

白及泛酸激酶基因的生物信息学分析

李清清¹, 黄恻隐¹, 范娠源¹, 刘海^{2*}, 汪牧非¹, 徐德林¹

¹遵义医科大学细胞生物学教研室 贵州遵义

²贵州省现代中药材研究所 贵州贵阳

【摘要】基于白及的转录组测序数据,采用生物信息学方法对白及 *PanK* 进行基因鉴定和序列分析。分析了白及 PANK 蛋白的等电点、亲疏水性以及二级结构、三级结构和结构域,进行亚细胞定位并预测其磷酸化位点、信号肽和跨膜区,最后,构建进化树分析。结果表明:白及 *PanK* 基因包含 2697 bp 的开放阅读框,共编码 898 个氨基酸,该蛋白相对分子质量为 99405.55 Da,是等电点 (pI) 为 5.79 的亲水蛋白质。该蛋白有 11 个苏氨酸磷酸化位点,具有泛酸激酶 (PLN02902 家族蛋白) 保守结构域,含 10 个功能 motif; 无信号肽及跨膜区,亚细胞定位发现该蛋白定位于细胞质中;进化树分析表明白及与铁皮石斛的 *PanK* 基因亲缘关系最近。结论:该基因具有泛酸激酶结构特征,氨基酸序列及核酸序列与同源物种相似度高,表明该基因在同源物种进化上高度保守,推测其在细胞质中起作用,可能与细胞代谢中的能量提供有关。

【关键词】白及;泛酸激酶基因;结构特征;功能预测

【基金项目】国家自然科学基金项目 (31560079、31560102),贵州省科学技术基金项目 (黔科合平台人才[2017]5733-050,黔科合成果[2019]4316 号),遵义市科技局遵义医科大学科学技术联合资金招标专项 (遵市科合 HZ 字 (2020) 91 号)

Bioinformatics Analysis of *PANK* in *Bletilla striata*

Qingqing Li¹, Ceyin Huang¹, Shenyuan Fan¹, Hai Liu^{2*}, Mufei Wang¹, Delin Xu¹

¹Department of Cell Biology, Zunyi Medical University, Zunyi, Guizhou, China

²Institute of modern traditional Chinese medicine of Guizhou Province, Guiyang, Guizhou, China

【Abstract】Based on transcriptome sequencing data of *Bletilla striata* (*B. striata*), we used bioinformatics methods to analyze gene identification and sequence of *PanK*. First, we analyzed the isoelectric point, hydrophilicity and hydrophobicity, secondary and tertiary structure and domain of PANK protein in *B. striata*. Then we analyzed its subcellular localization, predicted phosphorylation site, signal peptide and transmembrane region. Finally, we built the evolutionary tree. The results showed that the *PanK* gene of *B. striata* contained an ORF of 2697 bp and encoded a total of 898 amino acids. The protein had a relative molecular mass of 99405.55 Da, which predicted to be a hydrophilic protein with an isoelectric point (pI) of 5.79. The protein had 11 threonine phosphorylation sites, and a conserved domain of pantothenate kinase (PLN02902 family protein), which contained 10 motifs. Then we found that this protein had no signal peptide and transmembrane region, and it was located in the cytoplasm through subcellular localization. The phylogenetic tree analysis showed that the *PanK* gene of *B. striata* was more closed to *Dendrobium officinale* (*D. officinale*). In conclusion, this research showed that the *PanK* gene had the structural characteristics of pantothenate kinase, and the sequence of amino and nucleic acid were highly similar to homologous species, which meant it was highly conservative in evolution of homologous species. So we speculated that it played a role in the cytoplasm and might be related to the energy supply in cell metabolism.

第一作者:李清清 (1999-),女,在读硕士,研究方向为中药材细胞工程。

*通讯作者:刘海 (1984-),男,研究实习生,研究方向为中药材资源开发与利用,电子邮箱:liuhai200321@163.com。

【Keywords】 *Bletilla striata*; *PanK*; Structure Characteristics; Functional Prediction

白及 (*Bletilla striata* (Thunb.) Reichb.f) 是原产于我国长江流域^[1]的一种兰科浅根系多年生草本植物, 含有联苜类、二氢菲类、联菲类等多种化合物。现代药理研究表明, 白及具有止血、抗菌、抗肿瘤、抗溃疡、预防肠粘连、促进伤口愈合、抗失血性休克、保护皮肤、延缓衰老等功效^[2], 具有较高的药用价值和观赏价值^[3]。

