



(12)发明专利申请

(10)申请公布号 CN 111243676 A

(43)申请公布日 2020.06.05

(21)申请号 202010160122.2

(22)申请日 2020.03.10

(71)申请人 南京农业大学

地址 211225 江苏省南京市溧水区白马镇
国家农业科技园南京农业大学基地

(72)发明人 袁军 沈其荣 文涛

(74)专利代理机构 南京天华专利代理有限责任
公司 32218

代理人 莫英妍 徐冬涛

(51)Int.Cl.

G16B 40/00(2019.01)

G16B 30/00(2019.01)

G16B 5/00(2019.01)

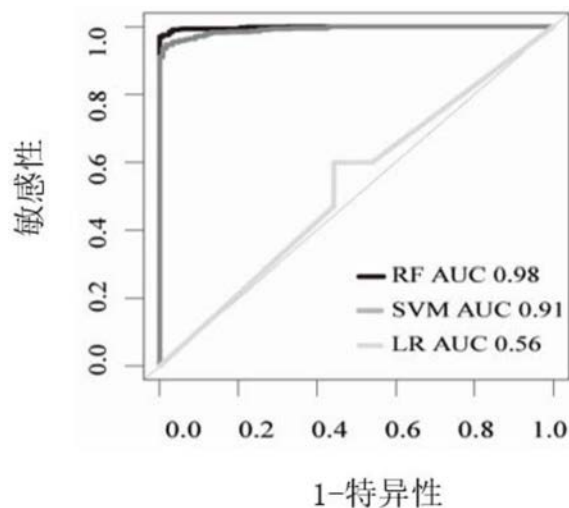
权利要求书2页 说明书8页 附图2页

(54)发明名称

一种基于高通量测序数据的枯萎病发病预测模型及应用

(57)摘要

本发明提出一种基于高通量测序数据的枯萎病发病预测模型及应用,通过整合不同地域、不同作物的枯萎病患病土壤和健康土壤的微生物高通量测序数据,利用机器学习方法分析患病土壤和健康土壤中细菌群落和真菌群落的特征,获取45个细菌和40个真菌的分类操作单元相对丰度特征用于辨别健康土壤和枯萎病发病土壤,并分别建立细菌模型和真菌模型,实现准确判别土壤枯萎病的发病情况,揭示枯萎病患病土壤微生物群落特征。



1. 一种基于高通量测序数据的枯萎病发病预测模型,其特征在于,包括以下步骤:

步骤1:数据收集:使用关键词从数据库中获取枯萎病相关的土壤微生物群落的测序样品登录号,进而获取枯萎病相关的土壤微生物群落的原始测序数据;

步骤2:数据处理:将土壤微生物群落的测序数据分为细菌原始测序数据和真菌原始测序数据,分别进行处理和分析;

步骤3:将土壤中的细菌群落和真菌群落作为生物标志物来区分枯萎病发病土壤和健康土壤,应用机器学习模型随机森林RF对区分枯萎病发病土壤和健康土壤的细菌群落和真菌群落分别建立分类模型:

步骤3-1:构建用于区分枯萎病发病土壤微生物群落和健康土壤微生物群落的细菌模型:

使用随机森林创建分类模型,利用版本号为v.4.6-14的RF软件包中的randomForest函数构建随机森林模型并设置参数ntree=1000,其他参数保持默认,将细菌群落组成的相对丰度按门、纲、目、科、属、种和OTU等级别进行建模,获得最佳预测模型的细菌分类级别;

对于每个细菌群落分类级别,将所有整合得到的细菌测序样本随机分为两部分,一部分占全部样本的三分之二,用于模型构建,另一部分占总样本三分之一,用于细菌模型预测,这个过程重复五次,通过模型预测数据评估最佳模型,从而获得预测准确率最高的用于建模的细菌分类级别;

在准确率最高的分类级别上,将步骤2中的整合的全部细菌原始测序数据组成训练集,使用相同的模型并设置相同的参数继续训练模型,进而优化所生成的预测枯萎病发病的细菌模型;

