

曲霉来源的 α -L-鼠李糖苷酶结构特征规律分析

蔡雨晨¹, 李孟泽¹, 李利君^{1,2}, 倪辉^{1,2}

(1. 集美大学食品与生物工程学院, 福建厦门 361021)

(2. 福建省食品微生物与酶工程重点实验室, 福建厦门 361021)

摘要: 为了研究所有曲霉来源的 α -L-鼠李糖苷酶的结构特征, 利用 NCBI 数据库收集到 291 条无重复的曲霉来源 α -L-鼠李糖苷酶核酸序列并通过进化树筛选得到 21 条具有代表性的蛋白序列, 通过序列比对分析得到这 21 条序列相互独立具有可靠的代表性, 并利用生物信息学工具对其理化性质、跨膜区、信号肽进行分析发现, 曲霉来源 α -L-鼠李糖苷酶的等电点 (pI) 范围为 4.66~7.17, 在氨基酸数量、分子量、原子总数上波动较大, 其中有 10 条序列拥有信号肽, 1 条序列为两次跨膜蛋白, 21 条序列都为亲水性蛋白; 对其进行进化树构建、三级结构建模及结构叠合发现, 这 21 条代表性序列的蛋白结构可以被分成两大类型, 第一大类型含有 1 个 $(\alpha/\alpha)_6$ 桶状结构和在桶底的 1 个 β 片层结构, 并根据额外含有的 β 片层数量的不同再被分成 4 个小类; 第二大类型拥有 1 个 $(\alpha/\beta)_8$ 结构和环绕在桶装结构域周围的 β 片层结构, 阐明了曲霉来源的 α -L-鼠李糖苷酶的蛋白结构特征, 这有助于更好的明确 α -L-鼠李糖苷酶的共同规律, 为改造该酶提供理论指导。

关键字: α -L-鼠李糖苷酶; 进化树; 代表性序列; 结构特征

文章编号: 1673-9078(2021)01-38-46

DOI: 10.13982/j.mfst.1673-9078.2021.01.0648

Analysis of Structural Characteristics of α -L-rhamnosidase from *Aspergillus* Species

CAI Yu-chen¹, LI Meng-ze¹, LI Li-jun^{1,2}, NI Hui^{1,2}

(1. College of Food and Biological Engineering, Jimei University, Xiamen 361021, China)

(2. Fujian Provincial Key Laboratory of Food Microbiology and Engineering, Xiamen 361021, China)

Abstract: In order to study the structural characteristics of all *Aspergillus*-derived α -L-rhamnosidase, 300 non-repeated *Aspergillus* derived α -L-rhamnosidase nucleic acid sequences were collected using NCBI database and 21 of them were selected through evolutionary tree selection. Through sequence alignment analysis, the 21 sequences are independently and reliably representative. Using bioinformatics tools to analyze the physical and chemical properties, transmembrane region, and signal peptide, it was found that the isoelectric point (pI) of α -L-rhamnosidase derived from *Aspergillus* species was 4.66 to 7.17, and the number of amino acids, molecular weight and total number of atoms fluctuates greatly. Ten sequences have signal peptide sequences, 1 sequence is twice transmembrane protein, and 21 sequences are all hydrophilic proteins. The phylogenetic tree construction, three-level structure modeling and structure superposition show that these 21 representative sequences can be divided into two types. The first type contains one $(\alpha/\alpha)_6$ barrel structure and one β sheet structure at the bottom of the barrel, and can be further divided into 4 sub-categories according to the number of additional β sheets contained; The second type has one $(\alpha/\beta)_8$ structure and β sheet structure surrounding the barrel structure domain. It elucidates the protein structure characteristics of α -L-rhamnosidase derived from *Aspergillus*, which helps to better clarify the common law of α -L-rhamnosidase and provide theoretical guidance for the modification of this enzyme.

引文格式:

蔡雨晨, 李孟泽, 李利君, 等. 曲霉来源的 α -L-鼠李糖苷酶结构特征规律分析[J]. 现代食品科技, 2021, 37(1): 38-46

CAI Yu-chen, LI Meng-ze, LI Li-jun, et al. Analysis of structural characteristics of α -L-rhamnosidase from *Aspergillus* species [J]. Modern Food Science and Technology, 2021, 37(1): 38-46

收稿日期: 2020-07-11

基金项目: 国家自然科学基金项目 (U1805235)

作者简介: 蔡雨晨 (1998-), 女, 硕士研究生, 研究方向: 食品化学与食品生物技术

通讯作者: 李利君 (1973-), 女, 博士, 教授, 研究方向: 食品加工中的生物学问题及相关应用技术研究

Key words: α -L-rhamnosidase; phylogenetic tree; representative sequence; structure characteristics

