(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
2023 年 6 月 15 日 (15.06.2023)

WIPO | PCT

(10) 国际公布号
WO 2023/102770 A1

(51) 国际专利分类号:
C12N 15/10 (2006.01)

(21) 国际申请号: PCT/CN2021/136423

(22) 国际申请日: 2021 年 12 月 8 日 (08.12.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 中国科学院深圳先进技术研究院 (**SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY CHINESE ACADEMY OF SCIENCES**) [CN/CN]; 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。 森瑞斯生物科技（深圳）有限公司 (**SENRIS BIOTECHNOLOGY (SHENZHEN) CO., LTD**) [CN/CN]; 中国广东省深圳市南山区招商街道沿山社区南海大道1019号南山医疗器械产业园A、B座B402, Guangdong 518000 (CN)。

(72) 发明人: 罗小舟 (**LUO, Xiaozhou**); 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。 邓华祥 (**DENG, Huaxiang**); 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。 邓艳午 (**DENG, Yanwu**); 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。 余函 (**YU, Han**); 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。 邱玉兰 (**QIU, Yulan**); 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。

WO 2023/102770 A1

(54) **Title:** HIGH-ADDED-VALUE METABOLITE CHASSIS CREATED BASED ON ARTIFICIAL INTELLIGENCE

(54) 发明名称: 基于人工智能创制的高附加值代谢物底盘

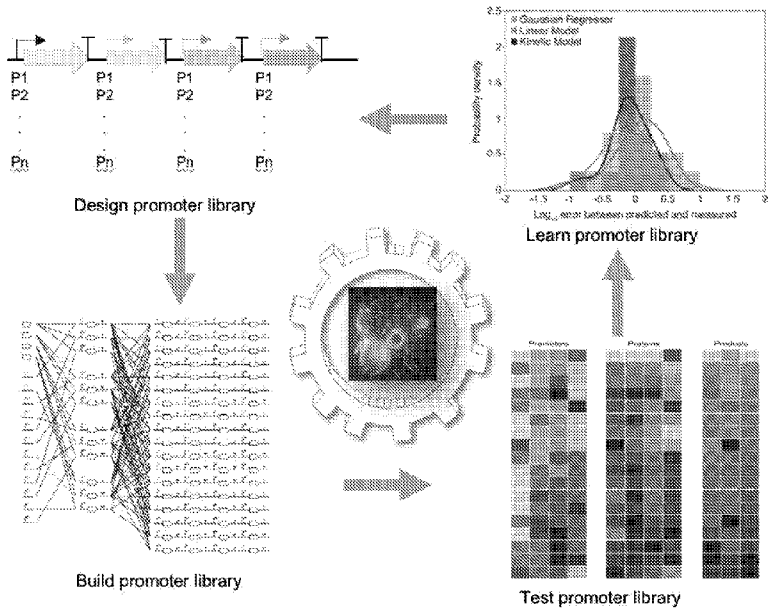


图 1

(57) **Abstract:** A high-added-value metabolite chassis created based on artificial intelligence, relating to the technical fields of synthetic biology and metabolic engineering. By using artificial intelligence technology, by means of few-shot sampling, "design-build-test-learn" closed-loop, cyclic learning of promoter and target metabolite yield data are achieved, a promoter library of high-added-value core metabolic flow key genes is continuously acclimatized, target core metabolic pathway resource allocation is optimized, and a high-yield and high-added-value metabolite chassis is obtained. Also provided are a creation method and a use of said high-added-value metabolite chassis created based on artificial intelligence. The problem that a high-added-value metabolite has no high-throughput molecular probe

[见续页]

(74) 代理人：北京市诚辉律师事务所 (BEIJING CHENGHUI LAW FIRM)；中国北京市朝阳区朝阳北路99号楼2单元905, Beijing 100123 (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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, 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。

(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, 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, 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)。

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))。

or molecular probe screening is unstable can be effectively solved, the labor cost is reduced, the microbial metabolic regulation time is shortened, an inhibitory effect between an intermediate product and a final product is eliminated, the core metabolic flux is enhanced, and efficient synthesis of high-added-value compounds is achieved.

(57) 摘要：基于人工智能创制的高附加值代谢物底盘，属于合成生物学与代谢工程技术领域。提供了采用人工智能技术，通过小样本采样，实现设计-构建-测试-学习闭环式、循环式学习启动子与目标代谢物产量数据，连续驯化高附加值核心代谢流关键性基因的启动子文库，优化目标核心代谢途径资源分配，得到高产高附加值代谢物底盘。还提供了基于人工智能创制的高附加值代谢物底盘的创制方法及其应用。可有效解决高附加值代谢物无高通量分子探针或者分子探针筛选不稳定的问题，减少人力成本，缩短微生物代谢调控时间，消除中间产物与终产物抑制效应，增强核心代谢通量，实现高附加值化合物高效合成。