通过交叉检验提取出来对模型影响最大的若干个OTU。

步骤3-2:构建用于区分枯萎病发病土壤微生物群落和健康土壤微生物群落的真菌模型:

利用R语言中的版本号为v.4.6-14的RF软件包中的randomForest函数构建随机森林模型并设置参数ntree=1000,其他参数保持默认,将真菌群落组成的相对丰度按门、纲、目、科、属、种和OTU等级别进行分类建模,获得最佳预测模型的真菌分类级别;

对于每个真菌群落分类级别,将所有整合得到的真菌测序样本随机分为两部分,一部分占全部样本的三分之二,用于模型构建,另一部分占总样本三分之一,用于真菌模型预测,这个过程重复五次,通过模型预测数据评估最佳模型,从而获得预测准确率最高的用于建模的真菌分类级别;

在准确率最高的分类级别上,将步骤2中的整合的全部真菌原始测序数据组成训练集,使用相同的模型并设置相同的参数继续训练模型,进而优化所生成的预测枯萎病发病的真菌模型。

通过交叉检验提取出来对真菌模型影响最大的若干个OTU。

2. 根据权利要求1所述的基于高通量测序数据的枯萎病发病预测模型,其特征在于,所述步骤1中的关键词包括枯萎病发病和健康土壤的微生物群落、枯萎病土壤微生物结构。

3. 根据权利要求1所述的基于高通量测序数据的枯萎病发病预测模型,其特征在于,所述步骤3中的细菌模型和真菌模型分别包含45个细菌OTU和40个真菌OTU的分类及丰度信息。

4. 根据权利要求1所述的基于高通量测序数据的枯萎病发病预测模型,其特征在於,所述步骤2中对测序数据的处理和分析具体包括:拼接双端原始测序数据,过滤低质量序列,并基于参考数据库聚类操作分类单元OTU,将OTU分配到微生物门类信息。

5. 根据权利要求4所述的基于高通量测序数据的枯萎病发病预测模型,其特征在於,所述参考数据库包括细菌参考数据库Greengene V 13.5和真菌参考数据库Unite。

6. 一种基于权利要求1-5任一基于高通量测序数据的枯萎病发病预测模型在土壤枯萎病发病预测中的应用。

一种基于高通量测序数据的枯萎病发病预测模型及应用

技术领域

[0001] 本发明属于农业植物保护技术领域,尤其是一种基于高通量测序数据的枯萎病发病预测模型及应用。

背景技术

[0002] 微生物对人类活动至关重要。土壤微生物群落与我们的生活密切相关,特别是与我们的农业实践密切相关,深刻影响农产品的质量和数量。镰刀菌枯萎病作为土传病害的典型代表,在世界范围内极大的影响农作物的生产。在田间已分离和试验了数百种对枯萎病菌和镰刀菌有抑制作用的有益微生物。镰刀菌枯萎病虽然是由尖孢镰刀菌引起的真菌病害,病原菌数量增加必然导致发病率上升。但是目前仍然无法以病原菌数量来预测病害的发生。因为不同土壤条件、土壤间微生物的相互作用均能影响发病情况。

[0003] 在过去的十几年中,通过高通量测序(HTS)方法逐渐明确了镰刀菌(Fusarium)发病和健康土壤微生物群落的特征。因此,我们在全世界范围内,通过公开发表的镰刀菌枯萎病相关研究获取枯萎病土壤微生物群落的原始测序数据,并考虑了方法学和技术上的缺陷,使用多种机器学习方法来分析患病或健康土壤中细菌和真菌群落的特征,并基于此构建区分枯萎病发病和健康微生物群落的分类模型。

发明内容

[0004] 本发明所解决的技术问题在于提供一种基于高通量测序数据的枯萎病发病预测模型,通过整合不同地域、不同作物的枯萎病患者土壤和健康土壤的微生物高通量测序数据,利用机器学习方法分析患病土壤和健康土壤中细菌群落和真菌群落的特征,并分别建立细菌模型和真菌模型,实现准确判别土壤枯萎病的发病情况。