泛酸激酶 (pantothenate kinase, 即原核生物中的 CoaA, 真核生物中的 PanK) 又称 ATP:D-泛酸-4-磷酸转移酶, 是参与辅酶 A (coenzyme A, CoA) 生物合成途径第一步的限速酶, 在生物体代谢中起着非常重要的作用。发现的泛酸激酶有 I 型、II 型和 III 型三种类型, 大肠埃希菌 (*Escherichiacoli*, *E.coli*) 的泛酸激酶 (*EcCoaA*) 是 I 型的代表; 金葡菌 (*Staphylococcus aureus*) 的泛酸激酶 (*SaCoaA*) 是 II 型的代表, 真核生物 PanKII 酶常以不同的亚型出现, 它具有分别由 *PanK1-4* 编码的四个同工型: mPanK1、mPanK2、mPanK3 和 mPanK4; III 型主要存在于幽门螺旋杆菌 (*Helicobacter pylori*)、结核分枝杆菌 (*Mycobacterium tuberculosis*, Mtb)、炭疽芽孢杆菌 (*Anthraxis*) 等微生物中, 也称 CoaX^[4]。在一个生物体中, 可以存在多种类型的泛酸激酶, 比如 Mtb 和枯草芽孢杆菌 (*Bacillus subtilis*) 同时含有 I 型和 III 型, 但缺乏 II 型^[5, 6]。目前, 对基因序列进行比较基因组学对比发现, 包括古细菌在内的几乎所有生物体中都包含泛酸激酶相关基因^[7, 8]。泛酸激酶在进化上高度分化, 但功能保守^[9], 真核生物和原核生物的泛酸激酶序列同源性较低。Awasthy 等人的研究表明 I 型同工酶 *CoaA* 是结核分枝杆菌中必不可少的^[5], 张鹭等人研究得到了高表达、高纯度的重组结核分枝杆菌的泛酸激酶^[10], 且与人类的泛酸激酶基因同源性极低^[11], 因此泛酸激酶被认为是抗菌抗结核药物开发的潜在新靶点, 在人体内该基因的突变则会引发神经退行性病变^[12, 13], 研究白及的泛酸激酶基因对于探索人类相关疾病的发病机制进而治疗该疾病具有一定的参考价值。

对于泛酸激酶已经在原核生物大肠埃希菌、肠沙门氏菌 (*Salmonella enteritidis*) 和鞘氨醇单胞菌 (*Sphingomonas paucimobilis*) 等革兰氏阴性菌以及酿酒酵母 (*Saccharomyces cerevisiae*) 等真核生物中

得到鉴定^[14], 对于人类的泛酸激酶神经性退行疾病也有研究^[15], 但对于白及泛酸激酶基因的结构和功能的研究还未见报道。本文对白及中的泛酸激酶基因进行分析, 对揭示泛酸激酶在白及中的生物学功能以及了解其在整个生物体代谢中的作用具有重要意义, 也为深入分析其它物种泛酸激酶的结构与功能关系提供参考, 有利于白及的开发利用, 并为研究与泛酸激酶相关的疾病提供科学依据。

1 材料与方法

1.1 材料

采用中国贵州省正安县 (北纬 28 度 56 分, 东经 107 度 43 分) 的白及成熟种子为材料, 诱导白及愈伤组织并进行共 45 天的悬浮培养^[16], 从 0 天开始每 3 天随机取样一次 (每个时间点取 3 个重复), 液氮研磨提取各个样本的总 RNA, 等量混合各个样本的 RNA 成一个 mix 用 PacBio 测序平台的 Iso-Seq 转录组测序^[17]。从测序结果中筛选出白及的 *PanK* 基因序列用于后续分析。

