类药物的设计与筛选增加了困难^[26]。因此利用 *BmaaRS* 的抑制剂研究, 将可能为筛选以原核生物 aaRS 为靶点设计的药物在真核生物毒理学研究方面带来方便, 同时也可能为鳞翅目昆虫害虫的防治提供依据。

参考文献 (References):

1. Martinis SA, Plateau P, Cavarelli J, Florentz C. Aminoacyl-tRNA synthetases: a family of expanding functions. *EMBO J*, 1999, 18(17): 4591-4596
2. 贾捷, 金由辛. 氨酰tRNA合成酶的分子网络和功能. 生物化学与生物物理进展, 2004, 31(4): 291-295.
3. Andreoli C, Prokisch H, Hörtnagel K, Mueller JC, Münsterkötter M, Scharfe C, Meitinger T. MitoP2, an integrated

- database on mitochondrial proteins in yeast and man. *Nucleic Acids Res*, 2004, 32: D459–D462.
4. Hengartner MO. The biochemistry of apoptosis. *Nature*, 2000, 407(6805): 770–776.
 5. Taylor RW, Turnbull DM. Mitochondrial DNA mutations in human disease. *Nat Rev Genet*, 2005, 6(5): 389–402.
 6. 周小龙, 王恩多. 与人类疾病相关的几种线粒体氨基酰-tRNA 合成酶. *生物化学与生物物理进展*, 2008, 35(8): 853–858.
 7. Eriani G, Delarue M, Poch O, Gangloff J, Moras D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, 1990, 347(6289): 203–206.
 8. Curnow AW, Ibba M, Söll D. tRNA-dependent asparagine formation. *Nature*, 1996, 382(6592): 589–590.
 9. Ibba M, Söll D. Aminoacyl-tRNA synthesis. *Annu Rev Biochem*, 2000, 69: 617–650.
 10. Brindefalk B, Viklund J, Larsson D, Thollessen M, Andersson SG. Origin and evolution of the mitochondrial aminoacyl-tRNA synthetases. *Mol Biol and Evol*, 2007, 24(3): 743–756.
 11. Chang PK, Dignam JD. Primary structure of alanyl-tRNA synthetase and the regulation of its mRNA levels in *Bombyx mori*. *J Biol Chem*, 1990, 265(34): 20898–20906.
 12. Nada S, Chang PK, Dignam JD. Primary structure of the gene for glycyl-tRNA synthetase from *Bombyx mori*. *J Biol Chem*, 1993, 268(11): 7660–7667.
 13. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 2006, 22(2): 195–201.
 14. Martinis SA, Plateau P, Cavarelli J, Florentz C. Aminoacyl-tRNA synthetases: a new image for a classical family. *Biochimie*, 1999, 81(7): 683–700.
