

蓝细菌中与 P₀ 蛋白或 NtcA 蛋白相互作用的 PipX 蛋白的生物信息学分析*

张 鹏 王 莉 陈雯莉**

华中农业大学农业微生物学国家重点实验室, 武汉 430070

摘要 应用 NCBI、Expasy 等在线生物信息学网站与生物信息学软件对蓝细菌各菌株的 PipX 进行多序列同源对比分析, 获得保守序列, 预测了二级结构与结构域, 分析了 pipX 的基因簇等。分别构建了蓝细菌的 16S rDNA 进化树以及基于 PipX 的分子进化树, 同时也对与 PipX 相互作用的 NtcA 和 P₀ 构建分子进化树进行了比较分析, 了解它们之间起源与进化的关系。由此推测 PipX 可能是蓝细菌氮代谢途径中除 NtcA、P₀ 外的另一个关键调控蛋白, 有着重要的生理功能。

关键词 PipX; NtcA; P₀; 生物信息学; 蓝细菌

中图分类号 Q 939.9 **文献标识码** A **文章编号** 1000-2421(2010)02-0048-07

蓝细菌(Cyanobacteria)是一类极具多样性的光合自养原核生物^[1], 是地球上最为古老的物种之一, 出现在大约 30 亿年前。蓝细菌是重要的固氮生物, 由于固氮酶对氧极其敏感, 在进化过程中不同的固氮蓝细菌为了保护固氮酶的活性有着不同的机制: 有的丝状蓝细菌会在其菌丝体上将部分细胞分化成专门执行固氮功能的异形胞(heterocyst); 其它没有异形胞分化的蓝细菌中, 有的将固氮作用与光合作用分开进行(黑暗下进行固氮作用, 光照下进行光合作用)、有的通过提高胞内的过氧化物酶或超氧化物歧化酶活性以解除氧毒害。

鱼腥蓝细菌(*Anabaena*) PCC7120 为一种有异形胞分化的蓝细菌, 在缺乏化合态氮源时可以形成终端分化的异形胞进行固氮, 是一种非常重要的固氮研究模式菌。异形胞的发育是一个复杂的调控过程^[2], 涉及到许多基因和小分子信号。其中, Δ 酮戊二酸(2OG)作为氮源缺乏的初始信号, 诱导异形胞的分化^[3,5]; NtcA 是最重要的 2OG 接受蛋白, 在蓝细菌氮代谢及异形胞分化发育中起到全局性的调控作用。另外, P₀ 蛋白也可以与 2OG 结合参与细胞内代谢的调节, 是重要的碳/氮平衡调节信号, 能够与 NAGK、PphA 和 Pama 这 3 个蛋白相互结合^[6]。随后在没有异形胞分化的聚球蓝细菌(*Syn2*

echococcus) PCC7942 中发现 PipX 与 P₀ 及 NtcA 之间有相互作用^[7], Espinosa 还提出 PipX 蛋白可以通过蛋白质相互作用将 NtcA 与 P₀ 这 2 个氮平衡调节蛋白联系在一起, 而这一过程受到胞内 2OG 浓度水平的调节。PipX 还可通过介导 NtcA 蛋白的调控跟氮代谢有关^[8]。但是对于蓝细菌中 PipX 在进化过程中的信息, 它的功能结构等方面的研究则没有进一步的报道。

笔者通过生物信息学的方法试图了解蓝细菌中 NtcA2PipX2P₀ 三者之间的起源与进化关系, 并预测蓝细菌中 PipX 蛋白的氨基酸序列的保守位点与 PipX 的结构、功能, 旨在为进一步了解蓝细菌的氮代谢与异形胞分化提供有价值的信息。

1 材料与方法

1.1 材料

用于分析的不同物种中 PipX、NtcA 和 P₀ 的氨基酸序列与 16S rDNA 的核酸序列来源于 NCBI GenBank 数据库和 Cyanobase 数据库, 包括变形藻门(Proteobacteria): a. 变形藻纲的 *Gluconacetobacter diazotrophicus* PA15、b. 变形藻纲的 *Ralstonia pickettii* 12D、g. 变形藻纲的 *Shewanella denitrificans* OS217; 产水菌门(Aquificales): 产水菌纲

