



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0165645
(43) 공개일자 2022년12월15일

(51) 국제특허분류(Int. Cl.)
C12N 9/00 (2006.01) C07K 14/41 (2006.01)
C12N 1/14 (2018.01) C12N 15/80 (2006.01)
C12P 7/22 (2006.01) C12R 1/645 (2006.01)
(52) CPC특허분류
C12N 9/93 (2013.01)
C07K 14/41 (2013.01)
(21) 출원번호 10-2022-0059678
(22) 출원일자 2022년05월16일
심사청구일자 2022년05월16일
(30) 우선권주장
1020210073669 2021년06월07일 대한민국(KR)

(71) 출원인
순천대학교 산학협력단
전라남도 순천시 중앙로 255(매곡동)
(72) 발명자
허재선
전라남도 순천시 해룡면 신대로 96 312-1301
김원용
전라남도 순천시 왕지5길 9-26 두산위브2차아파트
204동 503호
(뒷면에 계속)
(74) 대리인
정성중

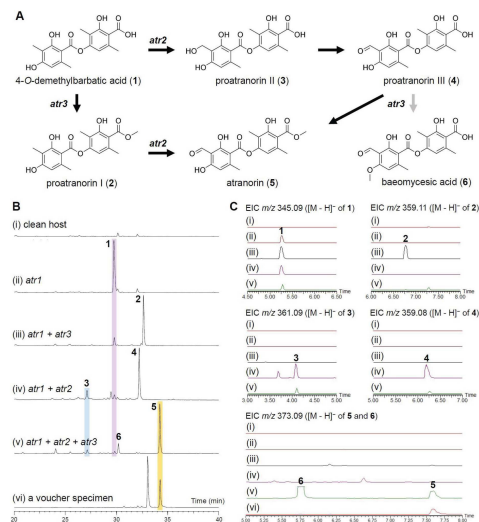
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 지의류로부터 유래된 아트라노린 생합성 유전자 및 이의 용도

(57) 요약

본 발명은 지의류인 스테레오카울론 알피눔(*Stereocaulon alpinum*)으로부터 유래된 아트라노린 생합성 유전자, 상기 유전자가 도입된 아라트라노린 생산용 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP) 및 상기 균주를 이용한 아트라노린 생산 방법에 관한 것이다. 본 발명에 따른 아트라노린 생합성 유전자가 도입된 아스코키타 라비에이 ATR-11 균주로부터 지의류 유래 2차 대사산물, 특히 아트라노린을 생산할 수 있다. 상기 균주로부터 생산된 아트라노린은 약학 조성물, 식품 조성물, 건강기능식품, 사료 조성물 등에 다양하게 활용 및 산업화할 수 있다.

대표도 - 도4



(52) CPC특허분류

C12N 1/145 (2021.05)

C12N 15/80 (2013.01)

C12P 7/22 (2013.01)

C12R 2001/645 (2021.05)

(72) 발명자

하형호

전라남도 순천시 왕지3길 55 두산위브 101-1404호

김항건

광주광역시 서구 시청로96번길 15, 1103호 (치평동, 로잔티움파크)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711121840

과제번호 2018H1D3A1A01074888

부처명 과학기술정보통신부

과제관리(전문)기관명 한국연구재단

연구사업명 국제연구인력교류(R&D)

연구과제명 유전체와 전사체 분석을 통한 지의류 형성곰팡이 2차대사산물 생합성 PKS 유전자 발굴 및 이형질체 발현을 통한 유용물질 대량생산 기반 구축

기 여 율 1/1

과제수행기관명 순천대학교

연구기간 2020.01.01 ~ 2020.12.31

명세서

청구범위

청구항 1

서열번호 1, 서열번호 3 및 서열번호 5로 이루어진 군으로부터 선택되는 1종 이상으로 이루어진, 아트라노린 생합성 유전자.

청구항 2

제1항에 있어서, 상기 아트라노린 생합성 유전자는 지의류로부터 유래된 것인, 아트라노린 생합성 유전자.

청구항 3

제2항에 있어서, 상기 지의류는 스테레오카울론 알피눔(*Stereocaulon alpinum*)인 것인, 아트라노린 생합성 유전자.

청구항 4

서열번호 1, 서열번호 3 및 서열번호 5로 이루어진 군으로부터 선택되는 1종 이상으로 이루어진 아트라노린 생합성 유전자를 포함하는, 지의류 유래의 대사산물 생산용 재조합 발현 벡터.

청구항 5

제4항에 있어서, 상기 지의류 유래의 대사산물은 4-*O*-디메틸바르바트산(4-*O*-demethylbarbatic acid), 프로아트라노린 I(proatranorin I), 프로아트라노린 II(proatranorin II), 프로아트라노린 III(Proatranorin III), 아트라노린(atranorin) 및 베오미세스산(baeomycesic acid)으로 이루어진 군으로부터 하나 이상 선택되는 것인, 지의류 유래의 대사산물 생산용 재조합 발현 벡터.

청구항 6

제4항 또는 제5항의 재조합 발현 벡터로 형질전환된 재조합 미생물.

청구항 7

제6항의 재조합 미생물을 배양하는 단계를 포함하는, 지의류 유래의 대사산물 생산 방법.

청구항 8

제7항에 있어서, 상기 지의류 유래의 대사산물은 4-*O*-디메틸바르바트산(4-*O*-demethylbarbatic acid), 프로아트라노린 I(proatranorin I), 프로아트라노린 II(proatranorin II), 프로아트라노린 III(Proatranorin III), 아트라노린(atranorin) 및 베오미세스산(baeomycesic acid)으로 이루어진 군으로부터 하나 이상 선택되는 것인, 지의류 유래의 대사산물 생산 방법.

청구항 9

아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP).

청구항 10

제9항에 있어서, 상기 균주는 아트라노린 생산용인, 균주.

청구항 11

제9항에 있어서, 상기 균주는 서열번호 1, 서열번호 3 및 서열번호 5로 이루어진 군으로부터 선택되는 1종 이상으로 이루어진, 아트라노린 생합성 유전자가 도입된 것인, 균주.

청구항 12

제9항 내지 제11항 중 어느 한 항의 균주를 배양하는 단계를 포함하는, 아트라노린 생산 방법.

발명의 설명

기술 분야

[0001] 본 발명은 지의류로부터 유래된 아트라노린 생합성 유전자 및 이의 용도에 관한 것이다. 보다 상세하게는, 스테레오카울론 알피눔(*Stereocaulon alpinum*)으로부터 유래된 아트라노린 생합성 유전자, 상기 유전자가 도입된 아트라노린 생산용 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC 83048BP) 및 상기 균주를 이용한 아트라노린 생산 방법에 관한 것이다.

배경 기술

[0002] 지의류는 생태계 기능에서 중요한 역할을 하며 알려진 모든 균류의 약 20%를 차지한다. 폴리케타이드(Polyketide) 유래 천연물은 지의류 지의체(lichen thalli)의 피질층과 수질층에 축적되며, 이 중 일부는 생물 및 비생물적 스트레스(초식동물 공격 및 자외선 조사)로부터 보호하는 데 중요한 역할을 한다. 그러나, 현재까지 단일 지의류 제품은 유전적 증거가 있는 각각의 생합성 유전자와 연결되어 있지 않다.

[0003] 지의류 형 성곰팡이(Lichen-forming fungi, LFF)는 광합성 파트너인 녹조류 또는 남조류와 공생하며, 때로는 두 가지 모두와 공생하며 LFF는 현재 알려진 모든 균류의 약 20%를 구성하는 것으로 추정된다. 지의류 공생은 이러한 유기체가 극도로 가혹한 서식지에 적응할 수 있게 하는 가장 성공적인 공생 중 하나이다. 조밀한 곰팡이 균사로 구성된 응집되고 종종 착색된 피질층은 광합성 파트너에게 기계적 안정화를 제공하고 수질의 균사로 둘러싸인 조류 세포는 광합성을 통해 LFF에 영양분을 제공한다. 폴리케타이드(polyketide) 기원의 이차 대사산물(SM), 즉 안트라퀴논, 테사이드, 테시돈 및 디벤조푸란(예컨대, 우스닉산)은 지의류 몸체의 피질 또는 수질층에 축적된다(Calcott MJ *et.al.*, *Chem Soc Rev.*, 47:1730-1760, 2018). 이 지의류 SM의 생태학적 역할은 거의 알려져 있지 않지만, 일부 연구에서는 피질 물질이 서식지 확장 및 초식 동물의 공격으로부터의 방어에 기여하였다는 증거를 제공한다.

[0004] 테사이드 및 테시돈 계열 화합물은 지의류에서 피질 및 수질 물질로 널리 퍼져 있으며, 이들 중 다수는 지의류에서만 독점적으로 발견된다. 지의류 테사이드는 오르셀린산 또는 3-메틸오르셀린산(3MOA)의 이량체화에 의해 형성되고, 테사이드의 후속 산화는 테시돈에 대한 삼환식 스캐폴드를 제공한다. 이들 화합물은 추가로 알킬화, 염소화, 수산화 및 *O*-메틸화와 같은 오르셀린산 및 3MOA 모이어티 내에서 다양한 변형 조합을 거쳐 수백 가지의 구조적으로 다양한 화합물을 생성한다. 이를 위해, 지의류에서 테사이드/테시돈 생합성 경로는 화학다양성을 초래하는 생합성 효소의 기질 특이성을 연구하기 위한 모델 시스템이 될 수 있다. 그러나, 지금까지 지의류에서 특정 생합성 유전자를 피질 또는 수질 물질에 귀속시키는 것은 LFF에 대한 유전 정보가 부족하고, 유전자 변형에 대한 내성이 있는 LFF를 조작하기 위한 분자 도구의 부족으로 인해 지의류에서 저조하였다.

[0005] 비환원 반복 유형 I 폴리케타이드 합성 효소(Nonreducing iterative type I polyketide synthases, NR-PKS)는 다중 도메인 효소이며, 단백질 서열 유사성 및 PKS 도메인 구조에 따라 7개 또는 8개의 주요 그룹으로 분류된다. 오르셀린산 및 3MOA는 박테리아, 곰팡이 및 식물의 다양한 SM에 대한 기본 골격이며, 사상균인 아스페르길루스 니둘란스(*Aspergillus nidulans*)에서 NR-PKS에 의해 생합성되나, 두 개의 NR-PKS가 지의류에서 기본 단위의 생합성을 담당할 것이라는 예측만 있었다. 아트라노린(3MOA 유래 테사이드) 및 우스닉산(디벤조푸란)은 거대지의류의 가장 흔한 피질 물질이며, 이들의 분류학적, 생태학적 및 약학적 중요성 때문에 큰 주목을 받았다. 여러 연구에서 NR-PKS는 가능성이 높은 우스닉산에 기인한다. 그러나, 아스페르길루스 오리제(*A. oryzae*)에서 추정되는 우스닉산 PKS 유전자의 이중 발현에 대한 노력은 성공하지 못하여 정확한 기능은 아직 밝혀지지 않았다.

[0006] 최근 몇 년 동안, SM 연구는 박테리아, 곰팡이 및 식물에서 게놈 마이닝 접근법으로부터 광범위하게 혜택을 받았다. 남극 지역 지의류인 사슴지의과(Cladoniaceae)는 LFF의 가장 큰 과(family) 중 하나이며, 나무지의속(*Stereocaulaceae*)과 밀접한 관련이 있다. 역사적으로, 화학분류법은 사슴지의과에서 종의 경계를 해결하고 구분하기 위한 다상 접근법으로 사용되어 왔다. 또한, 클라도니아속(*Cladonia*)은 피질 및 수질 물질의 생합성을 위한 PKS 유전자 연구를 위한 모델로 이용되어 왔다(Armaleo D *et.al.*, *Mycologia*, 103:741-754, 2011).

[0007] 이에 본 발명자들은 지의류 폴리케타이드 합성 효소(PKS)를 분류하고, 이중 숙주에서 지의류의 주요 피질 물질인 아트라노린의 생합성 경로를 재구성하였다. 이에 따라, 지의류 피질 물질인 아트라노린의 생합성을 담당하는

유전자를 확보하고, 이를 도입한 새로운 이중 발현 시스템을 구축하여 신규한 아트라노린 생산용 균주를 개발함으로써, 본 발명을 완성하였다.

선행기술문헌

비특허문헌

- [0008] (비특허문헌 0001) Calcott MJ et.al., Chem Soc Rev., 47:1730-1760, 2018
(비특허문헌 0002) Armaleo D et.al., Mycologia, 103:741-754, 2011

발명의 내용

해결하려는 과제

- [0009] 본 발명의 목적은 서열번호 1, 서열번호 3 및 서열번호 5로 이루어진 균으로부터 선택되는 1종 이상으로 이루어진, 아트라노린 생합성 유전자를 제공하기 위한 것이다.
- [0010] 본 발명의 다른 목적은 상기 아트라노린 생합성 유전자를 포함하는, 지의류 유래의 대사산물 생산용 재조합 발현 벡터를 제공하기 위한 것이다.
- [0011] 본 발명의 또 다른 목적은 상기 재조합 발현 벡터로 형질전환된 재조합 미생물을 제공하기 위한 것이다.
- [0012] 본 발명의 또 다른 목적은 상기 재조합 미생물을 배양하는 단계를 포함하는, 지의류 유래의 대사산물 생산 방법을 제공하기 위한 것이다.
- [0013] 본 발명의 또 다른 목적은 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)를 제공하기 위한 것이다.
- [0014] 본 발명의 또 다른 목적은 상기 균주를 배양하는 단계를 포함하는, 아트라노린 생산 방법을 제공하기 위한 것이다.

과제의 해결 수단

- [0015] 상기 목적을 달성하기 위해, 본 발명은 서열번호 1, 서열번호 3 및 서열번호 5로 이루어진 균으로부터 선택되는 1종 이상으로 이루어진, 아트라노린 생합성 유전자를 제공한다.
- [0016] 또한, 본 발명은 상기 아트라노린 생합성 유전자를 포함하는, 지의류 유래의 대사산물 생산용 재조합 발현 벡터를 제공한다.
- [0017] 또한, 본 발명은 상기 재조합 발현 벡터로 형질전환된 재조합 미생물을 제공한다.
- [0018] 또한, 본 발명은 상기 재조합 미생물을 배양하는 단계를 포함하는, 지의류 유래의 대사산물 생산 방법을 제공한다.
- [0019] 또한, 본 발명은 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)를 제공한다.
- [0020] 또한, 본 발명은 상기 균주를 배양하는 단계를 포함하는, 아트라노린 생산 방법을 제공한다.

발명의 효과

- [0021] 본 발명에 따른 지의류 유래 아트라노린 생합성 유전자가 도입된 이중 발현 시스템을 구축하고, 이를 적용한 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주로부터 지의류 유래 대사산물, 특히 아트라노린을 생합성할 수 있다. 상기 균주로부터 생산된 아트라노린은 약학 조성물, 식품 조성물, 건강기능식품, 사료 조성물 등에 다양하게 활용 및 산업화할 수 있다.

도면의 간단한 설명

- [0022] 도 1은 지의류의 게놈-인코딩 대사 잠재력을 나타낸 도이다. 29종의 지의류-형성 곰팡이 및 시아노데르멜라 아스테리스(*Cyanoderma asteris*)(식물 내생식물)에 대해 병합-기반 종 트리를 재구성하였다. 이 트리는 엔도카

르폰 푸실룸(*Endocarpon pusillum*)에 뿌리를 두고 있다. 노드 지원에 대한 모든 로컬 사후 확률은 화살표로 표시된 경우를 제외하고 0.98보다 높다. 접시지의강(Lecanoromycetes)의 다섯 목은 다른 색상으로 강조 표시되었으며, 한국 지의류 연구소에서 시퀀싱된 지의류 계통은 굵게 표시되어 있다. BUSCO는 계통 어셈블리의 완전성을 설명한다. 생합성 유전자 클러스터(BGC)의 수는 antiSMASH에 의해 예측되고, 거품 도표는 핵심 생합성 효소 범주의 상대적 풍부함을 나타낸다: NR-PKS, 비환원형 I PKS; R-PKS, 환원형 I PKS(PKS-NRPS 하이브리드 효소 포함); PKSIII, 유형 III PKS; NRPS, 비리보솜 펩티드 합성 효소(NRPS-유사 효소 포함); 및 테르펜, 테르페노이드 생산과 관련된 생합성 효소.

도 2는 클라도니아(*Cladonia*) 종에서의 폴리케타이드(polyketide) BGC 다양성을 나타낸 도면이다. (A) 6종의 클라도니아속(*Cladonia* spp.) 및 스테레오카울론 알피눔(*Stereocaulon alpinum*, S. alpinum)에서 유전자 클러스터 네트워크 분석에 의해 식별된 45개의 유전자 클러스터 패밀리(GCF)를 나타낸 것이다. 여기에서, PKS 패밀리와 관련된 GCF를 표지하였으며, 빨간색 숫자는 MIBiG 데이터베이스에 기록된 특성화된 BGC에 연결된 PKS 패밀리를 나타낸 것이다. 노드(node)는 BGC를 나타내며, 종별로 색상으로 구분하였다. PKS8, PKS14 및 PKS16 하위 네트워크에서 빨간색 점선 원으로 둘러싸인 노드는 MIBiG 데이터베이스의 지의류 BGC이다. 알려진 BGC에 연결된 싱글톤 BGC는 M, 모나스코루브린(monascorubrin) BGC; H, 하이포테마이신(hypothemycin) BGC; S, 소르비실린(sorbicillin) BGC; T, 테레산(terreic acid) BGC; B, 베타에논(betaenones) BGC; D, 테푸데신(depudecin) BGC; C, 커브팔리드(curvupallides) BGC로 표지하였다. (B) 유전자 클러스터 네트워크 분석에 의해 입증된 PKS 패밀리의 계통 분포를 나타낸 것이다. 보라색과 황토색의 숫자는 각각 NR-PKS 및 R-PKS 제품군을 나타낸 것이다. 검은색 사각형은 PKS 패밀리의 존재를 나타내고, 회색 사각형은 불완전한 유전자 주석으로 인한 가유전자 또는 부분 유전자를 나타낸 것이다. 유전적 복제 제거에 의해 클라도니아속(*Cladonia*) PKS 패밀리와 연결된 시그니처 2차 대사산물은 볼드체로 표시하였다.

도 3은 추정되는 아트라노린 BGC의 구성을 나타낸 도이다. CORASON 분석 파이프라인을 사용하여, 30개의 계통에서 검출된 1,527개의 BGC로부터 C. 랑기페리나(*C. rangiferina*)의 PKS23 BGC(Crg, 빨간색으로 표시) 중 하나에 상당한 히트를 나타내는 O-메틸트랜스퍼라제(O-methyltransferase)를 포함하는 15개의 BGC가 식별되었다. BGC를 보유하는 지의류 종은 도 1과 같이 약칭으로 표시하였다. 파란색 숫자는 Crg에 있는 것과 상동성 OMT의 단백질 서열 동일성을 백분율로 나타낸 것이다. 15개의 BGC 중, 7개의 BGC는 4개의 합성 유전자(회색 박스 표시): *atr1* (PKS23), *atr2*, *atr3* 및 *atr4*를 보유하였다. 2차 대사와 관련된 유전자는 예측된 기능에 따라 색상으로 구분하였다: NR-PKS, 비환원형 I PKS; R-PKS, 환원형 I PRS; NRPS, 비리보솜 펩티드 합성 효소; TF, 전사 인자; RTA, 7-아미노콜레스테롤에 대한 내성.

도 4는 아트라노린 BGC의 기능적 검증 결과를 나타낸 도이다. (A) 아트라노린의 제안된 생합성 경로를 나타낸 것이다. (B) 아트라노린 BGC의 다양한 생합성 유전자 세트를 발현하는 "클린 숙주"의 배양 추출물의 HPLC 프로파일을 나타낸 것이다: 도입된 유전자 없음 (i), *atr1*만 있음 (ii), *atr1* 및 *atr3* 있음(iii), *atr1* 및 *atr2* 있음 (iv), 및 *atr1*, *atr2* 및 *atr3* 있음(v). 계통-시퀀싱된 스테레오카울론 알피눔(*Stereocaulon alpinum*)에 대한 확증 표본의 아세톤 추출물에는 아트라노린 및 로바르산(*tR* = 33.1분) (vi)이 포함되어 있다. (C) 화합물 1 내지 6에 대해 *m/z* 값을 표시한 추출 이온 크로마토그램(EIC, Extracted ion chromatogram) 결과를 나타낸 것이다.

도 5는 곰팡이 비환원성 PKS 유전자의 9개의 계통군을 나타낸 도이다. 6개의 클라도니아속(*Cladonia* spp.) 및 스테레오카울론 알피눔(*Stereocaulon alpinum*)에서 103개의 NR-PKS의 최대 우도(likelihood) 트리와 비-지의류화된(non-lichenized) 곰팡이의 알려진 화합물에 연결된 82개의 NR-PKS는 케토신타제(ketosynthase) 및 생성물 주형 도메인의 연결된 시퀀스를 사용하여 재구성하였다. 외부 그룹(Outgroup)은 클라도니아속에서 발견되는 6-메틸살리실산 합성 효소(6-methylsalicylic acid synthase, 6MSAS)와 페니실리움 엑스판숨(*Penicillium expansum*)에서 파툴린(patulin) 생합성을 위한 6MSAS로 설정되었다. 분기는 부트스트랩(BS) 값을 기반으로 색상으로 구분하였다. 척도는 부위당 0.5개의 아미노산 서열 치환을 나타낸다. 가장 바깥쪽 영역에 관련 번호가 있는 파란색 및 빨간색 스트립은 유전자 클러스터 네트워크 분석에 의해 식별된 PKS 패밀리를 나타낸다(도 2 참조). 9개의 NR-PKS 그룹(그룹 I-IX)을 나타내는 계통발생 분기군은 다른 색상으로 음영 처리되었다. NR-PKS 그룹은 그룹 II를 제외하고, 75보다 큰 BS 값에 의해 지원된다. 새로 확인된 NR-PKS 그룹(그룹 IX)에는 지의류의 피질 및 수질 물질에 대한 후보 PKS인 PKS1, PKS2 및 PKS23 패밀리가 포함된다. 빨간색 화살표는 NR-PKS의 C-메틸트랜스퍼라제 도메인의 다계통성 손실을 나타낸다.