 15. Chihade JW, Brown JR, Schimmel PR, Ribas De Pouplana L. Origin of mitochondria in relation to evolutionary history of eukaryotic alanyl-tRNA synthetase. *Proc Natl Acad Sci USA*, 2000, 97(22): 12153–12157.
 16. Ribas de Pouplana L, Schimmel P. A view into the origin of life: aminoacyl-tRNA synthetases. *Cell Mol Life Sci*, 2000, 57(6): 865–870.
 17. Hashimoto T, Sánchez LB, Shirakura T, Müller M, Hasegawa M. Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. *Proc Natl Acad Sci USA*, 1998, 95(12): 6860–6865.
 18. 林军, 黄京飞. 氨基酰-tRNA 合成酶专一性识别的进化. *生物化学与生物物理进展*, 2002, 29(6): 837–841.
 19. Tolkunova E, Park H, Xia J, King MP, Davidson E. The human lysyl-tRNA synthetase gene encodes both the cytoplasmic and mitochondrial enzymes by means of an unusual alternative splicing of the primary transcript. *J Biol Chem*, 2000, 275(45): 35063–35069.
 20. Natsoulis G, Hilger F, Fink GR. The HTS1 gene encodes both the cytoplasmic and mitochondrial histidine tRNA synthetases of *S. cerevisiae*. *Cell*, 1986, 46(2): 235–243.
 21. Shiba K, Motegi H, Yoshida M, Noda T. Human asparainyl-tRNA synthetase: molecular cloning and the inference of the evolutionary history of Asx-tRNA synthetase family. *Nucleic Acids Res*, 1998, 26(22): 5045–5051.
 22. Brown JR, Doolittle WF. Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutamyl-tRNA synthetases. *J Mol Evol*, 1999, 49(4): 485–495.
 23. Woese CR, Olsen GJ, Ibba M, Söll D. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev*, 2000, 64(1): 202–236.
 24. Kim S, Lee SW, Choi EC, Choi SY. Aminoacyl-tRNA synthetases and their inhibitors as a novel family of antibiotics. *Appl Microbiol Biotechnol*, 2003, 61(4): 278–288.
 25. Rock FL, Mao W, Yaremchuk A, Tukalo M, Crépin T, Zhou H, Zhang YK, Hernandez V, Akama T, Baker SJ, Plattner JJ, Shapiro L, Martinis SA, Benkovic SJ, Cusack S, Alley MR. An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing site. *Science*, 2007, 316(5832): 1759–1761.
 26. 王恩多. 哺乳动物氨基酰-tRNA 合成酶的研究. *生命科学*, 2006, 18(3): 219–203.

