(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0113531
(43) 공개일자 2011년10월17일

(51) Int. Cl.
G01N 30/72 (2006.01) *G01N 33/15* (2006.01)
G01N 33/02 (2006.01)
(21) 출원번호 10-2010-0032980
(22) 출원일자 2010년04월09일
심사청구일자 2010년04월09일

(71) 출원인
고려대학교 산학협력단
서울 성북구 안암동5가 1
대한민국(농촌진흥청장)
경기 수원시 권선구 서둔동 250번지
(72) 발명자
이동호
서울 성북구 돈암2동 한진아파트 208동 1802호
김나현
서울특별시 송파구 잠실2동 리센츠 251동 2203호
(뒷면에 계속)
(74) 대리인
김선애

전체 청구항 수 : 총 19 항

(54) 크로마토그래피－질량분석을 이용한 인삼 연근 판별 방법

(57) 요 약

본 발명은 인삼 시료로부터 대사체를 추출하는 제1단계; 상기 추출된 대사체를 LC/MS 또는 GC/MS를 이용하여 분석하는 제2단계; 상기 LC/MS 또는 GC/MS 분석 결과를 통계처리 가능한 수치로 변환하는 제3단계; 및 상기 변환된 수치를 이용하여 통계학적 방법으로 분석하여 인삼 연근을 판별하는 제4단계를 포함하여 구성되는 크로마토그래피-질량분석을 이용하는 인삼 연근 판별 방법에 관한 것이다. 본 발명은 LC/MS 또는 GC/MS를 이용한 대사체학 분석기법을 이용하여 인삼의 대사체를 분석함으로써 인삼의 연근을 단시간 내에 정확하고 정밀하게 판별할 수 있고, 세근을 이용할 경우 인삼 뿌리의 손실을 최소화할 수 있는 이점이 있다.

대 표 도 - 도1



연근	부위	개체수	채집지역	채집날짜
1	주근	10	농진청 포장	2007.01.12
2	주근	10	농진청 포장	2007.01.12
3	주근	10	농진청 포장	2007.01.12
4	주근	10	농진청 포장	2007.01.12
	세근	10	농진청 포장	2007.01.12
5	주근	10	농진청 포장	2007.01.12
	세근	10	농진청 포장	2007.01.12
6	주근	10	농진청 포장	2007.01.12
	세근	10	농진청 포장	2007.01.12

(72) 발명자

김금옥

서울특별시 성북구 안암동5가 고려대학교자연계캠퍼스 생명과학대 서관 L513

이재원

서울특별시 서초구 반포동 반포@ 84-401

방경환

경기도 수원시 팔달구 화서2동 금강아파트 151동 101호

김영창

경기도 수원시 권선구 구운동 삼환아파트 11동 1501호

신유수

서울특별시 금천구 시흥2동 벽산1단지아파트 116동 1203호

차선우

경기도 수원시 권선구 금곡동 262-18 신미주아파트 4동 1501호

이 발명을 지원한 국가연구개발사업

과제고유번호 20070501034007009030O

부처명 농촌진흥청

연구관리전문기관

연구사업명 BIO그린21사업

연구과제명 [협동] 대사체학을 이용한 인삼 연근 판별 기술 개발

기여율

주관기관 고려대학교

연구기간 2009년 01월 01일 ~ 2009년 12월 31일

특허청구의 범위

청구항 1

인삼 시료로부터 대사체를 추출하는 제1단계;

상기 추출된 대사체를 액체크로마토그래피-질량분석(LC/MS)을 이용하여 분석하는 제2단계;

상기 액체크로마토그래피-질량분석(LC/MS) 분석 결과를 통계처리 가능한 수치로 변환하는 제3단계; 및

상기 변환된 수치를 이용하여 통계학적 방법으로 분석하여 인삼 연근을 판별하는 제4단계를 포함하여 구성되는 크로마토그래피-질량분석을 이용한 인삼 연근 판별 방법.

청구항 2

제1항에 있어서, 상기 통계적인 방법은 주성분 분석 (Principal Component Analysis, PCA) 또는 계층적군집 분석 (Hierarchical Cluster Analysis, HCA)인 것을 특징으로 인삼 연근 판별 방법.

청구항 3

제1항 또는 제2항에 있어서, 상기 인삼 시료는 주근이며 1-3년근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 4

제1항 또는 제2항에 있어서, 상기 인삼 시료는 세근이며 4-6년근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 5

제1항 또는 제2항에 있어서, 상기 대사체 중 의미 있는 영향력을 주는 대사체를 선택하여 분석하는 특징 선택 (feature selection) 단계를 더 포함하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 6

제5항에 있어서, 상기 인삼 시료는 주근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 PLS-DA(Partial Least Squre-Discriminant Analysis) 방법을 이용하여 대사체를 선택하고 분석하여 인삼 연근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 7

제5항에 있어서, 상기 인삼 시료는 주근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 RF (Random Forest), PAM (Predicion Analysis for Microarray) 및 PLS-DA (Partial Least Squre-Discriminant Analysis)로 이루어진 그룹에서 선택되는 2이상의 방법에 의해 공통으로 선택된 대사체를 이용하여 분석하여 인삼 연근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 8

제5항에 있어서, 상기 인삼 시료는 세근이고 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 RF (Random Forest), PAM (Predicion Analysis for Microarray) 및 PLS-DA (Partial Least Squre-Discriminant Analysis)로 이루어진 그룹에서 선택되는 1 이상의 방법을 이용하여 대사체를 선택하고 분석하여 인삼 연근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 9

제5항에 있어서, 상기 인삼 시료는 세근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 RF (Random Forest), PAM (Predicion Analysis for Microarray) 및 PLS-DA (Partial Least Squre-Discriminant Analysis)로 이루어진 그룹에서 선택되는 2 이상의 방법에 의해 공통으로 선택된 대사체를 이용하

여 분석하여 인삼 연근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 10

인삼 시료로부터 분리된 대사체들에 대하여 액체크로마토그래피-질량분석 장치에 의해 미리 구한 머무름 시간 및 분자량을 포함하는 기준데이터를 기억하는 메모리; 및

상기 기준데이터와, 판별대상인 인삼시료에 대하여 액체크로마토그래피-질량분석 장치에 의해 구한 머무름 시간 및 분자량을 포함하는 측정데이터를 비교하는 분석수단을 포함하여 구성되는 것을 특징으로 하는 인삼 연근 판별 장치.

청구항 11

인삼 시료로부터 대사체를 추출하는 제1단계;

상기 추출된 대사체를 가스크로마토그래피-질량분석 (GC/MS)을 이용하여 분석하는 제2단계;

상기 가스크로마토그래피-질량분석 (GC/MS) 분석 결과를 통계처리 가능한 수치로 변환하는 제3단계; 및

상기 변환된 수치를 이용하여 통계학적 방법으로 분석하여 인삼 연근을 판별하는 제4단계를 포함하여 구성되는 크로마토그래피-질량분석을 이용하는 인삼 연근 판별 방법.

청구항 12

제11항에 있어서, 상기 통계적인 방법은 주성분 분석 (Principal Component Analysis, PCA) 또는 계층적군집분석 (Hierarchical Cluster Analysis, HCA)인 것을 특징으로 인삼 연근 판별 방법.

청구항 13

제11항 또는 제12항에 있어서, 상기 인삼 시료는 주근이며 1년근 또는 5년근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 14

제11항 또는 제12항에 있어서, 상기 대사체 중 의미 있는 영향력을 주는 대사체를 선택하여 분석하는 특징 선택 (feature selection) 단계를 더 포함하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 15

제14항에 있어서, 상기 인삼 시료는 주근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 PLS-DA(Partial Least Squre-Discriminant Analysis) 방법을 이용하여 대사체를 선택하고 분석하여 인삼 연근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 16

제14항에 있어서, 상기 인삼 시료는 주근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 RF (Random Forest), PAM (Predicion Analysis for Microarray) 및 PLS-DA (Partial Least Squre-Discriminant Analysis) 방법에 의해 공통으로 선택된 대사체를 이용하여 분석하여 인삼 연근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 17

제14항에 있어서, 상기 인삼 시료는 세근이고 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 PLS-DA (Partial Least Squre-Discriminant Analysis) 방법을 이용하여 대사체를 선택하여 분석하여 인삼 연근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 18

제14항에 있어서, 상기 인삼 시료는 세근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 RF (Random Forest), PAM (Predicion Analysis for Microarray) 및 PLS-DA (Partial Least Squre-Discriminant Analysis)로 이루어진 그룹에서 선택되는 2이상의 방법에 의해 공통으로 선택된 대사체를 이용하

여 분석하여 인삼 연근을 판별하는 것을 특징으로 하는 인삼 연근 판별 방법.

청구항 19

인삼 시료로부터 분리된 대사체들에 대하여 기체크로마토그래피-질량분석 장치에 의해 미리 구한 머무름 시간 및 분자량을 포함하는 기준데이터를 기억하는 메모리; 및

상기 기준데이터와, 판별대상인 인삼시료에 대하여 기체크로마토그래피-질량분석 장치에 의해 구한 머무름 시간 및 분자량을 포함하는 측정데이터를 비교하는 분석수단을 포함하여 구성되는 것을 특징으로 하는 인삼 연근 판별 장치.

명 세 서

기 술 분 야

[0001] 본 발명은 크로마토그래피-질량분석을 이용한 인삼 연근 판별 방법에 관한 것이다. 더욱 상세하게 본 발명은 액체크로마토그래피-질량분석 (LC/MS) 또는 기체크로마토그래피-질량분석 (GC/MS)을 이용한 대사체학 분석기법을 이용하여 인삼의 대사체를 분석함으로써 인삼의 연근을 정확하고 정밀하고 신속하게 판별할 수 있는 방법에 관한 것으로, 이를 이용하여 인삼의 연근 관리에 객관적인 검증 시스템을 마련함으로써 인삼 유통에 대한 신뢰성을 향상시킬 수 있다.

배 경 기 술

[0002] 인삼은 두릅나무과 (Araliaceae)의 깊은 산악지에서 자라는 다년초로서 흔히 재배하고 있으며 높이가 60cm에 달하고 근경은 짧으며 곧거나 비스듬히 서고 밑에서 도라지같은 뿌리가 발달하며 근경 끝에서 1개의 원줄기가 나오고 끝에서 3~4개의 잎이 윤생하며 긴 엽병 끝에 5개의 장상복엽이 달린다. 소엽은 난형 또는 도란형이고 끝이 뾰족하며 밑부분이 좁고 표면맥 위에 잔털이 약간 있으며 가장자리에 잔톱니가 있다. 꽃은 4월에 피고 연한 녹색이며 산형화서에 달리고 화서는 윤생엽의 중앙부에서 1개가 나온다. 꽃받침, 꽃잎 및 수술은 각각 5개이며 암술대는 2개이고 열매는 둥글며 여러 개가 산형화서에 모여 달리고 적색으로 익는다. 뿌리를 약용으로 한다 (대한식물도감, 1993).

[0003] 인삼은 예로부터 한방약재로 널리 쓰이며 원기를 보하고 신체허약, 권태, 피로, 식욕부진, 구토, 설사에 쓰이며 폐기능을 도우며 진액을 생성하고 안신작용 및 신기능을 높여 준다고 알려져 있다. 인삼의 약리작용은 대뇌피질 흥분과 억제, 평형, 항피로, 항노화, 면역증강, 심장수축, 성선촉진, 고혈당억제, 단백질합성촉진, 항상성유지, 항암, 해독작용 등이 보고되었다.

[0004] 한국에서 재배되는 인삼의 뿌리는 비대근으로 원뿌리와 2~5개의 지근(支根)으로 되어 있고 미황백색이다. 분지성이 강한 식물이며 그 뿌리의 형태는 나이에 따라 차이가 있고 수확은 4~6년근 때에 한다. 한국 전매품인 홍삼의 원료로 쓰이는 것은 모두 6년근이다. 6년근은 길이 7~10cm, 지름 2.5cm 내외이고 뿌리 길이는 34cm, 무게 80g 정도이다. 인삼은 매년 땅속줄기에서 싹이 나오고 가을에는 줄기와 잎이 고사한다.

[0005] 인삼은 4~6년근이 주로 유통되는데, 가을에 재배한 6년근 인삼의 약효가 최고조라고 한다. 이를 바탕으로 인삼의 유통 역시 6년근을 위주로 이루어지는데, 4년근과 5년근에 비해 공급량이 턱없이 부족한 현실이다. 하지만 수요량에 비해 공급량이 절대적으로 부족한 생산 현실과는 반대로 시중에서는 6년근 인삼이 넘쳐나고 있다. 이는 4, 5년근 인삼을 6년근으로 속여서 팔기 때문이다. 하지만 이러한 일이 빈번하게 일어남에도 아직까지 뚜렷한 연근 판별 및 관리 제도가 없기 때문에 체계적인 관리가 힘든 실정이다.

[0006] 기존의 인삼의 연근 판별은 인삼뿌리의 머리, 줄기 흔적 확인에 의한 판별, 지근 발달에 의한 판별, 체형에 의한 판별 등의 육안 판별 방법과 염색에 의한 나이테 판별의 실험적 판별 방법 등의 주로 외부형태 특징에 의존한 판별을 하고, 근래 NIR 분석법 또는 NMR에 의한 연근 판별을 시도하고 있으나, 정확한 판별에는 한계가 있고 파괴적이거나 장시간이 소요되어 적용이 어렵다. 최근 언론에서도 인삼의 불법 유통에 관한 보도가 증가되는 등의 상황을 고려해 볼 때, 인삼의 연근에 대한 정확한 판별기준을 마련하여 유통의 체계적인 관리 시스템을 확립하는 것이 시급하다.

[0007] 한편, 대사체학 (metabolomics)은 다양한 유전적, 환경적 조건에서 세포나 조직 내에 존재하는 저분자량 대사물질 (metabolome)의 구성과 농도를 Mass나 NMR과 같은 다양한 분석기법을 사용하여 분석함으로써 생명현상을 총체적으로 연구하기 위한 새로운 학문으로서, 총체적 대사체 분석/프로파일 구축, 해독 및 이용 기술을 적용하여, 기존의 지노믹스 (genomics) 및 프로티오믹스 (proteomics) 기술을 실현하고, 보다 직접적이고 구체적인 이용 방법을 수립할 수 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 상술한 문제점을 해결하기 위해 안출된 것으로서, 본 발명의 목적은 LC/MS 또는 GC/MS 대사체학을 응용하여 인삼의 대사체를 분석함으로써 정밀하고 정확한 인삼 연근 판별 방법을 제공하여, 인삼의 연근 관리에 객관적인 검증 시스템을 마련함으로써 인삼 유통에 대한 신뢰성을 향상시키는 것이다.

[0009] 본 발명의 다른 목적은 비파괴적인 방법으로 뿌리의 손실을 최소화하여 극소량으로 단시간 내에 인삼 연근을 판별할 수 있는 인삼 연근 판별 방법을 제공하는 것이다.

[0010] 본 발명의 다른 목적 및 이점은 하기의 발명의 상세한 설명, 청구범위 및 도면에 의해 더욱 명확하게 된다.

과제의 해결 수단

[0011] 본 발명의 일 양태에 의하면, 본 발명은 인삼 시료로부터 대사체를 추출하는 제1단계; 상기 추출된 대사체를 액체크로마토그래피-질량분석(LC/MS)을 이용하여 분석하는 제2단계; 상기 액체크로마토그래피-질량분석(LC/MS) 분석 결과를 통계처리 가능한 수치로 변환하는 제3단계; 및 상기 변환된 수치를 이용하여 통계학적 방법으로 분석하여 인삼 연근을 판별하는 제4단계를 포함하여 구성되는 크로마토그래피-질량분석을 이용한 인삼 연근 판별 방법을 제공한다.

[0012] 본 발명에 의하면 크로마토그래피-질량분석 및 통계학적 방법을 이용하여 누구든지 최소의 시료로 과학적이며 정확하게 인삼 연근을 신속히 판별할 수 있는 이점이 있다.

[0013] 상기 제1단계에서 상기 인삼 시료의 대상체 분석을 위해 본 발명이 속하는 기술 분야의 공지된 추출법으로 적용할 수 있으며, 최적의 추출법으로 70% MeOH를 이용하여 추출할 수 있다.