收稿日期: 200920212; 修回日期: 2009010010

* 国家自然科学基金项目(30670046)资助

** 通讯作者。E-mail: wlchen@mail.hzau.edu.cn

张 鹏, 女, 1984 年生, 硕士研究生。研究方向: 蓝细菌分子遗传学。E-mail: zhangpeng726@gmail.com

的 *Sulfurihydrogenibium azorense* Az2Fu1; 蓝细菌门的绝大部分的菌株, 如 *Synechococcus elongatus* PCC6301、*Synechococcus elongatus* PCC7942、*Nostoc punctiforme* ATCC29133、*Anabaena variabilis* ATCC29413、*Anabaena* sp. PCC7120、*Cyanothece* sp. ATCC51142、*Microcystis aeruginosa* NIES2843、*Synechocystis* sp. PCC6803、*Cyanothece* sp. PCC8801、*Acaryochloris marina* MBIC11017、*Thermosynechococcus elongatus* BP21、*Trichodesmium erythraeum* IMS101、*Synechococcus* sp. JA223B、*Synechococcus* sp. JA2323Ab、*Prochlorococcus marinus* MIT9313、*Gloeobacter violaceus* PCC7421、*Synechococcus* sp. RCC307、*Nodularia spumigena* CCY9414、*Lyngbya* sp. PCC8106、*Arthrospira maxima* CS2328、*Crocospira watsonii* WH8501、*Microcoleus chthonoplastes* PCC7420、*Cyanothece* sp. PCC7424 等。

1.2 方法

通过 Cyanobase、NCBI 网站获取各序列, 在网站 (<http://www.phylogeny.fr>)^[9] 采用 Clustal W 程序进行多序列比对, 然后通过 NJ (neighbor2joining) 方法构建了分子进化树, 最后在软件 MEGA4 中进行处理。在分析 PipX 的保守序列时, 采用 Clustal W 程序进行多序列比对, 再在网站 (<http://weblogo.berkeley.edu/logo.cgi>) 将数据图式化。PipX 蛋白的二级结构通过 PHD^[10] 网站 ([\[npsa2pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html\]\(http://npsa2pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html\)\) 提供的在线工具分析完成。结构域的分析利用了 Pfam^{\[11\]} 网站 \(<http://pfam.sanger.ac.uk/>\) 和 Smart 网站 \(<http://smart.embl-heidelberg.de/>\) 提供的在线工具完成, 蛋白功能预测在 ProtFun^{\[12\]} 网站 \(<http://www.cbs.dtu.dk/services/ProtFun/>\) 完成。PipX 所处基因组的上下游基因信息在 Cyanobase 及 NCBI 上获得, 再利用软件 Enhance 作图。](http://</p>
</div>
<div data-bbox=)

2 结果与分析

2.1 保守序列及二级结构分析

在 NCBI GenBank 与 Cyanobase 数据库中, 以聚球蓝细菌 PCC7942 中的 PipX 氨基酸序列为基准对数据库进行 Blastp (protein2protein BLAST), 从结果中选取了 23 种代表性菌属的序列进行多序列比对, 分析出保守序列。结果见图 1。利用 PHD 在线工具对聚球蓝细菌 PCC7942 中的 PipX 氨基酸序列进行二级结构的预测, 发现氨基酸序列中进行 α 螺旋 (alpha helix) 的氨基酸残基有 50 个, 占多肽全长的 56.18%; 进行无规则卷曲 (random coil) 的有 32 个, 占 35.96%; 剩下 7 个则是以延伸链 (extended strand) 的形式存在, 占 7.86%, 结果见图 1, 将预测出的二级结构标注在对应氨基酸序列下面, 发现 3 个 α 螺旋的氨基酸序列都相对比较保守, 而其中第 1 个 α 螺旋的氨基酸序列特别保守。

