分类号：　　密　　级：公开

学　号：20212006054　　单位代码：10759

石河子大学

硕 士 学 位 论 文

马盲肠微生物新型碳水化合物活性酶的发现及功能分析

学 位 申 请 人　胡灵灵

指 导 教 师　倪伟 胡圣伟 李晓悦

申请学位门类级别　理学 硕士

学科、专业名称　生物学

研 究 方 向　生物化学与分子生物学

所 在 学 院　生命科学学院

中国·新疆·石河子

2024 年 6 月

分类号：　　密　　级：公开

学　号：20212006054　　单位代码：10759

石河子大学

硕士学位论文

马盲肠微生物新型碳水化合物活性酶的发现及功能分析

学位申请人	胡灵灵
指导教师	倪伟　胡圣伟　李晓悦
申请学位门类级别	理学 硕士
学科、专业名称	生物学
研究方向	生物化学与分子生物学
所在学院	生命科学学院

中国·新疆·石河子

2024 年　6 月

Discovery and functional analysis of novel carbohydrate-active enzymes in horse cecum microbiota

A Dissertation Submitted to

In Partial Fulfillment of the Requirements
for the Degree of
Master of Science

By

(Biochemistry and Molecular Biology)

Dissertation Supervisor:Prof. Ni Wei
Prof. Hu Shengwei
Prof. Li Xiaoyue

June, 2024

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所呈交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：胡灵灵　　时间： 2024 年 5 月 15 日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：胡灵灵　　时间：2024 年 5 月 15 日

导师签名：　　时间：2024 年 5 月 15 日

摘要

目的：碳水化合物活性酶（CAZymes）能够将复杂的多糖分解成更简单的形式。动物肠道微生物是碳水化合物酶的重要资源库，深入研究动物肠道微生物的基因组并从中挖掘出能够高效降解纤维素的新型碳水化合物活性酶具有重要的科研价值和现实生产意义。因此，本研究的目的是从马盲肠脆弱拟杆菌的 MAG117.bin.13 基因组中发现新型碳水化合物活性酶。应用生物信息学手段，挖掘出属于 GH 家族（Glycosidehydrolase, GHs）的新型多功能 CAZymes，能高效水解寡糖和多糖。从而为探索肠道微生物群编码的新型 CAZymes 提供参考，为开发新型工业酶和饲料添加剂奠定基础。

方法：（1）基于本实验室已有的研究基础，从马盲肠脆弱拟杆菌 MAG117.bin.13 基因组中以序列相似性、功能相似性及酶催化协同作用的特性为依据筛选得到假设蛋白。（2）采用 ProtParam、ProtScale、Signal P 3.0、TMHMM、NetNGlyc 1.0、NetPhos 2.0、SOPMA 和 SWISS- MODEL 软件对假设蛋白进行了理化性质分析、信号肽预测、跨膜结构及亲/疏水性预测、糖基化及磷酸化位点预测、二级结构和三级结构预测。（3）将假设蛋白送至上海生工进行合成，从克隆菌株中提取重组质粒，经 *EcoR*I、*Ba*m*H*I 双酶切鉴定后将测序正确的重组质粒转化到大肠杆菌菌株 BL21（DE3）中进行表达条件的优化及目的蛋白的纯化，而后使用 12%的聚丙烯酰胺凝胶电泳（SDS-PAGE）鉴定。（4）制作葡萄糖及 N-乙酰-D-葡糖胺标准曲线，采用二硝基水杨酸（DNS）法检测酶活力，将所得值与标准曲线进行比较，计算酶比活力，评价假设蛋白对一系列底物的催化降解能力。（5）将假设蛋白与马肠道宏基因组和 NR 数据库进行比对，以评价其家族成员组成的数量及序列多样性。