[0014] 상기 제2단계에서 LC/MS 분석을 위한 적정량의 시료를 결정하기 하기 위해 기존 문헌을 참고하여 추출량, 분석농도, 주입량 등을 확립할 수 있다. 또한, 상기 제2단계에서 2 종류의 극성이 다른 이동상을 사용할 수 있다. 한 종류(이동상 A)는 완충용액을 사용하고, 다른 한 종류(이동상 B)는 유기용매를 사용할 수 있으며, 이들을 시간에 따라 농도구배를 가지는 상태로 사용하는 것이 바람직하다. 이동상 A는 0.1% 포름산이 첨가된 물을 사용하는 것이 바람직하다. 이동상 B는 극성이 높은 유기 용매를 사용할 수 있으며, 0.1% 포름산이 첨가된 아세토니트릴을 사용할 수 있지만 이에 한정되는 것은 아니며, 당업자는 임의의 목적에 따라 통상적인 유기 용매를 사용할 수 있다. 이동상의 유량은 해당 컬럼의 경우 200~600μL/min을 사용할 수 있으며, 해당실험에서는 500μL/min가 분석에 적합한 유량으로 설정되었다. 총 분석시간은 12분으로 B의 비율을 처음 0.5분간 10%로 유지하고, 2.5분까지 30%, 6분까지 60%, 9분까지 90%로 올려주고 10.5분까지 100%로 유지하며 12분까지 초기 조건인 10%로 흘려 초기 조건으로 안정화 시키는 것이 바람직하다. 이때 컬럼의 온도는 35℃를 유지할 수 있다.

[0015] 한편, 고성능 액체크로마토그래피-질량분석기를 이용하여 시료를 분석하는 경우, 먼저 액체크로마토그래피의 분석컬럼의 고정상과 이동상 사이에서 흡착성 또는 분배계수의 차를 이용하여 분리된 성분들은 머무름 시간의 차이를 두고 질량분석기로 도입된다. 도입된 시료는 이온화 장치에 의하여 이동상은 제거되고 분석물질들이 이온화된다. 본 발명에서는 상기 제2단계에서 분석컬럼으로 역상컬럼을 사용함으로써, 분석물질을 바람직하게 검출할 수 있다. 이때 역상컬럼으로는 C18과 C8 컬럼을 사용할 수 있으며, C18 컬럼을 사용하는 것이 바람직하며, 이를 사용하는 경우, 분리능(resolution)과 인텐서티의 증가를 이룰 수 있어 검출감도 개선을 제공한다. 상기 액체크로마토그래피의 분석컬럼을 통과하여 분리된 시료는 질량분석기로 도입하여 질량분석기에 장착된 전자분무(electrospray) 이온화장치를 이용하여 이온화시킬 수 있다. 상기 질량분석기로 분석하는 단계에서 정량분석

시 질량분석기의 MRM(multiple reaction monitoring) 모드에서 정량하는 것은 신호대잡음비(signal to noise ratio)를 개선하기 위한 것이다.

[0016] 또한, 질량분석기를 사용하는 경우 네거티브 (negative) 와 포지티브 모드 (positive mode)에서 각각 인삼 대사체 분석을 위한 최적의 조건을 확립한다. 예를 들어 네거티브 모드 (Negative mode)에서의 조건은 다음과 같다; capillary voltage 2800, cone voltage 30, collision energy 6, desolvation temperature 300℃, source temperature 120℃.

[0017] 본 발명의 구체적인 일 양태에 의하면, 상기 통계적인 방법은 주성분 분석 (Principal Component Analysis, PCA) 또는 계층적군집 분석 (Hierarchical Cluster Analysis, HCA)인 것을 특징으로 한다. 상기 주성분 분석 (PCA)이란 원래 변수들의 선형결합으로 표시되는 새로운 주성분을 찾고, 이를 통하여 자료의 요약과 용이한 해석을 목적으로 하는 통계적 기법이다. 즉 차후의 분석을 위한 수단을 제공하여 주는 분석방법이다. 계층적군집 분석(HCA)이란 모든 대상들이 각각 하나의 군집을 이루고 있는 상태에서 출발하여 순차적으로 가장 유사한 대상들끼리 군집화함으로써 유사성이 높은 군집을 이루며, 최종적으로 모든 대상이 하나의 군집으로 묶여지는 방법이다.

[0018] 본 발명의 구체적인 일 양태에 의하면, 상기 인삼 시료는 주근이며 1-3년근을 판별하는 것을 특징으로 한다. 본 발명에 의하면, 상기 인삼의 주근 시료로부터 분리된 LC/MS 분석 원 데이터만으로도 인삼의 1-3년근을 정확하게 구분할 수 있어, 신속하고 간편한 이점이 있다.

[0019] 본 발명의 구체적인 일 양태에 의하면, 상기 인삼 시료는 세근이며 4-6년근을 판별하는 것을 특징으로 한다. 본 발명에 의하면, 상기 인삼의 세근 시료로부터 분리된 LC/MS 분석 원 데이터만으로도 인삼의 4-6년근을 정확하게 구분할 수 있다. 상기와 같이 인삼의 세근을 이용하면 판별하고자 하는 인삼의 손실을 최소화하면서 인삼 연근을 판별할 수 있고 이로 인해 시장에서 활용도가 높을 것으로 기대된다.

[0020] 본 발명의 구체적인 일 양태에 의하면, 상기 대사체 중 의미 있는 영향력을 주는 대사체를 선택하여 분석하는 특징 선택 (feature selection) 단계를 더 포함하는 것을 특징으로 한다. 대사체의 일부만을 선택하여 보다 정확하고 신속하게 인삼 연근을 판별할 수 있는 이점이 있다.

[0021] 본 발명의 보다 구체적인 일 양태에 의하면, 본 발명에서는 RF (Random Forest, Y. Qiu et al. Metabolomics (2008) 4:337-346), PAM (Prediction Analysis for Microarray, Y. Qiu et al. Metabolomics (2008) 4:337-346), PLSDA (Partial Least Square-Discriminant Analysis, Y. Qiu et al. Metabolomics (2008) 4:337-346) 등 세 가지의 특징 선택 (feature selection) 방법들을 사용하였고, 각 특징 선택 (feature selection) 방법들은 각 방법의 고유한 알고리즘에 따라 대사체들에 중요도 점수를 부여하여 대사체를 선택하였다. RF는 Importance Score를 이용하여 대사체의 순위를 평가하는 것으로, 각 대사체마다 자신의 importance라는 값을 갖게 되고 importance값은 그 대사체가 인삼의 연근을 판별하는 정확도에 얼마나 큰 영향을 주는지를 수치화한다. PAM은 class centroid와 overall centroid와의 차이를 계산하여 대사체의 순위를 평가하는 것으로, 각 대사체마다 연근별 평균과 전체 평균을 계산하고 이들의 차이가 클수록 큰 점수를 준다. 예를 들어, 1번 대사체의 3년근에서의 평균, 4년근에서의 평균, 5년근에서의 평균, 6년근에서의 평균을 계산하고 전체 샘플에서의 평균을 계산한 다음, 연근별 평균과 전체 샘플과의 차이를 계산하고 이 차이가 클수록 영향력 있는 대사체라고 간주하였다. PLSDA는 회귀계수의 가중합을 이용하여 대사체의 순위를 평가하는 것으로, PLS 회귀분석을 하면 각 대사체 마다 회귀계수가 나오게 되고, 이 회귀계수의 절대값이 클수록 중요한 대사체라고 할 수 있다. 여기서 회귀계수는 대사체가 그룹을 판별하는데 미치는 영향력을 수치화 한 것이다.

[0022] 본 발명의 보다 구체적인 일 양태에 의하면, 상기 인삼 시료는 주근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 PLS-DA(Partial Least Squre-Discriminant Analysis) 방법을 이용하여 대사체를 선택하고 분석하여 인삼 연근을 정확하게 판별할 수 있다.

[0023] 본 발명의 보다 구체적인 일 양태에 의하면, 상기 인삼 시료는 주근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 RF (Random Forest), PAM (Predicion Analysis for Microarray) 및 PLS-DA (Partial Least Squre-Discriminant Analysis)로 이루어진 그룹에서 선택되는 2 이상의 방법에 의해 공통으로 선택된 대사체를 이용하여 분석하여 인삼 연근을 정확하게 판별할 수 있다.

[0024] 본 발명의 보다 구체적인 일 양태에 의하면, 상기 인삼 시료는 세근이고 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 RF (Random Forest), PAM (Predicion Analysis for Microarray) 및 PLS-DA (Partial Least Squre-Discriminant Analysis)로 이루어진 그룹에서 선택되는 1 이상의 방법을 이용하여 대사체를 선택하고 분석하여 인삼 연근을 판별할 수 있다.

[0025] 본 발명의 보다 구체적인 일 양태에 의하면, 상기 인삼 시료는 세근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 RF (Random Forest), PAM (Predicion Analysis for Microarray) 및 PLS-DA (Partial Least Squre-Discriminant Analysis)로 이루어진 그룹에서 선택되는 2 이상의 방법에 의해 공통으로 선택된 대사체를 이용하여 분석하여 인삼 연근을 정확하게 판별할 수있다.

[0026] 본 발명의 보다 구체적인 일 양태에 의하면, 상기 대사체는 머무름 시간(retention time, min.) 및 이온(ion, *m/z*)이 각각 4.1 및 1105, 3.2 및 1436, 4.4 및 971, 2.2 및 499, 3.6 및 883, 4.4 및 971, 4.2 및 841, 4.3 및 1143, 2.9 및 861인 것을 특징으로 한다 (표 14 참조). 상기와 같은 구성에 의하면 9개의 대사체를 이용하여 인삼 연근을 객관적이며 정확하고 신속하게 판별할 수 있다.

[0027] 본 발명의 다른 양태에 의하면, 본 발명은 상기 인삼 시료로부터 분리된 대사체들에 대하여 액체크로마토그래피-질량분석 장치에 의해 미리 구한 머무름 시간 및 분자량을 포함하는 기준데이터를 기억하는 메모리 및 상기 기준데이터와, 판별대상인 인삼시료에 대하여 액체크로마토그래피-질량분석 장치에 의해 구한 머무름 시간 및 분자량을 포함하는 측정데이터를 비교하는 분석수단을 포함하여 구성되는 것을 특징으로 하는 인삼 연근 판별 장치를 제공한다. 상기한 구성에 의하면, 소수개의 대사체를 분석하여 정확하고 신속하게 인삼 연근을 판별할 수 있다. 나아가 상기 장치를 휴대용으로 제작하면 장소에 구애받지 않고 현장에서 세근을 이용하여 인삼 연근을 바로 판별할 수도 있다. 상기 대사체들은 표 14에 기재된 대사체들인 것이 바람직하다.

[0028] 본 발명의 다른 양태에 따르면, 본 발명은 인삼 시료로부터 대사체를 추출하는 제1단계; 상기 추출된 대사체를 가스크로마토그래피-질량분석 (GC/MS)을 이용하여 분석하는 제2단계; 상기 가스크로마토그래피-질량분석 (GC/MS) 분석 결과를 통계처리 가능한 수치로 변환하는 제3단계; 및 상기 변환된 수치를 이용하여 통계학적 방법으로 분석하여 인삼 연근을 판별하는 제4단계를 포함하여 구성되는 크로마토그래피-질량분석을 이용한 인삼 연근 판별 방법을 제공한다.

[0029] 상기 제1단계에서 GC/MS를 이용한 대사체 분석을 위해 본 발명이 속하는 기술 분야의 공지된 추출법으로 적용할 수 있으며, 최적의 추출법으로 $CHCl_3$: MeOH (1:1) 를 이용하여 추출할 수 있다.

[0030] 상기 제2단계에서 GC/MS 분석을 위한 적정량의 시료를 결정하기 위해 기존 문헌을 참고하여 추출량, 분석농도, 주입량 (injection volume) 등을 확립할 수 있다. 연근별 상이성을 밝혀내고자 하는 이 실험에서는 과량의 공통 주성분 (major compound)이 검출 (detection)되는 시간인 23분부터 24분 50초 사이의 실험값을 제외함으로써 미량 성분들의 검출 (detection) 비율을 높이는 것이 바람직하다. 상기 제2단계에서 기체크로마토그래피와 질량분석기를 이용하여 시료를 분석하는 경우, 먼저 기체크로마토그래피에 장착된 모세분리관의 고정상과 이동상 사이에서 흡착성 또는 분배계수의 차를 이용하여 분리된 성분들은 머무름 시간의 차이를 두고 질량분석기로 도입된다. 도입된 시료는 이온화 장치에 의하여 분석물질들이 이온화된다.

[0031] 본 발명의 구체적인 일 양태에 의하면, 상기 통계적인 방법은 주성분 분석 (Principal Component Analysis, PCA) 또는 계층적군집 분석 (Hierarchical Cluster Analysis, HCA)인 것을 특징으로 한다.

[0032] 본 발명의 구체적인 일 양태에 의하면, 상기 인삼 시료는 주근이며 1년근 또는 5년근을 판별하는 것을 특징으로 한다.

[0033] 본 발명의 구체적인 일 양태에 의하면, 상기 대사체 중 의미 있는 영향력을 주는 대사체를 선택하여 분석하는

특징 선택 (feature selection) 단계를 더 포함하는 것을 특징으로 한다.

[0034] 본 발명의 보다 구체적인 일 양태에 의하면, 상기 인삼 시료는 주근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 PLS-DA(Partial Least Squre-Discriminant Analysis) 방법을 이용하여 대사체를 선택하고 분석하여 인삼 연근을 판별할 수 있다.

[0035] 본 발명의 보다 구체적인 일 양태에 의하면, 상기 인삼 시료는 주근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 RF (Random Forest), PAM (Predicion Analysis for Microarray) 및 PLS-DA (Partial Least Squre-Discriminant Analysis) 방법에 의해 공통으로 선택된 대사체를 이용하여 분석하여 인삼 연근을 판별할 수 있다.

[0036] 본 발명의 보다 구체적인 일 양태에 의하면, 상기 인삼 시료는 세근이고 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 PLS-DA (Partial Least Squre-Discriminant Analysis) 방법을 이용하여 대사체를 선택하여 분석하여 인삼 연근을 판별할 수 있다.

[0037] 본 발명의 보다 구체적인 일 양태에 의하면, 상기 인삼 시료는 세근이며 4-6년근을 판별하기 위해 상기 특징 선택 (feature selection) 단계에서 RF (Random Forest), PAM (Predicion Analysis for Microarray) 및 PLS-DA (Partial Least Squre-Discriminant Analysis)로 이루어진 그룹에서 선택되는 2 이상의 방법에 의해 공통으로 선택된 대사체를 이용하여 분석하여 인삼 연근을 판별할 수 있다.

[0038] 본 발명의 보다 구체적인 일 양태에 의하면, 상기 대사체는 머무름 시간(retention time, min.) 및 이온(ion, *m/z*)이 각각 16.4 및 73, 26.0 및 204, 9.5 및 73, 20.8 및 204, 26.5 및 73, 6.2 및 57, 3.4 및 244, 31.8 및 217, 9.5 및 147, 11.3 및 147, 22.4 및 73, 19.0 및 149, 16.6 및 71, 7.5 및 57, 3.7 및 171, 32.0 및 441, 30.5 및 217, 18.3 및 73, 12.7 및 73, 9.8 및 133, 11.0 및 142인 것을 특징으로 한다(표 28 참조). 상기와 같은 구성에 의하면 21개의 대사체를 이용하여 인삼 연근을 객관적이며 정확하고 신속하게 판별할 수 있다.

[0039] 본 발명의 다른 양태에 의하면, 본 발명은 상기 인삼 시료로부터 분리된 대사체들에 대하여 기체크로마토그래피-질량분석 장치에 의해 미리 구한 머무름 시간 및 분자량을 포함하는 기준데이터를 기억하는 메모리 및 상기 기준데이터와, 판별대상인 인삼시료에 대하여 기체크로마토그래피-질량분석 장치에 의해 구한 머무름 시간 및 분자량을 포함하는 측정데이터를 비교하는 분석수단을 포함하여 구성되는 것을 특징으로 하는 인삼 연근 판별 장치를 제공한다. 상기한 구성에 의하면, 소수개의 대사체를 분석하여 정확하고 신속하게 인삼 연근을 판별할 수 있다. 나아가 상기 장치를 휴대용으로 제작하면 장소에 구애받지 않고 현장에서 세근을 이용하여 인삼 연근을 바로 판별할 수도 있다. 상기 대사체들은 표 28에 기재된 대사체들인 것이 바람직하다.