α -L-鼠李糖苷酶 (α -L-rhamnosidase, EC3.2.1.40) 是一种专一性水解鼠李糖苷键的酶^[1], 近年来, 在食品加工特别是饮料加工中被广泛应用, 如水解柚子中的苦味物质制作柚子汁, 改善酿造酒及果汁的风味等^[2], 此外, α -L-鼠李糖苷酶还可作为食品添加剂改变流体性质^[3]。在自然界中, α -L-鼠李糖苷酶主要来源于细菌^[4]和霉菌^[5], 最初是从青霉菌和曲霉菌代谢生产的酶制剂中纯化得到, 国外关于真菌中 α -L-鼠李糖苷酶基因的克隆研究也主要集中于曲霉来源的 α -L-鼠李糖苷酶^[6]; 在碳水化合物活性酶数据库 (CAZy)^[7]中, α -L-鼠李糖苷酶主要存在于糖苷水解酶第 78 家族 (glucoside hydrolase family 78, GH78), 少量存在于 GH28 和 GH106 家族中, 且来源于真菌的晶体结构有且仅有一个, 即 *Aspergillus terreus* 来源的 α -L-鼠李糖苷酶 (PDB:6gsz)^[8], 因此对曲霉来源的 α -L-鼠李糖苷酶在蛋白结构方面上缺乏系统性研究。

序列比对是解决进化树构建、保守区和保守位点分析等众多问题的开端和基础步骤^[9]。系统进化树能够展示蛋白的系统进化关系, 描述发生或进化顺序, 是系统性分析蛋白或基因序列的重要手段^[10]。同源建模是利用蛋白质的三级结构比一级结构更保守的原理, 使用已经确定结构的模板蛋白对未知结构的蛋白进行三级结构构建^[11]。穿线法建模则利用自然界中蛋白质折叠类型数目是一定的, 且相似性比较低的氨基酸序列可能对应着一致的折叠类型这一原理, 弥补同源建模中必须有相似度较高的模板的这一不足^[12,13]。

通过三级结构建模对蛋白结构进行直观的观察^[14], 并结合结构叠合的方法进一步对蛋白结构进行分类, 可以分析得到结构进化的规律。因此, 本文通过序列比对方法及进化树构建技术对曲霉来源的 α -L-鼠李糖苷酶进行研究, 并利用生物信息学手段对 α -L-鼠李糖苷酶蛋白序列的一级序列和二级结构进行分析, 再采用同源建模与穿线法建模的方法进行三级结构建模, 结合结构叠合的结果 综合分析探索曲霉来源 α -L-鼠李糖苷酶的蛋白结构特征。

1 材料与方法

1.1 实验数据来源

进入美国国家生物信息中心 NCBI 网站 (<http://www.ncbi.nlm.nih.gov/>), 输入检索词“(α -L-rhamnosidase) AND “*Aspergillus*””, 下载 NCBI 核苷酸数据库中曲霉来源 α -L-鼠李糖苷酶核酸序列的

FASTA 文件和 GenBank 文档, 并使用 NCBI Blast+ 2.10.0^[15]筛选出非重复序列。

1.2 分析方法

1.2.1 系统进化树构建

利用 ClustalX 2.0^[16]软件进行核酸和蛋白质的多序列比对; 运用 MEGAX 6.0^[17]软件对得到的多序列比对结果分别构建核酸和蛋白质序列进化树。

1.2.2 蛋白质一级结构和二级结构分析

(1) 通过 ProtParam^[18] (<https://web.expasy.org/protparam/>) 工具预测蛋白的理化性质;

(2) 利用 ProtScale^[19] (<http://www.expasy.org/cgi-bin/protscale.pl>) 工具进行蛋白质疏水性分析;

(3) 利用 TMHMM 2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0>) 工具寻找蛋白质的跨膜区域并进行分析;

(4) 使用 SignalP 5.0^[20] (<http://www.cbs.dtu.dk/services/SignalP/>) 工具进行蛋白的信号肽预测分析。

1.2.3 蛋白质三级结构分析

利用同源建模在线服务器 Swiss-Model^[11] (<https://swissmodel.expasy.org/>), 穿针引线法建模在线服务器 I-TASSER^[12] (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) 和 Phyre2^[13] (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) 进行蛋白的三级结构建模, 并使用工具 Verify3D^[21] (<https://servicesn.mbi.ucla.edu/Verify3D/>) 对模型的质量进行评价分析; 运用 UCSF Chimera 1.14^[22]软件进行蛋白质的三维结构结合叠合和比对。

1.3 数据处理

所有数据都根据实验方法中各个软件及网站收集得来, 最多保留两位小数; 序列比对结果使用 Esprout 3.0^[23] (<http://esprout.ibcp.fr/ESProut/cgi-bin/ESProut.cgi>) 进行美化; 281 条核酸序列进化树使用 ITOL (<https://itol.embl.de/>) 进行美化; 21 条蛋白序列进化树使用 MEGAX 6.0 软件自动生成的图片结果; 蛋白三维结构模型及结构叠合结果使用 UCSF Chimera 1.14 软件对进行渲染。

2 结果与讨论

2.1 曲霉来源 α -L-鼠李糖苷酶的核酸序列分析

2.1.1 核酸序列数据收集

表 1 NCBI 中 291 条曲霉来源 α -L-鼠李糖苷酶核酸序列统计信息Table 1 Sequence statistics of α -L-rhamnosidase from 310 *Aspergillus* species in NCBI