基于人工智能创制的高附加值代谢物底盘

技术领域

本发明属于合成生物学与代谢工程技术领域，具体涉及基于人工智能创制的高附加值代谢物底盘。

背景技术

萜类、生物碱、聚酮类化合物、黄酮类化合物等天然产物，具有抗菌抗癌抗病毒等生理活性，已广泛应用于食品、农业、医药行业。然而，由于环境条件复杂多变、提取成本高与纯度低等因素，严重阻碍这些高附加值代谢物的市场供给。这些化合物存在结构复杂，化学合成法也面临步骤冗长、合成效率低、副产物多等问题。相比植物提取与化学合成法，微生物合成法因发酵周期短、发酵成本低、环境友好等优势而被广泛关注。随着测序技术、基因编辑技术的跨越式发展，国内外研究者已系统性构建各类底盘（如大肠杆菌、酵母菌、放线菌）的分子操作工具，便于外源途径在微生物异源表达。

外源代谢流引入底盘微生物时，易诱发宿主内源代谢网络扰动，打破细胞氧化还原稳态，且外源途径各基因表达分配不平衡，易导致中间产物积累、代谢溢流等问题。因此，理性设计微生物内源代谢网络与外源途径，可有效扭转微生物目标代谢物产量低的现状。系统生物学技术如基因组代谢网络技术可全局分析微生物代谢网络，便于后期理性改造关键性节点，进而提高目标代谢物产量。例如，Fowler 团队借助CiED 模型预测全局代谢流，结合理性代谢改造关键性节点，实现柚皮素产量达到270 mg/L。因此，智能分配微生物内源代谢资源为目标化合物高效合成的重要基石。

合成生物学成为微生物设计与改造的有效工具，保障高附加值代谢物高效合成。例如，Jae Kyung Sohng团队也系统性的阐释了利用合成生物学工具生产各类天然与非天然的黄酮类化合物。Luo等人借助合成生物学技术在酵母中实现天然与非天然大麻素的高效合成。微生物合成法可有效克服植物来源天然产物供应不足的技术瓶颈。值得注意的是，微生物内源代谢网络复杂，且代谢物合成的细胞生理环境实时变化。当外源高附加值化合物代谢流引入工程菌株，易诱发宿主内源代谢网络扰动，打破细胞氧化还原稳态，诱发毒理环境，限制细胞正常生理活动。此外，外源途径各基因表达分配不平衡，易导致中间产物积累、代谢溢流等问题。因此，需借助系统生物学技术，平衡细胞内源与外源代谢流的资源分配，但是系统生物学技术仅从全局水平提供可调控的关键性节点，后期仍需基因操作技术予以验证。

针对该瓶颈，近期崛起的系统基因组网络操作技术与合成生物学技术加速了微生物高附加值智能细胞工厂的进程。其中，合成生物学的人工智能技术，以“设计-构建-测试-学习”

的循环式理念，革新了微生物代谢调控进程，理性平衡核心代谢流关键性基因表达平衡，可攻克核心代谢流不平衡、效能低等壁垒，智能分配生长代谢与目标代谢流资源，保障高效合成目标代谢物。实现从上游基因元件设计、组装与下游产物测试、再设计等迭代循环回路，有效增强合成生物学理性改造微生物代谢流的深度与广度，实现高附加值代谢物高效合成。目前，运用启动子工程策略为平衡高附加值物质代谢流的有效策略，但是启动子库容量宏大、且无高效、稳定的分子探针系统或化学方法，仅借助高效液相层析、液质联用等技术，难以短时间内攻坚代谢流资源分配平衡的难题。若能借助人工智能策略，即可实现载体构建、菌体培养、小样本采样、数据学习等迭代循环式操作流程，保障高附加值目标代谢通路基因表达水平的连续循环循环过程，解决无高通量筛选方法获取突变体的瓶颈性问题，同时可解放人力，减少人为错误与时间成本，最终，跨越式提升高附加值代谢物产量，满足其市场需求，解决国民健康危机，提高国民生活品质。因此，借助人工智能技术，连续闭环式学习目标代谢通路关键性酶表达与产量数据，可有效平衡细胞生长与核心代谢物合成代谢的资源分配，减少代谢溢流，并理性强化目标代谢物通路，创制微生物智能细胞工厂，实现高附加值化合物高效合成。

植物中约有7000种黄酮类化合物（主要为柚皮素及其衍生物）被验证具有抗氧化活性，可延缓衰老，广泛应用于食品与护肤品行业；也具有抗菌消炎、抗癌症等生理活性，可缓解糖尿病、高血压、高血脂等病痛。综上所述，黄酮类化合物展现出极高的经济效益与医药战略价值。作为黄酮类基本骨架结构，柚皮素合成途径需要4个酶参与，包括酪氨酸氨裂解酶（tyrosine ammonia lyase，TAL）、4-香豆酸CoA连接酶（4-coumarate:CoA ligase，4CL）、查尔酮合成酶（chalcone synthase，CHS）、查尔酮异构酶（chalcone isomerase，CHI）。因此，本发明以柚皮素代谢流为实例，借助人工智能技术连续驯化柚皮素代谢通路基因表达，解除其代谢屏障，实现黄酮类化合物高效合成。本发明可为其余各类高附加值代谢物核心代谢基因表达资源分配，提供理论支撑与技术保障；有效解决其市场供应，保障国民生活品质与健康水准。

发明内容

针对上述现有技术中存在的问题，本发明的目的在于设计提供基于人工智能创制的高附加值代谢物底盘。本发明运用人工智能技术，连续驯化高附加值核心代谢流关键性基因的启动子文库，解除代谢屏障，优化目标核心代谢途径资源分配，平衡其代谢流，实现目标代谢物高效合成。