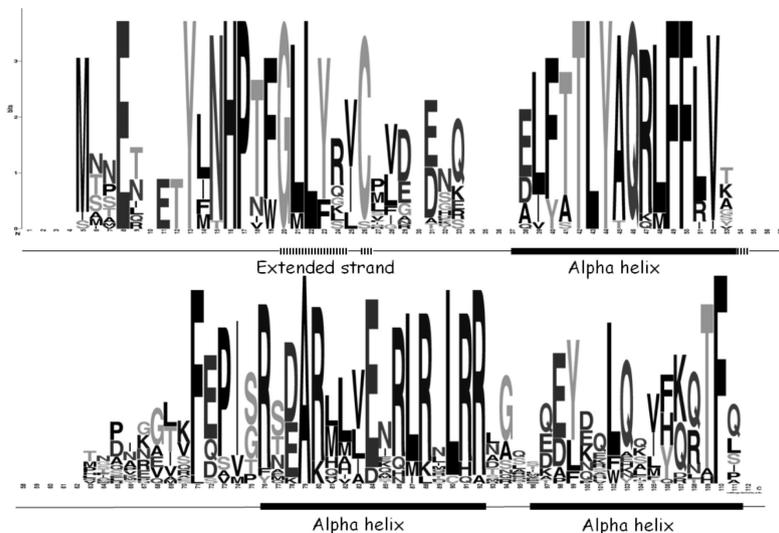


图1 PipX 的保守序列及二级结构

Fig. 1 The conservative sequences and secondary structure of PipX

2.2 蓝细菌 PipX 的序列及理化性质分析

在上面的 23 种菌株中再次选取 12 种有代表性的菌株进行序列与理化性质的分析。在 Cyanobase 与 NCBI GenBank 数据库中查找获取 PipX 的核苷酸和氨基酸序列,用 ProParam 在线软件分析氨基酸序列的理化性质,结果(表 1)显示:蓝细菌 PipX 平均核苷酸对数为 273 bp, G+ C 摩尔分数为 42.83%;多肽大小为 90 个氨基酸(约 10.7 ku);理论等电点约为 8.62,除了在 *Crocospaera watsonii* WH8501 中 PipX 蛋白为酸性蛋白外,其他都为碱性蛋白质;不稳定指数约为 53.06,为不稳定蛋白质;亲水性平均系数为-0.345,表明它们为亲水性蛋白质。

表 1 蓝细菌 PipX 的序列及理化性质分析

Table 1 Analysis of sequence and physicochemical property of PipX in cyanobacteria

物种 Species	核苷酸数 Number of nucleotides	(G+ C) 摩尔分数 $X_{(G+C)} / \%$	氨基酸数 Number of amino acids	分子量 Molecular weight/ku	理论等电点 Theoretical pI	不稳定指数 Instability index	脂肪族系数 Aliphatic index	亲水性平均系数 Grand average of hydropathicity
<i>Synechococcus elongatus</i> PCC7942	270	52.59	89	10.53	8.97	45.91	83.37	-0.330
<i>Anabaena</i> sp. PCC7120	279	41.94	92	11.02	7.91	52.57	80.54	-0.599
<i>Trichodesmium erythraeum</i> IMS101	288	36.46	95	11.51	9.77	61.23	93.26	-0.503
<i>Nodularia spumigena</i> CCY9414	279	39.80	92	10.95	7.92	47.16	75.22	-0.527
<i>Lyngbya</i> sp. PCC8106	270	42.22	89	10.56	8.98	59.41	85.39	-0.331
<i>Microcoleus chthonoplastes</i> PCC7420	264	43.18	87	10.27	9.03	53.66	85.17	-0.294
<i>Crocospaera watsonii</i> WH8501	270	35.93	89	10.67	6.72	56.07	100.79	-0.135
<i>Cyanothece</i> sp. ATCC51142	270	33.70	89	10.77	8.82	47.00	87.53	-0.335
<i>Synechocystis</i> sp. PCC6803	267	40.07	88	10.45	8.85	39.78	93.18	-0.220
<i>Synechococcus</i> sp. RCC307	267	56.55	88	10.42	8.98	73.31	68.75	-0.426
<i>Prochlorococcus marinus</i> MIT 9211	273	38.83	90	10.67	7.94	53.61	71.56	-0.443
<i>Gloeobacter violaceus</i> PCC7421	276	52.54	91	10.68	9.51	47.04	107.14	0.001
平均值 Mean	273	42.83	90	10.71	8.62	53.06	85.99	-0.345