[0005] 实现本发明目的的技术解决方案为:

[0006] 一种基于高通量测序数据的枯萎病发病预测模型,包括以下步骤:

[0007] 步骤1:数据收集:使用关键词从数据库中获取枯萎病相关的土壤微生物群落的测序样品登录号,进而获取枯萎病相关的土壤微生物群落的原始测序数据;

[0008] 步骤2:数据处理:将土壤微生物群落的测序数据分为细菌原始测序数据和真菌原始测序数据,分别进行处理和分析;

[0009] 步骤3:将土壤中的细菌群落和真菌群落作为生物标志物来区分枯萎病发病土壤和健康土壤,应用机器学习模型随机森林RF对区分枯萎病发病土壤和健康土壤的细菌群落和真菌群落分别建立分类模型:

[0010] 步骤3-1:构建用于区分枯萎病发病土壤微生物群落和健康土壤微生物群落的细菌模型:

[0011] 使用随机森林创建分类模型,利用RF软件包(版本号:v.4.6-14)的randomForest函数构建随机森林模型并设置参数ntree=1000,其他参数保持默认,将细菌群落组成的相对丰度按门、纲、目、科、属、种和OTU等级别进行建模,获得最佳预测模型的细菌分类级别;

[0012] 对于每个细菌群落分类级别,将所有整合得到的细菌测序样本随机分为两部分,一部分占全部样本的三分之二,用于模型构建,另一部分占总样本三分之一,用于细菌模型预测,这个过程重复五次,通过模型预测数据评估最佳模型,从而获得预测准确率最高的分类级别;

[0013] 在准确率最高的分类级别上,将步骤2中的整合的全部细菌原始测序数据组成训练集,使用相同的模型并设置相同的参数继续训练模型,进而优化所生成的预测枯萎病发病的细菌模型;

[0014] 通过交叉检验提取出来对模型影响最大的若干个OTU。

[0015] 步骤3-2:构建用于区分枯萎病发病土壤微生物群落和健康土壤微生物群落的真菌模型:

[0016] 利用R语言中的RF软件包(版本号:v.4.6-14)的randomForest函数构建随机森林模型并设置参数ntree=1000,其他参数保持默认,将真菌群落组成的相对丰度按门、纲、目、科、属、种和OTU等级别进行分类建模,获得最佳预测模型的真菌分类级别;

[0017] 对于每个真菌群落分类级别,将所有整合得到的真菌测序样本随机分为两部分,一部分占全部样本的三分之二,用于模型构建,另一部分占总样本三分之一,用于真菌模型预测,这个过程重复五次,通过模型预测数据评估最佳模型,从而获得预测准确率最高的用于建模的真菌分类级别;

[0018] 在准确率最高的分类级别上,将步骤2中的整合的全部真菌原始测序数据组成训练集,使用相同的模型并设置相同的参数继续训练模型,进而优化所生成的预测枯萎病发病的真菌模型。

[0019] 通过交叉检验提取出来对真菌模型影响最大的若干个OTU。

[0020] 进一步的,本发明的基于高通量测序数据的枯萎病发病预测模型,所述步骤1中的关键词包括枯萎病发病和健康土壤的微生物群落、枯萎病土壤微生物结构。

[0021] 进一步的,本发明的基于高通量测序数据的枯萎病发病预测模型,所述步骤3中的细菌模型和真菌模型分别包含45个细菌OTU和40个真菌OTU的分类及丰度信息。

[0022] 进一步的,本发明的基于高通量测序数据的枯萎病发病预测模型,所述步骤2中对测序数据的处理和分析具体包括:拼接双端原始测序数据,过滤低质量序列,并基于参考数据库聚类操作分类单元OTU,将OTU分配到微生物门类信息。

[0023] 进一步的,本发明的基于高通量测序数据的枯萎病发病预测模型,所述参考数据库包括细菌参考数据库Greengene V 13.5和真菌参考数据库Unite。