结果：（1）通过分析马盲肠脆弱拟杆菌的 MAG117.bin.13 基因组，研究了该基因组中 CAZymes 数量和种类的多样性。筛选得到了 18 个假设蛋白并从中确定到了 BfLac2275（命名为 BfLac2275）。（2）生物信息学分析表明 BfLac2275 在 N 端有一个由 19 个氨基酸残基组成的潜在信号序列且是一个非跨膜亲水蛋白。同时，三级结构预测结果表明 BfLac2275 蛋白中存在一个近似环形的凹槽，该凹槽可能是底物结合的活性位点。（3）将 BfLac2275 在大肠杆菌中诱导表达并优化条件，结果表明诱导重组蛋白 BfLac2275 表达的最适诱导浓度、温度和时间分别为 0.8 mmol/L、37 ℃及 16 h。（4）测定了 BfLac2275 对一系列底物的水解活性，结果发现 BfLac2275 对α-乳糖（156.94 U/mg)、D-（+）-麦芽糖（92.59 U/mg)、棉子糖（86.81 U/mg）和透明质酸钠（5.71 U/mg）具有催化活性，表现出了多功能酶的特征。通过测定金属离子对 BfLac2275 降解各种底物活性的影响，结果发现 Na^{2+}、Fe^{2+} 增强了 BfLac2275 对α-乳糖的相对活性。K^{2+}、Na^{2+}和 Mn^{2+}的加入使 BfLac2275 对 D-(+)-麦芽糖和棉子糖的相对活性大幅增加。并且添加单糖后，5mM 葡萄糖对 BfLac2275 降解 D-(+)-麦芽糖的相对活性有所增加。（5）从马肠道 MAGs 中预测到了 36 个 BfLac2275 的同源序列，同时，在 NR 数据库中预测到了 3488 个的同源序列。

结论：本试验从马盲肠脆弱拟杆菌 MAG117.bin.13 基因组中筛选到了碳水化合物活性酶假设蛋白 BfLac2275，通过生物信息学手段预测其理化性质及结构，经异源表达及纯化后验证其酶学特征。结果发现 BfLac2275 是一个能降解多种底物的新型多功能碳水化合物活性酶。本研究不仅为微生物编码的新型碳水化合物活性酶的挖掘提供了参考，也为深入研究肠道宏基因组假设蛋白的功能注释提供了范例。

关键词：新型碳水化合物活性酶；肠道微生物；原核表达；活性分析；脆弱拟杆菌

Abstract

Object: Carbohydrate-active enzymes (CAZymes) are capable of decomposing complex polysaccharides into simpler forms. Animal gut microorganisms are an important resource pool of carbohydrate enzymes, and it is of great scientific value and practical production significance to study the genomes of animal gut microorganisms in depth and to mine novel carbohydrate-active enzymes from them that can efficiently degrade cellulose. Therefore, the aim of this study was to discover novel carbohydrate-active enzymes from the MAG117.bin.13 genome of *Bacteroides fragilis* from the equine cecum. Bioinformatics was applied to unearth novel multifunctional CAZymes belonging to the GH family (Glycosidehydrolase, GHs) that efficiently hydrolyze oligosaccharides and polysaccharides. This will provide a reference for exploring novel CAZymes encoded by gut microbiota and lay a foundation for the development of novel industrial enzymes and feed additives.

Methods: (1) Based on the existing research foundation in our laboratory, the hypothetical proteins were screened from the genome of equine caecum *Bacteroides fragilis* MAG117.bin.13 on the basis of sequence similarity, functional similarity, and enzyme-catalyzed synergistic properties. (2) Physicochemical property analysis, signal peptide prediction, transmembrane structure and hydrophilic/hydrophobicity prediction, glycosylation and phosphorylation site prediction of the hypothetical proteins were carried out using ProtParam, ProtScale, Signal P 3.0, TMHMM, NetNGlyc 1.0, NetPhos 2.0, SOPMA and SWISS- MODEL software, secondary and tertiary structure prediction. (3) The hypothetical protein was sent to Shanghai Biotechnology for synthesis, and the recombinant plasmid was extracted from the cloned strain, and the correctly sequenced recombinant plasmid was transformed into *E. coli* strain BL21 (DE3) for optimization of the expression conditions and purification of the target protein by *EcoR*I and *BamH*I, and then identified by 12% polyacrylamide gel electrophoresis (SDS-PAGE). (4) The standard curves of glucose and N-acetyl-D-glucosamine were prepared, and the enzyme activity was detected by the dinitrosalicylic acid (DNS) method, and the obtained values were compared with the standard curves to calculate the specific activity of the enzyme, and to evaluate the catalytic degradation ability of the hypothetical protein to a series of substrates. (5) The hypothetical proteins were compared with the equine intestinal macrogenome and NR database to evaluate the number of family members comprising them and their sequence diversity.