种类	数量	种类	数量
<i>Aspergillus niger</i>	22	<i>Aspergillus nomiae</i> NRRL13137	8
<i>Aspergillus flavus</i>	15	<i>Aspergillus neoniger</i> CBS 115656	8
<i>Aspergillus steynii</i> IBT 23096	14	<i>Aspergillus uvarum</i> CBS 121591	8
<i>Aspergillus mulundensis</i>	14	<i>Aspergillus brunneoviolaceus</i> CBS 621.78	8
<i>Aspergillus oryzae</i> RIB40	12	<i>Aspergillus eucalypticola</i> CBS 122712	8
<i>Aspergillus nidulans</i>	11	<i>Aspergillus sclerotioniger</i> CBS 115572	7
<i>Aspergillus welwitschiae</i>	11	<i>Aspergillus ibericus</i> CBS 121593	7
<i>Aspergillus aculeatus</i>	10	<i>Aspergillus saccharolyticus</i> JOP 1030-1	6
<i>Aspergillus aculeatinus</i> CBS 121060	10	<i>Aspergillus terreus</i>	5
<i>Aspergillus homomorphus</i> CBS 101889	10	<i>Aspergillus glaucus</i> CBS 516.65	3
<i>Aspergillus piperis</i> CBS 112811	10	<i>Aspergillus heteromorphus</i> CBS 117.55	3
<i>Aspergillus thermomutatus</i>	9	<i>Aspergillus tubingensis</i>	3
<i>Aspergillus novofumigatus</i> IBT 16806	9	<i>Aspergillus alliaceus</i>	3
<i>Aspergillus vadensis</i> CBS 113365	9	<i>Aspergillus lentulus</i>	1
<i>Aspergillus japonicus</i> CBS 114.51	9	<i>Aspergillus candidus</i>	1
<i>Aspergillus costaricensis</i> CBS 115574	9	<i>Aspergillus campestris</i> IBT 28561	1
<i>Aspergillus bombycis</i>	9	<i>Aspergillus kawachii</i>	1
<i>Aspergillus fumigatus</i>	8	uncultured <i>Aspergillus</i>	1
<i>Aspergillus fischeri</i> NRRL 181	8		

表 2 21 条代表性核酸序列对应蛋白质登录号

Table 2 The corresponding protein registration numbers of 21 representative nucleic acid sequences

核酸登录号	蛋白质登录号	核酸登录号	蛋白质登录号
XM_001727082.1	XP_001727134.1	XM_026748155.1	XP_026603987.1
FR873475.1	CCB96437.1	XM_026748323.1	XP_026602627.1
XM_001395598.2	XP_001395635.2	XM_026748328.1	XP_026602632.1
XM_001398901.2	XP_001398938.2	XM_026749312.1	XP_026602402.1
XM_002383100.1	XP_002383141.1	XM_026772831.1	XP_026628025.1
XM_002385006.2	XP_002385047.1	XM_655143.1	XP_660235.1
XM_654718.1	XP_659810.1	XM_659441.1	XP_664533.1
XM_022542231.1	XP_022403539.1	XM_676642.1	XP_681734.1
XM_022538567.1	XP_022383582.1	XM_743517.1	XP_748610.1
XM_025655360.1	XP_025515427.1	XM_744823.1	XP_749916.1
XM_026747695.1	XP_026603527.1		

在 NCBI 核苷酸数据库中, 检索到 291 条曲霉来源的 α -L-鼠李糖苷酶核酸序列, 经过统计将它们分成了 37 个种类 (表 1)。发现几乎所有种类的曲霉都有分离纯化得到 α -L-鼠李糖苷酶的记录。

2.1.2 核酸系统进化树构建及代表性序列筛选

筛除掉 10 条完全重复序列, 最终得到 281 条未重复曲霉来源的 α -L-鼠李糖苷酶核酸序列, 然后对 281 条核酸序列建立进化树 (图 1), 并根据系统进化树筛选出 21 条代表性的核酸序列 (图 1 中的橙色标识), 并得到了对应的蛋白质序列 (表 2)。

