



(12)发明专利申请

(10)申请公布号 CN 108398491 A

(43)申请公布日 2018.08.14

(21)申请号 201710064160.6

(22)申请日 2017.02.04

(71)申请人 北京蓝标一成科技有限公司

地址 100176 北京市大兴区北京经济技术
开发区科创十四街20号院16号楼4单
元一层174室

(72)发明人 赵田 邹婷婷

(74)专利代理机构 北京易正达专利代理有限公
司 11518

代理人 路远

(51)Int.Cl.

G01N 30/02(2006.01)

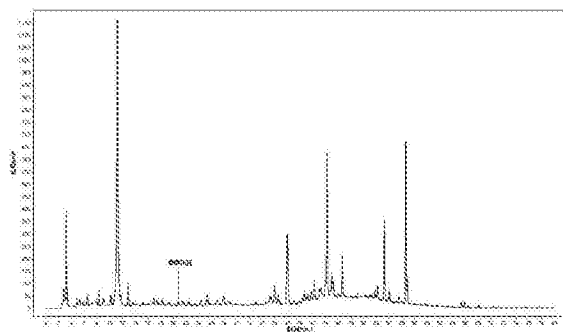
权利要求书2页 说明书10页 附图3页

(54)发明名称

一种美花石斛的质量检测方法

(57)摘要

本发明公开美花石斛的质量检测方法,包括:1)以ITS-26SE和ITS-17SE为引物,进行测序,以鉴定待测石斛药材的品种;2)对样本容量为n的样本进行色谱检测,获取以化学小分子成分夏佛托昔和/或柚皮素作为参照成分的检测数据;3)对样本分别进行指纹图谱检测,获取美花石斛的全化学成分的指纹图谱峰面积值;4)以色谱数据中夏佛托昔和/或柚皮素的成分含量值作为响应变量,将指纹图谱中的其它成分的峰面积值作为自变量建立分析模型通过Lasso方法筛选变量建立化学小分子成分的相关特征指纹图谱模型。通过一代测序及特征指纹图谱,精确地鉴别和控制美花石斛药材的质量。



1. 一种美花石斛的质量检测方法,其特征在于,包括:

1) 以 ITS-26SE:5' GAATCCCCGGTTCGCTCGCCGTTAC 3' 和

ITS-17SE:5' ACGAATTCATGGTCCGGTGAAGTGTTTCG 3' 为引物,进行PCR扩增测序,以鉴定待测石斛药材的品种为美花石斛;

2) 对样本容量为n的样本进行色谱检测,获取以化学小分子成分夏佛托昔和/或柚皮素作为参照成分的检测数据;

3) 对样本分别进行指纹图谱检测,获取美花石斛的全化学成分的指纹图谱峰面积值;

4) 以色谱数据中夏佛托昔和/或柚皮素的成分含量值作为响应变量,将指纹图谱中的其它成分的峰面积值作为自变量建立分析模型,通过Lasso (The Least Absolute Shrinkage and Selection Operator) 方法筛选变量建立化学小分子成分的相关特征指纹图谱模型,其基本模型为:

$$y = X^T \beta + \varepsilon$$

其中,y为响应变量, $y = (y_1, y_2, \dots, y_n)^T$; X为矩阵, $X = (x_1, x_2, \dots, x_n)^T$; $E(\varepsilon) = 0$; $\text{Var}(\varepsilon) = \sigma^2 I_n$; ε 为模型的随机误差项; σ 是随机误差项的标准差;n为样本量; I_n 是一个n×n的单位阵。

2. 根据权利要求1所述的方法,其特征在于,所述Lasso方法是通过式I计算实现的:

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^{p^*} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p^*} |\beta_j|$$

式 I

在式I中,n为样本量;p*为变量数;p为样本的维数; $y = (y_1, y_2, \dots, y_n)^T \in R^n$ 为响应变量; $x = (x_1, x_2, \dots, x_n)^T$ 为n×p的设计矩阵,包含对响应变量有影响的所有候选自变量; λ 为调整参数; $\sum_{j=1}^{p^*} |\beta_j|$ 为惩罚函数; β_0 的含义为公式的截距项,也就是当所有自变量x为0时响应变量y的值; β_j 的含义是自变量 x_j 的系数,即自变量 x_j 对响应变量y的影响程度。

3. 根据权利要求2所述的方法,其特征在于,所述 λ 的选择方法为K折交叉验证法:

K-fold CV:

$$CV_K = \frac{1}{n} \sum_{k=1}^K \left\| y_{(-k)} - \hat{y}_{(-k)} \right\|^2$$

其中,K为5或10。

4. 根据权利要求3所述的方法,其特征在于,所述分析模型为取CV值最小的子模型。

5. 根据权利要求2所述的方法,其特征在于,所述 λ 的选择遵循GCV准则,所述GCV准则定

义为: $GCV = \frac{1}{n} \frac{SSE_k}{(1 - df/n)^2}$

其中,SSE_k是含有k个变量的CV子模型的残差平方和,df = trace {P(λ)}。

6. 根据权利要求5所述的方法,其特征在于,所述分析模型为取GCV值最小的子模型。

7. 根据权利要求2所述的方法,当色谱数据呈现超高维情形时,首先采用以下SIS (Sure

Independence Screening) 方法筛选变量,再利用Lasso方法处理;

SIS: $M_\gamma = \{1 \leq i \leq p: |\omega_i| \text{ 是前 } |\gamma n| \text{ 个比较大的}\}$

其中, $M_* = \{1 \leq i \leq p: \beta_i \neq 0\}$ 表示真模型中非零系数的下标集; $s = |M_*|$ 表示非零系数的个数; $\omega = (\omega_1, \omega_2, \dots, \omega_p)^T = X^T y$; 对于任意给定的 $\gamma \in (0, 1)$, ω 的 p 个元素按绝对值从大到小排列并且定义; 此时 $|\gamma n| < n$, 选取 M_γ 中下标对应的自变量, 使超高维降到 d ($d \leq n$) 维; 其中, $d = n$ 或者 $d = \lceil n / \log n \rceil$ 。