[0040] 상술한 바와 같이, 본 발명은 과학적으로 정확하고 신속하며 인삼의 손실을 최소화하면서 인삼 연근을 판별하기 위해 LC/MS 또는 GC/MS를 이용해 최적의 분석 조건 하에서 인삼 주근 및 세근의 대사체를 분석할 수 있었다. 그 결과 분석된 데이터를 바탕으로 통계학적 분석을 통해 연근 판별을 시도하여 1) LC/MS을 이용하여 주근의 경우 1~3년근은 뚜렷이 판별되었고, 4~6년근의 경우 특징 선택 (feature selection) 과정을 거친 후에 연근을 정확하게 판별할 수 있었고, 2) LC/MS을 이용하여 세근의 경우 특징 선택 (feature selection) 과정 없이도 4~6년근을 정확하게 판별할 수 있었고, 3) GC/MS을 이용하여 주근의 경우 특징 선택 (feature selection) 과정을 거친 후에 4~6년근 판별할 수 있었고, 4) GC/MS을 이용하여 세근의 경우 특징 선택 (feature selection) 과정을 거친 후에 4~6년근을 정확하게 판별할 수 있었다.

발명의 효과

[0041] 상술한 바와 같이, 본 발명은 LC/MS 또는 GC/MS 대사체학 분석기법을 이용하여 인삼의 대사체를 분석함으로써 인삼의 연근을 정확하고 정밀하게 판별할 수 있는 기술을 개발하여 인삼의 연근 관리에 객관적인 검증 시스템을 마련함으로써 인삼 유통에 대한 신뢰성을 향상시킬 수 있는 효과를 도모할 수 있다.

[0042] 또한, 본 발명의 인삼 연근 판별 방법은 세근을 이용할 수 있으므로 인삼의 손실을 최소화할 수 있고 극소량으로 인삼 연근을 정확하게 판별할 수 있다.

[0043] 나아가, 본 발명의 인삼 연근 판별 방법은 LC/MS 또는 GC/MS 대사체학 분석기법을 이용하여 2시간 정도의 단시

간 내에 인삼 연근을 정확하게 판별할 수 있어 시장에서 활용도가 높다 할 것이다.

도면의 간단한 설명

[0044] 도 1은 본 발명에 이용된 시료용 인삼 주근 및 세근을 나타낸다.

도 2는 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법을 개략적으로 나타내는 개략적 순서도이다.

도 3은 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 LC/MS 분석 조건을 나타낸다.

도 4는 본 발명의 인삼 연근 판별 방법에 의한 GC/MS 분석용 시료 준비 단계를 나타내는 개략적 순서도이다.

도 5는 본 발명의 인삼 연근 판별 방법에 의한 GC/MS 분석 조건을 나타낸다.

도 6은 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 인삼의 주근(a) 및 세근(b)의 LC/MS 분석 전체 이온 크로마토그램을 나타낸다.

도 7은 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 1-6년근의 전체 대사체(metabolites)에 대한 PCA 플롯(PCA plot)을 나타낸다.

도 8은 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 1-6년근의 전체 대사체(metabolites)에 대한 HCA 플롯(HCA plot)을 나타낸다.

도 9는 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 4-6년근의 전체 대사체(metabolites)에 대한 PCA 플롯(PCA plot)을 나타낸다.

도 10은 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 4-6년근의 전체 대사체(metabolites)에 대한 HCA 플롯(HCA plot)을 나타낸다.

도 11은 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 4-6년근의 RF, PAM, PLSDA 방법에 의해 선정된 대사체(metabolites)에 대한 2D(a) 및 3D(b) PCA 플롯(PCA plot)을 나타낸다.

도 12는 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 4-6년근의 RF, PAM, PLSDA 방법에 의해 선정된 대사체(metabolites)에 대한 HCA 플롯(HCA plot)을 나타낸다.

도 13은 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 세근 4-6년근의 전체 대사체에 대한 2D(a) 및 3D(b) PCA 플롯(PCA plot)을 나타낸다.

도 14는 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 세근 4-6년근의 전체 대사체에 대한 HCA 플롯(HCA plot)을 나타낸다.

도 15는 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 세근 4-6년근의 RF, PAM, PLSDA 방법에 의해 선정된 대사체(metabolites)에 대한 2D(a) 및 3D(b) PCA 플롯(PCA plot)을 나타낸다.

도 16은 본 발명의 LC/MS를 이용한 인삼 연근 판별 방법에 의한 세근 4-6년근의 RF, PAM, PLSDA 방법에 의해 선정된 대사체(metabolites)에 대한 HCA 플롯(HCA plot)을 나타낸다.

도 17은 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 인삼의 주근(a) 및 세근(b)의 GC/MS 분석 전체 이온 크로마토그램을 나타낸다.

도 18은 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 1-6년근의 전체 대사체(metabolites)에 대한 PCA 플롯(PCA plot)을 나타낸다.

도 19는 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 1-6년근의 전체 대사체(metabolites)에 대한 HCA 플롯(HCA plot)을 나타낸다.

도 20은 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 4-6년근의 전체 대사체(metabolites)에 대한 PCA 플롯(PCA plot)을 나타낸다.

도 21은 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 4-6년근의 전체 대사체(metabolites)에 대한 HCA 플롯(HCA plot)을 나타낸다.

도 22는 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 4-6년근의 RF, PAM, PLSDA 방법에 의해 선정된 대사체(metabolites)에 대한 2D(a) 및 3D(b) PCA 플롯(PCA plot)을 나타낸다.

도 23은 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 주근 4-6년근의 RF, PAM, PLSDA 방법에 의해 선정된 대사체(metabolites)에 대한 HCA 플롯(HCA plot)을 나타낸다.

도 24는 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 세근 4-6년근의 전체 대사체에 대한 2D(a) 및 3D(b) PCA 플롯(PCA plot)을 나타낸다.

도 25는 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 세근 4-6년근의 전체 대사체에 대한 HCA 플롯(HCA plot)을 나타낸다.

도 26은 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 세근 4-6년근의 RF, PAM, PLSDA 방법에 의해 선정된 대사체(metabolites)에 대한 2D(a) 및 3D(b) PCA 플롯(PCA plot)을 나타낸다.

도 27은 본 발명의 GC/MS를 이용한 인삼 연근 판별 방법에 의한 세근 4-6년근의 RF, PAM, PLSDA 방법에 의해 선정된 대사체(metabolites)에 대한 HCA 플롯(HCA plot)을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0045] 이하, 실시예를 통해 본 발명을 상세히 설명한다. 그러나 하기 실시예들은 본 발명을 설명하기 위한 것으로, 본 발명의 권리범위가 하기 실시예들에 의하여 한정되는 것은 아니다.

[0046] **실시예 1. 인삼 시료의 준비(도 1 참조)**

[0047] 2007년 01월 12일 대한민국 수원 농업진흥청 포장에서 천풍산 고려인삼 1~6 년근 주근 각 10 개체와 천풍산 고려인삼 4~6 년근 세근 각 10 개체를 각각 채집하였다.

[0048] **실시예 2. LC/MS 분석용 시료의 준비**

[0049] LC/MS는 UPLC/Q-ToF MS를 이용하여 실시하였다. LC/MS 분석용 시료 준비, 조건 및 통계분석은 다음과 같았다.

[0050] 1) LC/MS 분석용 시료 준비 (도 2 참조)

[0051] LC/MS를 이용한 대사체 분석을 위한 최적으로 추출법으로 70% aqueous MeOH를 이용하여 추출하였다. LC/MS 분석을 위한 적정량의 시료를 결정하기 기존 문헌을 참고하여 추출량, 분석농도, 주입량 (injection volume) 등을 확립하였다. 채집 후 바로 동결 건조한 후 분쇄하여 50㎎을 70% aqueous MeOH 500㎕로 20분간 초음파 처리 (sonication)한 후 2,000rpm에서 10분간 원심분리 (centrifugation)를 하였다. 상등액을 0.2㎛ GHP 멤브레인 필터 (membrane filter)를 이용해 여과 (filtration)를 하고, 최종 농도 2㎎/㎖이 되도록 희석하였다.

[0052] 2) LC/MS 조건 확립 (도 3 참조)

[0053] (1) UPLC

[0054] Waters ACQUITY UPLC™ system (Waters Corp., MA, USA)을 이용해 수행하였다. ACQUITY UPLC BEH C18 (2.1 x 100mm, 1.7) 컬럼 (column)을 사용하고 이동상으로는 0.1% 포름산 (formic acid)이 첨가된 물 (A)과 0.1% 포름산 (formic acid)이 첨가된 아세토니트릴 (acetonitrile; B)을 사용하였다. 총 분석시간은 12분으로 B의 비율을 처음 0.5분간 10%로 유지하고, 2.5분까지 30%, 6분까지 60%, 9분까지 90%로 올려주었고 10.5분까지 100%로 유지하였으며 12분까지 초기 조건인 10%로 흘려 초기 조건으로 안정화 시켰다. 유속은 500㎕/min, 주입량은 2㎕, 컬럼 온도는 35℃를 유지하였다.

[0055] (2) Q-ToF MS

[0056] Q-TOF Micro mass detector (Waters, Manchester, UK)를 사용해 네거티브 (negative) 와 포지티브 모드 (positive mode)에서 각각 인삼 대사체 분석을 위한 최적의 조건을 확립하였다. 네거티브 모드 (Negative mode)에서의 조건은 다음과 같다; capillary voltage 2800, cone voltage 30, collision energy 6, desolvation temperature 300℃, source temperature 120℃.

[0057] 3) 통계분석

[0058] LC/MS 로 분석한 원 데이터와 특징 선택 (feature selection) 과정을 거친 데이터는 MarkerLynx XS (Waters, Manchester, UK)와 R version 2.6.1 (R Foundation for Statistical Computing, Vienna, Austria)을 이용해 주성분분석 (Principal Component Analysis, PCA)과 계층적군집분석 (Hierarchical Cluster Analysis, HCA) 을 실시하여 연근판별을 시도하였다.

[0059] 특징 선택 (Feature selection)은 변수 선택 (variable selection)이라고도 하는데, 분류 (classification) 방법을 사용할 때 보다 나은 결과를 얻기 위해 변수 (대사체) 를 선택하는 과정이다. 즉, 인삼의 연근을 판별하는데 있어 별 영향을 주지 못하는 대사체는 제거하고, 의미 있는 영향력을 주는 대사체를 선택하여 분석을 하였다.

[0060] 본 발명에서는 RF (Random Forest), PAM (Prediction Analysis for Microarray), PLSDA (Partial Least Square-Discriminant Analysis) 등 세 가지의 특징 선택 (feature selection) 방법들을 사용하였고, 각 특징 선택 (feature selection) 방법들은 각 방법의 고유한 알고리즘에 따라 대사체들에 중요도 점수를 부여하여 대사체를 선택하였다. RF는 Importance Score를 이용하여 대사체의 순위를 평가하는 것으로, 각 대사체마다 자신의 importance라는 값을 갖게 되고 importance값은 그 대사체가 인삼의 연근을 판별하는 정확도에 얼마나 큰 영향을 주는지를 수치화하였다. PAM은 class centroid와 overall centroid와의 차이를 계산하여 대사체의 순위를 평가하는 것으로, 각 대사체마다 연근별 평균과 전체 평균을 계산하고 이들의 차이가 클수록 큰 점수를 주었다. 예를 들어, 1번 대사체의 3년근에서의 평균, 4년근에서의 평균, 5년근에서의 평균, 6년근에서의 평균을 계산하고 전체 샘플에서의 평균을 계산한 다음, 연근별 평균과 전체 샘플과의 차이를 계산하고 이 차이가 클수록 영향력 있는 대사체라고 간주하였다. PLSDA는 회귀계수의 가중합을 이용하여 대사체의 순위를 평가하는 것으로, PLS 회귀분석을 하면 각 대사체 마다 회귀계수가 나오게 되고, 이 회귀계수의 절대값이 클수록 중요한 대사체라고 할 수 있다. 여기서 회귀계수는 대사체가 그룹을 판별하는데 미치는 영향력을 수치화 한 것이다.

[0061] **실시예 3. GC/MS 분석용 시료의 준비**

[0062] GC/MS는 GC/MSD를 이용하여 실시하였다. GC/MS 분석용 시료 준비 및 조건은 다음과 같았다.

[0063] 1) GC/MS 분석용 시료 준비 (도 4)

[0064] GC/MS를 이용한 대사체 분석을 위한 최적의 추출법으로 $CHCl_3$: MeOH (1:1) 를 이용하여 추출하였다. GC/MS 분석을 위한 적정량의 시료를 결정하기 위해 기존 문헌을 참고하여 추출량, 분석농도, 주입량 (injection volume) 등을 확립하였고 각각 5반복 실험 샘플 (sample)을 준비하였다. 먼저 채집 후 바로 동결 건조한 뒤 -80℃에 보관한 분말 (powder) 상태의 샘플 (sample) 10mg을 $CHCl_3$: MeOH (1:1) 1ml 로 40분간 초음파처리 (sonication) 하여 추출하고, 10,000 rpm에서 5분간 원심분리하여 상등액 200μl를 농축하고 200μl의 BSTFA를 첨가하여 70℃의 수조 (water bath) 에서 40분 동안 가열하여 시릴레이션 (silylation)을 진행하였다.

[0065] 2) GC/MS 조건 확립 (도 5)

[0066] 샘플이 주입된 후 5분간은 용매 (solvent)에 의한 기기의 무리를 감소시키기 위하여 매스 (mass)값을 검출 (detection)하지 않았다. 오븐 (oven)의 온도는 70℃에서 5분간 유지하다가 10℃/min 씩 280℃까지 올린 후 280℃에서 300℃까지 20℃/min씩 올린다. ion source temperature는 200℃이고 주입량 (injection volume) 1㎕를 split ratio 20:1로 희석하여 주입하며 분자량 측정 범위 (mass detection range)는 *m/z* 50-550까지 설정하였다. 분석된 전체 이온 크로마토그램 (ion chromatogram)에서 23분부터 24분 50초의 1분 50초간 분석되는 데이터 (data)의 경우 주성분 (major compound)으로 다른 피크 (peak) 들에 비해 상대적으로 매우 높은 abundance 값을 나타내기 때문에 이외의 화합물 (compound)들의 abundance값이 너무 적게 표현된다. 따라서 연근별 상이성을 밝혀내고자 하는 이 실험에서는 과량의 공통 주성분 (major compound)이 검출 (detection)되는 시간인 23분부터 24분 50초 사이의 실험값을 제외함으로써 미량 성분들의 검출 (detection) 비율을 높일 수 있음을 확인하였다.

[0067] 3) 통계분석

[0068] LC/MS 분석과 마찬가지로 GC/MS로 분석한 원 데이터는 Auto Mass Spectral Deconvolution & Identification System와 Spectconnect (http// spectconnect.mit.edu/)에 의해 deconvolution, assignment 되고 특징 선택

(feature selection) 과정 (RF, PAM, PLSDA)을 거쳐 MarkerLynx XS (Waters, Manchester, UK)와 R version 2.6.1 (R Foundation for Statistical Computing, Vienna, Austria) 을 이용해 주성분분석 (Principal Component Analysis, PCA)과 계층적군집분석 (Hierarchical Cluster Analysis, HCA) 을 실시하여 연근판별을 시도하였다.

[0069] **실시예 4. LC/MS의 데이터를 이용하여 인삼 연근 판별 분석**

[0070] 1) LC/MS 분석 결과

[0071] 인삼 주근과 세근 시료를 개체별로 LC/MS 분석 결과를 검토해 본 결과 육안으로 각 연근별 차이를 판별하기는 어려움이 있었다 (도 6).

[0072] 2) LC/MS 데이터의 통계분석 (chemometric analysis)

[0073] 인삼 주근과 세근 시료의 LC/MS 분석 대사체를 개체별로 분석하여 PCA와 HCA 방법으로 분석하였다. PCA는 다변량 통계 분석방법 중 가장 많이 쓰이는 unsupervised clustering method의 한 방법으로, 실험 그룹 간의 차이를 확인할 수 있고, HCA는 개체를 유사한 성격을 가지는 몇 개의 그룹으로 집단화하고 각 그룹의 특성을 파악하여 그룹들 사이의 관계를 분석할 수 있다.

[0074] (1) 인삼 주근의 대사체 LC/MS 데이터 분석

[0075] 인삼 1년근에서 6년근까지 각각 주근의 대사체 분석 데이터를 이용해 PCA와 HCA를 실시한 결과, 1~3년근은 LC/MS 분석 원 데이터만으로도 각각 구분이 잘 되었다(도 7 및 8)

[0076] 인삼 1년근에서 6년근까지 각각 주근의 대사체 분석 데이터를 이용해 각 연근의 분류 정확도를 RF, PAM, PLS-DA 등 세 가지의 특징 선택 (feature selection) 방법들을 사용하여 분석하였으며 대략 90% 이상의 확률로 연근을 분류할 수 있었다 (표 1-4). 표 1은 인삼 주근의 RF, PAM, PLS-DA 방법에 의한 주근 1-6 연근 분류 정확도를 나타내고, 표 2는 RF 분류방법의 주근 1-6 연근별 confusion table이고, 표 3은 PAM 분류방법의 주근 1-6 연근별 confusion table이고, 표 4는 PLS-DA 분류방법의 주근 1-6 연근별 confusion table이다.

표 1

[0077]

분류방법	선택된 대사체 개수	CV 정확도 (개체수=59)						
		1 년근	2 년근	3 년근	4 년근	5 년근	6 년근	전체평균
RF	119	1.000	1.000	1.000	0.998	0.900	0.948	0.974
PAM	1146	1.000	0.900	1.000	0.900	0.796	0.966	0.926
PLSDA	198	1.000	1.000	1.000	1.000	0.996	1.000	0.999

표 2

[0078]

		Predicted Class						Prediction Accuracy
		1Y	2Y	3Y	4Y	5Y	6Y	
True Class	1Y	450	0	0	0	0	0	1.000
	2Y	0	500	0	0	0	0	1.000
	3Y	0	0	500	0	0	0	1.000
	4Y	0	0	0	499	1	0	0.998
	5Y	0	0	0	9	450	41	0.900
	6Y	0	0	0	0	26	474	0.948

표 3

		Predicted Class						Prediction Accuracy
		1Y	2Y	3Y	4Y	5Y	6Y	
True Class	1Y	450	0	0	0	0	0	1.000
	2Y	50	450	0	0	0	0	0.900
	3Y	0	0	500	0	0	0	1.000
	4Y	0	0	0	450	50	0	0.900
	5Y	0	0	0	0	398	102	0.796
	6Y	0	0	0	0	17	483	0.966

[0079]

표 4

		Predicted Class						Prediction Accuracy
		1Y	2Y	3Y	4Y	5Y	6Y	
True Class	1Y	450	0	0	0	0	0	1.000
	2Y	0	500	0	0	0	0	1.000
	3Y	0	0	500	0	0	0	1.000
	4Y	0	0	0	500	0	0	1.000
	5Y	0	0	2	0	498	0	0.996
	6Y	0	0	0	0	0	500	1.000

[0080]

[0081] 인삼 연근 판별의 주 관심 부분인 4~6년근 만의 데이터(data)를 선별하여 위에서 수행한 통계분석을 실시하였으나, 전체 대사체 데이터로 완벽하게 4~6년근을 구분하기에는 다소 어려움이 있었다 (도 9 및 도 10 참조).

[0082] 그러나, 인삼 4~6년근 각각 주근의 대사체 분석 데이터를 이용해 각 연근의 분류 정확도를 RF, PAM, PLS-DA 등 세 가지의 특징 선택 (feature selection) 방법들을 사용하여 분석한 결과 PLS-DA 방법으로 4~6년근을 정확히 분류할 수 있었다 (표 5-8). 표 5는 인삼 주근의 RF, PAM, PLS-DA 방법에 의한 주근 4-6 연근 분류 정확도를 나타내고, 표 6은 RF 분류방법의 주근 4~6 연근별 confusion table이고, 표 7은 PAM 분류방법의 주근 4~6 연근별 confusion table이고, 표 8은 PLS-DA 분류방법의 주근 4~6 연근별 confusion table이다.

표 5

분류방법	선택된 대사체 개수	CV 정확도 (개체수=30)			
		4 년근	5 년근	6 년근	전체평균
RF	73	0.995	0.804	0.974	0.924
PAM	725	1.000	0.999	0.879	0.959
PLSDA	605	1.000	1.000	1.000	1.000

[0083]

표 6

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	995	5	0	0.995
	5Y	125	804	71	0.804
	6Y	3	23	974	0.974

[0084]

표 7

[0085]

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	1000	0	0	1.000
	5Y	0	999	1	0.999
	6Y	0	121	879	0.879

표 8

[0086]

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	1000	0	0	1.000
	5Y	0	1000	0	1.000
	6Y	0	0	1000	1.000

[0087] 또한, 인삼 4~6년근 각각 주근의 대사체 분석 데이터 중 RF, PAM, PLSDA 방법들에서 적어도 두 가지 방법에 의해 선택된 606개의 대사체 (표 9)를 이용하여 PCA 및 HCA분석을 수행하여 4~6년근을 정확히 구분할 수 있었다 (도 11 및 도 12 참조).

표 9

[0088]

NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)
1	358	2.7	815	21	888	3.1	1223	41	789	4.0	1160	61	569	3.6	1007	81	651	4.2	1077
2	307	3.2	695	22	202	2.6	565	42	698	4.1	1113	62	1148	4.3	1434	82	379	2.8	829
3	645	4.3	1074	23	1026	4.0	1296	43	359	4.1	815	63	580	4.3	1023	83	133	3.4	491
4	496	2.6	961	24	574	4.6	1015	44	562	3.3	1005	64	265	4.1	627	84	738	4.2	1130
5	811	3.0	1167	25	417	2.7	861	45	1142	3.8	1428	65	349	4.1	805	85	50	2.6	311
6	694	4.0	1107	26	654	2.9	1077	46	443	4.3	901	66	567	2.8	1007	86	356	4.1	815
7	823	4.0	1174	27	509	4.3	971	47	944	4.4	1252	67	590	3.3	1031	87	469	3.3	945
8	815	3.0	1169	28	1068	4.2	1326	48	227	4.2	584	68	409	3.6	855	88	690	3.8	1103
9	224	3.0	583	29	1067	3.8	1326	49	154	3.5	515	69	470	4.6	945	89	551	2.8	997
10	272	3.8	642	30	210	2.6	575	50	486	5.1	957	70	160	2.2	531	90	432	3.2	885
11	772	4.2	1143	31	102	3.4	445	51	1088	3.1	1346	71	502	4.6	965	91	771	4.0	1143
12	221	4.2	582	32	717	3.0	1121	52	842	4.0	1191	72	226	3.0	584	92	19	3.0	240
13	539	3.3	987	33	849	4.2	1193	53	700	4.2	1113	73	299	4.6	679	93	526	3.0	981
14	984	3.1	1273	34	58	2.5	323	54	494	2.8	961	74	381	3.8	829	94	970	4.0	1268
15	400	4.0	845	35	384	2.9	831	55	538	4.8	987	75	907	3.9	1238	95	555	5.1	1001
16	568	2.6	1007	36	475	3.2	953	56	890	4.0	1225	76	455	3.3	927	96	458	2.7	929
17	212	4.0	576	37	812	3.0	1169	57	124	2.1	473	77	545	4.6	991	97	1038	3.8	1300
18	237	4.0	599	38	763	4.7	1139	58	186	3.0	560	78	193	4.2	561	98	544	3.0	991
19	179	4.0	553	39	175	2.8	549	59	1095	3.8	1354	79	610	3.0	1047	99	859	4.0	1201
20	782	4.0	1153	40	1040	3.8	1300	60	768	3.3	1139	80	259	3.8	619	100	647	2.8	1075

[0089]

NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)
101	865	4.2	1210	121	726	4.2	1123	141	153	4.7	514	161	445	4.0	913	181	332	4.2	783
102	155	3.3	516	122	1069	4.0	1326	142	912	3.8	1240	162	171	4.0	540	182	774	4.2	1149
103	1131	4.0	1405	123	401	3.1	845	143	999	4.4	1280	163	425	6.1	869	183	225	4.2	584
104	449	4.6	919	124	507	3.4	971	144	262	4.3	627	164	1058	4.0	1324	184	378	4.2	829
105	1096	4.0	1354	125	533	3.2	985	145	436	4.0	889	165	941	4.5	1252	185	411	3.2	859
106	177	4.0	552	126	781	4.0	1153	146	862	4.1	1209	166	1064	4.3	1326	186	845	4.1	1193
107	586	4.7	1029	127	554	5.3	1001	147	517	5.1	975	167	165	3.6	531	187	846	4.2	1193
108	291	3.8	665	128	505	4.7	969	148	184	2.6	559	168	546	3.4	993	188	390	4.2	841
109	209	2.1	575	129	327	4.0	773	149	712	4.2	1119	169	503	2.8	967	189	20	3.4	240
110	480	4.2	955	130	882	4.7	1217	150	987	4.0	1276	170	547	3.4	993	190	464	4.4	939
111	345	4.0	799	131	1034	3.1	1296	151	414	5.7	861	171	1065	4.1	1326	191	688	4.2	1101
112	342	2.8	795	132	887	3.1	1223	152	799	4.5	1163	172	122	2.4	473	192	194	4.3	561
113	82	3.0	375	133	1018	3.8	1293	153	206	3.8	573	173	48	2.7	305	193	948	4.2	1255
114	683	3.3	1099	134	515	3.8	973	154	218	3.6	577	174	1008	3.8	1286	194	382	4.2	829
115	125	3.4	473	135	519	4.5	975	155	477	3.2	953	175	1011	4.2	1288	195	329	3.4	781
116	91	2.6	405	136	116	2.1	461	156	1071	3.8	1326	176	335	4.1	789	196	352	4.3	811
117	1070	3.9	1326	137	407	2.7	851	157	203	2.4	567	177	751	4.2	1135	197	88	2.5	395
118	205	4.0	571	138	658	4.0	1081	158	374	4.7	825	178	24	4.8	240	198	693	4.0	1107
119	18	3.3	240	139	325	4.1	769	159	1100	3.8	1358	179	973	4.2	1269	199	977	4.1	1270
120	86	2.7	387	140	988	3.8	1276	160	468	3.0	945	180	324	4.1	769	200	271	4.0	642

NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)
201	399	3.2	845	221	1144	3.8	1428	241	783	4.4	1155	261	128	2.9	487	281	540	4.6	987
202	94	6.2	413	222	577	3.0	1017	242	68	2.1	341	262	1031	4.0	1296	282	514	3.3	973
203	863	4.4	1210	223	1133	4.0	1405	243	595	3.4	1031	263	1123	3.8	1394	283	976	4.2	1270
204	959	3.8	1260	224	444	4.0	913	244	251	2.3	607	264	573	5.5	1015	284	816	4.7	1170
205	1017	4.0	1293	225	387	3.1	835	245	714	4.3	1119	265	348	4.3	803	285	937	4.4	1251
206	465	4.4	941	226	364	4.0	819	246	837	4.1	1183	266	363	4.3	819	286	331	3.4	781
207	985	3.1	1273	227	892	3.3	1226	247	855	3.1	1200	267	525	4.6	981	287	40	2.7	290
208	270	3.8	637	228	344	4.0	799	248	110	3.0	457	268	341	4.6	793	288	239	2.8	601
209	1084	3.8	1342	229	350	4.1	805	249	169	4.3	538	269	362	4.0	819	289	215	3.8	577
210	518	3.8	975	230	422	4.4	868	250	664	4.7	1091	270	1124	3.8	1394	290	10	2.2	240
211	49	2.2	307	231	614	4.2	1051	251	456	2.7	929	271	81	6.0	369	291	1000	4.2	1280
212	950	4.4	1256	232	921	4.3	1246	252	1063	3.8	1326	272	334	3.3	783	292	684	4.8	1099
213	508	3.6	971	233	780	4.0	1151	253	158	2.2	523	273	59	2.7	323	293	570	3.4	1009
214	1130	4.0	1405	234	366	4.4	823	254	1036	3.1	1296	274	104	3.4	447	294	579	3.3	1023
215	758	4.2	1137	235	353	3.6	811	255	1029	4.4	1296	275	733	4.0	1127	295	1081	4.2	1340
216	260	4.4	620	236	523	2.8	977	256	1097	3.8	1354	276	1137	3.0	1419	296	983	4.2	1273
217	309	3.7	708	237	691	4.0	1105	257	848	4.1	1193	277	560	4.8	1005	297	386	4.0	835
218	398	2.5	843	238	306	3.8	693	258	241	3.0	603	278	340	5.7	793	298	1002	3.8	1282
219	706	4.7	1117	239	442	4.0	899	259	114	2.1	459	279	428	3.4	883	299	1086	3.1	1346
220	500	2.7	963	240	199	4.0	563	260	866	4.1	1210	280	310	2.8	713	300	880	4.1	1217

NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)
301	899	4.4	1230	321	643	4.7	1073	341	247	2.5	605	361	832	4.4	1179	381	666	4.7	1091
302	1152	3.8	1458	322	236	3.8	599	342	627	2.9	1060	362	1106	3.1	1369	382	375	3.8	827
303	372	6.1	825	323	993	4.1	1278	343	244	4.4	604	363	172	6.5	545	383	430	4.1	885
304	510	3.4	971	324	531	3.2	984	344	920	4.1	1246	364	867	4.0	1210	384	678	4.2	1097
305	383	4.3	829	325	410	2.9	857	345	396	3.4	841	365	238	2.8	601	385	728	2.9	1124
306	600	4.4	1041	326	36	4.6	290	346	246	3.2	605	366	501	3.0	965	386	482	4.2	955
307	990	3.8	1276	327	972	4.0	1269	347	187	2.5	561	367	594	4.6	1031	387	242	4.2	604
308	454	3.4	927	328	112	6.3	457	348	275	2.2	645	368	493	5.1	961	388	979	4.2	1270
309	127	7.0	476	329	140	4.7	493	349	178	2.8	553	369	182	2.1	557	389	856	4.1	1201
310	276	2.7	645	330	222	4.3	582	350	38	3.0	290	370	640	3.4	1072	390	485	5.2	957
311	28	3.2	244	331	318	3.6	739	351	231	2.5	595	371	711	4.5	1119	391	189	4.2	561
312	157	3.6	517	332	584	3.4	1025	352	99	2.8	443	372	214	3.8	577	392	207	4.2	574
313	113	6.6	458	333	431	4.1	885	353	132	2.1	491	373	1145	4.3	1434	393	650	4.2	1077
314	42	2.9	293	334	960	4.3	1262	354	269	3.4	637	374	216	2.8	577	394	720	4.2	1123
315	662	4.7	1084	335	457	4.0	929	355	67	2.4	341	375	301	2.5	683	395	1023	4.2	1295
316	183	3.0	559	336	773	4.2	1144	356	957	3.8	1259	376	735	3.3	1129	396	796	4.2	1163
317	829	4.4	1178	337	208	4.1	574	357	268	3.8	637	377	729	4.0	1125	397	346	4.4	803
318	43	3.0	293	338	759	4.5	1137	358	852	4.4	1195	378	111	2.1	457	398	929	4.3	1249
319	484	4.4	955	339	653	4.3	1077	359	689	3.8	1103	379	625	4.6	1059	399	857	4.0	1201
320	441	4.3	897	340	596	3.3	1031	360	235	4.3	597	380	191	4.4	561	400	377	4.3	829

NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)
401	365	4.4	823	421	806	4.1	1167	441	1059	3.8	1324	461	788	4.0	1159	481	139	3.4	493
402	479	4.3	955	422	248	3.8	606	442	404	4.0	845	462	593	4.9	1031	482	565	2.8	1007
403	351	3.6	811	423	672	2.6	1093	443	791	4.4	1161	463	274	2.3	643	483	746	4.0	1131
404	450	4.4	925	424	419	4.0	862	444	343	2.8	795	464	646	4.3	1074	484	451	4.8	925
405	889	4.4	1223	425	151	3.2	513	445	913	4.0	1240	465	536	3.3	985	485	879	4.0	1217
406	566	3.6	1007	426	285	2.6	653	446	619	4.3	1052	466	1043	3.8	1300	486	840	4.4	1187
407	245	4.2	604	427	1098	3.7	1356	447	1077	3.2	1329	467	1116	3.8	1379	487	174	4.2	548
408	367	3.9	823	428	16	2.8	240	448	736	4.2	1130	468	1007	4.0	1286	488	481	4.4	955
409	996	4.3	1279	429	217	2.8	577	449	142	2.3	497	469	991	4.0	1276	489	452	4.4	925
410	1103	4.2	1364	430	321	2.8	740	450	657	4.2	1078	470	173	2.7	547	490	553	5.1	1001
411	675	4.0	1094	431	333	4.2	783	451	373	4.4	825	471	330	4.7	781	491	648	2.8	1075
412	919	4.1	1246	432	1140	3.8	1426	452	730	4.2	1125	472	472	4.0	947	492	1110	3.7	1372
413	975	4.4	1269	433	743	3.3	1131	453	659	4.0	1081	473	429	3.6	883	493	300	4.4	679
414	810	3.4	1167	434	549	2.8	997	454	1004	4.4	1282	474	44	8.1	293	494	699	4.3	1113
415	161	3.6	531	435	1061	3.8	1324	455	1030	4.2	1296	475	1102	4.2	1364	495	93	3.0	409
416	989	4.0	1276	436	513	3.0	973	456	314	2.8	723	476	719	3.0	1122	496	1075	3.2	1329
417	121	3.3	472	437	677	4.0	1095	457	830	4.3	1178	477	326	4.0	773	497	200	7.2	564
418	96	2.2	415	438	146	2.9	503	458	296	2.8	677	478	836	4.2	1183	498	971	4.0	1268
419	911	4.0	1240	439	292	3.8	670	459	273	3.9	642	479	1037	4.5	1298	499	228	6.6	589
420	511	3.3	973	440	1012	4.4	1288	460	1153	3.8	1458	480	420	4.0	862	500	1089	4.3	1348

NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)	NO	Meta #	Retention Time (min.)	Ion (m/z)
501	637	3.4	1070	521	506	3.6	971	541	11	2.7	240	561	949	4.0	1256	581	850	4.3	1194	601	940	4.4	1252
502	615	4.3	1051	522	280	3.0	649	542	1039	4.0	1300	562	196	2.6	563	582	833	4.3	1180	602	466	4.0	944
503	250	3.8	606	523	281	2.8	649	543	824	4.0	1174	563	903	4.5	1232	583	286	2.5	653	603	433	3.4	885
504	328	2.8	775	524	557	3.0	1003	544	1083	3.8	1342	564	1003	3.8	1282	584	724	4.3	1128	604	80	2.3	367
505	740	4.0	1130	525	578	4.3	1023	545	843	4.0	1192	565	777	4.2	1149	585	747	3.3	1133	605	76	3.0	341
506	1006	3.9	1284	526	258	4.0	619	546	864	4.0	1209	566	85	3.8	385	586	62	5.7	325	606	872	3.8	1213
507	966	4.3	1263	527	1028	4.5	1296	547	7	3.0	235	567	942	4.2	1252	587	416	3.8	861				
508	15	2.9	240	528	550	4.4	997	548	787	3.0	1157	568	906	4.5	1236	588	616	2.9	1051				
509	873	4.3	1216	529	798	4.2	1163	549	639	3.4	1071	569	289	4.4	653	589	1051	3.6	1316				
510	572	2.2	1009	530	393	4.1	841	550	755	4.4	1137	570	623	4.3	1059	590	542	4.3	991				
511	1055	3.8	1318	531	1009	4.0	1286	551	958	3.8	1260	571	878	4.0	1216	591	1046	3.8	1303				
512	670	4.5	1093	532	697	2.8	1109	552	875	4.4	1216	572	143	4.7	501	592	12	2.1	240				
513	220	2.7	579	533	75	2.2	341	553	1080	3.8	1338	573	1005	3.8	1282	593	655	2.8	1077				
514	1114	3.8	1379	534	277	2.2	645	554	512	3.8	973	574	391	3.2	841	594	130	2.2	489				
515	760	4.4	1137	535	708	3.4	1117	555	1118	4.0	1388	575	311	4.4	723	595	607	3.8	1047				
516	211	4.0	575	536	968	4.5	1265	556	461	2.8	931	576	752	4.2	1135	596	576	3.0	1017				
517	814	4.4	1169	537	1060	4.2	1324	557	267	3.8	629	577	339	2.8	791	597	162	2.8	531				
518	757	4.2	1137	538	219	3.0	577	558	483	4.3	955	578	704	3.4	1117	598	371	3.4	823				
519	820	4.7	1171	539	894	4.4	1226	559	108	3.0	455	579	355	3.2	815	599	195	3.4	561				
520	946	2.8	1254	540	201	7.0	564	560	45	6.6	297	580	790	4.0	1160	600	370	3.9	823				

[0090] (2) 인삼 세근의 대사체 LC/MS 데이터 분석

[0091] 이에 반해 세근의 경우 LC/MS 분석하여 얻어진 전체 대사체 (total metabolites) 를 이용하여 PCA 및 HCA를 실시하였을 때 4~6년근 각각의 시료가 명확히 분리되었다 (도 13 및 도 14 참조).

[0092] 역시, 인삼 4~6년근 각각 세근의 대사체 분석 데이터를 이용해 각 연근의 분류 정확도를 RF, PAM, PLS-DA등 세 가지의 특징 선택 (feature selection) 방법들을 사용하여 분석한 결과 사용한 모든 방법으로 4~6년근을 정확히 분류되었다 (표 10-13). 표 10은 인삼 세근의 RF, PAM, PLS-DA 방법에 의한 연근 분류 정확도를 나타내고, 표 11은 RF 분류방법의 세근 4-6 연근별 confusion table이고, 표 12는 PAM 분류방법의 세근 4-6 연근별 confusion table이고, 표 13은 PLS-DA 분류방법의 세근 4-6 연근별 confusion table이다.

표 10

[0093]

분류방법	선택된 대사체 개수	CV 정확도 (개체수=30)			
		4 년근	5 년근	6 년근	전체평균
RF	8	1.000	1.000	1.000	1.000
PAM	11	1.000	1.000	1.000	1.000
PLSDA	16	1.000	1.000	1.000	1.000

표 11

[0094]

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	500	0	0	1.000
	5Y	0	500	0	1.000
	6Y	0	0	500	1.000

표 12

[0095]

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	500	0	0	1.000
	5Y	0	500	0	1.000
	6Y	0	0	500	1.000

표 13

[0096]

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	500	0	0	1.000
	5Y	0	500	0	1.000
	6Y	0	0	500	1.000

[0097] 또한, 인삼 세근 4~6년근의 전체 대사체 (total metabolite) 중 RF, PAM, PLSDA 방법들에서 적어도 두 가지 방법에 의해 선택된 9개의 대사체 (표 14)를 이용하여 PCA 및 HCA 분석을 수행한 결과 아주 유의성 있게 4~6년근을 구분할 수 있었다 (도 15 및 도 16 참조).

표 14

[0098]

NO	Meta #	Retention Time (min.)	Ion (*m/z*)
1	152	4.1	1105
2	244	3.2	1436
3	121	4.4	971
4	35	2.2	499
5	108	3.6	883
6	122	4.4	971
7	82	4.2	841
8	164	4.3	1143
9	109	2.9	861

[0099] **실시예 5. GC/MS의 데이터를 이용하여 인삼 연근 판별 분석**

[0100] 1) GC/MS 분석 결과

[0101] 인삼 주근과 세근 시료를 개체별로 GC/MS 분석 결과를 검토해 본 결과 육안으로 각 연근별 차이를 판별하기는 어려움이 있었다 (도 17).

[0102] 2) GC/MS 데이터의 통계분석 (chemometric analysis)

[0103] 인삼 주근과 세근 시료의 GC/MS 분석 대사체를 개체별로 분석하여 PCA와 HCA 방법으로 분석하였다. PCA는 다변량 통계 분석방법 중 가장 많이 쓰이는 unsupervised clustering method의 한 방법으로, 실험 그룹 간의 차이를 확인할 수 있고, HCA는 개체를 유사한 성격을 가지는 몇 개의 그룹으로 집단화하고 각 그룹의 특성을 파악하여 그룹들 사이의 관계를 분석할 수 있다.

[0104] (1) 인삼 주근의 대사체 GC/MS 데이터 분석

[0105] 인삼 1년근에서 6년근까지 각각 주근의 대사체 분석 데이터를 이용해 PCA와 HCA를 실시한 결과, 1년근과 5년근은 GC/MS 분석 원 데이터만으로도 각각 구분이 잘 되었다 (도 18 및 도 19).

[0106] 인삼 1년근에서 6년근까지 각각 주근의 대사체 분석 데이터를 이용해 각 연근의 분류 정확도를 RF, PAM, PLS-DA 등 세 가지의 특징 선택 (feature selection) 방법들을 사용하여 분석하였으며 대략 80% 이상의 확률로 연근을 분류할 수 있었다 (표 15-18). 표 15는 인삼 주근의 RF, PAM, PLS-DA 방법에 의한 주근 1-6 연근 분류 정확도를 나타내고, 표 16은 RF 분류방법의 주근 1-6 연근별 confusion table이고, 표 17은 PAM 분류방법의 주근 1-6 연근별 confusion table이고, 표 18은 PLS-DA 분류방법의 주근 1-6 연근별 confusion table이다.

표 15

[0107]

분류방법	선택된 대사체 개수	CV 정확도 (개체수=30)						
		1 년근	2 년근	3 년근	4 년근	5 년근	6 년근	전체평균
RF	50	1.000	0.972	0.786	0.690	1.000	0.736	0.864
PAM	185	1.000	0.970	0.790	0.992	1.000	0.796	0.925
PLSDA	37	1.000	1.000	0.920	0.996	1.000	1.000	0.986

표 16

[0108]

		Predicted Class						Classification Accuracy
		1Y	2Y	3Y	4Y	5Y	6Y	
True Class	1Y	500	0	0	0	0	0	1.000
	2Y	0	486	8	6	0	0	0.972
	3Y	0	107	393	0	0	0	0.786
	4Y	0	0	0	345	0	155	0.690
	5Y	0	0	0	0	500	0	1.000
	6Y	0	0	1	121	10	368	0.736

표 17

[0109]

		Predicted Class						Classification Accuracy
		1Y	2Y	3Y	4Y	5Y	6Y	
True Class	1Y	500	0	0	0	0	0	1.000
	2Y	0	485	15	0	0	0	0.970
	3Y	0	105	395	0	0	0	0.790
	4Y	0	0	0	496	0	4	0.992
	5Y	0	0	0	0	500	0	1.000
	6Y	0	0	0	101	1	398	0.796

표 18

		Predicted Class						Classification Accuracy
		1Y	2Y	3Y	4Y	5Y	6Y	
True Class	1Y	500	0	0	0	0	0	1.000
	2Y	0	500	0	0	0	0	1.000
	3Y	0	40	460	0	0	0	0.920
	4Y	0	0	0	498	0	2	0.996
	5Y	0	0	0	0	500	0	1.000
	6Y	0	0	0	0	0	500	1.000

[0110]

[0111] 인삼 연근 판별의 주 관심 부분인 4~6년근만의 데이터를 선별하여 위에서 수행한 통계분석을 실시하였으나 전체 대사체의 데이터로 완벽하게 4~6년근을 구분하는 데에는 다소 어려움이 있었다 (도 20 및 도 21).

[0112] 그러나 인삼 4~6년근 각각 주근의 대사체 분석 데이터를 이용해 각 연근의 분류 정확도를 RF, PAM, PLS-DA등 세 가지의 특징 선택 (feature selection) 방법들을 사용하여 분석 한 결과 PLS-DA방법으로 4~6년근을 정확히 분류할 수 있었다 (표 19-22). 표 19는 인삼 주근의 RF, PAM, PLS-DA 방법에 의한 주근 4-6 연근 분류 정확도를 나타내고, 표 20은 RF 분류방법의 주근 4~6 연근별 confusion table이고, 표 21은 PAM 분류방법의 주근 4~6 연근별 confusion table이고, 표 22는 PLS-DA 분류방법의 주근 4~6 연근별 confusion table이다.

표 19

분류방법	선택된 대사체 개수	CV 정확도 (개체수=15)			
		4 년근	5 년근	6 년근	전체평균
RF	25	1.000	1.000	0.800	0.933
PAM	150	0.976	1.000	0.796	0.924
PLSDA	31	1.000	1.000	1.000	1.000

[0113]

표 20

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	250	0	0	1.000
	5Y	0	250	0	1.000
	6Y	50	0	200	0.800

[0114]

표 21

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	244	0	6	0.976
	5Y	0	250	0	1.000
	6Y	51	0	199	0.796

[0115]

표 22

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	250	0	0	1.000
	5Y	0	250	0	1.000
	6Y	0	0	250	1.000

[0116]

[0117] 또한, 인삼 4~6년근 각각 주근의 대사체 분석 데이터 중 RF, PAM, PLSDA 방법들에서 세 가지 방법 모두에 의해 선택된 13개의 대사체 (표 23)를 이용하여 PCA 및 HCA분석을 수행하여 4~6년근을 정확히 구분할 수 있었다 (도 22 및 도 23).

표 23

NO	Meta #	Retention Time (min.)	Ion (*m/z*)	NIST Library
1	117	4.0	105	-
2	157	20.1	73	-
3	139	21.3	73	trimethylsilyl 1-trimethylsilyl-5-trimethylsiloxy-3-(2-trimethylsilylamino)indolepropionate
4	140	29.8	217	-
5	168	27.9	217	-
6	199	15.5	217	-
7	148	9.5	73	-
8	185	12.5	287	-
9	172	13.8	57	-
10	65	3.4	70	2-Thiazolidinone, 3-(1-methylethyl)-4-methyl-
11	105	21.2	75	9,12-Octadecadienoic acid (Z,Z)-, trimethylsilyl ester
12	40	13.7	73	-
13	162	14.3	71	-

[0118]

[0119] (2) 인삼 세근의 대사체 GC/MS 데이터 분석

[0120] 이에 반해 세근의 경우 GC/MS 분석하여 얻어진 전체 대사체 (total metabolites)를 이용하여 PCA 및 HCA를 실시하였으나 완벽한 4~6년근의 구분에는 다소 어려움이 있었다 (도 24 및 도 25).

[0121] 역시, 인삼 4~6년근 각각 세근의 대사체 분석 데이터를 이용해 각 연근의 분류 정확도를 RF, PAM, PLS-DA등 세 가지의 특징 선택 (feature selection) 방법들을 사용하여 분석 한 결과 PLS-DA 방법으로 4~6년근을 정확히 분류할 수 있었다 (표 24-27). 표 24는 인삼 세근의 RF, PAM, PLS-DA 방법에 의한 세근 4-6 연근 분류 정확도를 나타내고, 표 25는 RF 분류방법의 세근 4-6 연근별 confusion table이고, 표 26은 PAM 분류방법의 세근 4-6 연근별 confusion table이고, 표 27은 PLS-DA 분류방법의 세근 4-6 연근별 confusion table이다.