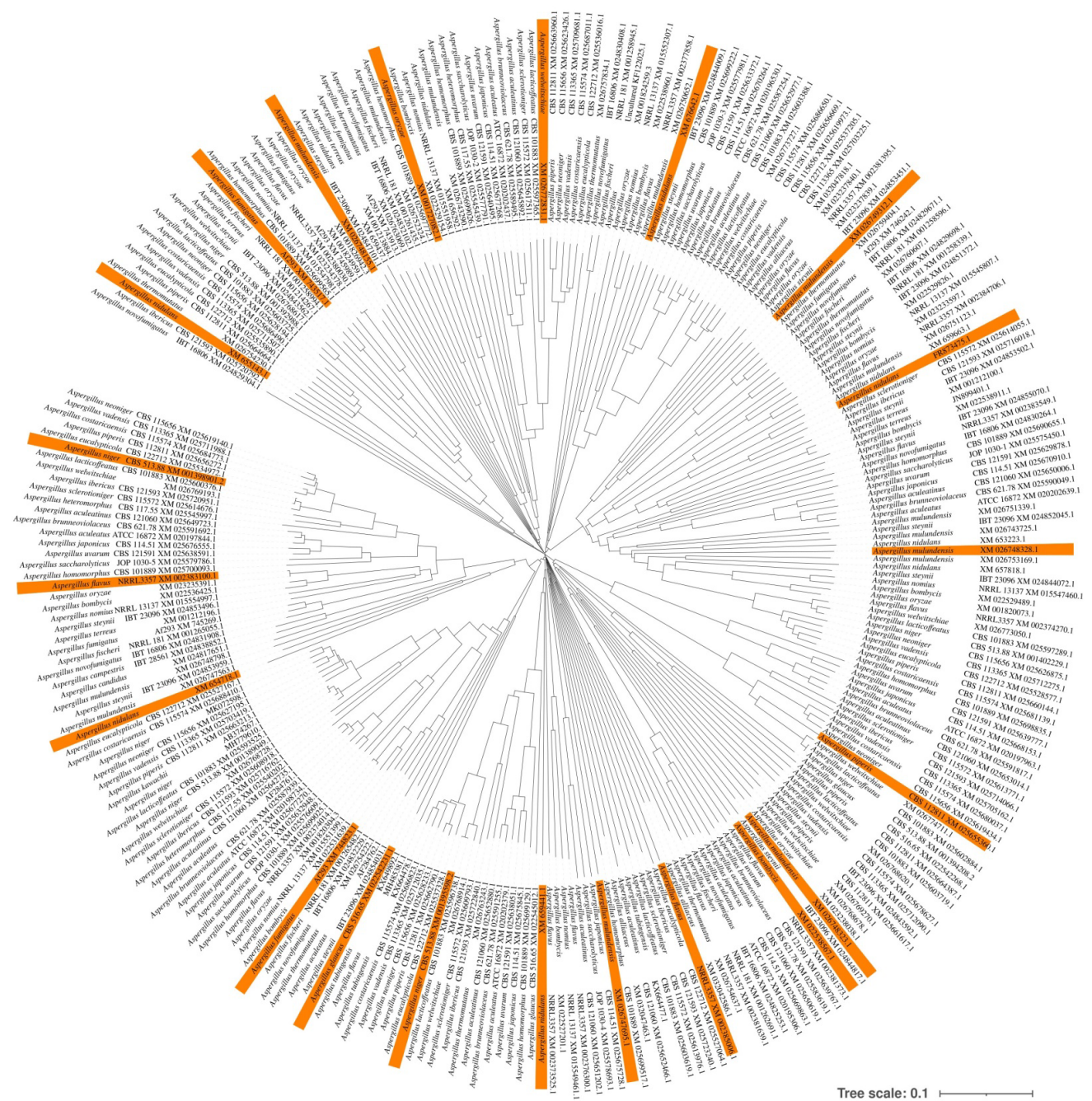


图1 281 条曲霉来源 α -L-鼠李糖苷酶核酸序列构建的系统进化树

Fig.1 Phylogenetic tree of 281 *Aspergillus* derived α -L-rhamnosidase nucleic acid sequences

2.2 曲霉来源 α -L-鼠李糖苷酶的蛋白序列分析

2.2.1 多序列比对分析

多序列比对结果显示了 α -L-鼠李糖苷酶蛋白质序列的保守氨基酸位点 (图2), 发现这 21 条 α -L-鼠李糖苷酶代表序列的氨基酸保守位点较为分散且仅有 22 个较保守位点, 分别位于 285、359、392、507、510、574、595、596、626、631、634、648、644、651、678、681、684、685、825、831、844 和 852 位, 推测这些氨基酸是与保持结构或催化功能密切相关的极其重要

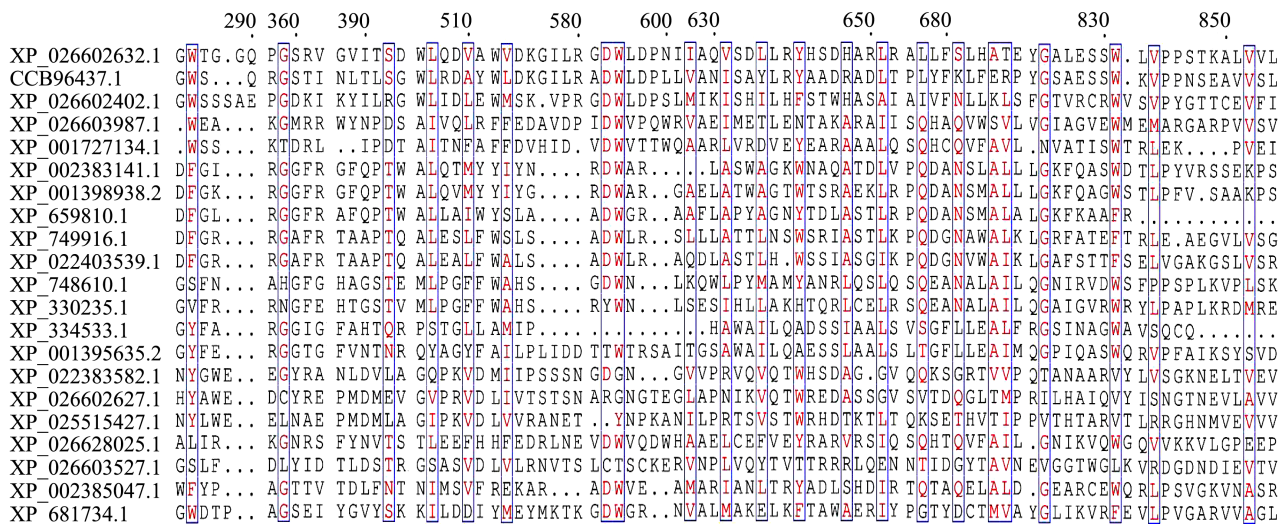
关键性氨基酸, 且并没有发现有非常保守的位点和保守区存在, 说明选择的这 21 条序列相互独立, 可以通过对这 21 条代表性序列的分析可以基本概括出所有曲霉来源 α -L-鼠李糖苷酶的相关规律。