Results: (1) The diversity in the number and species of CAZymes in the MAG117.bin.13 genome of *Bacteroides fragilis* from the equine caecum was investigated by analyzing the genome. Screening yielded 18 hypothetical proteins and identified to BfLac2275 (named BfLac2275) from them. (2) Bioinformatics analysis showed that BfLac2275 has a potential signal sequence consisting of 19 amino acid residues at the N-terminus and is a non-transmembrane hydrophilic protein. Meanwhile, tertiary structure prediction

showed that BfLac2275 has a nearly circular groove in the protein, which may be the active site for substrate binding. (3) BfLac2275 was induced to be expressed in *E. coli* and the conditions were optimized. The results showed that the optimal induction concentration, temperature and time using BfLac2275 were 0.8 mmol/L, 37 ℃ and 16 h. (4) The hydrolytic activity of BfLac2275 on a series of substrates was determined, and the results showed that BfLac2275 was active on α-lactose (156.94 U/mg), D-(+)-maltose (92.59 U/mg), cotton-nut sugar (86.81 U/mg) and sodium hyaluronate (5.71 U/mg) with catalytic activity, exhibiting the characteristics of a multifunctional enzyme. By determining the effects of metal ions on the activity of BfLac2275 in degrading various substrates, it was found that Na^{2+} and Fe^{2+} enhanced the relative activity of BfLac2275 towards α-lactose. the addition of K^{2+}, Na^{2+} and Mn^{2+} increased the relative activity of BfLac2275 towards D-(+)-maltose and raffinose dramatically. And the addition of monosaccharides increased the relative activity of 5 mM glucose on the degradation of D-(+)-maltose by BfLac2275. (5) Homologous sequences of 36 BfLac2275 were predicted from equine intestinal MAGs, while homologous sequences of 3488 were predicted from the NR database.

Conclusion: In this experiment, the carbohydrate-active enzyme hypothetical protein BfLac2275 was screened from the MAG117.bin.13 genome of *Bacteroides fragilis* from the equine appendix, and its physicochemical properties and structure were predicted by bioinformatics, and its enzymatic characteristics were verified after heterologous expression and purification. As a result, BfLac2275 was found to be a novel multifunctional carbohydrate-active enzyme capable of degrading multiple substrates. This study not only provides a reference for the mining of novel carbohydrate-active enzymes encoded by microorganisms, but also provides a paradigm for in-depth study of the functional annotation of hypothetical proteins in the intestinal macrogenome.

Key words: novel carbohydrate-active enzymes; gut microorganisms; prokaryotic expression; activity analysis; Bacteroides fragilis

目录

中英文缩略词对照表

缩写词	英文	中文
CAZymes	carbohydrate-active enzymes	碳水化合物活性酶
MAGs	metagenome-assembled genomes	元基因组组装基因组
PULs	Polysaccharide utilization sites	多糖利用位点
CAZy	carbohydrate-active enzymes database	碳水化合物活性酶数据库
GHs	glycoside hydrolases	糖苷水解酶
GTs	glycosyl transferases	糖基转移酶
PLs	polysaccharide lyases	多糖裂解酶
CEs	carbohydrate esterases	碳水化合物酯酶
AAs	auxiliary enzyme families	辅助酶家族
CBMs	carbohydrate binding modules	碳水化合物结合模块
SCFA	short-chain fatty acids	短链脂肪酸
BLASTP	Protein Basic Local Alignment Search Tool	蛋白质基本局部比对搜索工具
NCBI	National Center for Biotechnology Information	国家生物技术信息中心
Mw	molecular weight	分子量
pI	isoelectric point	等电点
ExPASy	Expert Protein Analysis System	专家蛋白质分析系统
TMHMM	Transmembrane Hidden Markov Model	隐马尔可夫模型
SOPMA	Self-Optimized Prediction Method with Alignment	带对齐功能的自优化预测方法
DNA	deoxyribonucleic acid	脱氧核糖核酸
LB	Luria-Bertani	溶菌肉汤
IPTG	Isopropyl-beta-D-thiogalactopyranoside	异丙基-β-D-硫代半乳糖苷
BCA	Bicinchoninic Acid Assay	蛋白质定量分析
BSA	Bovine serum albumin	牛血清白蛋白
DNS	3,5-dinitrosalicylic	3,5-二硝基水杨酸
SPSS	statistical package for the social sciences	社会科学统计软件包
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis	十二烷基硫酸钠--聚丙烯酰胺凝胶电泳
PBS	Phosphate Buffer Solution	磷酸盐缓冲溶液
NR	Non-Redundant Protein Sequence Database	非冗余蛋白库
ML	maximum likelihood	最大似然法
h	hour	小时
min	minute	分钟
g	gram	克
rpm	Revolutions Per Minute	转每分
bp	base pair	碱基对
Da	Dalton	道尔顿
mol	mole	摩尔
pH	Pondus Hydrogenii	酸碱度
mL	Milliliter	毫升