도 6은 지의류의 피질 및 수질 물질의 화학 구조를 나타낸 도이다. 화학 구조에 따라, 9개의 화학 그룹으로 분

류하였다 (표 2 참조). 여기에서, 3MOA는 3-메틸오르셀린산의 약어이다.

도 7은 PKS23의 이중 발현 검증 결과를 나타낸 도이다. (A) 스테레오카울론 알피눔(*Stereocaulon alpinum*)에서 PKS23(*atr1*) 유전자 구조의 개략도를 나타낸 것이다. 여기에서, 검은색 박스는 엑손이고, 회색 선은 인트론이다. 숫자는 염기쌍(bp)에서 엑손과 인트론의 크기를 나타낸 것이다. P1 및 P2는 RT-PCR 분석에 사용되는 프라이머 쌍이다. 상기 프라이머 쌍은 mRNA 발현이 게놈 DNA(gDNA, genomic DNA) 증폭과 구별될 수 있도록 하나 이상의 인트론 영역을 포함하도록 설계되었다. (B) 플라스미드 *sol1::atr1/pDS35*를 운반하는 2개의 추정 형질전환체 (T16 및 T25)의 RT-PCR 분석 결과를 나타낸 것이다. 액틴 1(Actin 1) 참조 유전자(mRNA의 경우 633 bp)에 대한 gDNA 앰플리콘에 해당하는 밴드가 관찰되지 않아, 게놈 DNA 오염이 없음을 나타낸다(좌측 하단 패널). 형질전환체 T25는 형질전환체 T16 (좌측 상단 패널)보다 도입된 *atr1*의 더 강한 발현을 나타내었다. *atr1*에서 4개의 인트론의 스플라이싱은 P1 및 P2 프라이머 쌍을 사용하여 플라스미드 *sol1::atr1/pDS35*(C)와 RNA 샘플(T25) 사이의 밴드 크기를 비교하여 확인하였다. mRNA 발현 및 플라스미드 DNA 증폭에 대한 예상 밴드 크기는 bp 단위로 겔 이미지의 좌측 및 우측에 숫자로 표시하였다. 여기에서, M은 100 bp DNA 크기 마커이다.

발명을 실시하기 위한 구체적인 내용

[0023]

폴리케타이드 기원의 템사이드 및 템시돈 계열 화합물은 지의류 몸체(lichen thalli)의 피질 또는 수질층에 축적된다. 지의류 화학의 분류학적 및 생태학적 중요성과 약학적 잠재력에도 불구하고, 생합성 유전자와 지의류 물질을 연결하는 단일 유전적 증거는 없다. 따라서, 본 발명자들은 템사이드/템시돈 생산에 관여하는 생합성 유전자 클러스터(BGC)의 분류 및 식별을 위해 지의류 폴리케타이드 합성 효소(PKS, polyketide synthase)를 체계적으로 분석하였다. 클라도니아속(*Cladonia*)의 중간 PKS 다양성 및 관련된 남극 지의류 스테레오카울론 알피눔(*Stereocaulon alpinum*)에 대한 본 발명자들의 심층 분석은 45개의 BGC 패밀리를 식별하여, 지의류 PKS를 비-지의류화된 곰팡이에서 이전에 특성화된 15개의 PKS와 연결하였다. 이들 중에서, 본 발명자들은 아트라노린(템사이드)을 생산하는 지의류에서만 독점적으로 발견되는 고도로 합성된 BGC를 확인하였다. 추정 아트라노린 PKS(*atr1* 생성)의 이중 발현은 많은 지의류에서 전구체 화합물로서 발견되는 4-O-디메틸바르바트산(4-O-demethylbarbatic acid)을 생성하였으며, 이는 템사이드 형성을 위한 Atr1의 분자간 가교 활성을 나타낸다. 이중 숙주에 맞춤 효소를 계속 도입하면, 거대 지의류의 가장 흔한 피질 물질 중 하나인 아트라노린이 생성되었다. 곰팡이 PKS의 계통발생 분석은 Atr1이 템사이드 및 템시돈의 생합성에 관여할 가능성이 있는 두 개의 보존된 지의류-특이적 PKS 패밀리를 포함하는 새로운 PKS 계통군에 배치되는 것으로 나타났다. 본 발명에서 본 발명자들은 클라도니아속(*Cladonia*)의 PKS 패밀리의 포괄적인 카탈로그를 제공하고, 기능적으로 지의류의 생합성 유전자 클러스터를 특성화하여 다양한 지의류 물질의 유전학 및 화학적 진화를 연구하기 위한 초석을 확립하였다.

[0025]

신규의 아트라노린 생합성 유전자

[0026]

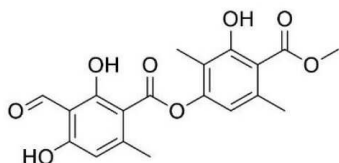
하나의 양태로서, 본 발명은 서열번호 1, 서열번호 3 및 서열번호 5로 이루어진 군으로부터 선택되는 1종 이상으로 이루어진, 아트라노린 생합성 유전자를 제공한다.

[0027]

본 명세서에 사용된 용어 "아트라노린(atranorin)"은 화학식명이 3-하이드록시-4-메톡시카보닐-2,5-다이메틸페닐-3-포밀-2,4-다이하이드록시-6-메틸벤조에이트이며, 하기 [화학식 1]의 구조식을 가진다. 이러한 아트라노린은 지의체의 이차 대사산물로서 지의체로부터 유래될 수 있다. 상기 아트라노린은 폐암 전이 억제 효능을 갖는 것으로 보고된 바 있다.

[0028]

[화학식 1]



[0029]

본 발명의 아트라노린은 아트라노린의 유사체, 유도체를 포함하는 의미로 사용될 수 있다.

[0030]

[0031]

본 명세서에 사용된 용어 "지의체(thallus)"는 지의류 공생 조류 또는 시아노박테리아의 세포, 및 지의류 형성 곰팡이의 균사로 이루어지고, 분아, 열아 또는 영양번식을 위한 지의류의 영양체를 말한다.

- [0032] 본 명세서에 사용된 용어 "지의류(lichen)"는 곰팡이(fungi, mycobionts)와 조류(algae, photobionts)가 복합체가 되어 생활하는 공생생물로서 모양, 크기 및 색깔이 매우 다양하며 주로 바위, 수피 및 토양에서 서식한다. 이러한 지의류는 디디믹 산(didymic acid), 스트렙실린(strepsilin), 소듐 우스네이트(sodium usnate), 레카노릭 산(lecanoric acid), 프소로믹 산(psoromic acid) 등 1, 2차 대사산물들을 생성하여 향미생물(곰팡이, 세균, 바이러스), 제초제, 향암, 면역강화, 정신질환 개선 등에 우수한 생리활성을 가지므로 예로부터 동서양을 막론하고 전통적으로 식품 또는 약재의 원료로 사용되어 왔다.
- [0033] 본 발명에 있어서, 상기 아트라노린 생합성 유전자는 지의류로부터 유래된 2차 대사산물, 일례로 아트라노린 발현을 유도할 수 있는 유전자로, 서열번호 1, 서열번호 3 및 서열번호 5로 이루어진 군으로부터 선택되는 1종 이상으로 이루어질 수 있다. 상기 지의류는 이에 제한되지는 않으나, 스테레오카울론 알피눔(*Stereocaulon alpinum*)일 수 있다.
- [0034] 상기 서열번호 1로 표시되는 염기서열은 *atr1* 유전자(NCBI Genbank No. MZ277879.1)로, 서열번호 3으로 표시되는 염기서열은 *atr2* 유전자(NCBI Genbank No. MZ277878.1)로, 서열번호 5로 표시되는 염기서열은 *atr3* 유전자(NCBI Genbank No. MZ277877.1)로 각각 명명될 수 있다.
- [0035] 또한, 상기 서열번호 1로 표시되는 염기서열은 서열번호 2의 아미노산 서열로 이루어진 *atr1* 단백질(NCBI Genbank No. QXF68953.1)로 암호화될 수 있다. 상기 서열번호 3으로 표시되는 염기서열은 서열번호 4의 아미노산 서열로 이루어진 *atr2* 단백질(NCBI Genbank No. QXF68952.1)로 암호화될 수 있다. 상기 서열번호 5로 표시되는 염기서열은 서열번호 6의 아미노산 서열로 이루어진 *atr3* 단백질(NCBI Genbank No. QXF68951.1)로 암호화될 수 있다.
- [0037] **지의류 유래 대사산물 생산용 재조합 발현 벡터**
- [0038] 다른 하나의 양태로서, 본 발명은 서열번호 1, 서열번호 3 및 서열번호 5로 이루어진 군으로부터 선택되는 1종 이상으로 이루어진 아트라노린 생합성 유전자를 포함하는, 지의류 유래의 대사산물 생산용 재조합 발현 벡터를 제공한다.
- [0039] 상기 "아트라노린 생합성 유전자"는 전술한 바와 같다.
- [0040] 본 명세서에 사용된 용어 "벡터(vector)"는 적합한 숙주 내에서 목적 단백질을 발현시킬 수 있도록 적절한 조절 서열에 작동 가능하게 연결된 특정 유전자를 함유하는 DNA 제조물로, 상기 조절 서열은 전사를 개시할 수 있는 프로모터, 그러한 전사를 조절하기 위한 임의의 오퍼레이터 서열, 적합한 mRNA 리보솜 결합부위를 코딩하는 서열, 및 전사 및 해독의 종결을 조절하는 서열을 포함한다. 벡터는 적당한 숙주세포 내로 형질전환된 후, 숙주 세포와 무관하게 복제되거나 기능할 수 있으며, 세포 그 자체에 통합될 수 있다. 본 발명의 벡터는 당해 기술 분야에서 잘 알려진 유전자 재조합 기술을 이용하여 제조할 수 있으며, 부위-특이적 DNA 절단 및 연결은 당해 기술 분야에서 일반적으로 알려진 효소 등을 사용한다.
- [0041] 본 명세서에 사용된 용어 "발현 벡터"는 지의류 유래 아트라노린 생합성 유전자가 도입된 벡터로, 이종(heterologous)인 미생물 또는 숙주세포에서 자체적으로 발현시킬 수 없는 지의류 유래 대사산물을 인위적으로 발현시키기 위해, 지의류 유래 아트라노린 생합성 유전자를 이종 균주 내로 도입하여 지의류 유래 대사산물을 발현시킬 수 있는 재조합 발현 벡터를 의미한다.
- [0042] 구체적으로, 상기 도입된 아트라노린 생합성 유전자에 의해 지의류 유래의 2차 대사산물인 아트라노린(atranorin)을 비롯하여 4-O-디메틸바르바트산(4-O-demethylbarbatic acid), 프로아트라노린 I(proatranorin I), 프로아트라노린 II(proatranorin II), 프로아트라노린 III(Proatranorin III) 및 베오미세스산(baeomycesic acid) 등의 지의류 유래 대사산물을 이종 미생물 또는 숙주세포에서 발현시킬 수 있는 벡터를 의미한다. 본 발명에서 발현 벡터는 '벡터' 또는 '재조합 발현 벡터'와 혼용될 수 있다.
- [0043] 상기 발현 벡터는 바람직하게는 하나 이상의 선택성 마커를 포함할 수 있다. 상기 마커는 통상적으로 화학적인 방법으로 선택될 수 있는 특성을 갖는 핵산 서열로, 형질전환된 세포를 비 형질전환 세포로부터 구별할 수 있는 모든 유전자가 이에 해당된다. 그 예로는 암피실린(ampicilin), 카나마이신(kanamycin), G418, 블레오마이신(Bleomycin), 하이그로마이신(hygromycin), 클로람페니콜(chloramphenicol), 아프라마이신(apramycin)과 같은 항생제 내성 유전자가 있으나, 이에 한정되는 것은 아니며, 당업자에 의해 적절히 선택 가능하다.
- [0044] 본 발명의 일 구체예에서, 스테레오카울론 알피눔(*Stereocaulon alpinum*) 유래 아트라노린 생합성 유전자가 도

입된 벡터는 이종(heterologous)인 병아리 콩 역병 곰팡이균 아스코키타 라비에이(*Ascochyta rabiei*)에 형질전환되어 지의류 유래의 대사산물을 이종 발현시킬 수 있다.

[0046] **아트라노린 생합성 유전자 함유 재조합 발현 벡터가 적재된 재조합 미생물**

[0047] 또 다른 하나의 양태로서, 본 발명은 상기 재조합 발현 벡터로 형질전환된 재조합 미생물을 제공한다.

[0048] 상기 재조합 발현 벡터는 전술한 바와 같다.

[0049] 본 명세서에 사용된 용어 "형질전환"은 외부로부터 주어진 DNA에 의하여 생물의 유전적인 성질이 변하는 것을 의미하며, 특히 본 발명에서는 특정 유전자, 즉 지의류 유래 아트라노린 생합성 유전자를 포함하는 벡터를 특정 이종 미생물 내에 도입하여 특정 이종 미생물 내에서 상기 유전자가 발현될 수 있도록 하는 것을 의미한다.

[0050] 상기 재조합 발현 벡터는 특정 이종 미생물에 형질전환하여 재조합 미생물 균주를 제조하는데, 이때 상기 형질전환 방법은 공지된 기술이라면 특별히 이에 제한되지 않고 사용할 수 있다. 예를 들면 칼슘 클로라이드 방법 또는 전기천공법(Neumann, *et al.*, *EMBO J.*, 1:841, 1982)을 형질전환에 사용할 수 있다.

[0051] 상기 특정 이종 미생물은 당업계에 널리 알려져 있는 미생물 또는 숙주세포라면 제한없이 사용할 수 있다. 본 발명의 지의류 유래 아트라노린 생합성 유전자의 도입 효율과 발현 효율이 높은 미생물 또는 숙주세포를 사용할 수 있는데, 예를 들면, 곰팡이, 박테리아, 효모를 포함할 수 있다.

[0052] 구체적으로 상기 재조합 미생물은 곰팡이일 수 있고, 보다 구체적으로 아스코키타 라비에이(*Ascochyta rabiei*)일 수 있다. 보다 더 구체적으로, 상기 재조합 미생물은 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)일 수 있다.

[0054] **지의류 유래의 대사산물의 생산 방법**

[0055] 또 다른 하나의 양태로서, 본 발명은 상기 재조합 미생물을 배양하는 단계를 포함하는, 지의류 유래의 대사산물 생산 방법을 제공한다.

[0056] 상기 재조합 미생물은 전술한 바와 같이 서열번호 1, 서열번호 3 및 서열번호 5로 이루어진 균으로부터 선택되는 1종 이상으로 이루어진 아트라노린 생합성 유전자를 포함하는, 지의류 유래의 대사산물 생산용 재조합 발현 벡터가 도입된 재조합 미생물일 수 있다.

[0057] 상기 재조합 미생물은 이에 제한되지는 않으나, 4-*O*-디메틸바르바트산(4-*O*-demethylbarbatic acid), 프로아트라노린 I(proatranorin I), 프로아트라노린 II(proatranorin II), 프로아트라노린 III(Proatranorin III), 아트라노린(atranorin) 및 베오미세스산(baeomycesic acid)으로 이루어진 균으로부터 하나 이상 선택되는, 지의류 유래의 대사산물을 생산하는 것일 수 있다.

[0058] 상기 지의류 유래의 대사산물 생산용 재조합 미생물의 배양 조건은 형질전환된 재조합 미생물 또는 숙주세포의 종류에 따라 달라질 수 있다. 상기 재조합 미생물 또는 숙주세포의 종류에 따라 당업자에게 공지된 조건에 따라, 즉, 통상의 온도 조건, 예컨대 25 내지 40℃ 및 pH 조건, 예컨대 pH 6 내지 8의 조건 하에서 배양될 수 있다. 이러한 조건 하에서 재조합 미생물 또는 숙주세포로부터 전술한 바와 같은 지의류 유래의 대사산물을 높은 수율로 대량 생산할 수 있다.

[0059] 상기 재조합 미생물 또는 숙주세포는 전술한 바와 같이 지의류 유래 아트라노린 생합성 유전자의 도입 효율과 발현 효율이 높은 미생물 또는 숙주세포로서 당업계에 공지된 미생물 또는 숙주세포를 제한없이 사용할 수 있다. 구체적으로 상기 재조합 미생물은 곰팡이일 수 있고, 보다 구체적으로 아스코키타 라비에이(*Ascochyta rabiei*)일 수 있다. 보다 더 구체적으로, 상기 재조합 미생물은 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)일 수 있다.

[0060] 또한, 상기 재조합 미생물 또는 숙주세포의 배양(또는 발효)에 사용되는 배지는 특정한 미생물 균주 또는 숙주세포의 요구조건을 적절하게 만족시키는 것을 적절히 선택할 수 있다. 상기 배지는 다양한 탄소원, 질소원, 인원 및 미량원소 성분을 포함할 수 있다. 배지 내 탄소원으로는 글루코즈, 사카로즈, 락토즈, 프락토즈, 말토즈, 전분, 셀룰로즈와 같은 당 및 탄수화물, 대두유, 해바라기유, 피마자유, 코코넛유 등과 같은 오일 및 지방, 팔미트산, 스테아린산, 리놀레산과 같은 지방산, 글리세롤, 에탄올과 같은 알코올, 아세트산과 같은 유기산을 예

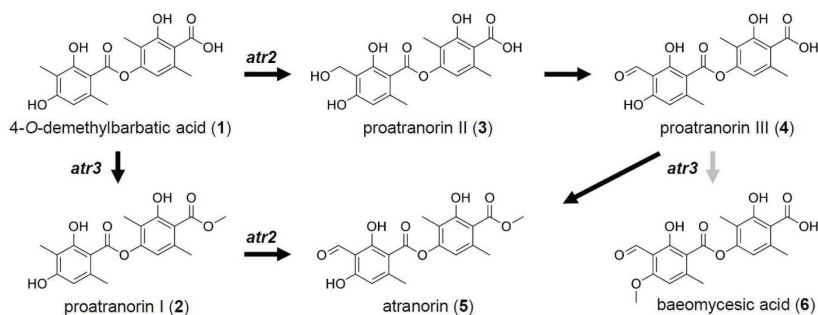
시할 수 있으나, 이에 제한되는 것은 아니다. 이들 물질은 개별적으로 또는 혼합물로서 사용될 수 있다. 상기 탄소원은 글리세롤, 포도당 또는 이들의 조합일 수 있다. 배지 내 질소원으로는 펩톤, 효모 추출물, 육즙, 맥아 추출물, 옥수수 침지액, 대두밀 및 요소 또는 무기 화합물, 예를 들면 황산암모늄, 염화암모늄, 인산암모늄, 탄산암모늄 및 질산암모늄을 예시할 수 있으나, 이에 제한되는 것은 아니다. 질소원 또한 개별적으로 또는 혼합물로서 사용할 수 있다. 배지 내 인원으로서는 인산수소소칼륨 또는 인산수소이칼륨 또는 상응하는 나트륨-함유 염을 예시할 수 있으나, 이에 제한되는 것은 아니다. 또한, 배양 배지는 성장에 필요한 황산마그네슘 또는 황산철과 같은 금속염을 포함하거나, 아미노산 및 비타민과 같은 필수 성장 물질을 포함할 수 있으나, 이에 제한되는 것은 아니다. 상술된 원료들은 배양 과정에서 배양물에 적절한 방식에 의해 회분식으로 또는 연속식으로 첨가될 수 있다.

[0061] 또한, 필요에 따라, 수산화나트륨, 수산화칼륨, 암모니아와 같은 기초 화합물 또는 인산 또는 황산과 같은 산 화합물을 적절한 방식으로 사용하여 배양물의 pH를 조절할 수 있다. 또한, 지방산 폴리글리콜 에스테르와 같은 소포제를 사용하여 기포 생성을 억제할 수 있다. 배양물의 온도는 통상 20℃ 내지 45℃, 또는 25℃ 내지 40℃를 유지할 수 있다. 배양은 원하는 대사산물, 예컨대 4-O-디메틸바르바트산, 프로아트라노린 I, 프로아트라노린 II, 프로아트라노린 III, 아트라노린 및 베오미세스산으로 이루어진 군으로부터 하나 이상 선택되는 지의류 유래의 2차 대사산물의 생산량이 최대로 얻어질 때까지 계속될 수 있다.

[0062] 본 명세서에서 제공되는 지의류 유래의 대사산물의 생산 방법은, 상기한 배양 단계 이후에, 생산된 지의류 유래의 2차 대사산물을 회수 또는 분리하는 단계를 추가로 포함할 수 있다. 재조합 균주 또는 이의 배양물로부터 생산된 지의류 유래의 2차 대사산물을 회수 또는 분리하는 방법은 당업계에 널리 알려져 있는 방법 중에서 선택될 수 있다. 예컨대, 4-O-디메틸바르바트산, 프로아트라노린 I, 프로아트라노린 II, 프로아트라노린 III, 아트라노린 또는 베오미세스산의 회수 방법의 예로서, 원심분리, 초음파파쇄, 여과, 이온교환 크로마토그래피, 고성능 액체 크로마토그래피(high performance liquid chromatography: HPLC), 가스 크로마토그래피(gas chromatography: GC) 등의 방법이 있으나, 이들 예에 한정되는 것은 아니다.

[0063] 본 명세서에서 제공되는 지의류 유래의 대사산물은 전술한 아트라노린 생합성 유전자들 중 1종 이상을 선택적으로 도입하여 형질전환시킨 재조합 미생물로부터 생산될 수 있다(하기 [반응식 1] 참조).

[0064] [반응식 1]



[0065] 본 발명의 일 구체예에서, 상기 서열번호 1의 염기서열로 이루어진 atr1 유전자가 도입된 재조합 미생물을 소정의 조건 하에서 배양함으로써 4-O-디메틸바르바트산을 생산할 수 있다. 이때, 재조합 미생물은 이에 제한되지는 않으나 식물 병원성 곰팡이 아스코키타 라비에이(*As. rabiei*)일 수 있다.

