



(12) 发明专利申请

(10) 申请公布号 CN 115956129 A

(43) 申请公布日 2023.04.11

(21) 申请号 202180035290.1

(51) Int. Cl.

(22) 申请日 2021.02.03

G12Q 1/6869 (2006.01)

(85) PCT国际申请进入国家阶段日

2022.11.14

(86) PCT国际申请的申请数据

PCT/JP2021/003874 2021.02.03

(87) PCT国际申请的公布数据

W02022/168195 JA 2022.08.11

(71) 申请人 国立大学法人东北大学

地址 日本宫城县

(72) 发明人 松尾步 陶山佳久 佐藤光彦

广田峻

(74) 专利代理机构 北京润平知识产权代理有限公司

11283

专利代理师 刘兵

权利要求书2页 说明书23页

序列表5页 附图14页

(54) 发明名称

遗传信息分析系统和遗传信息分析方法

(57) 摘要

本发明的一个方面涉及一种遗传信息分析系统,其包括:比较侧信息获取部,用于获取表示比较对象序列的比较对象序列信息,该比较对象序列是比较对象的遗传信息的碱基序列;以及准确度获取部,用于基于参考序列信息和所述比较对象序列信息来获取所述比较对象的植物种类与分析对象的植物种类相同的准确度,所述参考序列信息表示从所述分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择的参考序列。

1. 一种遗传信息分析系统,其特征在于,包括:

比较侧信息获取部,用于获取表示比较对象序列的比较对象序列信息,该比较对象序列是比较对象的遗传信息的碱基序列;以及

准确度获取部,用于基于参考序列信息和所述比较对象序列信息来获取所述比较对象的遗传信息与分析对象的遗传信息相同的准确度,所述参考序列信息表示从所述分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择的参考序列。

2. 根据权利要求1所述的遗传信息分析系统,其特征在于,

至少基于第一碱基序列信息以及第二碱基序列信息获取所述参考序列信息,该第一碱基序列信息表示分析对象的遗传信息的独立获取的2个或更多的碱基序列中的一个,该第二碱基序列信息表示所述2个或更多的碱基序列中的另一个碱基序列。

3. 根据权利要求1或2所述的遗传信息分析系统,其特征在于,

所述比较对象的碱基序列的序列信息量比所述参考序列的序列信息量多。

4. 一种遗传信息分析方法,其特征在于,包括:

获取表示比较对象序列的比较对象序列信息的步骤,该比较对象序列是比较对象的遗传信息的碱基序列;以及

基于参考序列信息和所述比较对象序列信息来获取所述比较对象的遗传信息与分析对象的遗传信息相同的准确度的步骤,所述参考序列信息表示从所述分析对象的遗传信息的独立获取的2个或更多的碱基序列中共同的碱基序列中选择的参考序列。

5. 根据权利要求4所述的遗传信息分析方法,其特征在于,

至少基于第一碱基序列信息以及第二碱基序列信息获取所述参考序列信息,该第一碱基序列信息表示分析对象的遗传信息的独立获取的2个或更多的碱基序列中的一个,该第二碱基序列信息表示所述2个或更多的碱基序列中的另一个碱基序列。

6. 根据权利要求4或5所述的遗传信息分析方法,其特征在于,

所述比较对象的碱基序列的序列信息量比所述参考序列的序列信息量多。

7. 一种参考序列信息获取装置,其特征在于,包括:

参考序列信息获取部,用于获取表示参考序列的参考序列信息,该参考序列从分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择。

8. 一种参考序列信息获取方法,其特征在于,包括:

获取表示参考序列的参考序列信息的步骤,该参考序列从分析对象的遗传信息的独立获取的2个或更多的碱基序列中共同的碱基序列中选择。

9. 一种特异序列信息获取系统,其特征在于,包括:

参考序列信息获取部,用于获取表示参考序列的参考序列信息,该参考序列从分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择;以及

特异序列信息获取部,用于从所述参考序列信息中获取特异序列信息,该特异序列信息表示特异性地存在于所述参考序列中的碱基序列。

10. 一种特异序列信息获取方法,其特征在于,包括:

获取表示参考序列的参考序列信息的步骤,该参考序列从分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择;以及

从所述参考序列信息中获取特异序列信息的步骤,该特异序列信息表示特异性地存在

于所述参考序列中的碱基序列。

11. 一种检测用核酸的制造方法,其特征在于,包括:

利用权利要求10所述的方法获取特异序列信息的步骤;以及

根据所述特异序列信息来制造所述分析对象的检测用核酸的步骤。

遗传信息分析系统和遗传信息分析方法

技术领域

[0001] 本发明涉及一种遗传信息分析系统和遗传信息分析方法。

背景技术

[0002] 在农林水产品中准确、迅速且廉价地识别种类、产地、个体、和品种的差异的技术的开发在防止假冒品种(特别是侵害品种培育者权、假冒产地、假冒品质等)方面成为紧迫的课题。以往使用的用于识别种类、产地、个体和品种的DNA异同的方法主要分为以下两种。

[0003] 一种是不需要事先获取DNA信息的方法。作为这种方法,可列举出:(1)检测基于限制酶处理的DNA片段的长度差异的方法(RFLP法等);(2)比较有无利用PCR的扩增片段的方法(RAPD法、ISSR法等);(3)通过上述两种方法的组合利用凝胶电泳等检测DNA片段的有无的方法(AFLP法等)等。这些方法因不需要事先获取信息的作业而具有通用性高的优点,但是,由于缺乏重现性和识别能力等,目前很少使用。

[0004] 近年来,提出了通过使用二代测序仪的基因组分析来进行SNP(s)基因型分型(genotyping)的方法(例如非专利文献1),其被期待为一种迅速且准确地识别品种的差异的方法。

[0005] 另一种是事先获取分析对象种的DNA信息,在对象种的基因组DNA中找出识别有效的区域,制作仅检测(扩增)该区域的PCR引物用于识别的方法(例如,专利文献1、非专利文献1和2)。即,是一种利用仅识别用的标记(DNA标志物)区域的DNA信息进行差异识别的微卫星分析法等方法。在该方法中,一般使用事先确认了能够可靠地检测的区域的DNA变异信息,因此具有检测数据的重现性高的优点。另一方面,该方法存在许多缺点,例如:DNA标志物缺乏通用性,即需要按每个分类群开发标志物;用于选定具有实效性的DNA标志物的筛选作业需要花费大量的时间和费用;如果开发出的DNA标志物的变异性低,则其可能不足以作为识别信息等。

[0006] 现有技术文献

[0007] 专利文献

[0008] 专利文献1:日本专利6220332号公报

[0009] 非专利文献

[0010] 非专利文献1:Suyama Y,Matsuki Y(2015)MIG-seq:an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform.Scientific Reports 5:16963.

发明内容

[0011] 发明要解决的问题

[0012] 近年来,通过使用二代测序仪的SNP基因型分型来识别品种的差异的方法备受瞩目。但是,从二代测序仪获得的碱基序列信息具有错误率高的缺点(错误率:0.1%~16%; Travis C Glenn,Mol Ecol Resour.2011Sep;11(5):759-69.)。此外,由于能够从基因组中

的极多数的区域中得到庞大数量的序列,因此获得最初存在于基因组中多个不同区域中的相同或几乎相同的序列并不罕见。因此,除了已经得到了全基因组信息的生物种以外,特别是在有基因组重复(一种基因组内的一部分被复制而在多个部位存在大致相同的序列的现象)的情况、以及是高阶倍体(基因组整体重复多次的结构)的情况等下,在比较不同的样品间的序列数据时,会产生数据分析上的困难。

[0013] 而且,在使用二代测序仪获取多个全基因组区域的碱基序列数据的方法中,并不是始终得到所有区域的序列数据,通常会发生数据丢失。因此,无法根据碱基序列数据的有无来简单地评价样品之间的基因组的近似性。

[0014] 不限于二代测序仪,由测序仪得到的碱基序列中包含不少错误。因此,基于测序仪获得的碱基序列数据来评价样品之间的相似性,需要选择能够承受近似性评价的高精度的碱基序列数据。

[0015] 因此,本发明的目的在于提供一种能够评价多个生物样品之间的近似性的遗传信息分析系统和遗传信息分析方法。此外,本发明的目的在于提供一种能够提供错误少、精度高的碱基序列数据的参考序列信息获取装置以及参考序列信息获取方法。此外,本发明的课题在于提供一种特异序列信息获取系统和特异序列信息获取方法,用于获取特异于通过上述参考序列信息获取装置或参考序列信息获取方法获取的参考序列的碱基序列。

[0016] 用于解决问题的方案

[0017] 本发明的一个方面涉及一种遗传信息分析系统,其特征在于,包括:比较侧信息获取部,用于获取表示比较对象序列的比较对象序列信息,该比较对象序列是比较对象的遗传信息的碱基序列;以及准确度获取部,用于基于参考序列信息和所述比较对象序列信息来获取所述比较对象的遗传信息与分析对象的遗传信息相同的准确度,所述参考序列信息表示从所述分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择的参考序列。

[0018] 本发明的一个方面涉及上述的基因信息分析系统,其特征在于,至少基于第一碱基序列信息以及第二碱基序列信息获取所述参考序列信息,该第一碱基序列信息表示分析对象的遗传信息的独立获取的2个或更多的碱基序列中的一个,该第二碱基序列信息表示所述2个或更多的碱基序列中的另一个碱基序列。

[0019] 本发明的一个方面涉及上述的基因信息分析系统,其特征在于,所述比较对象的碱基序列的序列信息量比所述参考序列的序列信息量多。

[0020] 本发明的一个方面涉及一种遗传信息分析方法,其特征在于,包括:获取表示比较对象序列的比较对象序列信息的步骤,该比较对象序列是比较对象的遗传信息的碱基序列;以及基于参考序列信息和所述比较对象序列信息来获取所述比较对象的遗传信息与分析对象的遗传信息相同的准确度的步骤,所述参考序列信息表示从所述分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择的参考序列。

[0021] 本发明的一个方面涉及上述的基因信息分析方法,其特征在于,至少基于第一碱基序列信息以及第二碱基序列信息获取所述参考序列信息,该第一碱基序列信息表示分析对象的遗传信息的独立获取的2个或更多的碱基序列中的一个,该第二碱基序列信息表示所述2个或更多的碱基序列中的另一个碱基序列。

[0022] 本发明的一个方面涉及上述的基因信息分析方法,其特征在于,所述比较对象的

碱基序列的序列信息量比所述参考序列的序列信息量多。

[0023] 本发明的一个方面涉及一种参考序列信息获取装置,其特征在于,包括:参考序列信息获取部,用于获取表示参考序列的参考序列信息,该参考序列从分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择。

[0024] 本发明的一个方面涉及一种参考序列信息获取方法,其特征在于,包括:使用上述的参考序列信息获取装置,获取表示参考序列的参考序列信息的步骤,该参考序列从分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择。

[0025] 本发明的一个方面涉及一种特异序列信息获取系统,其特征在于,包括:参考序列信息获取部,用于获取表示参考序列的参考序列信息,该参考序列从分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择;以及特异序列信息获取部,用于从所述参考序列信息中获取特异序列信息,该特异序列信息表示特异性地存在于所述参考序列中的碱基序列。

[0026] 本发明的一个方面涉及一种特异序列信息获取方法,其特征在于,包括:获取表示参考序列的参考序列信息的步骤,该参考序列从分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择;以及从所述参考序列信息中获取特异序列信息的步骤,该特异序列信息表示特异性地存在于所述参考序列中的碱基序列。

[0027] 本发明的一个方面涉及一种检测用核酸的制造方法,包括:利用上述的参考序列信息获取方法获取特异序列信息的步骤;以及根据所述特异序列信息来制造所述分析对象的检测用核酸的步骤。

[0028] 发明效果

[0029] 根据本发明,提供一种能够评价多个生物样品之间的近似性的遗传信息分析系统以及遗传信息分析方法。此外,提供一种能够提供错误少、精度高的碱基序列数据的参考序列信息获取装置以及参考序列信息获取方法。此外,提供一种特异序列信息获取系统和特异序列信息获取方法,用于获取特异于通过上述参考序列信息获取装置或参考序列信息获取方法获取的参考序列的碱基序列。进而,提供一种检测用核酸的制造方法等,该检测用核酸能够基于所述特异序列高精度地检测对象的遗传信息。

附图说明

[0030] 图1是示出了实施方式的遗传信息分析系统100的结构的示例的图。

[0031] 图2是示出了实施方式中的参考序列的获取方法的示例的图。

[0032] 图3是示出了实施方式中的参考序列信息获取装置1的硬件结构的示例的图。

[0033] 图4是示出了实施方式中的评价装置2的硬件结构的示例的图。

[0034] 图5是示出了实施方式中的控制部11的功能结构的示例的图。

[0035] 图6是示出了实施方式中的控制部21的功能结构的示例的图。

[0036] 图7是示出了实施方式中的参考序列信息获取装置1执行的处理流程的示例的流程图。

[0037] 图8是示出了实施方式中的评价装置2执行的处理流程的示例的流程图。

[0038] 图9是示出了实施方式中的特异序列信息获取系统200的结构的示例的图。

[0039] 图10是示出了实施方式中的特异序列信息获取装置3的硬件结构的示例的图。

[0040] 图11是示出了实施方式中的控制部31的功能结构的示例的图。

[0041] 图12是示出了实施方式中的特异序列信息获取装置3执行的处理流程的示例的流程图。

[0042] 图13是示出了MIG-seq法的概要的图(改进自陶山佳永.森林遗传育种学研究中的MIG-seq法的利用.森林遗传育种.2019,8(2):85-89.)。SSR:简单重复序列,ISSR:简单重复序列间,多重PCR:在同一反应中扩增2种以上基因的PCR。

[0043] 图14是示出了实施例中使用的香菇品种的亲子关系的图。

[0044] 图15是示出了在实施例1中使用从C-1获取的参考序列信息进行近似性评价的结果的图。

[0045] 图16是示出了在实施例1中使用从C-2获取的参考序列信息进行近似性评价的结果的图。

[0046] 图17是示出了在实施例1中使用从品种D获取的参考序列信息进行近似性评价的结果的图。

[0047] 图18是示出了使用实施例2中获取的品种D特异性引物,以品种A~E的基因组DNA为模板进行PCR的结果的图。

具体实施方式

[0048] [基因信息分析系统]

[0049] 为了便于说明,以下以分析对象的样品为生物样品的情况为例进行说明。然而,分析对象只要是具有遗传信息的分析对象,就没有特别限定。遗传信息通常是DNA或RNA等碱基序列信息。更具体而言,分析对象包括例如病毒、细菌、古细菌、菌类、藻类、原生动物、植物、动物等。遗传信息包括上述病毒或生物的基因组DNA、基因组RNA、或作为它们的转录产物的mRNA等。下述说明中的“生物样品”是包含来自分析对象生物的遗传信息(基因组DNA、mRNA等)的样品,并且可以是任何形式,只要遗传信息保留即可。生物样品包括但不限于活生物体、干燥体、干燥粉末、生物加工产品、培养细胞等。

[0050] 图1是示出了实施方式的遗传信息分析系统100的结构示例的图。遗传信息分析系统100获取匹配准确度。匹配准确度是指作为分析对象的生物的种类与作为比较对象的生物的种类相同的准确度。即,匹配准确度是表示作为比较对象的生物与作为分析对象的生物的近似程度的大小的评价结果是正确的准确度的信息。匹配准确度示出了作为比较对象的生物的遗传信息与作为分析对象的生物的遗传信息相同的准确度。