표 24

분류방법	선택된 대사체 개수	CV 정확도 (개체수=15)			
		4 년근	5 년근	6 년근	전체평균
RF	7	0.988	1.000	0.836	0.941
PAM	56	0.832	1.000	0.828	0.887
PLSDA	19	1.000	1.000	1.000	1.000

[0122]

표 25

[0123]

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	247	0	3	0.988
	5Y	0	250	0	1.000
	6Y	41	0	209	0.836

표 26

[0124]

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	208	0	42	0.832
	5Y	0	250	0	1.000
	6Y	43	0	207	0.828

표 27

[0125]

		Predicted Class			Classification Accuracy
		4Y	5Y	6Y	
True Class	4Y	250	0	0	1.000
	5Y	0	250	0	1.000
	6Y	0	0	250	1.000

[0126] 또한, 인삼 세근 4~6년근의 전체 대사체 (total metabolite) 중 RF, PAM, PLSDA 방법들에서 적어도 두 방법에 의해 선택된 21개의 대사체 (표 28)를 이용하여 PCA 및 HCA분석을 수행한 결과 아주 유의성 있게 4~6년근을 구분할 수 있었다 (도 26 및 도 27).

표 28

[0127]

NO	Meta #	Retention Time (min.)	Ion (*m/z*)	NIST Library
1	183	16.4	73	–
2	106	26.0	204	–
3	148	9.5	73	–
4	192	20.8	204	Urea, N,N'-bis(trimethylsilyl)-
5	60	26.5	73	–
6	14	6.2	57	–
7	4	3.4	244	–
8	64	31.8	217	αD-Glucopyranoside, 1,3,4,6-tetrakis-O-(trimethylsilyl)-βD-fructofuranosyl 2,3,4,6-tetrakis-O-(trimethylsilyl)-
9	197	9.5	147	–
10	36	11.3	147	Butanedioic acid, bis(trimethylsilyl) ester
11	187	22.4	73	dl-2-Benzylaminooctanol
12	47	19.0	149	–
13	200	16.6	71	–
14	68	7.5	57	3-Ethyl-3-methylheptane
15	9	3.7	171	Silanamine, N,N'-methanetetraylbis[1,1,1-trimethyl-
16	188	32.0	441	–
17	175	30.5	217	–
18	123	18.3	73	–
19	151	12.7	73	–
20	29	9.8	133	–
21	35	11.0	142	L-Proline, 1-(trimethylsilyl)-, trimethylsilyl ester

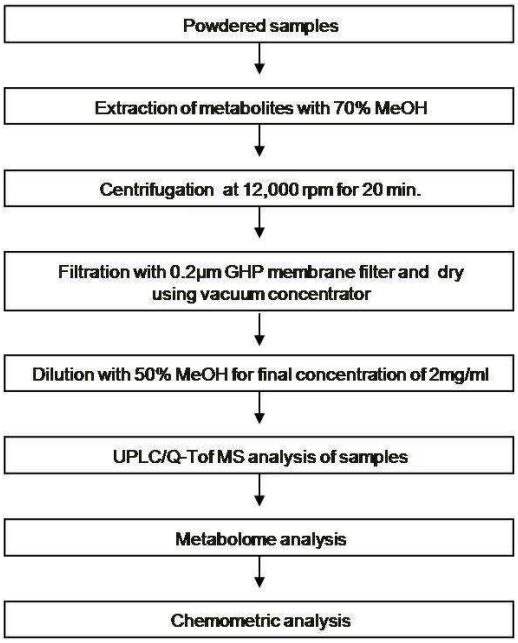
도면

도면1



연근	부위	개체수	채집 지역	채집 날짜
1	주근	10	농진청 포장	2007.01.12
2	주근	10	농진청 포장	2007.01.12
3	주근	10	농진청 포장	2007.01.12
4	주근	10	농진청 포장	2007.01.12
	세근	10	농진청 포장	2007.01.12
5	주근	10	농진청 포장	2007.01.12
	세근	10	농진청 포장	2007.01.12
6	주근	10	농진청 포장	2007.01.12
	세근	10	농진청 포장	2007.01.12

도면2



도면3

UPLC

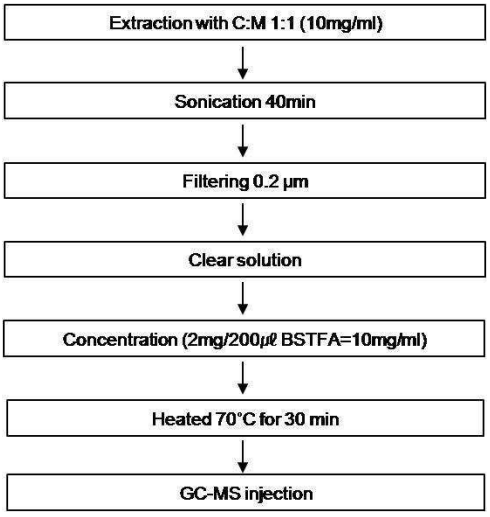
- Column: ACQUITY UPLC BEH C18 2.1 x 100mm, 1.7μm
- Mobile phase A: 0.1% FA in Water
- Mobile phase B: 0.1% FA in Acetonitrile
- Flow rate: 500μl/min
- Injection volume: 5μl
- Column temperature: 35℃
- Gradient condition:

Time (min)	Flow rate (μl/min)	% A	% B	Curve
Initial	500	90.0	10.0	
0.50	500	90.0	10.0	6
2.50	500	70.0	30.0	6
6.00	500	40.0	60.0	6
9.00	500	10.0	90.0	6
10.50	500	0.0	100.0	1
15.00	500	80.0	10.0	1

Q-Tof

• Capillary (V)	2800.0
• Sample Cone (V)	30.0
• Extraction Cone (V)	1.0
• Desolvation Temp (°C)	250.0
• Source Temp (°C)	120.0
• TOF Flight Tube (V)	5630.0
• Reflectron (V)	1780.0
• Pusher	860.4
• Puller	650.0
• Start Mass	200.0
• End Mass	1500.0
• Start Time (min)	0.0
• End Time (min)	15.0
• Scan Time (sec)	0.4
• InterScan Time (sec)	0.1

도면4



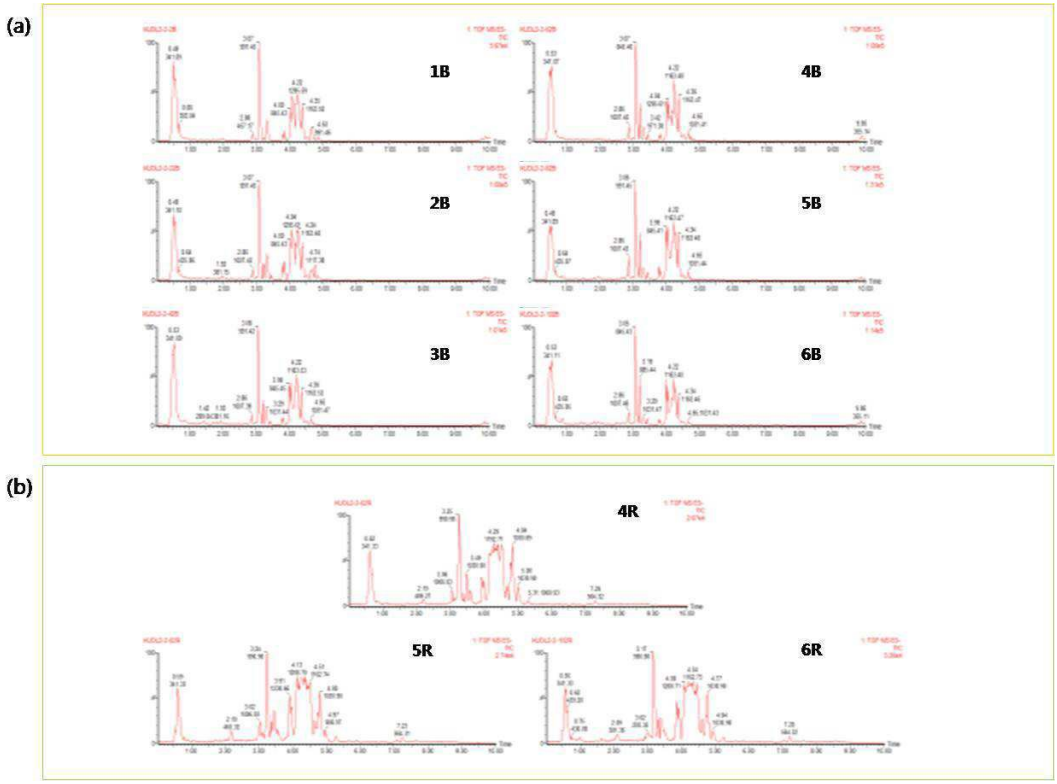
도면5

GC-MS condition

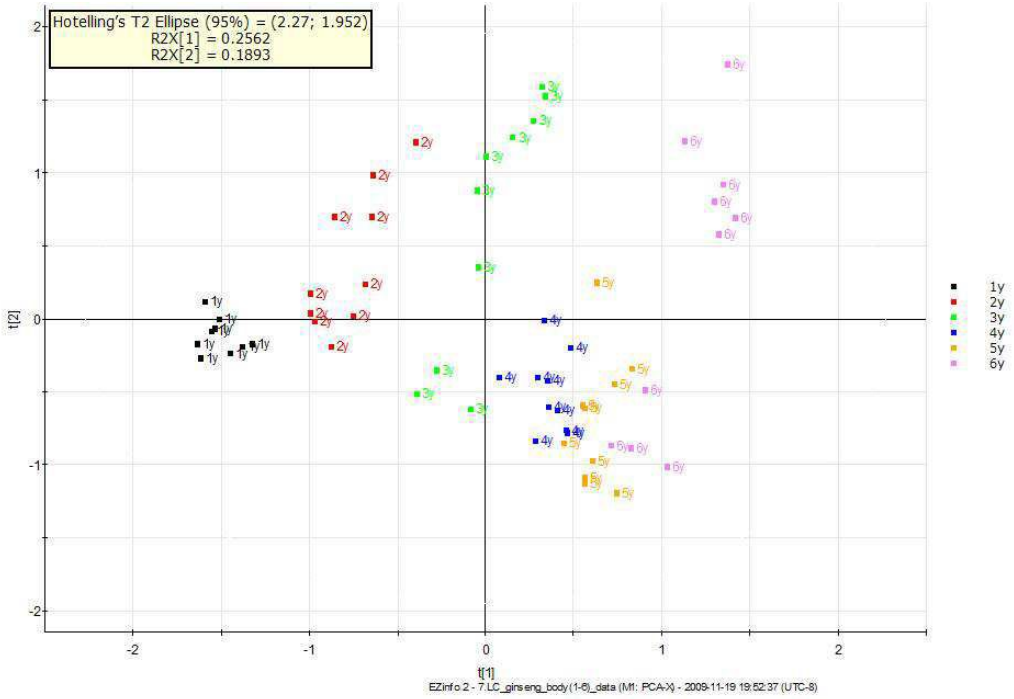
- Heater 230 ℃
- Interface 250 ℃
- MS Quad 150 ℃
- MS Source 200 ℃
- Split ratio 20:1
- Solvent delay 3min
- Mass range 50-550 a.m.u
- Post run 5min
- RUN TIME 33 min
- Time event 23-24.5 min
- DB 5-MS

Rate	Temp.	Time
-	70℃	5min
10℃/min	280℃	5min
20℃/min	300℃	-

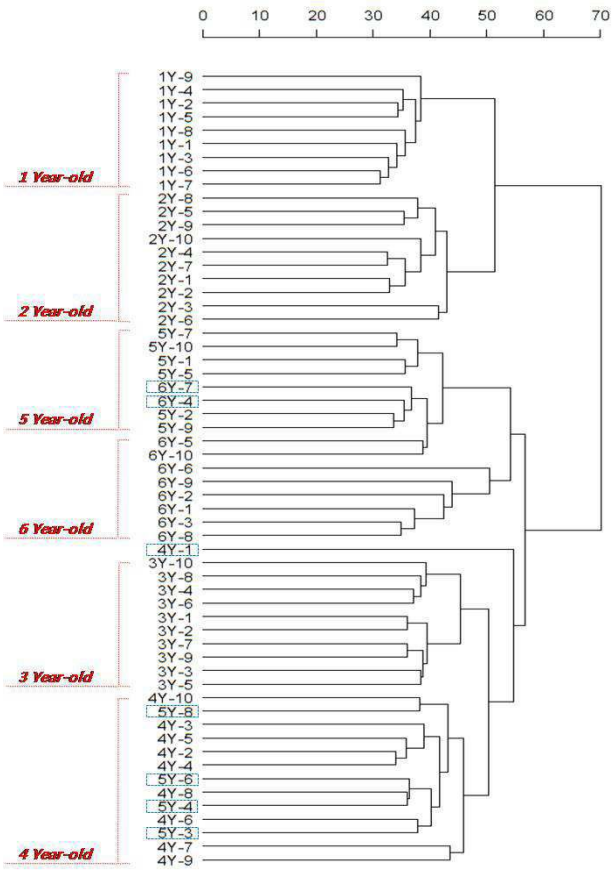
도면6



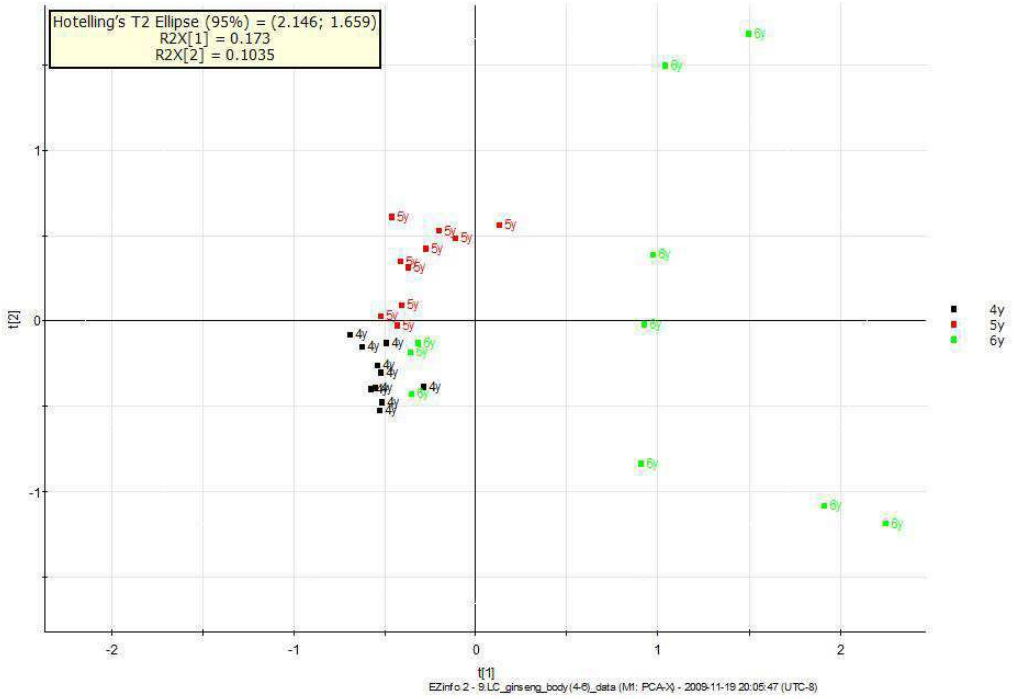
도면7


Hotelling's T2 Ellipse (95%) = (2.27; 1.952)
R2X[1] = 0.2562
R2X[2] = 0.1893
t[1]
t[2]
1y
2y
3y
4y
5y
6y
EZinfo 2 - 7.LC_ginseng_body (1-6)_data (M1: PCA-X) - 2009-11-19 19:52:37 (UTC-8)

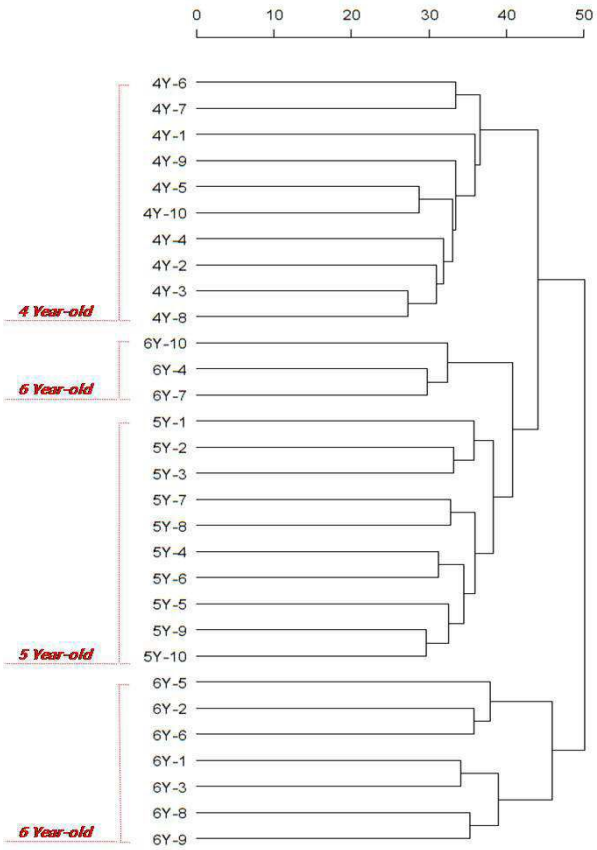
도면8


0 10 20 30 40 50 60 70
1 Year-old
2 Year-old
5 Year-old
6 Year-old
3 Year-old
4 Year-old
1Y-9
1Y-4
1Y-2
1Y-5
1Y-8
1Y-1
1Y-3
1Y-6
1Y-7
2Y-8
2Y-5
2Y-9
2Y-10
2Y-4
2Y-7
2Y-1
2Y-2
2Y-3
2Y-6
5Y-7
5Y-10
5Y-1
5Y-5
6Y-7
6Y-4
5Y-2
5Y-9
6Y-5
6Y-10
6Y-6
6Y-9
6Y-2
6Y-1
6Y-3
6Y-8
4Y-1
3Y-10
3Y-8
3Y-4
3Y-6
3Y-1
3Y-2
3Y-7
3Y-9
3Y-3
3Y-5
4Y-10
5Y-8
4Y-3
4Y-5
4Y-2
4Y-4
5Y-6
4Y-8
5Y-4
4Y-6
5Y-3
4Y-7
4Y-9

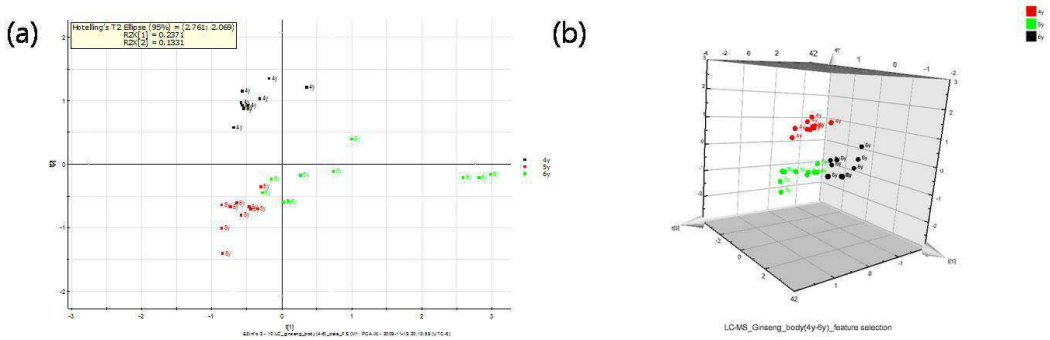
도면9


Hotelling's T2 Ellipse (95%) = (2.146; 1.659)
R2X[1] = 0.173
R2X[2] = 0.1035
t[2]
t[1]
4y
5y
6y
EZinfo 2 - 9.LC_ginseng_body (4-6)_data (M1: PCA-X) - 2009-11-19 20:05:47 (UTC-8)

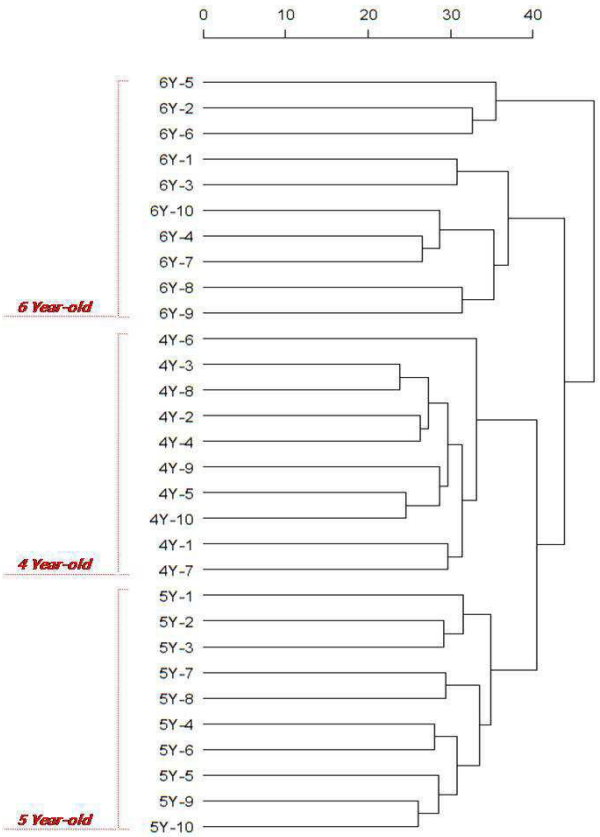
도면10


0
10
20
30
40
50
4Y-6
4Y-7
4Y-1
4Y-9
4Y-5
4Y-10
4Y-4
4Y-2
4Y-3
4Y-8
4 Year-old
6Y-10
6Y-4
6Y-7
6 Year-old
5Y-1
5Y-2
5Y-3
5Y-7
5Y-8
5Y-4
5Y-6
5Y-5
5Y-9
5Y-10
5 Year-old
6Y-5
6Y-2
6Y-6
6Y-1
6Y-3
6Y-8
6Y-9
6 Year-old

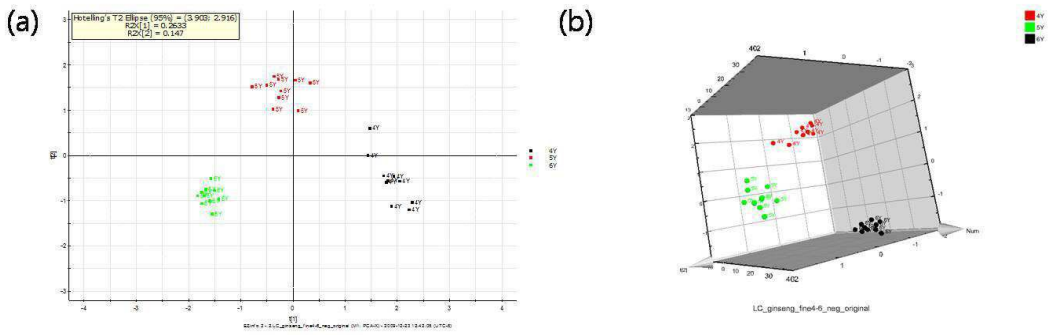
도면11



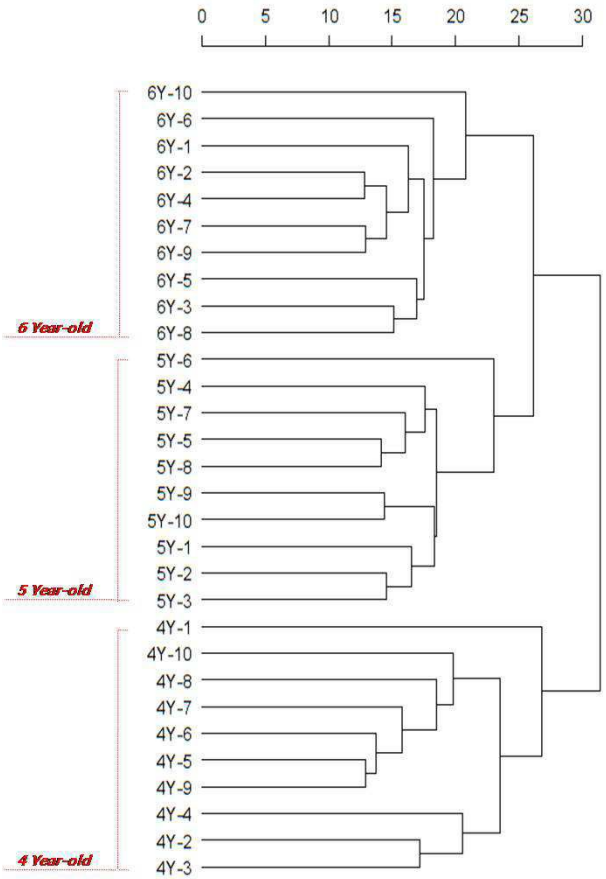
도면12



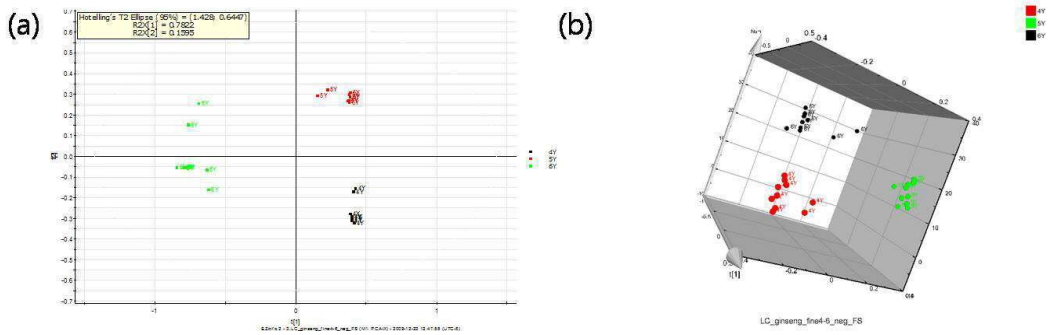
도면13



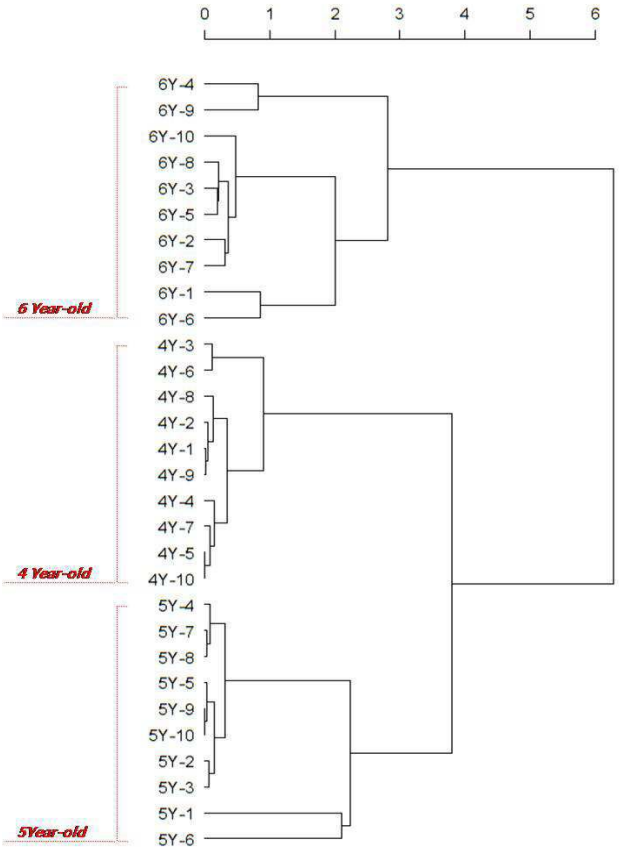
도면14


0
5
10
15
20
25
30
6Y-10
6Y-6
6Y-1
6Y-2
6Y-4
6Y-7
6Y-9
6Y-5
6Y-3
6Y-8
6 Year-old
5Y-6
5Y-4
5Y-7
5Y-5
5Y-8
5Y-9
5Y-10
5Y-1
5Y-2
5Y-3
5 Year-old
4Y-1
4Y-10
4Y-8
4Y-7
4Y-6
4Y-5
4Y-9
4Y-4
4Y-2
4Y-3
4 Year-old

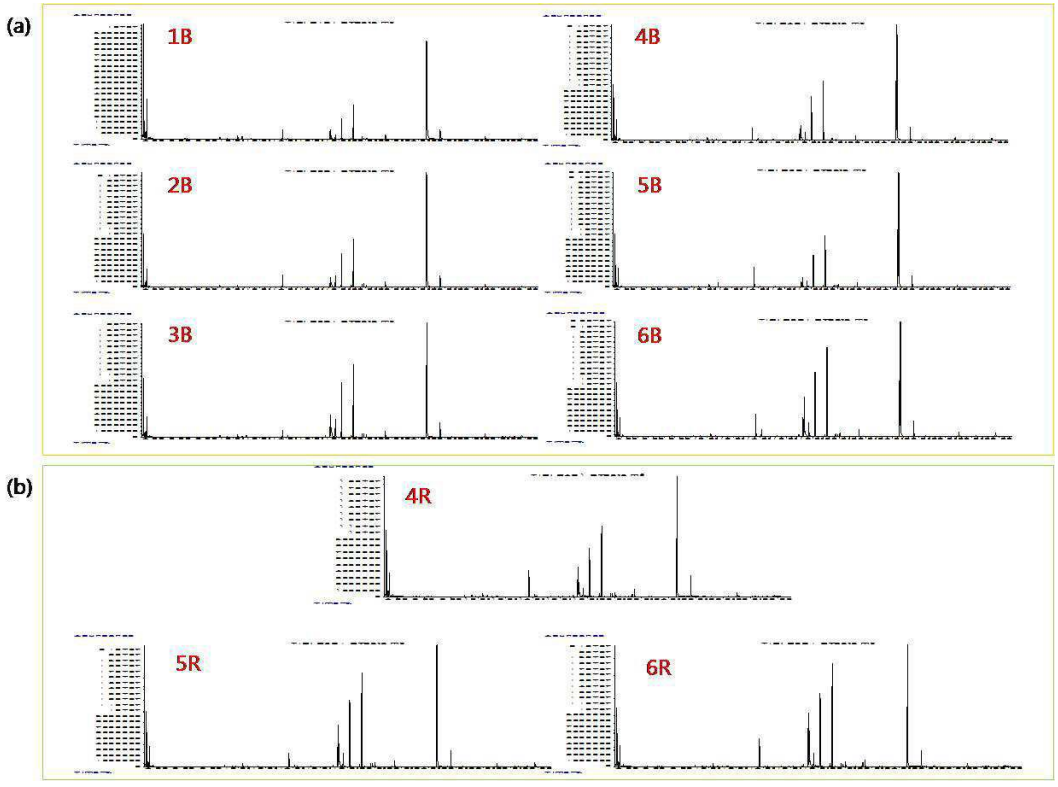
도면15


(a)
(b)

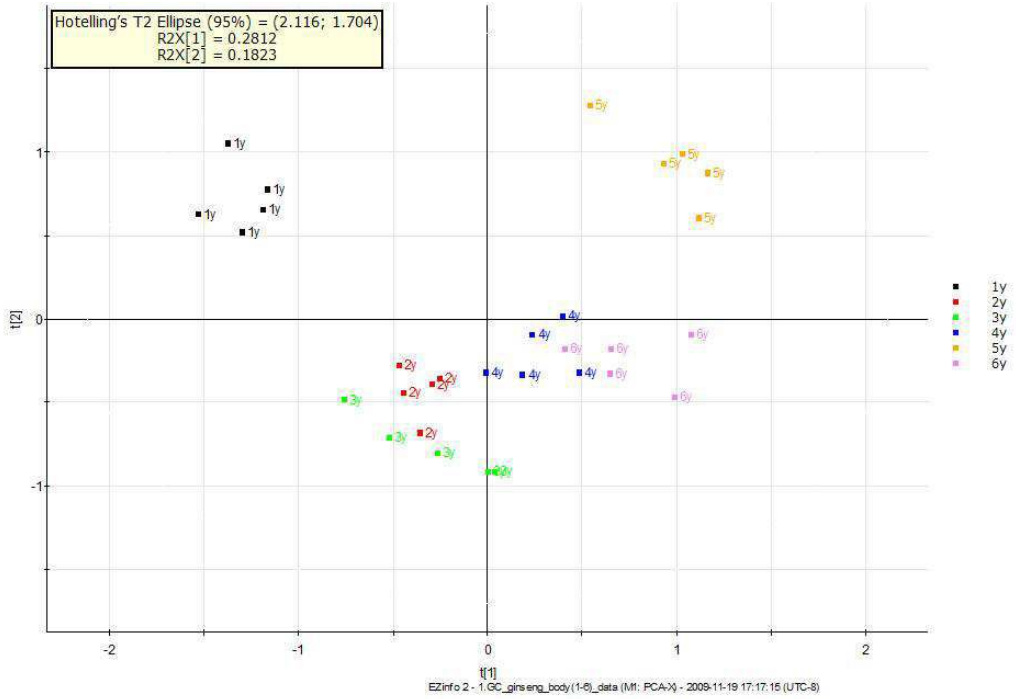
도면16


0
1
2
3
4
5
6
6Y-4
6Y-9
6Y-10
6Y-8
6Y-3
6Y-5
6Y-2
6Y-7
6Y-1
6Y-6
6 Year-old
4Y-3
4Y-6
4Y-8
4Y-2
4Y-1
4Y-9
4Y-4
4Y-7
4Y-5
4Y-10
4 Year-old
5Y-4
5Y-7
5Y-8
5Y-5
5Y-9
5Y-10
5Y-2
5Y-3
5Y-1
5Y-6
5Year-old