[0066] 본 발명의 일 구체예에서, 상기 서열번호 1의 염기서열로 이루어진 atr1 유전자 및 서열번호 5의 염기서열로 이루어진 atr3 유전자가 도입된 재조합 미생물을 소정의 조건 하에서 배양함으로써 프로아트라노린 I을 생산할 수 있다. 이때, 재조합 미생물은 이에 제한되지는 않으나 식물 병원성 곰팡이 아스코키타 라비에이(*As. rabiei*)일 수 있다.

[0067] 본 발명의 일 구체예에서, 상기 상기 서열번호 1의 염기서열로 이루어진 atr1 유전자 및 서열번호 3의 염기서열로 이루어진 atr2 유전자가 도입된 재조합 미생물을 소정의 조건 하에서 배양함으로써 프로아트라노린 II 및 프로아트라노린 III을 생산할 수 있다. 이때, 재조합 미생물은 이에 제한되지는 않으나 식물 병원성 곰팡이 아스코키타 라비에이(*As. rabiei*)일 수 있다.

[0068] 본 발명의 일 구체예에서, 상기 상기 서열번호 1의 염기서열로 이루어진 atr1 유전자, 서열번호 3의 염기서열로 이루어진 atr2 유전자 및 서열번호 5의 염기서열로 이루어진 atr3 유전자가 도입된 재조합 미생물을 소정의 조

건 하에서 배양함으로써 아트라노린을 생산할 수 있다. 이때, 재조합 미생물은 이에 제한되지는 않으나 식물 병원성 곰팡이 아스코키타 라비에이(*As. rabiei*)일 수 있다.

[0071] **아트라노린 생산용 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주**

[0072] 또 다른 하나의 양태로서, 본 발명은 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)를 제공한다.

[0073] 본 명세서에 사용된 용어 "아스코키타 라비에이(*Ascochyta rabiei*)"는 병아리 콩 역병을 유발하는 식물 병원성 곰팡이균으로, 전술한 지의류 유래의 아트라노린 생합성 유전자를 도입하는 데 사용된 이종의 숙주 미생물이다.

[0074] 상기 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)는 국립농업과학원 농업미생물은행(KACC)에 2021년 5월 28일자로 기탁번호 KACC 83048BP로 기탁된 아트라노린 생산용 균주이다.

[0075] 구체적으로, 상기 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주는 전술한 서열번호 1, 서열번호 3 및 서열번호 5로 이루어진 균으로부터 선택되는 1종 이상으로 이루어진, 아트라노린 생합성 유전자가 도입된 균주로, 아트라노린을 생산할 수 있다. 상기 균주로부터 생산된 아트라노린은 약학 조성물, 식품 조성물, 건강기능식품, 사료 조성물 등에 다양하게 활용 및 산업화할 수 있다.

[0077] **아스코키타 라비에이 ATR-11 균주를 이용한 아트라노린 생산 방법**

[0078] 또 다른 하나의 양태로서, 본 발명은 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)를 배양하는 단계를 포함하는, 아트라노린 생산 방법을 제공한다.

[0079] 상기 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)는 전술한 바와 같다.

[0080] 상기 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)의 배양 조건은 특별히 이에 제한되지 않으나, 15 내지 25℃의 온도, 배양 시간 7일 내지 21일일 수 있고, 바람직하게는 18 내지 22℃의 온도, 배양 시간 14일 내지 18일일 수 있다.

[0081] 또한, 상기 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)의 배양은 상기 균주가 효율적으로 이용할 수 있는 탄소원, 질소원, 무기염 등을 포함하는 천연 배지 또는 합성 배지를 사용할 수 있다. 사용될 수 있는 탄소원은 글루코오스, 프럭토오스, 수크로오스, 콩가루, 땅콩가루, 밀가루, 옥수수가루, 텍스트린, 콘밀, 과당 시럽과 같은 탄수화물; 녹말, 녹말의 가수분해물; 아세트산 및 프로피온산과 같은 유기산; 에탄올, 프로판올, 글리세롤과 같은 알코올; 콩기름, 올리브기름, 카놀라유, 땅콩기름, 생선기름 등의 기름류 등을 포함한다. 질소원은 암모니아; 염화암모늄, 암모늄설페이트, 암모늄아세테이트 및 암모늄포스페이트와 같은 무기산 또는 유기산의 암모늄염; 펩톤, 육추출물(meat extract), 이스트추출물, 옥수수 침지액, 카제인 가수분해물, 대두추출물, 대두가수분해물; 다양한 발효된 세포 및 이들의 분해물 등을 포함한다. 아미노산으로, 글루타민산나트륨, 메티오닌, 라이신, 루이신, 시스테인, 발린 등을 포함한다. 무기염은 포타슘디하이드로젠 포스페이트, 다이포타슘하이드로젠 포스페이트, 마그네슘 포스페이트, 마그네슘 설페이트, 소듐 클로라이드, 망간 설페이트, 구리 설페이트, 칼슘 카보네이트 등을 포함한다.

[0082] 상기 배지의 pH는 배양 중에서 바람직하게는 3.0 내지 9.0의 범위를 유지한다. 배지의 pH는 무기 또는 유기산, 알칼리 용액, 우레아, 칼슘 카보네이트, 암모니아 등으로 조절할 수 있다.

[0083] 상기 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)를 배양한 후, 이로부터 아트라노린을 분리, 정제 및 회수하기 위해서는 공지의 다양한 방법을 이용할 수 있다. 예를 들면 균체를 원심분리를 통해 제거한 후, 염색, 에탄올 침전, 한외여과막, 겔 여과 크로마토그래피, 이온교환 칼럼 크로마토그래피 등의 공지된 적절한 방법이나 이들을 조합하여 분리 정제할 수 있다.

[0084] 상기 아스코키타 라비에이(*Ascochyta rabiei*) ATR-11 균주(기탁번호: KACC83048BP)를 활용한 아트라노린의 생산은 실험실 규모에서 뿐만 아니라, 대규모 생산에도 적합하므로 상업적 활용에 용이하다.

[0085] 상기 아트라노린은 이에 제한되지는 않으나, 지의류 스테레오카울론 알피넘(*Stereocaulon alpinum*)으로부터 유래한 아트라노린일 수 있다.

[0087] 이하, 실시예를 통하여 본 발명의 구성 및 효과를 더욱 상세히 설명하고자 한다. 이들 실시예는 오로지 본 발명을 예시하기 위한 것일 뿐, 본 발명의 범위가 이들 실시예에 의해 한정되는 것은 아니다.

[0089] 실시예 1: 계통 분석

[0090] 계놈 시퀀싱된 29종의 지의류-형성 곰팡이(Lichen-forming fungi, LFF)와 내생 곰팡이(endophytic fungus)인 시아노테르멜라 아스테리스(*Cyanoderrella asteris*)의 계통발생을 유추하기 위해 본 발명자들은 병합-기반 계통 발생학 접근법을 추구하였다. LFF에 대한 배양, 계놈 시퀀싱 및 주석 절차는 하기 실시예 8에 설명되어 있다. 30개의 주석이 달린 계놈에서 추론된 단백질 서열의 단일-사본 오르토로그 클러스터(Single-copy ortholog cluster, SCO)는 인플레이션 계수가 2.5인 OrthoMCL(v2.0.9)을 사용하여 식별되었다. 393개의 SCO 각각에 대해, '자동' 설정과 함께 MAFFT(v7.310)를 사용하여 단백질 서열을 정렬하였으며, 매개변수: '-b4=5'과 함께 Gblock(v0.91b)을 사용하여 잘못 정렬된 영역에 대해 결과 393개의 다중 서열 정렬을 트리밍하였다. RAXML 프로그램(v82)은 각 다중 서열 정렬에 대해 100개의 최대 우도(likelihood) 트리를 계산하는 데 사용되었다. 병합-기반 트리를 생성하기 위해 두 개의 다중 위치 부트스트래핑 옵션(사이트 전용 리샘플링 및 유전자/사이트 리샘플링)이 있고 부트스트랩 옵션이 없는 ASTRAL-III(v5.7.4)를 사용하였다. 모든 트리는 동일한 토폴로지(topology)를 보여주었다.

[0092] 실시예 2: 생합성 유전자 클러스터 패밀리 분석

[0093] 30종의 계놈에서 BGC 식별을 위해, 계놈 어셈블리 및 주석 파일은 '--최소' 매개변수 설정으로 antiSMASH 프로그램(v5.0+)에 의해 처리되었다. 6종의 클라도니아속(*Cladonia* spp.) 및 남극 지의류 스테레오카울론 알피눔(*Stereocaulon alpinum*, *S. alpinum*)에서 발견된 BGC의 유전적 복제 제거를 위해 본 발명자들은 MIBiG 데이터 베이스(v1.4)를 참조하여 BiG-SCAPE 프로그램을 사용하였다. 하나 이상의 반복 유형 I PKS를 포함하는 BGC를 분석하기 위해, 반복 유형 I PKS(PF00109 및 PF02801)와 관련된 Pfam 도메인만 포함하도록 구성 파일 'domain_whitelist.txt'를 수정하고 선택적 인수 '-믹스' 및 '-- 하이브리드-오프'로 BiG-SCAPE 프로그램을 실행하였다. 도메인 유형의 자카드(Jaccard) 인덱스, 도메인 시퀀스 유사성 및 도메인 인접성 인덱스를 기반으로 BiG-SCAPE 프로그램은 더 작은 값이 더 큰 BGC 유사성을 나타내는 클러스터의 페어별 조합 간의 유사성 매트릭스를 계산하였다. 본 발명자들은 0.3-0.8의 엣지-길이 컷오프(edge-length cutoff)를 0.1 단계로 사용하여 BGC 네트워크를 평가하였고, 6종의 클라도니아속(*Cladonia* spp.) 및 *S. alpinum*에서 GCF의 대표로서 0.5의 컷오프(cutoff) 값을 사용하는 네트워크를 고려하였다. PKS 패밀리에 대한 개별 네트워크는 Python package NetworkX(v2.5)를 사용하여 시각화되었으며, BiG-SCAPE에 의해 계산된 BGC 쌍 간의 유사도 제공으로 엣지-너비에 가중치가 부여되었다. 본 발명자들은 지의류 물질 생산을 담당하는 PKS 패밀리를 포함하는 BGC의 다중-위치, 대략 최대-우도, 계통수를 생성하는 CORASON 프로그램을 사용하였다. PKS23 패밀리의 경우 *C. 랑기페리나*(*C. rangiferina*)의 OMT(Crg06815)를 쿼리 유전자로 사용하여 30개의 시퀀싱된 종에서 검출된 총 1,527개의 BGC에서 상동성 BGC를 검색하였다. PKS1 패밀리의 경우 *C. 랑기페리나*(*C. rangiferina*)의 GAL4 유형 전사 인자(유전자 ID: Crg07068)를 쿼리 유전자로 사용하였다.

[0094] 실시예 3: 곰팡이 NR-PKS의 계통발생 분석

[0095] 반복 유형 I PKS의 KS 도메인은 진화적으로 보존된 것으로 간주되어, 전체 PKS의 유사성에 대한 프록시 역할을 할 수 있다. 본 발명자들은 6종의 클라도니아속(*Cladonia* spp.) 및 *S. alpinum*의 계놈에서 총 242개의 PKS를 확인하였으며, 그 중 4개의 PKS(Cgr01615, Cgr03964, Cgr08611 및 Cmt10189)에 KS 도메인이 누락되었다. KS 도메인 서열은 온라인 도구인 NaPDos를 사용하여 238개의 PKS에서 추출하고 MUSCLE(v3.8.31)을 사용하여 정렬하였다. 클러스터링 분석을 위해, BuddySuite 프로그램의 AlignBuddy 기능을 사용하여 선택적 인수인 '-pi'를 사용하여 238개 PKS의 KS 도메인에 대한 전체-대-전체(all-versus-all) 유사성 매트릭스를 계산하였다. R package Superheat를 사용하여 *k*-평균값에 의해 클러스터링된 KS 도메인의 유사도 백분율을 보여주는 히트맵이 생성되었다. 곰팡이 NR351 PKS 계통발생의 경우, 본 발명자들은 6종의 클라도니아속(*Cladonia* spp.) 및 *S. alpinum*에서 발견된 103개의 NR-PKS의 KS 및 PT 도메인의 연결된 단백질 서열과 지의류가 없는 곰팡이의 알려진 화합물에 연결된 82개의 NR-PKS를 사용하였다. 본 발명자들은 처음에 6종의 클라도니아속(*Cladonia* spp.) 및 *S. alpinum*에서 106개의 NR-PKS를 확인하였으며, 여기에는 현재 계놈 어셈블리(유사유전자)에서 전체 서열을 안정적으로 정

의할 수 없는 7개의 NR-PKS가 포함된다. 이러한 부분 PKS 중, 3개의 NR-PKS(Cbo04702, Cma06590 및 Cmt06606)는 KS 또는 PT 도메인이 없었고, 계통발생 분석에서 제외되었다. 지의류 NR-PKS의 PT 도메인은 이전에 특성화된 PKS의 도메인과 정렬하여 식별되었다. 페니실리움 익스판숨(*Penicillium expansum*)에서 파툴린(UniProtKB: A0A075TRC0) 생합성을 담당하는 6-메틸살리실산 합성 효소(6-methylsalicylic acid synthase, 6MSAS)는 곰팡이 NR-PKS 계통발생에 대한 외부 그룹으로 설정되었다. 또한, 4개의 클라도니아속(*Cladonia* spp.)에서 발견된 4개의 6MSAS가 분석에 포함되었다(Cbo07291, Cgr05254, Cmt10005 및 Cuc03485). KS 및 PT 도메인의 단백질 서열은 MAFFT(v7.310)를 사용하여 '자동' 설정으로 정렬하고, 각 도메인의 거짓된 서열 또는 잘못 정렬된 영역은 'gappyout' 매개변수로 trimAl 프로그램(v1.2)을 사용하여 트리밍하였다. KS 및 PT 도메인에 대한 결과 다중 서열 정렬이 FASconCAT-G(v1.04)와 연결되었다. 연결된 서열에서 RAxML(v8.2)을 사용하여 매개변수 설정이 '-m PROTGAMMAWAG'인 사이트의 대체율에 대한 감마 분포를 사용하여 최대-우도 트리를 계산하였다. 노드 지원은 1,000개의 부트스트랩 복제에 의해 평가되었다. 최종 트리는 6MSAS 아웃그룹에 뿌리를 두고 iTOL(v5.7)로 주석을 달았다.

[0096] 실시예 4: 아스코키타 라비에이(*Ascochyta rabiei*)에서 "클린 숙주"의 생성

[0097] 분할-마커 전략을 사용하여 솔라나피론-마이너스 돌연변이체("클린 숙주")를 생성하였다. *sol1*의 다운스트림 영역(1,208 bp) 및 *sol3*의 업스트림 영역(1,459 bp)은 각각 L5/L3 및 R5/R3 프라이머 쌍을 사용하여 *As. rabiei* 분리주 AR628의 gDNA로부터 증폭되었다. 하이그로마이신 포스포트랜스퍼라제 유전자(hph) 카세트는 프라이머 쌍 HYG5/HYG3을 사용하여 pCB1004 플라스미드로부터 증폭되었다. L3 및 R5 프라이머는 각각 *hph* 카세트의 3' 및 5' 말단에 상보적인 27개의 nt long 오버행 서열을 가지므로, 증폭된 *sol1* 다운스트림 및 *sol3* 업스트림 영역이 중첩 확장 PCR에 의해 *hph* 카세트에 융합될 수 있다. 마지막으로, N5/HY-R 및 YG-F/N3 프라이머 쌍을 사용하여 융합된 구성에서 분할 마커 구성을 증폭하였다. 분할 마커 구성물은 Hallen-Adams 등에 기술된 바와 같이 폴리에틸렌 글리콜-매개 유전자 변형에 의해 *As. rabiei*의 원형질체에 도입되었다. 유전자 교체에 사용된 프라이머는 하기 표 1에 나열되어 있다.

[0098] 본 발명에 사용된 프라이머 정보를 하기 표 1에 정리하여 나타내었다.

표 1

[0099]

프라이머	서열 (5'-3')	서열번호	설명
Sol1PpDS33_fwd	AGATCTCGGCACGTTTTGATCGCAAAG	7	pDS33으로 <i>sol1</i> 프로모터 클로닝
Sol1PpDS33_rev	AAGCTTAGTAATATCTGGCGATGAAGCGTCAG	8	pDS33으로 <i>sol1</i> 프로모터 클로닝
Inf_Sta046440_fwd	CAGATATTACTAAGCTCATGGCGTCTAACCATGAGGAG	9	pDS35로 PKS23 클로닝
Inf_Sta046440_rev	GTAACGTTAAGTGGATCCTTACAACATAAGCTCTCAATCAACCC	10	pDS35로 PKS23 클로닝
Sta046440_check1F	GGACGATGAGAACGTCCTAG	11	PKS23 시퀀싱
Sta046440_check2F	CGAGTCAGGTGATCCAGTG	12	PKS23 시퀀싱
Sta046440_check3F	CCAATCATACTGAGCGCTG	13	PKS23 시퀀싱
Sta046440_check4F	CACATGTCTGTGGCTTGAG	14	PKS23 시퀀싱
Sta046440_check5F	CAACTCCGGTGATGATTG	15	PKS23 시퀀싱
Sta046440_check6F	CAACCCTTCCACATATTGAG	16	PKS23 시퀀싱
pII99_sol15_fwd	GAATTTCAGAGCGACAGTGAAAGC	17	pII99로 <i>sol15</i> 프로모터 클로닝
pII99_sol15_rev	AGATCTAGCTAGCTTGGCGTGATGTCAGTCAATGCTG	18	pII99로 <i>sol15</i> 프로모터 클로닝
pII99_tef1a_fwd	GAATTCCAGCCGAGACAGCAGAATCAC	19	pII99로 <i>tef1a</i> 프로모터 클로닝
pII99_tef1a_rev	AGATCTAGCTAGCGTGTATGTTTGTGGAATAAAAAGGG	20	pII99로 <i>tef1a</i> 프로모터 클로닝
Inf95_Sta04642_fwd	ATCAGCCAAGCTAGCAGCATGACTTCCGTTGATACAATG	21	pII95로 <i>atr3</i> 클로닝
Inf98_Sta04642_fwd	AAACATAACACGCTAGCAGCATGACTTCCGTTGATACAATG	22	pII98로 <i>atr3</i> 클로닝
Inf99_Sta04642_rev	GGAGACCGGCAGATCCCGTCACATGTCCGTCATAATAAC	23	pII95 및 pII98로 <i>atr3</i> 클로닝
Inf98_Sta04643_fwd	AAACATAACACGCTAGAACCATGGCTCTCCTAGACACAATTG	24	pII95로 <i>atr2</i> 클로닝
Inf95_Sta04643_fwd	ATCAGCCAAGCTAGAACCATGGCTCTCCTAGACACAATTG	25	pII98로 <i>atr2</i> 클로닝

Inf99_Sta04643_rev	GGAGACCGGCAGATCGAGGAGCAAAGGAACAAG GAACAG	26	pII95 및 pII98로 <i>atr2</i> 클로닝
pII99_MK5	CGTCAAGAGACCTACGAGACTG	27	플라스미드 <i>tefla::atr2</i> /pII98에서 <i>nptII</i> 를 <i>ble</i> 로 교체
pII99_MK3	GCTATACTTCTAGGTCTTGAAGAG	28	플라스미드 <i>tefla::atr2</i> /pII98에서 <i>nptII</i> 를 <i>ble</i> 로 교체
Inf_BLE_pII99_fwd	GTAGGTCTCTTGACGCATTAAGACCTCAGCGCT AGTGG	29	플라스미드 <i>tefla::atr2</i> /pII98에서 <i>nptII</i> 를 <i>ble</i> 로 교체
Inf_BLE_pII99_rev	GACCTAGAAGTATAGCGGTGTACGGAGCATTC ACTAGG	30	플라스미드 <i>tefla::atr2</i> /pII98에서 <i>nptII</i> 를 <i>ble</i> 로 교체
L5	GAGGTCTTCACACCACAAGTTCG	31	"클린 숙주" 생성
L3	GGCAAAGGAATAGAGTAGATGCCGACGGGCTC TTGCGTAGTGGTACAGATG	32	"클린 숙주" 생성
R5	GTTGACCTCCACTAGCTCCAGCCAAGCGAATAG GTAACAAAGCCAGCCGAG	33	"클린 숙주" 생성
R3	GAGAATTGCGGCGCAGGATGTTC	34	"클린 숙주" 생성
N5	CTGACTTTCTTCTTGACGCCCTGC	35	"클린 숙주" 생성
N3	CTAACGATGTCGTTAGTCGCCTTG	36	"클린 숙주" 생성
HYG-F	GCTTGGCTGGAGCTAGTGGAG	37	"클린 숙주" 생성
HYG-R	CGGTCGGCATCTACTCTATTCCTT	38	"클린 숙주" 생성
YG-F	CGATGTAGGAGGCGTGGATATGTC	39	"클린 숙주" 생성
HY-R	TGTAGTGTATTGACCGATTCTTGCG	40	"클린 숙주" 생성
P1_fwd	GGACGATGAGAACGTCCTAG	41	RT-PCR 분석
P1_rev	GAGAGTTCCTCATGCTCGTC	42	RT-PCR 분석
P2_fwd	CATCGACGACACGGAGATACTG	43	RT-PCR 분석
P2_rev	GACCGTGAGATTCTTTGAAGAGG	44	RT-PCR 분석
Actin1_fwd	CAATGGTTTCGGGTATGTGCAAG	45	RT-PCR 분석
Actin1_rev	GAAGAGCGAAACCCTCGTAGAT	46	RT-PCR 분석

[0100] 실시예 5: *atr1* 및 RT-PCR 분석의 이종(Heterologous) 발현

[0101] "클린 숙주"에서 외래 유전자의 효율적인 발현을 위해, *sol1* 유전자 프로모터(pDS35), *sol5* 유전자 프로모터(pII95) 또는 번역 연장 인자 1 알파 유전자 프로모터(pII98)를 운반하는 발현 벡터를 구성하였다. 그 다음, *atr1*, *atr2* 및 *atr3* 유전자를 발현 벡터에 개별적으로 클로닝하였다. 발현 벡터 구성 및 클로닝 절차에 대한 자세한 내용은 실시예 9 및 10에 기술되어 있다. *atr1*의 이종 발현을 위해, *sol1::atr1*/pDS35 플라스미드(5-10 μ g)를 사용하여 "클린 숙주"를 형질전환하였다. 추정 형질전환체를 100 μ g/ml의 노르세오티리신 선택배양(Sta04644-T16 및 Sta04644-T25로 명명된 2개의 형질전환체는 선택배양에 내성을 보였고, RNA 추출을 위해 나일론 멤브레인으로 덮인 PDA에서 계대배양하였다. 배양 2주 후, PDA에서 자라는 균사체를 나일론 멤브레인으로부터 편날 면도날로 긁어내어 총 RNA를 추출하였다. 제조업체의 지침에 따라 TRIzol 시약(Thermo Fisher Scientific)을 사용하여 추가 추출 단계와 함께, 액체 질소에서 분쇄된 균사로부터 총 RNA를 추출하였다. 추가 추출 단계는 1회의 페놀(pH 4.6)-클로로포름-이소아밀 알코올(25:24:1) 추출에 뒤이어 초기 TRIzol-클로로포름 상 분리 후 1회의 클로로포름 추출 단계 RNA 펠렛을 88 μ L의 뉴클레아제가 없는 물에 용해시키고, DNase 처리로 게놈 DNA를 절단하여 수행하였다(Qiagen). 이후, RNA Clean & Concentrator (Zymo research)를 사용하여 RNA 샘플을 농축하였다. 본 발명자들은 RT-PCR 분석에서 4개의 인트론으로 *atr1*의 발현과 인트론 접합을 확인하였다. 두 개의 프라이머 세트가 설계되었다: (i) *atr1* 코딩 서열의 5' 영역에 위치한 첫 번째 인트론의 측면 영역을 증폭하고, (ii) *atr1* 코딩 서열의 3' 영역에서 두 번째, 세 번째 및 네 번째(마지막) 인트론을 포함하는 영역을 증폭하였다. OneStep RT-PCR 키트(Qiagen)를 사용하여, 200 나노그램의 총 RNA를 역전사하고 *atr1* 및 액틴1(양성 대조군)을 증폭하였다. RT-PCR 분석 및 플라스미드 구성에 사용된 프라이머는 상기 표 1에 나열되어 있다.