[0051] 遗传信息分析系统100包括参考序列信息获取装置1和评价装置2。

[0052] 参考序列信息获取装置1执行参考序列信息获取处理。参考序列信息获取处理是至少根据第一碱基序列信息和第二碱基序列信息这2个分析序列信息,获取表示参考序列的信息(以下称为“参考序列信息”)的处理。

[0053] 分析序列信息是从分析对象的生物样品中获取的遗传信息即核酸的碱基序列。作为遗传信息的核酸可以是DNA或RNA,但优选是基因组DNA。第一碱基序列信息和第二碱基序列信息都是分析序列信息的示例。

[0054] 从分析对象的生物样品中获取分析序列信息的方法没有特别限定,可以使用已知的方法。例如,可以通过从生物样品中提取核酸,根据已知的核酸扩增法(PCR法、等温扩增

法等)等制备序列分析用样品,然后使用测序仪进行序列分析来获取分析序列信息。作为分析序列信息而获取碱基序列的核酸区域没有特别限定,可以获取任意的区域的碱基序列。例如,可以是已知存在基因多态性的基因组区域,或者是全基因组。基因多态性可以是碱基的替换、插入、缺失、以及重复数的变动中的任意一种,也可以是它们的组合。基因多态性可以是单核苷酸多态性(SNP)。作为示例,序列分析用样品也可以包含通过以分析对象的基因组DNA为模板、使用由多个简单重复序列设计的引物进行多重PCR(将2种以上的碱基序列通过同一反应进行扩增的PCR)扩增的DNA片段。

[0055] 序列分析中使用的测序仪可以是基于桑格法(Sanger法)的测序仪,也可以是二代测序仪或三代测序仪,但优选为二代测序仪。二代测序仪是指与使用桑格法进行毛细管电泳的测序仪,即“第一代测序仪”进行对比而使用的术语,是同时并行地处理大量的DNA片段(数千万至数亿个)来确定碱基序列的测序仪。二代测序仪中使用的测序技术包括但不限于离子半导体测序、焦磷酸测序(Pyrosequencing)、基于使用可逆染料终止子的合成的测序、基于连接(ligation)的测序等。三代测序仪中使用的测序技术包括但不限于纳米孔测序(Nat.Biolotechnol.26,1146-1153(2008);SCIENCE,303,1189-1192(2004);Surface Science 432,L611-L616(1999);Nano Lett.,279-285(2011);Science Reports,2,501(2012))等。

[0056] 分析序列信息是基于从同一生物样品中独立获取的2个或更多碱基序列获取的。第一碱基序列信息表示第一碱基序列。第一碱基序列是分析对象的遗传信息的独立获取的2个或更多碱基序列中的一个。第二碱基序列信息表示第二碱基序列。第二碱基序列是分析对象的独立获取的2个或更多碱基序列中的另一个。“独立获取”是指在基于测序仪的序列分析中,作为独立的样品进行序列分析。独立进行序列分析的方法包括:例如在测序仪中分别进行序列分析的方法;以及添加不同的标记序列,使用二代测序仪进行序列分析,并且根据标记序列对获取的碱基序列进行分类的方法等。第一碱基序列以及第二碱基序列例如可以如下获取。

[0057] a) 从生物样品中提取核酸,制备核酸样品,制备序列分析用样品,然后将样品分成2个或更多,分别进行序列分析,获取每个序列分析的碱基序列。将得到的碱基序列中的一个设为第一碱基序列,将另一个设为第二碱基序列。另外,在制备核酸样品后,可以将核酸样品分割为2个或更多,分别进行序列分析用样品的制备和序列分析。此外,也可以将生物样品分成2个或更多,分别进行核酸样品的制备、序列分析用样品的制备和序列分析。

[0058] b) 从生物样品中提取核酸,制备核酸样品,将所述核酸样品分成2个或更多,分别添加不同的标记序列,制备各序列分析用样品,将它们汇总,用二代测序仪进行序列分析。根据标记序列对得到的碱基序列进行分类,按每个标记序列获取碱基序列。将得到的碱基序列中的一个设为第一碱基序列,将另一个设为第二碱基序列。另外,也可以将生物样品分成2个或更多,分别进行核酸样品的制备和序列分析用样品的制备。

[0059] 在上述a)和b)的任一个中,从核酸的提取到碱基序列的获取为止进行的所有操作在第一碱基序列和第二碱基序列的获取中优选在相同条件下进行。

[0060] 第一碱基序列以及第二碱基序列优选通过MIG-seq法(Multiplexed ISSR Genotyping by sequencing;Suyama Y,Matsuki Y(2015)Sciencic Reports 5:16963.)来获取。作为具体例,可举出实施例中记载的方法。

[0061] 当第一碱基序列和第二碱基序列是通过二代测序仪获取的碱基序列时,可以分别选择读取数(读取同一碱基序列的次数)为规定的参考值以上的碱基序列作为第一碱基序列和第二碱基序列。可以认为,读取数越多,碱基序列包含测序错误的风险越低,越可靠。也可以说,读取数越多,碱基序列越能代表分析对象的生物样品的遗传信息。读取数的参考值的实例包括5次或更多、7次或更多、或10次或更多。基于读取数的碱基序列的筛选可以在后述的选择参考序列时进行。

[0062] 为了便于说明,以下以根据第一碱基序列信息和第二碱基序列信息获取参考序列信息的情况为例,对遗传信息分析系统100进行说明。

[0063] 参考序列是从第一碱基序列和第二碱基序列中共同的碱基序列中选择的碱基序列。

[0064] 更具体地,参考序列信息获取处理例如是比较第一碱基序列信息和第二碱基序列信息,检测匹配的碱基序列,获取匹配的碱基序列作为参考序列信息的处理。匹配的碱基序列为参考序列。

[0065] 第一碱基序列以及第二碱基序列可以分别是由多个碱基序列构成的碱基序列的集合。例如,第一碱基序列可以是由碱基序列(1-1)~碱基序列(1-m)构成的m个碱基序列的集合。第二碱基序列可以是由碱基序列(2-1)~碱基序列(2-n)构成的n个碱基序列的集合。参考序列可以是将(1-1)~碱基序列(1-m)与碱基序列(2-1)~碱基序列(2-n)进行对照,选择完全匹配的碱基序列而得到的碱基序列的集合(参见图2)。构成第一碱基序列、第二碱基序列以及参考序列的各个碱基序列的长度没有特别限定,例如可列举出100~500个碱基、100~400个碱基、100~300个碱基、或100~200个碱基等。

[0066] 如果在获取第一碱基序列以及第二碱基序列时没有进行基于读取数的筛选,可以在检测到第一碱基序列以及第二碱基序列的共同序列后,进行前述的基于读取数的筛选。

[0067] 评价装置2执行评价处理。评价处理是基于参考序列信息和表示作为比较对象的生物的碱基序列(以下称为“比较对象序列”)的信息(以下称为“比较对象序列信息”)来获取匹配准确度的处理。匹配准确度是表示近似程度的大小的评价结果是正确的准确度的信息,因此评价处理是进行近似性的评价的处理的示例。

[0068] 比较对象序列是想要获取其与已获取了参考序列的生物样品的匹配准确度的生物的遗传信息的碱基序列。例如,在对附有特定植物品种名的产品进行品种鉴定时,可以从所述产品中获取参考序列,使用从已知为所述植物品种的植物个体获取的碱基序列作为比较对象序列。或者,比较对象序列也可以是作为所述植物品种的碱基序列公开的碱基序列。碱基序列的公开信息例如可以从基因数据库(GenBank、DDBJ等)或文献等中得到。

[0069] 比较对象序列的序列信息量优选与第一碱基序列以及第二碱基序列的序列信息量相同或更多。由此,能够无浪费地使用第一碱基序列和第二碱基序列的序列信息,能够获取更准确的匹配准确度。例如,比较对象序列可以使用包含第一碱基序列和第二碱基序列的核酸区域(例如,基因组DNA区域)的核酸区域的碱基序列。比较对象序列例如能够如下获取。

[0070] A) 从比较对象的生物样品中提取核酸,在与第一碱基序列以及第二碱基序列相同的条件(相同的引物、相同的扩增条件、相同的序列分析条件)下获取碱基序列。

[0071] B) 从比较对象的生物样品中提取核酸,使用与第一碱基序列和第二碱基序列相同

的引物,以比第一碱基序列和第二碱基序列的退火温度更低的退火温度进行核酸扩增,制备序列分析用样品,在与第一碱基序列和第二碱基序列相同的序列分析条件下进行序列分析,获取碱基序列。

[0072] C) 从比较对象的生物样品中提取核酸,在与第一碱基序列和第二碱基序列相同的条件(相同的引物、相同的扩增条件)下进行核酸扩增,制备序列分析用样品,在比第一碱基序列和第二碱基序列更温和的序列分析条件下进行序列分析,获取碱基序列。例如,在比较对象序列中,不进行由测序仪(例如,二代测序仪)读取的序列的选择,或者使选择的程度低于第一碱基序列和第二碱基序列。

[0073] D) 对于作为比较对象的生物,当包含第一碱基序列和第二碱基序列的基因组区域的碱基序列为已知时,该已知的碱基序列从基因数据库、文献等中获取。

[0074] E) 使用作为比较对象的生物的全基因组数据作为比较对象序列。全基因组数据可以从比较对象的生物样品中新获取,也可以从基因数据库或文献等中获取。

[0075] 比较对象序列优选通过MIG-seq法(Multiplexed ISSR Genotyping by sequencing;Suyama Y,Matsuki Y(2015)Sciencic Reports 5:16963.)来获取。作为具体例,可举出实施例中记载的方法。

[0076] 图3是示出了实施方式中的参考序列信息获取装置1的硬件结构的示例的图。参考序列信息获取装置1包括具备经由总线连接的CPU等处理器91和存储器92的控制部11,执行程序。参考序列信息获取装置1通过执行程序,用作包括控制部11、输入部12、通信部13、存储部14、以及输出部15的装置发挥功能。

[0077] 更具体而言,处理器91读取存储于存储部14的程序,并将读取的程序存储于存储器92。通过处理器91执行存储于存储器92的程序,参考序列信息获取装置1用作包括控制部11、输入部12、通信部13、存储部14、以及输出部15的装置发挥功能。

[0078] 控制部11控制参考序列信息获取装置1所具备的各种功能部的操作。例如,控制部11执行参考序列信息获取处理。例如,控制部11将通过执行参考序列信息获取处理获得的参考序列信息记录于存储部14。

[0079] 输入部12被配置为包括鼠标、键盘、触摸面板等输入装置。输入部12也可以被配置为将这些输入装置连接到参考序列信息获取装置1的接口。输入部12接收针对参考序列信息获取装置1的各种信息的输入。

[0080] 通信部13被配置为包括用于将参考序列信息获取装置1连接到外部装置的通信接口。通信部13经由有线或无线与外部装置进行通信。外部装置例如是评价装置2。通信部13通过与评价装置2之间的通信将参考序列信息输出到评价装置2。外部装置例如是第一碱基序列信息和第二碱基序列信息的发送源装置。通信部13通过与第一碱基序列信息和第二碱基序列信息的发送源装置之间的通信,获取第一碱基序列信息和第二碱基序列信息。

[0081] 另外,第一碱基序列信息和第二碱基序列信息不一定需要由参考序列信息获取装置1通过通信部13获取。第一碱基序列信息和第二碱基序列信息可以通过将第一碱基序列信息和第二碱基序列信息输入到输入部12由参考序列信息获取装置1获取。

[0082] 存储部14使用诸如磁性硬盘装置或半导体存储装置等非暂时性计算机可读存储介质装置构成。存储部14存储与参考序列信息获取装置1相关的各种信息。存储部14存储通过输入部12或通信部13输入的信息,例如第一碱基序列信息和第二碱基序列信息。存储部

14存储例如通过执行参考序列信息获取处理而得到的参考序列信息。

[0083] 输出部15输出各种信息。输出部15被配置为包括例如CRT (Cathode Ray Tube: 阴极射线管) 显示器、液晶显示器、有机EL (Electro-Luminescence: 电致发光) 显示器等显示装置。输出部15也可以被配置为将这些显示装置连接到参考序列信息获取装置1的接口。输出部15例如将输入到输入部12或通信部13的信息输出。输出部15也可以显示例如通过执行参考序列信息获取处理而得到的参考序列信息。

[0084] 图4是示出了实施方式中的评价装置2的硬件结构的示例的图。评价装置2包括具备经由总线连接的CPU等处理器93和存储器94的控制部21, 执行程序。评价装置2通过执行程序, 用作包括控制部21、输入部22、通信部23、存储部24、以及输出部25的装置发挥功能。

[0085] 更具体而言, 处理器93读取存储于存储部24的程序, 并将读取的程序存储于存储器94。通过处理器93执行存储于存储器94的程序, 评价装置2用作包括控制部21、输入部22、通信部23、存储部24、以及输出部25的装置发挥功能。

[0086] 控制部21控制评价装置2所具备的各种功能部的操作。例如, 控制部21执行评价处理。控制部21例如将评价处理的执行结果记录于存储部24。

[0087] 输入部22被配置为包括鼠标、键盘、触摸面板等输入装置。输入部22也可以被配置为将这些输入装置连接到评价装置2的接口。输入部22接收针对评价装置2的各种信息的输入。

[0088] 通信部23被配置为包括用于将评价装置2连接到外部装置的通信接口。通信部23经由有线或无线与外部装置进行通信。外部装置例如是参考序列信息获取装置1。通信部23通过与参考序列信息获取装置1之间的通信来获取参考序列信息。外部装置是比较对象序列信息的发送源装置。通信部23通过与比较对象序列信息的发送源装置之间的通信, 获取比较对象序列信息。

[0089] 另外, 参考序列信息和比较对象序列信息不一定需要由评价装置2通过通信部23获取。参考序列信息和比较对象序列信息可以由评价装置2通过将参考序列信息和比较对象序列信息输入到输入部22来获取。

[0090] 存储部24使用诸如磁性硬盘装置或半导体存储装置等非暂时性计算机可读存储介质装置构成。存储部24存储与评价装置2相关的各种信息。存储部24存储通过输入部22或通信部23输入的信息, 例如参考序列信息、比较对象序列信息和参考序列信息等。存储部24例如存储评价处理的执行结果。

[0091] 输出部25输出各种信息。输出部25被配置为包括例如CRT显示器、液晶显示器、有机EL显示器等显示装置。输出部25也可以被配置为将这些显示装置连接到评价装置2的接口。输出部25例如将输入到输入部22或通信部23的信息输出。输出部25也可以显示例如通过执行评价处理而得到的结果。

[0092] 图5是示出了实施方式中的控制部11的功能结构的示例的图。控制部11包括参考侧信息获取部111、参考序列信息获取部112、记录部113、以及输出控制部114。

[0093] 参考侧信息获取部111获取输入到输入部12或通信部13的第一碱基序列信息和第二碱基序列信息。另外, 在第一碱基序列信息以及第二碱基序列信息预先存储于存储部14的情况下, 参考侧信息获取部111也可以通过从存储部14读取第一碱基序列信息以及第二碱基序列信息来获取第一碱基序列信息和第二碱基序列信息。

[0094] 参考序列信息获取部112执行参考序列信息获取处理。参考序列信息获取部112通过执行参考序列信息获取处理,基于第一碱基序列信息和第二碱基序列信息获取参考序列信息。