一种美花石斛的质量检测方法

技术领域

[0001] 本发明属于中药成份检测分析领域,具体涉及一种美花石斛的质量检测方法。

背景技术

[0002] 美花石斛(*Dendrobium loddigesii* Rolfe.):又名环草石斛、蟠龙石斛。茎柔弱,常下垂,细圆柱形,长10cm~30cm,直径0.2cm~0.4cm,有时分枝,具多节,节间长1.5cm~2cm,干后金黄色;叶基部具鞘,叶鞘膜质,干后鞘口常张开。总状花序侧生于具叶的老茎上部,具1朵~2朵花;花序柄长0.2cm~0.3cm,基部具1枚~2枚短的、杯状膜质鞘;花白色或紫红色;花苞片膜质,卵形,长0.2cm,先端钝;花梗和子房淡绿色,长2cm~3cm;中萼片卵状长圆形,长1.7cm~2cm,宽0.6cm~0.7cm,先端锐尖,具5条脉;侧萼片披针形,长1.7cm~2cm,宽0.6cm~0.7cm,先端急尖,具5条脉,基部歪斜;萼囊近球形,长约0.5cm;花瓣椭圆形,与中萼片等长,宽0.8cm~0.9cm,先端稍钝,全缘,具3条~5条脉;唇瓣近圆形,直径1.7cm~2cm,上面中央金黄色,周边淡紫红色,稍凹,边缘具短流苏,两面密布短柔毛;蕊柱白色,正面两侧具红色条纹,长约0.4cm;药帽白色,近圆锥形,密布细乳突状毛,前端边缘具不整齐的齿。美花石斛的茎可入药,味甘,微寒。益胃生津,滋阴清热。用于阴伤津亏,口干烦渴,食少干呕,病后虚热,目睛不明。

[0003] 石斛一直被人们视为珍贵的中草药,具有十分重要的滋补功效。在临床上,石斛被用于治疗多种疾病,具有增强免疫力、抗氧化、降血糖和抑制癌症等药理功效。由于人为长期无节制采挖及不合理利用石斛,其野生资源日趋减少,市场上出现了一些以假乱真、以次充好的现象。此外,由于石斛品种较多,其品种间的杂交使得其近缘的种存在性状交叉现象,分类区别比较困难。因此,有必要建立石斛的特征指纹图谱对石斛的药材质量进行评价。

[0004] 色谱指纹图谱是一种综合的、量化的鉴别手段,作为一种全局分析的全景模式,反映的是样品的整体情况。但在色谱指纹图谱分析过程中,很多数据都是高维的,即数据包含很多属性或特征,比如有关美花石斛色谱指纹图谱,就能更好地对美花石斛进行描述,但在实际应用中,对高维数据直接进行操作将会面临“维数灾难”的问题,“维数灾难”会导致建模过程所需要的样本数随着维数升高而呈指数级增长。面对高维数据,常规的最小二乘法不再适用,为了提高模型的可解释性和预测的准确度,变量选择变得很重要。如何高效地从众多的变量中筛选出对因变量有重要作用的若干个变量,是在对指纹图谱进行分析时亟需解决的问题。

[0005] 目前国家药典采用定量测定中药材中某一活性成分或有效成分即小分子成分的含量的高低来评价其质量。但研究证明,中药的疗效是来自其多种“活性成分”之间的协同作用,甚至是被普遍公认的有效“活性成分”与“非活性成分”之间的协同作用或“生克作用”才能达到中药的疗效,而不是某一活性成分单独作用的结果。在中医理论指导下的中药,任何一种活性成分均不能全面反映中医用药所体现的整体疗效。

发明内容

[0006] 本发明提供了一种美花石斛的质量检测方法,通过一代测序鉴定石斛药材的品种,利用Lasso方法筛选变量建立美花石斛中的化学小分子成分的相关特征指纹图谱模型,通过一代测序和相关特征指纹图谱模型准确评价美花石斛的药材的质量。

[0007] 本发明的目的是通过以下技术方案实现的:

[0008] 一种美花石斛的质量检测方法,包括:

[0009] 1) 以ITS-26SE:5' GAATTCCTCCGGTTCGCTCGCCGTTAC 3' ;

[0010] ITS-17SE:5' ACGAATTCATGGTCCGGTGAAGTGTTTCG 3' 为引物,进行PCR扩增测序,以鉴定待测石斛药材的品种为美花石斛样本;

[0011] 2) 对样本容量为n的美花石斛样本进行色谱检测,获取以化学小分子成分夏佛托昔和/或柚皮素作为参照成分的检测数据;

[0012] 3) 对样本分别进行指纹图谱检测,获取美花石斛的全化学成分指纹图谱峰面积值;

[0013] 4) 以色谱数据中夏佛托昔和/或柚皮素的成分含量值作为响应变量,将指纹图谱中的其它成分的峰面积值作为自变量建立分析模型,通过Lasso (The Least Absolute Shrinkage and Selection Operator) 方法筛选变量建立化学小分子成分的相关特征指纹图谱模型,其基本模型为:

$$[0014] \quad y = X^T \beta + \varepsilon$$

[0015] 其中,y为响应变量, $y = (y_1, y_2, \dots, y_n)^T$;X为矩阵, $X = (x_1, x_2, \dots, x_n)^T$; $E(\varepsilon) = 0$; $\text{Var}(\varepsilon) = \sigma^2 I_n$; ε 为模型的随机误差项; σ 是随机误差项的标准差;n为样本量; I_n 是一个n×n的单位阵。

[0016] 假定随机项服从古典假定,即:

[0017] (1) 随机项具有零均值, $E(\varepsilon_i | x_i) = 0$;

[0018] (2) 随机项具有同方差, $\text{Var}(\varepsilon_i | x_i) = \sigma^2$;

[0019] (3) 随机项无序列相关性, $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$;

[0020] (4) ε 服从正态分布, $\varepsilon_i \sim N(0, \sigma^2)$ 。

[0021] 随机项的方差矩阵是一个对角线为 σ^2 ,其他地方为0的方阵,如下所示:

$$[0022] \quad \text{Var}(\varepsilon) = \sigma^2 I_n = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}_{n \times n}$$

[0023] 其中, I_n 是一个n×n的单位阵,n为数据的样本量, $I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$ 。

[0024] 进一步地,所述Lasso方法是通过式I计算实现的:

$$[0025] \quad \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^{p^*} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p^*} |\beta_j|$$