[0102] 실시예 6: 아트라노린을 생산하는 이종 숙주의 생성

[0103] *atr1*-발현 균주(Sta04644-T25)에 맞춤형 효소를 도입하기 위해, 4개의 생성된 플라스미드(*sol5::atr2*/pII95, *tefla::atr2*/pII98, *sol5::atr3*/pII95 및 *tefla::atr3*/pII98)를 Sta04644-T25 균주로 개별적으로 형질전환시키고, 추정 형질전환체를 200 μ g/ml의 G418 디설파이트(Sigma-Aldrich)를 함유하는 PDA 상에서 계대배양하였다. 대사산물 식별을 위해 G418 디설파이트에 내성이 있는 형질전환체를 선택하였다. *sol1::*

atr1/pDS35 및 *sol5:: atr3/pII95*를 모두 보유하는 균주를 아트라노린-생산 균주의 생성에 사용하였고, 이는 아트라노린의 직접적인 전구체인 화합물 **2**(프로아트라노린 I)의 가장 큰 생산을 나타내었다. G418 디설페이트에 내성을 부여하는 네오마이신 포스포트랜스퍼라제 II 유전자(*nptII*) 카세트가 이미 화합물 **2**를 생성하는 균주에 통합되었기 때문에, 본 발명자들은 *tef1a::atr2/pII98* 플라스미드의 *nptII* 카세트를 블레오마이신 내성 단백질 유전자(*ble*)로 교체하였으며, *tef1a::atr2/pII98* 플라스미드는 프라이머 쌍 'Inf_BLE_pII99_fwd' 및 'Inf_BLE_pII99_rev'를 사용하여 증폭되었고, *ble*는 pAC1750에서 증폭되었다. 이후, In-Fusion HD 클로닝 키트(Takara)를 사용하여 두 PCR 산물을 융합하였다. 생성된 플라스미드를 균주 생산 화합물 **2**로 형질전환하고, 추정 형질전환체를 200 µg/ml의 제오신(ThermoFisher Scientific에서 유통하는 블레오마이신 패밀리의 구성원)을 함유하는 PDA에서 계대배양하였다. 아트라노린을 생산하는 형질전환체를 선별하기 위해, 제오신 내성 형질전환체의 화학적 프로파일은 계능 염기서열 LFF가 원래 분리된 *S. alpinum* 바우처 표본(국립수목원 등록번호: KHL0017342)과 비교하였다. HPLC, LC-MS/MS 및 배양 추출물 및 정제된 화합물에 대한 NMR 분광 분석에 대한 자세한 방법론은 실시예 11 내지 13에 자세히 기술되어 있다.

[0104] 실시예 7: 데이터 유효성

[0105] 이 전체 계능 샷건 프로젝트는 *C. borealis*(*C. borealis*), 파멜리아 cf. 스퀴로사(*Parmelia* cf. *squarrosa*) 및 *S. 알피눔*(*S. alpinum*)에 대해 각각 JAFK000000000, JAEUB000000000 및 JAEUBA000000000 (바이오프로젝트: PRJNA693578, PRJNA693575, 및 PRJNA693574) 수탁하에 DDBJ/ENA/GenBank에 기탁되었다. LC-MS/MS 원 데이터는 MSV000087081이라는 등록번호로 MassIVE 데이터베이스(<http://massive.ucsd.edu>)에 기탁되었다.

[0106] 실시예 8: 계능 시퀀싱 및 주석

[0107] LFF는 남극 대륙 킹 조지 섬(S62° 13'51" W58° 46'32")에서 채집한 *C. borealis*(*C. borealis*) 및 *S. 알피눔*(*S. alpinum*)과 한국 덕유산(N35° 51'24" E127° 44'54")에서 채집한 파멜리아 cf. 스퀴로사(*Parmelia* cf. *squarrosa*)에서 분리되었다. LFF는 맥아 추출물 한천 배지(BD Biosciences)에서 3-6개월 동안 배양되었다. 표준 페놀/클로로포름 추출 방법을 사용하여 계능 DNA를 추출하고, Illumina HiSeq2000 또는 NovaSeq6000 기기에서 시퀀싱하여 150개의 뉴클레오타이드 페어-엔드 리드를 생성하였다. Illumina 어댑터를 트리밍하고 판독 품질을 평가하였으며, 콘티그(contig)를 Macrogen Inc.(Seoul, South Korea)에서 조립하였다. NCBI 데이터베이스에서 검색된 계능 어셈블리는 GenSAS (v.6.0) 주석 파이프라인을 사용하여 주석을 달았다. 간단히 말해서, DNA 소스를 'Fungi'로 설정하여 RepeatModeler (v1.0.11) 및 RepeatMasker (v4.0.7) (www.repeatmasker.org)를 사용하여 복잡성이 낮은 영역과 반복을 마스킹하였다. *ab initio* 유전자 예측이 하기 도구를 사용하여 수행된 마스킹된 컨센서스 시퀀스가 생성되었다: (i) Augustus (v3.3.1), *A. 니돌란스*(*A. nidulans*)를 트레이닝된 유기체로 선택, (ii) GeneMarkES (v4.33), (iii) 인간 및 기타 척추동물에 대한 매개변수 설정을 사용하는 Genscan (v1.0) 및 (iv) GlimmerM (v2.5.1), *아스페르길루스*(*Aspergillus*)를 트레이닝된 유기체로 선택. 상동성-기반 예측의 경우, (v) BLAST+ (v2.7.1) 및 (vi) DIAMOND (v0.9.22)를 사용하여 곰팡이(Fungi)에 대한 NCBI 참조 전사체 및 단백질 데이터베이스를 각각 검색하였다. EvidenceModeler (v06-25-2012)를 사용한 컨센서스 유전자 모델 예측의 경우, 전술한 독립형 유전자 예측에 다음과 같이 가중치를 부여하였다: (i)-5, (ii)-10, (iii)-1, (iv)-1, (v)-5 및 (vi)-5. 계능 주석은 Ascomycota_odb9 데이터 세트를 사용하여 BUSCO (v5b)로 평가되었다.

[0108] 실시예 9: PKS 발현을 위한 발현 벡터 구성

[0109] pDS23 플라스미드는 선택 마커로 노르세오트리신 아세틸트랜스퍼라제(nourseothricin acetyltransferase, *nat*) 유전자 카세트 및 *아스페르길루스* *니돌란스*(*Aspergillus nidulans*)에서 파생된 글리세르알데히드-3-인산디히드로게나아제(glyceraldehyde-3-phosphate dehydrogenase) 유전자(*G3PD*)의 구성적 프로모터의 제어 하에 향상된 녹색 형광 단백질(*eGFP*)을 전달한다. pDS23에서 *G3PD* 프로모터를 다중 복제 부위로 대체하기 위해, 플라스미드를 AvrII 및 HindIII 제한 효소와 여러 제한 효소 부위(NheI, AatII, SacI, PmeI, BglII 및 SacII)는 플라스미드의 AvrII와 HindIII 부위 사이에 클로닝되었다. 생성된 플라스미드를 pDS33으로 명명하였다. *As. rabiei*의 고유 프로모터를 운반하는 발현 벡터를 생성하기 위해, *As. rabiei* 분리물 AR628의 gDNA에서 증폭된 *sol1* (솔라나 피론 합성 효소 인코딩) 유전자 프로모터를 포함하는 1,270-bp DNA 단편을 BglII 및 HindIII로 미리 절단된 pDS33에 클로닝하여 pDS35 플라스미드를 생성하였다. *atr1*의 코딩 서열(Sta04644)은 프라이머 쌍 'Inf_Sta046440_fwd' 및 'Inf_Sta046440_rev' (7,829 bp)로 증폭되었으며, In-Fusion HD 클로닝 키트(Takara)를 사용하여 eGFP 대신 HindIII 및 BamHI로 선형화된 pDS35에 클로닝하여 플라스미드 *sol1::atr1/pDS35*를 생성하였다.

[0110] 실시예 10: 효소 발현 맞춤을 위한 발현 벡터 구축

[0111] pII99 플라스미드는 선택 마커로서 네오마이신 포스포트랜스퍼라제 II 유전자(*nptII*) 카세트를 운반한다. *sol5*의 프로모터(솔라나피론 생합성의 마지막 단계를 촉매하는 디엘스-알더레이즈(Diels-Alderase)를 인코딩함)를 운반하는 발현 벡터를 생성하기 위해, *As. rabiei* AR628 균주의 gDNA로부터 증폭된 *sol5* 프로모터를 포함하는 579-bp DNA 단편이 EcoRI 및 BglIII로 미리 절단된 pII99에 클로닝하여 pII95 플라스미드를 생성하였다. 추가 제한 효소 부위를 도입하기 위해, *sol5* 프로모터 증폭을 위한 역방향 프라이머가 NheI 부위를 갖도록 설계되었다(상기 표 1 참조). pII95 플라스미드를 NheI 및 BglIII로 절단하고, *S. 알피눔*(*S. alpinum*)의 gDNA로부터 증폭된 *atr3* (Sta04642) 또는 *atr2* (Sta04643)의 PCR 산물을 미리 절단된(precut) 플라스미드에 융합하여, 플라스미드 *sol5::atr3/pII95* 및 *sol5::atr2/pII95*를 생성하였다. 유사하게, *A. 니돌란스*(*A. nidulans*)로부터 유래된 번역 연장 인자 1 알파 유전자(translation elongation factor 1 alpha gene, *tef1a*)의 구성적 프로모터를 운반하는 발현 벡터를 생성하기 위해, 플라스미드 pII98은 pAC1750 플라스미드로부터 증폭된 *tef1a* 프로모터를 포함하는 880-bp DNA 단편을 pII99로 클로닝하여 생성되었으며, 코딩 서열과 *atr3* 및 *atr2*의 종결자 영역은 NheI 및 BglIII로 미리 절단된 pII98에 개별적으로 융합되어 플라스미드 *tef1a::atr3/pII98* 및 *tef1a::atr2/pII98*을 생성하였다.

[0112] 실시예 11: HPLC 분석 및 대사산물 정제

[0113] 대사산물 추출 및 프로파일링을 위해, 형질전환체를 2-3주 동안 PDA에서 성장시키고 곰팡이 콜로니를 포함하는 전체 한천을 펠트리 접시에서 잘라내고 50 ml 팔콘 튜브에 10 ml의 에틸 아세테이트에 담그었다. 30분 동안 초음파 처리한 후, 용액 1 ml의 분취량을 1.5 ml 미세원심분리 튜브로 옮기고, 증발 건조시키고, 100 μ l의 메탄올로 재구성하고, 고성능 액체 크로마토그래피(HPLC) 분석을 수행하였다. *S. 알피눔*(*S. alpinum*) 바우처 표본(20-30 mg) 조각을 2 ml 미세원심분리기 튜브에서 300 μ l의 아세톤에 담그었다. 30분 동안 초음파 처리한 후, 샘플을 HPLC에 주입하기 전에 0.4 mm 주사기 멤브레인을 통해 여과하였다. Prominence Modular LC-20A HPLC 기기(Shimadzu)를 사용하여 형질전환체의 화학적 프로파일링을 수행하였다. 샘플은 YMC-Pack ODS-A (컬럼 크기, 150 $\times$ 4.6 mm; 입자 크기, 5 μ m; 기공 크기, 12 nm; 40 $^{\circ}$ C)를 사용하고, 254 nm에서 포토다이오드 어레이 검출기(SPD-M20A, 범위 180-700 nm)를 사용하여 분석하였다. 이동상은 펌프 A의 경우 증류수/트리플루오로아세트산(99.9:0.1, v/v)과 펌프 B의 경우 메탄올/트리플루오로아세트산(99.9:0.1, v/v)으로 구성되었다. HPLC 분석은 1.0 ml/분: 0-30분, 20-100%; 30-40분, 100%; 40-52분, 펌프 B의 20% 유속에서 구배 프로그램을 사용하여 수행되었다. 본 발명자들은 지의류 표본에서 발견되는 대사산물을 식별하기 위해 HPLC 분석을 위해 지의류 물질 데이터베이스를 사용하였다. 구조적 식별을 위해, 화합물을 주요 대사산물로 생산하는 형질전환체의 배양 추출물에서 화합물 1, 2 및 4를 정제하였다. 20개의 PDA 배양물을 에틸 아세테이트로 3회 추출하였다. 배양 추출물을 회전 증발 농축기(Buchi)에서 증발 건조시키고, 메탄올로 재구성하고, 분취용 HPLC에 적용하였다. 254 nm에서 UV 모니터링과 함께 Kromasil C18-컬럼(250 $\times$ 10 mm, 5 μ m, 10 nm; 40 $^{\circ}$ C)을 사용하여 2.2 ml/분의 유속으로 HPLC에서 분리가 이루어졌다. 화합물 1을 66% 아세토니트릴을 사용하여 정제하였다(t_R = 9.86분; 9.5 mg); 2는 85% 아세토니트릴을 사용하여 정제되었다(t_R = 9.77분; 6.2 mg); 4를 80% 아세토니트릴을 사용하여 정제하였다(t_R = 9.60분; 2.1 mg). 용매를 0.1% 트리플루오로아세트산으로 완충하였다. 정제된 화합물을 NMR 분광 분석에 적용하였다.

[0114] 실시예 12: 생성물 식별을 위한 LC-MS/MS 분석

[0115] 배양 추출물을 메탄올로 20배 희석하고, 0.22 μ m PTFE 필터로 여과하였다. LC-MS 분석은 Waters VION IMS QTOF 질량 분석기(Waters Co.)에 연결된 Waters Acquity I-Class UPLC 시스템에서 수행되었으며, 이는 전자분무 이온화 인터페이스가 장착되어 있다. 이동상은 물 중 0.1% 포름산(펌프 A) 및 아세토니트릴(펌프 B)로 구성되었다. 0.3 ml/분의 일정한 유속에서 단계적 구배 방법을 사용하여 펌프 B의 20-100%(0-9분), 3분 세척 및 3분 재생 조건으로 컬럼을 용리하였다. 주입 부피는 2 μ l이었다. 컬럼 온도 및 샘플 오거나이저를 각각 40 $^{\circ}$ C 및 20 $^{\circ}$ C로 유지하였다. 탠덤 MS 분석은 m/z 50-1500 Da 범위의 음이온 모드 및 0.2초의 수집 시간을 갖는 데이터-독립 수집(MSE)에서 수행되었다. 전구체 이온의 검출을 위한 낮은 충돌 에너지는 6 eV로 설정되었고, 단편화를 위한 높은 충돌 에너지는 20-40 eV로 설정되었다. 이온화 조건은 다음과 같이 설정되었다: 모세관 전압은 2.5kV이고, 콘 전압은 40V이고, 소스 온도는 100 $^{\circ}$ C이고, 탈용매 온도는 250 $^{\circ}$ C이고, 콘 가스 유량은 0 l/h이고, 탈용매 가스 유량은 800 l/h이었다. 네블라이저 및 보조 가스는 고순도 질소를 사용하였고, 충돌 가스는 아르곤을 사용하였다. m/z 554.2615에서 류신 엔케팔린의 $[M-H]^{-}$ 이온은 질량 정확도 및 재현성을 보장하기 위해 잠금 질량으로 사용되었다. 형질전환체에서 검출된 주요 화합물의 MS/MS 스펙트럼 데이터 매칭은 MS-DIAL을 사용한 전처리 후 GNPS의 기능 기반-분자 네트워킹 워크플로에 의해 수행되었다.

[0116] 실시예 13: NMR 분광 데이터 요약

[0117] 실시예 13-1: 4-O-테메틸바르바르트산(화합물 1)

[0118] ^1H NMR (400 MHz, CD_3COCD_3): δ_{H} 11.65 (s, OH), 6.73 (s, 1H), 6.47 (s, 1H), 2.68 (s, 3H), 2.63 (s, 3H), 2.12 (각각 s, 3H x 2); ^{13}C NMR (100 MHz, CD_3COCD_3): δ_{C} 170.4, 164.0, 163.3, 160.6, 153.0, 140.9, 140.7, 116.7, 116.6, 111.3, 109.6, 109.2, 103.6, 23.8, 23.1, 8.6, 7.3; 카르복실기에 대한 ^{13}C NMR 신호는 관찰되지 않았거나 탄소 핵의 긴 이완 시간으로 인해 매우 낮은 강도일 수 있으며, 공개된 데이터에서 카르복실기의 δ_{C} 는 174.1이었다. ^1H NMR 및 ^{13}C NMR 데이터는 공개된 데이터와 비슷하였다. HRESIMS m/z 3450972 $[\text{M}-\text{H}]^-$ ($\text{C}_{18}\text{H}_{17}\text{O}_7$ 에 대한 계산치, 345.0974); MS/MS 스펙트럼은 GNPS 스펙트럼 라이브러리에 보관되어 있다.

[0119] 실시예 13-2: 프로아트라노린 I(화합물 2)

[0120] ^1H NMR (400 MHz, CDCl_3): δ_{H} 11.90(s, OH), 11.70(s, OH), 6.50(s, 1H), 6.29(s, 1H), 3.96(s, 3H), 2.60(s, 3H), 2.52(s, 3H), 2.12(s, 3H), 2.07(s, 3H); ^{13}C NMR (100 MHz, CDCl_3): δ_{C} 172.4, 170.3, 164.2, 162.9, 159.0, 152.6, 140.7, 139.8, 117.0, 116.4, 111.2, 110.1, 109.0, 104.2, 52.3, 24.6, 24.1, 9.4, 7.7; ^1H NMR (400 MHz, CDCl_3) 데이터는 공개된 데이터와 비슷하였다. ^{13}C NMR (100 MHz, CDCl_3) 데이터는 화합물 2에 하나 이상의 산소-메틸 탄소(δ_{C} 52.3)가 있다는 점을 제외하고는 4-O-테메틸바르바르트산 데이터와 비슷하였다. HRESIMS m/z 3591132 $[\text{M}-\text{H}]^-$ ($\text{C}_{19}\text{H}_{19}\text{O}_7$ 에 대한 계산치, 359.1131); MS/MS 스펙트럼은 GNPS 스펙트럼 라이브러리에 보관되어 있다.

[0121] 실시예 13-3: 프로아트라노린 III (화합물 4)

[0122] ^1H NMR (400 MHz, CDCl_3): δ_{H} 12.54 (s, OH), 12.45 (s, OH), 11.77(s, OH), 10.35(s, CHO), 6.54(s, 1H), 6.40(s, 1H), 2.68(s, 3H), 2.59(s, 3H), 2.08(s, 3H); ^1H NMR (400 MHz, CDCl_3) 데이터는 프로아트라노린 III (화합물 4)의 카르복실기에서 메틸기(δ_{H} 3.99)가 누락된 것을 제외하고는 아트라노린(화합물 5)의 공개된 데이터와 비슷하였다. HRESIMS m/z 3590771 $[\text{M}-\text{H}]^-$ ($\text{C}_{18}\text{H}_{15}\text{O}_8$ 에 대한 계산치, 359.0767); MS/MS 스펙트럼은 GNPS 스펙트럼 라이브러리에 보관되어 있다.