[0095] 记录部113将各种信息存储于存储部14。记录部113例如将输入到输入部12或通信部13的第一碱基序列信息和第二碱基序列信息记录于存储部14。记录部113例如将参考序列信息记录于存储部14。输出控制部114控制输出部15的操作。通过输出控制部114的操作的控制,输出部15显示例如第一碱基序列信息和第二碱基序列信息。

[0096] 图6是示出了实施方式中的控制部21的功能结构的示例的图。控制部21包括比较侧信息获取部211、准确度获取部212、记录部213、以及输出控制部214。

[0097] 比较侧信息获取部211获取输入到输入部22或通信部23的参考序列信息和比较对象序列信息。另外,在参考序列信息预先存储于存储部24的情况下,比较侧信息获取部211也可以通过从存储部24读取参考序列信息来获取参考序列信息。

[0098] 准确度获取部212执行评价处理。准确度获取部212通过执行评价处理来获取参考序列与分析对象的碱基序列之间的匹配准确度。

[0099] 记录部213将各种信息存储于存储部24。记录部213例如将输入到输入部22或通信部23的参考序列信息和比较对象序列信息记录于存储部24。记录部213例如将参考序列信息记录于存储部24。输出控制部214控制输出部25的操作。通过输出控制部214的操作的控制,输出部25例如显示参考序列信息和比较对象序列信息。

[0100] 图7是示出了实施方式中的参考序列信息获取装置1执行的处理流程的示例的流程图。参考侧信息获取部111获取第一碱基序列信息和第二碱基序列信息(步骤S101)。接着,参考序列信息获取部112执行参考序列信息获取处理(步骤S102)。通过步骤S102的处理,获取参考序列信息。接着,记录部213将参考序列信息记录于存储部14(步骤S103)。

[0101] 图8是示出了实施方式中的评价装置2执行的处理流程的示例的流程图。比较侧信息获取部211获取参考序列信息和比较对象序列信息(步骤S201)。接着,准确度获取部212执行评价处理(步骤S202)。通过步骤S202的处理,获取参考序列与比较对象的碱基序列之间的匹配准确度。接着,输出控制部214控制输出部25的操作以在输出部25上显示匹配准确度(步骤S203)。

[0102] (应用例1)

[0103] 当怀疑以特定品种名流通的产品是否属于该品种时,可以利用遗传信息分析系统100获得匹配准确度,从而进行所述产品的品种鉴定。例如,从产品样品中获取第一碱基序列和第二碱基序列,并使用该品种种的碱基序列作为比较对象序列,从而使用遗传信息分析系统100获取匹配准确度。匹配准确度越高,表示该产品与该品种越近似。如果匹配准确度约为100%(例如,96%或更高、97%或更高、98%或更高、或99%或更高),则可以将该产品鉴定为该品种。

[0104] (应用例2)

[0105] 当认为分析对象样品为特定生物品种中混入了其他品种时,可以通过使用遗传信息分析系统100得到匹配准确度来获得作为分析对象的样品中的其他品种的混入比例。例如,在谷物产品等粒状产品或粉状产品的情况下,有时会混入与品牌名称不同的品种来加量。通过从产品样品获取第一碱基序列和第二碱基序列,并使用该品牌谷物的碱基序列作

为比较对象序列,能够使用遗传信息分析系统100来获取非品牌品种的混入比例。

[0106] (应用例3)

[0107] 在分析对象样品所来自的生物个体存在多个候选亲本的情况下,可以通过使用遗传信息分析系统100得到匹配准确度来进行亲子鉴定。例如,从作为分析对象的生物个体获取第一碱基序列和第二碱基序列,获取参考序列信息。从候选父亲以及候选母亲的各生物个体中分别获取碱基序列,获取这些碱基序列的集合作为比较对象序列信息。使用所述参考序列信息和所述比较对象序列信息,通过遗传信息分析系统100获取匹配准确度。如果匹配准确度约为100% (例如,96%或更高、97%或更高、98%或更高、或99%或更高),则能够判定候选亲本的组合是该生物个体的亲本组合。

[0108] 以这种方式配置的遗传信息分析系统100通过将参考序列与比较对象的碱基序列进行比较来评价分析对象与比较对象的遗传信息的近似性(即近似的程度)或者混入比例。

[0109] 如上所述,参考序列是第一碱基序列和第二碱基序列中共同的碱基序列。因此,即使在用于制备序列分析用样品的核酸扩增、以及测序仪的序列分析后,参考序列也是在两个碱基序列之间匹配的碱基序列。因此,参考序列是比第一碱基序列或第二碱基序列更不易受由于核酸扩增和/或序列分析引起的扩增错误或读取错误的影响的碱基序列。

[0110] 因此,遗传信息分析系统100,其使用参考序列来评价分析对象与比较对象的种类的近似性,能够以更高的精度评价分析对象与比较对象之间的近似性或混入比例。因此,遗传信息分析系统100可以选择能够承受近似性或混入的评价的高精度的碱基序列数据。

[0111] (变形例)

[0112] 另外,参考序列不一定必须是第一碱基序列和第二碱基序列这两个碱基序列中共同的碱基序列,其可以是通过3个或更多独立的序列获取操作从分析对象的样品中获得的3个或更多碱基序列中共同的碱基序列。

[0113] 因此,在参考序列信息获取处理中,不一定仅基于第一碱基序列信息和第二碱基序列信息这2个分析序列信息来获取参考序列信息,也可以基于3个或更多的分析序列信息来获取参考序列信息。

[0114] 另外,参考序列信息获取处理可以反复执行。在这样的情况下,第一碱基序列或第二碱基序列中的至少一个可以是通过紧接在前的参考序列信息获取处理获得的参考序列。这样,参考序列信息获取装置1获取的参考序列可以是参考序列信息获取处理反复执行规定的次数而得到的结果。

[0115] 另外,比较侧信息获取部211可以经由通信部23从存储遗传信息的规定的部211获取比较对象序列信息。

[0116] 从特定的分析对象获取的参考序列信息记录于存储部24,并在从不同的分析对象获取参考序列信息并进行评价处理时可以用作比较对象序列信息。

[0117] 另外,参考序列信息获取装置1和评价装置2不必安装于不同的外壳中,也可以安装为一个装置。例如,控制部11还可以包括比较侧信息获取部211和准确度获取部212,以执行参考序列信息获取处理和评价处理。在这样的情况下,参考序列信息获取装置1用作不仅执行参考序列信息获取处理而且还执行评价处理的装置。

[0118] 另外,遗传信息分析系统100、参考序列信息获取装置1以及评价装置2可以分别使用通过网络可通信地连接的多台信息处理装置来实现。在该情况下,遗传信息分析系统

100、参考序列信息获取装置1以及评价装置2各自所具备的各功能部可以分布安装在多个信息处理装置中。

[0119] 另外,遗传信息分析系统100、参考序列信息获取装置1以及评价装置2的各功能的全部或者部分可以使用ASIC(Application Specific Integrated Circuit:应用规范集成电路)、PLD(Programmable Logic Device:可编程逻辑器件)、FPGA(Field Programmable Gate Array:现场可编程门阵列)等硬件来实现。程序可以记录于计算机可读的记录介质。计算机可读的记录介质例如是软盘、磁光盘、ROM、CD-ROM之类的便携式介质、以及内置于计算机系统内的硬盘等存储装置。程序也可以经由电气通信线路发送。

[0120] [特异序列信息获取系统]

[0121] 利用参考序列,可以识别能够将与分析对象相同的遗传信息与其他遗传信息识别开来的特异序列。为了便于说明,以下以特异序列的识别对象为生物品种的情况为例进行说明。

[0122] 图9是示出了实施方式中的特异序列信息获取系统200的结构示例的图。特异序列信息获取系统200获取表示特异序列的信息,即特异序列信息。特异序列是特异性地存在于分析对象所具有的遗传信息中的序列。例如,在分析对象是特定品种的情况下,特异序列是能够特异性地检测所述特定品种的碱基序列。即,特异序列是能够检测特定品种的遗传信息的检测标志物。

[0123] 特异序列信息获取系统200包括参考序列信息获取装置1和特异序列信息获取装置3。

[0124] 参考序列信息获取装置1与上述相同。

[0125] 特异序列信息获取装置3执行特异序列信息获取处理。特异序列信息获取处理是根据参考序列信息和表示其他品种的碱基序列(以下称为“其他品种序列”)的信息(以下称为“其他品种序列信息”),获取特异于参考序列的碱基序列(仅存在于参考序列,不存在于其他品种序列中的碱基序列)的信息的处理。

[0126] 其他品种序列是与获取参考序列的分析对象的品种不同品种的遗传信息的碱基序列。其他品种序列优选来自多个品种。其他品种序列优选地从与分析对象的品种亲缘度较近的多个品种中获取。由此,可以获得能够将分析对象的品种与高亲缘度的品种识别开来的特异序列。其他品种序列例如能够从碱基序列的公开信息中获取。碱基序列的公开信息可以从基因数据库(GenBank、DDBJ等)或文献等中得到。

[0127] 图10是示出了实施方式中的特异序列信息获取装置3的硬件结构的示例的图。特异序列信息获取装置3包括具备经由总线连接的CPU等处理器95和存储器96的控制部21,执行程序。评价装置2通过执行程序,用作包括控制部21、输入部22、通信部23、存储部24、以及输出部25的装置发挥功能。

[0128] 更具体而言,处理器95读取存储于存储部34的程序,并将读取的程序存储于存储器96。通过处理器95执行存储于存储器96的程序,特异序列信息获取装置3用作包括控制部31、输入部32、通信部33、存储部34、以及输出部25的装置发挥功能。

[0129] 控制部31控制特异序列信息获取装置3所具备的各种功能部的操作。控制部31例如执行特异序列信息获取处理。控制部31例如将特异序列信息获取处理的执行的结果记录于存储部34。

[0130] 输入部32被配置为包括鼠标、键盘、触摸面板等输入装置。输入部32也可以被配置为将这些输入装置连接到特异序列信息获取装置3的接口。输入部32接收针对特异序列信息获取装置3的各种信息的输入。

[0131] 通信部33被配置为包括用于将特异序列信息获取装置3连接到外部装置的通信接口。通信部33经由有线或无线与外部装置进行通信。外部装置例如是参考序列信息获取装置1。通信部33通过与参考序列信息获取装置1之间的通信来获取参考序列信息。外部装置是其他品种序列信息的发送源装置。通信部33通过与其他品种序列信息的发送源装置之间的通信,获取其他品种序列信息。图9中的序列数据库900是其他品种序列信息的发送源的示例。

[0132] 另外,参考序列信息和其他品种序列信息不一定需要由特异序列信息获取装置3通过通信部33获取。参考序列信息和其他品种序列信息可以由特异序列信息获取装置3通过将参考序列信息和其他品种序列信息输入到输入部32获取。

[0133] 存储部34使用诸如磁性硬盘装置或半导体存储装置等非暂时性计算机可读存储介质装置构成。存储部34存储与特异序列信息获取装置3相关的各种信息。存储部34存储通过输入部32或通信部33输入的信息,例如参考序列信息、其他品种序列信息和参考序列信息等。存储部34例如存储特异序列信息获取处理的执行结果。

[0134] 输出部35输出各种信息。输出部35被配置为包括例如CRT显示器、液晶显示器、有机EL显示器等显示装置。输出部35可以被配置为将这些显示装置连接到特异序列信息获取装置3的接口。输出部35例如将输入到输入部32或通信部33的信息输出。输出部35也可以显示例如通过执行特异序列信息获取处理而得到的结果。

[0135] 图11是示出了实施方式中的控制部31的功能结构的示例的图。控制部31包括其他品种侧信息获取部311、特异序列信息获取部312、记录部313、以及输出控制部314。

[0136] 其他品种侧信息获取部311获取输入到输入部32或通信部33的参考序列信息和其他品种序列信息。另外,在参考序列信息预先存储于存储部34的情况下,其他品种侧信息获取部311可以通过从存储部34读取参考序列信息来获取参考序列信息。

[0137] 特异序列信息获取部312执行特异序列信息获取处理。特异序列信息获取部312通过执行特异序列信息获取处理,获取表示仅存在于参考序列中而不存在于其他品种序列中的特异序列的特异序列信息。

[0138] 记录部313将各种信息存储于存储部34。记录部313例如将输入到输入部32或通信部33的参考序列信息和其他品种序列信息记录于存储部34。记录部313例如将参考序列信息记录于存储部34。输出控制部314控制输出部35的操作。通过输出控制部314的操作的控制,输出部35例如显示参考序列信息和其他品种序列信息。

[0139] 图12是示出了实施方式中的特异序列信息获取装置3执行的处理流程的示例的流程图。其他品种侧信息获取部311获取参考序列信息和其他品种序列信息(步骤S301)。接着,特异序列信息获取部312执行特异序列信息获取处理(步骤S302)。通过步骤S302的处理,获取了表示特异序列的特异序列信息,该特异序列可以将参考序列与其他品种序列识别开来。接着,输出控制部314控制输出部35的操作以在输出部35上显示特异序列信息(步骤S303)。

[0140] 以这种方式配置的特异序列信息获取系统200通过将参考序列与其他品种序列进

行比较,获取表示特异序列的特异序列信息,该特异序列为特异性地存在于分析对象中的碱基序列。

[0141] 如上所述,参考序列是比第一碱基序列或第二碱基序列更不易受由于核酸扩增和/或序列分析引起的扩增错误或读取错误的影响的碱基序列。因此,使用参考序列来获取特异序列信息的特异序列信息获取系统200能够以更高的精度识别特异序列。因此,由特异序列信息获取系统200获取的特异序列能够提供高可靠性的分析对象检测技术。

[0142] (变形例)

[0143] 另外,其他品种侧信息获取部311可以经由通信部33从存储其他品种的遗传信息的规定的外部装置获取其他品种序列信息。

[0144] 另外,参考序列信息获取装置1和特异序列信息获取装置3不必安装于不同的外壳中,也可以安装为一个装置。例如,控制部11还可以包括其他品种侧信息获取部311和特异序列信息获取部312,以执行参考序列信息获取处理和特异序列识别处理。在这样的情况下,参考序列信息获取装置1用作不仅执行参考序列信息获取处理而且执行特异序列识别处理的装置。

[0145] 另外,特异序列信息获取系统200、参考序列信息获取装置1以及特异序列信息获取装置3可以分别使用通过网络可通信地连接的多台信息处理装置来安装。在该情况下,特异序列信息获取系统200、参考序列信息获取装置1以及特异序列信息获取装置3各自所具备的各功能部可以分布安装在多个信息处理装置中。

[0146] 另外,特异序列信息获取系统200、参考序列信息获取装置1以及特异序列信息获取装置3的各功能的全部或者部分可以使用ASIC(Application Specific Integrated Circuit:应用规范集成电路)、PLD(Programmable Logic Device:可编程逻辑器件)、FPGA(Field Programmable Gate Array:现场可编程门阵列)等硬件来实现。程序可以记录于计算机可读的记录介质。计算机可读的记录介质例如是软盘、磁光盘、ROM、CD-ROM等便携式介质、以及内置于计算机系统内的硬盘等存储装置。程序也可以经由电气通信线路发送。

[0147] [遗传信息分析方法、参考序列信息获取方法、特异序列信息获取方法]