2.2.2 理化性质分析

对 21 条代表性蛋白质序列进行理化性质分析(表3) 发现它们氨基酸数量范围为 556~1032 个, 分子量 (Mr) 的极差为 37606 u, 原子总数的极差为 5225 个, 在氨基酸数量、分子量、原子总数上波动较大。对负电荷残基总数(Asp+Glu)和正电荷残基总数(Arg+lys)

进行比较,发现负电荷残基总数略多,大多带负电荷。理论等电点 (pI) 的范围是 4.66~7.17,除了 XP_660235.1 的 pI 为 7.13、XP_664533.1 为 7.17,是弱碱性蛋白质;其余 α -L-鼠李糖苷酶蛋白质的理论等电点都小于 7,属于酸性蛋白质,与张霞^[6]总结的关于

真菌来源的 α -L-鼠李糖苷酶的 pI 实验值一致; α 螺旋指数的范围是 71.36~86.19;亲水性平均系数的范围是 -0.355 到-0.051,说明曲霉来源的 α -L-鼠李糖苷酶为亲水性蛋白^[24]。



2.2.3 疏水性和跨膜区分析

氨基酸的疏水性反映 α -L-鼠李糖苷酶的折叠情况,在潜在的跨膜区域会出现疏水区。对亲水性平均系数比较小的 XP_660235.1 蛋白进行疏水性分析和跨膜区预测。使用 ProtScale 得到了对 XP_660235.1 蛋白的分析结果(图 3)。发现 XP_660235.1 在 100~200 之间有明显的两个疏水峰,是潜在的跨膜区域^[25]。

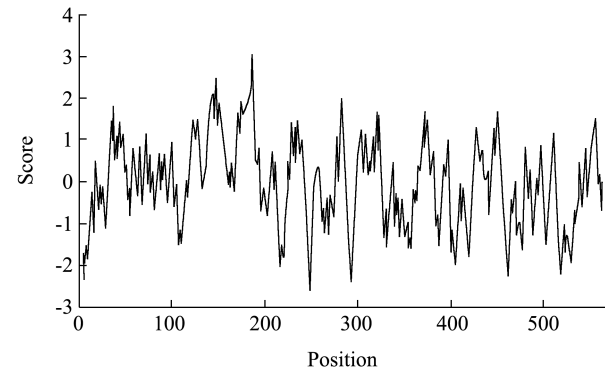


图 3 XP_660235.1 蛋白的疏水性预测图

Fig.3 Hydrophobicity prediction map of XP_660235.1 protein

注:以 0 为基准,在 0 以下的峰为亲水峰和 0 以上的峰则为疏水峰。

使用 TMHMM 工具预测 XP_660235.1 蛋白的跨膜区,发现 136 到 158 位和 171 到 193 位拥有跨膜螺旋区的可能性接近 1(图 4),说明 XP_660235.1 蛋白有两部分位于细胞膜表面,分别是氨基酸序列的 1 到 135 位和 194 到 568 位,XP_660235.1 为两次跨膜的蛋白质,这与 ProtScale 预测结果一致。对剩余的 20 条蛋白质序列进行预测,结果显示无跨膜区存在。

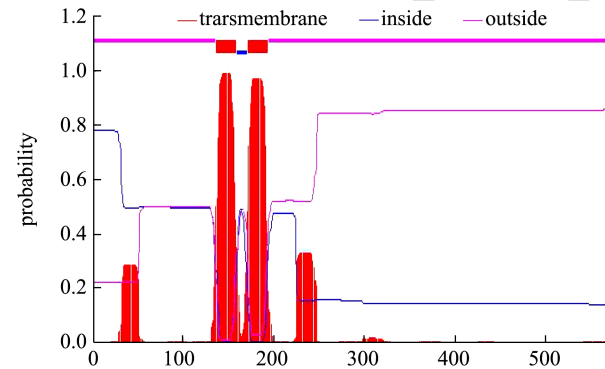


图 4 XP_660235.1 蛋白的跨膜区预测图

Fig.4 Prediction map of transmembrane region of XP_660235.1 protein

注:纵坐标代表可能性,红色峰代表预测的跨膜区。

2.2.4 蛋白进化树分析

对这 21 条曲霉来源 α -L-鼠李糖苷酶蛋白质序列进行系统进化树构建分析(图 5),根据进化树结果可以将 21 条代表性序列分为两组,第一组包含 17 个序列为 XP_660235.1、XP_664533.1、XP_748610.1、

XP_749916.1、XP_001395635.2、XP_001398938.2、XP_002383141.1、XP_002385047.1、XP_022403539.1、XP_001727134.1、CCB96437.1、XP_681734.1、XP_659810.1、XP_026602402.1、XP_026602632.1、XP_026603987.1、XP_026628025.1,第二组包含 4 个序列为 XP_022383582.1、XP_025515427.1、XP_026602627.1、XP_026603527.1。