[0123] 실시예 14: 클라도니아속(*Cladonia*) PKS 패밀리와 유전적 복제 제거에 대한 주석

[0124] 실시예 14-1: 클라도니아속(*Cladonia* spp.)의 PKS 패밀리와

[0125] Big-SCAPE 프로그램을 사용하여, 적어도 하나의 반복적인 I형 PKS 유전자를 포함하는 45개의 유전자 클러스터 패밀리(GCF)가 확인되었다. PKS1, PKS2, PKS4, PKS5, PKS15 및 PKS16 패밀리의 생합성 유전자 클러스터(BGC)는 6종의 클라도니아(*Cladonia*) 종과 *S. alpinum*에서 보존되었고, PKS3 및 PKS10 패밀리의 BGC는 6개의 클라도니아 종에서 보존되었다. 이전에, Timsina 등은 클로도니아 클로로피아(*Cladonia chlorophaea*) 종 복합체(PKS1, PKS2, PKS3, PKS5, PKS7, PKS10, PKS11, PKS12, PKS13, PKS14, PKS15, PKS16 및 MSAS)의 종에서 보존된 13개의 PKS 패밀리를 확인하였다. 본 발명자들은 또한 유전자 클러스터 네트워크 분석에서 이러한 PKS 패밀리를 식별하였다(도 2의 A). 그러나, 본 발명자들은 6종의 클라도니아(*Cladonia*) 종의 게놈에서 PKS12 상동체를 검출하는데 실패하였으며, PKS12 패밀리가 원래 확인된 *C. 그레이*(*C. grayi*) 게놈에서도 마찬가지였다. 유전자 주석이 불완전하고 단편화되어 antiSMASH 프로그램이 세가지 PKS, 즉 Cma06514(PKS3), Cgr07347(PKS10) 및 Cuc08442(PKS19)를 포함하는 BGC를 검출하지 못했기 때문에, 3개의 PKS가 수동으로 검출되어 각 PKS 패밀리에 연결된다.

[0126] 실시예 15: 클라도니아 생합성 유전자 클러스터(BGC)의 유전적 복제 제거

[0127] 6개의 클라도니아 속(*Cladonia* spp.) 및 *S. alpinum*의 알려지지 않은 생합성 유전자 클러스터를 다른 곰팡이의 알려진 화합물에 연결하기 위해, Big-SCAPE 프로그램에 의해 예측된 GCF는 MiBiG(Minimal Information about a Biosynthetic Gene cluster) 데이터베이스(v.1.4)에 기탁된 이전에 특성화된 BGC와 결합되었다. 그 결과, 8개

의 GCF 및 7개의 싱글톤(singleton) BGC가 비-지의류화된(non-lichenized) 곰팡이에서 발견되는 알려진 BGC와 연결되었다(도 2의 A). 계통 발생학적으로 먼 종의 이러한 보존된 생합성 유전자 클러스터(BGC)는 지의류 및 비-지의류화된 곰팡이 사이의 생합성 경로의 공유 조상을 제시하였다. 그럼에도 불구하고, 이러한 보존된 BGC는 높은 수준의 유전자 함량 변이를 나타내어 화학적 다양성으로 이어질 수 있는 분기 경로를 제시하였다. MIBiG 데이터베이스에 3개의 지의류 BGC가 나열되어 있다. 구체적으로, *C. uncialis*의 PKS8, *C. uncialis*의 PKS14 및 *C. grayi*의 PKS16에 대한 BGC는 도 2의 A에서 빨간색 점선으로 표시되어 있다.

[0128] 실시예 15-1: PKS8

[0129] *C. uncialis*에서 메틸플로로아세토펜 합성 효소(methylphloroacetophenone synthase, MPAS)로도 알려진 PKS8은 계통발생 분석 및 유전자 발현 연구를 기반으로 하여 우스닉산의 생합성에 관여하는 것으로 알려져 있다. PKS8/MPAS 패밀리의 GCF는 *C. 보레알리스(C. borealis)*, *C. 메타코랄리페라(C. metacorallifera)*, *C. 랑기페리나(C. rangiferina)* 및 *C. 운시알리스(C. uncialis)*에 의해 공유되었다. 그러나, 이 4가지 클라도니아(*Cladonia*) 종 중에서 우스닉산은 *C. 랑기페리나(C. rangiferina)*에서 검출되지 않았다.

[0130] 실시예 15-2: PKS14

[0131] *C. uncialis* (BGC0001489)에서 발견된 추정 6-하이드록시멜레인 BGC(6-hydroxymellein BGC)는 *C. borealis* 및 *C. grayi*에서 PKS14 BGC와 함께 GCF를 형성하였다. 6-하이드록시멜레인 BGC의 유전자 구성원의 역할은 테레인(terrein) 생합성 경로를 기반으로 잠정적으로 지정되었으며, 여기서 *terA* 및 *terB*는 *A. terreus*에서 6-하이드록시멜레인을 생합성하기 위해 협력한다. 본 발명자들의 네트워크 분석에서, 테레인 BGC(MIBiG accession: BGC0000161)는 3개의 클라도니아 종의 6-하이드록시멜레인 BGC와 연결되지 않았으며, 아마도 이들 사이에서 관찰된 유전자 함량 변화 때문일 수 있다: *A. terreus*의 테레인 BGC의 11개 유전자 구성원 중 3개의 유전자(*terA*, *terB* 및 *terD*)만이 3개의 클라도니아 종의 6-하이드록시멜레인 BGC에 보존되었다.

[0132] 실시예 15-3: PKS13 및 PKS15

[0133] 6개의 클라도니아 속(*Cladonia* spp.)에 보존된 PKS15 패밀리의 BGC 및 *S. alpinum*은 다양한 곰팡이에서 멜라닌 생성을 담당하는 1,3,6,8-테트라하이드록시나프탈렌(T4HN) 합성 효소를 포함하여 이전에 특성화된 많은 BGC에 연결되었다. 간결함을 위해, 도 2의 A는 PKS15 패밀리의 BGC에 연결된 T4HN 합성 효소(BGC0001257)를 포함하는 하나의 BGC만 보여주었다. 본 발명자들은 PKS13 BGC가 T4HN 합성 효소를 함유하는 알려진 BGC에도 연결되어 있음을 확인하였다. 6개의 클라도니아 속 및 *S. alpinum*에 보존된 PKS15 패밀리와 달리, PKS13 패밀리는 클라도니아 속 및 *S. alpinum*에서 발견되지 않았으며, *Exophiala lecanii-corni*에서 멜라닌 생성을 담당하는 ElcPKS1과 51-54% 단백질 서열 동일성을 보여주었다(도 5). ElcPKS1 및 PKS13 패밀리는 PKS15 패밀리와 기타 기존 T4HN과 구별되는 고도로 지원되는 계통군(99%의 부트스트랩 값)을 형성하였으며, 이는 일부 클라도니아 종이 다른 기원의 멜라닌 PKS의 두 복사본을 가지고 있음을 나타낸다.

[0134] 실시예 15-4: PKS16

[0135] *C. grayi* (BGC0001266)에서 발견된 추정 회색조 BGC는 6개의 클라도니아 속 및 *S. alpinum*에서 보존된다. PKS16 BGC는 *C. grayi*에서 발견되는 오르셀린산(orsellinic acid)으로부터 유래된 그레이아니산(덱시돈)을 생산하기 위해 제안되었다.

[0136] 실시예 15-5: PKS17

[0137] *C. 보레알리스(C. borealis)*, *C. 마실렌타(C. macilenta)* 및 *C. 메타코랄리페라(C. metacorallifera)*가 공유하는 PKS17 패밀리의 GCF는 다양한 곰팡이가 생산하는 안트라퀴논인 에모딘의 생합성을 위한 BGC(BGC0001583)에 연결되었다. 에모딘 생합성에 필요하고 충분한 4가지 핵심 효소(NRPKS, 메탈로-β-락타마제, 디카르복실라제 및 안트론 옥시게나제) 외에도, 사이토크롬 P450 모노옥시게나제를 포함한 몇 가지 맞춤 효소가 상동 클라도니아 BGC에서 발견되었다. P450 모노옥시게나제는 클라도스포리움 풀범(*Cladosporium fulvum*)에서 2개의 에모딘 유도체의 커플링을 매개하는 Clam과 39% 서열 동일성을 보여 P450 효소가 스카이린의 생합성에 관여할 수 있음을 시사한다. 흥미롭게도, 에모딘 생합성을 위한 4개의 핵심 유전자를 보유하고 있는 또 다른 BGC가 PKS17 BGC와 배열이 다르지만 *C. 마실렌타(C. macilenta)*의 계놈에서 발견되었다. 추정되는 에모딘 생합성을 위한 *C. 마실렌타*의 두 번째 BGC는 *C. 운시알리스(C. uncialis)*의 PKS18 BGC와 GCF를 형성하였으며, 이는 페스탈로티옵시스 픽시(*Pestalotiopsis fici*)에서 발견되는 에모딘 유도체인 페스테산(pestheic acid)의 BGC와 상동성이다.

[0138] 실시예 15-6: PKS34

- [0139] C. 마실렌타(*C. macilenta*) 및 C. 메타코랄리페라(*C. metacoralifera*)가 공유하는 PKS34 계열의 GCF는 푸사리움 푸지쿠로이(*Fusarium fujikuroi*)가 생산하는 적색 균사체 색소인 푸사루빈(fusarubin)의 생합성을 위한 BGC(BGC0001242)와 연결되었다. 그러나, BGC 내용은 매우 다양하였다. 푸사루빈 BGC에서 *fsr1*(NR-PKS 인코딩), *fsr4*(알코올 탈수소효소 인코딩) 및 *fsr5*(단쇄 탈수소효소 인코딩)의 추정 상동체만 PKS34 BGC에서 찾을 수 있었다.
- [0140] **실시예 15-7: PKS37**
- [0141] 페니실리움 엑스판숨(*Penicillium expansum*)의 파툴린 BGC(BGC0000120)는 C. 보레알리스(*C. borealis*), C. 메타코랄리페라(*C. metacoralifera*) 및 C. 운시알리스(*C. uncialis*)가 공유하는 PKS37 패밀리의 GCF에 연결되었다. 파툴린 BGC의 15개 유전자 구성원 중, *patF-patO*와 상동성인 10개의 유전자는 파툴린의 직접적인 전구체인 네오파툴린의 생합성에 필요한 3개의 클라도니아 속(*C. 보레알리스*의 *patF* 제외)에서 찾을 수 있다. 네오파툴린의 파툴린으로의 최종 전환을 담당하는 *patD* 및 *patE*의 상동체는 클라도니아 계통에 없었다. 대신, 발견된 다른 맞춤 효소는 C. 메타코랄리페라 및 C. 운시알리스에서 에피머라제, C. 보레알리스 및 C. 운시알리스에서 타우D 유사 디옥시게나제(TauDlike dioxygenase), C. 운시알리스에서 두 개의 단쇄 탈수소효소이었다.
- [0142] **실시예 16: MIBiG 데이터베이스에 연결된 싱글톤 BGC**
- [0143] 이전에 특성화된 여러 BGC가 단일 종(싱글톤 BGC)에서 발견된 BGC에 연결되었다. 이러한 싱글톤 BGC는 비-지의류화된 곰팡이에서 발견되는 상응하는 BGC에 대해 유전자 함량 변화가 거의 없고 서열 유사성이 높은 경향이 있어 이러한 BGC가 지의류에 최근 도입되었음을 시사한다. 비-지의류화된 곰팡이로부터 수평으로 옮겨졌을 가능성이 있는 BGC는 다음과 같다: C. 마실렌타의 BGC에 연결된 소르비실린 BGC(BGC0001404), C. 그레이(*C. grayi*)의 BGC에 연결된 테레산 BGC(BGC0000160), C. 랑기페리나(*C. rangiferina*)의 BGC에 연결된 하이포테마이신 BGC(BGC0000076), C. 운시알리스의 BGC에 연결된 쿠르부팔리테스(*curvupallides*) BGC (BGC0001563), 및 S. 알피눔(*S. alpinum*)의 BGC에 연결된 테푸데신 BGC(BGC0000046). 모나스코루브린(*monascorubin*) BGC(BGC0000099) 및 베타에논(*betaenone*) BGC(BGC0001264)는 본 발명자들의 네트워크 분석 뿐만 아니라 C. 운시알리스 PKS의 상동성 매핑에 대한 초기 연구에서 C. 운시알리스의 BGC에도 연결되었다.
- [0144] **실시예 17: 클라도니아속(*Cladonia*) PKS 패밀리의 상동성 매핑에 대한 주석**
- [0145] **실시예 17-1: 보존된 지의류 PKS 패밀리의 상동성 매핑**
- [0146] 유전자 클러스터 네트워크 분석에서 많은 지의류 BGC가 높은 우도를 가진 알려진 화합물로 귀속되었지만, 높은 유전자 함량 변이로 인해 비-지의류화된 곰팡이에서 이전에 특성화된 BGC와 연결되지 않은 더 상동성인 BGC가 있을 수 있다. 따라서, 본 발명자들은 NCBI 비중복 단백질 서열 데이터베이스에 대해 45개의 PKS 패밀리에 대한 BLAST 검색을 수행하여 알려진 화합물이 있는 PKS 유전자와 상동성인 클라도니아속(*Cladonia*) PKS 패밀리를 식별하였다.
- [0147] **실시예 17-2: PKS4**
- [0148] PKS4 패밀리는 6개의 클라도니아 속 및 *S. alpinum*에 보존된 R-PKS 효소를 인코딩한다. PKS4 패밀리는 지의류화 및 비-지의류화된 곰팡이에서 찾을 수 있으며, 물리적으로 연관된 유형 III PKS와 공-진화(co-evolution)를 겪은 것으로 보인다.
- [0149] **실시예 17-3: PKS5**
- [0150] PKS5 패밀리는 글라레아 로조엔시스(*Glarea lozoyensis*)에서 제노로조에논(*xenolozoyenone*) PKS와 높은 서열 유사성을 나타내는 PKS-NRPS 하이브리드 효소를 인코딩하는 6개의 클라도니아 속 및 *S. alpinum*에서 보존되었다(단백질 서열 수준에서 56%). PKS5 패밀리의 상동체는 다양한 곰팡이에서 찾을 수 있다. 흥미롭게도, *G. lozoyensis*에서 제노로조에논 생합성을 위한 R-PKS 및 NRPS 효소를 인코딩하는 두 개의 직렬 유전자는 동일한 프로모터의 제어 하에 하나의 디시스트로닉(dicistronic) mRNA로 공동 전사된다. 따라서, 지의류의 PKS5 패밀리가 R-PKS 및 NRPS 효소에 대한 두 개의 개별 폴리펩티드 또는 PKS-NRPS 하이브리드 효소에 대한 단일 펩티드를 코딩하는지 여부를 조사해야 한다.
- [0151] **실시예 17-4: PKS9**
- [0152] PKS9 패밀리는 C. 보레알리스(*C. borealis*), C. 마실렌타(*C. macilenta*), C. 메타코랄리페라(*C. metacoralifera*) 및 C. 운시알리스(*C. uncialis*)에 의해 공유되었으며, 이는 이전의 상동성 매핑 연구에서 기

술한 바와 같이 페니실리움 브레비콤팩툼(*Penicillium brevicompactum*)의 5-메틸오르셀린산(5-methylorsellinic acid)으로부터 유래된 마이코페놀산(mycophenolic acid)의 생합성을 담당하는 NR-PKS(MpaC)와 유사하다 (44% 단백질 서열 동일성). 또한, PKS9 패밀리는 아스페르길루스(*Aspergillus*) 및 페니실리움(*Penicillium*) 종의 3,5-디메틸오르셀린산 유래 메로테르페노이드의 생합성에 관여하는 NR-PKS(AdrD, AndM, AusA, NvfA, PrhL 및 Trt4)에 대해 43-45% 단백질 서열 동일성을 나타내었다. 그러나, 클라도니아속(*Cladonia*) 계통의 PKS9 BGC에는 테르펜 생합성과 관련된 생합성 유전자, 예컨대 프레닐전달효소(prenyltransferase) 및 테르펜 시클라아제(terpene cyclase)가 부족하였다.

[0153] 실시예 17-5: PKS18

[0154] C. 마실렌타(*C. macilenta*) 및 C. 운시알리스(*C. uncialis*)에서 공유하는 PKS18 패밀리는 이전 상동성 매핑 연구에서 기술한 바와 같이 페스탈로티옵시스 픽시(*Pestalotiopsis fici*)에서 에모딘 유도체인 페스테산의 생합성에 관여하는 NR-PKS(PtaA)에 대해 64-68% 단백질 서열 동일성을 보여주었다.

[0155] 실시예 17-6: PKS21

[0156] 보라색 색소 비루로퀸(biruloquinone)은 엽상 지의류(foliose lichen) 파멜리아 비룰라에(*Parmelia birulae*)에서 처음 발견되었다. 비루로퀸은 자연계의 클라도니아 종에서 발견되지 않았지만, 본 발명자들은 이전에 무균 배양에서 비루로퀸을 생산하는 C. 마실렌타(*C. macilenta*)에서 분리된 계놈-시퀀싱된 균체(mycobiont)의 화학적 변이체를 확인하였다. 본 발명자들은 비루로퀸 생성자와 비-생성자 사이의 전체-전사체 비교에 의해 비루로퀸의 생합성에 PKS21 패밀리를 잠정적으로 지정하였다.

[0157] 실시예 17-7: PKS22

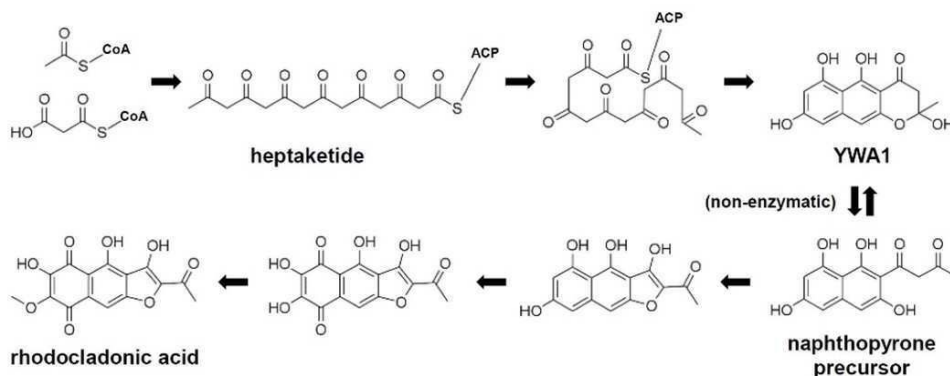
[0158] 나프타자린 유도체인 적색 색소 크리스타자린은 원래 C. 크리스타텔라(*C. cristatella*)의 균체 배양에서 확인되었다. 본 발명자들은 계놈-시퀀싱된 C. 메타코랄리페라(*C. metacorallifera*)에서 분리된 균체에서 크리스타자린 생산을 위한 최적의 배양 조건을 설정하고, PKS23이 C. 메타코랄리페라의 크리스타자린-유도 조건에서 고도로 상향조절된다는 관찰에 기초하여 PKS22 패밀리를 크리스타자린의 생합성으로 귀속시킨다.

[0159] 실시예 17-8: PKS24

[0160] PKS24 패밀리는 C. 보레알리스(*C. borealis*), C. 마실렌타(*C. macilenta*), C. 메타코랄리페라(*C. metacorallifera*) 및 S. 알피눔(*S. alpinum*)에 의해 공유되었다. C. 메타코랄리페라(*C. metacorallifera*)의 PKS24 BGC는 단편화된 PKS24 유전자를 가진 다른 3종의 BGC와 약간 달랐다. PKS24 패밀리는 FgPKS14와 46-48% 단백질 서열 동일성을 보여주는, 푸사리움 그라미네아룸(*Fusarium graminearum*)에서 오르셀린산 생합성에 관여하는 FgPKS14와 함께 100% 부트스트랩에 의해 최대로 지지되는 단일계통 계통군을 형성하였다(도 5).

[0161] 실시예 17-9: PKS30

[0162] PKS30 패밀리는 A. 니둘란스(*A. nidulans*)가 생산하는 나프토피론(naphthopyrone)인 분생포자 황색 색소(일명 YWA1)의 생합성을 담당하는 NR-PKS(wA)와 52-53%의 단백질 서열 동일성을 보여주었다. C. 보레알리스(*C. borealis*), C. 마실렌타(*C. macilenta*), C. 메타코랄리페라(*C. metacorallifera*) 및 C. 운시알리스(*C. uncialis*)에 보존된 PKS30 BGC는 C. 운시알리스에서 색소가 검출되지는 않았지만 적색 자낭반(apothecial) 색소인 로도클라돈산(rhodocladonic acid)의 생합성에 관여할 가능성이 있다. C. 보레알리스(*C. borealis*), C. 마실렌타(*C. macilenta*) 및 C. 메타코랄리페라(*C. metacorallifera*)라는 세 가지 적색 자낭반 색소 생성자가 공유하는 두 개의 GCF(PKS21 및 PKS30 패밀리)만 있었다. PKS21 및 PKS30 패밀리는 각각 NR-PKS 그룹 III 및 IV에 속하며, 이는 나프토피론-유래 화합물의 생합성에 관여하는 NR-PKS를 포함한다(도 5). PKS21 패밀리가 비루로퀸(페난트렌퀴논)의 생합성에 잠정적으로 지정되었기 때문에, PKS30 패밀리는 세 개의 적색 자낭반 색소 생성자가 공유하는 마지막으로 남은 나프토피론 PKS이다. 이러한 추정적 지정은 PKS30 BGC에 있는 O-메틸트랜스페라제 및 2개의 모노옥시게나제와 같은 로도클라돈산의 생합성에 필요한 효소의 존재에 의해 강화되었다. 나프토피론 전구체로부터 로도클라돈산에 대한 제안된 생합성 경로는 하기 나타낸 바와 같다.



실시예 17-10: PkeA

C. 보레알리스(*C. borealis*) 및 C. 메타코랄리페라(*C. metacoralifera*)가 공유하는 BGC에는 A. 니둘란스(*A. nidulans*)에서 펠론(felinone)의 생합성을 담당하는 PkeA(DbaI으로도 알려짐)에 대해 63-65% 단백질 서열 동일성을 나타내는 NR-PKS가 포함되었다. 본 발명자들은 PkeA/DbaI와 상동성인 NR-PKS에 대한 PKS 패밀리 번호를 지정하지 않았으며, 이는 그룹 VII에 속하는 다른 클라도니아속(*Cladonia*) NR-PKS(예를 들어, C. 마실렌타(*C. macilenta*)의 소르비실린 BGC)와 같이 클라도니아속(*Cladonia*)의 특정 계통에 특이적일 수 있다.