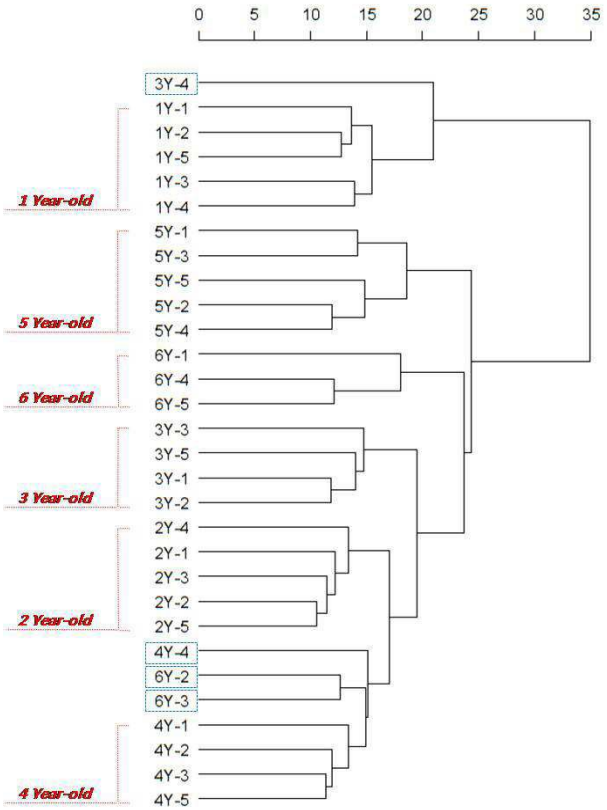
도면17


(a)
1B
4B
2B
5B
3B
6B
(b)
4R
5R
6R

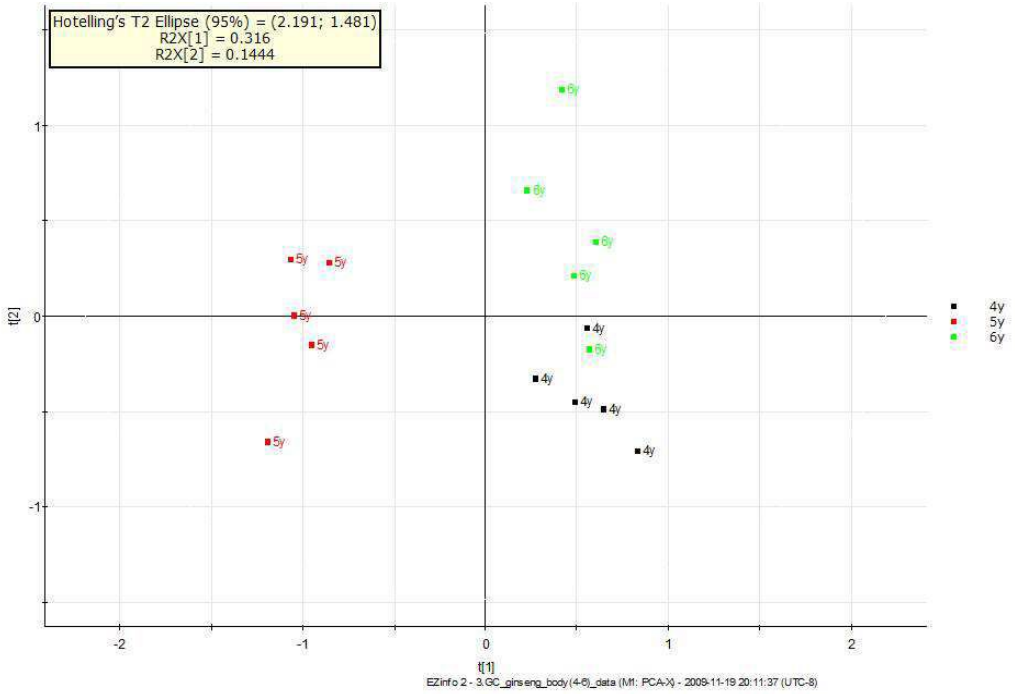
도면18


Hotelling's T2 Ellipse (95%) = (2.116; 1.704)
R2X[1] = 0.2812
R2X[2] = 0.1823
t[1]
t[2]
1y
2y
3y
4y
5y
6y
EZinfo 2 - 1.GC_ginseng_body (1-6)_data (M1: PCA-X) - 2009-11-19 17:17:15 (UTC-8)

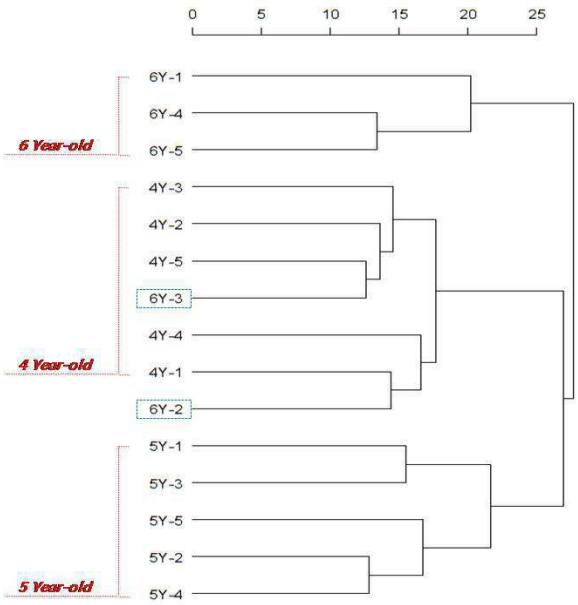
도면19


0
5
10
15
20
25
30
35
3Y-4
1Y-1
1Y-2
1Y-5
1Y-3
1Y-4
1 Year-old
5Y-1
5Y-3
5Y-5
5Y-2
5Y-4
5 Year-old
6Y-1
6Y-4
6Y-5
6 Year-old
3Y-3
3Y-5
3Y-1
3Y-2
3 Year-old
2Y-4
2Y-1
2Y-3
2Y-2
2Y-5
2 Year-old
4Y-4
6Y-2
6Y-3
4Y-1
4Y-2
4Y-3
4Y-5
4 Year-old

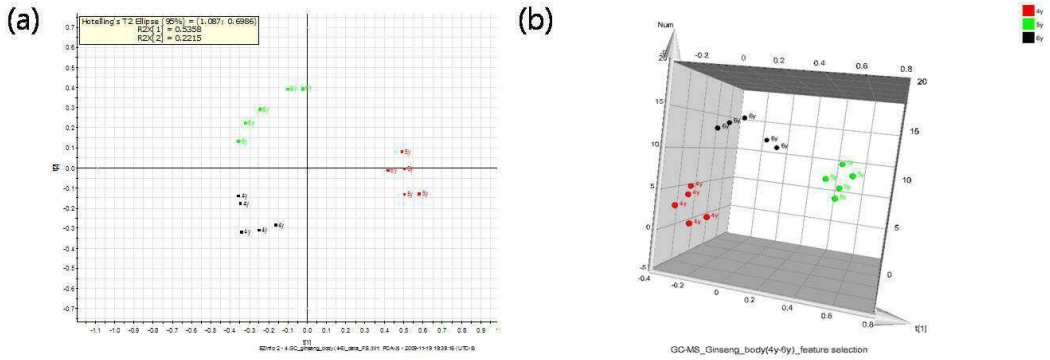
도면20


Hotelling's T2 Ellipse (95%) = (2.191; 1.481)
R2X[1] = 0.316
R2X[2] = 0.1444
t[2]
t[1]
4y
5y
6y
EZinfo 2 - 3.GC_ginseng_body (4-6)_data (M1: PCA-X) - 2009-11-19 20:11:37 (UTC-8)

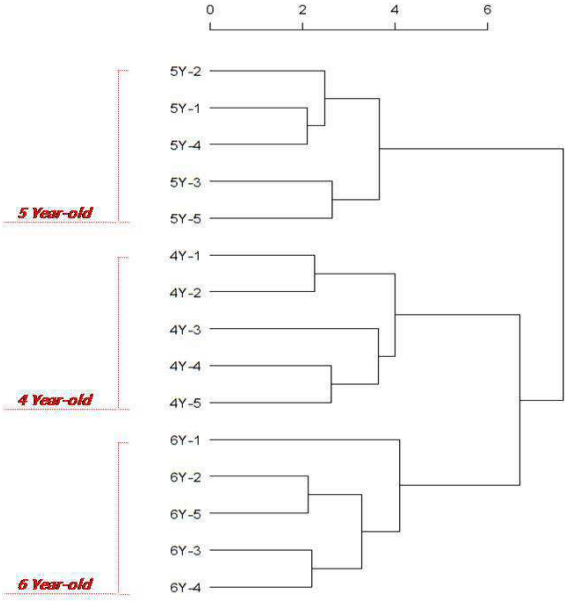
도면21


0
5
10
15
20
25
6Y-1
6Y-4
6Y-5
6 Year-old
4Y-3
4Y-2
4Y-5
6Y-3
4Y-4
4Y-1
6Y-2
4 Year-old
5Y-1
5Y-3
5Y-5
5Y-2
5Y-4
5 Year-old

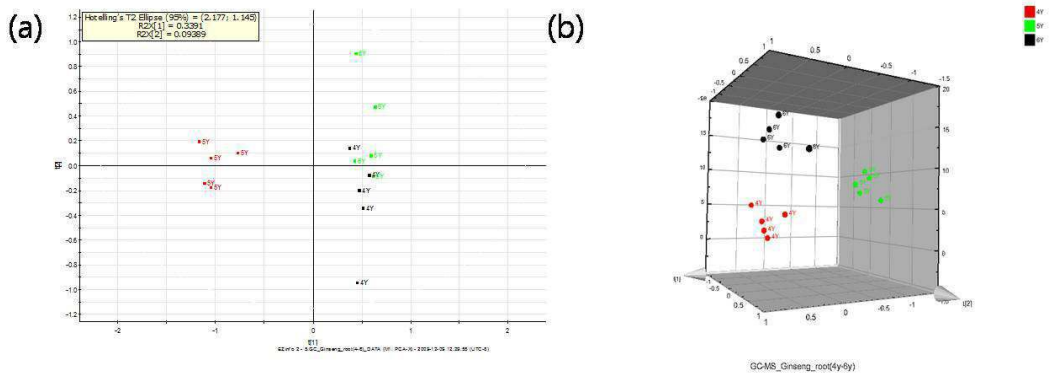
도면22



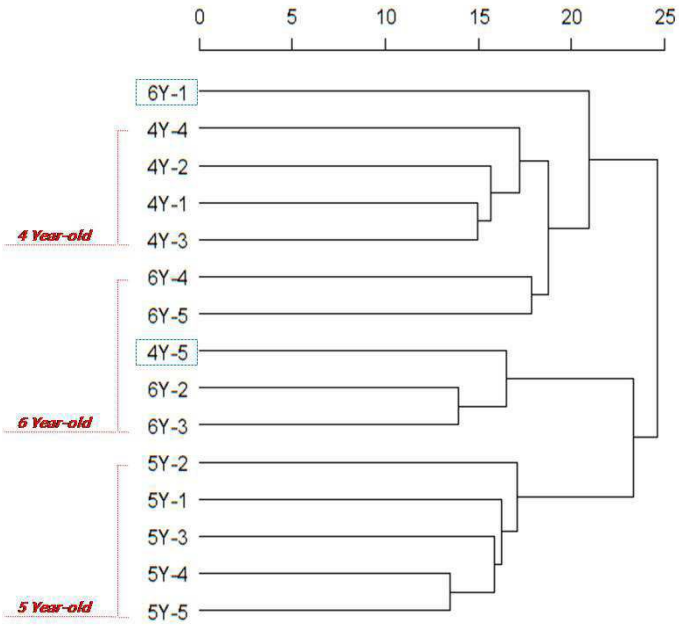
도면23



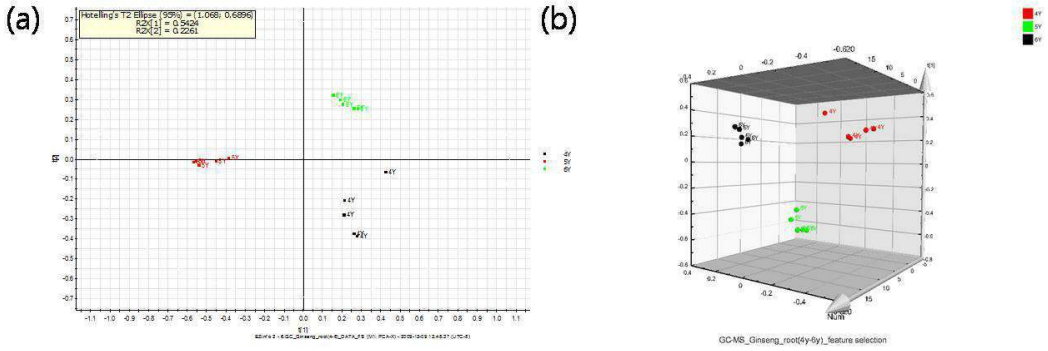
도면24



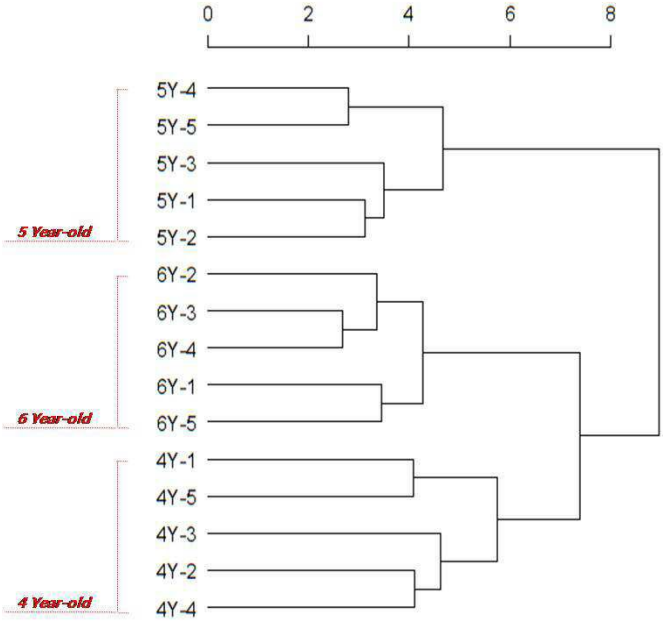
도면25


0
5
10
15
20
25
6Y-1
4Y-4
4Y-2
4Y-1
4Y-3
4 Year-old
6Y-4
6Y-5
4Y-5
6Y-2
6Y-3
6 Year-old
5Y-2
5Y-1
5Y-3
5Y-4
5Y-5
5 Year-old

도면26


(a)
(b)

도면27


0
2
4
6
8
5Y-4
5Y-5
5Y-3
5Y-1
5Y-2
5 Year-old
6Y-2
6Y-3
6Y-4
6Y-1
6Y-5
6 Year-old
4Y-1
4Y-5
4Y-3
4Y-2
4Y-4
4 Year-old