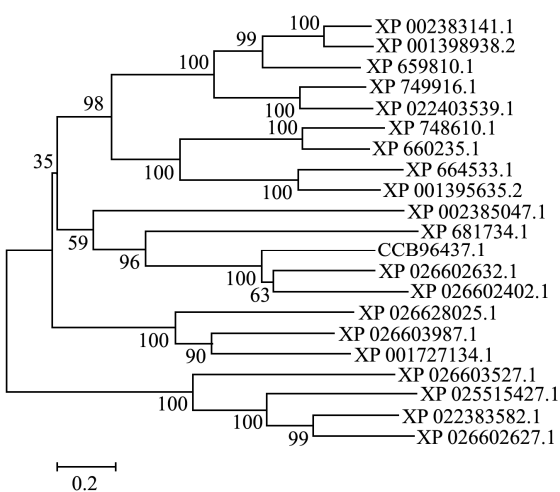


图 5 21 条曲霉来源 α -L-鼠李糖苷酶蛋白质序列构建的邻接法进化树

Fig.5 Phylogenetic tree of 21 *Aspergillus* derived α -L-rhamnosidase protein sequences

注:分点附近的数值为 bootstrap 值,代表进化树分支的可信度;左下标尺表示该长度代表的遗传距离。

2.2.5 信号肽预测分析

表 4 信号肽预测结果

Table 4 Prediction results of signal peptide

蛋白质登录号	可能性	裂解位点	裂解位点	可能性
XP_001395635.2	0.98	22 and 23	IGA-QR	0.50
XP_001398938.2	0.78	22 and 23	VCG-TD	0.19
XP_022403539.1	0.98	19 and 20	SQA-GK	0.85
XP_002383141.1	0.99	25 and 26	ARA-AK	0.96
XP_022383582.1	1.00	19 and 20	GIA-AD	0.83
XP_025515427.1	0.85	19 and 20	VWS-VG	0.53
XP_659810.1	0.99	18 and 19	ASA-GR	0.82
XP_026602627.1	0.99	19 and 20	GIA-FP	0.50
XP_026603527.1	0.78	22 and 23	IVA-SK	0.24
XP_748610.1	0.95	20 and 21	GRA-TI	0.54

在 SignalP 5.0 的预测结果中(表 4),XP_002383141.1、XP_022403539.1、XP_659810.1、XP_001395635.2、XP_022383582.1、XP_748610.1、XP_001398938.2、XP_025515427.1、XP_026602627.1 和 XP_026603527.1 共 10 条 α -L-鼠李糖苷酶蛋白序列含有信号肽。与进化树分类结果结合发现,含有信号

肽的序列均匀的分布在第一组与第二大组中,说明曲霉来源的 α -L-鼠李糖苷酶存在胞外酶与胞内酶两种,且信号肽的有无不能反映出曲霉来源 α -L-鼠李糖苷酶的进化规律。

2.3 曲霉来源 α -L-鼠李糖苷酶的蛋白三级结构分析

2.3.1 三级结构建模分析

表5 Verify3D 建模评分表

Table 5 Verify3D modeling scoring table

蛋白质登录号	得分比/%	蛋白质登录号	得分比/%
XP_001395635.2	88.1	XP_002385047.1	88.4
XP_001727134.1	80.57	XP_022403539.1	81.7
XP_660235.1	86.4	XP_659810.1	84.45
XP_748610.1	83.2	XP_026602402.1	88.18
XP_022383582.1	84.3	XP_026602627.1	82.5
XP_025515427.1	82.11	XP_026602632.1	98.14
CCB96437.1	99.07	XP_026603527.1	83.5
XP_001398938.2	81.5	XP_026603987.1	80.0
XP_002383141.1	88.16	XP_026628025.1	82.66
XP_749916.1	85.97	XP_660235.1	82.2
XP_681734.1	84.2		

对21条曲霉来源 α -L-鼠李糖苷酶的蛋白质序列进行三维建模,其中5条采用同源建模法,另外16条采用穿针引线法进行建模。对所有蛋白质三级结构建模结果进行Verify3D评价。从结果(表5)中可以看到蛋白质序列的得分比最小值为80.0%,说明 α -L-鼠李糖苷酶的蛋白质序列的建模结果良好。