실시예 18: 지의류의 대사 잠재력

LFF의 진화적 관계를 조사하고 계놈-인코딩된 대사 잠재력을 비교하기 위해, 8개의 계놈을 사용하여 393개의 단일 복사본 이중 상동 유전자의 병합-기반 트리를 재구성하였으며, 본 발명자들은 2020년 9월 현재 NCBI 데이터 베이스 및 JGI 웹사이트에서 사용할 수 있는 시퀀싱된 22개의 계놈 어셈블리를 가지고 있다. 시퀀싱된 종의 대다수(30개 중 26개)는 접시지의강(Lecanoromycetes)에 속하며, 4개의 LFF는 접시지의강 외부에 배치되었다: 아르토니아 라디아타(*Arthonia radiata*) (아르토니아강), 엔도카폰 푸실룸(*Endocarpon pusillum*) (유로티오미세테스), 스크레로포라 산귀니아(*Sclerophora sanguinea*) (코니오사이보마이세테스(Coniocybomycetes)) 및 비리도텔리움 비렌스(*Viridothelium virens*) (도티데오마이세테스(Dothideomycetes)). 알렉토리아 사르멘토사(*Alectoria sarmentosa*) 및 세트라도니아 리니어리스(*Cetradonia linearis*)의 계놈을 제외하고 BUSCO 분석에 의해 평가된 계놈의 완성도는 91% 이상이었다(도 1). 멸종 위기에 처한 지의류 세트라도니아 리니어리스(*Ce. linearis*) 및 스크레로포라 산귀니아(*Sc. sanguinea*)의 계놈은 20 Mbp보다 짧다(도 1). 이러한 두 종을 제외하고, 모든 계놈은 7,927-14,537개의 오픈 리딩 프레임(open reading frame)을 포함하는 26.3-56.1 Mbp 크기였다(도 1). 30종의 계놈-인코딩된 대사 가능성은 antiSMASH를 사용하여 BGC에 대한 계놈을 마이닝하여 조사되었다. 시퀀싱된 LFF의 대사 잠재력은 16-108의 BGC 수로 매우 가변적이었다(도 1); 파멜리아속(*Parmelia* sp.) KoLRI021559는 가장 큰 BGC 다양성을 보인 반면, 세트라도니아 리니어리스(*Ce. linearis*) 및 스크레로포라 산귀니아(*Sc. sanguinea*)의 계놈은 수축된 계놈 크기에 기인할 수 있는 가장 작은 BGC 세트를 보유하였다. 특히, 6개의 클라도니아속(*Cladonia*) 계놈은 52-65개의 BGC를 보유하고 유사한 수의 PKS를 인코딩한다; 평균 16개의 NR-PKS 및 20개의 환원형 PKS(R-PKS)(도 1). 6개의 클라도니아속(*Cladonia* spp.) 폴리케타이드 기원의 독특한 SM 세트를 생산하여(표 2 및 도 6 참조), 중간(inter-species) PKS 유전자 다양성을 연구하고 비교 분석을 통해 지의류의 피질 또는 수질 물질의 생합성을 담당하는 BGC를 식별하는 데 이상적이다.

하기 표 2에 30종의 계놈-염기서열 분석에서 보고된 지의류 물질을 정리하여 나타내었다.

표 2

종(species)	지의류 물질(Lichen substances)
클라도니아 마실렌타(<i>Cladonia macilenta</i>)	바르바트산(barbatic acid), 디디믹산(didymic acid)
클라도니아 메타코랄리페라(<i>Cladonia metacoralifera</i>)	탐놀산(thamnolic acid), 우스닉산(usnic acid), 디디믹산(didymic acid)
클라도니아 보레알리스(<i>Cladonia borealis</i>)	바르바트산(barbatic acid), 우스닉산(usnic acid)
클라도니아 운시알리스(<i>Cladonia uncialis</i>)	스퀘민산(squamatic acid), 우스닉산(usnic acid)

클라도니아 랑기페리나 (<i>Cladonia rangiferina</i>)	아트라노린(atranorin), 푸마프로토세트라르산(fumaprotocetraric acid)
클라도니아 그레이이 (<i>Cladonia grayi</i>)	그레이아닉산(grayanic acid), 푸마프로토세트라르산(fumaprotocetraric acid)
세트라도니아 리니어리스 (<i>Cetradonia linearis</i>)	-
스테레오카울론 알피눔 (<i>Stereocaulon alpinum</i>)	로바르산(lobaric acid), 아트라노린(atranorin), 스틱트산(stictic acid)
레타리아 루피나 (<i>Letharia lupina</i>)	아트라노린(atranorin), 노르스틱트산(norstictic acid)
레타리아 컬럼비아나 (<i>Letharia columbiana</i>)	아트라노린(atranorin)
슈테베르니아 푸르푸라시아 (<i>Pseudevernia furfuracea</i>)	올리벤토르산(olivetoric acid), 피소드산(physodic acid), 아트라노린(atranorin)/클로로아트라노린(chloroatranorin)
알렉토리아 사르멘토사 (<i>Alectoria sarmentosa</i>)	알렉토론산(alectoronic acid), 우스닉산(usnic acid)
에버니아 프루나스트리 (<i>Evernia prunastri</i>)	레카노르산(lecanoric acid)/에베른산(evernic acid), 피소드산(physodic acid), 아트라노린(atranorin)/클로로아트라노린(chloroatranorin), 살라진산(salazinic acid), 우스닉산(usnic acid)
파멜리아속 KoRLI021559 (<i>Parmelia</i> sp. KoRLI021559)	아트라노린(atranorin)/클로로아트라노린(chloroatranorin), 살라진산(salazinic acid)
우스네아 플로리다 (<i>Usnea florida</i>)	탐놀산(thamnolic acid), 노르스틱트산(norstictic acid)/살라진산(salazinic acid), 우스닉산(usnic acid)
우스네아 하코넨시스 (<i>Usnea hakonensis</i>)	우스닉산(usnic acid)
라말리나 인터미디아 (<i>Ramalina intermedia</i>)	석화산(sekikaic acid), 아트라노린(atranorin), 우스닉산(usnic acid)
라말리나 페루비아나 (<i>Ramalina peruviana</i>)	석화산(sekikaic acid), 우스닉산(usnic acid)
크산토리아 파리에티나 (<i>Xanthoria parietina</i>)	파리에틴(parietin)/팔라시놀(fallacinol)/파리에틴산(parietinic acid)/에모딘(emodin)
크산토리아 엘레간스 (<i>Xanthoria elegans</i>)	파리에틴(parietin)
기알로레키아 플레이버베센스 (<i>Gyalolechia flavorubescens</i>)	파리에틴(parietin)/팔라시놀(fallacinol)/프라길린(fragilin)/에모딘(emodin)
로바리아 풀모나리아 (<i>Lobaria pulmonaria</i>)	기로포르산(gyrophoric acid), 아트라노린(atranorin), 노르스틱트산(norstictic acid)/살라진산(salazinic acid)
엄빌리카리아 퍼스투라타 (<i>Umbilicaria pustulata</i>)	레카노르산(lecanoric acid)/기로포르산(gyrophoric acid)/움빌리카르산(umbilicaric acid), 스카이린(skyrin)
엄빌리카리아 히스패니카 (<i>Umbilicaria hispanica</i>)	레카노르산(lecanoric acid)/기로포르산(gyrophoric acid)/움빌리카르산(umbilicaric acid), 스카이린(skyrin)
엄빌리카리아 무엘렌베르기 (<i>Umbilicaria muehlenbergii</i>)	레카노르산(lecanoric acid)/기로포르산(gyrophoric acid)
시아노더멜라 아스테리스 (<i>Cyanodermella asteris</i>)	스카이린(skyrin)
스클레로포라 산귀니아 (<i>Sclerophora sanguinea</i>)	-
아르소니아 라디아타 (<i>Arthonia radiata</i>)	검출되지 않음
비리도텔리움 비렌스 (<i>Viridothelium virens</i>)	리헷산톤(lichexanthone)
엔도카폰 푸실룸 (<i>Endocarpon pusillum</i>)	-

[0170] 실시예 19: 클라도니아속(*Cladonia*) PKS의 유전적 복제 제거(dereplication)

[0171] 지의류 폴리케타이드, 즉 템사이드 및 템시돈 계열 화합물의 생합성에 관여하는 PKS 유전자를 검색하기 위해,

본 발명자들은 6개의 클라도니아속(*Cladonia* spp.)과 관련 남극 지의류 *S. 알피눔*(*S. alpinum*)에서 적어도 하나의 PKS를 보유하는 226개의 BGC에 초점을 맞추었으며, 이로부터 총 242개의 PKS가 확인되었다. 본 발명자들은 먼저 각 PKS의 보존된 케토신타제(KS) 도메인에 대한 클러스터링 분석을 수행하고, 상동성 관계를 암시하는 많은 클러스터를 식별하였다. 클로도니아 클로로피아(*Cladonia chlorophaea*) 종 복합체에 대한 이전 연구에서 설명된 12개의 PKS 패밀리를 클러스터링 분석에서 검출되었으며, 이는 PKS 유전자의 상당 부분이 클라도니아속에 보존되어 있음을 나타낸다. 그러나, 이들의 생성물에 대해서는 알려진 바가 거의 없다. 따라서, 본 발명자들은 BGC 다양성을 서열 유사성 네트워크에 매핑하는 Big-SCAPE 프로그램을 사용하여 BGC를 비-지의류화된 곰팡이의 알려진 화합물에 연결하였다. 네트워크 분석은 BGC의 세 가지 속성을 그래픽으로 요약하는 데 사용되었다: (i) 네트워크와 관련된 PKS 생성물 (번호), (ii) 네트워크를 통한 종 분포 (노드), 및 (iii) BGC 쌍 간의 유사성 정도 (가장자리)(도 2의 A). 유전자 클러스터 네트워크에 의해 묘사된 바와 같이, 적어도 하나의 PKS 유전자를 보유하는 BGC는 45개의 유전자 클러스터 패밀리(GCF)로 그룹화되었다. 8개의 GCF 및 7개의 종별 BGC(빨간색으로 표시됨)는 MIBiG 데이터베이스에 기탁된 이전에 특성화된 BGC와 결합되었다. 도 2의 B는 클라도니아속(*Cladonia*) PKS 패밀리와 연결된 시그니처 SM을 나타내는 클라도니아속 PKS 패밀리의 계통 분포를 요약하였다.

[0172] 실시예 20: 추정되는 아트라노린 BGC의 식별

[0173] 6개의 클라도니아속(*Cladonia* spp.) 및 *S. 알피눔*(*S. alpinum*)의 유전자 클러스터 네트워크 분석 결과 45개의 GCF 중 PKS23 패밀리가 아트라노린의 생합성에 관여할 수 있음이 밝혀졌으며, 이는 두 아트라노린 생성자인 *C. 랑기페리나*(*C. rangiferina*) 및 *S. 알피눔*(*S. alpinum*)이 공유하는 유일한 GCF이기 때문에 다양한 거대 지의류의 주요 피질 물질이다 (도 2). 아트라노린은 다른 템사이드 및 템시돈 계열 화합물에서 거의 발견되지 않는 3MOA 모이어티 내에 메톡시카르보닐기를 갖는 구조가 독특하다. *C. 랑기페리나*(*C. rangiferina*)의 PKS23 BGC에서 *O*-메틸트랜스퍼라제(OMT)를 사용한 BLAST 검색은 테레토닌 (메로테르페노이드)의 생합성 동안 아스페르길루스 테레우스(*A. terreus*)의 3,5-디메틸오르셀린산 모이어티 내에서 메톡시카르보닐 그룹의 형성을 매개하는 OMT인 Trt5(UniProtKB: Q0C8A3)와 37% 단백질 서열 동일성을 보여주었다 (표 3).

[0174] 하기 표 3에 클라도니아 랑기페리나의 PKS23 생합성 유전자 클러스터 정보를 정리하여 나타내었다.

표 3

[0175]

ORF ^a	크기 (aa)	BLASTP 동족체(homolog) ^b	동일성(%)	보존된 도메인 (Conserved domain)	E-값
06811	744	MFS 당당류 수송체	71	당 수송체 (Pfam00083)	1e-105
06812	306	수송 제어 단백질(export control protein) CHS7-유사	68	키틴 합성 효소 III 촉매 소단위 (Pfam12271)	1e-131
06813	495	TFIIH 복합체 p47 서브유닛	58	TFIIH 복합체 소단위 Ssl1-유사 (Pfam04056)	1e-82
06814	549	MFS 다제 수송체 (<i>atr4</i>)	50	주요 촉진자 슈퍼패밀리 (Pfam07690)	9e-31
06815	346	Trt5, <i>O</i> -메틸트랜스퍼라제(<i>atr3</i>)	37	SAM-의존성 메틸트랜스퍼라제 (Pfam08241)	1e-04
06816	508	시토크롬 P450 모노옥시게나제 (<i>atr2</i>)	43	시토크롬 P450 (Pfam00067)	3e-57
06817	2,513	비환원성 폴리케타이드 합성 효소 (<i>atr1</i>)	44	SAT-KS-AT-PT-ACP-MT-TE	
06818	472	가상 단백질	33	인산전이효소 패밀리 (Pfam01636)	4e-08
06819	2,377	환원성 폴리케타이드 합성 효소(R-PKS)	40	KS-AT-DH-ER-KR-ACP	
06820	184	가상 단백질	40	미지 기능의 도메인, DUF3237 (Pfam11578)	5e-22
06821	301	가상 단백질	34	미지 기능의 도메인, DUF2306 (Pfam10067)	2e-09
06822	236	가상 단백질	28	검출되지 않음	
06823	356	가상 단백질	32	검출되지 않음	
06824	565	추정 디메틸아닐린 모노옥시게나제	41	플라빈-결합 모노옥시게나제-유사 (Pfam00743)	3e-14

[0176] ^aORF: *C. 랑기페리나*(*C. rangiferina*) 오픈 리딩 프레임(open reading frame). 합성 핵심 유전자는 볼드체로 강

조 표시됨.

[0177] ^bBLAST 검색의 경우, 4개의 아스페르길루스(*Aspergillus*) 종에 대한 NCBI 비중복 단백질 서열 데이터베이스가 사용됨: 아스페르길루스 푸미가투스(*Aspergillus fumigatus*; taxid 746128), A. 니둘란스(*A. nidulans*; taxid 162425), A. 니제르(*A. niger*; taxid 5061) 및 A. 테레우스(*A. terreus*; taxid 33178).

[0178] 아트라노린을 생산하는 다른 지의류에서 보존된 PKS23 BGC의 존재를 조사하기 위해, 본 발명자들은 GCF 내 및 전체에서 BGC의 보존 및 변이를 연구하는 데 유용한 CORASON 파이프라인을 채택하였다. C. 랑기페리나(*C. rangiferina*)의 PKS23 BGC에서 OMT의 단백질 서열을 30개의 게놈에서 검출된 총 1,527개의 BGC에 대해 질의하였다. 본 발명자들은 각각 상당한 히트(>33% 단백질 서열 동일성)를 나타내는 OMT를 포함하는 15개의 BGC를 확인하였다(도 3). 15개의 BGC 중 7개의 BGC에는 4개의 합성 핵심 유전자가 포함되어 있다: PKS23(*atr1* 합성), 시토크롬 P450 모노옥시게나제(*atr2*), OMT(*atr3*) 및 수송체(*atr4*)(도 3 및 표 3). 이러한 합성 BGC는 30개의 시퀀싱된 종 중에서 아트라노린을 생성하는 지의류에서만 독점적으로 발견되었다(표 2 참조).

[0179] **실시예 21: 아트라노린 생합성 경로의 재구성**

[0180] 아스페르길루스 니둘란스(*A. nidulans*) 및 아스페르길루스 오리제(*A. oryzae*)와 같은 잘 확립된 이중 숙주에서 지의류 PKS 유전자의 이중 발현을 위한 힘든 노력에도 불구하고, 시도는 아직 알려지지 않은 이유로 성공하지 못하였다. 따라서, 본 발명자들은 식물 병원성 곰팡이 아스코키타 라비에이(*Ascochyta rabiei*) (도티데오마이세테스(*Dothideomycetes*))를 사용하여 새로운 이중 발현 시스템을 구축하기 시작하였다. 추정되는 아트라노린 BGC에서 세 가지 생합성 유전자의 화학적 구조 및 예측된 기능을 기반으로 아트라노린에 대한 폴리케타이드 경로를 구상할 수 있었다(도 4의 A). 개별 유전자의 역할을 조사하기 위해, 먼저 본 발명자들은 야생형 아스코키타 라비에이(*As. rabiei*)에서 솔라나피론의 생합성을 위한 BGC를 제거하여 상당한 대사산물 생산을 보이지 않는 "클린 숙주"를 생성하였다(도 4의 B). 그런 다음 S. 알피눔(*S. alpinum*) gDNA에서 복제된 *atr1*을 "클린 숙주"에 도입하였다. *atr1*의 발현 및 올바른 인트론 스플라이싱은 RT-PCR 분석에 의해 확인되었다(도 7 참조). LC-MS/MS 및 NMR 분석은 알려진 지의류 템사이드, 4-O-데메틸바르바트산(4-O-demethylbarbatic acid, 화합물 1)이 *atr1*을 발현하는 균주에 의해 생성되었음을 나타내었고 (도 4의 B 및 도 4의 C 참조), 이는 2개의 3MOA 단위가 *Atr1* 자체에 의해 템사이드로 결합되었음을 시사한다.

[0181] 화합물 1로부터 아트라노린의 생합성은 2개의 3MOA 단위 각각 내에서 메틸기의 산화 및 카르복실기의 메틸화를 필요로 하기 때문에(도 4의 A), 본 발명자들은 *atr2* 및 *atr3*을 코딩하는 유전자를 *atr1* 발현 균주에 개별적으로 도입하였다. *atr3*과 *atr1*의 동시-발현은 화합물 1의 7'-O-메틸화된 유사체, 프로아트라노린 I(proatranorin I) (화합물 2)을 산출하였으며, 이는 아스페르길루스 테레우스(*A. terreus*)의 Trt5와 마찬가지로 *Atr3*이 실제로 카르복실 메틸라제임을 나타낸다(도 4의 B 및 도 4의 C 참조). LC-MS/MS 분석은 *atr1*과 *atr2*의 동시-발현으로 화합물 3 및 4가 생성되었으며, 각각 화합물 1의 하이드록실화 유사체(프로아트라노린 II) 및 추가로 산화된 알데히드(프로아트라노린 III)로 주석이 달린 것으로 나타났다 (도 4의 B 및 도 4의 C 참조). 분리된 프로아트라노린 III(화합물 4)에 대한 NMR 분석은 *Atr2*에 의한 산화가 화합물 1의 C-9에서 발생함을 확인하였다. 마지막으로 *atr1*, *atr2* 및 *atr3*의 세 가지 유전자를 발현하는 균주를 생성하였다. 아트라노린(화합물 5)의 생성은 MS/MS 스펙트럼을 지의류 데이터베이스에서 아트라노린의 참조 스펙트럼과 일치시켜 확인하였다. 이는 또한 각각 피질 및 수질 물질로서 아트라노린(화합물 5)과 로바릭산(lobaric acid)을 생산하는 시퀀싱된 S. 알피눔(*S. alpinum*)에 대한 실제 바우처 표본과 화학적 프로파일을 비교하여 확인되었다(도 4의 B 및 도 4의 C). 흥미롭게도, 화합물 6의 MS/MS 스펙트럼 매칭은 자연에서 S. *alpinum*에서 보고되지 않은 또 다른 지의류 템사이드인 베오미세스산(baeomycesic acid)임을 확인하였으며, 이는 *Atr3*이 이중 숙주에서 이완된 기질 특이성을 나타냄을 시사한다.

[0182] **실시예 22: 지의류 물질의 생합성을 담당하는 PKS의 새로운 계통발생 분기군**

[0183] 곰팡이 NR-PKS와 아트라노린 생산을 담당하는 *Atr1*의 진화적 관계를 연구하기 위해, 본 발명자들은 6종의 클라도니아속(*Cladonia* spp.) 및 S. *alpinum*에서 발견된 103개의 NR-PKS와 비-지의류화된 곰팡이의 알려진 화합물에 연결된 82개의 NR-PKS의 보존된 KS 및 생성물 주형 (PT) 도메인의 연결된 시퀀스의 계통발생 트리를 재구성하였다. 곰팡이 NR-PKS는 이전의 계통발생학적 분석에서 지금까지 크게 8개 그룹(그룹 I-VIII)으로 분류되었다. 본 발명에서, 본 발명자들은 PKS23 패밀리를 뿐만 아니라 PKS1 및 PKS2 패밀리를 포함하는 100% 부트스트랩이 지원하는 새로운 NR-PKS 그룹(그룹 IX)을 확인하였다(도 5). PKS1 패밀리는 이전에 C. 랑기페리나(*C. rangiferina*)에서 3MOA-유래 지의류 물질을 생합성하기 위해 제안되었다. 그룹 IX의 지의류 NR-PKS의 경우, S. *alpinum*의 PKS2를 제외한 모든 것은 폴리케타이드 중간체의 메틸화와 관련된 C-메틸트랜스퍼라제(cMT) 도메인을 보유하였다.

흥미롭게도, 새로 확인된 그룹 IX에는 비-지의류화된 곰팡이에서 이전에 특성화된 3개의 NR-PKS인 AscC, StbA 및 TmPKS12가 포함되었다(도 5). 이러한 NR-PKS에는 cMT 도메인이 없으며, 오르셀린산을 생성하는 것으로 알려져 있다. 또한, 그룹 IX에 대한 그룹 VIII 기저에는 버섯-형성 곰팡이(주름균강(Agaricomycetes)) 및 푸사리움 그라미네아룸(*Fusarium graminearum*)의 오르셀린산-유래 화합물과 연결된 여러 NR-PKS가 포함되어 있다는 점에 주목해야 하며(도 5 참조), 이는 PKS1 및 PKS2 패밀리가 PKS23 패밀리와 같이 메틸화된 형태의 오르셀린산에서 유래된 지의류 물질의 생합성에 관여할 가능성이 있음을 나타낸다.