为了实现上述目的，本发明采用以下技术方案：

基于人工智能创制的高附加值代谢物底盘，其特征在于采用人工智能技术，通过小样本采样，实现设计-构建-测试-学习闭环式、循环式学习启动子与目标代谢物产量数据，连续驯化高附加值核心代谢流关键性基因的启动子文库，优化目标核心代谢途径资源分配，得到高产高附加值代谢物底盘。

所述的基于人工智能创制的高附加值代谢物底盘，其特征在于所述高附加值代谢物包括类黄酮化合物、类黄酮化合物衍生物、类黄酮化合物经修饰后的化合物，所述类黄酮化合物包括柚皮素、花青素、黄芩素或生松素。

所述的高附加值代谢物底盘的创制方法，其特征在于包括以下步骤：

（1）验证不同强度的启动子，建立目标代谢物的关键性基因不同表达水平的启动子文库，采用 Golden gate 酶切与连接技术，在抗生素平板上筛选关键性基因的启动子文库的工程菌株文库；

（2）挑取所述步骤（1）得到的工程菌株文库中的工程菌至深孔板，置于高速摇床，培养过夜，以 0.5-3%接种量接种至发酵培养基中，培养 1-3 天，得到启动子工程菌；

（3）采用化学法或分子探针法，分别在紫外信号与荧光信号初步筛选所述步骤（2）得到的启动子工程菌的目标代谢物产量；

（4）采用 Sanger 测序技术测定不同工程菌的启动子序列，采用高效液相层析技术在紫外信号条件下，测定不同工程菌的目标代谢物产量；

（5）基于机器学习模型学习不同强度启动子的组合分布，得到产量最高的目标代谢物的启动子工程菌。

所述的创制方法，其特征在于所述步骤（1）中验证不同强度的启动子的方法为：以 mKate2 荧光蛋白为分子探针，在 588/633nm 紫外波长下检测不同启动子的荧光信号，分别以阿拉伯糖与 T7 启动子为弱与强的标准启动子，对上述启动子归类为低、中、高三种强度。

所述的创制方法，其特征在于所述步骤（1）中启动子文库的建立方法为：采用高保真、高亲和力的 linker 序列连接关键性基因的表达框，以 ccdB 基因和/或 mKate2 基因作为背景质粒，采用 Golden gate 酶切与连接技术，制备得到启动子文库。

所述的创制方法，其特征在于所述步骤（1）中 Golden gate 酶切与连接的条件为：以高保真、高亲和力的 linker 序列连接各表达框，以 37℃ 5min，16℃ 5min，50 次循环为优化连接条件。

所述的创制方法，其特征在于所述步骤（2）采用自动化技术完成。

所述的创制方法，其特征在于所述步骤（3）中化学法包括 Al^{3+}化学法。

所述的创制方法，其特征在于所述步骤（5）的具体操作方法为：以关键性基因的不同强度的启动子作为模型输入，对应的目标代谢物产量作为模型的输出，基于机器学习模型学习二者的映射关系，对所有组合情况进行预测，产量从高到低排序，选取产量最高的样品实验验证，采用 Golden gate 技术组装样品，重复步骤（2）-（4），验证机器学习样品数据，输入步骤（4）得到的数据，迭代机器学习，得到高产高附加值代谢物底盘。

所述的基于人工智能创制的高附加值代谢物底盘在高效合成黄酮类化合物上的应用。

人工智能策略循环式学习各类高附加值化合物关键性基因表达的普适性技术与方法：小样本采样不同代谢物产量的工程菌株，以不同强度的启动子作为模型的输入，对应的目标产物产量为模型的输出，基于机器学习模型学习它们之间的映射关系，预测所有可能的组合情况，从高到底排序，选取产量最高的样本进行验证；借助高效率组装技术、代谢物初筛技术、HPLC定量技术等验证机器学习样品数据；迭代学习以上所有数据，进而平衡各基因表达水平，增强高附加值代谢通量，实现代谢物高产。

机器学习策略平衡柚皮素关键性基因表达的普适性策略：小样本采样不同柚皮素产量的工程菌，以各工程菌株不同强度的启动子作为模型的输入信号，对应柚皮素产量为模型的输出信号，基于机器学习模型学习对应映射关系，并预测所有可能的组合情况，从高到底排序，选取产量最高的样本进行检验；借助高效率组装技术、代谢物初筛技术、HPLC定量技术等检验机器学习样品数据；迭代学习以上所有数据，进而平衡柚皮素各关键性基因表达水平，实现其高产。

与现有技术相比，本发明具有的有益效果：

（1）本发明借助人工智能技术，实现“设计-构建-测试-学习”四位一体的自主学习过程，为各类高附加值代谢物途径平衡，提供新思路，可有效解决高附加值代谢物无高通量分子探针或者分子探针筛选不稳定的问题，减少人力成本，缩短微生物代谢调控时间，消除中间产物与终产物抑制效应，增强核心代谢通量，实现高附加值化合物高效合成。

（2）本发明的机器学习技术，采用小样本采样及数据迭代训练的闭环式流程，平衡核心代谢流基因表达水平，高产目标代谢物，可有效减少样品采样，降低微生物法合成高附加值物质的人力与时间成本。

（3）本发明的闭环式学习策略可高产柚皮素等高附加值化合物，根本性解决黄酮类化合物市场供应不足的问题，缓解全民健康的危机，减少我国癌症人口数量，抵御新冠病毒的威胁。