表 1 电子克隆氨酰-tRNA 合成酶基因序列来源

| 基因名称 | ORF 序列来源 (外显子的连接区域) |
|--------------------|---|
| <i>BmAlaRS-1</i> | BABH01005062 (6 860–9 763) |
| <i>BmAlaRS-2</i> | AADK01008277 (9 819–9 998, 11 032–11 205, 11 672–11 811, 12 337–12 492, 12 738–12 882), BABH01025225 (290–438, 1 331–1 634, 2 004–2 806, 3 926–4 079) |
| <i>BmArgRS-1</i> | BABH01012472 (22 435–20 384) |
| <i>BmArgRS-2</i> | BABH01001356 (6 454–5 383, 5 166–4 667) |
| <i>BmAsnRS-1</i> | BABH01038527 (5 498–5 525, 5 918–6 265, 6 815–7 035, 7 143–7 261, 8 592–8 765, 9 375–9 503), BABH01038528 (947–1 091, 2 075–2 179, 2 716–2 926, 4 292–4 476) |
| <i>BmAsnRS-2</i> | BABH01027222 (8 430–8 368, 6 938–5 934, 5 676–5 419) |
| <i>BmAspRS-1</i> | AADK01020556 (2 053–2 167, 2 897–3 060, 5 219–5 354), BABH01017798 (2 360–2 537, 3 401–3 545), AADK01014086 (6 561–6 410, 5 263–5 097), BABH01017800 (1 841–1 956, 3 397–3 435) + EST: CK516452 (379–419) + BABH01017801 (1–78, 1 254–1 365), BABH01017802 (764–823, 1 188–1 291) |
| <i>BmAspRS-2</i> | AADK01004194 (2 638–4524) |
| <i>BmCysRS-1</i> | BABH01001027 (3 693–3 874, 4 026–4 292), BABH01001029 (688–904, 1 142–1 352, 2 113–2 345, 3 697–3 930), BABH01001030 (2 806–2 941, 4 748–4 953, 5 703–5 865, 6 951–7 147), AADK01047617 (389–496) |
| <i>BmCysRS-2</i> | AADK01014672 (3 981–5 516) |
| <i>BmGlnRS-1</i> | BABH01000834 (13 827–13 943, 14 035–14 433, 15 022–15 235, 15 697–16 062, 16 132–16 322, 16 476–16 705, 16 795–16 931), AADK01036179 (2 585–2 733) + BABH01000835 (1–23, 1 515–1 752, 2 186–2 331, 3 145–3 334, 4 660–4 707) |
| <i>BmGlu-ProRS</i> | BABH01019736 (10 864–10 903, 10 991–11 327, 11 428–11 571, 11 877–11 993, 12 218–12 344, 12 740–12 932, 13 149–13 318, 14 074–14 232, 14 888–14 962, 14 285–15 429, 16 791–17 038, 17 333–17 544), BABH01019737 (1 576–1 784), BABH01019741 (1 281–1 484, 3 574–3 777, 4 885–5 085, 6 269–6 427, 7 779–7 910, 8 595–8 727, 9 676–9 824, 10 329–10 529, 11 408–11 525, 12 046–12 185, 13 098–13 224, 13 684–13 925, 14 486–14 646, 14 938–15 081, 15 340–15 505) |
| <i>BmGluRS-2</i> | BABH01024010 (8 480–8 826, 9 124–9 319, 9 788–10 002, 10 693–10 808, 10 909–11 018, 11 370–11 903) |
| <i>BmGlyRS-1</i> | BABH01003232 (16 809–14 611) |
| <i>BmGlyRS-2</i> | DNA polymerase accessory subunit BABH01034689 (1 139–1 528, 2 157–2 291, 2 503–2 659, 3 793–3 905, 4 679–4 783, 6 274–6 330) |
| <i>BmHisRS-1</i> | eukaryotic translation initiation factor 2-alpha kinase BABH01004455 (934–874), BABH01004453 (4 271–2 932, 2 209–1 874) |
| <i>BmHisRS-2</i> | BABH01000658 (1 761–1 669, 614–402), EST: AU004124 (286–507), BABH01000657 (20 667–20 461, 19 854–19 560, 18 440–18 340, 17 954–17 772, 16 848–16 615) |
| <i>BmIleRS-1</i> | BABH01006370 (360–487, 1 348–1 453, 1 595–1 796, 1 868–5 070) |
| <i>BmIleRS-2</i> | AADK01006474 (2 703–2 891, 3 034–3 156, 3 362–3 723, 5 168–5 281, 6 142–6 341, 6 744–6 914, 7 770–7 999, 9 544–9 680, 11 026–11 158, 11 915–12 069), BABH01035807 (5 779–5 639, 4 757–3 899) |
| <i>BmLeuRS-1</i> | BABH01008012 (18 495–18 403, 18 161–17 712, 17 438–16 094, 15 992–15 754, 15 231–14 995, 13 996–13 781) |
| <i>BmLeuRS-2</i> | BABH01000422 (680–678), BAAB01110645 (914–708), AADK01012972 (3 767–4 103, 4 338–4 578, 4 885–5 783, 6 536–6 794, 6 877–7 241), BAAB01087117 (1 984–1 716, |