[0024] 一种基于上述任一的基于高通量测序数据的枯萎病发病预测模型在土壤枯萎病预测中的应用。

[0025] 本发明采用以上技术方案与现有技术相比,具有以下技术效果:

[0026] 1、本发明的基于高通量测序数据的枯萎病发病预测模型充分考虑了细菌群落和真菌群落的特征,预测结果可靠。

[0027] 2、本发明的基于高通量测序数据的枯萎病发病预测模型在土壤枯萎病发病预测中的应用操作简便,无需种植植物,只需提取土壤DNA,进行高通量测序即可。

[0028] 3、本发明的基于高通量测序数据的枯萎病发病预测模型适用于所有镰刀菌枯萎病土壤的发病预测。

附图说明

[0029] 图1是随机森林、支持向量机和逻辑回归建立细菌模型的ROC曲线；

[0030] 图2是随机森林、支持向量机和逻辑回归建立真菌模型的ROC曲线。

具体实施方式

[0031] 下面详细描述本发明的实施方式,所述实施方式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方式是示例性的,仅用于解释本发明,而不能解释为对本发明的限制。

[0032] 实施例1

[0033] 一种基于高通量测序数据的枯萎病发病预测模型,包括以下步骤:

[0034] 1、数据收集

[0035] 基于已发表研究,使用关键词枯萎病群落(Fusarium wilt community)或者枯萎病结构(Fusarium wilt structure)或者枯萎病菌群(Fusarium wilt microbiome)从谷歌学术查找真菌枯萎病和微生物群落相关的研究论文,得到枯萎病相关的土壤微生物群落的测序样品登录号。大部分相关研究的原始数据存储在NCBI中,也有一部分存储在DDBJ和欧洲核酸数据库中,均可以从中获取测序数据信息。

[0036] 目前NCBI上存储了大量尚未公开发表文章的数据,往往这些数据的分组及其采样信息被描述的很清楚,可通过测序数据存储登记的项目号中搜索关键词Fusarium wilt community或者Fusarium wilt structure或者Fusarium wilt microbiome,获得枯萎病研究相关的土壤微生物群落的原始测序数据。

[0037] 2、数据处理:将土壤微生物群落的测序数据分为细菌原始测序数据和真菌原始测序数据,分别进行处理和分析。

[0038] (1)在谷歌、NCBI、欧洲核酸数据库,日本核酸数据库上下载每个项目中全部样品的列表文件;

[0039] (2)使用prefetch工具根据list文件下载原始测序数据;

[0040] (3)使用fastq-dump工具解压数据并区分细菌真菌测序结果;

[0041] (4)使用fastqc工具结合shell脚本对测序数据进行质量评估;

[0042] (5)通过Rstudio整合Qiime+usearch+seqkit+R等工具形成两套基于Greengene V 13.5(细菌)和Unite(真菌)的Pipeline进行自动化原始数据处理;

[0043] (6)基于biom格式otu表格合并和qiime标准格式mapping文件合并。

[0044] 3、将土壤中的细菌群落和真菌群落作为生物标志物来区分枯萎病发病土壤和健康土壤,应用机器学习模型随机森林(RF)对区分枯萎病发病土壤和健康土壤的细菌群落和真菌群落分别建立分类模型:

[0045] 3-1、构建用于区分枯萎病发病土壤微生物群落和健康土壤微生物群落的细菌模型:

[0046] 1)使用机器学习算法随机森林创建了一个分类模型,该模型适用于微生物种群数据。在此基础上,利用R语言中RF软件包(版本号:v.4.6-14)的randomForest函数构建随机森林模型并设置参数ntree=1000,其他参数保持默认,将细菌群落组成的相对丰度按门、纲、目、科、属、种和OUT等级别进行建模,以获得最佳预测模型的细菌分类级别。