2.4 pipX 基因簇的分析

在细菌基因组上,许多基因紧密成簇排列,基因与基因之间可能紧密连锁且功能密切相关,而不同物种间基因簇的排列规律还能透露出该物种进化的信息。为了进一步了解 pipX 的功能及其在蓝细菌中的进化信息,本研究分析了 13 种蓝细菌菌株(图 2),分别代表蓝细菌的各种生理特征。每个菌中都列出 pipX 基因及其上游和下游的 5 个基因共 11 个基因所组成的基因连锁图,这里将其定义为 pipX 基因簇。结果显示,除 *Trichodesmium erythraeum*

2.3 结构域分析

选取了表 1 中 13 种菌株的 PipX 氨基酸序列,利用 Pfam 网站提供的工具进行功能域分析,发现在 Pfam 数据库中,已经为这个序列家族定义了一个结构域,但是其具体功能还没有被注释,该结构域在 PfamB 数据库,对应编码为 PB003564。利用 Smart 网站的分析软件分析发现这些序列中都不包含跨膜结构、卷曲螺旋和信号肽片段。最后利用 ProtFun 网站提供的工具分析预测 PipX 的功能,发现蓝细菌中 PipX 的功能最可能的是参与能量代谢或是参与翻译过程,另外,它还可能作为生物合成的配体,参与复制与转录过程或参与氨基酸合成。

IMS101 外,在大部分固氮蓝细菌中 pipX 基因下游紧邻一个 650 bp 左右的基因,很有可能与 pipX 共用一个启动子进行转录,并且发现该基因在这些固氮蓝细菌中同源性极高,另外, Cyanobase 网站对某些菌中该基因的注释为含有丙胺酸消旋酶结构域。在固氮蓝细菌中 pipX 基因簇间有很大的差别,可以分为 3 类:丝状产异形胞菌属如 *Anabaena* sp. PCC7120、*Anabaena variabilis* ATCC29413 和 *Nostoc punctiforme* ATCC29133 这 3 种为一类基因簇;丝状不产异形胞蓝细菌如 *Trichodesmium*

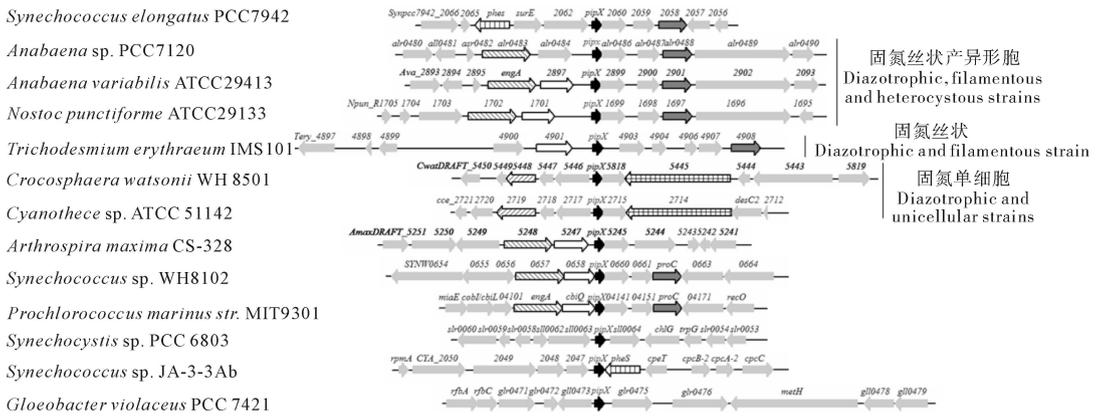


图 2 蓝细菌 13 个菌株中 pipX 基因簇信息

Fig. 2 Alignment for pipX clusters in 13 species of cyanobacteria

erythraeum IMS101 为一类基因簇; 单细胞不产异形胞的蓝细菌如 Crocosphaera watsonii WH8501 和 Cyanothece sp. ATCC51142 为一类基因簇。

2.5 进化树分析

对 16S rDNA、PipX、P₀ 和 NtcA 分别进行了进化树分析, 后者蛋白在进行预分析时选取的物种为 NCBI 数据库 Blastp 结果中几乎所有的物种, 仅除去全序列同源性 30% 以下的物种。然后, 在预