| | |
|-------------------------------------|--|
| | 999-745), BABH01000420 (5 606-5 397, 4 484-4 250, 2 385-2 177) |
| <i>BmLysRS-1</i> | BABH01010489 (21 366-21 332, 21 154-20 782, 19 834-19 583, 18 555-18 163, 18 064-17 882, 17 113-16 953, 16 368-16 128, 16 039-15 935) |
| <i>BmLysRS-2</i> | BABH01010489 (21 309-21 221, 21 154-20 782, 19 834-19 583, 18 555-18 163, 18 064-17 882, 17 113-16 953, 16 368-16 128, 16 039-15 935) |
| <i>BmMetRS-1</i> | BABH01031383 (17 063-17 165, 19 000-19 127, 20 623-20 773, 21 374-21 521, 21 766-21 863, 23 362-23 484, 24 034-24 150, 24 943-25 018), BABH01031384 (649-776), BABH01031385 (898-1 008, 1 676-1 853, 3 149-3 364, 4 143-4 244, 5 655-5 772, 6 599-6 669, 7 231-7 328), BABH01031386 (2 073-2 221, 2 873-3 005, 4 122-4 265, 4 785-4 923, 5 586-5 639, 6 182-6 364, 7 692-7 874, 8 278-8 394) |
| <i>BmMetRS-2</i> | BABH01016667 (8 279-8 363, 9 718-9 816, 10 291-10 472, 10 951-11 074), BABH01016668 (2 020-2 121, 2 682-2 800, 4 315-4 466, 5 185-5 394), BABH01016669 (90-262, 737-1 107, 1 340-1 348) |
| <i>BmPheRS-1α</i> | BABH01052419 (79-594), AADK01009324 (780-1 134, 1 766-2 043, 4 002-4 142), BABH01044388 (782-989, 1 814-2 034) |
| <i>BmPheRS-1β</i> | BABH01008991 (615-558, 484-429, 247-93), BABH01008990 (7 538-7 382, 5 756-5 580, 5 309-5 130, 5 590-4 415, 3 910-3 761, 2 352-2 220, 1 198-988), BABH01008989 (4 591-4 445, 2 735-2 584) |
| <i>BmPheRS-2</i> | BABH01032560 (513-1 031, 1 633-2 277, 3 216-3 383) |
| <i>BmProRS-1</i> | glutamyl-prolyl tRNA synthetase, bifunctional aminoacyl tRNA synthetase |
| <i>BmProRS-2</i> | BABH01036404 (16 229-16 325, 17 918-18 159, 19 262-19 352, 20 736-20 906, 21 688-21 912, 23 490-23 618, 23 902-23 998), BABH01036405 (792-944) |
| <i>BmSerRS-1</i> | BABH01021731 (7 419-7 099, 6 561-6 345, 6 254-5 094, 4 884-4 703, 3 599-3 509, 2 733-2 477, 2 265-2 139, 1 681-1 553) |
| <i>BmSerRS-2a</i> | BABH01003440 (29 341-30 703) |
| <i>BmSerRS-2b</i> | BABH01013633 (21 883-22 491), EST: CK489569 (320-328) + BABH01013634 (1-179, 393-560, 1 040-1 358) |
| <i>BmThrRS-1</i> | BABH01018235 (7 249-7 199, 5 596-5 352, 4 558-4 447, 3 948-3 871, 3 590-3 333, 2 657-2 511, 1 920-1 811, 1 145-990), AADK01003314 (2 773-2 950, 3 413-3 535, 4 246-4 424, 5 519-5 693, 6 569-6 683, 7 882-8 036) |
| <i>BmThrRS-2</i> | mitochondrial ribosomal protein L39 AADK01003988 (3 347-3 269, 3 186-3 007, 2 817-2 717), BABH01035297 (14 707-14 566, 13 903-13 685, 12 944-12 831, 11 150-11 029) |
| <i>BmTrpRS-1</i> | BABH01019781 (3 510-3 621), BABH01019782 (2 840-2 908, 3 069-3 228, 3 562-3 744, 3 910-4 087, 4 797-5 003, 6 129-6 260, 7 427-7 588) |
| <i>BmTrpRS-2</i> | AADK01002137 (15 720-15 786, 16 213-16 408, 16 912-17 151, 18 155-18 299, 19 026-19 177, 20 579-20 917) |
| <i>BmTyrRS-1</i> | BABH01006157 (32 517-33 128) + AADK01047046 (237-1 201) |
| <i>BmTyrRS-2</i> | BABH01041812 (3 435-3 562, 3 888-4 035, 4 231-4 409, 4 490-4 705, 4 840-4 918, 5 415-5 690, 6 096-6 284, 6 590-6 753) |
| <i>BmValRS-1</i> | BABH01026625 (5 366-5 205, 4 243-4 076, 3 689-3 556, 3 066-2 908, 2 375-2 114, 1 392-1 262, 792-695), BABH01026624 (11 739-11 615, 10 586-10 441, 9 960-9 765, 7 960-7 779, 6 811-6 605, 6 211-5 984, 5 481-5 340, 4 222-4 052, 3 057-2 863, 1 844-1 560, 880-665) |
| <i>BmValRS-2</i> | BABH01024198 (6 831-7 077, 7 472-7 581, 7 671-8 161, 8 872-10 993) |