附图说明

图 1 为本发明流程图；

图 2 为黄酮类化合物基本骨架结构；

图 3 为柚皮素（Naringenin）合成路径；

图 4 为自动化技术连续进化核心代谢流关键性基因后代谢物产量结果图。

具体实施方式

以下将通过附图和实施例对本发明作进一步说明。

实施例1：

本发明旨在运用人工智能策略，闭环式学习核心代谢流基因启动子文库与产量数据，借以最优分配关键性基因资源，高效合成目标代谢物。通过小样本采样，实现“设计-构建-测试-学习”闭环式、循环式学习启动子与目标代谢物产量数据，借以平衡各类高附加值化合物代谢流关键性基因表达水平，增强核心代谢流通量，解除中间产物、终产物对关键性酶抑制等屏障，高效合成高附加值代谢物。

详细阐述机器学习闭环式、迭代式学习各关键性基因表达水平的普适性策略，流程图如图1所示，其具体技术方案如下所示：

（1）验证不同启动子表达强度。以mKate2荧光蛋白为分子探针，在588/633nm，检测不同启动子荧光信号，并将备选启动子分为低、中、高三种不同强度。

（2）制备关键性基因不同表达水平的启动子文库。以高保真性、高亲和力的linker序列连接各关键性基因的表达框；背景质粒选取ccdB基因与mKate2基因；优化Golden gate酶切与连接条件，在相应抗生素平板筛选不同启动子工程菌文库。

（3）提供基因突变文库挑取与培养技术，即挑取克隆子至96深孔板，放置高速摇床培养过夜。以1%接种量，转移至发酵培养基，培养2天。

（4）提供小样本采样的初筛方法。借助化学法或分子探针法，根据特定荧光与紫外信号，筛选出不同产量的启动子工程菌。

（5）提供不同工程菌株各基因启动子与目标代谢物产量测定的方法。运用Sanger测序技术测定不同工程菌株启动子序列信息；借助高效液相层析技术测定目标高附加值代谢物产量。

（6）基于机器学习模型学习不同强度启动子的组合分布，寻找高附加值物质代谢通路平衡最优解的方法。以各个关键性基因工程菌对应的不同强度的启动子作为模型的输入，对应的目标产物产量为模型的输出，基于机器学习模型学习它们之间的映射关系，并对所有可能的组合情况进行预测，从高到底排序，选取产量最高的样本实验验证。借助Golden gate技术

组装各个样品，根据本发明第三到五方面培养及测定方法，验证机器学习样品数据。重新输入本发明第六方面数据，迭代机器学习，最终，增强高附加值代谢通量，实现代谢物高产。

实施例2：

以柚皮素高附加值代谢物为实例，举例说明人工智能技术闭环式、迭代式学习小样本来源的柚皮素关键性基因启动子文库与产量数据，借以平衡代谢流，最优分配柚皮素关键性基因表达水平，创制高产黄酮类化合物的底盘微生物。黄酮类化合物基本骨架结构如图2所示，柚皮素（Naringenin）合成路径如图3所示。

详细说明机器学习技术连续驯化各类关键性代谢流基因表达的普适性策略，以高附加值的黄酮类化合物（柚皮素）为实例，详实说明机器学习技术闭环式、迭代式学习核心代谢流基因表达文库的流程。

具体包括如下步骤：

（1）本发明提供柚皮素核心代谢流的4个关键性酶：*Rhodotorula glutinis*来源的酪氨酸氨裂解酶（tyrosine ammonia lyase，TAL；KF765779），其核苷酸序列如SEQ ID NO.1所示；*Petroselinum crispum* 来源的4-香豆酸CoA连接酶（4-coumarate:CoA ligase，4CL；KF765780)），其核苷酸序列如SEQ ID NO.2所示；*Petunia X hybrida*来源的查尔酮合成酶（chalcone synthase，CHS；KF765781），其核苷酸序列如SEQ ID NO.3所示；*Medicago sativa*来源的查尔酮异构酶（chalcone isomerase，CHI；KF765782），其核苷酸序列如SEQ ID NO.4所示。

（2）制备柚皮素关键性基因不同表达水平的启动子文库。以高保真性、高亲和力的linker序列连接各关键性基因的表达框；背景质粒选取ccdB基因与mKate2基因；运用Golden gate技术，制备启动子文库，并在相应抗生素平板筛选柚皮素关键性基因不同启动子工程菌株文库。

（3）提供基因突变文库挑取与培养技术，即挑取克隆子至96深孔板，并放置高速摇床培养过夜。以1%接种量，转移至发酵培养基，培养2天。

（4）提供小样本采样的初筛方法。借助Al^{3+}化学法，分别在373 nm紫外信号与382/505 nm的荧光信号筛选不同柚皮素产量的启动子工程菌。

（5）提供不同工程菌株各基因启动子与目标代谢物产量测定的方法。运用Sanger测序技术测定不同工程菌株启动子序列信息；借助高效液相层析技术，在290 nm紫外信号条件下，测定不同工程菌柚皮素产量。

（6）基于机器学习模型学习不同强度启动子的组合分布，寻找柚皮素代谢通路平衡最优解的方法。以柚皮素 4 个关键性基因工程菌对应的不同强度的启动子作为模型的输入，对应的目标产物产量为模型的输出，基于机器学习模型学习它们之间的映射关系，并对所有可能