第 1 章 文献综述

1.1 碳水化合物活性酶概述

1.1.1 碳水化合物活性酶家族的组成

自然界中存在大量的植物、动物和微生物来源的碳水化合物，而肠道微生物在分解这些复杂的聚糖聚合物中扮演着关键角色[1]。由于这种挑战的复杂性，需要有互补特异性的酶联合体来将多糖和聚糖完全转化为单糖，以便进行进一步的代谢[2]。在各种动植物病原真菌和细菌中，碳水化合物活性酶类蛋白（CAZymes）发挥着重要作用，对它们的生长和发育起着关键作用[3]。因此碳水化合物活性酶数据库（Carbohydrate-Active enZYmes Database）以涵盖木质纤维素降解所需要的相关酶类进行分类，主要涉及以下 6 大类，糖苷水解酶（Glycoside Hydrolases，GHs）[4]、糖基转移酶（Glycosyl Transferases，GTs）[5]、多糖裂解酶（Polysaccharide Lyases，PLs）[6]、碳水化合物酯酶(Carbohydrate Esterases，CEs)[7]、辅助酶类家族（Auxiliary Activities，AAs）[8]以及碳水化合物结构域（CarbohydrateBinding Modules，CBMs）[9]。其中糖苷水解酶是种类和数量最丰富的家族，GH 家族主要参与碳水化合物主链的降解[10]。GH 家族广泛应用于各行各业，包括化学、生物燃料、食品、饲料和制药等[11-12]。因此，从自然界探索新型高效的 CAZymes 对于充分利用复杂的多糖和聚糖资源以及开发工业酶和饲料添加剂至关重要。

碳水化合物活性酶是自然界中存在并且分布最广的一类重要有机化合物，是所有生物体维持生命活动所需能量的主要来源[13]。糖苷水解酶家族是一组分布比较广的酶，它能水解两种或多种碳水化合物之间的糖苷键，以基于序列相似性为基础的分类方法可以将不同来源或者不同功能的糖苷水解酶分配到不同的家族中[14]。目前糖苷水解酶已被分为 189 个家族。碳水化合物结合模块是具有碳水化合物结合活性的离散折叠的连续氨基酸序列[15]。近几年来，由于宏基因组测序数据的迅速增长，碳水化合物活性酶数据库正在不断地对所新发现的家族进行排序分类，并对数据库定时更新。其中 CAZy 数据库中种类最多以及生物特性中叙述最详尽的蛋白质类型是糖苷水解蛋白质。第 1 个被发现并报道的纤维素酶家族是 GH5 家族，GH5 家族是 CAZy 库中最大的一个糖苷水解酶家族，由于这个家族第一个被发现的，所以该家族曾经被命名为“纤维素酶家族 A”[16]。

1.1.2 碳水化合物活性酶家族的多糖降解作用

近几年，由于新发现的 CAZymes 被不断报道，极大的丰富了肠道微生物 CAZymes

体系[17-18]。随着宏基因组技术的发展，碳水化合物活性酶的新家族不断被发现挖掘。相关研究表明，结构多糖的酶促分解依赖于特定糖苷水解酶的产生，例如，GH5 家族的成员至少有 20 种不同的酶活性，这些酶的产生对于宿主的健康及代谢至关重要。有研究表明部分肠道微生物 CAZymes 的表达与肥胖等具有相关性[19]。频繁的使用抗生素治疗同样也会影响 GHs 的活性及碳水化合物的代谢[20]。

碳水化合物结合模块通常与碳水化合物活性酶中的催化模块相关联，CBM 有三种结合底物的方式，分别是能够与结晶多糖的平坦表面相互作用、结合聚糖链或小糖分子的末端和与碳水化合物配体结合来促进多糖酶降解的催化[21]。这种结合模块折叠成的特定三维空间结构，具有结合碳水化合物的功能[22]。相关研究表明，碳水化合物结合结构域可以通过结合糖苷水解酶家族的底物，以此来提高作用于底物的催化活性[23]。