表 2 BmaaRS 结构域

| I 类 BmaaRS | CCD 编号 | II 类 BmaaRS | CCD 编号 |
|-------------|--|--------------------|--|
| BmArgRS-1 | cd00671; pfam05746; pfam03485; PRK01611 | BmAlaRS-1 | cd00673; pfam07973; pfam02272 |
| BmArgRS-2 | cl00015; pfam05746 | BmAlaRS-2 | cd00673; PRK00252 |
| BmCysRS-1 | cd00672; cl00015; COG0215 | BmAsnRS-1 | cd0077; cd04323; COG0017 |
| BmCysRS-2 | cd00672; cl00015; COG0215 | BmAsnRS-2 | cd00776; cd04318; PRK03932 |
| BmGlnRS-1 | cd00807; pfam03950; pfam04558; PRK05347 | BmAspRS-1 | cd00776; cd04320; PRK05159 |
| BmGlu-ProRS | cd00778; cd00807; cd00862; cl00015; cd00936; pfam03950; cl00349; cl02776; PRK08661; pfam00749 | BmAspRS-2 | cl00268; cd04317; PRK00476 |
| BmGluRS-2 | cd00808; PRK01406 | BmGlyRS-1 | cd00774; cd00858; cd00935; cl00268; COG0423 |
| BmIleRS-1 | cd00818; pfam08264; PRK06039 | BmGlyRS-2 | cl00266 |
| BmIleRS-2 | cl00015; pfam08264; PRK05743 | BmHisRS-1 | cl09925 |
| BmLeuRS-1 | cd00812; cl00015; PRK00390 | BmHisRS-2 | cd00773; cd00859; COG0124 |
| BmLeuRS-2 | cd00812; cl00015; pfam08264; PRK12300 | BmLysRS-1 | cd00775; cd04322; COG1190 |
| BmMetRS-1 | cd00814; cd00939; cd00299; cl00349; PRK12268 | BmLysRS-2 | cd00775; cd04322; COG1190 |
| BmMetRS-2 | cd00814; PRK11893 | BmPheRS-1 α | cd00496; PRK04172 |
| BmTrpRS-1 | cd00806 | BmPheRS-1 β | cd00769; cl01340; pfam03484; PRK09616 |
| BmTrpRS-2 | cd00806; PRK12282 | BmPheRS-2 | cd00496; pfam03147 |
| BmTyrRS-1 | cd00805; cd02799 | BmGlu-ProRS | cd00778; cd00807; cd00862; cl00015; cd00936; pfam03950; cl00349; cl02776; PRK08661; pfam00749 |
| BmTyrRS-2 | cd00805; PRK05912 | BmProRS-2 | cl00268; PRK09194 |
| BmValRS-1 | cd00817; cl00015; pfam08264; PRK05729 | BmSerRS-1 | cd00770; pfam02403; PRK05431 |
| BmValRS-2 | cd00817; cl00015; pfam08264; PRK05729 | BmSerRS-2a | cl00268; pfam02403; COG0172 |
| | | BmSerRS-2b | cl00268 |
| | | BmThrRS-1 | cd00771; cd00860; cd01667; pfam07973; PRK00413 |
| | | BmThrRS-2 | cd01667; PRK00413 |