[0047] 2) 对于每个细菌群落分类级别,将所有整合得到的细菌测序样本随机分为两部分,一部分占全部样本的三分之二,用于模型构建,另一部分占总样本三分之一,用于细菌模型预测。对上述过程重复五次,通过模型预测数据评估最佳模型,从而获得预测准确率最高的用于建模的细菌分类级别。在准确率最高的分类级别上,将步骤2中的整合的全部细菌原始测序数据组成训练集,使用相同的模型并设置相同的参数继续训练模型,进而优化所生成的预测枯萎病发病的细菌模型。通过交叉检验提取出来对模型影响最大的若干个OTU,一共得到45个OTU。如表1,细菌类群在OTU水平上的建模最明显地区别于枯萎病发病土壤和健康土壤。

[0048] 3) 在OTU水平上使用随机森林(重要性=TRUE,接近性=TRUE)函数对来自24个研究(758个样本)的细菌原始测序数据进行训练,生成枯萎病发病土壤和健康土壤的细菌分类模型。如表1,本细菌模型获得了98.94%的高准确度。

[0049] 表1细菌模型分级预测精度

	分类级别	疾病样本的准确率 (%)	健康样本的准确率 (%)	总样本 (%)	
	门	79.5±1.3	98.9±1.2	95.78±1.25	
	纲	86.03±2.3	99.37±2.3	97.23±2.3	
	目	88.53±5.6	99.21±1.9	97.49±3.75	
[0050]	细菌	科	89.34±2.5	99.22±1.6	97.63±2.05
		属	87.7±1.9	98.74±1.4	96.97±1.65
		种	71.31±5.3	98.11±1.6	93.79±3.45
		OTU	95.9±3.1	99.53±1	98.94±2.05

[0051] 3.2构建用于区分枯萎病发病土壤微生物群落和健康土壤微生物群落的真菌模型:

[0052] 1) 利用R语言中RF软件包(版本号:v.4.6-14)的randomForest函数构建随机森林模型并设置参数ntree=1000,其他参数保持默认,将真菌群落的相对丰度分为门、纲、目、科、属、种和OTU等级别进行分类建模,获得最佳预测模型的真菌分类级别。

[0053] 2) 对于每个真菌群落分类级别,将所有整合得到的真菌测序样本随机分为两部分,一部分占全部样本的三分之二,用于模型构建,另一部分占总样本三分之一,用于真菌模型预测。对上述处理进行五次重复,通过模型预测数据评估最佳模型,从而获得预测准确率最高的用于建模的真菌分类级别。在准确率最高的分类级别上,将步骤2中的整合的全部真菌原始测序数据组成训练集,使用相同的模型并设置相同的参数继续训练模型,进而优化所生成的预测枯萎病发病的真菌模型。通过交叉检验提取出来对真菌模型影响最大的若干个OUT,一共得到40个OTU。如表2,真菌类群在OTU水平上的建模最明显地区别于枯萎病发病土壤和健康土壤。与细菌类群一样,在OTU水平上建模最能清楚地区分枯萎病发病土壤和健康土壤,因此在该水平上进行进一步分析。

[0054] 3) 在OTU水平上利用随机森林(重要性=TRUE,接近性=TRUE)函数对来自22个研究(279个样本)的真菌原始测序数据进行训练,生成枯萎病发病土壤和健康土壤的真菌分类模型,其准确率高达96.42%,见表2。

[0055] 表2真菌模型分级预测精度

	分类级别	疾病样本的准确率 (%)	健康样本的准确率 (%)	总样本 (%)
	门	79.51±4.6	84.08±5.3	82.08±4.95
	纲	87.71±2.3	94.27±4.8	91.4±3.55
	目	91.17±3.1	93.63±1.9	92.11±2.5
[0056]	真菌	科	94.91±2.3	95.7±2.2
		属	95.55±0.5	96.42±0.75
		种	97.46±1.6	96.42±1.1
		OTU	96.82±0.3	96.42±0.7