式 I

[0026] 在式I中, n 为样本量; p^* 为变量数; p 为样本的维数; $y = (y_1, y_2, \dots, y_n)^T \in R^n$ 为响应变量; $x = (x_1, x_2, \dots, x_n)^T$ 为 $n \times p$ 的设计矩阵, 包含对响应变量有影响的所有候选自变量; λ 为调整参数; $\sum_{j=1}^{p^*} |\beta_j|$ 为惩罚函数; β_0 的含义为公式的截距项, 也就是当所有自变量 x 为0时响应变量 y 的值; β_j 的含义是自变量 x_j 的系数, 即自变量 x_j 对响应变量 y 的影响程度。

[0027] 进一步地, 所述 λ 的选择方法为K折交叉验证法:

$$[0028] \quad K\text{-fold CV: } CV_K = \frac{1}{n} \sum_{k=1}^K \|y_{(k)} - \hat{y}_{(-k)}\|^2$$

[0029] 其中, K 为5或10。

[0030] 进一步地, 所述分析模型为取CV值最小的子模型。

[0031] 进一步地, 所述 λ 的选择遵循GCV准则, 所述GCV准则定义为:

$$[0032] \quad GCV = \frac{1}{n} \frac{SSE_k}{(1 - df/n)^2}$$

[0033] 其中, SSE_k 是含有 k 个变量的CV子模型的残差平方和, $df = \text{trace}\{P(\lambda)\}$; trace 表示矩阵的迹。在线性代数中, 一个 $n \times n$ 的矩阵 A 的主对角线(从左上方至右下方的对角线)上各个元素的总和被称为矩阵 A 的迹(或迹数), 一般记作 $\text{tr}(A)$ 。也就是说, df 等于矩阵 $P(\lambda)$ 中主对角线上所有元素的和。

[0034] 进一步地, 所述分析模型为取GCV值最小的子模型。

[0035] 进一步地, 当色谱数据呈现超高维情形时, 首先采用以下SIS(Sure Independence Screening)方法筛选变量, 再利用Lasso方法处理:

[0036] SIS: $M_\gamma = \{1 \leq i \leq p: |\omega_i| \text{ 是前 } |\gamma n| \text{ 个比较大的}\}$

[0037] 其中, $M_* = \{1 \leq i \leq p: \beta_i \neq 0\}$ 表示真模型中非零系数的下标集; $s = |M_*|$ 表示非零系数的个数; $\omega = (\omega_1, \omega_2, \dots, \omega_p)^T = X^T y$; 对于任意给定的 $\gamma \in (0, 1)$, ω 的 p 个元素按绝对值从大到小排列并且定义; 此时 $|\gamma n| < n$, 选取 M_γ 中下标对应的自变量, 是超高维降到 d ($d \leq n$) 维; 其中, $d = n$ 或者 $d = \lceil n / \log n \rceil$ 。

[0038] 本发明还提供了上述方法在美花石斛质量控制中的应用。

[0039] 与现有技术相比, 本发明至少具有以下优点:

[0040] (a) 本发明中先以ITS-26SE和ITS-17SE为引物测定石斛药材的特征序列, 以确定该石斛药材为美花石斛品种; 其后以夏佛托昔和/或柚皮素的色谱数据作为自变量, 将上述检测数据中的其它成分的色谱数据作为因变量建立小分子成分与指纹图谱线性回归模型, 使石斛药材的质量评价更为精确;

[0041] (b) 本发明采用Lasso方法对美花石斛指纹图谱进行变量选择, 有效的解决了“维数灾难”的问题;

[0042] (c) 本发明对原有指纹图谱进行降维,建立美花石斛化学小分子成分的相关特征指纹图谱,对单一化学小分子成分的含量解释针对性和适用性更强;

[0043] (d) 本发明通过美花石斛化学小分子成分的相关特征指纹图谱,实现化学小分子成分含量的关联性分析,能有效鉴别和控制美花石斛药材的质量;

[0044] (e) 当色谱数据呈现超高维情形时,首先采用SIS方法进行降维,再利用Lasso方法处理。

附图说明

[0045] 图1为美花石斛全成分指纹图谱色谱图;

[0046] 图2为夏佛托昔化学小分子物质标准样品的色谱图;

[0047] 图3为柚皮素化学小分子物质标准样品的色谱图;

[0048] 图4为美花石斛中夏佛托昔化学小分子的色谱图;

[0049] 图5为美花石斛中柚皮素化学小分子的色谱图。

[0050] 注:1号峰夏佛托昔;2号峰柚皮素。

具体实施方式

[0051] 下面结合附图和实施例对本发明作进一步详述,以下实施例只是描述性的,不是限定性的,不能以此限定本发明的保护范围。

[0052] 实施例1美花石斛的一代测序

[0053] 一代测序引物序列:

[0054] ITS-26SE:5' GAATTCCCGGTTTCGCTCGCCGTTAC 3' ;

[0055] ITS-17SE:5' ACGAATTCATGGTCCGGTGAAGTGTTTCG 3' 。

[0056] 扩增测序参数为:98℃变性2min后进行PCR循环,PCR循环参数为98℃20s;52℃30s;68℃1min,38个循环,68℃7min,扩增结束后设置4℃保温,并进行一代分子测序。

[0057] 通过一代测序,鉴别待测石斛药材的品种为美花石斛。

[0058] 实施例2美花石斛的提取方法

[0059] 取美花石斛干燥样品,用粉碎机粉碎,过药典筛(孔径0.335mm),精密称取石斛粉末1.000g(称量误差不能超过0.2%),置于100ml锥形瓶中,分别加入50mL 75%甲醇(V水:V甲醇=25:75),室温下超声30min后取出,过滤,滤液旋蒸浓缩至干,用75%甲醇溶剂(V水:V甲醇=25:75)溶解,最后转移至10ml容量瓶中定容,摇匀,以0.45μm微孔滤膜过滤,即得美花石斛样品溶液。

[0060] 实施例3美花石斛提取物的色谱检测方法

[0061] ①对照品溶液制备

[0062] 分别精密称取夏佛托昔4.10mg和柚皮素4.08mg,分别置于10ml容量瓶中,加75%(V/V)甲醇溶解稀释,摇匀,作为储备液。于4℃冰箱内冷藏备用。