1.2 方法

用 NCBI 的 BLAST 在线软件对白及 *PanK* 基因的核苷酸序列进行比对分析, 并利用 ORF finder 分析这一核苷酸序列的开放阅读框; 采用 ExPasyProt Param 分析白及 PANK 的相对分子质量、等电点和氨基酸组成; 使用 Prot Scale 软件进行亲疏水性分析; 用 NetPhos3.1 Server 对白及 PANK 的磷酸化位点进行预测; 使用在线工具 SOPMA 分析白及 *PanK* 基因的二级结构, SWISS MODEL 进行序列的三维结构预测; 用 NCBI 分析其蛋白质保守结构域; 利用在线软件 Signal P4.0 Server 对信号肽进行预测, TMHMM 2.0 Server 对跨膜区进行预测; 应用在线工具 Plant-mPLoc 和 GenScript 对亚细胞进行定位分析, 运用 MEME5.0.5 在线软件 (<http://meme-suite.org/>) 分析蛋白质的基序, 设定基序宽度最大值为 50, 基序数为 20^[18], 其余参数为默认值。利用在线软件 ProtFun 以及 String 对白及 *PanK* 基因进行功能分类预测; 使用 MEGA6.0 软件的邻接法 (Neighbor-joining, NJ), 参数设置 SUBSTITUTION MODEL 选择 p-distance、DATA SUBSET TO USE 选择 Partial deletion、Site Coverage Cutoff=50% 和 Bootstrap method=500, 构建多种 *PanK* 序列的系

统发生树。

2 结果与分析

2.1 *PanK* 基因核苷酸与氨基酸序列的理化性质特征分析

从白及悬浮培养细胞的核苷酸数据库文库中筛选出 *PanK* 基因序列, 全长为 3075bp, ORF finder 发现该序列包含 1 个完整的开放阅读框, 全长 2697 bp, 共编码 898 个氨基酸序列残基 (Amino acid, AA)。ExpasyProt Param 软件预测显示白及 PANK 由 20 种氨基酸组成, 其中 Leu 含量最高, 占比为 12.1%, Gly 含量次之, 占比为 9.2%, 不含 Pyl 和 Sec。该蛋白相对分子质量为 99405.55 Da, 理论等电点 (pI) 为 5.79, N 末端为 Met, 原子组成为 $C_{4414}H_{6980}N_{1216}O_{1321}S_{37}$, 负电荷残基 (Asp+Glu) 122 个, 正电荷残基 (Arg+Lys) 105 个, 不稳定系数为 39.55 (<40), 表明该蛋白质结构稳定。

2.2 疏水性分析

利用在线软件 Prot Scale 对白及 PANK 蛋白的亲疏水性进行预测分析, 其中分值越低代表亲水性越强, 分值越高其疏水性越强。分析结果表明, 白及 PANK 亲水性最大的氨基酸是精氨酸, 分值为-4.5, 疏水性最大的氨基酸是异亮氨酸, 分值为 4.5。从整体上看, 白及 PANK 的亲水性氨基酸要多于疏水性氨基酸, 其亲水性评估系数为-0.221。因此, 白及 PANK 是一个亲水的蛋白质 (图 1)。

2.3 磷酸化修饰检测

蛋白磷酸化是一种最普遍的翻译后修饰, 主要磷酸化位点有丝氨酸、苏氨酸以及酪氨酸残基。蛋白磷酸化修饰几乎参与调节生命活动的整个过程, 因此蛋白质磷酸化修饰的分析对于了解其在生命活动中的作用至关重要。一般来说, 多肽链中的氨基酸潜在的磷酸化位点越多, 发挥更多功能的可能性就越大。用 NetPhos3.1 Server 对白及 PANK 的蛋白质序列进行潜在磷酸化位点分析, 结果表明白及 PANK 一共有 11 个苏氨酸磷酸化位点 (图 2)。

2.4 蛋白质结构预测分析

用 SOPMA 对白及 *PanK* 基因的二级结构进行预测, 分析表明主要的结构元件是 α -螺旋和无规卷曲 (图 3 A)。从分布比例上看, 该基因 898 个氨基酸残基中, 384 个氨基酸残基为 α -螺旋占 42.76%, 126 个氨基酸残基为线性结构占 14.03%, 332 个氨基酸残基为无规卷曲占 36.97%, 56 个氨基酸残基为 β -转角占 6.24%。利用 SWISS-MODEL 绘制出白及 PANK 的三级结构图, 结果显示蛋白螺旋与无规则卷曲占据了该蛋白高级结构的大部分空间 (图 3 B), 这与白及 PANK 的二级结构相吻合。

2.5 基序和保守结构域分析

利用 MEME 分析白及 PANK 氨基酸序列得到 10 个 motif (图 4), 利用 SMART 分析白及 PANK 蛋白功能结构域, 结果表明: 该基因的第 23 到 896 个氨基酸是泛酸激酶 (PLN02902 家族蛋白) 保守结构域 (图 5)。