[0057] 土壤的细菌群落和真菌群落都可以作为生物标志物来区分枯萎病发病土壤和健康土壤,为了确定这一结果我们使用随机森林 (RF)、支持向量机 (SVM) 和逻辑回归 (LR) 建立了三个机器学习模型。

[0058] 如图1、图2所示,图1展示了通过三种分类方法随机森林、支持向量机、逻辑回归进行枯萎病发病和健康土壤细菌群落的分类和预测,发现随机森林模型对枯萎病发病和健康土壤的细菌群落分类和预测效果最好;图2展示了通过三种分类方法随机森林、支持向量机、逻辑回归进行枯萎病发病和健康土壤真菌群落的分类和预测,发现随机森林模型对枯萎病发病和健康土壤的真菌群落分类和预测效果最好。由三种模型的ROC曲线可知,RF模型预测样本分类的准确率最高。接下来,利用细菌数据和真菌数据建立了在OTU水平上训练的模型,它们的分类准确率最高,分别为95.00%和92.00%。

[0059] 三种机器学习方法建立的模型在微生物 (C:细菌性,D:真菌性) 群落OTU水平上的ROC曲线,其中RF模型预测样本分类准确率最高 (AUC=0.98)。

[0060] 实施例2

[0061] 为了验证细菌模型准确性,接下来分析来自NCBI SRA数据库的13个独立的细菌数据集,包括26个枯萎病发病土壤样本和321个健康土壤样本。所有采集的样本,细菌模型的平均准确率为94.46%,患病样本的准确率为91.75%,健康样本的准确率为96.45% (见表3)。

[0062] 同时,使用了来自NCBI SRA数据库的6个独立的真菌数据集,包括21个枯萎病发病土壤样本和144个健康土样本,进行真菌模型验证。真菌模型对所有采集的样本的平均准确率为93.05%,对患病样本的准确率为91.67%,对健康样本的准确率为95.25% (见表4)。

[0063] 表3 NCBI数据集的细菌模型预测精度

细菌模型预测								
名称	来源	健康样 本数目	健康样本 正确率/%	患病样 本数目	患病样本 正确率/%	所有样 本数目	总样本 正确率/%	
	A1_13	数据库	8	100.00	-	-	8	100.00
	No_3	数据库	7	100.00	2	100.00	9	100.00
	NCBI_32	数据库	12	100.00	-	-	12	100.00
	NCBI_31	数据库	33	100.00	-	-	33	100.00
	NCBI_30	数据库	6	100.00	-	-	6	100.00
[0064]	NCBI_26	数据库	9	100.00	-	-	9	100.00
	NCBI_19	数据库	18	61.00	-	-	18	61.00
	NCBI_11	数据库	15	100.00	-	-	15	100.00
	NCBI_21	数据库	155	100.00	-	-	155	100.00
	NCBI_24	数据库	52	100.00	-	-	52	100.00
	No_2	数据库	6	100.00	6	100.00	12	100.00
	No_217	数据库	-	-	12	67.00	12	67.00
	No_224	数据库	-	-	6	100.00	6	100.00
	总计	-	321	96.45	26	91.75	347	94.46

[0065] 表4 NCBI数据集的真菌模型预测精度

真菌模型预测								
名称	来源	健康样 本数目	健康样本 正确率/%	患病样 本数目	患病样本 正确率/%	所有样 本数目	总样本 正确率/%	
	NCBI_14	数据库	30	100.00	-	-	30	100.00
[0066]	NCBI_18	数据库	24	100.00	-	-	24	100.00
	No_71	数据库	27	81.00	3	100.00	30	83.30
	NCBI_12	数据库	63	100.00	-	-	63	100.00
	No_217	数据库	-	-	12	75.00	12	75.00
	No_224	数据库	-	-	6	100.00	6	100.00
[0067]	总计	-	144	95.25	21	91.67	165	93.05

[0068] 实施例3

[0069] 土壤样品采集：于2019年4月下旬分别采集了海南、广东、北京、江苏的香蕉、黄瓜、西瓜、百合的4种不同作物的田间土壤样品。对于病虫害的土壤收集，选择了连作且遭受枯