[0063] 再分别精密吸取一定量的对照品储备溶液,利用75%甲醇稀释,准确配制夏佛托昔和柚皮素混合对照品溶液。通过不同的稀释比例,稀释制备成分7个浓度点。注入高效液相色谱仪。

[0064] ②美花石斛的小分子成分含量测定样品提取处理方法:

[0065] 取本品粉末(过三号筛) 1.00g,精密称定,置100ml容量瓶中,精密加甲醇-水(75:25) 50ml,超声处理(功率250W,频率40kHz) 30分钟,放冷,过滤,滤液旋蒸浓缩至干,用5ml甲醇-水(75:25)溶解,上清液过0.45 μ m微孔滤膜,取续滤液,即得。

[0066] ③美花石斛小分子成分含量测定色谱条件:

[0067] 色谱条件:

[0068] 含量测定色谱条件:Grace Allitima C18色谱柱(250mm*4.6mm,5 μ m);流动相采用二元梯度洗脱系统,A相:0.2%醋酸-水,B相:乙腈;梯度洗脱程序如表1;以波长290nm测定柚皮素,以波长334nm测定夏佛托苷;参比波长为500nm,柱温30 $^{\circ}$ C;流速1.0mL/min,进样量20 μ L。

[0069] 指纹图谱色谱条件:Grace Allitima C18色谱柱,优选为250mm \times 4.6mm,5 μ m规格的色谱柱;流动相:A相:0.4%乙酸+20mmol/L醋酸铵水溶液,B相:乙腈;梯度洗脱:0~12min:2%~15%B相,12~35min:15%~24%B相,35~45min:24%~36%B相,45~60min:36%~75%B相,60~80min:75%~95%B相;流速1.0mL/min;柱温30 $^{\circ}$ C;进样量20 μ L;检测波长280nm。

[0070] 表1美花石斛小分子成分含量测定的洗脱梯度

	T(min)	B相(乙腈)
	0	15%
[0071]	12	20%
	35	60%
	45	95%

[0072] 图1为美花石斛全成分指纹图谱色谱图,检测波长280nm;

[0073] 图2为夏佛托苷(1号峰)标准样品的检测图谱图;

[0074] 图3为柚皮素(2号峰)标准样品的检测图谱图;

[0075] 图4为美花石斛中夏佛托苷(1号峰)的检测图谱;

[0076] 图5为美花石斛中柚皮素(2号峰)的检测图谱。

[0077] 实施例4美花石斛中与夏佛托苷的相关特征指纹图谱的建立

[0078] 1. 石斛样品溶液制备

[0079] 取石斛干燥样品,用粉碎机粉碎,过药典筛(孔径0.335mm),精密称取石斛粉末1.000g(称量误差不能超过0.2%),置于100ml锥形瓶中,分别加入50mL75%甲醇(V水:V甲醇=25:75),室温下超声30min后取出,过滤,滤液旋蒸浓缩至干,用75%甲醇溶剂(V水:V甲醇=25:75)溶解,最后转移至10ml容量瓶中定容,摇匀,以0.45 μ m微孔滤膜过滤,即得。表2为得到的对照品夏佛托苷的线性关系。

[0080] 表2对照品夏佛托苷线性关系表

[0081]

组分	线性方程	线性范围(μ g/ml)	相关系数 R^2	检测限(μ g/ml)	定量限(μ g/ml)
夏佛托苷	$y = 39.876x - 3.8086$	1.025~41	0.9999	0.0169	0.0846

[0082] 2. 美花石斛夏佛托苷的相关特征指纹图谱的建立方法

[0083] 第一步:计算所有协变量 x 与 y 的相关系数;

[0084] 第二步:将相关系数的绝对值从大到小进行排列,选出前 $2\sqrt{n}$ 个协变量,记为 x_1, x_2, \dots, x_p ;

[0085] 第三步:将 y 与 x_1, x_2, \dots, x_p 进行线性回归,采用Lasso方法,进行变量选择。

[0086] Lasso (Least Absolute Shrinkage and Selection Operator) 函数的第一部分表示模型拟合的优良性,第二部分可以视为惩罚。该方法把小的系数往0压缩,一旦某个系数被压缩到0,对应的变量就被删除。就好像用“筛子”过滤,把影响小的变量一次就筛掉了。 λ 越小,模型中的变量越多 λ 越大,收缩量越大,选出的变量就越少。而Lasso方法是一种连续的、有序的过程,方差较小。当调节参数足够大时,惩罚项具有将其中某些系数的估计值强制设定为0的作用,因而Lasso方法可以进行变量选择,能够得到稀疏模型。

[0087] 当自变量为 p ,样本量为 n ,当 $p \gg n$,首先采用SIS方法降维,再采用Lasso方法筛选变量。

[0088] $SIS M_\gamma = \{1 \leq i \leq p: |\omega_i| \text{ 是前 } |\gamma n| \text{ 个比较大的}\}$

[0089] 其中, $M_* = \{1 \leq i \leq p: \beta_i \neq 0\}$ 表示真模型中非零系数的下标集; $s = |M_*|$ 表示非零系数的个数; $\omega = (\omega_1, \omega_2, \dots, \omega_p)^T = X^T y$;对于任意给定的 $\gamma \in (0, 1)$, ω 的 p 个元素按绝对值从大到小排列并且定义;此时 $|\gamma n| < n$,选取 M_γ 中下标对应的自变量,使超高维降到 d ($d \leq n$)维;其中, $d = n$ 或者 $d = \lfloor n / \log n \rfloor$ 。

[0090] Lasso方法筛选的线性模型为: $y_i = X_i^T \beta + \varepsilon_i, i = 0, \dots, n$

[0091] 其中, y_i 为第 i 个响应变量, $y = (y_1, y_2, \dots, y_n)'$; X_i 是 $P_n \times 1$ 阶的协变量, $X = (x_1, x_2, \dots, x_n)'$; ε_i 是均值为0,方差为 σ^2 的i.i.d的随机误差项, $E(\varepsilon) = 0, \text{Var}(\varepsilon) = \sigma^2 I_n$ 。

[0092] 假定随机项服从古典假定,即:

[0093] (1) 随机项具有零均值, $E(\varepsilon_i | x_i) = 0$;

[0094] (2) 随机项具有同方差, $\text{Var}(\varepsilon_i | x_i) = \sigma^2$;

[0095] (3) 随机项无序列相关性, $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$;

[0096] (4) ε 服从正态分布, $\varepsilon_i \sim N(0, \sigma^2)$ 。

[0097] 随机项的方差矩阵是一个对角线为 σ^2 ,其他地方为0的方阵,如下所示:

$$[0098] \quad \text{Var}(\varepsilon) = \sigma^2 I_n = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}_{n \times n}$$

$$[0099] \quad \text{其中, } I_n \text{ 是一个 } n \times n \text{ 的单位阵, } n \text{ 为数据的样本量, } I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}。$$

[0100] 为了同时进行变量选择和对参数进行估计,Lasso方法通过惩罚最小二乘目标函数式 I 的最小化来实现。

[0101]
$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^{p^*} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p^*} |\beta_j|$$

[0102] 式 I

[0103] 其中, $y = (y_1, y_2, \dots, y_n)^T \in R^n$ 为响应变量向量。以石斛数据为例, 每种石斛有两个响应变量序列(夏佛托苷 ($\mu\text{g/g}$)、柚皮素 ($\mu\text{g/g}$))。响应变量受自变量的影响, 一般情况下 y 为连续变量。

[0104] $x = (x_1, x_2, \dots, x_n)^T$ 为 $n \times p$ 的设计矩阵, 包含对响应变量有影响的所有候选自变量。

[0105] p 为样本的维数, n 为样本容量。在石斛数据中, 维数 p 远大于样本容量 n , 因此最小二乘估计不再适用, 需要采用变量选择的方法来进行模型估计。

[0106] $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ 是一个 p 维的参数。

[0107] $\sum_{j=1}^{p^*} |\beta_j|$ 为惩罚函数, λ 为调整参数。在变量选择方法中, 模型拟合的优良程度与对于入选变量个数惩罚的力度之间的平衡通过不同的准则来体现, 而这里是通过直接选取调节参数来实现的, 不同的 λ 值对应不同的惩罚力度。 λ 越大, 压缩的程度越强, 最后估计得到的非零参数越少, 选择 λ 最常见的方法是 K 折交叉验证法:

[0108] K-fold CV: $CV_K = \frac{1}{n} \sum_{k=1}^K \|y^{(k)} - \hat{y}^{(-k)}\|^2$

[0109] 一般, K 可取为 5 或 10。

[0110] GCV 准则是 CV 准则中的 K 取 n 时的一种近似情形, 定义为:

[0111] $GCV = \frac{1}{n} \frac{SSE_k}{(1 - df/n)^2}$

[0112] 其中, SSE_k 是含有 k 个变量的 CV 子模型的残差平方和, $df = \text{trace}\{P(\lambda)\}$ 。对于最终最优模型的选择, 可以取 CV 值或 GCV 值最小的子模型。

[0113] 采用线性模型, 由于初始自变量 $p = 434$ 个, 样本量 $n = 13$ 个, $p \gg n$, 故变量筛选是较为重要的工作。先经 SIS 降维后, 采用 Lasso 方法筛选变量。

[0114] Lasso 方法筛选所得自变量有 4 个, R-square 为 0.8725。

[0115] Lasso 方法筛选所得自变量及其相应系数如下表 3, 其中第一列为入选自变量编号, 第二列为相应系数, 第三列为系数方差, 第四列为检验 P 值。

[0116] 表 3 Lasso 方法所得入选自变量及其相应系数

[0117]

x_model (选择的自变量 X)	β (自变量 X 的系数)	Varbeta(β 的方差)	p-value (P 值)
常数项	8.916578	0.689085	2.69E-38
241	-0.04671	0.011519	5.01E-05

[0118]

148	0.254178	0.016036	1.38E-56
183	-0.27674	0.02044	9.17E-42
383	-0.0121	0.002136	1.46E-08

[0119] 表3的结果显示,以夏佛托昔为响应变量的筛选结果:第1列为采用Lasso方法选出的变量,即选出了变量241、148、183、383,其所对应的p值(列4)均小于显著性水平0.05,具有显著性差异。上述自变量的含义:美花石斛n批次样本(n不小于10)的指纹图谱根据保留时间对齐后的指纹图谱峰面积值。

[0120] 列2为每个变量具体对应的 β 参数值。 β 值为正说明该变量对石斛夏佛托昔小分子存在正向影响; β 值为负说明该变量对石斛夏佛托昔小分子存在负向影响。 β 值的绝对值大小显示的是该变量对石斛夏佛托昔小分子影响程度的大小。具体来说,在表3中,变量148对石斛夏佛托昔小分子的影响为正;变量241、183、383对石斛夏佛托昔小分子的影响为负,其中变量183的负向影响较大。

[0121] 实施例5美花石斛柚皮素的相关特征指纹图谱的建立

[0122] 1. 对照品溶液制备

[0123] 精密称定柚皮素4.08mg,置于10ml容量瓶,加75%甲醇溶解稀释,摇匀,作为储备液。于4℃冰箱内冷藏备用。再精密吸取一定量的对照品储备溶液,加75%甲醇稀释,准确配制柚皮素对照品溶液。通过不同的稀释比例,稀释制备成分7个浓度点,注入高效液相色谱仪。表4为得到的对照品柚皮素的线性关系。

[0124] 表4对照品柚皮素线性关系表

[0125]

组分	线性方程	线性范围 ($\mu\text{g/ml}$)	相关系数 R^2	检测线 ($\mu\text{g/ml}$)	定量限 ($\mu\text{g/ml}$)
柚皮素	$y = 64.534x - 7.5584$	1.02~40.8	0.9998	0.016	0.082

[0126] 2. 柚皮素相关特征指纹图谱的建立

[0127] 采用线性模型,由于初始自变量 $p=434$ 个,样本量 $n=17$ 个, $p \gg n$,故变量筛选是较为重要的工作。本发明采用Lasso方法筛选变量。