2.3.2 结构叠合分析

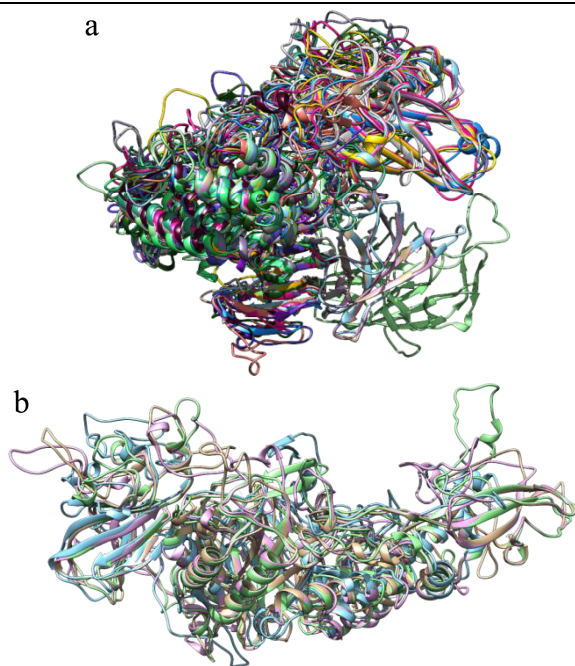
结构叠合的分类结果显示(表6),这21个曲霉

表6 结构分类表

Table 6 Structure classification table

蛋白质登录号	所属类别	蛋白质登录号	所属类别
XP_664533.1	第一大类第一小类	XP_748610.1	第一大类第三小类
XP_660235.1	第一大类第一小类	XP_002385047.1	第一大类第三小类
XP_022403539.1	第一大类第二小类	XP_681743.1	第一大类第四小类
XP_749916.1	第一大类第二小类	XP_026602402.1	第一大类第四小类
XP_001395635.2	第一大类第二小类	XP_026602632.1	第一大类第四小类
XP_001398938.2	第一大类第二小类	CCB96437.1	第一大类第四小类
XP_002383141.1	第一大类第二小类	XP_022383582.1	第二大类
XP_659810.1	第一大类第三小类	XP_026602627.1	第二大类
XP_001727134.1	第一大类第三小类	XP_025515427.1	第二大类
XP_026603987.1	第一大类第三小类	XP_026603527.1	第二大类
XP_026628025.1	第一大类第三小类		

来源 α -L-鼠李糖苷酶可分为两大类(图6), XP_681743.1、XP_664533.1、XP_660235.1、XP_026602402.1、XP_022403539.1、XP_659810.1、XP_001727134.1、XP_026602632.1、XP_026603987.1、XP_026628025.1、XP_748610.1、XP_749916.1、XP_001395635.2、XP_001398938.2、XP_002383141.1、XP_002385047.1与CCB96437.1共17条序列组成第一大类,每条序列都拥有一个 $(\alpha/\alpha)_6$ 桶状结构和桶底的一个 β 折叠结构这一基本结构,根据除基本结构外的 β 片层结构的数量,又可以将第一大类分为4个小类(图7),其中XP_664533.1与XP_660235.1组成无额外 β 片层的第一个小类;XP_022383582.1、XP_749916.1、XP_001395635.2、XP_001398938.2与XP_002383141.1共5条序列组成有1个额外 β 片层的第二小类;XP_659810.1、XP_001727134.1、XP_026603987.1、XP_026628025.1、XP_748610.1与XP_002385047.1共6条序列组成有2个额外 β 折叠的第三小类;XP_681743.1、XP_026602402.1、XP_026602632.1与CCB96437.1共4条序列组成有3个额外 β 折叠的第四小类,目前已报道的曲霉来源的晶体结构(PDB:6gsz)^[8]就属于这一类;XP_022383582.1、XP_026602627.1、XP_025515427.1与XP_026603527.1共4条序列组成第二大类,与第一大类不同,第二大类4个序列都属于GH106家族,因此拥有GH106的基本结构 $(\alpha/\beta)_8$ 结构和环绕在桶装结构域周围的 β 折叠结构^[26]。第一大类与第二大类的分类结果与进化树分类相符,说明蛋白的进化规律会一定程度的体现在其三级结构上;第一大类的四个小类与进化树分类不符,说明 β 折叠结构的数量并不能反映曲霉来源 α -L-鼠李糖苷酶的进化规律。

图6 α -L-鼠李糖苷酶的两个大类叠合图Fig.6 Two kinds of superposition graphs of α -L-rhamnosidase

注: a: 第一大类叠合情况; b: 第二大类叠合情况。

图7 α -L-鼠李糖苷酶的第一大类的四个小类叠合图Fig.7 Four subclasses of the first class of α -L-rhamnosidase

注: a: 第一大类第一小类叠合情况; b: 第一大类第二小类叠合情况; c: 第一大类第三小类叠合情况; d: 第一大类第四小类叠合情况。

3 结语

本文通过 NCBI 数据库, 收集了所有非重复的曲霉来源 α -L-鼠李糖苷酶的核酸数据共 291 条, 并通过进化树筛选得到具有代表性的 21 条序列并预测了蛋白理化性质。通过对这 21 条代表性序列进行序列比对和进化树构建发现, 虽然这 21 条序列具有非常少的保守位点, 但是它们之间依然存在进化规律, 且根据这一进化规律, 可以将这些序列分为两组; 信号肽分析结果显示有 10 条序列含有信号肽, 说明曲霉来源的 α -L-鼠李糖苷酶有胞外酶与胞内酶两种; 跨膜区分析发现 1 条来源于 *Aspergillus nidulans* 的 α -L-鼠李糖苷酶为二次跨膜蛋白; 通过蛋白的三级结构建模及叠合, 发现 21 条 α -L-鼠李糖苷酶主要存在于 GH78 与 GH106 家族中, 将 21 条 α -L-鼠李糖苷酶分为两个类型, 第一大类都含有一个 $(\alpha/\alpha)_6$ 桶状结构和桶底的一个 β 片层结构, 并根据额外含有的 β 片层结构的数量进一步分成 4 个小类; 第二大类则含有 1 个 $(\alpha/\beta)_8$ 结构和环绕在桶装结构域周围的 β 片层结构, 且结构叠合分类与进化树的分类一致, 说明蛋白的进化规律会一定程度的体现在其三级结构上, 而小类的分类结果说明 β 折叠结构的数量并不能作为说明曲霉来源 α -L-鼠李糖苷酶进化规律的依据。本文通过筛选出 21 条代表性序列阐明曲霉来源的 α -L-鼠李糖苷酶蛋白序列性质及结构特征, 系统性的分析了其结构规律, 为该酶的定向进化和分子改造提供了强有力的参考。