分析结果的基础上进行数据优化, 得到最终的进化树。

构建 16S rDNA 序列的进化树, 选取了 23 种蓝细菌的全序列进行分析, 结果见图 3。结果显示固氮丝状产异形胞蓝细菌位于一个分支, 固氮单细胞蓝细菌位于另一个分支。聚球蓝细菌和原绿蓝细菌 2 种海洋单细胞蓝细菌位于另一个遥远的分支, 其中包含多种亲缘关系非常近的蓝细菌, 在构建进化树时都进行了简化。

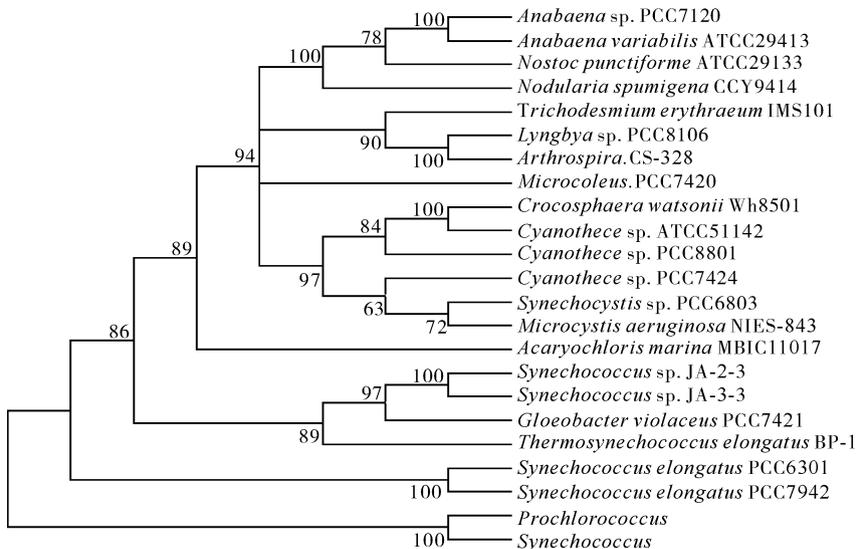


图 3 基于 16S rDNA 序列构建的蓝细菌的进化树

Fig. 3 Phylogenetic tree of 16S rDNA homologs

构建 PipX 氨基酸序列的进化树, 选取了同上 面一样的 23 种菌株的全序列进行分析, 结果见图 4。采取同样的方法构建 NtcA 氨基酸序列的进化树, 结果如图 5 所示。PipX 和 NtcA 这 2 个蛋白仅存在于蓝细菌中, 并且它们存在于蓝细菌绝大多数

的菌属之中, 包括在进化上很古老的 Gloeobacter violaceus PCC7421 和 Thermosynechococcus elongatus BP21, 这表明在所有蓝细菌的共同祖先中就已经存在 PipX 和 NtcA, 而且在随后漫长的进化过程中也没有丢失。NtcA 和 PipX 进化树的拓扑结

构与蓝细菌 16S rDNA 进化树的拓扑结构相似, 其中各类固氮菌属构成的拓扑结构相似程度较大。

构建 P₀ 氨基酸序列的进化树时, 菌株的选择与 PipX 和 NtcA 有一点差别。因为 PipX 和 NtcA 只存在于蓝细菌中, P₀ 不仅仅存在于蓝细菌还存在于其他物种中, 而且在蓝细菌的有些属中存在 2 个同源的 P₀ 蛋白。于是选取了上面 23 种蓝细菌, 其中有的菌株有 2 个 P₀ 氨基酸序列(如 *Cyanoth2*

ece sp. PCC7424、*Lyngbya* sp. PCC8106、*Acaryo2* *chloris marina* MBIC11017、*Gloeobacter violaceus* PCC7421 这 4 种蓝细菌), 另外还选取了变形蓝细菌的 4 个菌和产水菌中的 1 个菌, 结果见图 6。产水菌和变形菌都处在进化树上比较遥远的独立分支, 说明了其与蓝细菌的亲缘关系十分远。图 6 显示的 4 个蓝细菌有 2 个 P₀, 而 4 个菌中都有一个 P₀ 蛋白处于较远的分支, 另一个 P₀ 蛋白的分支则处于大的蓝细菌各菌属分支中。

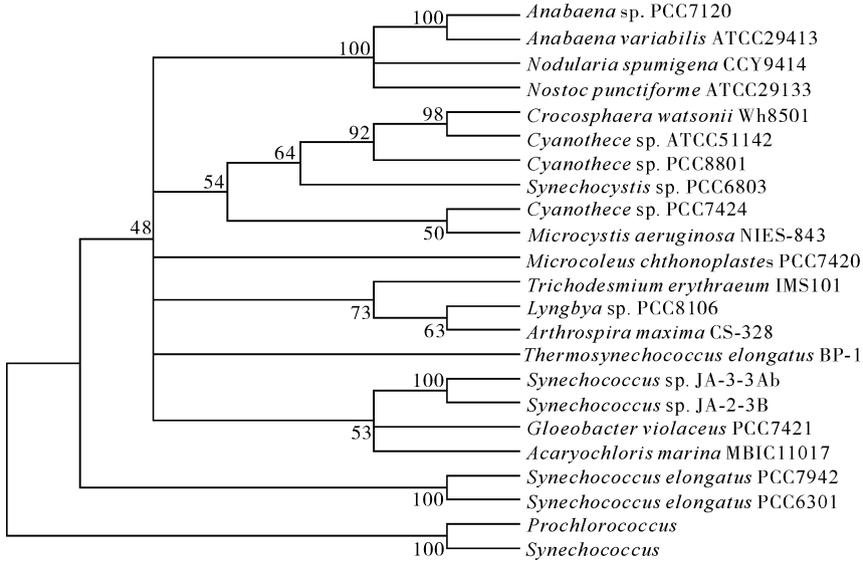


图 4 基于 PipX 构建的蓝细菌的进化树

Fig. 4 Phylogenetic tree of PipX homologs

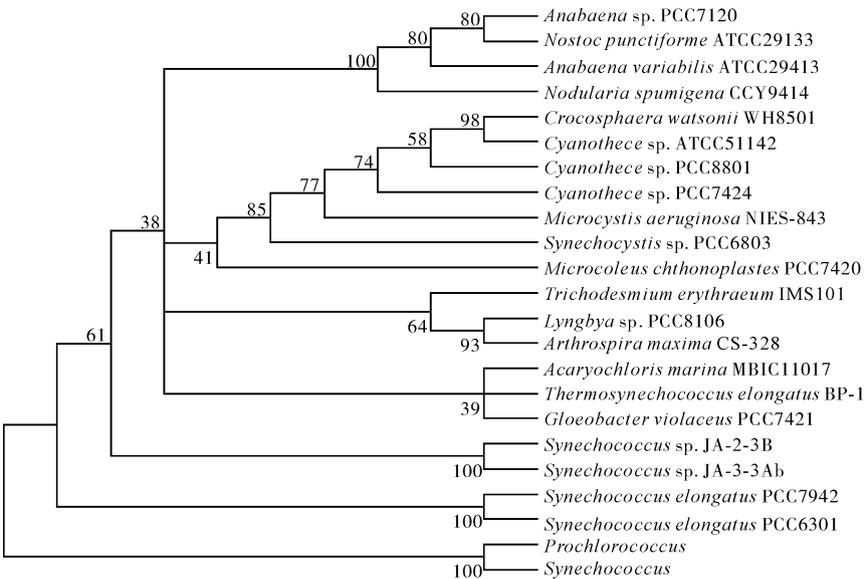
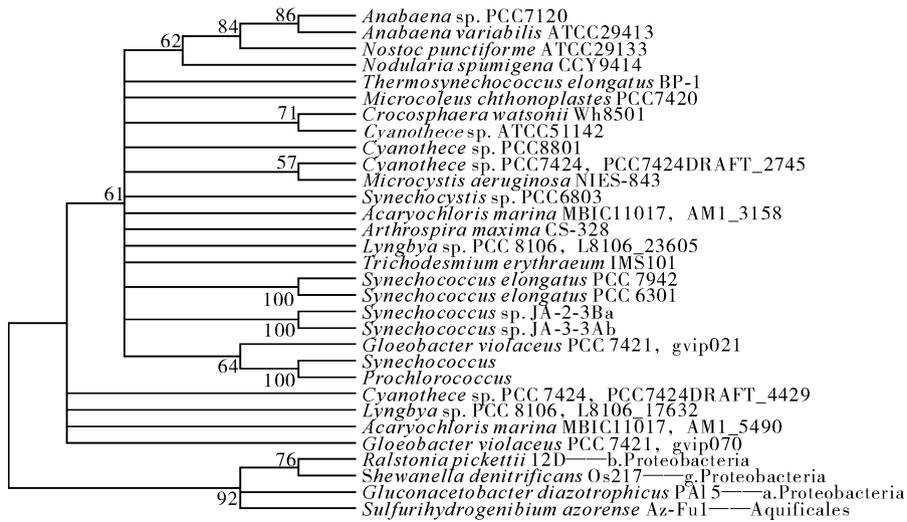


图 5 基于 NtcA 构建的蓝细菌的进化树

Fig. 5 Phylogenetic tree of NtcA homologs

图 6 基于 P₀ 构建的蓝细菌的进化树Fig. 6 Phylogenetic tree of P₀ homologs

综合分析 PipX 进化树、NtcA 进化树和 P₀ 进化树, 固氮丝状产异形胞的蓝细菌处在一个稳定的分支中, PipX 和 NtcA 进化树同 16S rDNA 进化树一样固氮丝状产异形胞蓝细菌、固氮丝状不产异形胞蓝细菌和固氮单细胞不产异形胞蓝细菌分别对应 3 个分支, 但每个进化树 3 个分支之间的拓扑结构有些许差别, 虽然 P₀ 进化树与其他进化树相比在固氮菌分支的拓扑结构上存在一些差异, 但是整个树形还是十分相似, 相对比较保守。另外还分析了 NtcA 蛋白结合配体结构域的氨基酸序列, 构建了对应的进化树(数据没有显示), 发现 NtcA 全蛋白序列的进化树与 NtcA 结合配体结构域序列的进化树也比较相似。

3 讨论

近年来, 伴随着基因组和功能基因组研究的广泛开展, 生物信息学理论和方法引起人们越来越重视的同时也得到了迅猛的发展, 为应用研究提供更多更加有用的信息。笔者利用各种生物信息学方法对蓝细菌 PipX 进行分析, 预测了其基本理化性质: 蓝细菌 PipX 为亲水性不稳定的蛋白质, 多肽大小为 90 个氨基酸(约 10.7 ku), 理论等电点约为 8.62, 除了在 *Crocosphaera watsonii* WH8501 菌属中 PipX 为酸性蛋白外, 其他都为碱性蛋白质; 二级结构以 α 螺旋和无规则卷曲为主, 零星散布着少量的延伸链; 序列中只含有一段功能没被注释的结构域, 无跨膜结构、无卷曲螺旋, 也无信号肽片段; 功能

是参与能量代谢或参与翻译过程, 也有可能作为生物合成的配体、参与复制与转录过程或参与氨基酸合成。

分析 pipX 基因簇发现: 在固氮单细胞蓝细菌和固氮丝状产异形胞蓝细菌中, pipX 基因下游紧邻一个 650 bp 左右的基因, 某些菌属中的该基因含有丙胺酸消旋酶结构域。在固氮蓝细菌中, pipX 基因簇有很大的差别, 基本可以分为丝状产异形胞蓝细菌、丝状不产异形胞蓝细菌和单细胞不产异形胞蓝细菌这 3 种基因簇, 推断 pipX 基因的进化与 16S rDNA 的进化存在着相似性。

对 16S rDNA 进化树与 PipX 进化树进行分析与比较, 发现 PipX 存在于几乎所有蓝细菌中, 基于它的进化树与 16S rDNA 进化树类似, 说明它可能起源于蓝细菌的共同祖先。对 16S rDNA、PipX、NtcA 和 P₀ 这 4 个进化树进行分析与比较, 发现 NtcA 进化树与 PipX 进化树类似于 16S rDNA 进化树, P₀ 进化树与 PipX 进化树和 16S rDNA 进化树的几个固氮菌属的拓扑结构有一点差别, 但是整个树形的拓扑结构还是十分相似的。所以, PipX 和 P₀ 或 NtcA 的相互作用在进化树上的分析显示: PipX 与 P₀ 在某些固氮菌属中的进化有一点差别, 但在大部分蓝细菌中的进化是一致的。另外, PipX 序列的保守性也说明其具有重要的生理功能, 结合 Espinosa 等人对聚球蓝细菌 PCC 7942 中 PipX 的研究结果^[8], 推测 PipX 可能是蓝细菌氮代谢途径中除 NtcA、P₀ 外的另一个重要调控蛋白。