肠道微生物群参与多种与健康和福祉相关的代谢和稳态功能。其组成因个体而异，并取决于与宿主和微生物群落相关的因素，这些因素需要适应利用肠道环境中存在的各种营养物质[24]。例如纤维素、半纤维素、淀粉或糖原类物质，此外还有一些能结合细胞表面的多糖，都需要肠道微生物所编码的酶进行催化[25]。除催化结构域外，一些木聚糖酶还含有非催化结构域，CBM 可用于提高它们与不溶性底物的结合能力[26]。这些辅助结构域最被认可的功能是结合多糖，使生物催化剂与其底物紧密和长期接近，允许碳水化合物水解，基于氨基酸的相似性，CBMs 被分为 55 个家族，这些家族在底物特异性方面表现出显著的差异[27]。由于大部分天然多糖的价格比较低廉，将其直接作为填料应用于蛋白纯化或者制备固定化酶方面等都具有巨大的价格优势[28]。碳水化合物活性酶家族在降解同一复杂多糖时，各类 CAZymes 体现出协同催化的特性[13]。当多糖结构越复杂，就需要多种碳水化合物活性酶联合起来共同作用，对于多糖主链的降解主要通过 GH 家族和 PL 家族来完成的；侧链的降解则需要包括木聚糖、葡聚糖及 CE 家族在内的大量水解酶的参与[10]。

纤维素作为自然界中最普遍的植物性多糖，其主要是植物细胞壁的组成成分，通常与半纤维素、木质素和果胶结合在一起，这种复杂顽固的结合方式以及结合程度使得其需要有特定的酶来催化才能被生物体吸收利用，对植物源食品的质地影响很大[29]。这就需要有专门降解纤维素的酶来使其分解为可以被吸收的成分。由于人体消化道内不存在纤维素酶，而纤维素又是一种重要的膳食纤维，对于帮助胃肠道消化有重要作用，它也是自然界中含量最多、分布最广的一种植物性多糖[30]。木质纤维素广泛存在于自然界中，木质纤维素的高效降解需要多种微生物的协同作用，这取决于不同微生物所编码的大量酶[31]。木质素降解主要是一个氧化过程，其中木质素过氧化物酶将聚合物消化成更小的片段，作为一种难以控制的成分，较高的木质素含量对工业用途的产品回收率提出了较高的挑战[32]。同时，纤维素酶的酶促水解可以将木质纤维素水解为可发酵的糖，这是将木质纤维素生物质转化为有价值的产品的前提，而发现并挖掘新的碳水化合物活性酶家

族可以用来克服木质纤维素的顽固性，这是可持续发展和生物基础经济的核心挑战[33]。

CAZymes 在将纤维素、聚糖、淀粉和糖原等复合碳水化合物分解成可以被肠上皮吸收的成分中起着至关重要的作用[34]。人类基因组仅仅只编码大约 17 种 CAZymes 的类型，而大多数在肠道内起作用的 CAZymes 是由肠道中的微生物群编码的，其中一种叫多形拟杆菌，可以编码大约 260 种酶[35]。加大对微生物群所编码的 CAZymes 的研究将有助于更好地理解人类复杂的碳水化合物代谢，以及出错时发生的疾病[36]。过去的研究也显示，不同的个体中的肠道微生物的 CAZyme 谱可能具有不同的碳水化合物代谢能力，这说明肠道中碳水化合物活性酶的数量及种类也可能较为不同[37]。

1.2 拟杆菌属多糖利用的机制概述

在拟杆菌属细菌中，多糖利用位点（Polysaccharide utilization sites，PUL）是共同调节的细菌基因，它们是严格调节的共定位基因簇，编码复杂碳水化合物催化所需的酶和蛋白质集合，可以感知营养物质并实现聚糖消化[38]。PUL 主要编码细胞表面聚糖结合蛋白（SGBP）、TonB 依赖性转运蛋白（TBDT）、CAZymes（最常见的是 GH，但在底物适用的情况下还有 PL 和 CE）和碳水化合物传感器/转录调节因子的补体[39]。PUL 的复杂性通常与其同源底物的复杂性成正比，并且可能包括辅助酶，例如蛋白酶[40]、硫酸酯酶[41-42]和磷酸酶[43]。相关研究表明人类肠道微生物组成员，特别是拟杆菌属，含有许多可利用聚糖并且塑造生态动态的 PUL[44]。拟杆菌属细菌具有广泛的 CAZymes，可以降解种类和结构多样化的聚糖，提供有益于宿主的独特代谢功能[45-47]。不同底物的酶法糖化对于促进各种复杂的微生物群落至关重要，例如海洋、土壤、反刍动物和单胃微生物群。因此，高度特异性的碳水化合物活性酶、识别蛋白和转运蛋白在某些物种的基因组中大量富集，在竞争环境中起着至关重要的作用。