参考文献

- [1] Vinita Y, Pramod K Y, Sarita Y, et al. α -L-rhamnosidase: a review [J]. Process Biochemistry, 2010, 45(8): 1226-1235
- [2] Stinco C M, Fernandez-Vazquez R, Hernanz D, et al. Industrial orange juice debittering: Impact on bioactive

- compounds and nutritional value [J]. Journal of Food Engineering, 2013, 116(1): 155-161
- [3] 朱运平, 郝梦露, 白雪, 等. 柚苷酶的生产及其在食品工业中的研究进展[J]. 中国食品添加剂, 2015, 6: 155-160
ZHU Yun-pin, XI Meng-lu, BAI Xue, et al. The production of naringinase and its research progress in food industry [J]. China Food Additives, 2015, 6: 155-160
- [4] Federica D L, Francesca M, Vincenzo T, et al. RHA-P: isolation, expression and characterization of a bacterial α -L-rhamnosidase from *Novosphingobium* sp. PP1Y [J]. Journal of Molecular Catalysis B: Enzymatic, 2016, 134: 136-147
- [5] Ge L, Chen A, Pei J, et al. Enhancing the thermostability of α -L-rhamnosidase from *Aspergillus terreus* and the enzymatic conversion of rutin to isoquercitrin by adding sorbitol [J]. BMC Biotechnology, 2017, 17(1): 21
- [6] 张霞, 李利君, 倪辉, 等. 微生物来源 α -L-鼠李糖苷酶的分子和结构生物学研究进展[J]. 生命科学研究, 2015, 19(1): 68-74
ZHANG Xia, LI Li-jun, NI Hui, et al. Progresses on molecular biology and structural biology of α -L-rhamnosidase from microbial sources [J]. Life Science Research, 2015, 19(1): 68-74
- [7] Brandi L C. The Carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics [J]. Nucleic Acids Research, 2009, 37(S1): D233-D238
- [8] Pachi P, Škerlová J, Šimčíková D, et al. Crystal structure of native α -L-rhamnosidase from *Aspergillus terreus* [J]. Acta Crystallographica Section D Structural Biology, 2018, 11: 1078-1084
- [9] 邓小燕, 徐胜超. 基于 Yarn 云平台的生物基因多序列比对并行算法[J]. 基因组学与应用生物学, 2019, 38(7): 3009-3015
DENG Xiao-yan, XU Sheng-chao. Parallel algorithm based yarn cloud platform for genetic multi-sequence alignment [J]. Genomics and Applied Biology, 2019, 38(7): 3009-3015
- [10] Scott J G, Maini P K, Anderson A R A, et al. Inferring tumor proliferative organization from phylogenetic tree measures in a computational model [J]. Systematic Biology, 2018, 69(4): 623-637
- [11] Andrew W, Martino B, Stefan B, et al. Swiss-model: homology modelling of protein structures and complexes [J]. Nucleic Acids Research, 2018, 46(1): 296-303
- [12] Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions [J]. Nucleic Acids Research, 2015, 43: 174-181
- [13] Kelley L A, Mezulis S, Yates C M, et al. The Phyre2 web portal for protein modeling, prediction and analysis [J]. Nature Protocols, 2015, 10(6): 845-858
- [14] 熊伟, 张晓娟, 张海洋, 等. 基于生物信息学方法预测人线粒体转录终止因子 3 蛋白的结构与功能[J]. 生物技术通讯, 2015, 26(3): 367-373
XIONG Wei, ZHANG Xiao-Juan, ZHANG Hai-Yang, et al. Analysis of the structure and function of human mitochondrial transcription termination factor 3 based on bioinformatic methods [J]. Letters in Biotechnology, 2015, 26(3): 367-373
- [15] Christiam C, George C, Vahram A, et al. Blast⁺: architecture and applications. BMC bioinformatics 10:421 [J]. BMC Bioinformatics, 2009, 10(1): 421-430
- [16] Thompson J D, Gibson T J, Frédéric P. The clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools [J]. Nucleic Acids Research, 1997, 25(24): 4876-4882
- [17] Tamura K, Stecher G, Peterson D, et al. MEGA6: molecular evolutionary genetics analysis version 6.0 [J]. Molecular Biology and Evolution, 2013, 30(12): 2725-2729
- [18] Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the Expasy server [J]. Methods in Molecular Biology, 1999, 112: 531-552
- [19] Madeira F, Park Y M, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019 [J]. Nucleic Acids Research, 2019, 47(1): 636-641
- [20] José J A, Armenteros K D, Tsirigos C K S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks [J]. Nature Biotechnology, 2019, 37: 420-423
- [21] Bowie J, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure [J]. Science, 1991, 253(5076): 164-170
- [22] Pettersen E F, Goddard T D, Huang C C, et al. UCSF chimera -a visualization system for exploratory research and analysis [J]. Journal of Computational Chemistry, 2004, 25(13): 1605-1612
- [23] Xavier, Robert, Patrice, et al. Deciphering key features in protein structures with the new END script server [J]. Nucleic Acids Research, 2014, 42(1): 320-324
- [24] 胡胜民, 王钰, 贺福初, 等. 蛋白质超家族分子进化研究的一种新方法[J]. 科学通报, 2000, 45(21): 2310-2315
HU Shengmin, WANG Yu, HE Fuqie, et al. A new method for molecular evolution of protein superfamily [J]. Chinese Science Bulletin, 2000, 45(21): 2310-2315