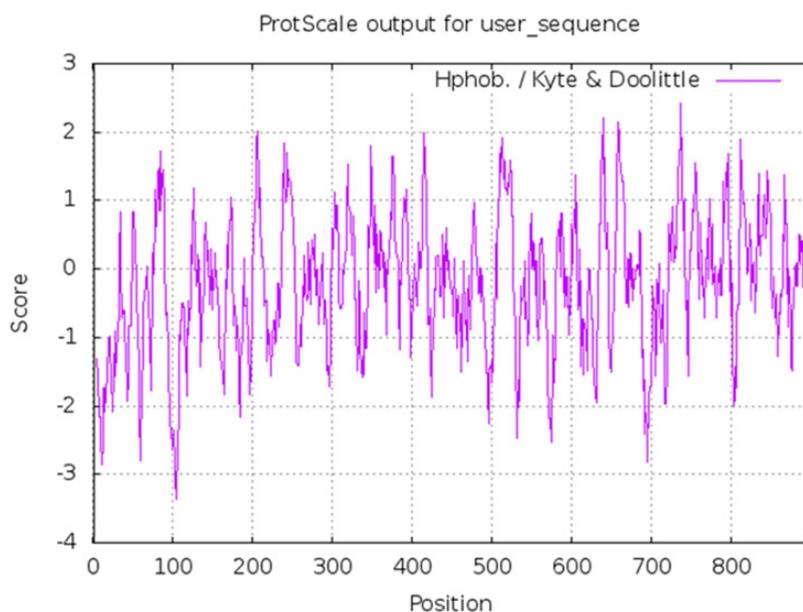


图 1 白及 PANK 疏水性分析结果

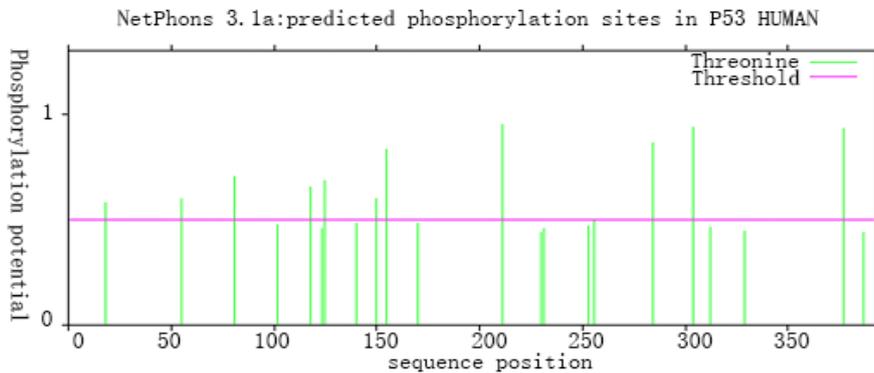
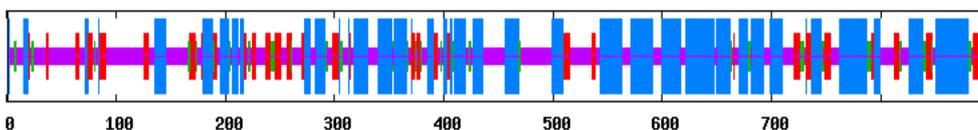
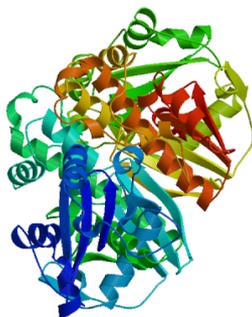