[0148] 本发明还提供一种遗传信息分析方法。遗传信息分析方法包括:获取表示比较对象序列的比较对象序列信息的步骤,该比较对象序列为比较对象的遗传信息的碱基序列;以及基于表示参考序列的参考序列信息和所述比较对象序列信息,获取表示所述比较对象的遗传信息与所述分析对象的遗传信息相同的准确度(匹配准确度)的步骤,该参考序列从分析对象的遗传信息的独立获取的2个或更多的碱基序列中共同的碱基序列中选择。遗传信息分析方法可以通过遗传信息分析系统100来实施。或者,也可以使用二代测序仪附带的软件等,将所述参考序列信息和比较对象序列信息相互比较,以检测与比较对象序列匹配的参考序列并获取匹配准确度。

[0149] 本发明还提供参考序列信息获取方法。参考序列信息获取方法包括获取表示参考序列的参考序列信息的步骤,该参考序列从分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择。上述步骤也可以通过参考序列信息获取装置1执行参考序列信息获取处理来进行。或者,也可以使用二代测序仪附带的软件等,将所述2个或更多碱基序列相互比较,以识别共同的碱基序列并获取参考序列信息。

[0150] 本发明还提供一种特异序列信息获取方法。特异序列信息获取方法包括:获取表

示参考序列的参考序列信息的步骤,该参考序列从分析对象的遗传信息的独立获取的2个或更多碱基序列中共同的碱基序列中选择;以及从所述参考序列信息中获取特异序列信息的步骤,该特异序列信息表示特异性地存在于所述参考序列中的碱基序列。特异序列信息获取方法也可以通过特异序列信息获取系统300来实施。或者,也可以使用安装了BLAST (Basic Local Alignment Search Tool) 等已知的序列对准算法的程序等,将所述参考序列信息与其他品种序列信息进行比较,检测仅存在于参考序列的特异序列,获取特异序列信息。

[0151] [检测用核酸的制造方法]

[0152] 基于通过特异序列信息获取系统200或特异序列信息获取方法获取的特异序列信息,可以制造特异性地检测分析对象的遗传信息的检测用核酸。检测用核酸可以通过亚磷酸酰胺法等已知的核酸合成法来制造。检测用核酸可以包含特异序列,也可以包含特异序列的互补序列。检测用核酸可以用作用于检测分析对象的遗传信息的引物、或探针等。检测用核酸可以是DNA,也可以是RNA,也可以是DNA与RNA的混合物。检测用核酸并不限于天然核酸,也可以含有人工核酸。人工核酸是指人工合成的具有与核酸相似功能的化合物。人工核酸包括但不限于LNA、BNA、PNA等。

[0153] 实施例

[0154] 以下,通过实施例对本发明进行说明,但本发明并不限于以下的实施例。

[0155] (a) 碱基序列数据的获取

[0156] 在实施例中,使用二代测序仪 (Next Generation Sequencer:NGS),通过MIG-seq法获取碱基序列数据。但是,碱基序列数据的获取方法并不限于此。

[0157] (序列分析用样品的制备)

[0158] 为了制备第一碱基序列获取用的序列分析用样品,从分析对象样品中提取DNA (第一碱基序列用样品)。

[0159] 为了制备第二碱基序列获取用的序列分析用样品,通过独立于所述第一碱基序列获取用样品的操作,从分析对象样品中提取DNA (第二碱基序列用样品)。

[0160] 为了制备比较对象序列获取用的序列分析用样品,从比较对象样品中提取DNA (比较对象序列用样品)。

[0161] 对于所述各样品,构建了DNA片段群(库)作为二代测序仪的序列分析用样品。根据Suyama&Matsuki (2015)的方法,使用MIG-seq法(参照图13:修改自陶山佳永.森林遗传育种学研究中的MIG-seq法的利用.森林遗传育种.2019,8(2):85-89.)进行库构建。其中,第一次PCR的退火温度为38℃。此外,在第二次PCR步骤中添加标记序列(index)是在两侧进行的。

[0162] 作为第一次PCR (ISSR区域的扩增),ISSR (Inter-Simple Sequence Repeat:简单重复序列间)区域之间的PCR扩增使用Suyama&Matsuki (2015)的set-1引物进行。表1显示了第一次PCR中使用的引物。

[0163] [表1]

序列名	序列(5'-3')	序列号
正向引物:	(尾+锚:CTG) + SSR + 锚	
(ACT) ₄ TG-f	<u>CGCTCTTCCGATCTCTG</u> ACTACTACTACTTG	1
(CTA) ₄ TG-f	<u>CGCTCTTCCGATCTCTG</u> CTACTACTACTATG	2
(TTG) ₄ AC-f	<u>CGCTCTTCCGATCTCTG</u> TGTGTTGTTGTTGAC	3
(GTT) ₄ CC-f	<u>CGCTCTTCCGATCTCTG</u> GTTGTTGTTGTTCC	4
(GTT) ₄ TC-f	<u>CGCTCTTCCGATCTCTG</u> GTTGTTGTTGTTTC	5
(GTG) ₄ AC-f	<u>CGCTCTTCCGATCTCTG</u> GTTGTTGTTGTTGAC	6
(GT) ₆ TC-f	<u>CGCTCTTCCGATCTCTG</u> GTTGTTGTTGTTTC	7
[0164] (TG) ₆ AC-f	<u>CGCTCTTCCGATCTCTG</u> TGTGTTGTTGTTGAC	8
反向引物:	(尾+锚:GAC) + SSR + 锚	
(ACT) ₄ TG-r	TGCTCTTCCGATCT <u>GAC</u> ACTACTACTACTTG	9
(CTA) ₄ TG-r	TGCTCTTCCGATCT <u>GAC</u> CTACTACTACTATG	10
(TTG) ₄ AC-r	TGCTCTTCCGATCT <u>GAC</u> TGTGTTGTTGTTGAC	11
(GTT) ₄ CC-r	TGCTCTTCCGATCT <u>GAC</u> GTTGTTGTTGTTCC	12
(GTT) ₄ TC-r	TGCTCTTCCGATCT <u>GAC</u> GTTGTTGTTGTTTC	13
(GTG) ₄ AC-r	TGCTCTTCCGATCT <u>GAC</u> GTTGTTGTTGTTGAC	14
(GT) ₆ TC-r	TGCTCTTCCGATCT <u>GAC</u> GTTGTTGTTGTTTC	15
(TG) ₆ AC-r	TGCTCTTCCGATCT <u>GAC</u> TGTGTTGTTGTTGAC	16

[0165] 作为第一次PCR的试剂,使用了Multiplex PCR Assay Kit Ver.2(Takara Bio)。将反应液组成和反应条件分别示于表2和表3。

[0166] [表2]

模板 DNA	1.00 μL
2x 多重 PCR 缓冲液	3.00 μL
第一次 PCR 引物混合物 (浓度: 各 20 μM)	0.96 μL (最终浓度: 各 0.2 μM)
多重 PCR 酶混合物	0.03 μL
ddH ₂ O	1.01 μl
总计	6.00 μL

[0168] [表3]

1 循环	
94°C	1 min
25 循环	
94°C	30 sec
38°C	1 min
72°C	1 min
1 循环	
72°C	10 min

[0170] 使用AMPure XP (BECKMAN COULTER) 进行珠粒纯化以去除残留在第一次PCR产物中

的夹杂物、选择大小(去除长度200bp以下的短片段)以及平均化样品之间的浓度。纯化步骤遵循该产品的标准协议。

[0171] 为了添加样品识别用的标记序列(Read-1侧9个碱基,Read-2侧5个碱基)和Illumina MiSeq Sequencer(Illumina)用的接头序列(P5和P7序列),使用具有表4所示组成的反应溶液,在表5所示反应条件下进行第二次PCR。用于第二次PCR的引物序列如下所示(由于标记序列按每个样品而不同,因此作为n序列示出)。

[0172] Read-1侧(正向引物):

[0173] 5'-AATGATACGGCGACCACCGAGATCTACACnnnnnnACACTCTTCCCTACACGACGCTCTTCCGATCTCTG-3'(序列号17)

[0174] Read-2侧(反向引物):

[0175] 5'-CAAGCAGAAGACGGCATAACGAGATnnnnnnnnnnGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAC-3'(序列号18)

[0176] [表4]

	纯化的第一次 PCR 产物	1.60 μ L
	第二次 PCR 正向引物 (浓度: 1 μ M)	1.20 μ L (最终浓度: 各 0.2 μ M)
	第二次 PCR 反向引物 (浓度: 1 μ M)	1.20 μ L (最终浓度: 各 0.2 μ M)
[0177]	5 \times PrimeSTAR GXL 缓冲液	1.20 μ L
	dNTP 混合物	0.48 μ L
	PrimerSTAR GXL 聚合酶	0.12 μ L
	ddH ₂ O	0.20 μ L
	总计	6.00 μ L

[0178] [表5]

	12 循环	
[0179]	98°C	10 sec
	54°C	15 sec
	68°C	1 min

[0180] 使用AMPure XP(BECKMAN COULTER)进行珠粒纯化以去除残留在第二次PCR产物中的夹杂物 and 选择大小。与第一次PCR产物不同的是,每个样品(第一碱基序列用样品、第二碱基序列用样品、比较对象序列用样品)的第二次PCR产物都添加了标记序列用于识别。因此,使用每个样品的等效混合物进行第二次PCR产物的纯化。选择的大小设置为400bp或更大,以使Read-1侧和Read-2侧的碱基序列不重叠。

[0181] 为了掌握去除夹杂物和大小选择后库的大致摩尔浓度,使用Qubit 2.0 fluorometer(Invitrogen,Life Technologies)测量DNA重量浓度,通过下式计算摩尔浓度。另外,平均片段大小(碱基序列长度)为550bp,平均分子量为660,其使用微芯片电泳装置MultiNA(岛津)测量。

[0182] [数1]

$$[0183] \quad \text{摩尔浓度}(nM) = \frac{\text{重量浓度} \left(\frac{ng}{\mu L} \right) \times 10^6}{550(\text{bp}) \times 660}$$

[0184] 库的准确的摩尔浓度测量使用Library Quantification Kit (Takara Bio), 在CFX Connect™实时PCR分析系统 (Bio-Rad) 中进行。作为测量的准备工作, 根据从Qubit测量值计算的摩尔浓度制备三个阶段的测量用稀释液, 以达到标准样品的摩尔浓度范围(0.01pM~10pM)。反应液组成和反应条件遵循Library Quantification Kit (TaKaRa Bio) 的标准协议, 标准样品和稀释溶液均重复三次进行测量。使用CFX Connect™实时PCR分析系统计算的摩尔浓度设置为假设平均片段大小为447bp。因此, 库的实际摩尔浓度是使用以下公式从测量的总共9个样品(3阶段的稀释液×3次重复)的平均值计算得出的。

[0185] [数2]

$$[0186] \quad \text{摩尔浓度}(nM) = \frac{\text{SQ(Starting Quantity)} \times (\text{稀释倍率})}{1000} \times \frac{447(\text{bp})}{550(\text{bp})}$$

[0187] 根据Illumina公司的标准协议, 将库验证调整为包含2%的Phi X的12pM库。

[0188] (序列分析)

[0189] 新一代测序使用MiSeq Reagent kit v2 (300cycle, Illumina), 用Illumina MiSeq Sequencer (Illumina) 进行。使用从两端读取库的Paired-End进行测序, 每侧进行117个循环的测序反应。本来, 每侧可以进行150个循环的测序反应, 但由于QV (Quality Value: 质量值) 随着循环次数的增加而降低, 因此在实施例中进行每侧117个循环的测序反应。此外, 由于用MIG-seq法构建的库两端的17个碱基是第一次PCR的引物序列, 其碱基平衡有偏差, 因此从Read-1侧和Read-2侧读取开始的17个碱基设置Dark Cycle, 未获取序列信号(图像)。即, 序列信号在每侧获取100个循环, 在两侧获取200个循环。

[0190] (序列数据的纯化)

[0191] 对如上所述获得的测序数据如下分析。首先, 作为第一步, 对每个样品进行文件分类和序列信息纯化处理(提取用于识别的碱基序列数据)。具体而言, 使用软件bcl2fastq (Illumina), 从所获取的各循环的信号信息生成Fastq文件(存储有由测序读取的碱基序列、各碱基的精度信息(QV值)的文件), 并根据样品识别用的标记序列对Fastq文件进行分类(解复用)。作为此时的选项设定, 将标记序列的允许失配数量从默认值1变更为0。此外, 在包含至少1个标记序列的QV值为30以下的碱基的情况下, 通过自制程序将由与该标记序列相同的集群得到的Read-1侧和Read-2侧的Fastq文件除去。

[0192] 作为第二步, 使用软件Trimmomatic version 0.32 (Anthony et al. 2014), 对每个样品进行生成的Multi-Fastq文件中存储的各碱基序列数据的质量筛选处理。详细的步骤如下所示。

[0193] 由于读取序列的前端部和后端部各有1个碱基的QV值通常较低, 因此将它们删除(删除后的序列长度为每侧98个碱基)。

[0194] 删除了98个以下碱基的短碱基序列和包含接头序列的碱基序列的数据。将Sliding window的参数从默认值4:15变更为4:30, 更严格地删除低QV值的碱基序列(在连续的4个碱基的平均QV值低于30的情况下, 将该碱基序列数据除去)。

[0195] 作为第三步, 将通过了质量筛选处理的Read-1侧的序列和Read-2侧的序列按照相

同的集群进行结合,生成了共计196个碱基的碱基序列数据。但是,在仅单侧的碱基序列数据通过了质量筛选处理的标准的情况下,去除两侧的碱基序列数据。

[0196] (b) 比较对象序列信息的获取

[0197] 对于从比较对象序列获取用样品中得到的数据组,将通过了上述质量筛选处理的碱基序列数据在从同一样品获取的两个或多个数据组之间进行整合,以生成更详尽的碱基序列数据组作为比较对象序列信息的碱基序列数据组。将其用作比较对象序列信息。另外,详尽的碱基序列数据组的碱基序列数与筛选处理后的碱基序列数相等。此外,该碱基序列数据组通常包含大量样品内和跨重复的完全匹配的碱基序列数据,但如果数据量不足以影响数据处理速度,则简单地整合处理。另外,在本实施例中,作为比较对象序列信息的碱基序列数据组,使用通过整合从同一样品获得的两个或多个数据组而获得的数据组,但也可以直接使用一个数据组。

[0198] (c) 参考序列信息的获取

[0199] 关于从第一碱基序列获取用样品和第二碱基序列获取用样品得到的数据组,对于通过上述质量筛选处理的每个碱基序列,为每个样品汇集具有196个碱基完全匹配的碱基序列,包括互补序列。接着,对汇集的每个碱基序列计算读取数(读取完全相同的碱基序列的次数),并将每个样品中读取数小于10次的碱基序列数据作为不可靠数据删除。最后,仅提取从第一碱基序列获取用样品得到的数据组、以及从第二碱基序列获取用样品得到的数据组这2个数据组之间完全匹配的碱基序列数据,生成参考序列信息的数据组。另外,在实施例中,使用第一碱基序列用样品和第二碱基序列获取用样品这2个样品获取参考序列信息,但也可以使用3个或更多样品获取参考序列信息。

[0200] (d) 匹配准确度的获取

[0201] 对通过上述得到的比较对象序列信息的数据组,对照参考序列信息的数据组,搜索参考序列在比较对象序列中完全匹配的碱基序列,求出其数目。接着,根据下述式(1)算出各样品间的近似度(近似性·匹配度)。

[0202] [数3]

$$[0203] \quad \text{匹配准确度}(\%) = \frac{\text{与比较对象序列信息数据组完全匹配的参考序列信息的碱基序列数}}{\text{参考序列信息中的碱基序列总数}} \times 100 \quad \dots (1)$$