萎病至少5年的地块,同时选择了没有发生枯萎病的新植地块进行健康的土壤收集。为了避免地理因素引起的差异,健康地选择在病害地附近。每种土样取0-20cm土层20个随机土芯(直径5cm×20cm),每个小区s型取样,合为一个土样。新收集的土壤经过2毫米的筛网筛选,以清除植物碎片和可能的小动物,然后在提取DNA之前储存在-80℃。

[0070] DNA提取、基因扩增和测序:(1)使用Power Lyzer PowerSoil DNA分离试剂盒(Qiagen,德国)按照制造商的说明书进行操作。(2)PCR扩增利用引物515F和806R合成针对细菌16S rRNA基因V4区域的PCR产物,获得292bp的扩增子,针对真菌群落,利用引物ITS3-F和ITS4-R对ITS2区进行PCR扩增;(3)采用PCR扩增仪进行PCR扩增,扩增周期为:95℃5min,94℃30s,52℃30s,72℃30s,72℃10min;(4)测序由专业公司进行精准测序。

[0071] 运用本发明的模型预测:细菌模型对所有样本的平均准确率为90.00%,患病样本的准确率为87.50%,健康样本的准确率为92.50%(见表5)。

[0072] 真菌模型对所有样本的平均准确率为80.00%,对患病土壤样本的准确率为97.50%,对健康土壤样本的准确率为62.50%(见表6)。研究表明,细菌模型和真菌模型都可以作为生物标志物来区分枯萎病和健康土壤。

[0073] 表5收集的样本的细菌模型预测精度

细菌模型预测							
名称	来源	健康样本数目	健康样本正确率/%	患病样本数目	患病样本正确率/%	所有样本数目	总样本正确率/%
[0074]	香蕉	5	90.00	5	60.00	10	75.00
	黄瓜	5	80.00	5	100.00	10	90.00
	西瓜	5	100.00	5	90.00	10	95.00
	百合	5	100.00	5	100.00	10	100.00
	总计	20	92.5	20	87.5	40	90

表6 收集的样本的真菌模型预测精度

真菌模型预测							
名称	来源	健康样本数目	健康样本正确率/%	患病样本数目	患病样本正确率/%	所有样本数目	总样本正确率/%
[0075]	香蕉	5	60.00	5	100.00	10	80.00
	黄瓜	5	90.00	5	100.00	10	95.00
	西瓜	5	0.00	5	100.00	10	50.00
	百合	5	100.00	5	90.00	10	95.00
	总计	20	62.5	20	97.5	40	80

[0076] 本发明通过将多个独立来源的镰刀菌枯萎病土壤样品的细菌数据和真菌数据结

合起来,用于检测病害土壤的微生物群落和建立用于预测土壤健康程度的微生物群落模型,通过机器学习RF模型方法选择的45个细菌和40个真菌OTUs及其分类、丰度信息作为预测模型。

[0077] 以上所述仅是本发明的部分实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进,这些改进应视为本发明的保护范围。

