

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024 年 4 月 4 日 (04.04.2024)



(10) 国际公布号
WO 2024/066461 A1

(51) 国际专利分类号:

C12Q 1/6869 (2018.01) **G16B 10/00** (2019.01)

C12Q 1/04 (2006.01) **G16B 20/00** (2019.01)

PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(21) 国际申请号: PCT/CN2023/099026

(22) 国际申请日: 2023 年 6 月 8 日 (08.06.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:

202211176719.1 2022年9月26日 (26.09.2022) CN

(71) 申请人: 华东理工大学(EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY) [CN/CN]; 中国上海市徐汇区梅陇路 130 号, Shanghai 200237 (CN)。

(72) 发明人: 牟伯中 (MU, Bozhong); 中国上海市徐汇区梅陇路 130 号, Shanghai 200237 (CN)。刘一凡 (LIU, Yifan); 中国上海市徐汇区梅陇路 130 号, Shanghai 200237 (CN)。寿利斌 (SHOU, Libin); 中国上海市徐汇区梅陇路 130 号, Shanghai 200237 (CN)。

(74) 代理人: 上海科盛知识产权代理有限公司等 (SHANGHAI KESHENG INTELLECTUAL PROPERTY AGENCY LTD. et al.); 中国上海市徐汇区零陵路 899 号飞洲国际大厦 18F 座, Shanghai 200030 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA,

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则4.17的声明:

— 发明人资格(细则4.17(iv))

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(54) **Title:** METHOD FOR IDENTIFYING MICROORGANISMS HAVING OIL RESERVOIR FLOODING FUNCTION BASED ON METAGENOMICS AND METATRANSCRIPTOMICS

(54) 发明名称: 基于宏基因组和宏转录组识别油藏驱油功能微生物的方法

(57) **Abstract:** A method for identifying microorganisms having an oil reservoir flooding function based on metagenomics and metatranscriptomics, which comprises the following steps: S1, extracting the total DNA and total RNA from a water sample of an oil reservoir; S2, sequencing the obtained total DNA and total RNA to obtain metagenomic and metatranscriptomic original data of the oil reservoir sample; and S3, analyzing metagenomic and metatranscriptomic results, and identifying microorganisms having a flooding function.

(57) 摘要: 一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法, 包括以下步骤: S1: 自油藏产出水样提取总DNA和总RNA; S2: 对获得的总DNA和总RNA进行测序, 获取油藏样品的宏基因组和宏转录组原始数据; S3: 通过对宏基因组和宏转录组结果进行分析, 识别具有驱油功能的微生物。

WO 2024/066461 A1

基于宏基因组和宏转录组识别油藏驱油功能微生物的方法

技术领域

本发明涉及微生物检测技术领域，尤其是涉及一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法。

背景技术

我国稠油资源量约为 198.7 亿吨，年产量高达 3087 万吨 (2017 年)，已占原油总产量的 16.2%。由于稠油粘度高 (50-10000mPa.s)，在油藏中流动性差，一般以蒸汽热采方法为主，而热采法耗能大、成本高、开采效果差。此外，油藏是一个天然生物反应发生器，同时蕴含了具有各种功能的好氧及厌氧微生物。通过利用微生物来采油的技术绿色环保、成本低，可用于稠油开采，其主要机理是通过微生物在油藏中乳化原油、产气、产表面活性剂、产多糖和降解烃等五方面功能降低稠油粘度、提高水驱效率。

近 20 年来，对油藏环境中微生物的研究已经从最初的纯菌分离培养模式过渡到了依赖测序的分子生物学研究方式。其中，测序手段可以大致分为依赖 PCR 扩增的测序技术和不依赖 PCR 扩增的宏基因组与宏转录组测序技术。前者以 16S rRNA 基因克隆文库方法为代表，在油藏环境样品中已经受到广泛应用，通过设计特异性的引物可以扩增出样品中的相关基因序列，从而在基因的水平上阐述微生物的潜在代谢功能。而后者不需要 PCR 扩增，可以同时测定样品中所有基因的序列信息。因此，采用宏基因组测序分析油藏环境样品可以深入解析样品中潜在的代谢网络，进一步将宏基因组技术和宏转录组技术结合能够得到代谢途径上各个基因的转录水平，从而推断油藏环境下的各种微生物代谢过程。

现有的宏基因组学分析手段 (例如一种基于宏转录组学和宏基因组学的环境中抗生素抗性基因的活性定量及宿主鉴定方法，申请号 202110740585.0) 已经可以根据需求对一些常规环境样品的目标功能基因和重要微生物进行分析。

但是地下油藏作为一个以厌氧条件为主的特殊环境，如果不对样品的采集和提取过程进行针对性的处理，样品中的微生物组成极易受到干扰而发生变化，RNA 也会发生降解，从而导致后续的分析无法获得真正的油藏原位微生物数据。并且油藏中微生物的功能多种多样，其中值得关注研究的种类繁多，单一数据库无法对这些功能进行有效的分析，因此必须结合多个公开数据库以及本地自建数据库才能更全面地注释和分析样品中的这些关键功能。

可见，基于目前的技术空白，亟需开发一种特别针对油藏环境而开发的基于宏基因组学和宏转录组学的微生物识别、分析方法。

发明内容

本发明的目的就是为了解决对现有技术中油藏微生物检测手段可以获得的信息有限的问题，提供了一种联合应用宏基因组和宏转录组准确地识别油藏中微生物和代谢功能的方法。

本发明的目的通过以下技术方案实现：

本发明的目的是提供一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法，包括以下步骤：

S1：自油藏产出水样提取总 DNA 和总 RNA；

S2：对获得的总 DNA 和总 RNA 进行测序，获取油藏样品的宏基因组和宏转录组原始数据；

S3：通过对宏基因组和宏转录组结果进行分析，识别具有驱油功能的微生物。

进一步地，S1 中，在提取总 DNA 和总 RNA 之前，在待提取 RNA 的样品中加入抑制剂以抑制厌氧微生物的 RNA 降解。

进一步地，S2 中，还包括：对油藏样品的宏基因组和宏转录组原始数据进行预处理，得到去除接头和低质量片段的目标数据；

所述预处理的过程包括：

利用 fastp 软件分别对油藏样品的宏基因组和宏转录组中各 DNA 或 RNA 链序列双端的原始数据进行滑窗质量剪裁，同时，根据序列首尾两端的引物信息，利用 cutadapt 软件去除引物，得到质控后的双端序列数据。

进一步地，滑窗质量剪裁的参数为-W 4，-M 20，即滑动窗大小为 4，平均质量值为 20。

进一步地，S3 中，对宏基因组和宏转录组结果进行分析的过程包括：

对质控后的双端序列数据进行组装、分箱并评估质量，去除冗余后提取其中高质量的 MAGs（宏基因组组装基因组）数据集；

根据构建的参考数据库对高质量的 MAGs 组装数据进行注释，识别出具有不同驱油功能的功能基因和 MAGs，并计算对应的测序深度和相对丰度；

对高质量的 MAGs 数据集做进化关系分析，进而做出驱油微生物的群落结构分析；

将宏转录组测序得到的序列短片段质控过滤后与 MAGs 数据对比，计算出各个基因的转录水平。

进一步地，S3 中，组装、分箱并评估质量的过程包括：

组装：使用拼接程序 SPADes 在 Meta 模式下，将质控后的双端序列数据样品短序列拼接成长度不一的 contigs（交替片段产物），然后根据双端测序的信息将不同 contigs 连接成有测序缺口的 scaffolds（骨架序列）；

分箱：采用 bowtie2 软件将质控后的 MAGs 短序列数据信息比对到长序列信息上，获得不同长序列的测序覆盖度信息，进一步同时采用 Maxbin2、Metabat2、CONCOCT 三种 Binning 手段从 MAGs 中分离出优势菌的基因组，并进一步导入 DAS_Tool 程序中进行评估，最终整合并提高不同方法生成基因组的质量；

评估质量：利用 dRep 软件将质控后的 MAGs 中相似度较高的基因组去除冗余，通过 CheckM 工具根据基因组中单拷贝标记基因的有无和数量来估算基因组的完整度和污染度。

进一步地，S3 中，对高质量的 MAGs 组装数据进行注释的过程包括：

使用 Prodigal 程序（预测开放阅读框）将拼接后的长序列翻译成编码蛋白序列（CDS），并提交到 KEGG 数据库，采用 GhostKOALA 工具进行功能性注释并获得代表不同旁系同源亚基的 KO 号；

同时采用本地软件 KofamKOALA 根据各个 KO 旁系同源家族蛋白的隐马可夫模型（HMM）和推荐的置信标准给各个蛋白序列注释 KO 号；

最后采用 EggNOG mapper 2 工具给蛋白序列注释 COG 号，再转换成 KO 号，使得最终每个蛋白质的 KO 号注释采用以下顺序：1) GhostKOALA KO, 2) KofamKOALA KO，3) EggNOG mapper KO。

进一步地，S3 中，所述驱油功能包括烷烃降解、产气、产乳化剂、产表面活性剂、产多糖中的一种或多种。

进一步地，S3 中，识别出具有不同驱油功能的功能基因和 MAGs，并结合 bowtie2 软件对比得到的各序列的测序覆盖度信息计算对应的测序深度和相对丰度的过程包括：

针对氢气还原酶，首先通过本地的氢气还原酶亚组的 HMM 模型比对找出潜在的功能基因蛋白；

之后将潜在的功能基因蛋白序列提交到 HydDB 数据库的在线分析软件进一步划分氢气还原酶的亚型；

针对基因组中潜在的编码次级代谢产物的功能，通过将基因组序列提交至 AntiSMASH 网站，结合不同的工具找到基因组中潜在编码次级代谢产物的基因组，并预测代谢产物的类型；

针对数据库中信息缺少的功能基因，单独构建本地的蛋白序列数据库，通过 BlasP (Blast Protein) 比对方法找到最相似的蛋白序列并进一步分析，所述数据库中信息缺少的功能基因包括厌氧烃降解初始活化基因 AssA、EbdA、AhyA、AncA，和细菌微室蛋白簇基因中的一种或多种。

进一步地，S3 中，对高质量的 MAGs 数据集做进化关系分析的过程包括：

首先将目标序列以及数据库中的相似参比序列下载并合并文件，合并后的序列首先在 MAFFT 上排列整齐，并以 80% 的阈值来选择保守位点，进而使用 IQ-tree (该算法采用最大似然法构建系统发育树) 来两两比对序列并生成最大似然进化树。

进一步地，S3 中，将宏转录组测序得到的序列短片段质控过滤后与 MAGs 数据对比，计算出各个基因的转录水平的过程包括：

采用 bowtie2 软件将高质量的 MAGs 数据集中的 cDNA 短片段比对到通过 S3 中组装分箱后的宏基因组拼接得到的 DNA 长片段上，计算出各个基因

的转录水平，以 TPM 值（Transcripts Per Million）来表示。

与现有技术相比，本技术方案的优势在于：

本技术方案是一种特别针对油藏环境而开发的基于宏基因组学和宏转录组学的微生物分析方法，整体过程中，自油藏产出水样提取总 DNA 和总 RNA，测序获取油藏样品的宏基因组和宏转录组原始数据，通过对宏基因组和宏转录组结果进行分析，识别具有驱油功能的微生物，整体过程不依赖传统的速度较慢的微生物单菌分离鉴定手段，适合处理未知物种较多的样本，且可检测极低丰度物种，检测全面、快速。

具体实施方式

下面结合具体实施例对本发明进行详细说明，但绝不是对本发明的限制。本技术方案中如未明确说明的软件/程序名称、控制方法、算法等特征，均视为现有技术中公开的常见技术特征。

本发明的油藏驱油功能微生物识别方法基于宏基因组和宏转录组测序进行检测，具体包括如下步骤：

步骤 s1：收集环境样品，在样品中加入 RNALater 试剂抑制 RNA 降解，然后分别用 DNA 和 RNA 提取试剂盒提取 DNA 和 RNA。

步骤 s2：进行宏基因组和宏转录组测序。

步骤 s3：宏基因组和宏转录组测序结果分析，包括以下步骤：

步骤 s31：使用 fastp 对测序得到的核算短片段进行质检并剔除质量较低的序列和参与的测序接头。

步骤 s32：使用拼接程序 SPADes 在‘Meta’模式下进行样品短序列拼接成长度不一的‘contigs’，然后根据双端测序的信息将不同‘contigs’连接成有测序缺口的‘scaffolds’。

步骤 s33：使用 Prodigal 程序将拼接后的长序列翻译成编码蛋白序列（CDS），采用 GhostKOALA，KofamKOALA 和 EggNOG mapper 2 三种工具给蛋白序列注释 COG 号，再转换成 KO 号。最终每个蛋白质的 KO 号注释采用以下顺序：1) GhostKOALA KO, 2) KofamKOALA KO 以及 3) EggNOG mapper KO。

步骤 s34：针对氢气还原酶，首先通过本地的氢气还原酶亚组的 HMM 模

型比对找出潜在的功能基因蛋白。之后将这些蛋白序列提交到 HydDB 数据库的在线分析软件进一步划分氢气还原酶的亚型。针对基因组中潜在的编码次级代谢产物的功能，主要通过将基因组序列提交至 AntiSMASH 网站，结合不同的工具找到基因组中潜在编码次级代谢产物的基因组，并预测代谢产物的类型。此外，针对数据库中信息较少的功能基因，如厌氧烃降解初始活化基因 AssA, EbdA, AhyA 和 AncA，以及细菌微室蛋白簇基因，专门构建本地的蛋白序列数据库，通过 BlasP 比对方法找到最相似的蛋白序列并进一步分析。

步骤 s35：使用 Maxbin2、Metabat2、CONCOCT 三种 Bining 手段从数据中分离出基因组，并导入 DAS_Tool 程序中进行评估，合并提取通过不同方法得到的高质量基因组。

步骤 s36：首先将目标序列以及数据库中的相似参比序列下载并合并文件，合并后的序列首先在 MAFFT 上排列整齐，并以 80%的阈值来选择保守位点，进而使用 IQ-tree 来两两比对序列并生成最大似然进化树。

步骤 s37：针对通过宏转录组测序得到的 mRNA 序列短片段。在质控后采用 bowtie2 等软件将 cDNA 短片段比对到宏基因组拼接得到的 DNA 长片段上，计算出各个基因的转录水平，以 TPM 值（Transcripts Per Million）来表示。

实施例 1

（1）样品采集及核酸提取

在 RNA 样品的桶中提前加入总体积 20%的 RNALater，取样时保证油藏样品充满全部桶体积以排除空气。在提取核酸前低温保存。

分别使用 PowerSoil Total DNA Kit (QIAGEN,美国)和 PowerSoil Total RNA Kit(QIAGEN,美国)提取 DNA 和 RNA，并使用 RNase-Free DNase set(QIAGEN, 美国)提纯 RNA，合成 cDNA 后-80℃保存。

（2）获得微生物的宏基因组和宏转录组数据

使用 NEB Next® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, 美国)试剂构建了带有索引码的测序文库。使用 Qubit 3.0 Fluorometer (Life Technologies, 美国) 和 Agilent 4200 (Agilent, 加拿大)共同评估了文库的质量。最后在 Illumina Hiseq X-ten 平台上，对文库进行了测序，得到了 150bp

的双端碱基序列。

(3) 数据分析

原始数据通过使用 fastp 进行质控和过滤。宏基因组数据通过 SPADes 经行组装,Maxbin2、Metabat2 和 CONCOCT 进行分箱。获得的基因组导入 DAS_Tool 后获得高质量的基因组。使用 prodigal 将长片段翻译成编码蛋白序列 CDS, 采用 GhostKOALA, KofamKOALA 和 EggNOG emapper 2 三种工具给蛋白序列注释,通过 HydDB 数据库 AntiSMASH 分析基因组的产氢气能力和次级代谢产物合成能力, 通过自建本地数据库分析厌氧烃降解能力。

宏转录组数据通过 Bowtie2 比对到宏基因组拼接的 DNA 长片段上, 计算出各个基因的转录水平 TPM。

结合样品中基因组的相对丰度和对应的代谢通路可以准确地识别油藏中的主要微生物和代谢功能, 再结合转录组数据可以得到这些微生物的表达活性。

以大庆油田的一份产出液样品为例, 表 1 中的数据为各基因组测序覆盖度和代谢通路。从表中可以得知该样品中最主要的微生物为变形菌门的 *Kerstersia_gyiorum* 和 *Serratia_nematodiphila* 以及厚壁菌门的 *Enterococcus*。样品中的微生物大都为具有硝酸盐还原能力的兼性厌氧微生物, 且样品中存在针对不同链长的好氧烃降解菌。近一半的微生物都具有完整的合成脂肽的代谢通路。此外, 大多数氢营养型微生物同时具有产氢能力。以上这些信息可以为油藏提高采收率和防腐蚀方面的决策提供理论依据。

基因组信息表 1

基因组	测序覆盖度	乙酸代谢	好氧	好氧烃降解	好氧长链烃降解	脂肪酸降解	糖类降解	硝酸盐还原	氢气营养型	产氢气	乙醇发酵	脂肽合成
<i>Alkaliphilus_oremlandii_bin_24</i>	0.82						√	√			√	
<i>Aquabacterium_bin_26</i>	3.99	√	√	√		√	√	√			√	
<i>Bacillus_cereus_AE_bin_2</i>	3.36	√	√				√				√	
<i>Bacillus_licheniformis_bin_11</i>	0.96	√	√				√	√			√	√
<i>Bacillus_paralicheniformis_bin_5</i>	0.22	√	√					√			√	√
<i>Bacillus_thermoamylovorans_bin_1</i>	0.19	√	√			√	√	√			√	
<i>Brevibacillus_borstelensis_bin_18</i>	0.88	√	√				√				√	√
<i>Brevibacillus_borstelensis_bin_6</i>	0.19	√									√	√
<i>Burkholderiaceae_bin_19</i>	7.84	√	√			√	√	√			√	
<i>Burkholderiaceae_bin_5</i>	15.83	√	√			√	√	√			√	

Clostridium_G_cochlearium_bin_14	21.26						√	√	√		√	
Clostridium_J_argentinense_bin_24	0.73						√	√	√		√	√
Enterococcus_faecium_bin_1	45.73											
Enterococcus_faecium_bin_8	323.94	√					√				√	
Escherichia_coli_bin_13	8.71	√	√			√	√	√	√	√	√	√
Escherichia_coli_bin_15	0.19	√	√			√		√	√	√	√	
Gemmobacter_nectariphilus_bin_2	6.78	√	√	√	√	√	√				√	
Kerstesia_gyiorum_bin_2	1772.74	√	√		√	√	√				√	
Leucobacter_bin_9	8.89	√	√		√	√	√	√			√	√
Pseudomonas_formosensis_bin_23	0.37	√	√			√	√	√			√	
Pseudomonas_formosensis_bin_24	0.24	√	√		√	√	√	√			√	
Pseudomonas_thermotolerans_bin_8	0.46	√	√	√	√	√		√			√	
Roseovarius_bin_1	3.56	√	√	√			√	√			√	
Serratia_nematodiphila_bin_12	193.18	√	√		√	√	√	√			√	√
Thauera_bin_4	41.11	√	√			√	√	√	√	√	√	
Thauera_phenylacetica_bin_19	6.45	√	√			√	√	√	√	√	√	√
Tissierellaceae_bin_14	0.47						√	√			√	

上述的对实施例的描述是为便于该技术领域的普通技术人员能理解和使用发明。熟悉本领域技术的人员显然可以容易地对这些实施例做出各种修改，并把在此说明的一般原理应用到其他实施例中而不必经过创造性的劳动。因此，本发明不限于上述实施例，本领域技术人员根据本发明的揭示，不脱离本发明范畴所做出的改进和修改都应该在本发明的保护范围之内。

权 利 要 求 书

1. 一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法，其特征在于，包括以下步骤：

S1：自油藏产出水样提取总 DNA 和总 RNA；

S2：对获得的总 DNA 和总 RNA 进行测序，获取油藏样品的宏基因组和宏转录组原始数据；

S3：通过对宏基因组和宏转录组结果进行分析，识别具有驱油功能的微生物。

2. 根据权利要求 1 所述的一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法，其特征在于，S1 中，在提取总 DNA 和总 RNA 之前，在待提取 RNA 的样品中加入抑制剂以抑制厌氧微生物的 RNA 降解。

3. 根据权利要求 1 所述的一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法，其特征在于，S2 中，还包括：对油藏样品的宏基因组和宏转录组原始数据进行预处理，得到去除接头和低质量片段的目标数据；

所述预处理的过程包括：

利用 fastp 软件分别对油藏样品的宏基因组和宏转录组中各 DNA 或 RNA 链序列双端的原始数据进行滑窗质量剪裁，同时，根据序列首尾两端的引物信息，利用 cutadapt 软件去除引物，得到质控后的双端序列数据。

4. 根据权利要求 3 所述的一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法，其特征在于，S3 中，对宏基因组和宏转录组结果进行分析的过程包括：

对质控后的双端序列数据进行组装、分箱并评估质量，去除冗余后提取其中高质量的 MAGs 数据集；

根据构建的参考数据库对高质量的 MAGs 组装数据进行注释，识别出具有不同驱油功能的功能基因和 MAGs，并计算对应的测序深度和相对丰度；

对高质量的 MAGs 数据集做进化关系分析，进而做出驱油微生物的群落结构分析；

将宏转录组测序得到的序列短片段质控过滤后与 MAGs 数据对比，计算

出各个基因的转录水平。

5. 根据权利要求4所述的一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法，其特征在于，S3中，组装、分箱并评估质量的过程包括：

组装：使用拼接程序 SPADes 在 Meta 模式下，将质控后的双端序列数据样品短序列拼接成长度不一的 contigs，然后根据双端测序的信息将不同 contigs 连接成有测序缺口的 scaffolds；

分箱：采用 bowtie2 软件将质控后的 MAGs 短序列数据信息比对到长序列信息上，获得不同长序列的测序覆盖度信息，进一步同时采用 Maxbin2、Metabat2、CONCOCT 三种 Binning 手段从 MAGs 中分离出优势菌的基因组，并进一步导入 DAS_Tool 程序中进行评估，最终整合并提高不同方法生成基因组的质量；

评估质量：利用 dRep 软件将质控后的 MAGs 中相似度较高的基因组去除冗余，通过 CheckM 工具根据基因组中单拷贝标记基因的有无和数量来估算基因组的完整度和污染度。

6. 根据权利要求4所述的一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法，其特征在于，S3中，对高质量的 MAGs 组装数据进行注释的过程包括：

使用 Prodigal 程序将拼接后的长序列翻译成编码蛋白序列，并提交到 KEGG 数据库，采用 GhostKOALA 工具进行功能性注释并获得代表不同旁系同源亚基的 KO 号；

同时采用本地软件 KofamKOALA 根据各个 KO 旁系同源家族蛋白的隐马可夫模型和推荐的置信标准给各个蛋白序列注释 KO 号；

最后采用 EggNOG mapper 2 工具给蛋白序列注释 COG 号，再转换成 KO 号，使得最终每个蛋白质的 KO 号注释采用以下顺序：1) GhostKOALA KO, 2) KofamKOALA KO，3) EggNOG mapper KO。

7. 根据权利要求4所述的一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法，其特征在于，S3中，所述驱油功能包括烷烃降解、产气、产乳化剂、产表面活性剂、产多糖中的一种或多种。

8. 根据权利要求7所述的一种基于宏基因组和宏转录组识别油藏驱油功

能微生物的方法，其特征在于，S3 中，识别出具有不同驱油功能的功能基因和 MAGs 的过程包括：

针对氢气还原酶，首先通过本地的氢气还原酶亚组的 HMM 模型比对找出潜在的功能基因蛋白；

之后将潜在的功能基因蛋白序列提交到 HydDB 数据库的在线分析软件进一步划分氢气还原酶的亚型；

针对基因组中潜在的编码次级代谢产物的功能，通过将基因组序列提交至 AntiSMASH 网站，结合不同的工具找到基因组中潜在编码次级代谢产物的基因组，并预测代谢产物的类型；

针对数据库中信息缺少的功能基因，单独构建本地的蛋白序列数据库，通过 BlastP 比对方法找到最相似的蛋白序列并进一步分析，所述数据库中信息缺少的功能基因包括厌氧烃降解初始活化基因 AssA、EbdA、AhyA、AncA，和细菌微室蛋白簇基因中的一种或多种。

9. 根据权利要求 4 所述的一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法，其特征在于，S3 中，对高质量的 MAGs 数据集做进化关系分析的过程包括：

首先将目标序列以及数据库中的相似参比序列下载并合并文件，合并后的序列首先在 MAFFT 上排列整齐，并以 80% 的阈值来选择保守位点，进而使用 IQ-tree 来两两比对序列并生成最大似然进化树。

10. 根据权利要求 4 所述的一种基于宏基因组和宏转录组识别油藏驱油功能微生物的方法，其特征在于，S3 中，将宏转录组测序得到的序列短片段质控过滤后与 MAGs 数据对比，计算出各个基因的转录水平的过程包括：

采用 bowtie2 软件将高质量的 MAGs 数据集中的 cDNA 短片段比对到宏基因组，并拼接得到的 DNA 长片段上，计算出各个基因的转录水平，以 TPM 值来表示。

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/099026

A. CLASSIFICATION OF SUBJECT MATTER

C12Q1/6869(2018.01)i; C12Q1/04(2006.01)i; G16B10/00(2019.01)i; G16B20/00(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: C12Q; G16B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

VEN, CNABS, PubMed, CNKI, ISI Web of Knowledge: 宏基因组, 宏转录组, 油藏, 驱油功能, 微生物, 总DNA, 总RNA, 测序, fastp, cutadapt, MAGs, 组装, 注释, contigs, bowtie2, Binning, KEGG, 氢气还原酶, 进化树, reservoir microbial, oil reservoir, metagenomics, metatranscriptomes, total DNA, total RNA, sequencing, microorganisms, assembly, annotation, hydrogen reductase, evolved tree

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 115678978 A (EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY) 03 February 2023 (2023-02-03) claims 1-10	1-10
A	CN 112786102 A (PEKING UNIVERSITY) 11 May 2021 (2021-05-11) entire document	1-10
A	CN 112599198 A (SHANGHAI PERSONAL BIOTECHNOLOGY CO., LTD.) 02 April 2021 (2021-04-02) entire document	1-10
A	CN 113337591 A (TSINGHUA SHENZHEN INTERNATIONAL GRADUATE SCHOOL) 03 September 2021 (2021-09-03) entire document	1-10
A	CN 113257348 A (NANKAI UNIVERSITY) 13 August 2021 (2021-08-13) entire document	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
 “D” document cited by the applicant in the international application
 “E” earlier application or patent but published on or after the international filing date
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 “O” document referring to an oral disclosure, use, exhibition or other means
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

12 September 2023

Date of mailing of the international search report

18 September 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/099026

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LIU, Yifan et al. "Metabolic capability and in situ activity of microorganisms in an oil reservoir" <i>Microbiome</i> , 05 January 2018 (2018-01-05), Abstract, Methods	1-2
Y	LIU, Yifan et al. "Metabolic capability and in situ activity of microorganisms in an oil reservoir" <i>Microbiome</i> , 05 January 2018 (2018-01-05), Abstract, Methods	3-10
Y	林佳伟等 (LIN, Jiawei et al.). "宏基因组技术在油藏微生物分析中的应用 (Application of Metagenomic Sequencing in Reservoir Microbial Analysis)" <i>计算机与应用化学 (Computers and Applied Chemistry)</i> , Vol. 35, No. 7, 28 July 2018 (2018-07-28), pages 565-570	3-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/099026

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	115678978	A	03 February 2023	None	
CN	112786102	A	11 May 2021	None	
CN	112599198	A	02 April 2021	None	
CN	113337591	A	03 September 2021	None	
CN	113257348	A	13 August 2021	None	

国际检索报告

国际申请号

PCT/CN2023/099026

A. 主题的分类

C12Q1/6869(2018.01)i; C12Q1/04(2006.01)i; G16B10/00(2019.01)i; G16B20/00(2019.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: C12Q; G16B

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

VEN, CNABS, PubMed, CNKI, ISI Web of Knowledge, 宏基因组, 宏转录组, 油藏, 驱油功能, 微生物, 总DNA, 总RNA, 测序, fastp, cutadapt, MAGs, 组装, 注释, contigs, bowtie2, Binning, KEGG, 氢气还原酶, 进化树, reservoir microbial, oil reservoir, metagenomics, metatranscriptomes, total DNA, total RNA, sequencing, microorganisms, assembly, annotation, hydrogen reductase, evolved tree

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 115678978 A (华东理工大学) 2023年2月3日 (2023 - 02 - 03) 权利要求1-10	1-10
A	CN 112786102 A (北京大学) 2021年5月11日 (2021 - 05 - 11) 全文	1-10
A	CN 112599198 A (上海派森诺生物科技股份有限公司) 2021年4月2日 (2021 - 04 - 02) 全文	1-10
A	CN 113337591 A (清华大学深圳国际研究生院) 2021年9月3日 (2021 - 09 - 03) 全文	1-10
A	CN 113257348 A (南开大学) 2021年8月13日 (2021 - 08 - 13) 全文	1-10
X	Liu, Y.F. 等. "Metabolic capability and in situ activity of microorganisms in an oil reservoir" Microbiome, 2018年1月5日 (2018 - 01 - 05), Abstract, Methods	1-2

☒ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2023年9月12日

国际检索报告邮寄日期

2023年9月18日

ISA/CN的名称和邮寄地址

中国国家知识产权局

中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

刘宝

电话号码 (+86) 010-53962038

C. 相关文件		
类 型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	Liu, Y.F. 等. "Metabolic capability and in situ activity of microorganisms in an oil reservoir" Microbiome, 2018年1月5日 (2018 - 01 - 05), Abstract, Methods	3-10
Y	林佳伟等. "宏基因组技术在油藏微生物分析中的应用" 计算机与应用化学, 第35卷, 第7期, 2018年7月28日 (2018 - 07 - 28), 565-570页	3-10

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/099026

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	115678978	A	2023年2月3日	无	
CN	112786102	A	2021年5月11日	无	
CN	112599198	A	2021年4月2日	无	
CN	113337591	A	2021年9月3日	无	
CN	113257348	A	2021年8月13日	无	