致谢 感谢华中农业大学农业微生物学国家重点实验室研究生张巨源、戟博阳、张少然在本研究中提供的技术支持与数据分析!

参 考 文 献

- [1] STANIER R Y, COHENBAZIRE G. Phototrophic prokaryotes: the cyanobacteria[J]. *Annu Rev Microbiol*, 1977, 31: 225-274.
- [2] GOLDEN J W, YOON H S. Heterocyst development in *Anabaena*[J]. *Curr Opin Microbiol*, 2003, 6: 557-563.
- [3] LI J H, LAURENT S, KONDE V, et al. An increase in the level of 2-oxoglutarate promotes heterocyst development in the cyanobacterium *Anabaena* sp. strain PCC 7120[J]. *Microbiology*, 2003, 149: 3257-3263.
- [4] LAURENT S, CHEN H, BEDU S, et al. Non-metabolizable analog of 2-oxoglutarate elicits heterocyst differentiation under repressive conditions in *Anabaena* sp. PCC 7120[J]. *Proc Natl Acad Sci USA*, 2005, 102: 9902-9912.
- [5] ZHANG C C, LAURENT S, SAKR S, et al. Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals[J]. *Mol Microbiol*, 2006, 59: 362-375.
- [6] OSANAI T, TANAKA K. Keeping in touch with P₀: P₀ interacting proteins in unicellular cyanobacteria[J]. *Plant Cell Physiol*, 2007, 48(7): 908-914.
- [7] ESPINOSA J, FORCHHAMMER K, BURILLO S, et al. Inter2 action network in cyanobacterial nitrogen regulation: PipX, a protein that interacts in a 2-oxoglutarate dependent manner with P₀ and NtcA[J]. *Mol Microbiol*, 2006, 61: 457-469.
- [8] ESPINOSA J, FORCHHAMMER K, and CONTRERAS A. Role of the *Synechococcus* PCC 7942 nitrogen regulator protein PipX in NtcA controlled processes[J]. *Microbiology*, 2007, 153: 712-718.
- [9] DEREPPER A, GUIGNON V, BLANC G, et al. Phylogeny. fr: robust phylogenetic analysis for the non-specialist[J]. *Nucleic Acids Research*, 2008, 36: 465-469.
- [10] COMBET C, BLANCHET C, GEOURJON C, et al. NPS@: network protein sequence analysis[J]. *Trends Biochem Sci*, 2000, 25(3): 142-150.
- [11] FINN R D, TATE J, MISTRY J, et al. The PFAM protein families database[J]. *Nucleic Acids Research*, 2008, 36: 2812-288.
- [12] JENSEN L J, STARFELDT H H, BRUNAK S. Prediction of human protein function according to gene ontology categories[J]. *Bioinformatics*, 2003, 19: 635-642.

Bioinformatics Analysis of PipX, a Protein Interacting with P₀ or NtcA in Cyanobacteria

ZHANG Peng WANG Li CHEN Wenli

State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China

Abstract The multiple sequence homologous alignment of PipX in cyanobacteria and the conserved sequences were analyzed by bioinformatics softwares and websites such as NCBI, ExPasy and so on. The secondary structure and structural domain of PipX was predicted and pipX gene clusters were dissected. The phylogenetic tree of 16S rDNA in cyanobacteria was constructed. In addition, the phylogenetic relationships among PipX, NtcA and P₀ were studied by constructing phylogenetic trees of these genes. The results indicated that PipX is an important regulation protein in the nitrogen metabolism of cyanobacteria in addition to NtcA and P₀, and might have important physiological functions.

Key words PipX; NtcA; P₀; bioinformatics; cyanobacteria

(责任编辑: 张志钰)