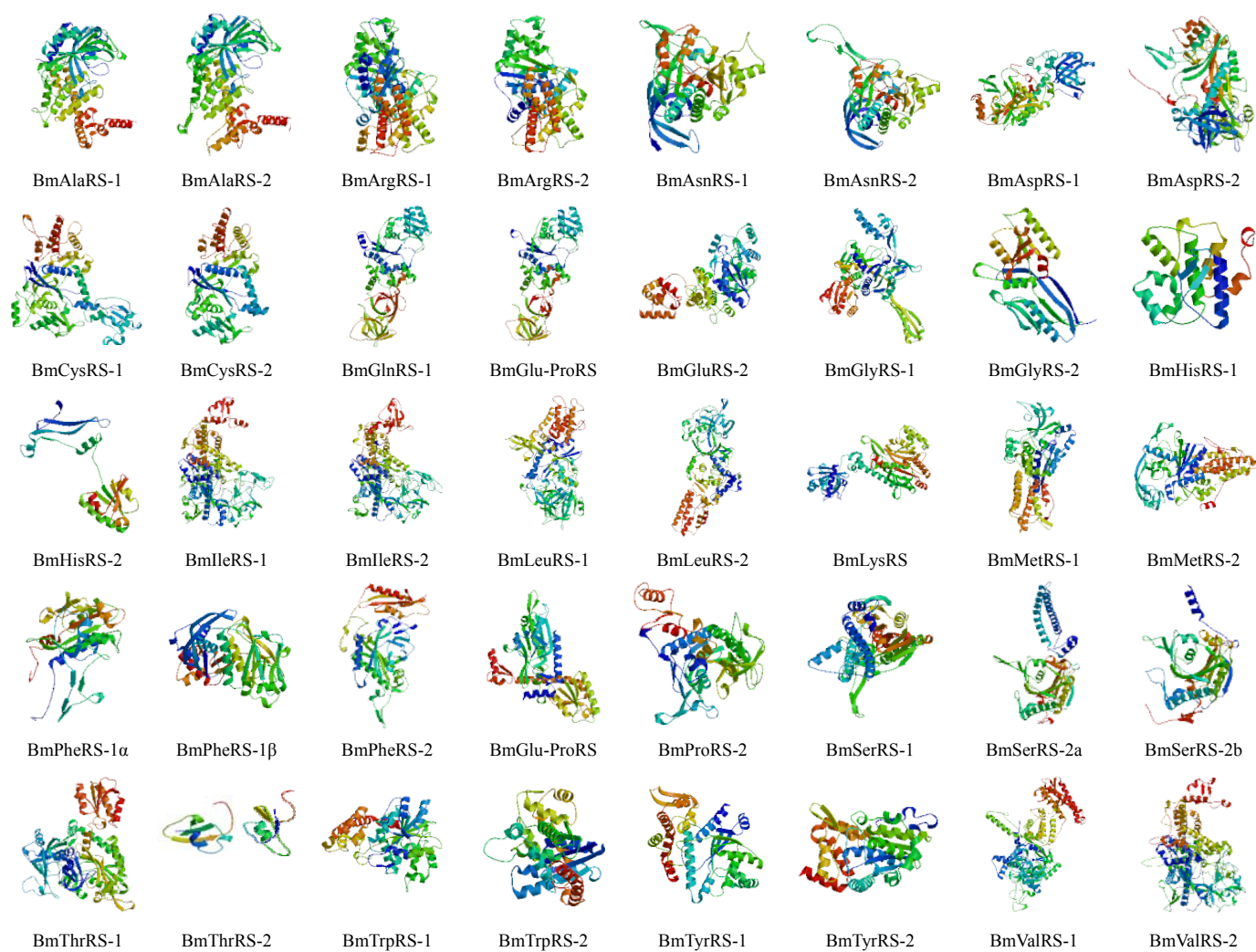


图 1 BmaRS 三级结构预测

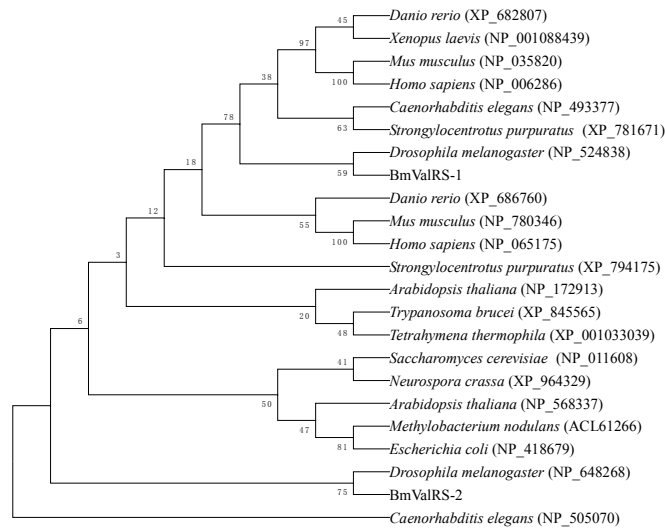


图 2 ValRS 的聚类分析