图2 NetPhos2.0 Server 对白及 PANK 的磷酸化位点预测



A

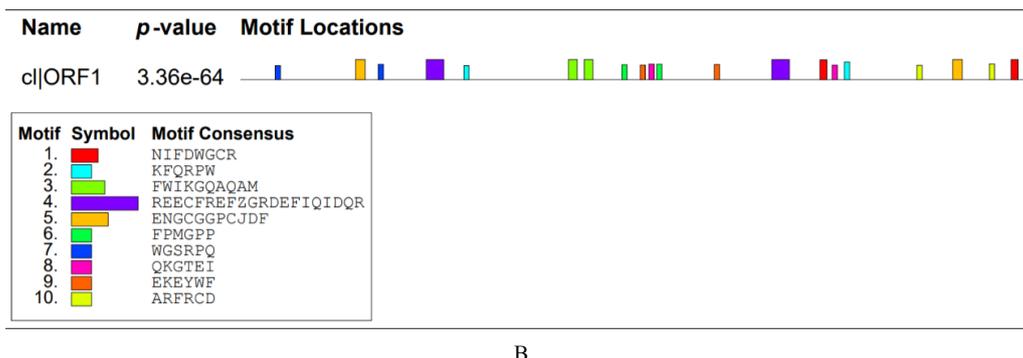


B

图3 白及 *PanK* 基因的结构预测。A.二级结构, B.三维结构。蓝色代表 α -螺旋, 红色代表延伸, 绿色代表 β -转角, 紫色代表无规卷曲。



A



不同色块的长度和顺序分别代表各基序在蛋白质序列中的实际大小和位置。

图 4 白及 PANK 蛋白的基序序列(A)及位置(B)

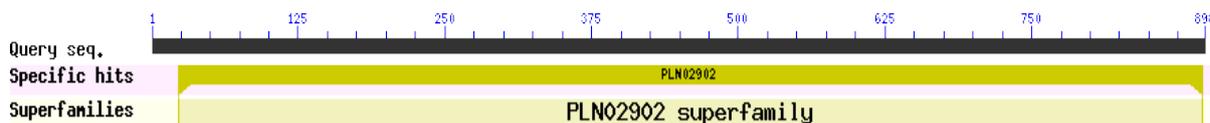


图 5 白及的 PANK 保守结构域

2.6 亚细胞定位、信号肽与跨膜区以及功能预测

不同的亚细胞定位决定植物不同的特异的生物学功能。使用在线工具 Psort 对白及 PANK 的亚细胞定位进行预测, 结果显示, 其可能定位于细胞质中。GenScribe 分析得到 56.5%定位于细胞质中, 26.1%定位于核中, 17.4%定位于线粒体中。Signal P4.0 Server 软件分析表明, *PanK* 基因编码的蛋白质无信号肽。TMHMM 2.0 Server 软件对白及 PANK 蛋白氨基酸序列的跨膜结构域进行在线预测表明, PANK 的整条肽链都位于细胞膜外, 不存在跨膜区。结合上述预测结果推测, 在细胞质基质中合成后的 PANK 很可能被直接锚定在细胞质基质特定部位行使功能, 不进行蛋白质转运。

2.7 白及 *PanK* 基因功能预测分析

采用 ProtFun 对白及 *PanK* 的功能进行预测, 结果显示 562-592 之间的氨基酸序列 (KAVASEGGTDDAKRRGEAFARAFSAHLARLM

) 有 FAD、NAD 以及 NADP 的相关功能。采用 String 在线软件对白及 *PanK* 的功能进行预测 (表 1)。

上述结果表明: 白及 *PanK* 基因可能具有泛酸激酶活性, 结合 ATP 功能, 还可能具有 I 型磷酸二酯酶/核苷酸焦磷酸酶、碳氮水解酶、泛酸激酶 II 型、磷酸转移酶等功能, 在细胞生长发育过程中起重要作用。

2.8 系统进化分析

利用 MEGA6.0 软件, 选择 NJ 法构建系统进化树, 比较分析不同物种中 *PanK* 基因的进化与亲缘关系, 共涉及 20 个物种 (相同物种取 Score 值最高的序列进行比对)。系统发育分析图中 (图 6) 白及和在铁皮石斛 (*Dendrobium catenatum*) 属于同一小分支, 与蝴蝶兰 (*Phalaenopsis equestris*) 同属第二小分支, 说明白及与铁皮石斛亲缘关系最近, 与蝴蝶兰的亲缘关系次之。

表 1 白及 *PanK* 的功能预测

Domain	description	strength
PF03630	Fumble	3.29
PF01663	Type I phosphodiesterase / nucleotide pyrophosphatase	3.05
PF00795	Carbon-nitrogen hydrolase	2.45
IPR004567	Type II pantothenate kinase	3.29
IPR002591	Type I phosphodiesterase/nucleotide pyrophosphatase/phosphate transferase	3.12
IPR017850	Alkaline-phosphatase-like, core domain superfamily	2.82
IPR036526	Carbon-nitrogen hydrolase superfamily	2.45
IPR003010	Carbon-nitrogen hydrolase	2.45