수탁번호

[0184]

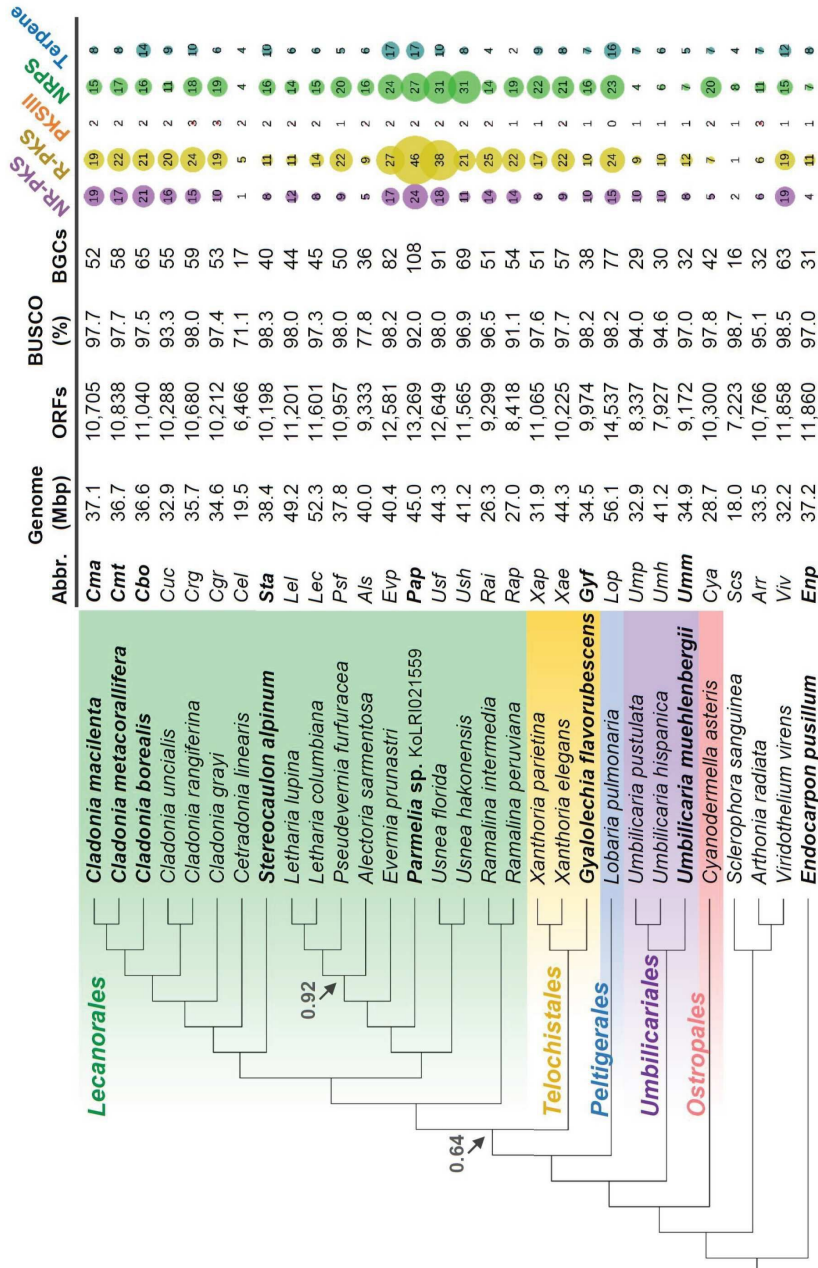
기탁기관명 : 농업생명공학연구원

수탁번호 : KACC83048BP

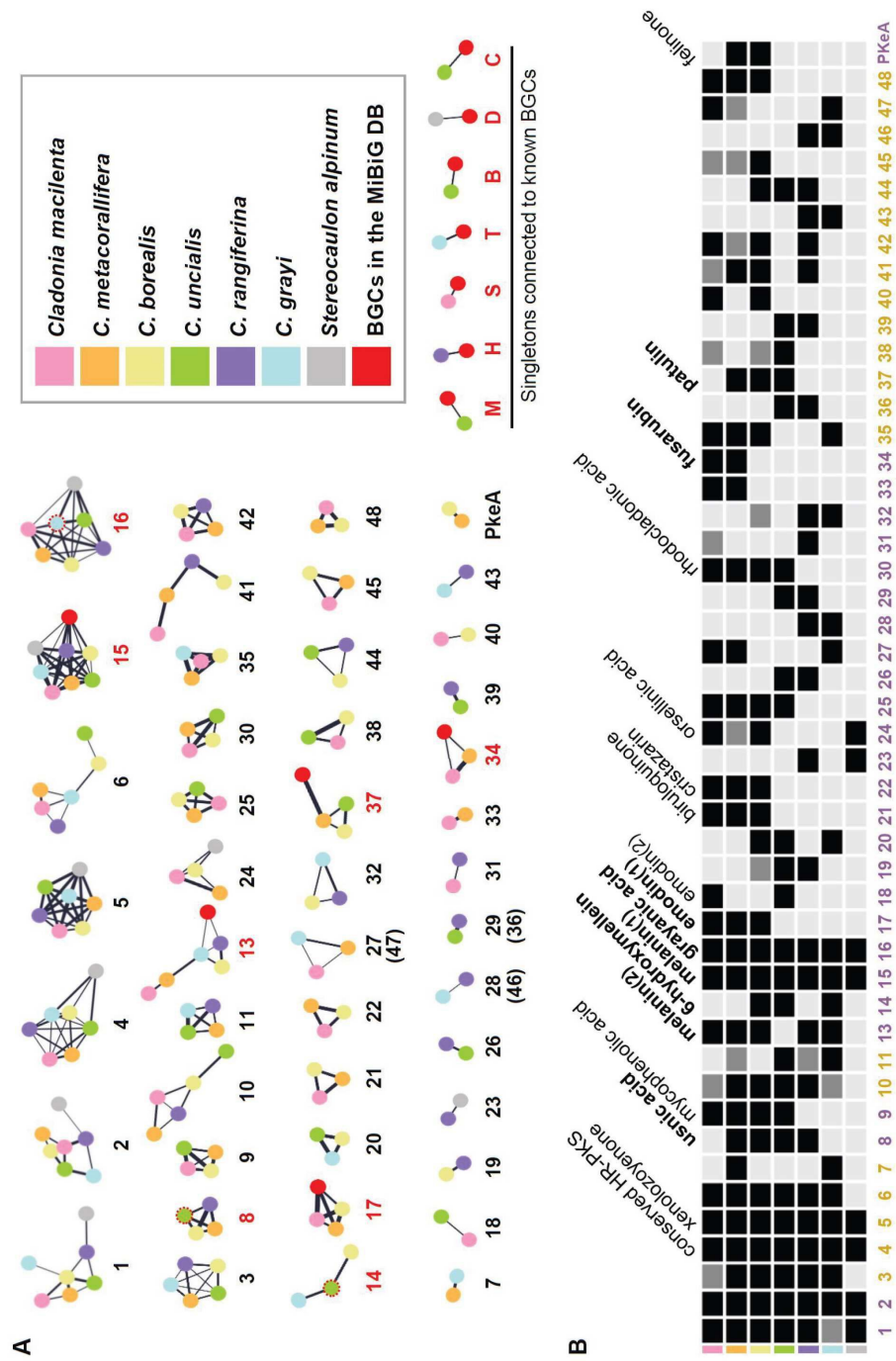
수탁일자 : 20210528

도면

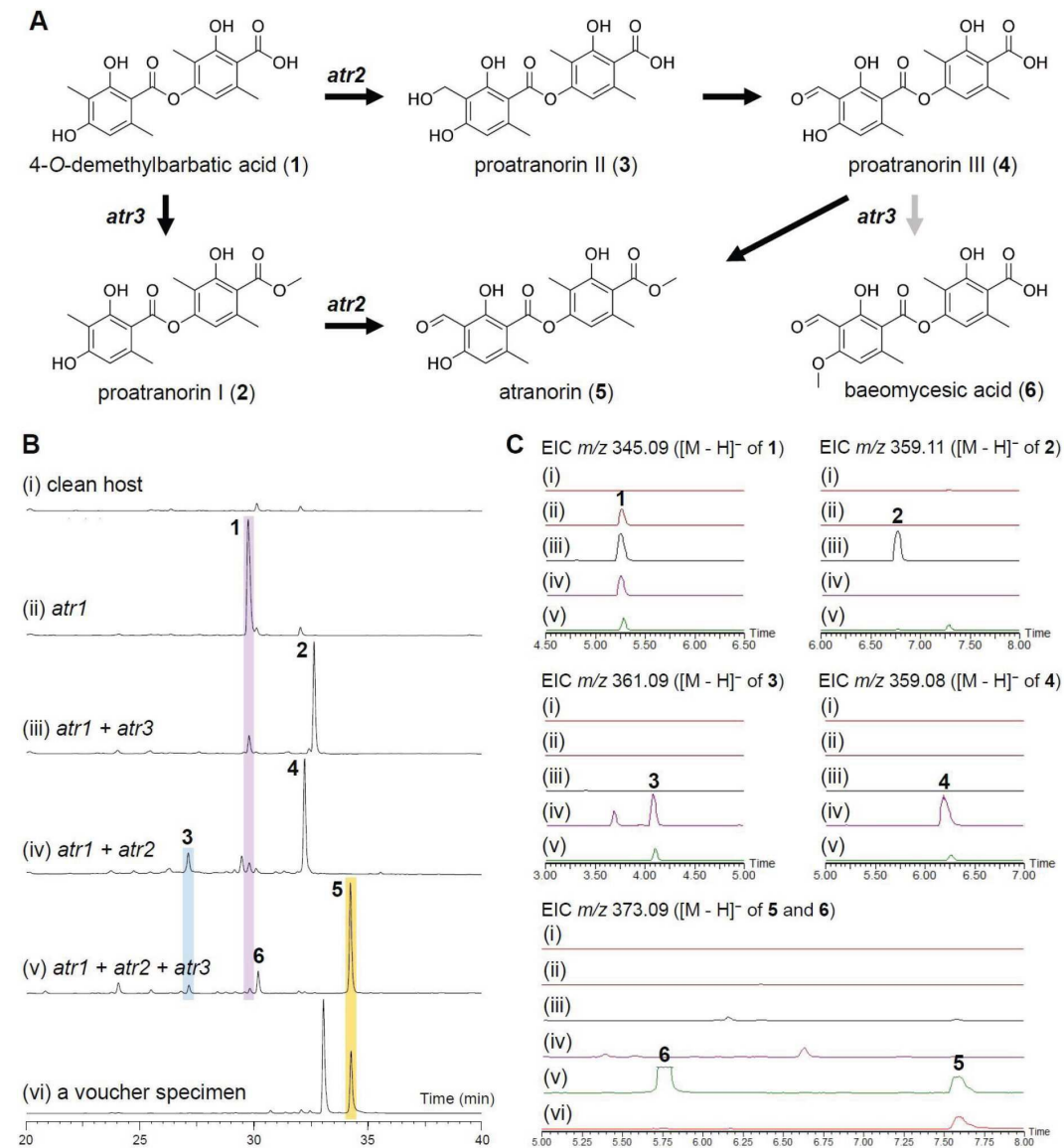
도면1



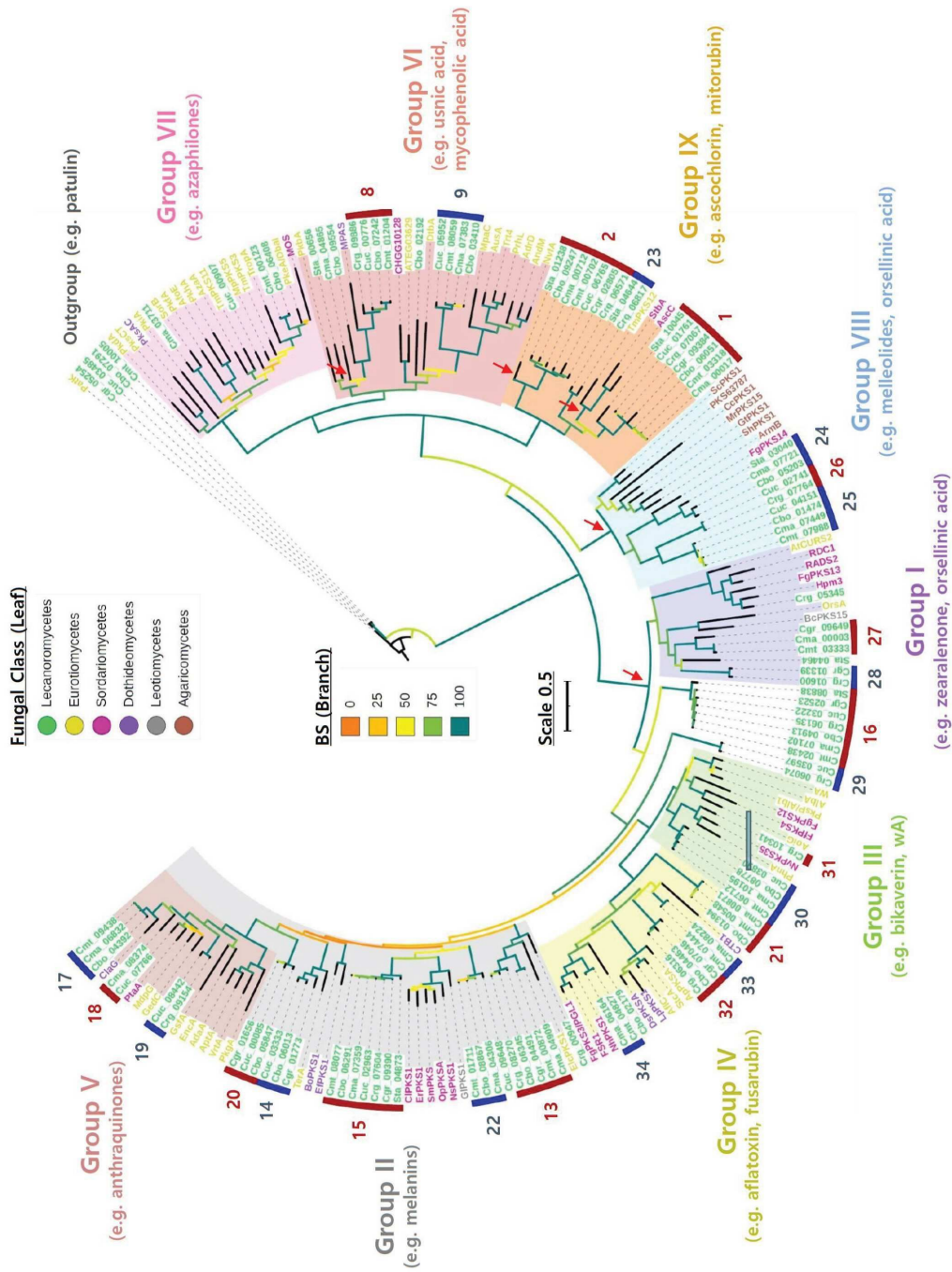
도면2



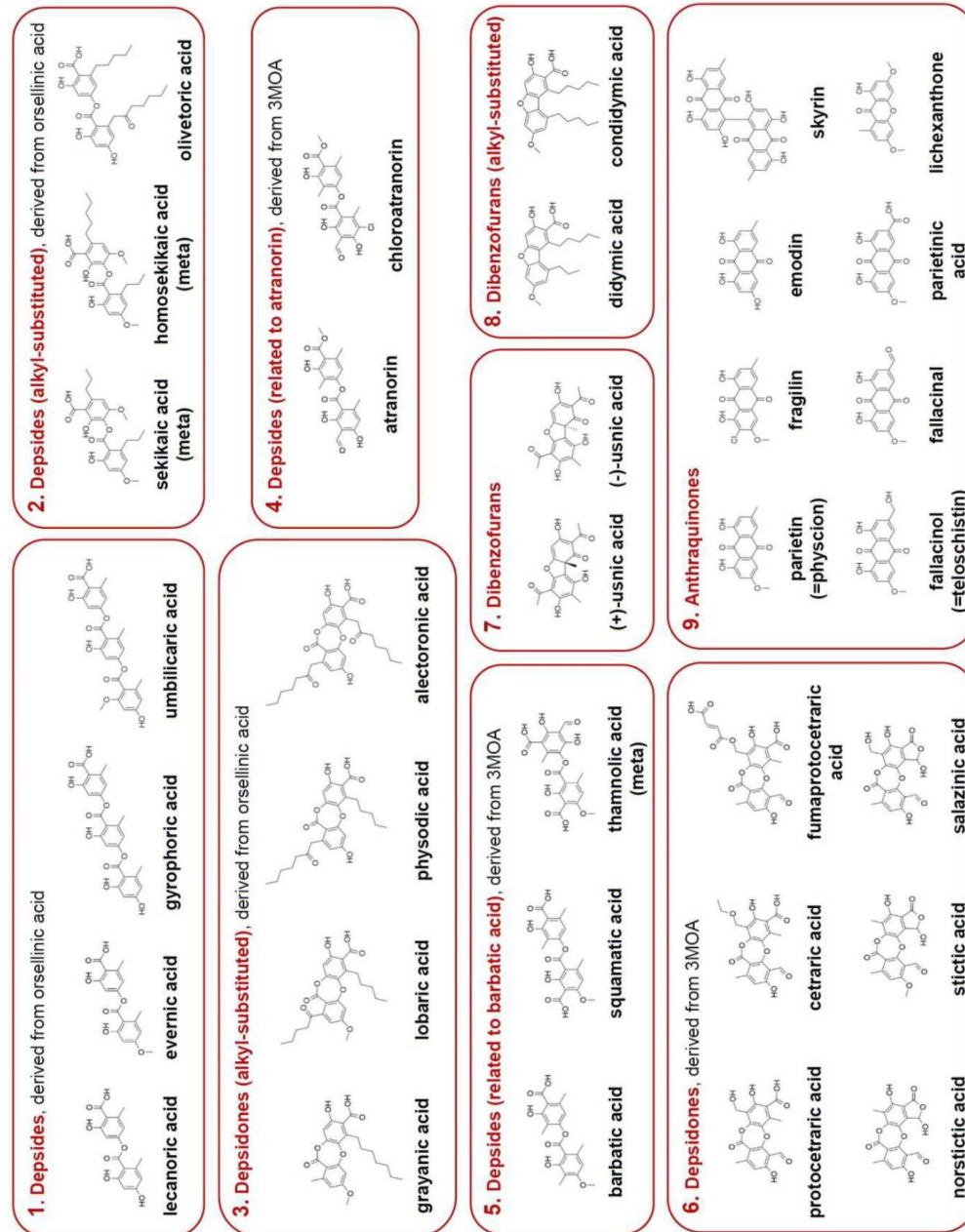
도면4



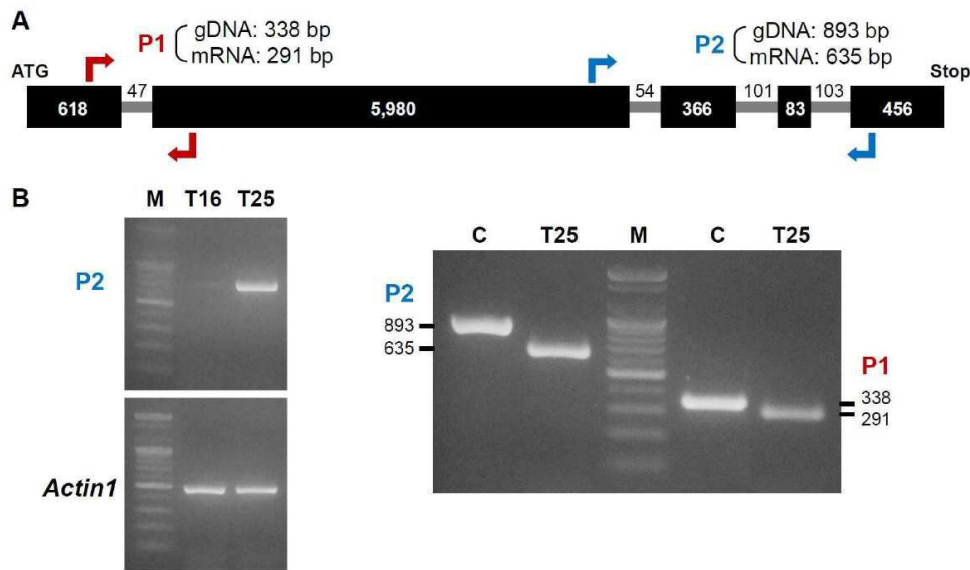
도면5



도면6



도면7



서열 목록

- <110> INDUSTRY-ACADEMIC COOPERATION FOUNDATION OF SUNCHON NATIONAL UNIVERSITY
- <120> Atranorin biosynthesis gene derived from lichens and uses thereof
- <130> PA-21-0182
- <150> KR 10-2021-1234567
- <151> 2021-02-12
- <160> 46
- <170> KoPatentIn 3.0
- <210> 1
- <211> 7808
- <212> DNA
- <213> Artificial Sequence
- <220><223> Stereocaulon alpinum atr1 gene(MZ277879)
- <400> 1

```

atggcgtcta accatgagga gctacattca atggtagtgt tcagcccga gagcaaagcc      60
ccaaaagaag gataacctga tgagttacgg tcatatctat gcggcaaggc tgaacttcga      120

cctttgcttg atggtatcga aaaccttcca aacacctggc caatctttgc tcaaaggaac      180
tccgacattg cagctctgac tcaaggaatc cggtacacac aggcactctc agactgggcc      240
aagcatggaa cttcgtcagg aatctcgaac gttatgtctg ggatcctatc gttgcctctc      300
ttgaccatca ttcaggtggc ccaatacttc caattccttg aagtgaaaaa gctccgtcac      360
agcgacttca tggagcgact gaggtgtagg ggtggagttc aagggtattg tggcggccta      420