复合碳水化合物以结构和储存多糖的形式存在，是生物圈中代谢可利用碳的最大储存库[48-49]，因此，需要相应的大量特定酶来实现完全糖化和促进初级代谢。肠道微生物群对单胃和反刍动物营养的复杂碳水化合物降解的贡献引起了人们的广泛兴趣[50]。由 Abigail Salyers 及其同事于 1980 年发起的研究鉴定出 8 个基因作为单个基因簇的一部分，统称为淀粉利用系统（Sus），建立了复合碳水化合物利用的新范式[51-52]。PUL 调节通常由以下三种机制之一介导：SusR 传感器/调节剂、胞浆外功能（ECF-σ）因子-抗σ因子对或混合双组分系统（HTCS）。如上所述，*susC-susD* 对是 PUL 的标志，已被用于在关键人类肠道共生体的基因组中定位 PUL 补体，包括 *Bacteroides thetaiotaomicron*（88 个 PUL）[53]、*Bacteroides ovatus*（126 个 PUL）[54]和 *Bacteroides cellulosilyticus* WH2（113 个 PUL）[55]。大规模（宏）基因组方法显然有助于 PUL 的发现，以及预测各种拟杆菌门细菌的代谢潜力。然而，在分子和细胞水平上精细化功能表征对于充分了解 PUL 在

微生物群落中的作用仍然至关重要。最近，一系列高影响力的研究结合了遗传学、酶学、生物物理和结构技术，这种方法的开创性研究描述了几种人类肠道共生拟杆菌属物种对果聚糖和菊粉的差异利用，揭示了每种物种都含有一组连锁特异性酶，这些酶有助于定义营养偏好[56]。

最近，针对其他复杂多糖的 PULs 的综合功能研究已有报道。研究发现，一对来自卵形芽孢杆菌的靶向木聚糖 PUL-XylS 和 PUL-XylL 可编码针对组成和分支不同的单个植物β-木聚糖量身定制的酶[57]。最近，来自卵形芽孢杆菌的半乳甘露聚糖特异性 PUL 的详细遗传、生化和酶结构表征揭示了两种甘露聚糖特异性 SGBP、两种 GH26 内β-甘露聚糖酶和一种 GH36 外 a-半乳糖苷酶在解构该植物细胞壁多糖中的相互作用[58-59]。在环境细菌中，来自约翰逊黄杆菌的复杂几丁质利用位点已被广泛地表征[60]。该系统的显着特征包括两对 SusC/SusD 同源物、由几丁质结合模块隔开的两个 GH18 模块组成的分泌型几丁质酶，以及细胞内氨基葡萄糖-6-磷酸脱氨酶。总而言之，这些研究凸显了基于系统的分析可以为各种生态系统生态学背景下的 PUL 结构功能研究带来相当大的洞察力[61]。对拟杆菌门[62]细菌中多糖利用位点（PULs）的探索和多糖酶氧化裂解的发现在很大程度上推动了新型降解 CAZymes 的发现[63]。

1.3 应用宏基因组学筛选新型碳水化合物活性酶

1.3.1 宏基因组学技术

微生物遍布自然界的每个角落，包括海洋、陆地和动物肠道等。大量的微生物很难通过经典的微生物技术来培养。宏基因组学允许直接从环境生态位中回收遗传物质或者其它样品，而无需任何分离培养技术。目前，宏基因组学作为一种有力工具，其被广泛用作从不可培养的微生物群落中分离和鉴定出具有新型生物催化活性的酶[64]。“宏基因组”的概念在 1998 年被提出，其定义为“the genomes of the total microbiota found in nature”，即自然界中所有微生物基因组的总和[65]。宏基因组学是将样品中的微生物群落作为整体进行研究的一门学科。宏基因组学技术（metagenomicnext-generation sequencing，mNGS）是一种基于二代测序技术的微生物检测技术，它使得研究微生物群落中不可培养的微生物基因组成为了可能[66]。

宏基因组学既描述了一个科学研究领域，也描述了一种技术，这种技术能够对任何环境样品中的微生物群落进行不依赖培养的分析。在过去的 10 到 15 年里，该领域产生的基因组数据呈指数级增长，让研究人员能够无可比拟地接触到“未表征的大多数” [67]。据估计，在许多环境中，超过 99%的微生物尚未培养[68]，而培养的原核生物代表绝大多数仅来自四个门：厚壁菌门、拟杆菌门、放线菌门和变形菌门[69]。宏基因组学提供了一