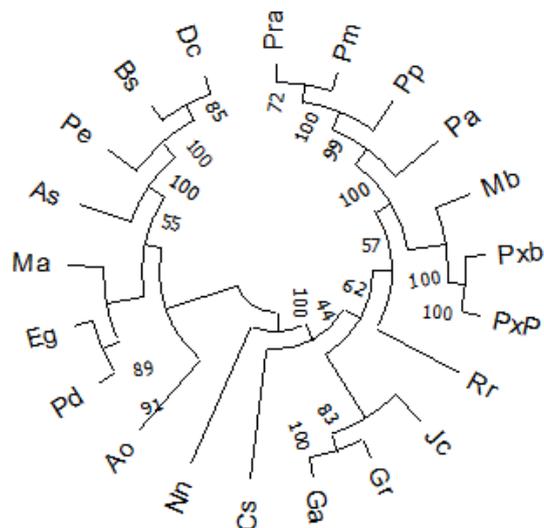


图6 *PanK* 基因的系统进化树 Bs: 白及; Dc: 铁皮石斛; Pe: 蝴蝶兰; As: 深圳拟兰; Pd: 海枣; Ao: 石刁柏; Ma: 香蕉亚种; Pa: 欧洲甜樱桃; Pra: 杏; Rr: 苞叶木; Jc: 麻风树; Pxb: 白梨; Pp: 巴旦木; Pm: 红梅; Mb: 山荆子; Nn: 荷花; Gr: 雷蒙德氏棉; Eg: 油棕; Cs: 茶; Pxp: 秋子梨; Ga: 棉树

3 讨论

泛酸激酶参与了 CoA 的生物合成的第一步, CoA 是生命活动必须的辅助因子, CoA 及其衍生因子直接或间接参与了生物体内众多重要的代谢反应^[19], 因此, *PanK* 基因是参与生物体内代谢的关键基因。本研究鉴定了白及的 *PanK* 基因, 分析表明该蛋白是一个保守的亲水性蛋白, 说明该基因能够减少水分丧失, 在脱水组织中防止蛋白质失活^[20]。与兰科植物铁皮石斛和蝴蝶兰同源性较高, 均在 91% 以上, 表现出高度的保守性, 推测其在不同植物中发挥相似的生理功能, 有利于植物遗传和生物学功能的稳定性, 但其在不同物种之间的序列差异较大^[21], 在进化上高度分化但功能保守, 根据这一特点, 该基因有可能成为研发抗菌药物的潜在靶点。

结构分析发现, 白及 PANK 的二级结构主要由 α -螺旋结构 (Alpha helix) 和无规则卷曲结构 (Random coil) 两个元件构成, 结合线性结构和 β 螺旋形成复杂的三级结构, 亚细胞定位于细胞质中。通过分析发现, *PanK* 具有典型的泛酸激酶蛋白家族结构域, 且具有结合 ATP 的功能, 有 11 个苏氨酸磷酸化位点, 其可能在白及的代谢过程中参与信号调控和能量供应等重要过程。因此, 推测白及 *PanK* 基因在生物体代谢过程中参与细胞代谢相关途径的能量供应, 在生物体生长发育以及代谢过程中发挥着重要的调控功能, 具体调控机制仍有待发掘。

本研究首次对白及 *PanK* 进行了结构分析和功能预测, 为进一步了解 *PanK* 的作用机制及功能应用奠定了理论基础, 同时为植物 *PanK* 基因的分子研究提供了理论依据和资料。

参考文献

- [1] 陈磊, 唐成林, 程万, 等. 贵州产白及组培瓶苗分级标准及其验证_陈磊 [J]. 贵州农业科学, 2019, 47(10):110-4.
- [2] 陈小璐, 喻韬, 罗蓉. 泛酸激酶相关神经变性病的临床特征及分子生物学发病机制 [J]. 中华医学遗传学杂志, 2019, 36(2): 175-8.
- [3] 万大群, 赵仁全, 刘海, 等. 白及的成分、药理作用和临床应用研究进展. 中国药业, 2017, 26(2):93-96.
- [4] Lim B C, Ki C S, Cho A, et al. Pantothenate kinase [J]. European Journal of Neurology, 2012, 19(4): 556-61.
- [5] Awasthy D, Ambady A, Bhat J, et al. Essentiality and functional analysis of type I and type III pantothenate kinases of Mycobacterium tuberculosis [J]. Microbiology, 2010, 156(9): 2691-2710.
- [6] Yuta O, Hiroki K, Munchiko A, et al. Role of prokaryotic type I and III pantothenate kinases in the coenzyme A biosynthetic pathway of *Bacillus subtilis* [J]. Canadian journal of microbiology, 2014, 60(5): 297-305.
- [7] Kita A, Kishimoto A, Shimosaka T, et al. Crystal structure of pantoate kinase from *Thermococcus kodakarensis* [J].