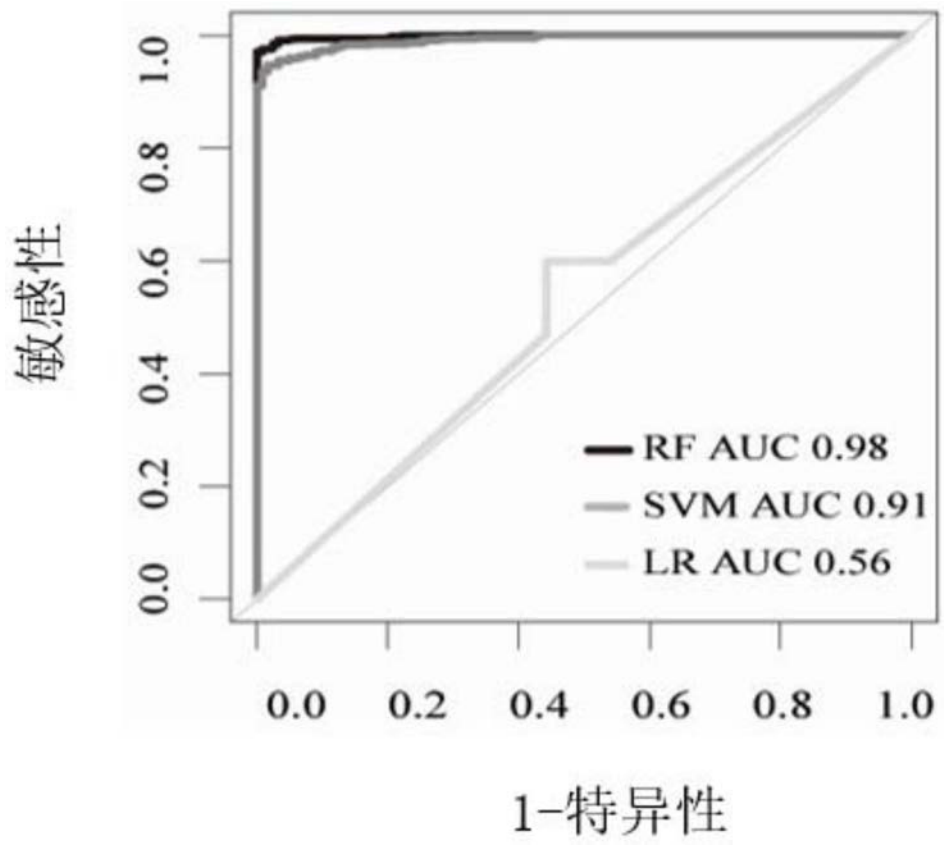


图1

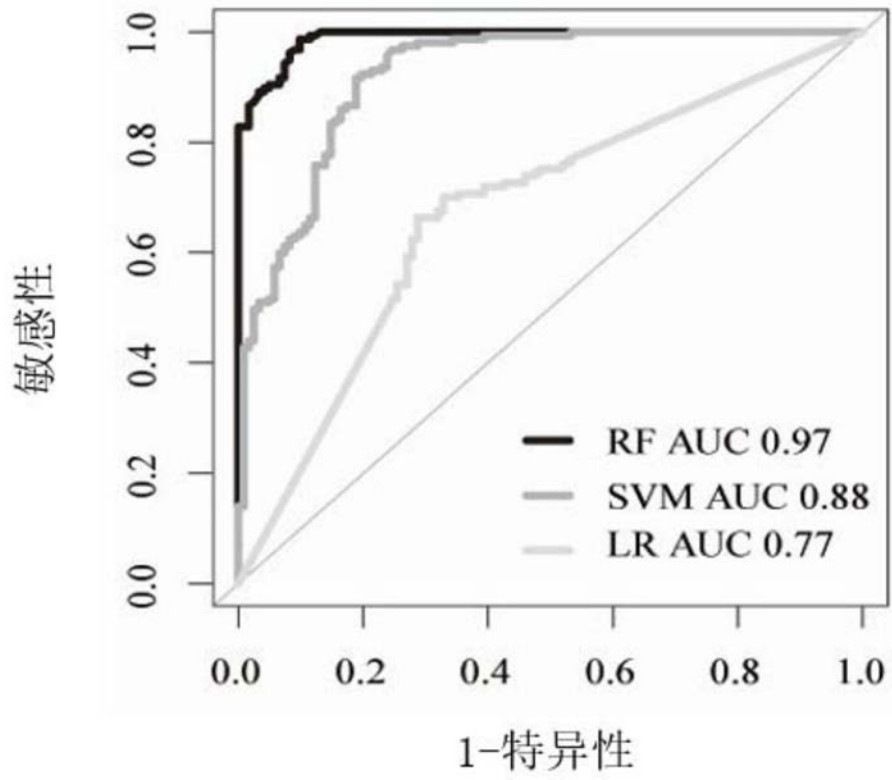


图2