

(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



(10) 国际公布号
WO 2021/196357 A1

(43) 国际公布日
2021 年 10 月 7 日 (07.10.2021)

(51) 国际专利分类号:
G16B 30/10 (2019.01) *G16B 40/10* (2019.01)

(21) 国际申请号: PCT/CN2020/090177

(22) 国际申请日: 2020 年 5 月 14 日 (14.05.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

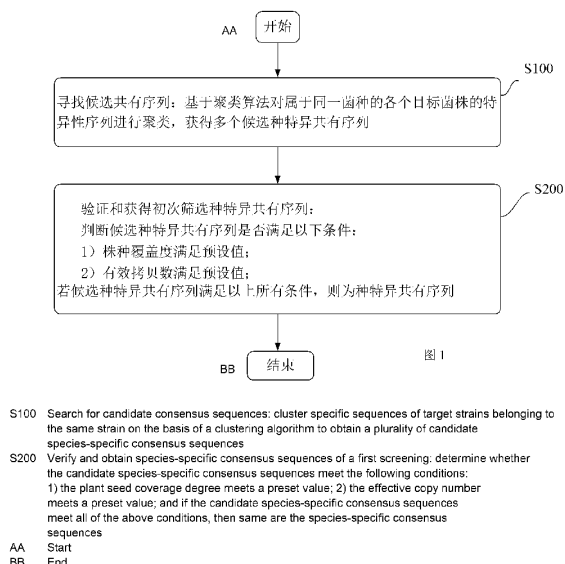
(30) 优先权:
202010254696.6 2020 年 4 月 2 日 (02.04.2020) CN

(71) 申请人: 上海之江生物科技股份有限公司(SHANGHAI ZJ BIO-TECH CO., LTD.) [CN/CN];
中国上海市浦东新区张江高科技园区瑞庆路 528 号 20 幢乙号 1 层、21 幢甲号 1 层, Shanghai 201114 (CN)。

(72) 发明人: 嵇匆(JI, Cong); 中国上海市浦东新区张江高科技园区瑞庆路 528 号 20 幢乙号 1 层、21 幢甲号 1 层, Shanghai 201114 (CN)。邵俊斌(SHAO, Junbin); 中国上海市浦东新区张江高科技园区瑞庆路 528 号 20 幢乙号 1 层、21 幢甲号 1 层, Shanghai 201114 (CN)。刘燕(LIU, Yan); 中国上海市浦东新区张江高科技园区瑞庆路 528 号 20 幢乙号 1 层、21 幢甲号 1 层, Shanghai 201114 (CN)。齐霞(QI, Xia); 中国上海市浦东新区张江高科技园区瑞庆路 528 号 20 幢乙号 1 层、21 幢甲号 1 层, Shanghai 201114 (CN)。金宇丹(JIN, Yudan); 中国上海市浦东新区张江高科技园区瑞庆路 528 号 20 幢乙号 1 层、21 幢甲号 1 层, Shanghai 201114 (CN)。李启腾(LI, Qiteng); 中国上海市浦东新区张江高科技园区瑞庆路 528 号 20 幢乙号 1 层、21 幢甲号 1 层, Shanghai 201114 (CN)。

(54) Title: METHOD AND DEVICE FOR OBTAINING SPECIES-SPECIFIC CONSENSUS SEQUENCES OF MICROORGANISMS AND APPLICATION

(54) 发明名称: 微生物的种特异共有序列的获得方法、装置及应用



(57) Abstract: Provided is a method for obtaining species-specific consensus sequences of microorganisms, at least comprising the following steps: S100, searching for candidate consensus sequences: clustering specific sequences of various target strains belonging to the same strain on the basis of a clustering algorithm to obtain a plurality of candidate species-specific consensus sequences; and S200, verifying and obtaining species-specific consensus sequences of a first screening: determining whether the candidate species-specific consensus sequences meet the following conditions: 1) the plant seed coverage degree meets a preset value; 2) the effective copy number

[见续页]

(74) 代理人: 上海光华专利事务所 (普通合伙) (J.Z.M.C. PATENT AND TRADEMARK LAW OFFICE (GENERAL PARTNERSHIP)); 中国上海市杨浦区国定路 335 号 5022 室余明伟, Shanghai 200433 (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, 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))。

meets a preset value; and if the candidate species-specific consensus sequences meet all of the above conditions, then same are the species-specific consensus sequences. The method is highly sensitive; a repetitive sequence may be found in an incompletely assembled motif; obtained species-specific consensus sequences are accurate, and may be identified to a the subspecies level; the conservative property of identified consensus sequences is high, and the maximum value of the plant seed coverage is achieved as much as possible with the fewest consensus sequences; and all logic modules have multiple verifications, so the accuracy is high.

(57) 摘要: 提供一种微生物的种特异共有序列的获得方法, 至少包括以下步骤: S100, 寻找候选共有序列: 基于聚类算法对属于同一菌种的各个目标菌株的特异性序列进行聚类, 获得多个候选种特异共有序列; S200, 验证和获得初次筛选种特异共有序列: 判断候选种特异共有序列是否满足以下条件: 1) 株种覆盖度满足预设值; 2) 有效拷贝数满足预设值; 若候选种特异共有序列满足以上所有条件, 则为种特异共有序列。方法灵敏度高; 可在组装不完整的基序中寻找重复序列; 获得的种特异性共有序列精确, 可识别到亚种水平; 识别的共有序列保守性强, 以最少的共有序列尽可能达到株种覆盖度的最大值; 所有的逻辑模块都带有多重验证, 准确性高。