[0128] Lasso方法筛选所得自变量有4个, R -square为0.7399。

[0129] Lasso方法筛选所得自变量及其相应系数如下表5,其中第一列为入选自变量编号,第二列为相应系数,第三列为系数方差,第四列为检验P值。

[0130] 表5 Lasso方法所得入选自变量及其相应系数

[0131]

x_model (选择的自变量 X)	β (自变量 X 的系数)	$Var\beta$ (β 的方差)	p-value (P 值)
常数项	10.71728	0.338727	1E-219

[0132]

262	-0.00123	0.00052	0.018105
324	0.012763	0.000997	1.64E-37
421	0.27064	0.031952	2.45E-17
392	0.007978	0.00146	4.64E-08

[0133] 表5的结果显示,以柚皮素为自变量结果:采用Lasso方法选出的变量即为列1所示,即筛选出了变量262、324、421、392,该变量的p值(列4)均小于显著性水平0.05,具有显著性差异。上述自变量的含义:美花石斛的n批次样本(n不小于10)的指纹图谱根据保留时间对齐后的指纹图谱峰面积值。

[0134] 列2给出的是每个变量具体对应的 β 参数值。 β 值为正说明该变量对石斛柚皮素小

分子存在正向影响;β值为负说明该变量对石斛柚皮素小分子存在负向影响。β值的绝对值大小显示的是该变量对石斛柚皮素小分子影响程度的大小。具体来说,在表5中,变量324、421、392对石斛柚皮素小分子的影响为正,其中变量421对石斛柚皮素小分子的影响较大;变量262对石斛柚皮素小分子的影响为负。

[0135] 根据一代分子数据的结果显示该技术可清晰区地把美花石斛与其他石斛区分开来,具有可鉴定种的作用,一代测序分子序列可以作为该种鉴定指标。同时,美花石斛的特征指纹图谱的两种化学小分子含量在种内各样品表现稳定,并能明确与其他石斛区分开来,也可以作为该种的鉴定指标之一。因此,一代数据和指纹图谱均可作为鉴定该种的指标,并且两者有以下关联关系:(1)当一代数据鉴定样品为美花石斛时,该样品的指纹图谱具有特定的特征,即由一代数据即可鉴定出该种,并可知该种的指纹图谱特征;(2)当指纹图谱鉴定样品为美花石斛时,也可以推知其一代数据。因此两者互为充分必要条件,并且由于该种的指纹图谱特征在各代表样品中表现稳定,可确定该种的药效成份含量。因此,一代测序数据及指纹图谱均可作为该种的鉴定指标,并可作为药材的评价指标。若在品种鉴定与质量评价中分别采用一代测序和指纹图谱进行测定时,可将两种方法的鉴定结果结合在一起,使得对石斛药材的鉴定结果更加的准确、可靠。

[0136] 实施例7美花石斛的一代测序及小分子成分的相关特征指纹图谱在石斛药材质量评价中的应用

[0137] 1.利用以下一代测序引物序列及扩增测序参数测定待测石斛药材的序列,将其鉴定为美花石斛药材。

[0138] 一代测序引物序列:

[0139] ITS-26SE:5' GAATTCCCCGGTTCGCTCGCCGTTAC 3' ;

[0140] ITS-17SE:5' ACGAATTCATGGTCCGGTGAAGTGTTTCG 3' ;

[0141] 扩增测序参数为:98℃变性2min后进行PCR循环,PCR循环参数为98℃20s;52℃30s;68℃1min,38个循环,68℃7min,扩增结束后设置4℃保温。

[0142] 通过以上测序,确定待测石斛药材为美花石斛品种。

[0143] 2.供试品测定样品的制备

[0144] 精密称取美花石斛粉末,置100ml容量瓶中,每1g样品精密加体积比为75:25的甲醇-水50ml,以250W功率、40kHz频率超声处理30分钟,冷却后过滤,将滤液旋蒸浓缩至干,每1g美花石斛粉末对应地用5ml体积比为75:25的甲醇-水溶解,上清液过0.45μm微孔滤膜,取续滤液,即得美花石斛的小分子成分含量测定样品。

[0145] 3.色谱检测

[0146] 色谱条件:

[0147] 色谱柱:Grace Allitima C18色谱柱(250mm×4.6mm,5μm);流动相:A相:0.4%乙酸+20mmol/L醋酸铵水溶液,B相:乙腈;梯度洗脱:0~12min,2%~15%B,12~35min,15%~24%B,35~45min,24%~36%B;45~60min,36%~75%B;60~80min,75%~95%B,流速1.0mL/min;柱温30℃;进样量20μL;检测波长280nm。

[0148] 样品制备方法:

[0149] 取本品粉末(过三号筛)1.00g,精密称定,置100ml容量瓶中,精密加甲醇-水(75:25)50ml,超声处理(功率250W,频率40kHz)30分钟,放冷,过滤,滤液旋蒸浓缩至干,用5ml甲

醇-水 (75:25) 溶解, 上清液过 $0.45\mu\text{m}$ 微孔滤膜, 取续滤液, 即得。

[0150] 在检测时以波长 280nm 测定全成分指纹色谱图, 将得到的全成分指纹图谱与图1为对照的指纹图谱进行相似度比对; 相似度大于 0.85 为质量合格。

[0151] 中药指纹图谱是分析仪器检测得到的各种反映中药材、半成品和中成药(或植物药)所含复杂化学物质成分分布的量化特征关联药效活性控制为特点, 从宏观上整体反映中药材、半成品和中成药(或植物药)中所含化学物质成分的种类、数量和含量特征, 并能量化揭示潜在复杂的生物活性信息特征的图谱。

[0152] 本发明首先通过一代测序确定石斛的种类, 再通过测定石斛小分子成分含量数据与石斛指纹图谱全部峰面积进行关联性研究, 通过关联数据建模, 找出石斛内在的关联性, 这样即可全面评价石斛质量, 且能有效精确地鉴别和控制美花石斛药材的质量, 使得分析结果更加可靠, 避免其他种类的石斛药材的干扰。

[0153] 以上, 仅为本发明较佳的具体实施方式, 但发明的保护范围并不局限于此, 任何熟悉本技术领域的技术人员在本发明揭露的技术范围内, 可轻易想到的变化或替换, 都应涵盖在本发明的保护范围之内。因此, 本发明的保护范围应该以权利要求书的保护范围为准。