[0204] (e) 特异序列信息的获取

[0205] 从参考序列信息中探索特异于分析对象的碱基序列,设计PCR用引物。为了调查所获取的PCR用引物的实效性,将作为比较对象的其他样品和目标样品的DNA样品作为模板,进行基于Fast Cycling PCR Kit(Qiagen)的PCR扩增,调查其扩增的有无或扩增序列的碱基序列信息。将Fast Cycling PCR的反应液组成和条件分别示于表6和表7。

[0206] [表6]

[0207]	模板 DNA	1.00 μL
	快速 PCR Master Mix	3.00 μL
	正向 PCR 引物 (浓度: 20 μM)	0.15 μL (最终浓度: 0.5 μM)
	反向 PCR 引物 (浓度: 1 μM)	0.15 μL (最终浓度: 0.5 μM)
	ddH ₂ O	1.70 μL
	总计	6.00 μL

[0208] [表7]

[0209]	1 循环	
	95°C	5 min
	25 循环	
	96°C	5 sec
	55°C	5 sec
	68°C	15 sec
	1 循环	
	72°C	1 min

[0210] <实施例1:香菇品种样品的基因组近似性分析>

[0211] 以8个品种的食用香菇为对象,实施近似性的评价。另外,在该实施例中,为了确认品种内、品种间的差异识别精度的高低,有意包括了极亲缘的品种对,其被认为难以识别。此外,作为模拟品种鉴定试验用样品,使用已知品种名的2个样品,实施样品间的品种识别。

[0212] (1) 供试菌株样品的选定

[0213] 选定创造品种时的谱系(杂交)关系明确的品种A~H(参照图14)。作为近缘的品种对,品种A~H包括单亲共同的品种A和B、处于亲子关系的品种B和C、同样处于亲子关系的品种G和D和E(G的杂交亲本为D和E)、以及所谓的自交(单核菌丝体和双核菌丝体的杂交育种)亲本品种D和作为其子的F。此外,收集了品种C的2个双核菌丝体样品(C-01和C-02),这些样品已在与创造该品种的种菌公司不同的位置进行继代培养,用于模拟品种鉴定试验。即,将不同品种的8个样品(品种A~H)和模拟品种鉴定用的2个样品(品种C)作为检查对象。

[0214] (2) DNA提取

[0215] 从收集的各菌体分别刮取菌丝,使用改进CTAB法(Murray and Thompson 1980, Saghai-Marroof et al.1986)提取DNA。

[0216] (3) 参考序列信息和比较对象序列信息的获取

[0217] 按照上述(a)~(c),获取参考序列信息和比较对象序列信息。参考序列信息的获取使用了品种C(C-01、C-02)和品种D。从C-01、C-02和品种D获取的参考序列数分别为2091、948和1658。此外,比较对象序列信息的获取使用品种A、B、E、F、G和H。比较对象序列数为407322~1612254。

[0218] (4) 品种间的基因组近似性

[0219] 以谱系关系明确的8个品种的样品为对象,通过上述式(1)算出针对各个品种的多品种的基因组的近似性,将其作为匹配准确度。

[0220] 首先,对品种C算出其模拟品种鉴定用样品即C-01和C-02的匹配准确度,其值接近100%,分别为99.3% (图15:C-01) 和99.9% (图16:C-02)。该结果有力地表明,模拟品种鉴定用样品C-01和C-02为品种C,是正确的品种鉴定结果。另一方面,C-01和C-02相对于品种C以外的品种的近似性的值在51.1%~82.2%的范围内。即,通过显示与其他品种的近似性明显较低的值,示出了模拟品种鉴定用样品C-01和C-02不是品种A、B、D~H中的任一种。

[0221] 此外,对于认为常规方法不能识别的极近缘的品种之间,将作为自交亲本的品种D作为参考侧,将品种D的子的品种F作为比较对象侧计算匹配准确度,结果显示:匹配准确度反映基因组的近似性,是比较高的值,但与上述的同一品种间的值(大致100%)相比,是明显低的值(90.7%),可以将品种D和品种F识别为不同的品种(图17)。另外,在其他亲子关系密切的品种间,它们的匹配准确度的值最大为85.2%,能够清楚地识别品种(图17)。

[0222] <实施例2:香菇品种特异性识别标志物的制作>

[0223] 利用在实施例1中获得的碱基序列数据,制作了目标特异性识别标志物。即,从上述品种中选择品种D作为用于简单地比较样品的基因组和目标品种的基因组的方法的对象。

[0224] (1) 品种D特异性碱基序列的识别和引物的选定

[0225] 通过搜索品种D特异性的碱基序列,以非特异性扩增的有无和扩增效率等信息作为判断标准,选定以下引物组。

[0226] 正向侧

[0227] ATGTATAACCTATCCCACC (序列号19)

[0228] 反向侧

[0229] TTGACGATAATTTTGCTGGG (序列号20)

[0230] (2) 使用品种特异性引物进行的PCR扩增和扩增产物的确认

[0231] 按照上述(e),使用特异于品种D的引物组,对从品种A~H制备的DNA样品进行PCR扩增(参照图14)。结果,在作为检查对象的品种中,仅在作为目标品种的品种D中检测到扩增产物(图18)。由此,可以说至少在对象品种中,产生了能够通过PCR容易地仅识别品种D的标志物。

[0232] <实施例3:海参的亲亲子鉴定>

[0233] (1) 检查对象

[0234] 将海参的20只雄性个体(F01~20)和11只雌性个体(M01~11)在同一个水池中饲养和繁殖。以通过繁殖获得的10只子个体(0-01~10)为检查对象进行亲亲子鉴定。

[0235] (2) DNA提取

[0236] 从20只雄性个体(F01~20)、11只雌性个体(M01~11)和10只子个体(0-01~10)中分别采集细胞,并使用改进CTAB法(Murray and Thompson 1980, Saghai-Marroof et al. 1986)提取DNA。

[0237] (3) 参考序列信息和比较对象序列信息的获取

[0238] 按照上述(a)~(c),获取参考序列信息和比较对象序列信息。参考序列信息的获取使用了10只子个体(0-01~10)。从子个体(0-01~10)获取的参考序列信息的读取数如表8~10所示。

[0239] 比较对象序列信息的获取使用了20只雄性个体(F01~20)和11只雌性个体(M01~

11)。20只雄性个体 (F01~20) 和11只雌性个体 (M01~11) 分别组合制作220种候选亲本的组合。从所述候选亲本组合的雄性个体和雌性个体中获取的比较对象序列信息的集合作为亲子鉴定用的比较对象序列信息数据组。比较对象序列信息数据组的读取数在220000~320000的范围内。

[0240] (4) 亲子鉴定

[0241] 对于从10只子个体 (0-01~10) 中获取的每一个参考序列信息,通过上式(1)分别计算与220个候选亲本组合的比较对象序列信息数据组的匹配准确度。表8~10示出了各子个体的匹配准确度最高的10个候选亲本的组合。

[0242] [表8]

子个体 读取数	O-01		O-02		O-03		O-04	
排名	候选亲本	匹配准确度 (%)	候选亲本	匹配准确度 (%)	候选亲本	匹配准确度 (%)	候选亲本	匹配准确度 (%)
1	F05-M04	100.00	F05-M10	100.00	F05-M11	100.00	F05-M06	100.00
2	F08-M04	88.11	F08-M10	88.92	F08-M11	85.60	F08-M06	89.17
3	F16-M04	87.84	F16-M10	86.43	F05-M03	82.83	F16-M06	86.67
4	F15-M04	83.24	F12-M10	84.49	F16-M11	81.99	F04-M06	83.89
5	F04-M04	82.97	F03-M10	82.83	F05-M08	81.72	F05-M01	83.89
6	F12-M04	82.97	F04-M10	82.83	F05-M10	81.44	F17-M06	83.89
7	F03-M04	81.89	F06-M10	82.55	F05-M02	80.89	F11-M06	83.61
8	F11-M04	81.89	F11-M10	82.55	F04-M11	80.61	F05-M07	83.33
9	F18-M04	81.89	F09-M10	82.27	F05-M07	80.33	F05-M04	83.06
10	F20-M04	81.89	F18-M10	82.27	F05-M09	80.33	F03-M06	82.78

[0244] [表9]

子个体 读取数	O-05		O-06		O-07		O-08	
排名	候选亲本	匹配准确度 (%)	候选亲本	匹配准确度 (%)	候选亲本	匹配准确度 (%)	候选亲本	匹配准确度 (%)
1	F17-M02	100.00	F05-M09	100.00	F05-M11	100.00	F05-M11	100.00
2	F08-M02	84.87	F08-M09	87.92	F08-M11	85.88	F05-M03	83.19
3	F04-M02	82.63	F16-M09	86.80	F16-M11	83.05	F08-M11	82.91
4	F17-M03	82.63	F05-M05	84.27	F15-M11	81.07	F05-M05	81.77
5	F02-M02	81.79	F06-M09	83.71	F04-M11	80.79	F05-M02	81.20
6	F09-M02	80.67	F03-M09	83.43	F13-M11	79.94	F05-M08	81.20
7	F14-M02	80.67	F12-M09	83.15	F12-M11	79.38	F05-M07	80.91
8	F06-M02	80.11	F17-M09	83.15	F03-M11	79.10	F16-M11	80.91
9	F17-M09	80.11	F18-M09	82.87	F17-M11	78.53	F05-M10	80.63
10	F16-M02	79.83	F04-M09	82.58	F06-M11	78.25	F05-M09	80.06

[0246] [表10]

子个体 读取数	O-09		O-10	
	348		244	
排名	候选亲本	匹配准确度 (%)	候选亲本	匹配准确度 (%)
1	F05-M02	100.00	F17-M11	100.00
2	F08-M02	88.22	F17-M03	86.07
3	F16-M02	86.21	F17-M07	86.07
4	F06-M02	83.62	F17-M04	84.02
5	F04-M02	82.76	F17-M06	83.20
6	F12-M02	82.47	F04-M11	82.79
7	F13-M02	82.47	F17-M08	82.79
8	F05-M03	81.61	F17-M02	82.38
9	F10-M02	81.61	F17-M05	82.38
10	F01-M02	81.32	F08-M11	81.97

[0247] 如表8~10所示,对于任意的子个体,存在1组匹配准确度100%的候选亲本的组合。推定该候选亲本的组合是子个体的亲本的组合。

[0248] 以上,参照附图对本发明的实施方式进行了详细叙述,但具体的结构不限于该实施方式,也包括不脱离本发明的主旨的范围的设计等。

[0249] 符号说明

[0250] 100 遗传信息分析系统

[0251] 1 参考序列信息获取装置

[0252] 2 评价装置

[0253] 3 特异序列信息获取装置

[0254] 11 控制部

[0255] 12 输入部

[0256] 13 通信部

[0257] 14 存储部

[0258] 15 输出部

[0259] 21 控制部

[0260] 22 输入部

[0261] 23 通信部

[0262] 24 存储部

[0263] 25 输出部

[0264] 31 控制部

[0265] 32 输入部

[0266] 33 通信部

[0267] 34 存储部

[0268] 35 输出部

[0269] 111 参考侧信息获取部

[0270] 112 参考序列信息获取部

[0271] 113 记录部

- [0273] 114 输出控制部
- [0274] 211 比较侧信息获取部
- [0275] 212 准确度获取部
- [0276] 213 记录部
- [0277] 214 输出控制部
- [0278] 311 其他品种侧信息获取部
- [0279] 312 特异序列信息获取部
- [0280] 313 记录部
- [0281] 314 输出控制部
- [0282] 91 处理器
- [0283] 92 存储器
- [0284] 93 处理器
- [0285] 94 存储器
- [0286] 95 处理器
- [0287] 96 存储器。

SEQUENCE LISTING

<110>	国立大学法人东北大学	
<120>	遗传信息分析系统和遗传信息分析方法	
<130>	PC-31039	
<160>	20	
<170>	PatentIn version 3.5	
<210>	1	
<211>	31	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	1	
	cgctcttccg atctctgact actactactt g	31
<210>	2	
<211>	31	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	2	
	cgctcttccg atctctgcta ctactactat g	31
<210>	3	
<211>	31	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	3	
	cgctcttccg atctctgttg ttggttga c	31
<210>	4	
<211>	31	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	4	
	cgctcttccg atctctggtt gttgttgttc c	31

<210>	5	
<211>	31	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	5	
	cgctcttccg atctctgggtt gttggtgttt c	31
<210>	6	
<211>	31	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	6	
	cgctcttccg atctctgggtg gtggtggtga c	31
<210>	7	
<211>	31	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	7	
	cgctcttccg atctctgggtg tgtgtgtgtt c	31
<210>	8	
<211>	31	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	8	
	cgctcttccg atctctgtgt gtgtgtgtga c	31
<210>	9	
<211>	31	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	9	

tgctcttccg atctgacact actactactt g	31
<210> 10	
<211> 31	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 10	
tgctcttccg atctgaccta ctactactat g	31
<210> 11	
<211> 31	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 11	
tgctcttccg atctgacttg ttgttgttga c	31
<210> 12	
<211> 31	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 12	
tgctcttccg atctgacgtt gttgttgttc c	31
<210> 13	
<211> 31	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 13	
tgctcttccg atctgacgtt gttgttgttt c	31
<210> 14	
<211> 31	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	

<400> 14	
tgctcttccg atctgacgtg gtggtggtga c	31
<210> 15	
<211> 31	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 15	
tgctcttccg atctgacgtg tgtgtgtggtt c	31
<210> 16	
<211> 31	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 16	
tgctcttccg atctgactgt gtgtgtgtga c	31
<210> 17	
<211> 70	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<220>	
<221> misc_feature	
<222> (30) .. (34)	
<223> n is a, c, g, or t	
<400> 17	
aatgatacgg cgaccaccga gatctacaen mnnnacactc tttccctaca cgacgctctt	60
ccgatctctg	70
<210> 18	
<211> 70	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<220>	
<221> misc_feature	

<222>	(25) .. (33)	
<223>	n is a, c, g, or t	
<400>	18	
	caagcagaag acggcatacg agatnnnnn nngtgactg gagttcagac gtgtgctctt	60
	ccgatctgac	70
<210>	19	
<211>	19	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	19	
	atgtataacc tatcccacc	19
<210>	20	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成引物	
<400>	20	
	ttgacgataa ttttgctggg	20