- Proteins. 2020, 88(5):718-724.
- [8] 杨延辉,赵建新,王浩,等. 泛酸激酶及其抑制剂的研究进展 [J]. 药学进展, 2014, 38: 641-8.
- [9] 沈丹. 泛酸激酶的克隆鉴定及其与泛酸类似物的作用 [J]. 清华大学学报, 2008.03, 48: 403-7.
- [10] 张鹭,王庆忠,徐颖,等. 结核分枝杆菌泛酸激酶的克隆表达及酶学性质 [J]. 第四军医大学学报, 2006, 27(23): 2135-8.
- [11] Arcus V L, Lott J S, Johnston J M, et al. The potential impact of structural genomics on tuberculosis drug discovery [J]. Drug Discovery Today, 2006, 11(1-2): 28-34.
- [12] Ma L Y, Wang L, Yang Y M, et al. Novel gene mutations and clinical features in patients with pantothenate kinase - associated neurodegeneration [J]. Clinical Genetics, 2015, 87(1): 93-5.
- [13] Tanteles G A, Spanou-Aristidou E, Antoniou C, et al. Novel homozygous PANK2 mutation causing atypical pantothenate kinase-associated neurodegeneration (PKAN) in a Cypriot family [J]. Journal of the Neurological Sciences, 2014, 340(1-2): 233-6.
- [14] Kimura Y, Kamimoto T, Tanaka N. Enzymatic Characteristics of a Polyphosphate/ATP-NAD Kinase, PanK, from *Myxococcus xanthus* [J]. Current Microbiology, 2019, 77(2): 173-178.
- [15] Jeong S Y, Hogarth P, Placzek A, et al. 4'-Phosphopantetheine corrects CoA, iron, and dopamine metabolic defects in mammalian models of PKAN [J]. EMBO Molecular Medicine, 2019, 11: 1-17.
- [16] Pan Y, Li L, Xiao S, et al. Callus growth kinetics and accumulation of secondary metabolites of *Bletilla striata* Rchb.f. using a callus suspension culture[J]. PLoS ONE, 2020, 15(2):e0220084.
- [17] Li L, Liu H, Wen W, et al. Full Transcriptome Analysis of Callus Suspension Culture System of *Bletilla striata* [J]. Frontiers in Genetics, 2020, 11: 995.
- [18] Liu H, Wu M, Zhu D, et al. Genome-Wide analysis of the AAAP gene family in moso bamboo (*Phyllostachys edulis*) [J]. BMC Plant Biology, 2017, 17(1): 1-18.
- [19] 郑培义. 酿酒酵母辅酶 A 生物合成蛋白 Cab2 (磷酸泛酰半胱氨酸合成酶) 结构生物学研究[D].中国科学技术大学,2019.
- [20] Temperature-Induced Extended Helix/Random Coil Transitions in a Group I Late Embryogenesis-Abundant Protein from Soybean [J]. Plant Physiology, 2002, 128(3): 822-832.
- [21] Yao J, Subramanian C, Rock CO, et al. Human pantothenate kinase 4is a pseudo-pantothenate kinase [J]. Protein Sci. 2019;28(6):1031-1047.

收稿日期: 2021 年 7 月 20 日

出刊日期: 2021 年 8 月 23 日

引用本文: 李清清, 黄恻隐, 范振源, 刘海, 汪牧非, 徐德林, 白及泛酸激酶基因的生物信息学分析[J]. 细胞与分子生物学研究, 2021, 1(1): 26-32.

DOI: 10.12208/j.ijcmbr.20210006

检索信息: 中国知网 (CNKI Scholar)、万方数据 (WANFANG DATA) 等数据库收录期刊

版权声明: ©2021 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。<http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS