

(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
2012 年 6 月 28 日 (28.06.2012)



(10) 国际公布号
WO 2012/083832 A1

- (51) 国际专利分类号:
C12N 15/10 (2006.01) *C07H 21/02* (2006.01)
C12N 15/11 (2006.01) *C12Q 1/68* (2006.01)
- (21) 国际申请号: PCT/CN2011/084237
- (22) 国际申请日: 2011 年 12 月 19 日 (19.12.2011)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201010600211.0 2010 年 12 月 22 日 (22.12.2010) CN
- (71) 申请人 (对除美国外的所有指定国): **深圳华大基因科技有限公司 (BGI SHENZHEN CO., LIMITED)**
[CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼 11F-3, Guangdong 518083 (CN)。
深圳华大基因研究院 (BGI SHENZHEN) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
- (72) 发明人: 及
- (75) 发明人/申请人 (仅对美国): **万成 (WAN, Cheng)**
[CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
刘晓 (LIU, Xiao)

[CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
陈利 (CHEN, Li)
[CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
刘雪莉 (LIU, Xueli)
[CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
王俊 (WANG, Jun)
[CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
汪建 (WANG, Jian)
[CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(74) 代理人: **北京清亦华知识产权代理事务所 (普通合伙) (TSINGYIHUA INTELLECTUAL PROPERTY LLC)**; 中国北京市海淀区清华园清华大学照澜院商业楼 301 室, Beijing 100084 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY,

[见续页]

(54) Title: METHOD FOR ENRICHING PROKARYOTIC MRNA AND USE THEREOF

(54) 发明名称: 富集原核生物 MRNA 的方法及其用途

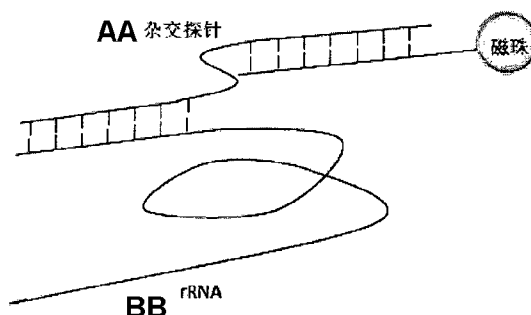


图 1 / Fig. 1

AA HYBRIDIZATION PROBE
BB MAGNETIC BEAD

(57) Abstract: Provided are a method for enriching prokaryotic mRNA, a method for constructing a transcriptome sequencing library, a transcriptome sequencing library, a method for determining prokaryotic mRNA sequence information, a probe and a use thereof in enriching mRNA, and a kit containing the probe and a use thereof in constructing a transcriptome sequencing library. The method for enriching prokaryotic mRNA comprises the following steps: hybridizing a nucleic acid sample with a probe, wherein the nucleic acid sample contains prokaryotic 16s rRNA, 23s rRNA, and mRNA, and the probe carries a marker and can specifically bind 16s rRNA and 23s rRNA, to form a probe-rRNA complex; and removing the probe-rRNA complex by using a molecule entity specifically binding the marker, to enrich the prokaryotic mRNA.

[见续页]



WO 2012/083832 A1



TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

- (84) **指定国** (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, 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, MK, MT, NL, NO,

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。
- 包括说明书序列表部分(细则 5.2(a))。

(57) **摘要:**

提供了富集原核生物 mRNA 的方法、构建转录组测序文库的方法、转录组测序文库、确定原核生物 mRNA 序列信息的方法、探针及其在富集 mRNA 中的用途以及含有该探针的试剂盒及其在构建转录组测序文库中的用途。富集原核生物 mRNA 的方法包括以下步骤: 将核酸样品与探针杂交, 其中该核酸样品含有原核生物的 16s rRNA、23s rRNA 和 mRNA, 该探针携带标记物, 并且该探针能够与 16s rRNA 和 23s rRNA 特异性结合以便形成探针-rRNA 复合物; 以及利用特异性结合标记物的分子实体去除探针-rRNA 复合物, 以便富集原核生物 mRNA。

富集原核生物 mRNA 的方法及其用途

优先权信息

本申请请求 2010 年 12 月 22 日向中国国家知识产权局提交的、专利申请号为 201010600211.0 的专利申请的优先权和权益，并且通过参照将其全文并入此处。

技术领域

本发明涉及基因组测序领域，特别是转录组测序领域。具体地，本发明涉及富集原核生物 mRNA 的方法及其用途。更具体地，本发明提供了一种富集原核生物 mRNA 的方法、一种构建转录组测序文库的方法、一种转录组测序文库、一种确定原核生物 mRNA 序列信息的方法、一种探针及其在富集 mRNA 中的用途以及一种试剂盒及其在构建转录组测序文库中的用途。

背景技术

原核生物是生物系统中的重要组成部分，包括细菌、蓝藻、立克次氏体和衣原体等多个门类。获得原核生物的 mRNA 信息，乃至对原核生物进行转录组研究，都对医药、食品、环境和疾病等许多领域具有重要意义。但是，在原核生物中，mRNA (messenger RNA, 信使 RNA) 的含量极低，仅占其总 RNA (total RNA) 的 2%-5% 左右，且绝大多数原核生物 mRNA 都不具有 3' 端的 poly(A) 结构，使得 mRNA 的分离有较大困难。因此，按照以真核生物 mRNA 特点而建立的转录组文库构建方法进行原核生物转录组 Solexa 高通量测序，难以获得理想的结果（例如参见 Arnoud H.M. van Vliet. *Next generation sequencing of microbial transcriptomes: challenges and opportunities*. Federation of European Microbiological Societies, Microbiol Lett 302 (2010) 1-7，通过参照将其全文并入本文）。因此，确定原核生物 mRNA 序列信息的方法仍有待改进。

发明内容

本发明旨在至少解决现有技术中存在的技术问题之一。为此，本发明提供了富集原核生物 mRNA 的方法及其用途。

根据本发明的一个方面，本发明提供了一种富集原核生物 mRNA 的方法。根据本发明的实施例，该方法包括以下步骤：将核酸样品与探针杂交，该核酸样品含有原核生物的 16S

rRNA、23 s rRNA 和 mRNA，该探针携带标记物，并且该探针能够与 16s rRNA 和 23 s rRNA 特异性结合以便形成探针-rRNA 复合物；以及利用特异性结合标记物的分子实体去除探针-rRNA 复合物，以便富集原核生物 mRNA。发明人惊奇地发现，利用根据本发明实施例的富集原核生物 mRNA 的方法，能够有效地富集原核生物的 mRNA，且方便、快速、可重复性好，去除 rRNA 的效率、效果好，相比现有技术，利用该方法富集得到的原核生物 mRNA 的纯度和比例显著提高。

根据本发明的又一方面，本发明提供了一种构建转录组测序文库的方法。根据本发明的实施例，该方法包括以下步骤：根据本发明实施例的富集原核生物 mRNA 的方法，富集 mRNA；以及针对所得到的 mRNA，构建测序文库，以便获得转录组测序文库。发明人惊奇地发现，利用根据本发明实施例的构建转录组测序文库的方法，能够方便有效地构建原核生物的转录组测序文库，且可重复性好、成本低、需时少、效率高，获得的文库质量好。

根据本发明的另一方面，本发明提供了一种转录组测序文库。根据本发明的实施例，该文库是由本发明的构建转录组测序文库的方法构建的。发明人惊奇地发现，根据本发明实施例的转录组测序文库，能够有效地应用于高通量测序平台例如 Illumina Solexa 测序平台，从而基于测序结果，能够准确有效地获得原核生物 mRNA 序列信息。

根据本发明的再一方面，本发明提供了一种确定原核生物 mRNA 序列信息的方法。根据本发明的实施例，该方法包括以下步骤：根据本发明实施例的构建转录组测序文库的方法构建转录组测序文库；以及对转录组测序文库进行测序，以便获得原核生物 mRNA 序列信息。发明人惊奇地发现，利用根据本发明实施例的确定原核生物 mRNA 序列信息的方法，能够准确有效地确定原核生物的 mRNA 序列信息，且成本低、用时少、效率高、可重复性好。

根据本发明的又一方面，本发明提供了一种探针。根据本发明的实施例，该探针携带标记物，并且能够与原核生物的 16s 和 23s rRNA 的保守区域特异性结合。根据本发明的实施例，本发明的探针能够用于富集原核生物 mRNA，由此，本发明的探针能够有效地应用于本发明的富集原核生物 mRNA 的方法、构建转录组测序文库的方法以及确定原核生物 mRNA 序列信息的方法中，从而能够使得原核生物的 mRNA 序列信息被准确有效地确定。

根据本发明的另一方面，本发明提供了根据本发明实施例的探针在富集 mRNA 中的用途。

根据本发明的再一方面，本发明提供了一种试剂盒。根据本发明的实施例，该试剂盒包括：根据本发明实施例的探针；以及分子实体，该分子实体特异性结合探针的标记物。发明人发现，利用根据本发明实施例的试剂盒，能够方便、快速地构建原核生物的转录组

测序文库，且文库质量非常好。

根据本发明的又一方面，本发明提供了根据本发明实施例的试剂盒在构建转录组测序文库中的用途。

5 本发明的附加方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本发明的实践了解到。

附图说明

本发明的上述和/或附加的方面和优点从结合下面附图对实施例的描述中将变得明显和容易理解，其中：

10 图 1：显示了利用杂交探针与去除原核生物 rRNA 的原理；

图 2：显示了利用 Agilent Bioanalyzer 2100d 对总 RNA 样品与根据本发明实施例的富集原核生物 mRNA 的方法富集的 mRNA 进行检测的结果；

图 3：显示了根据本发明一个实施例的大肠杆菌转录组 Solexa 测序结果；以及

15 图 4：显示了根据本发明一个实施例的 E.coli 1 转录组测序文库经 DSN 酶处理前后的测序结果。

发明详细描述

下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，仅用于解释本发明，而不能理解为对本发明的限制。

需要说明的是，在本发明的描述中，除非另有说明，“多个”的含义是两个或两个以上。

富集原核生物 mRNA 的方法

25 根据本发明的一个方面，本发明提供了一种富集原核生物 mRNA 的方法。根据本发明的实施例，该方法可以包括以下步骤：将核酸样品与探针杂交，该核酸样品含有原核生物的 16s rRNA、23 s rRNA 和 mRNA，该探针携带标记物，并且该探针能够与 16s rRNA 和 23 s rRNA 特异性结合以便形成探针-rRNA 复合物；以及利用特异性结合标记物的分子实体去除探针-rRNA 复合物，以便富集原核生物 mRNA。根据本发明的一个实施例，标记物为 polyA，分子实体为寡聚 dT。根据本发明的具体示例，该分子实体形成于磁珠上。根据本发
30 明的实施例，探针包括 SEQ ID NO: 1-10 所示的寡核苷酸。根据本发明的实施例，可以在 70-74℃ 下进行杂交 25-35 分钟。根据本发明的一个实施例，杂交是在 72℃ 下进行的。根据

本发明的具体示例，进行杂交 30 分钟。

在本文中所使用的术语“杂交”是指，相互间具有互补序列的两个单链核酸分子在一定条件下(适宜的温度及离子强度等)按碱基互补配对原则退火形成双链核酸的过程。核酸杂交可以在 DNA-DNA 之间进行，也可以在 DNA-RNA 或 RNA-RNA 之间进行，只要它们之间存在互补序列，可以进行碱基配对。通常，杂交的双方可以分别是待测核酸分子和已知核酸分子。在杂交体系中已知序列的核酸分子称作探针(probe)。核酸杂交可以包括固-液相杂交和液相杂交，其中液相杂交是在溶液中进行的杂交反应，其是指待测核酸分子与已知核酸分子(探针)在溶液中退火形成杂交复合物。当利用携带标记物的探针进行杂交时，在杂交反应结束后，能够利用探针上的标记物分离和检测杂交后的双链，其中，标记物是本领域内是已知的，可以包括但不限于放射性同位素或核素、生物素、吖啶翁酯(acridinium ester)或 polyA 等。根据所使用的标记物的类型，可以利用本领域内已知的相应的核酸分离和检测方法，包括但不限于羟基磷灰石(HAP)法或亲和吸附法，将杂交产物进行分离和检测。

发明人惊奇地发现，利用根据本发明实施例的富集原核生物 mRNA 的方法，能够快速、有效地富集原核生物的 mRNA，且可重复性好，去除 rRNA 的效率高、效果好，相比现有技术，利用该方法富集得到的原核生物 mRNA 的纯度和比例显著提高。根据本发明的实施例，利用本发明的富集原核生物 mRNA 的方法富集的原核生物 mRNA，能够有效地应用于原核生物的转录组测序文库的构建，且可重复性好，获得的文库质量好，进一步将所得的文库进行高通量测序，从而能够准确有效地获得原核生物 mRNA 序列信息。

根据本发明的又一方面，本发明提供了一种构建转录组测序文库的方法。根据本发明的实施例，该方法可以包括以下步骤：根据本发明实施例的富集原核生物 mRNA 的方法，富集 mRNA；以及针对所得到的 mRNA，构建测序文库，以便获得转录组测序文库。根据本发明的实施例，转录组测序文库适于利用选自 Illumina Solexa、ABI SOLiD 及 Roche 454 测序平台的至少一种进行测序。根据本发明的实施例，本发明的构建转录组测序文库的方法可以进一步包括：利用双链特异性核酸酶对所述转录组测序文库进行处理。

发明人惊奇地发现，利用根据本发明实施例的构建转录组测序文库的方法，能够有效构建原核生物的转录组测序文库，且可重复性好、成本低、用时少、效率高，获得的文库质量好。根据本发明的实施例，利用本发明的构建转录组测序文库的方法构建的原核生物的转录组测序文库，能够有效地应用于高通量测序平台例如 Illumina Solexa 测序平台，且测序成本低、用时少、效率高、可重复性好、测序结果准确，从而基于测序结果，能够准确有效地获得原核生物 mRNA 序列信息。

进一步, 根据本发明的实施例, 本发明的构建转录组测序文库的方法可以包括以下步骤: 提供与原核生物的 16s 和 23s rRNA 的保守区域杂交的携带标记物的探针; 使用该探针结合核酸样品中的 rRNA, 以便形成探针-rRNA 复合物; 利用与标记物特异性结合的分子实体结合探针-rRNA 复合物, 从而去除 rRNA-探针复合物而富集 mRNA, 其后将富集得到的 mRNA 用于构建文库。

其中, 根据本发明的实施例, 探针是针对 16s RNA 和 23s RNA 保守区设计的探针。根据本发明的实施例, 设计探针的方法不受特别限制。根据本发明的一个实施例, 标记物是 polyA, 并且分子实体为寡聚 dT (Oligo (dT))。根据本发明的具体示例, 寡聚 dT 可以形成于磁珠上。根据本发明的实施例, 标记物 polyA 的长度不受特别限制, 可以由本领域技术人员根据现有技术来确定。根据本发明的具体示例, polyA 可以是 5-30 个碱基 A, 优选 10-20 个碱基 A, 例如可以为 15、16、17、18 或 19 个碱基 A。具体地, 根据本发明的实施例, 探针包括 SEQ ID NO: 1-10 所示的寡核苷酸。根据本发明的实施例, 探针与 rRNA 杂交的温度是 70-74℃, 优选 72℃; 杂交时间是 25-35min, 优选 30min。根据本发明的实施例, 用于构建文库的方法可以为基于 Illumina solexa, ABI SOLiD 或 Roche 454 测序平台的建库方法, 优选基于 illumina solexa 测序平台的建库方法。其中, 上述三种测序平台均为已经商业化的测序平台, 其说明书中都提供有 RNA 相关的建库流程, 具体可参考商家的网站。根据本发明的实施例, 该方法还可以进一步包括使用 DSN 酶处理构建完成的文库的步骤。

根据本发明的另一方面, 本发明提供了一种转录组测序文库。根据本发明的实施例, 该文库是由本发明的构建转录组测序文库的方法构建的。发明人惊奇地发现, 根据本发明的实施例的转录组测序文库, 能够有效地应用于高通量测序平台例如 Illumina Solexa 测序平台, 从而基于测序结果, 能够准确地获得原核生物 mRNA 序列信息。

根据本发明的再一方面, 本发明提供了一种确定原核生物 mRNA 序列信息的方法。根据本发明的实施例, 该方法包括以下步骤: 根据本发明实施例的构建转录组测序文库的方法构建转录组测序文库; 以及对转录组测序文库进行测序, 以便获得原核生物 mRNA 序列信息。根据本发明的实施例, 测序的方法不受特别限制, 可以利用选自 Illumina Solexa、ABI SOLiD 及 Roche 454 测序平台的至少一种进行测序。根据本发明的具体示例, 可以利用 Illumina Solexa 测序平台进行测序。

发明人惊奇地发现, 利用根据本发明实施例的确定原核生物 mRNA 序列信息的方法, 能够有效地确定原核生物的 mRNA 序列信息, 且成本低、用时少、效率高、结果准确、可重复性好。

根据本发明的又一方面, 本发明提供了一种探针。根据本发明的实施例, 该探针携带

标记物，并且能够与原核生物的 16s 和 23s rRNA 的保守区域特异性结合。根据本发明的实施例，该探针由 SEQ ID NO: 1-10 所示的寡核苷酸构成。

发明人发现，根据本发明实施例的探针能够用于富集原核生物 mRNA，由此，本发明的探针能够有效地应用于本发明的富集原核生物 mRNA 的方法、构建转录组测序文库的方法以及确定原核生物 mRNA 序列信息的方法中，从而能够使得原核生物的 mRNA 序列信息
5 被有效地确定，且信息准确，可重复性好。

根据本发明的另一方面，本发明提供了根据本发明实施例的探针在富集 mRNA 中的用途。根据本发明的实施例，探针在富集 mRNA 中的用途主要表现为：在本发明的富集原核生物 mRNA 的方法中，利用本发明的携带标记物的探针能够与原核生物中的 16s
10 rRNA 和 23 s rRNA 特异性结合以便形成探针-rRNA 复合物，进一步利用特异性结合标记物的分子实体能够去除探针-rRNA 复合物，从而能够成功富集原核生物 mRNA。

根据本发明的再一方面，本发明提供了一种试剂盒。根据本发明的实施例，该试剂盒包括：根据本发明实施例的探针；以及分子实体，其特异性结合该探针的标记物。其中，关于根据本发明实施例的探针的特征、优点及用途，前面已经详细描述，在此不再赘述。
15 根据本发明的一个实施例，探针的标记物为 polyA，且分子实体为寡聚 dT。根据本发明的一个具体事例，分子实体形成于磁珠上。由此，利用根据本发明实施例的试剂盒，能够方便有效地构建原核生物的转录组测序文库，且文库质量非常好。当然，本领域技术人员能够理解，试剂盒中还可以包含其他用于构建原核生物的转录组测序文库的常规组件，根据本发明的实施例，本发明的试剂盒中还可以包括适于构建原核生物的转录组测序文库的
20 其他试剂。

具体地，根据本发明的实施例，本发明的试剂盒可以包括：与原核生物的 16s 和 23s rRNA 的保守区域杂交的经标记的探针，以及能与探针特异性结合的分子实体。根据本发明的一个实施例，该试剂盒中用于标记的标记物是 polyA，分子实体为寡聚 dT (Oligo (dT))。根据本发明的具体示例，该寡聚 dT 形成于磁珠上。根据本发明的一些
25 实施例，该试剂盒中的探针包括 SEQ ID NO: 1-10 所示的寡核苷酸。

根据本发明的又一方面，本发明提供了根据本发明实施例的试剂盒在构建转录组测序文库中的用途。

需要说明的是，根据本发明实施例的确定原核生物 mRNA 序列信息的方法是本发明的发明人经过艰苦的创造性劳动和优化工作才完成的。

30

下面将结合实施例对本发明的方案进行解释。本领域技术人员将会理解，下面的实施

例仅用于说明本发明，而不应视为限定本发明的范围。实施例中未注明具体技术或条件的，按照本领域内的文献所描述的技术或条件（例如参考 J.萨姆布鲁克等著，黄培堂等译的《分子克隆实验指南》，第三版，科学出版社）或者按照产品说明书进行。所用试剂或仪器未注明生产厂商者，均为可以通过市购获得的常规产品，例如可以采购自 Illumina 公司。

5 实施例 1: 大肠杆菌转录组 solexa 测序

一、样品处理

1. 利用本发明的方法处理样品

利用本发明的富集原核生物 mRNA 的方法，按照下列步骤，分别将三份大肠杆菌 (*E.coli*) 总 RNA 样品进行处理：

10 1.1 DNA 酶消化：

1) 取 10μg 总 RNA 样品（利用 TRIzol (invitrogen, 美国)，按照生产商说明书中的步骤，从大肠杆菌 (*E.coli*) 中提取）至一个无 RNA 酶的 1.5mL 离心管 (Axygen, 美国) 中，然后添加 1μL DNA 酶 (invitrogen, 美国)，混匀后依次于 37℃ 下进行反应 10min，80℃ 下加热 5min，然后终止反应；

15 2) 向上述离心管中添加 3 倍体积的无水乙醇 (sigma, 美国)，混匀，然后于 -80℃ 下静置 30min，接着于 4℃ 下进行高速离心 20min，吸去上清获得沉淀，然后添加 1mL 用 DEPC 水 (Ambion, 美国) 配成的 75% 乙醇将沉淀进行洗涤一次，然后于 4℃ 下进行高速离心 5min，弃上清并在空气中干燥去除残留的乙醇，以便获得干燥的沉淀；

3) 利用 10μL DEPC 水将得到的干燥的沉淀进行溶解，获得 DNA 酶消化过的总 RNA，
20 然后于 -80℃ 下进行保存，备用。

1.2 将探针与 rRNA 杂交：

其中，这里的探针由如 SEQ ID NO: 1-10 所示的寡核苷酸构成。

1) 取 200μL 杂交缓冲液至无 RNA 酶的 1.5mL EP 管 (Axygen, 美国) 中；

2) 向 EP 管中添加 15μL DNA 酶消化过的总 RNA (不超过 10μg)，混匀；

25 3) 向 EP 管中添加 4μL 探针，混匀，然后将其进行短暂离心使混合液位于管底；

4) 将 EP 管置于 72℃ 下进行变性 5min；

5) 将 EP 管置于 37℃ 下进行结合 30min 以便形成探针-rRNA 复合物，备用。

1.3 含寡聚 dT 的磁珠结合探针-rRNA 复合物：

1) 将洗脱液置于 37℃ 下进行温育 10min，备用；

30 2) 向上述含有探针-rRNA 复合物的 EP 管中添加 50μL 结合有寡聚 dT 的磁珠 (Dynabeads Oligo(dT)25, invitrogen, 61005)，混匀，然后进行短暂离心，并于 37℃ 下进行温育 30 min；

3) 将上述 EP 管置于磁力架 (invitrogen, 美国) 上, 静置 2~3min, 然后取上清液至收集管 (Collection Tube, Axygen, 美国) 中, 于冰上静置;

4) 利用预热的洗脱液对 EP 管中的磁珠进行洗涤: 向 EP 管中添加 100 μ L 预热的洗脱液, 混匀后于磁力架上静置 2~3min, 然后取上清液, 并将其与收集管中的溶液合并;

5) 向收集管中添加三倍体积的无水乙醇沉淀 RNA, 然后用适量 RNase-free 水将得到的 RNA 沉淀溶解, 由此, 获得大肠杆菌 (*E.coli*) 总 RNA 样品的经过本发明的方法处理的样品 (即富集的 mRNA)。并于 -80 $^{\circ}$ C 下保存, 备用。

2. 平行地, 采用切胶法 (例如参见 Ken C. McGrath, Skye R. Thomas-Hall, et al. *Isolation and analysis of mRNA from environmental microbial communities, Journal of Microbiological Methods*. 75 (2008), 172-176, 通过参照将其全文并入本文) 去除总 RNA 样品中的 rRNA, 以便获得经过切胶处理的样品, 备用。

3. 平行地, 直接将总 RNA 样品保存, 作为对照组, 备用。

利用 Agilent Bioanalyzer 2100 (Agilent, 美国) 分别将上述经过本发明的方法处理的样品和总 RNA 样品进行检测, 结果见图 2。图 2 显示了利用 Agilent Bioanalyzer 2100 将总 RNA 样品与根据本发明实施例的富集原核生物 mRNA 的方法富集的 mRNA 进行检测的结果。如图 2 所示, 其中, a 为总 RNA 样品的检测结果, b 为根据本发明实施例的富集原核生物 mRNA 的方法富集 mRNA 后的 RNA 样品的检测结果。由图 2 可知, 将总 RNA 样品通过本根据本发明实施例的富集原核生物 mRNA 的方法处理后, 2100 检测峰图上原有的 16s、23s rRNA 峰消失了, 结果表明该方法有效去除了总 RNA 样品中的大部分 rRNA。

二、建库并测序

将上述检测合格的经过本发明的方法处理的样品、经过切胶处理的样品以及总 RNA 样品分别按照 illumina solexa 平台转录组测序建库方法 (参考例如 illumina mRNA-Seq Sample Preparation Guide) 构建样品的转录组测序文库, 然后将所得的三个文库分别于 illumina HiSeq2000 测序仪上进行测序, 结果见图 3。图 3 显示了根据本发明一个实施例的大肠杆菌 (*E.coli*) 转录组 Solexa 测序结果。如图 3 所示, 其中, “E.coli 1-3” 分别为利用经过本发明的方法处理的样品 E.coli 1、E.coli2 和 E.coli3 所构建的 3 个转录组测序根据本发明实施例的方法所得到的测序结果, “切胶组” 为利用经过切胶处理的样品所构建的转录组测序文库的测序结果, “对照组” 为利用总 RNA 样品直接构建的转录组测序文库的测序结果。由图 3 可知, 利用三个经过本发明的方法处理的样品所构建的转录组测序文库的测序结果中的 mRNA 比例分别为 27.3%、35.5%、48.3%, 平均值为

37.0%，与利用经过切胶处理的样品所构建的转录组测序文库的测序结果相比 mRNA 比例提高了 3 倍左右，与利用总 RNA 样品直接构建的转录组测序文库的测序结果相比 mRNA 比例提高 30 倍左右。

图 2 和图 3 的结果表明，本发明的富集原核生物 mRNA 的方法，能够有效减少原核生物的总 RNA 中 rRNA 的含量，提高测序数据中的 mRNA 的比例。其中，实验数据表明，处理后的样品中 mRNA 比例相对未经处理的对照组提高 20 倍以上，与琼脂糖凝胶电泳回收法相比，mRNA 比例也提高了 2 倍以上。

三、DSN 酶处理文库

将上述利用经过本发明的方法处理的样品 E.coli 1 所构建的转录组测序文库（在本文中也称为 E.coli 1 文库或 E.coli 1 转录组测序文库），取等量的两份，将其中一份不经 DSN 酶处理，以其即 E.coli 1 文库作为对照，将另一份进行 DSN 酶（双链特异性核酸酶，Evrogen，俄罗斯）处理，以便获得经 DSN 酶处理的 E.coli 1 文库，然后将两个文库分别将于 Illumina HiSeq2000 测序仪上进行测序。

其中，将其中一份 E.coli 1 文库进行 DSN 酶处理是按照下列步骤进行的：

1) 取 4 μ L 上述 E.coli 1 文库（cDNA）至 1.5mL 离心管中，然后添加 4 μ L 杂交缓冲液和 8 μ L 超纯水，混匀后将其进行短暂离心使溶液位于管底；

2) 将上述离心管于 98 $^{\circ}$ C 下进行加热 2min，然后于 68 $^{\circ}$ C 下进行孵育 4~7 小时；

注意：以下的第 3、4 步的加液与混匀操作均需在恒温（68 $^{\circ}$ C）下进行，不能将离心管拿出加热仪（ABI 9700 型 PCR 仪，Applied Biosystems，美国）。

3) 向上述离心管中添加 20 μ L 预热至 68 $^{\circ}$ C 的 DSN 酶反应缓冲液，充分混匀后于 68 $^{\circ}$ C 下进行孵育 10min；

4) 向上述离心管中添加 4 μ L DSN 酶溶液，充分混匀后于 68 $^{\circ}$ C 下进行孵育 25min；

5) 向上述离心管中添加 40 μ L 终止缓冲液，充分混匀后于 68 $^{\circ}$ C 下进行孵育 5min，然后将离心管于冰上静置 5min，并添加 20 μ L 超纯水，混匀后于 -20 $^{\circ}$ C 下进行保存；

6) 取 20 μ L DSN 处理后的 cDNA 作为模板，进行 PCR 扩增（采用 Solexa 测序引物），以便获得扩增产物，即获得经 DSN 酶处理的 E.coli 1 文库。

两个文库的测序结果见图 4。图 4 显示了根据本发明一个实施例的 E.coli 1 转录组测序文库经 DSN 酶处理前后的测序结果。如图 4 所示，处理前的 E.coli 1 文库中 mRNA 比例为 27.3%，经过 DSN 酶处理后 E.coli 1 文库中 mRNA 比例提高至 61.33%，约提高了 2.25 倍，表明 DSN 酶处理对细菌转录组测序文库的构建具有显著影响，可提高文库中 mRNA 的比例，从而能够降低测序成本。

工业实用性

本发明的富集原核生物 mRNA 的方法、构建转录组测序文库的方法、转录组测序文库、确定原核生物 mRNA 序列信息的方法、探针及其在富集 mRNA 中的用途以及试剂盒及其在构建转录组测序文库中的用途，能够有效地应用于原核生物的转录组测序文库的构建以及测序，并且获得的文库质量好，测序结果准确。

尽管本发明的具体实施方式已经得到详细的描述，本领域技术人员将会理解。根据已经公开的所有教导，可以对那些细节进行各种修改和替换，这些改变均在本发明的保护范围之内。本发明的全部范围由所附权利要求及其任何等同物给出。

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示意性实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

权利要求书

1、一种富集原核生物 mRNA 的方法，其特征在于，包括以下步骤：

将核酸样品与探针杂交，所述核酸样品含有原核生物的 16s rRNA、23s rRNA 和所述 mRNA，所述探针携带标记物，并且所述探针能够与所述 16s rRNA 和 23s rRNA 特异性结合以便形成探针-rRNA 复合物；以及

利用特异性结合所述标记物的分子实体去除所述探针-rRNA 复合物，以便富集所述原核生物 mRNA。

2、根据权利要求 1 所述的方法，其特征在于，所述标记物为 polyA，所述分子实体为寡聚 dT。

3、根据权利要求 2 所述的方法，其特征在于，所述分子实体形成于磁珠上。

4、根据权利要求 1 所述的方法，其特征在于，所述探针包括 SEQ ID NO: 1-10 所示的寡核苷酸。

5、根据权利要求 1 所述的方法，其特征在于，在 70-74℃ 下进行所述杂交 25-35 分钟。

6、根据权利要求 5 所述的方法，其特征在于，在 72℃ 下进行所述杂交。

7、根据权利要求 5 所述的方法，其特征在于，进行所述杂交 30 分钟。

8、一种构建转录组测序文库的方法，其特征在于，包括以下步骤：

根据权利要求 1-7 任一项所述的方法，富集 mRNA；以及

针对所得到的 mRNA，构建测序文库，以便获得转录组测序文库。

9、根据权利要求 8 所述的方法，所述转录组测序文库适于利用选自 Illumina Solexa、ABI SOLiD 及 Roche 454 测序平台的至少一种进行测序。

10、根据权利要求 8 所述的方法，其特征在于，进一步包括：

利用双链特异性核酸酶对所述转录组测序文库进行处理。

11、一种转录组测序文库，其是由根据权利要求 8-10 任一项所述的方法构建的。

12、一种确定原核生物 mRNA 序列信息的方法，其特征在于，包括以下步骤：

根据权利要求 8-10 任一项所述的方法构建转录组测序文库；以及

对所述转录组测序文库进行测序，以便获得所述原核生物 mRNA 序列信息。

13、根据权利要求 12 所述的方法，其特征在于，利用选自 Illumina Solexa、ABI SOLiD 及 Roche 454 测序平台的至少一种进行测序。

14、根据权利要求 13 所述的方法，其特征在于，利用 Illumina Solexa 测序平台进行测序。

15、一种探针，其特征在于，所述探针携带标记物，并且能够与原核生物的 16s 和 23s

rRNA 的保守区域特异性结合。

16、根据权利要求 15 所述的探针，其特征在于，由 SEQ ID NO: 1-10 所示的寡核苷酸构成。

17、权利要求 15 或 16 所述的探针在富集 mRNA 中的用途。

5 18、一种试剂盒，其特征在于，包括：

权利要求 15 或 16 所述的探针；以及

分子实体，所述分子实体特异性结合所述探针的标记物。

19、根据权利要求 18 所述的试剂盒，其特征在于，所述标记物为 polyA，所述分子实体为寡聚 dT。

10 20、根据权利要求 19 所述的试剂盒，其特征在于，所述分子实体形成于磁珠上。

21、权利要求 18-20 任一项所述的试剂盒在构建转录组测序文库中的用途。

15

20

25

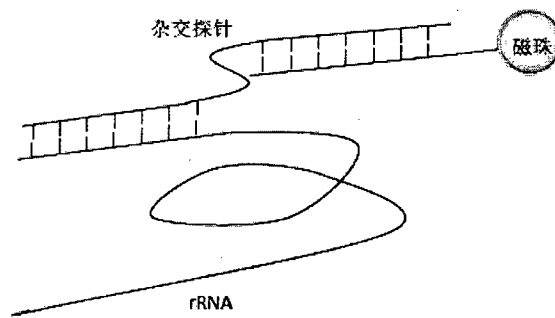


图 1

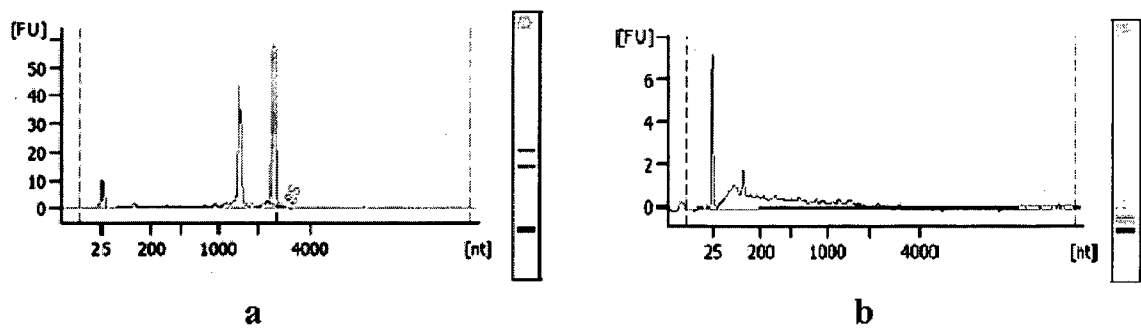
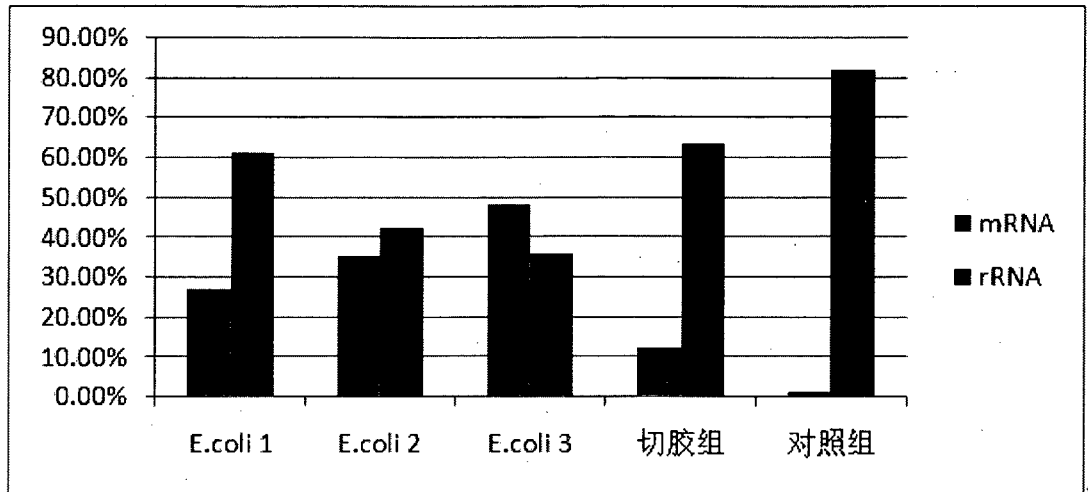
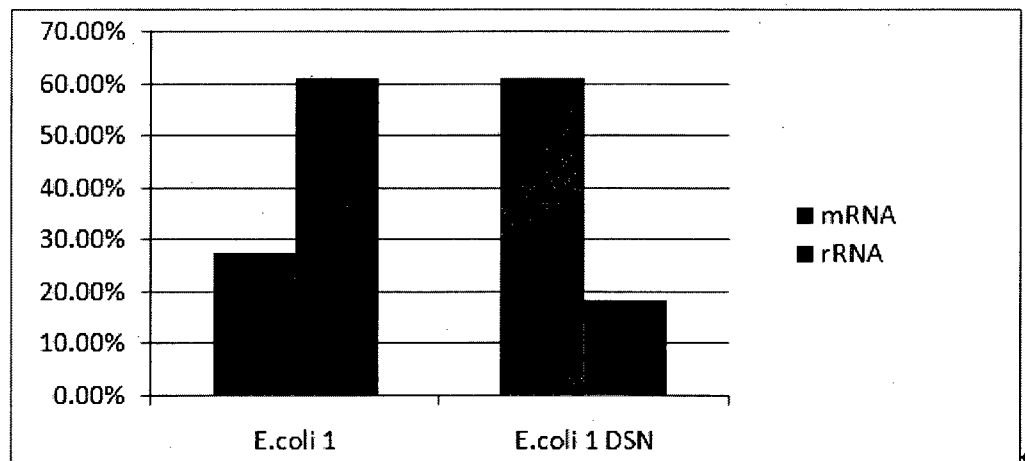


图 2



测序样品	总 reads 数	总比对 reads 数	比对率	比对 mRNA reads 数	比对 rRNA reads 数	mRNA 比例
E. coli 1	19933294	17640962	88.5%	5437742	12203220	27.3%
E. coli 2	27446636	21334636	77.7%	9688996	11645640	35.5%
E. coli 3	33390046	28231084	84.5%	16141336	12089748	48.3%
切胶组	21273866	16226655	76.3%	2647775	13578201	12.4%
对照组	18645908	15590871	83.6%	216783	15342051	1.2%
总计	120689750	99024208	82.1%	34132632	64858860	

图 3



测序样品	总 reads 数	总比对 reads 数	比对率	比对 mRNA reads 数	比对 rRNA reads 数	mRNA 比例
E. coli 1	19933294	17640962	88.5%	5437742	12203220	27.3%
E. coli 1 DSN	20657422	16429511	79.5%	12670167	3759344	61.3%
总计	40590716	34070473	84.0%	18107909	15962564	

图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2011/084237**A. CLASSIFICATION OF SUBJECT MATTER**

See the extra sheet

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: C12N, C07H, C12Q

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)

Data bases: CNABS, CJFD, CNKI, DWPI, SIPOABS, ISI WEB OF KNOWLEDGE, CA, EMBASE, MEDLINE, PUBMED, GENBANK+EMBL+DDBJ+PDB, CHINESE PATENT BIOLOGICAL SEQUENCE SEARCH SYSTEM

Search terms: prokaryon, prokaryot??, bacteri??, blue green algae, cyanobacteri??, rickettsia, chlamydia?, mRNA, isolat+, segregat+, concentrat+, enrich+, gather+, purif+, depurat+, probe?, rRNA, DSN, duplex-specific nuclease?, librar???, sequenc+; sequence search for SEQ ID NO: 1-10

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 1491955 A (INSTITUTE OF MICROBIOLOGY AND EPIDEMIOLOGY, PLA ACADEMY OF MILITARY MEDICAL SCIENCES), 28 April 2004 (28.04.2004) see abstract, and claims 1-10	1, 5-9, 11-15, 17, 18, 21
Y		2, 3, 10, 19, 20
Y	CN 101851786 A (CITY UNIVERSITY OF HONG KONG SHENZHEN RESEARCH INSTITUTE), 06 October 2010 (06.10.2010) see abstract, and embodiment 1	2, 3, 10, 19, 20
A	CN 101640085 A (NANKAI UNIVERSITY), 3 February 2010 (03.02.2010), see the whole document	1-21

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 05 March 2012 (05.03.2012)	Date of mailing of the international search report 15 March 2012 (15.03.2012)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer ZHAO, Shuo Telephone No.: (86-10) 62411070

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2011/084237

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 1882839 A (ATOMIC ENERGY COMMISSION), 20 December 2006 (20.12.2006), see the whole document	1-21
A	TIAN, Zhexian et al., Transcriptome analysis of Sinorhizobium meliloti nodule bacteria in nifA mutant background. CHINESE SCIENCE BULLETIN, 15 September 2006 (15.09.2006), vol. 51, no. 17, pages 2079-2086, ISSN 1001-6538, see the whole document	1-21

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/CN2011/084237

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 1491955 A	28.04.2004	CN 1491955 C	27.12.2006
CN 101851786 A	06.10.2010	None	
CN 101640085 A	03.02.2010	CN 101640085 B	01.06.2011
CN 1882839 A	20.12.2006	EP 1668365 A2	14.06.2006
		EP 2189792 A3	01.09.2010
		CA 2746552 A1	21.04.2005
		FR 2860006 A1	25.03.2005
		JP 2007506956 A	22.03.2007
		US 2010203540 A1	12.08.2010
		CN 1882839 B	10.08.2011
		CN 102225325 A	26.10.2011
		EP 1668365 B1	06.04.2011
		DE 602004032156 D1	19.05.2011
		WO 2005036170 A2	21.04.2005
		US 2006239864 A1	26.10.2006
		US 7687257 B2	30.03.2010
		EP 2189792 A2	26.05.2010
		JP 4659748 B2	30.03.2011

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2011/084237

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/10 (2006.01) i

C12N 15/11 (2006.01) i

C07H 21/02 (2006.01) i

C12Q 1/68 (2006.01) i

A. 主题的分类

参见附加页

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: C12N, C07H, C12Q

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))

数据库: CNABS, CJFD, CNKI, DWPI, SIPOABS, ISI WEB OF KNOWLEDGE, CA, EMBASE, MEDLINE, PUBMED, GENBANK+EMBL+DDBJ+PDB, 中国专利生物序列检索系统

检索词: 原核, 细菌, 蓝藻, 立克次, 衣原体, 信使 RNA, 分离, 富集, 纯化, 探针, 核糖体 RNA, 双链特异性核酸酶, 文库, 测序, prokaryot??. bacteri??. blue green algae, cyanobacteri??. rickettsia, chlamydia?, mRNA, isolat+, segregat+, concentrat+, enrich+, gather+, purif+, depurat+, probe?, rRNA, DSN, duplex-specific nuclease?, librar???, sequenc+; 对 SEQ ID NO: 1-10 的序列检索

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
X	CN1491955A (中国人民解放军军事医学科学院微生物流行病学研究所), 28.4 月 2004 (28.04.2004), 参见摘要, 权利要求 1-10	1,5-9,11-15,17,18,21
Y		2,3,10,19,20
Y	CN101851786A (香港城市大学深圳研究院), 6.10 月 2010 (06.10.2010), 参见摘要, 实施例一	2,3,10,19,20
A	CN101640085A (南开大学), 3.2 月 2010 (03.02.2010), 参见全文	1-21

☒ 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

“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

5.3 月 2012 (05.03.2012)

国际检索报告邮寄日期

15.3 月 2012 (15.03.2012)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局

中国北京市海淀区蓟门桥西土城路 6 号 100088

传真号: (86-10)62019451

受权官员

赵硕

电话号码: (86-10) 62411070

C(续). 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
A	CN1882839A (原子能委员会), 20.12 月 2006 (20.12.2006), 参见全文	1-21
A	TIAN, Zhexian 等, Transcriptome analysis of Sinorhizobium meliloti nodule bacteria in nifA mutant background. Chinese Science Bulletin, 15.9 月 2006 (15.09.2006), 第 51 卷, 第 17 期, 第 2079-2086 页, ISSN 1001-6538, 参见 全文	1-21

国际检索报告

关于同族专利的信息

国际申请号

PCT/CN2011/084237

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN1491955 A	28.04.2004	CN1491955 C	27.12.2006
CN101851786 A	06.10.2010	无	
CN101640085 A	03.02.2010	CN101640085 B	01.06.2011
CN1882839 A	20.12.2006	EP1668365 A2	14.06.2006
		EP2189792 A3	01.09.2010
		CA2746552 A1	21.04.2005
		FR2860006 A1	25.03.2005
		JP2007506956 A	22.03.2007
		US2010203540 A1	12.08.2010
		CN1882839 B	10.08.2011
		CN102225325 A	26.10.2011
		EP1668365 B1	06.04.2011
		DE602004032156 D1	19.05.2011
		WO2005036170 A2	21.04.2005
		US2006239864 A1	26.10.2006
		US7687257 B2	30.03.2010
		EP2189792 A2	26.05.2010
		JP4659748 B2	30.03.2011

A. 主题的分类

C12N 15/10 (2006.01)i

C12N 15/11 (2006.01)i

C07H 21/02 (2006.01)i

C12Q 1/68 (2006.01)i