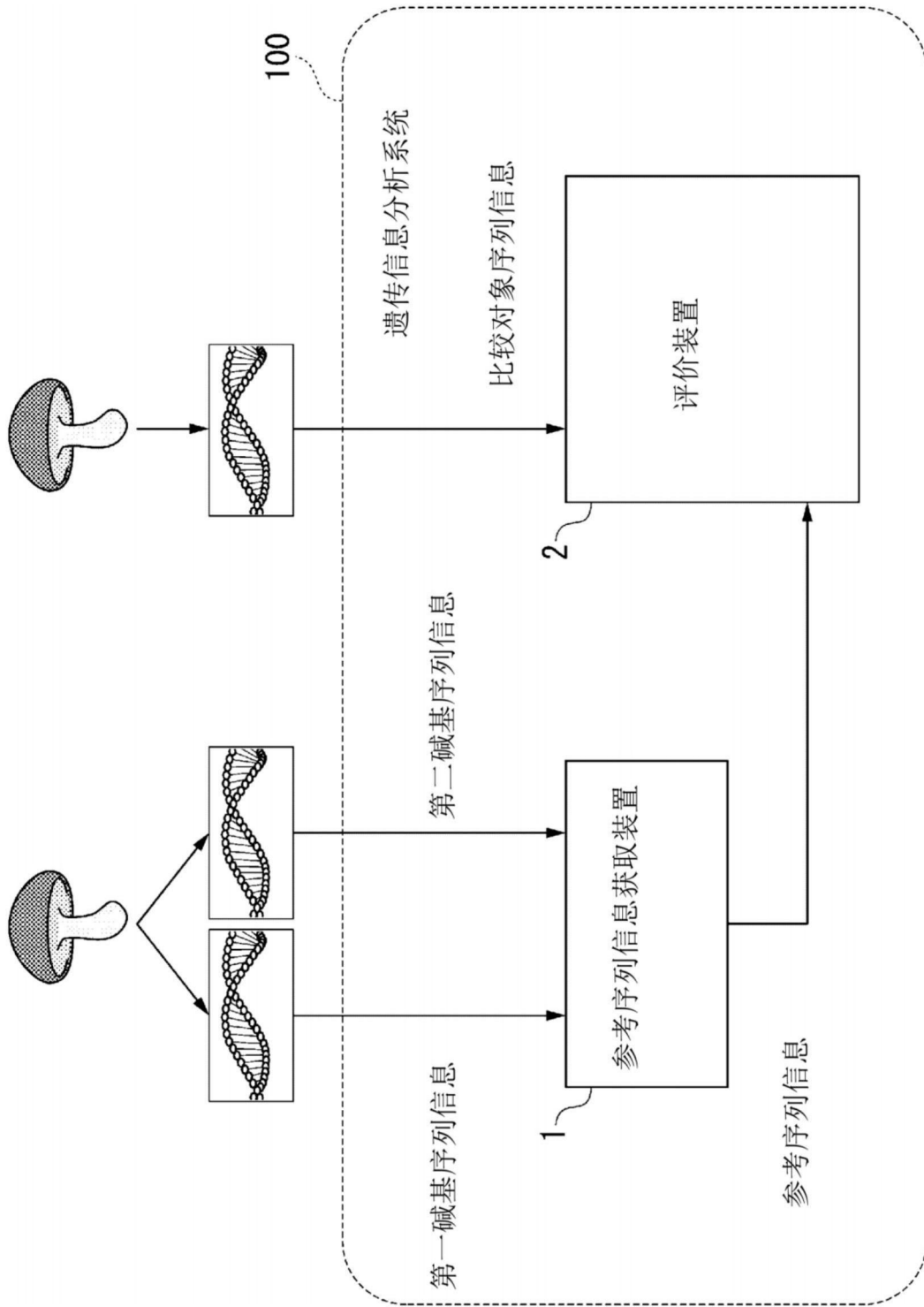


图1

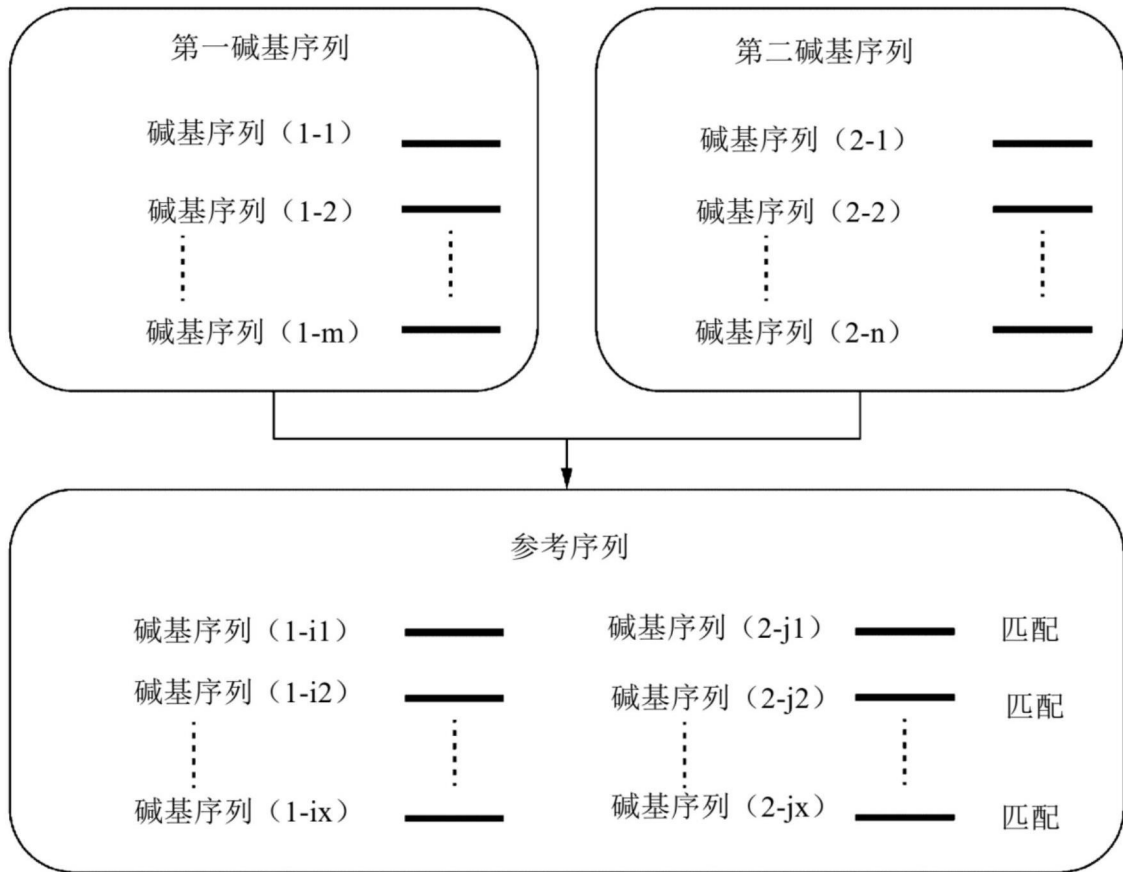


图2

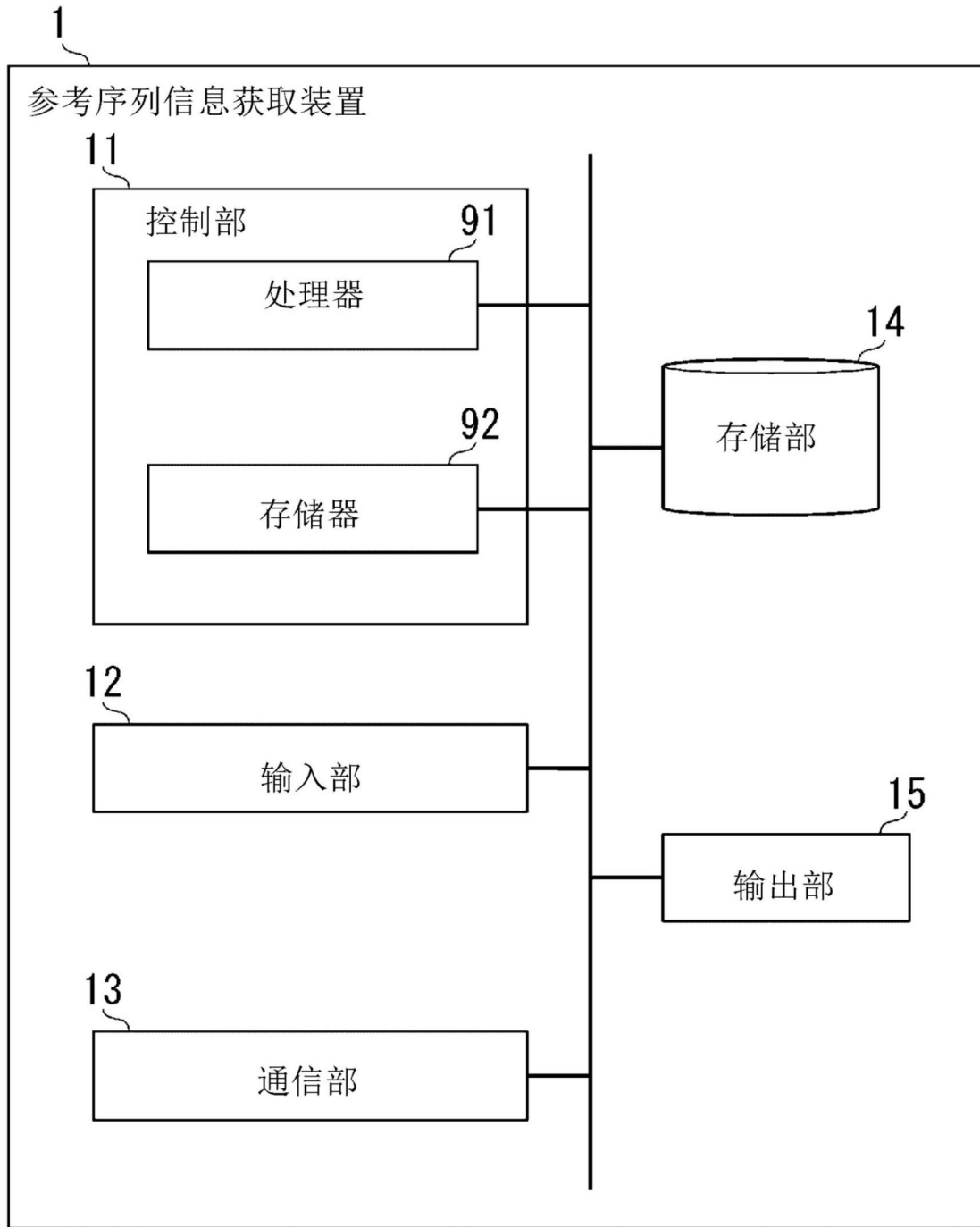


图3

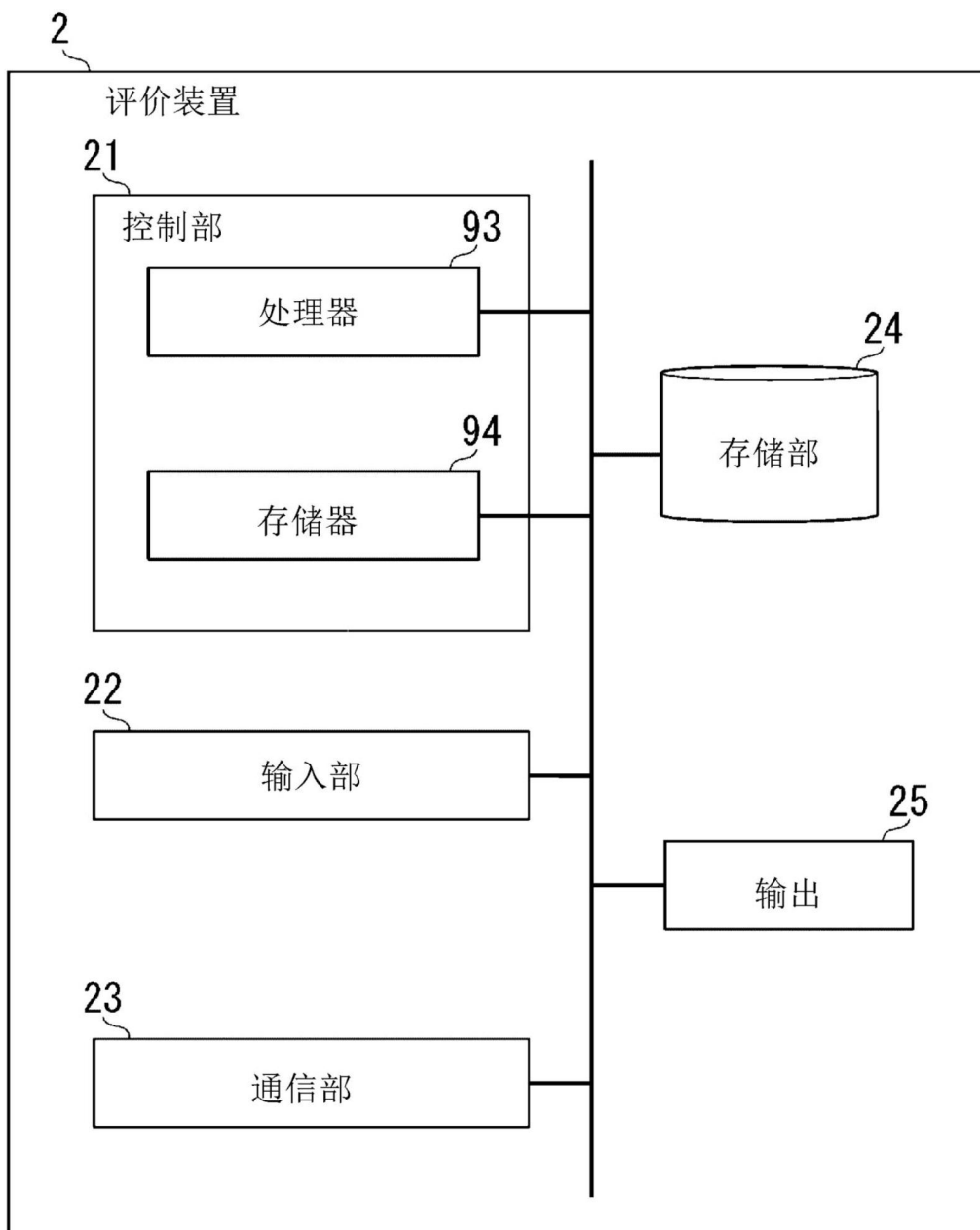


图4

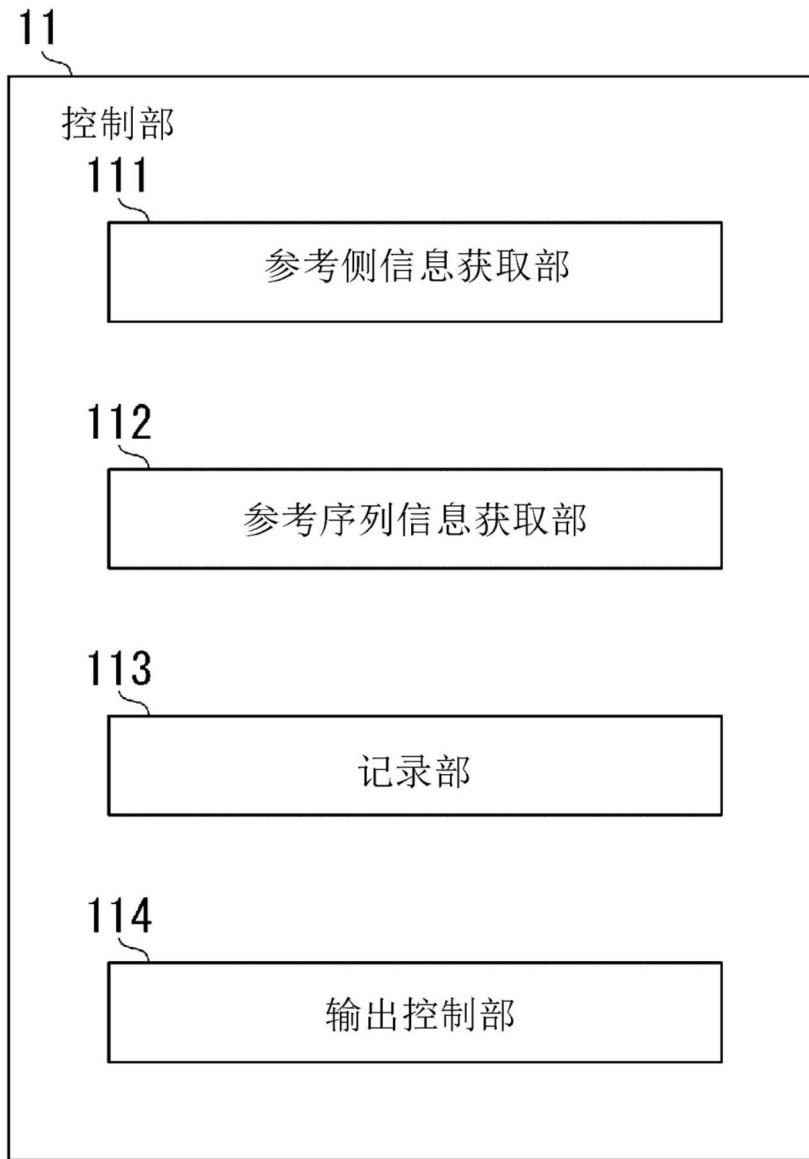


图5

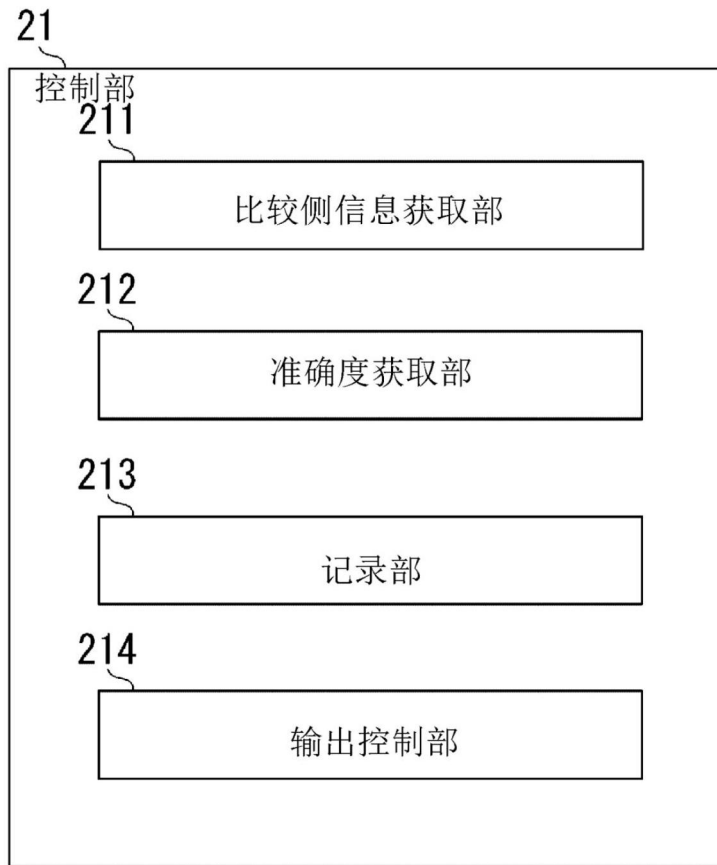


图6

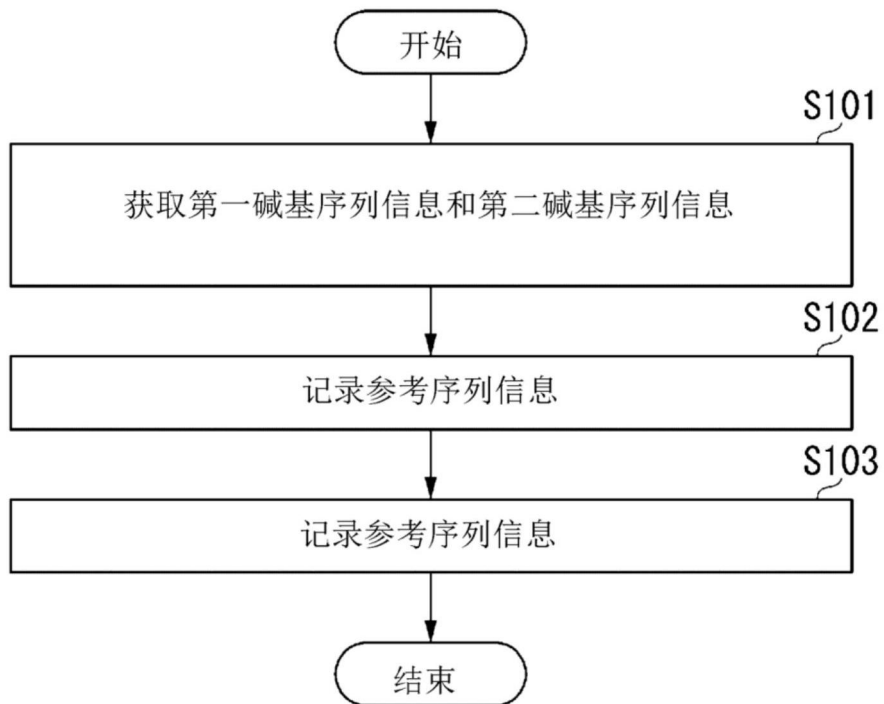


图7

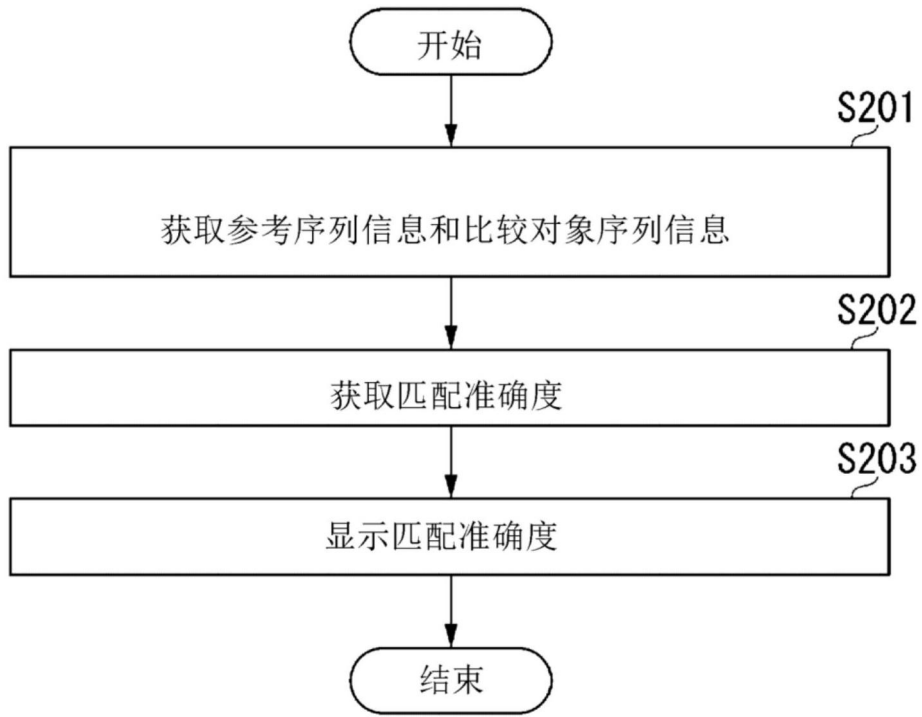


图8

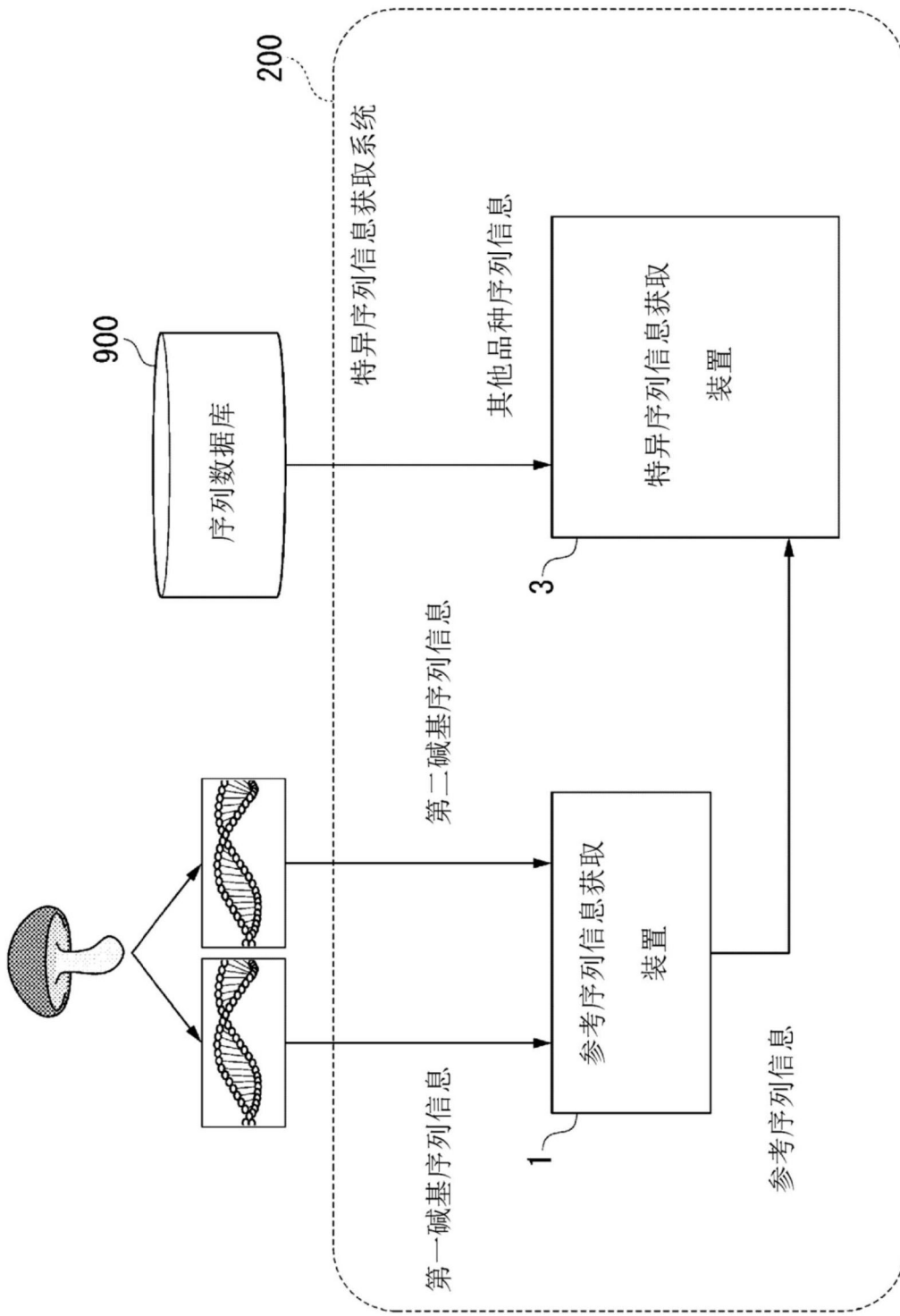


图9

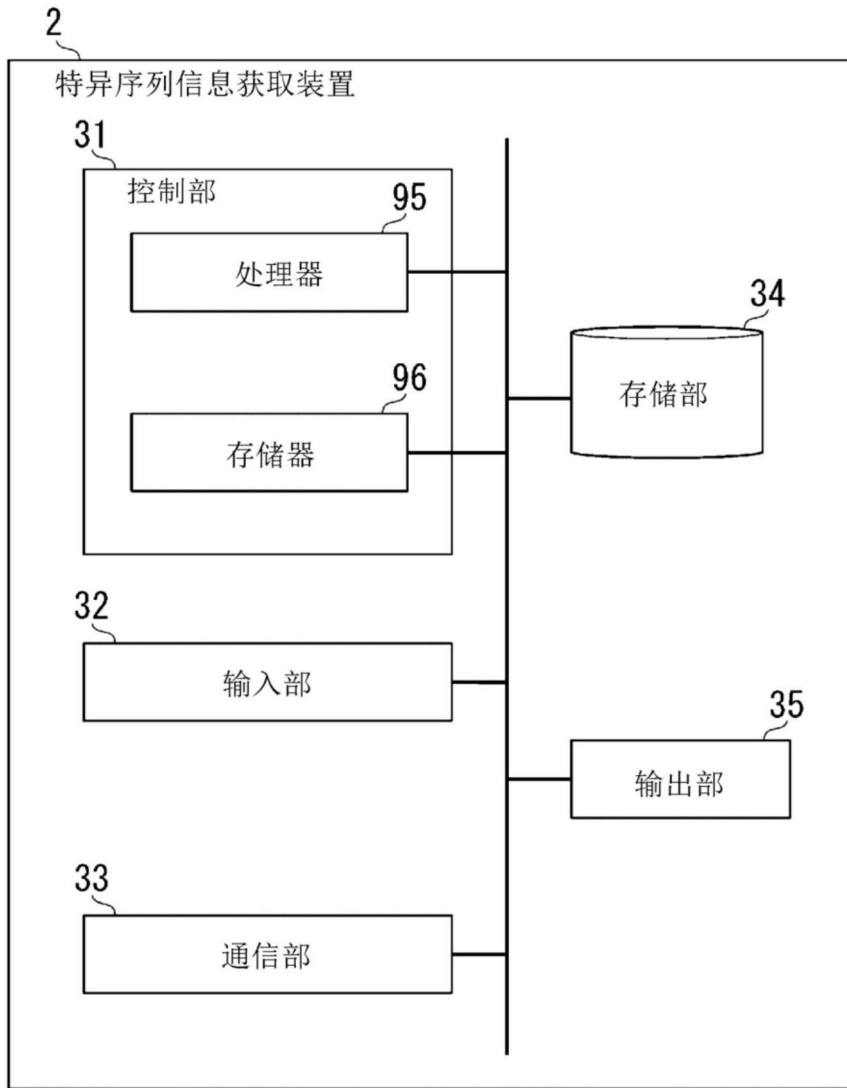


图10

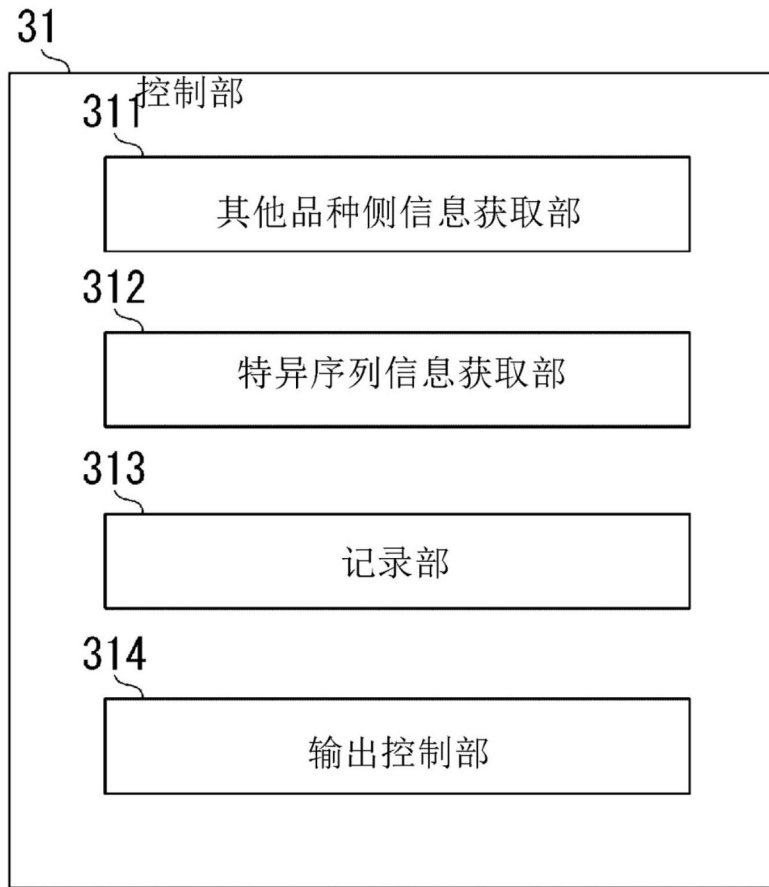


图11

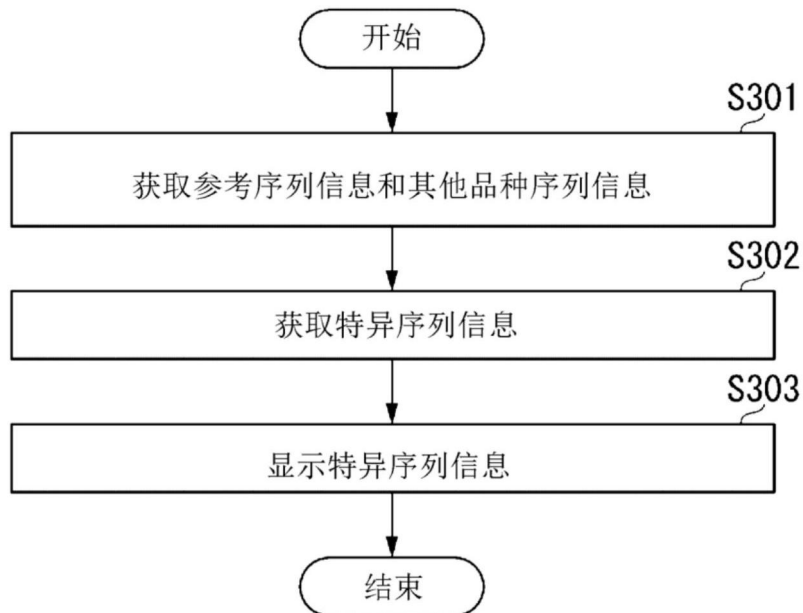


图12

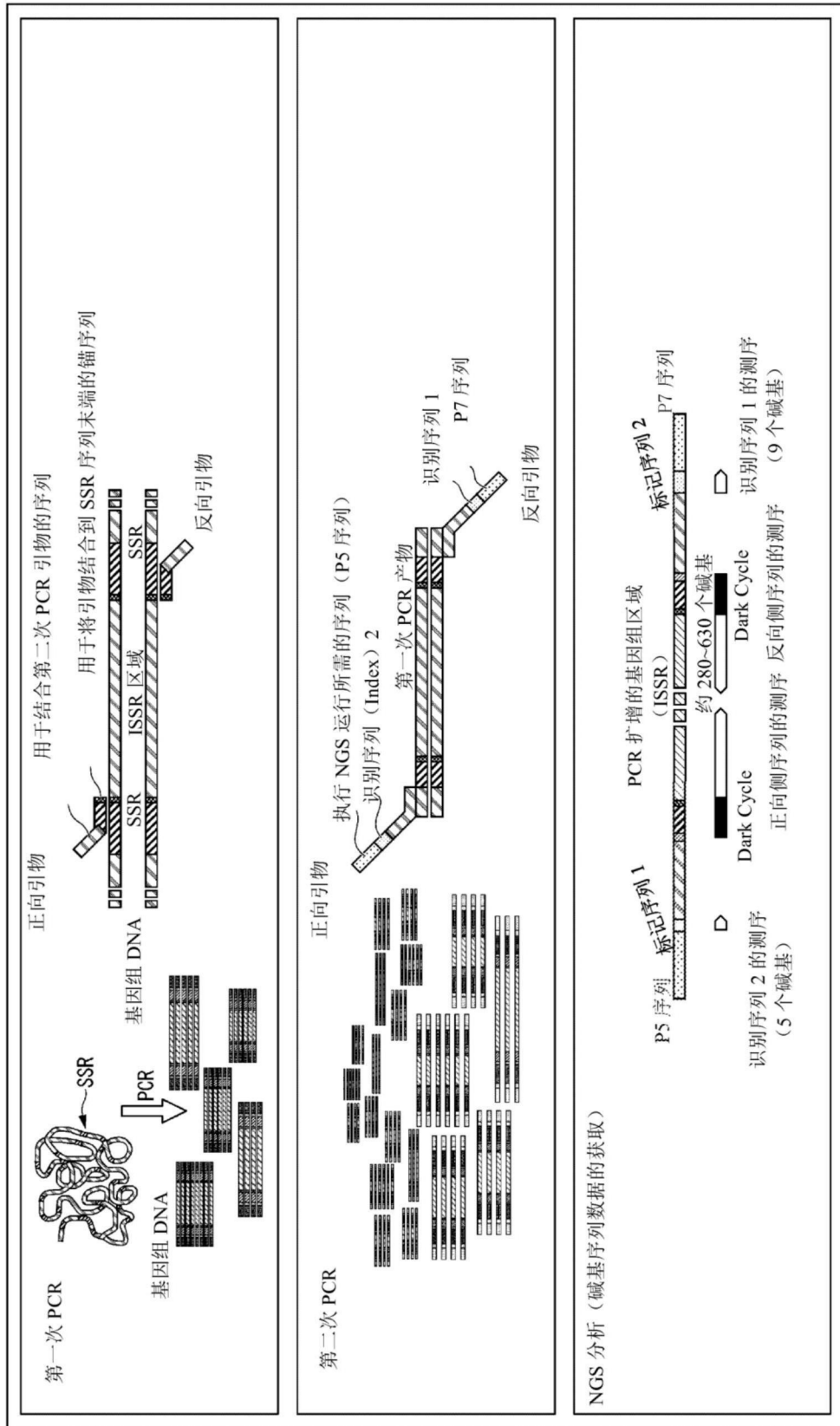


图13

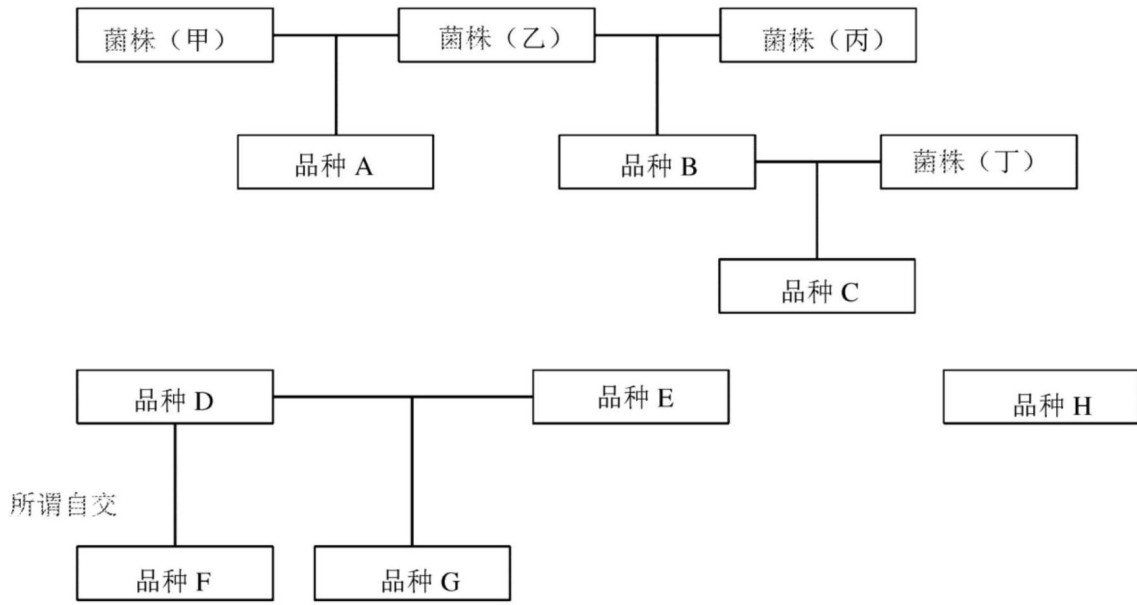


图14

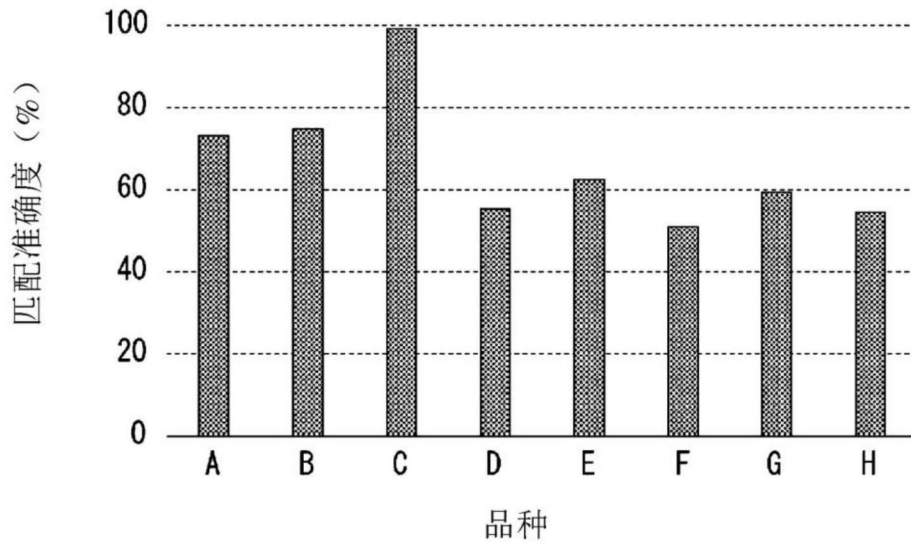


图15

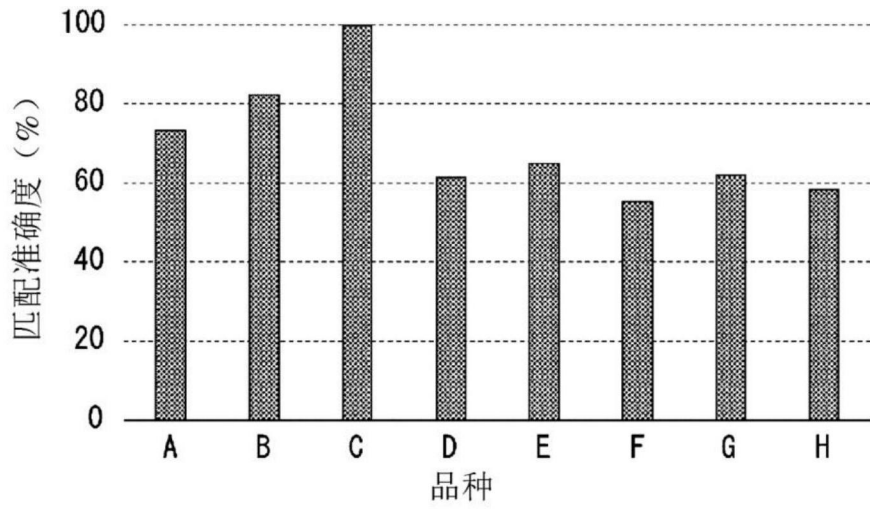


图16

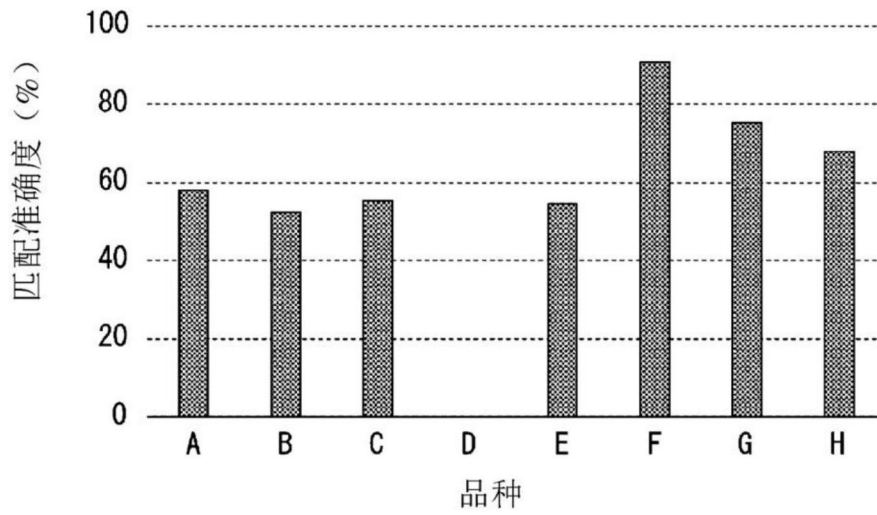


图17

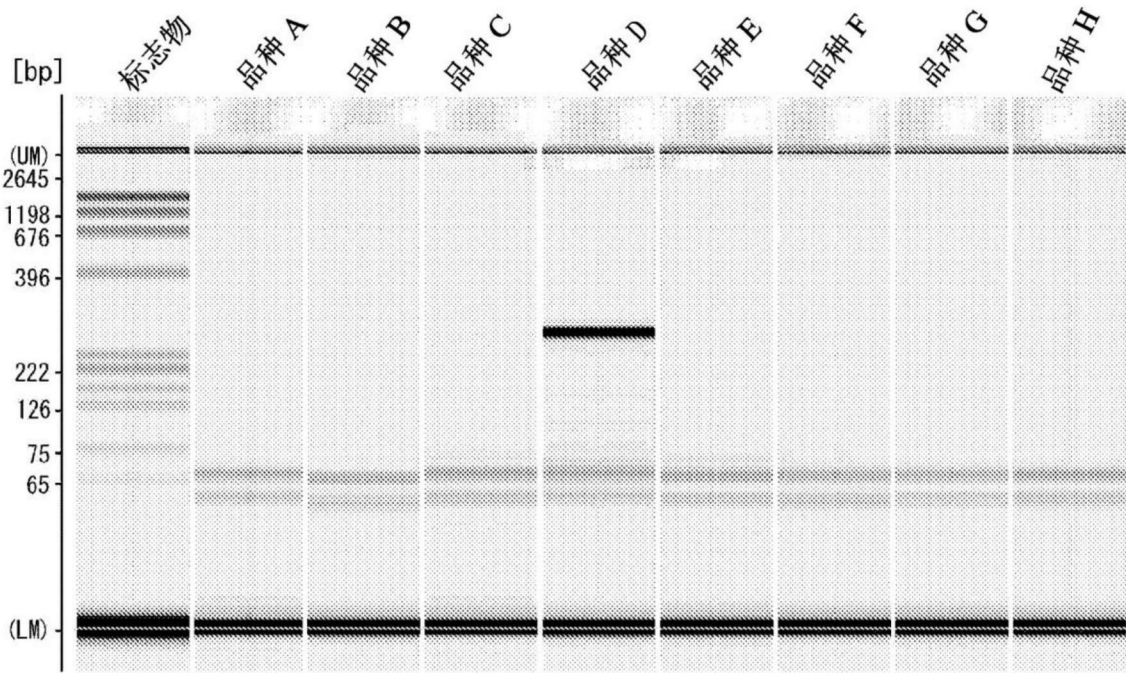


图18