的组合情况进行预测，从高到底排序，选取产量最高的样本实验验证。借助 Golden gate 技术组装各个样品，根据本发明第九到十一方面培养及测定方法，验证机器学习样品数据。重新输入本发明第十二方面数据，迭代机器学习，最终，增强高附加值代谢通量，实现代谢物高产。

实验结果证明：人工智能技术，使得柚皮素产量提高至 1.32 g/L，如图 4 所示，为目前文献报道的最高水平。

权 利 要 求 书

1、基于人工智能创制的高附加值代谢物底盘，其特征在于采用人工智能技术，通过小样本采样，实现设计-构建-测试-学习闭环式、循环式学习启动子与目标代谢物产量数据，连续驯化高附加值核心代谢流关键性基因的启动子文库，优化目标核心代谢途径资源分配，得到高产高附加值代谢物底盘。

2、如权利要求 1 所述的基于人工智能创制的高附加值代谢物底盘，其特征在于所述高附加值代谢物包括类黄酮化合物、类黄酮化合物衍生物、类黄酮化合物经修饰后的化合物，所述类黄酮化合物包括柚皮素、花青素、黄芩素或生松素。

3、如权利要求 1 或 2 所述的高附加值代谢物底盘的创制方法，其特征在于包括以下步骤：

（1）验证不同强度的启动子，建立目标代谢物的关键性基因不同表达水平的启动子文库，采用 Golden gate 酶切与连接技术，在抗生素平板上筛选关键性基因的启动子文库的工程菌株文库；

（2）挑取所述步骤（1）得到的工程菌株文库中的工程菌至深孔板，置于高速摇床，培养过夜，以 0.5-3%接种量接种至发酵培养基中，培养 1-3 天，得到启动子工程菌；

（3）采用化学法或分子探针法，分别在紫外信号与荧光信号初步筛选所述步骤（2）得到的启动子工程菌的目标代谢物产量；

（4）采用 Sanger 测序技术测定不同工程菌的启动子序列，采用高效液相层析技术在紫外信号条件下，测定不同工程菌的目标代谢物产量；

（5）基于机器学习模型学习不同强度启动子的组合分布，得到产量最高的目标代谢物的启动子工程菌。

4、如权利要求 3 所述的创制方法，其特征在于所述步骤（1）中验证不同强度的启动子的方法为：以 mKate2 荧光蛋白为分子探针，在 588/633nm 紫外波长下检测不同启动子的荧光信号，分别以阿拉伯糖与 T7 启动子为弱与强的标准启动子，对上述启动子归类为低、中、高三种强度。

5、如权利要求 3 所述的创制方法，其特征在于所述步骤（1）中启动子文库的建立方法为：采用高保真、高亲和力的 linker 序列连接关键性基因的表达框，以 ccdB 基因和/或 mKate2 基因作为背景质粒，采用 Golden gate 酶切与连接技术，制备得到启动子文库。

6、如权利要求 3 所述的创制方法，其特征在于所述步骤（1）中 Golden gate 酶切与连接的条件为：以高保真、高亲和力的 linker 序列连接启动子的表达框，以 37℃ 5min，16℃ 5min，50 次循环为优化连接条件。

7、如权利要求 3 所述的创制方法，其特征在于所述步骤（2）采用自动化技术完成。

8、如权利要求 3 所述的创制方法，其特征在于所述步骤（3）中化学法包括 Al^{3+}化学法。

9、如权利要求 3 所述的创制方法，其特征在于所述步骤（5）的具体操作方法为：以关键性基因的不同强度的启动子作为模型输入，对应的目标代谢物产量作为模型的输出，基于机器学习模型学习二者的映射关系，对所有组合情况进行预测，产量从高到低排序，选取产量最高的样品实验验证，采用 Golden gate 技术组装样品，重复步骤（2）-（4），验证机器学习样品数据，输入步骤（4）得到的数据，迭代机器学习，得到高产高附加值代谢物底盘。

10、如权利要求 1 或 2 所述的基于人工智能创制的高附加值代谢物底盘在高效合成黄酮类化合物上的应用。

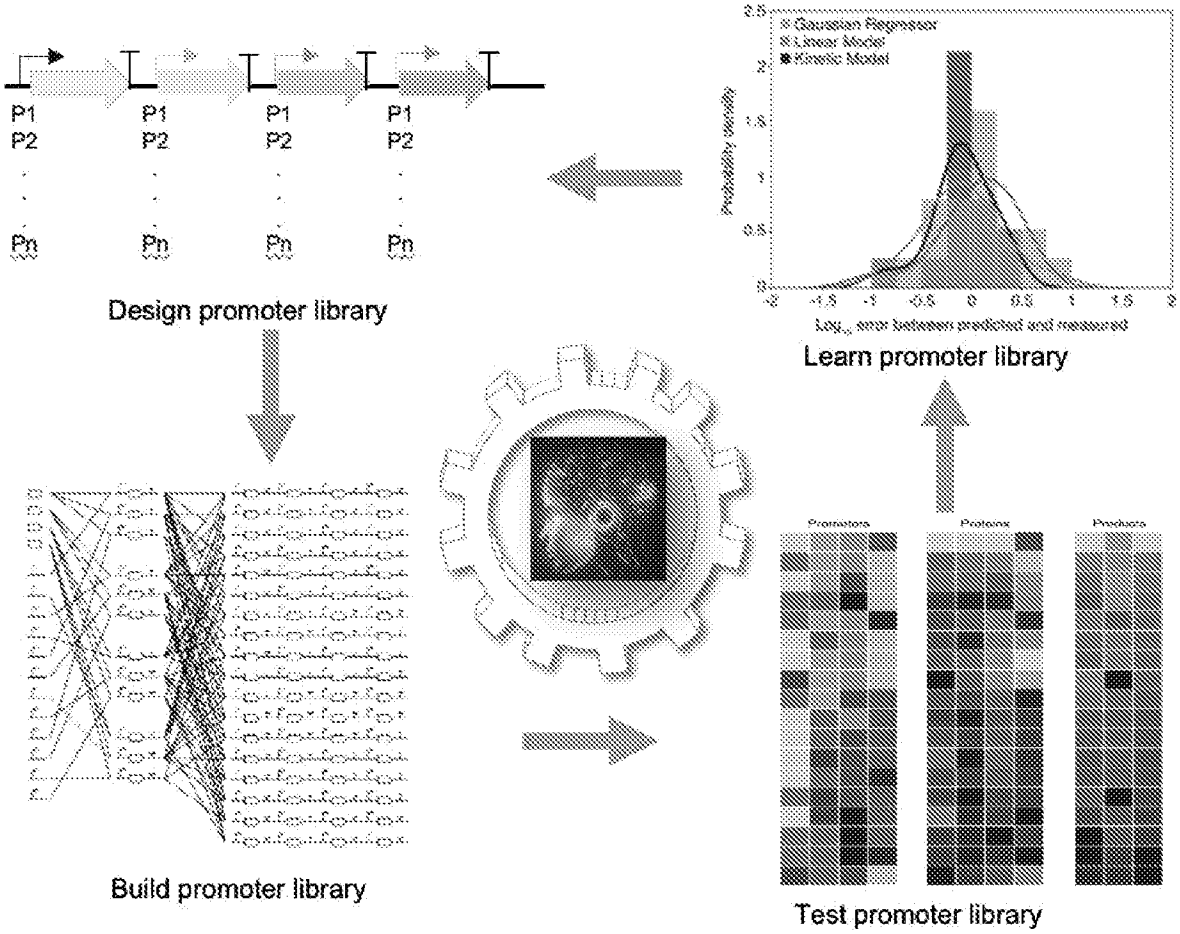

P1
P2
Pn
Design promoter library
Learn promoter library
Build promoter library
Test promoter library

图 1

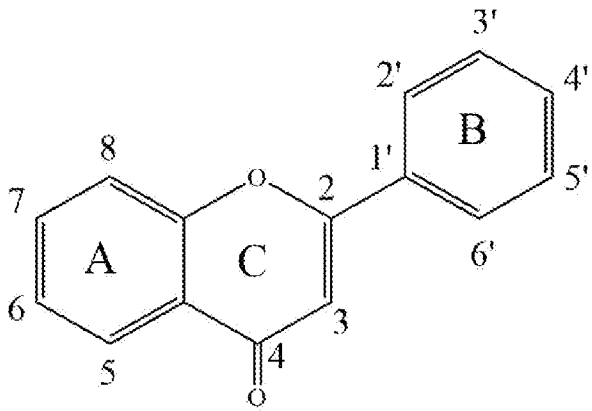

A
B
C
O

图 2

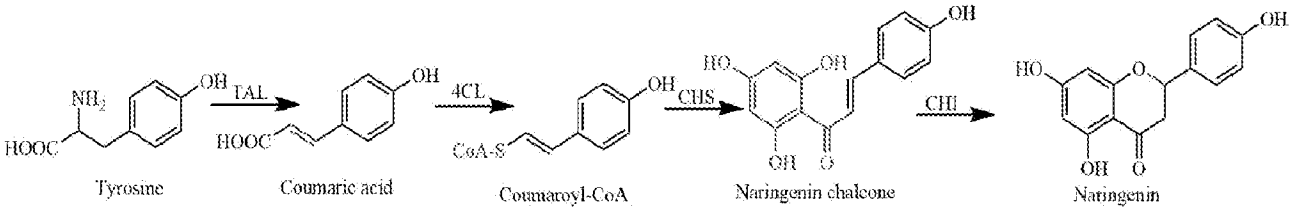

NH2
OH
HOOC
TAL
OH
HOOC
4CL
OH
CoA-S
CHS
HO
OH
OH
O
CHI
HO
O
OH
OH
O
Tyrosine
Coumaric acid
Coumaroyl-CoA
Naringenin chalcone
Naringenin

图 3

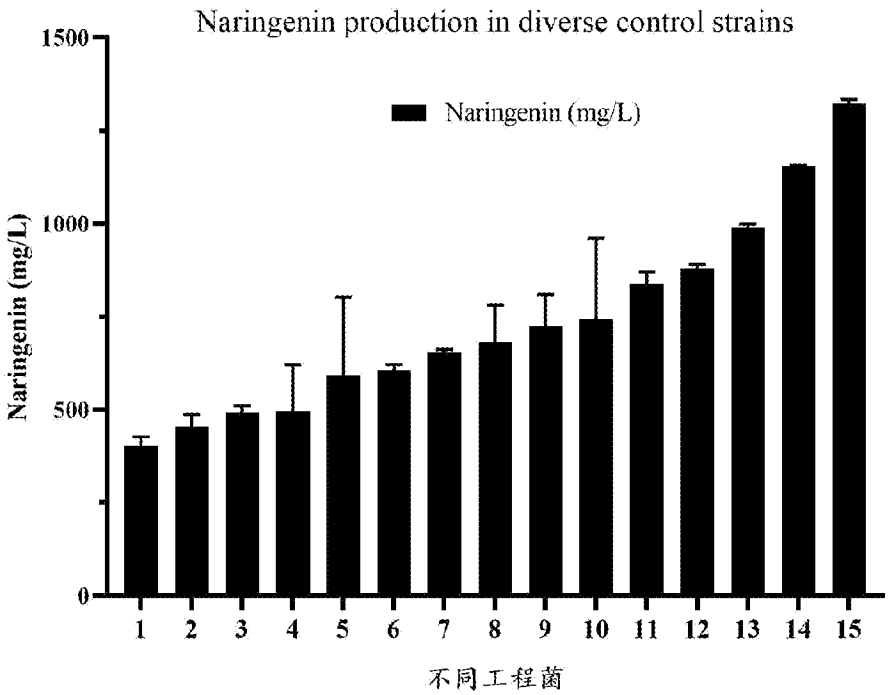

Naringenin production in diverse control strains
Naringenin (mg/L)
Naringenin (mg/L)
1500
1000
500
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
不同工程菌

图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2021/136423

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/10(2006.01)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)

C12N;G06N

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)

WPI; EPODOC; CNKI; CNPAT: 中国科学院深圳先进技术研究院, 森瑞斯生物科技, 罗小舟, 邓华祥, 邓艳午, 余函, 邱玉兰, 合成生物学, DBTL, 启动子, 强度, 筛选, 选择, 挑选, 选取, 产量, 量产, 高产, 工程菌, 菌株, 类黄酮, 柚皮素, 花青素, 黄芩素, 生松素, 代谢物, 代谢产物, 荧光, 萤光, 探针, mKate, 紫外, UV, HPLC, LC, 液相色谱, 学习, 预测, 模型, 发酵, 基因, 代谢途径, synthetic biology, promoter, enhance, strength, screening, selection, yield, production, quantity, bacteria, strain, flavonoid, naringin, anthocyanins, baicalein, sponecin, metabolite, fluorescence, probe, mKate, liquid chromatography, learning, prediction, model, fermentation, gene, pathway, Xiaozhou Luo

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	董俊凯 (DONG, Junkai). "基于酿酒酵母标准化生物元件的代谢途径组装方法及应用 (The Establishment and Applications of the Pathway Assembly Method Based on Standard Biological Elements in Saccharomyces Cerevisiae)" *中国博士学位论文全文数据库 基础科学辑 (Non-official translation: Basic & Science, China Doctoral Dissertations Full-text Database),* No. 04, 15 April 2020 (2020-04-15), abstract, and pp. 35-81	1
Y	董俊凯 (DONG, Junkai). "基于酿酒酵母标准化生物元件的代谢途径组装方法及应用 (The Establishment and Applications of the Pathway Assembly Method Based on Standard Biological Elements in Saccharomyces Cerevisiae)" *中国博士学位论文全文数据库 基础科学辑 (Non-official translation: Basic & Science, China Doctoral Dissertations Full-text Database),* No. 04, 15 April 2020 (2020-04-15), abstract, and pp. 35-81	2-10

☑ 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
"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	Date of mailing of the international search report
17 August 2022	**25 August 2022**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2021/136423

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	周胜虎 (ZHOU, Shenghu). "基于静态与动态调控策略的柚皮素生物合成研究 (Static and Dynamic Regulation Strategies Drive (2S)-Naringenin Biosynthesis)" *中国博士学位论文全文数据库 工程科技I辑 (Engineering Science & Technology I, China Doctoral Dissertations Full-Text Database),* No. 05, 15 May 2020 (2020-05-15), abstract, and pp. 8-67	2-10
A	CN 109554415 A (INSTITUTE OF MICROBIOLOGY, CHINESE ACADEMY OF SCIENCES) 02 April 2019 (2019-04-02) entire document	1-10
A	CN 110878262 A (ZHENGZHOU UNIVERSITY) 13 March 2020 (2020-03-13) entire document	1-10
A	CN 110869502 A (ZYMERGEN INC.) 06 March 2020 (2020-03-06) entire document	1-10
A	CN 110144626 A (ZHEJIANG UNIVERSITY OF TECHNOLOGY) 20 August 2019 (2019-08-20) entire document	1-10
A	US 2020291445 A1 (NORTHWESTERN UNIVERSITY) 17 September 2020 (2020-09-17) entire document	1-10
A	US 2012034661 A1 (THE RESEARCH FOUNDATION OF STATE UNIVERSITY OF NEW YORK et al.) 09 February 2012 (2012-02-09) entire document	1-10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2021/136423

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☑ forming part of the international application as filed:
      ☑ in the form of an Annex C/ST.25 text file.
      ☐ on paper or in the form of an image file.
   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2021/136423

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 109554415 A	02 April 2019	CN 109554415 B	29 April 2022
CN 110878262 A	13 March 2020	CN 110878262 B	18 May 2021
CN 110869502 A	06 March 2020	JP 2020524494 A	20 August 2020
		KR 20200014836 A	11 February 2020
		WO 2018226810 A1	13 December 2018
		EP 3635111 A1	15 April 2020
		US 2020102554 A1	02 April 2020
		CA 3064607 A1	13 December 2018
CN 110144626 A	20 August 2019	None	
US 2020291445 A1	17 September 2020	None	
US 2012034661 A1	09 February 2012	WO 2011140344 A1	10 November 2011
		US 9181539 B2	10 November 2015

国际检索报告

国际申请号 PCT/CN2021/136423

A. 主题的分类

C12N 15/10(2006.01)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)

C12N;G06N

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））

WPI;EPODOC;CNKI;CNPAT: 中国科学院深圳先进技术研究院，森瑞斯生物科技，罗小舟，邓华祥，邓艳午，余函，邱玉兰，合成生物学，DBTL，启动子，强度，筛选，选择，挑选，选取，产量，量产，高产，工程菌，菌株，类黄酮，柚皮素，花青素，黄芩素，生松素，代谢物，代谢产物，荧光，萤光，探针，mKate，紫外，UV，HPLC，LC，液相色谱，学习，预测，模型，发酵，基因，代谢途径，synthetic biology，promoter，enhance，strength，screening，selection，yield，production，quantity，bacteria，strain，flavonoid，naringin，anthocyanins，baicalein，sponecin，metabolite，fluorescence，probe，mKate，liquid chromatography，learning，prediction，model，fermentation，gene，pathway，Xiaozhou Luo

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	董俊凯. "基于酿酒酵母标准化生物元件的代谢途径组装方法及应用." 中国博士学位论文全文数据库 基础科学辑., 第04期，2020年4月15日 (2020-04-15)， 摘要，第35-81页	1
Y	董俊凯. "基于酿酒酵母标准化生物元件的代谢途径组装方法及应用." 中国博士学位论文全文数据库 基础科学辑., 第04期，2020年4月15日 (2020-04-15)， 摘要，第35-81页	2-10
Y	周胜虎. "基于静态与动态调控策略的柚皮素生物合成研究." 中国博士学位论文全文数据库 工程科技Ⅰ辑., 第05期，2020年5月15日 (2020-05-15)， 摘要，第8-67页	2-10

☑ 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	国际检索报告邮寄日期
2022年8月17日	2022年8月25日

ISA/CN的名称和邮寄地址	受权官员
中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	瓮龙明
传真号 (86-10)62019451	电话号码 86-(10)-53962423

国际检索报告

国际申请号
PCT/CN2021/136423

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
A	CN 109554415 A (中国科学院微生物研究所) 2019年4月2日 (2019-04-02) 全文	1-10
A	CN 110878262 A (郑州大学) 2020年3月13日 (2020-03-13) 全文	1-10
A	CN 110869502 A (齐默尔根公司) 2020年3月6日 (2020-03-06) 全文	1-10
A	CN 110144626 A (浙江工业大学) 2019年8月20日 (2019-08-20) 全文	1-10
A	US 2020291445 A1 (NORTHWESTERN UNIVERSITY) 2020年9月17日 (2020-09-17) 全文	1-10
A	US 2012034661 A1 (THE RESEARCH FOUNDATION OF STATE UNIVERSITY OF NEW YORK 等) 2012年2月9日 (2012-02-09) 全文	1-10

国际检索报告

国际申请号
PCT/CN2021/136423

第I栏　核苷酸和/或氨基酸序列(续第1页第1.c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列,国际检索是基于下列序列表进行的：

a. ☑ 作为国际申请的一部分提交的：
   - ☑ 附件C/ST.25文本文件形式
   - ☐ 纸件或图形文件形式

b. ☐ 根据细则13之三.1(a)仅为国际检索目的以附件C/ST.25文本文件形式与国际申请同时提交的：

c. ☐ 仅为国际检索目的在国际申请日之后提交的：
   - ☐ 附件C/ST.25文本文件形式（细则13之三.1(a)）
   - ☐ 纸件或图形文件形式（细则13之三.1(b)和行政规程第713段）

2. ☐ 另外，在提交/提供了多个版本或副本的序列表的情况下，提供了关于随后提交的或附加的副本中的信息与申请时提交的作为申请一部分的序列表的信息相同或未超出申请时提交的申请中的信息范围（如适用）的所需声明。

3. 补充意见：

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2021/136423

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	109554415	A	2019年4月2日	CN	109554415	B	2022年4月29日
CN	110878262	A	2020年3月13日	CN	110878262	B	2021年5月18日
CN	110869502	A	2020年3月6日	JP	2020524494	A	2020年8月20日
				KR	20200014836	A	2020年2月11日
				WO	2018226810	A1	2018年12月13日
				EP	3635111	A1	2020年4月15日
				US	2020102554	A1	2020年4月2日
				CA	3064607	A1	2018年12月13日
CN	110144626	A	2019年8月20日		无		
US	2020291445	A1	2020年9月17日		无		
US	2012034661	A1	2012年2月9日	WO	2011140344	A1	2011年11月10日
				US	9181539	B2	2015年11月10日