括号中数据为 GenBank 登录号。

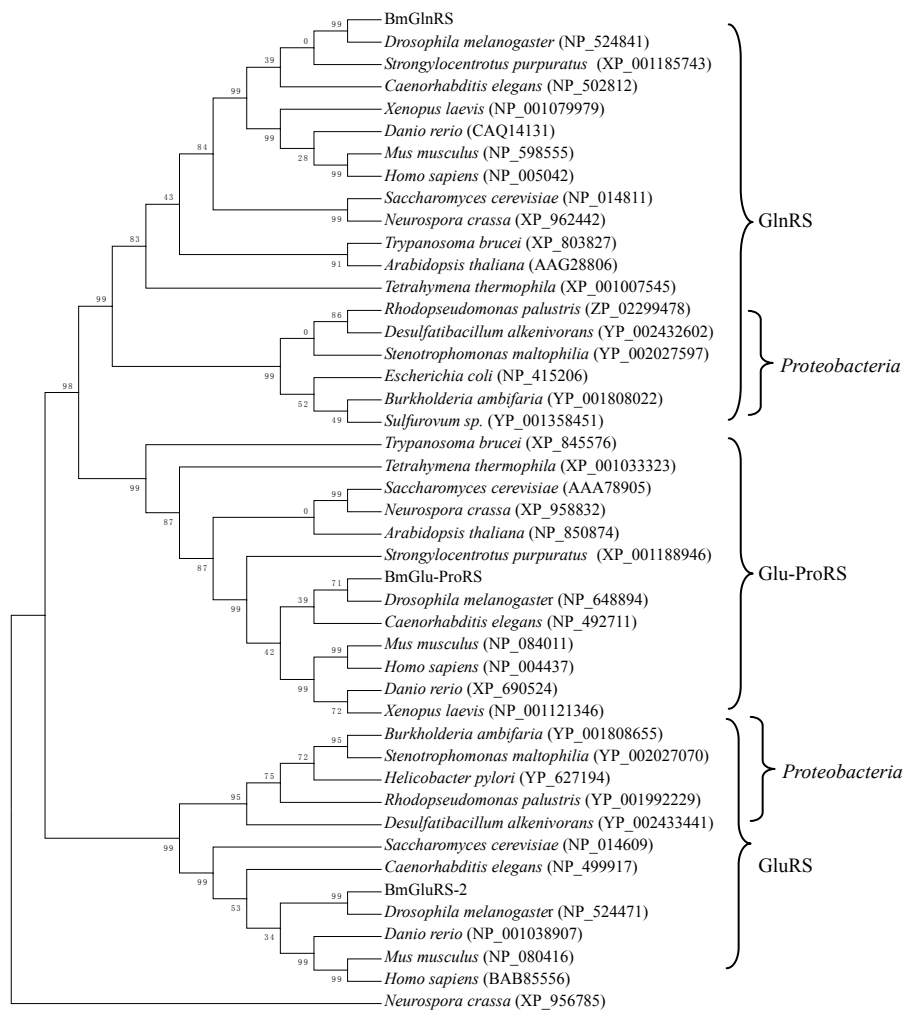


图 3 GlnRS / GluRS 的聚类分析

括号中数据为 GenBank 登录号。