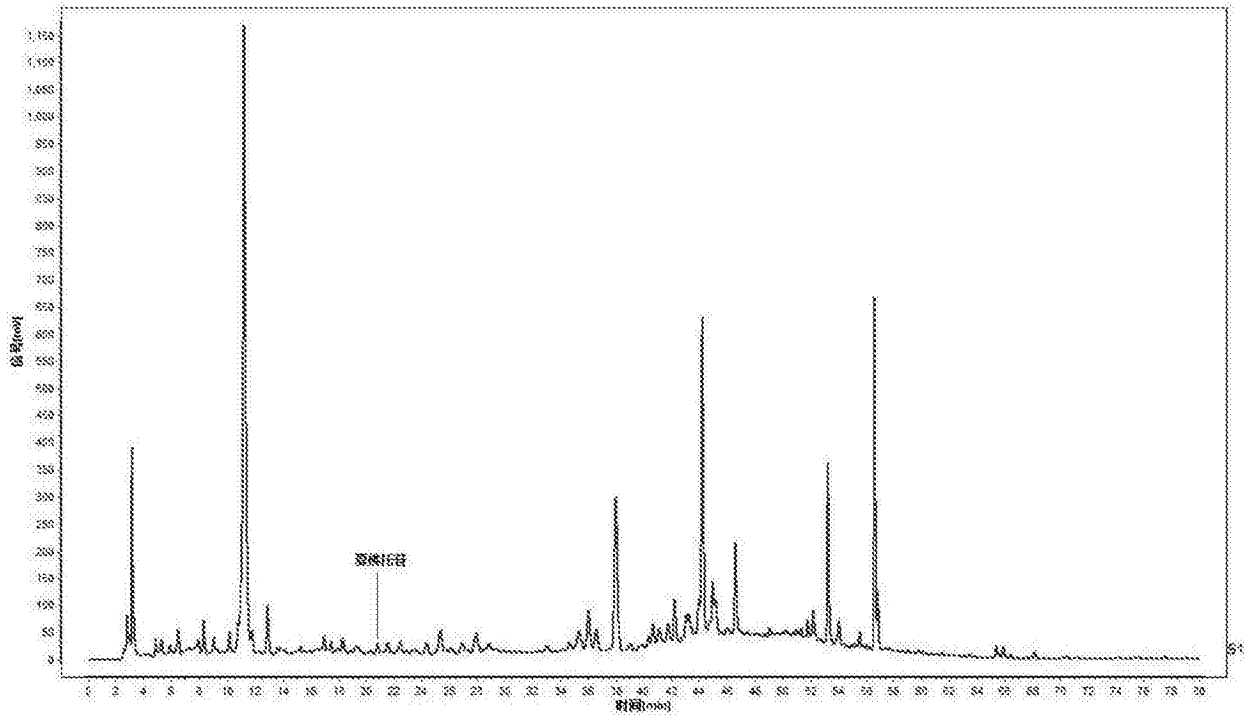


图1

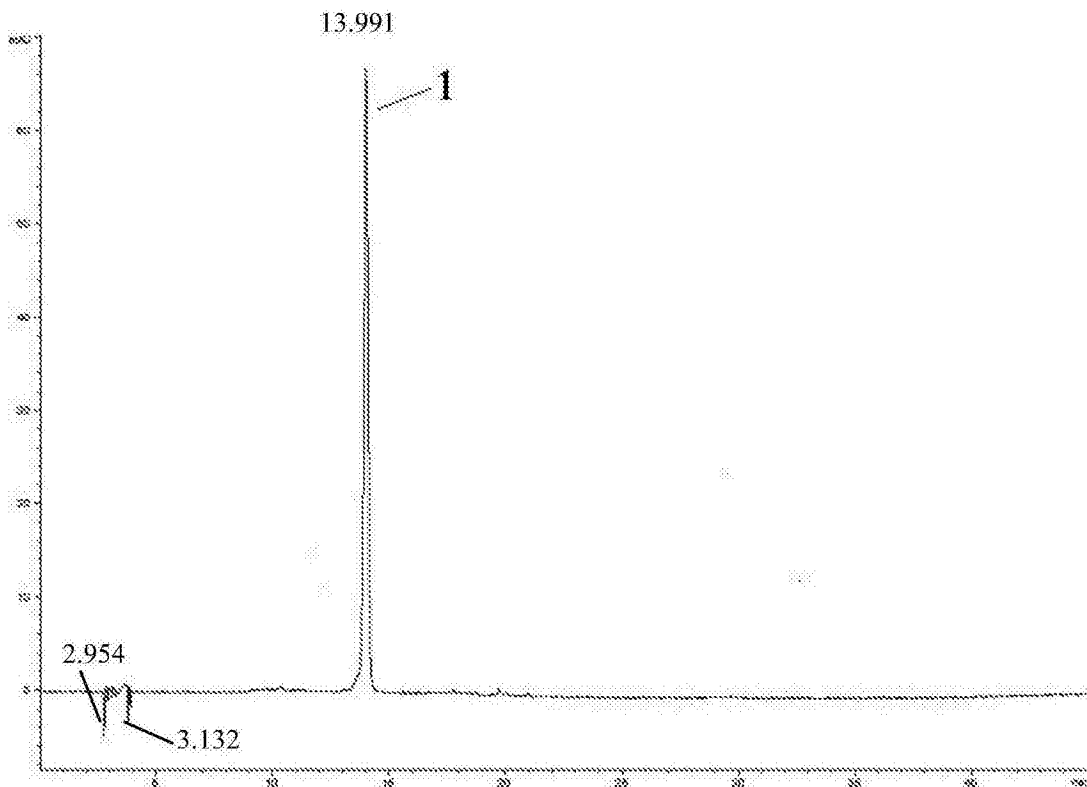


图2

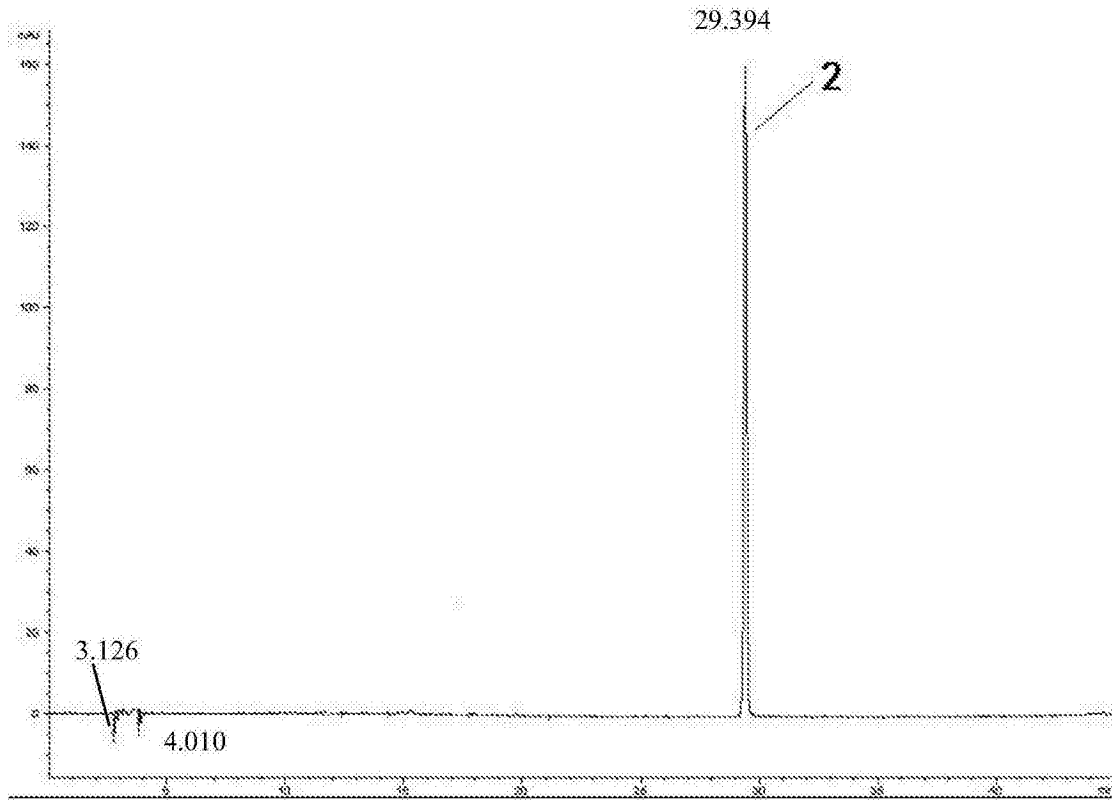


图3

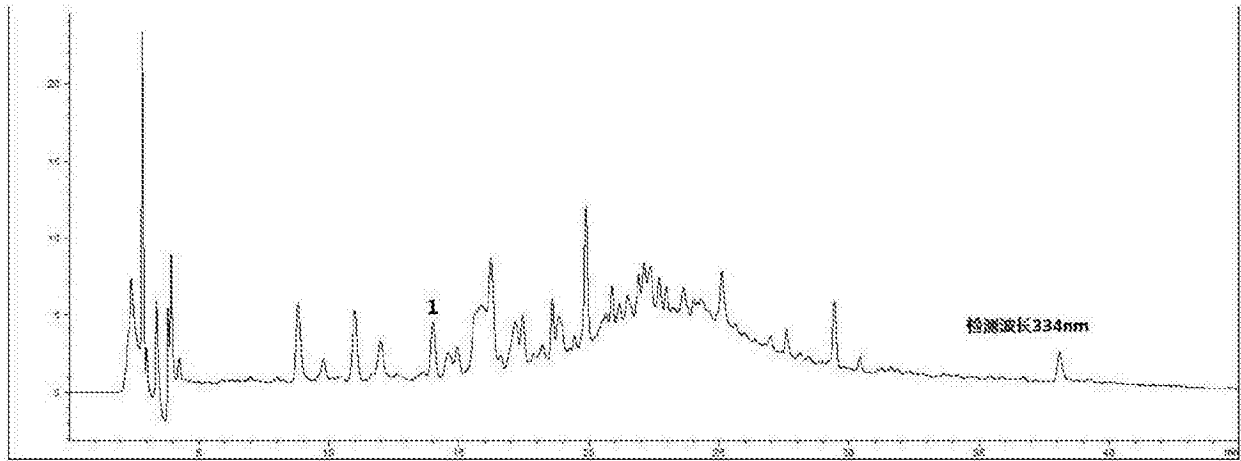


图4

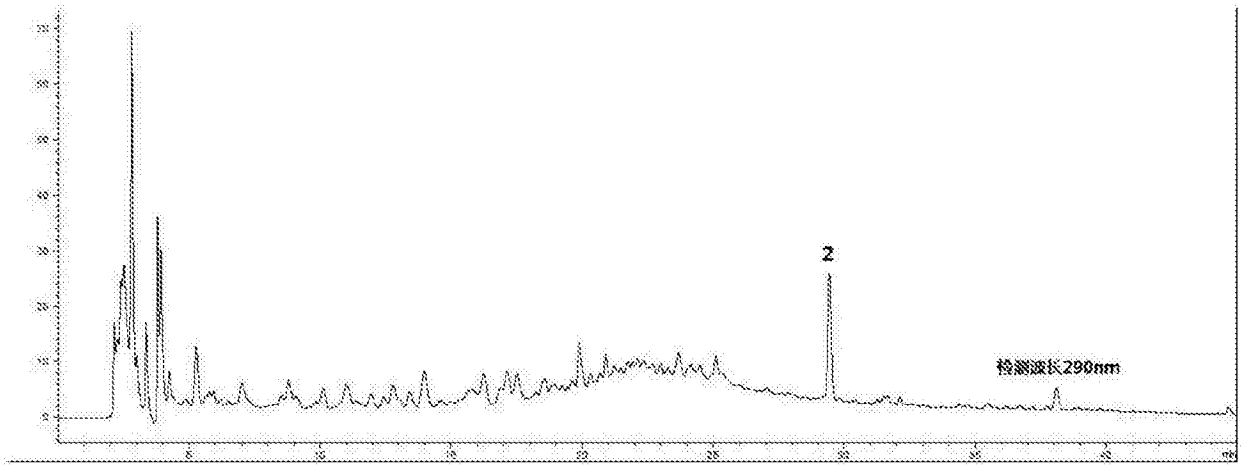


图5