```

atgcctgccca ttgccatcgc ctgctctgcg acagaagctg aggttggttac caacgctgtt	480
aaagcgatgg gtattgcact aggtgttggc gcttacgggg agcttgggga cgatgagaac	540
gtcctagggc caactacaat agttgtccgc ctcaagcagg agggccaagg tggagacatc	600
ataaaggact ttcttgacgt gagctgctcg ttccgcacct accaaactca ctaactttgt	660
tacaggcgca tatatcagct atcacggacc caaagaccgt cagcattgta gggtcagccc	720
catctttggc agagatacaa gcccagtgta agtctaacgg catgcaaacc caagccatgc	780
atctgcgtgg gaaagtgcatt aatcccgaaa acgctaattc agcgttgagg ctttgcatgc	840
tttgcgacga gcatgaggaa ctctctttgc caaacgcttc acacttgacg gctcctttga	900
gatctaacaa gacaggggaag aagttattag atgtctctct tactcatgaa gcgatcgaaa	960
caatcctggc atcctgctgt caatggtagc accttctcaa aggggtgtgc aaagacttgg	1020
agaaaacagg aactcaatca caccttttcg catcgttcgg cattggcgac tgcattccctc	1080
tgacaccgtt tcatcaggca gggctacaga ttaccaaatt ggacgtttctc tcctttgtca	1140
aagccttgat gccgcctgtt ctaccaatgt ccggtcacaa ccaccaatat gcttatccaa	1200
ctgacgctgt ggctgtcgtc ggaatggcgt gtcgcctacc gggtgccaac agcgtcgaag	1260
aactctggga tcgatctcg tccggtggtt caactgttac cctgtccca gaagatcgga	1320
tggatattgc aggtagcttt cgagctatgc aagatccgaa gtgggctgct aagcaacaat	1380
ggtggggcaa tttatctct gatattgccg gatttgacca cagcttcttc cgcattgagcc	1440
ctcgggaagc cgcattgatg gatccacaac aacgattct cctcgagaca gcataccaag	1500
caatggagtc gattggctac cttggttcgc atcgacgca gtcaggatgat ccagtgggag	1560
tattcttagg agcaagtctc gtcgagtacc tggacaatac ttctcaaacc ccaccgacag	1620
catacacctc cactggcaca atcagggcgt ttttgagcgg caagataagt tattactttg	1680
gctggaccgg tccttctgaa attctcgaca cagcatgctc gtcgtctctg gttgccatta	1740
acagggcctg caaggcgatt cagaacgatg agtgcccat ggcaattgct ggtggcgcta	1800
atctaattac tggatccac aattatttgg atctcgcaaa agcaggcttc cttagtccca	1860
gtggacaatg caagccgttt gatggcgctg ctgatggcta ctgccgttcg gaaggcgagg	1920
gtctcgttgt tctcaagcgg ttgagccagg cattgaccga tgggaatcag atccttggcg	1980
tcatcacggg ggctagcacc aatcaaggcg gtttaagtc atcgtcaca gtccgcaca	2040
gcgtgcccc ggicaaattg taccaaaaca ttcttcatca ggcgggaatg agaccggagc	2100
aagtttctta ttgcgagact catggcactg gcactcaagc aggcgatcct ctggagattg	2160

agagtgtgcg ggaggttttt ggtggtccta aaagacagga ttccatgcac ataggctcaa 2220

 tcaaaggcaa cattggtcac tgtgagactg ctgcaggagt cgctggcctt ctcaaagcgc 2280
 tcgtcatggt caacaaagct gcaataccgc ctctagcgag tcacaagagc ttgaatccga 2340
 agattgccgc cctcgaacct gataagcttg ctatttcaag ttgtcttgaa gactggagag 2400
 tctcgccgag agctgcattg gtaaacagct atggcgcggc cggaagtaat tctgctgtcc 2460
 tactatgcca ggccccgac attgacaacg cacctttaca ccgcgtagca accacagagc 2520
 atacctatcc aatcatactg agcgtgcat cgaaaccaag tctactttcg aatgctgaaa 2580
 acatagcaag ctatttgagg aaagcaactt ccaagtgcac gattgccgat gttgcgttca 2640

 ctctcgtcaa acaacgtaaa cgccatccgt tacaatggat aacaatggag tcaagcatcg 2700
 acggcttggc caagtcctg gggcttttgc aggaccctc aaatgctccg cagccaaaga 2760
 aggtggttat gacattttct ggccaaagtc gacaatcgat tgggctcaac aaagaatggt 2820
 atgactcttt ccctctgttc cggcgccatg tgaacgaatg cgacgacttg ctgcagcaga 2880
 gcggatttcc ttctgcaag tctgcaatgt tcgataagga gcctgctcgg gatgttgtgc 2940
 cactccagtg tgcgatgttt gctgttcagt acgcatctgc cttgtcttgg atcgattgtg 3000
 gacttcaagt agaagctgta attggccaca gcttcggtga gttgacagct ttggcagtgt 3060

 cgggtacctt ttccctcaaa gatgctctca atcttgttgc taccgagct acattgatgc 3120
 aatcaaaatg ggggcctcac aaagggacca tgctcttgat cagtgcggct actgagatgg 3180
 tacgtaaaat cattgcagga aaccgcgatg tcgagatcgc ttgccacaat gcaccaacaa 3240
 gccaaattgt tgtaggtagc caagctgcta tctcggagggt tgaaaaagtt ctcgaaaaca 3300
 atacggagta tcgcggcatc cagagtcagc gcctcaatgt tacgcacggc ttccattctc 3360
 aattcaccga gccattactc gagaatcttt cagagagtgc cagatctcta gtcttccatg 3420
 agccgaagat ttgtctagag tggtgcactc ttgaagaact gaaccatgtg gggcctgac 3480

 acttagcacg acacactcgg gagcctgttt acttctacca cgctgtacgc cgctcgagc 3540
 agcgtcttgg cacatgtctg tggtttgagg ctggtttcga ctctccgatt ataccaatga 3600
 ccaagagggc cgttgaattc ccagagagac acacctttct ggacatgaag actccgagtg 3660
 ggacgaatcc cactaagatg ctactaccg caacgatcaa cctttggcaa aacggtgcta 3720
 catectctt ctggggcttc cateccatag agacaacaaa cattaaacag gtgtggctgc 3780
 caccgtacca attcgatcgg acctccatt ggatgcctta taccgatcat gctttggaga 3840
 tgagcaaaat ccaagccgtc atcagcaatt ctgaaccctt ggtcgaatta tcaactaagc 3900

caccgcgact gggtgagcct cggacgaagc ccagtgagaa aggtgaattc tccatgaaca	3960
cacaagcgcg acgtatatac gaaattgttt ctggtcatgc tgttttgagt cgaccattgt	4020
gtcctgcagc tatgtacatg gaatgcgcaa tcatggccgc acagttgtct attggcaaca	4080
tcgtgggcca ggcgccgtgg ttcgagaatc tgacctttga ggctcccctc ggaatggatc	4140
ctgacaacga caccacagtg gtcttgaagg atgatgttc caaaagcaga tggtcatttg	4200
tcgcgaggag cagcagtagg agcaaccgca aacgaagcc agtcttgac gccaaagggg	4260
acttcggttt cactacacaa actcaagtgc atcgatatga gaggcttgta accgaccgca	4320
tgaggcattt gcagcactct aaaagcgaga cacttaagtc caagcgagct tatggacttt	4380
tctcccgaat cggtcggtat gcagagctat tgaagggaat ttcatcaatt actctaggtg	4440
attcagaagc atctgcaata atcgacgtac cgcttgccgc cagcaccgag gacagctcgg	4500
caacaggact ctgtgattgc gtggcacttg atgccttcat tcaagtgtgc ggcttattga	4560
tcaactccgg tgatgattgt gccgaagatg aggtttttgt tgcaactggt gtcgagaact	4620
tctcaatgic gttggcctgc gacttcgac gctgcagaac ttggcttgta ttcgcatgt	4680
tcacaccatc gggtaatggt aaagcgatgg gagatgtttt catcctgacg agggacaatg	4740
tctctggtcat gacaatcatg ggtgtacaat ttaccaagct tccgattaca aggctggaga	4800
agctattgga ctcgccaat cccaaagctc acaatacgcc aattttgaag tcaagtcagc	4860
aagactccat cgtatcgcc agcagtagct ctagtacgga gcatagcgat gacgacagcg	4920
aggacgatgg atcacggtct cccagccaca gtgatactc cgcttgacagc gactcagaag	4980
cgccagcgga caacggtgct gcaaagaaac taaagtctct cattgcatca tatgttggga	5040
tcgctgaaga cgctatttcc gatgatgcaa atattgcgga tcttggagtc gactctcttg	5100
ctgctacaga actagccgac gagatctcga acgacttcgc taaggaaatc gacggcggcg	5160
agcttcccat gatgacattt ggtgaacttt gtgcaatagt ggccccggag atggctgcaa	5220
agcctgcaaa ggcgaaaaag aagattccct acaaaggtaa agatgaagcg actgttgttg	5280
agtcacacc aggtaaatca cagagtgaga tcaaggactt gaaggctgtg gtggagcctt	5340
tgccaagatc gactcccatg ttatcagata ccaccgtcgt ccgatctgat cctaccaag	5400
tccttcgcca aattgatacc atgtttcagc catctgctga tacattcggc ttcacggact	5460
actggacagc ggtcgcccca aagcagaata agttagtcc ggcatatatt ggcgaggagt	5520
tcagaaagct aggactcgat ctctgggagg ttcaacctgg agcaaccctt ccacatattg	5580
agtaccttcc aaagcatgaa aaggttgtcc agagactgtg ggatattctt gctgatcatg	5640
gtattgtcta caattatgac gccgcaaagg tgcggtccag caaacctctt ccggacgctc	5700
cttcacggc tctcctcaat gagctcaacg ccttgttccc aaatttcgcc aatgagaatc	5760

ggctaattgtc tgtcacggcg cctcatttcg cggatggttt gaggggaaag actgataata	5820
ttagtctgct gtttgggagt caacgtggcc aggaatgcct gaatgacttc tataacaatt	5880
cgcctcagtt ggccgtgatg accgatcact tgcctacttt ctttaagcag ctctgaaag	5940
aagccccact tgaaggttca ctccgaatcc ttgaggtcgg tggcggcttt ggagggacga	6000
caaaacgact agcagaaatg ctggaggcac taggtcagcc ggtcagatg acgttcacgg	6060
atgtctcgct catgctagtg aaggaagcaa ggaaaaagt ttccaaacat tcttgatgg	6120
actttcagtc ctgaatctt gaaaaggacc ctccagcttc tcttcaaagg acatacgata	6180
ttgtgattgg gactaatgta gtgcacgcaa catcgaacat tgtgaattcg acgacacga	6240
tgcgaagtct tttaggaag ggtggcttta tcgtcttacc tgaagtcacc cgcattgttg	6300
actggtacga tctggtttat ggactccttg atggctgggtt ggccctcaaa gactctagga	6360
cgtatccttt gcagcctgcc gatgactggg tgagagacct catgaaagcg ggattcgaga	6420
ccgcatacta ctgagggga gactccgagg agtccaacac acagcagctg attattggct	6480
cgaccaggcc aagcaaggta gcgtctacca gcgggctttc ggaagccagg ctaagtaagt	6540
cctatagaat tgagacaatg ccatacaaga tcatcgacga cacggagata ctggcggacg	6600
tcttcttccc agagcatgaa gtggcctcgg aagccatgcc tataggtaat cttctccagt	6660
gtgtctagca atctcacga gccgtgact tctgacaagc cctcatgac cacggtggcg	6720
gttttatgac actctcaaaa acggccatca gacctacca gaccaattc ctctgggaaa	6780
acggctatct tcctatcagc atcgattacc gactttgccc cgagattgat ctaatcgctg	6840
gtccgatgac ggacgttcgt gacgtcttaa cctgggtgcg aaaacagctg cctgcgattg	6900
ctcgtacccg cgcatcaat gtggacccaa cgaagttgt agtcatcgga tggccacag	6960
gtggtcatct ggccctgact acagcgtgga catgtgagga cattggagaa gagccacctg	7020
ttgcagtcct tagctttac ggtccaacga acttcgaatc tgaaggtgcg cgtccattga	7080
cactcttttc tgggtgctag tgccacctgt ggtccaaacc gtagcttagg ctgcattcct	7140
acaagatgc taattgttca taacagacat tgatcgagg cgtgcggaac agtaccaga	7200
acggactatg tcttctgac ggatccgcaa gtcattatcc acgaagcctg tgagtatctc	7260
cttgctattg caaattctcc taggccttga cattgcttcc ccatagcatt gttactcttc	7320
tactcactgc taataccctc ttccacacgc agatcacttc ctatgactgt cccaccgga	7380
cagacagcac cggctctcggc tgggtgcgtc ccggcgaccc acgctctgaa ctctcttgt	7440
ccctcttcaa agaattctac ggtctgcaag tgatgcttaa cggctcttct gccgcagatt	7500
tggccaaacc tccgcatgt gctaaaattc aggcgatcag cccgatggcg cagctcaagg	7560
ccgggaggta caacgtccca acttttgtaa ttcatcaga ttgtgatgag atcgcccggt	7620

tcagggatag cgaggctttt gtggaggaac tggcgaggag aggggtgaag actggactag 7680

ggagggtgag gggcaagaag catattcatg acttggcggt gaagccggag aaggacggat 7740

gggtggacgg tgcgggcgtt ggatatgagt tcattttcga tgttgttga cgaggcgta 7800

ggggttga 7808

<210> 2

<211> 2500

<212> PRT

<213> Artificial Sequence

<220><223> Stereocaulon alpinum atr1(QXF68953)

<400> 2

Met Ala Ser Asn His Glu Glu Leu Pro Ser Met Val Val Phe Ser Pro

1 5 10 15

Gln Ser Lys Ala Pro Lys Glu Gly Tyr Leu Asp Glu Leu Arg Ser Tyr

20 25 30

Leu Cys Gly Lys Ala Glu Leu Arg Pro Leu Leu Asp Gly Ile Glu Asn

35 40 45

Leu Pro Asn Thr Trp Ser Ile Phe Ala Gln Arg Asn Ser Asp Ile Ala

50 55 60

Ala Leu Thr Gln Gly Ile Arg Tyr Thr Gln Ala Leu Ser Asp Trp Ala

65 70 75 80

Lys His Gly Thr Ser Ser Gly Ile Ser Asn Val Met Ser Gly Ile Leu

85 90 95

Ser Leu Pro Leu Leu Thr Ile Ile Gln Val Val Gln Tyr Phe Gln Phe

100 105 110

Leu Glu Val Lys Lys Leu Arg His Ser Asp Phe Met Glu Arg Leu Arg

115 120 125

Cys Arg Gly Gly Val Gln Gly Tyr Cys Gly Gly Leu Met Pro Ala Ile

130 135 140

Ala Ile Ala Cys Ser Ala Thr Glu Ala Glu Val Val Thr Asn Ala Val

145 150 155 160

Lys Ala Met Gly Ile Ala Leu Gly Val Gly Ala Tyr Gly Glu Leu Gly

165	170	175
Asp Asp Glu Asn Val Leu Gly Pro Thr Thr Ile Val Val Arg Leu Lys		
180	185	190
Gln Glu Gly Gln Gly Gly Asp Ile Ile Lys Asp Phe Pro Asp Ala His		
195	200	205
Ile Ser Ala Ile Thr Asp Pro Lys Thr Val Ser Ile Val Gly Ser Ala		
210	215	220
Pro Ser Leu Ala Glu Ile Gln Ala Arg Val Lys Ser Asn Gly Met Gln		
225	230	235
		240
Thr Gln Ala Met His Leu Arg Gly Lys Val His Asn Pro Glu Asn Ala		
245	250	255
Asn Leu Ala Leu Glu Leu Cys Met Leu Cys Asp Glu His Glu Glu Leu		
260	265	270
Ser Leu Pro Asn Ala Ser His Leu Thr Ala Pro Leu Arg Ser Asn Lys		
275	280	285
Thr Gly Lys Lys Leu Leu Asp Val Ser Leu Thr His Glu Ala Ile Glu		
290	295	300
Thr Ile Leu Ala Ser Cys Cys Gln Trp Tyr Asp Leu Leu Lys Gly Val		
305	310	315
		320
Cys Lys Asp Leu Glu Lys Thr Gly Thr Gln Ser His Leu Phe Ala Ser		
325	330	335
Phe Gly Ile Gly Asp Cys Ile Pro Leu Thr Pro Phe His Gln Ala Gly		
340	345	350
Leu Gln Ile Thr Lys Leu Asp Val Leu Ser Phe Val Lys Ala Leu Met		
355	360	365
Pro Pro Val Leu Pro Met Ser Gly His Asn His Gln Tyr Ala Tyr Pro		
370	375	380
Thr Asp Ala Val Ala Val Val Gly Met Ala Cys Arg Leu Pro Gly Ala		
385	390	395
		400
Asn Ser Val Glu Glu Leu Trp Asp Leu Ile Ser Ser Gly Gly Ser Thr		
405	410	415
Val Thr Pro Val Pro Glu Asp Arg Met Asp Ile Ala Gly Ser Phe Arg		

420 425 430
 Ala Met Gln Asp Pro Lys Trp Ala Ala Lys Gln Gln Trp Trp Gly Asn
 435 440 445
 Phe Ile Ser Asp Ile Ala Gly Phe Asp His Ser Phe Phe Arg Met Ser

 450 455 460
 Pro Arg Glu Ala Ala Ser Met Asp Pro Gln Gln Arg Ile Leu Leu Glu
 465 470 475 480
 Thr Ala Tyr Gln Ala Met Glu Ser Ser Gly Tyr Leu Gly Ser His Arg
 485 490 495
 Arg Glu Ser Gly Asp Pro Val Gly Val Phe Leu Gly Ala Ser Phe Val
 500 505 510
 Glu Tyr Leu Asp Asn Thr Ser Ser Asn Pro Pro Thr Ala Tyr Thr Ser
 515 520 525

 Thr Gly Thr Ile Arg Ala Phe Leu Ser Gly Lys Ile Ser Tyr Tyr Phe
 530 535 540
 Gly Trp Thr Gly Pro Ser Glu Ile Leu Asp Thr Ala Cys Ser Ser Ser
 545 550 555 560
 Leu Val Ala Ile Asn Arg Ala Cys Lys Ala Ile Gln Asn Asp Glu Cys
 565 570 575
 Pro Met Ala Leu Ala Gly Gly Val Asn Leu Ile Thr Gly Ile His Asn
 580 585 590
 Tyr Leu Asp Leu Ala Lys Ala Gly Phe Leu Ser Pro Ser Gly Gln Cys

 595 600 605
 Lys Pro Phe Asp Gly Ala Ala Asp Gly Tyr Cys Arg Ser Glu Gly Ala
 610 615 620
 Gly Leu Val Val Leu Lys Arg Leu Ser Gln Ala Leu Thr Asp Gly Asn
 625 630 635 640
 Gln Ile Leu Gly Val Ile Thr Gly Ala Ser Thr Asn Gln Gly Gly Leu
 645 650 655
 Ser Pro Ser Leu Thr Val Pro His Ser Ala Ala Gln Val Lys Leu Tyr
 660 665 670

Gln Asn Ile Leu His Gln Ala Gly Met Arg Pro Glu Gln Val Ser Tyr
675 680 685

Cys Glu Thr His Gly Thr Gly Thr Gln Ala Gly Asp Pro Leu Glu Ile
690 695 700

Glu Ser Val Arg Glu Val Phe Gly Gly Pro Lys Arg Gln Asp Ser Met
705 710 715 720

His Ile Gly Ser Ile Lys Gly Asn Ile Gly His Cys Glu Thr Ala Ala
725 730 735

Gly Val Ala Gly Leu Leu Lys Ala Leu Val Met Val Asn Lys Ala Ala
740 745 750

Ile Pro Pro Leu Ala Ser His Lys Ser Leu Asn Pro Lys Ile Ala Ala
755 760 765

Leu Glu Pro Asp Lys Leu Ala Ile Ser Ser Cys Leu Glu Asp Trp Arg
770 775 780

Val Ser Pro Arg Ala Ala Leu Val Asn Ser Tyr Gly Ala Ala Gly Ser
785 790 795 800

Asn Ser Ala Val Leu Leu Cys Gln Ala Pro Asp Ile Asp Asn Ala Pro
805 810 815

Leu His Arg Val Ala Thr Thr Glu His Thr Tyr Pro Ile Ile Leu Ser
820 825 830

Ala Ala Ser Lys Pro Ser Leu Leu Ser Asn Ala Glu Asn Ile Ala Ser
835 840 845

Tyr Leu Arg Lys Ala Thr Ser Lys Cys Thr Ile Ala Asp Val Ala Phe
850 855 860

Thr Leu Val Lys Gln Arg Lys Arg His Pro Leu Gln Trp Ile Thr Met
865 870 875 880

Glu Ser Ser Ile Asp Gly Leu Val Lys Ser Leu Gly Ser Leu Gln Asp
885 890 895

Pro Ser Asn Ala Pro Gln Pro Lys Lys Val Val Met Thr Phe Ser Gly
900 905 910

Gln Ser Arg Gln Ser Ile Gly Leu Asn Lys Glu Trp Tyr Asp Ser Phe
915 920 925

Pro Leu Phe Arg Arg His Val Asn Glu Cys Asp Asp Leu Leu Gln Gln
930 935 940

Ser Gly Phe Pro Ser Cys Lys Ser Ala Met Phe Asp Lys Glu Pro Ala
945 950 955 960

Arg Asp Val Val Pro Leu Gln Cys Ala Met Phe Ala Val Gln Tyr Ala
965 970 975

Ser Ala Leu Ser Trp Ile Asp Cys Gly Leu Gln Val Glu Ala Val Ile
980 985 990

Gly His Ser Phe Gly Glu Leu Thr Ala Leu Ala Val Ser Gly Thr Leu
995 1000 1005

Ser Leu Lys Asp Ala Leu Asn Leu Val Ala Thr Arg Ala Thr Leu Met
1010 1015 1020

Gln Ser Lys Trp Gly Pro His Lys Gly Thr Met Leu Leu Ile Ser Ala
1025 1030 1035 1040

Ala Thr Glu Met Val Arg Lys Ile Ile Ala Gly Asn Arg Asp Val Glu
1045 1050 1055

Ile Ala Cys His Asn Ala Pro Thr Ser Gln Ile Val Val Gly Thr Gln
1060 1065 1070

Ala Ala Ile Ser Glu Val Glu Lys Val Leu Glu Asn Asn Thr Glu Tyr
1075 1080 1085

Arg Gly Ile Gln Ser Gln Arg Leu Asn Val Thr His Gly Phe His Ser
1090 1095 1100

Gln Phe Thr Glu Pro Leu Leu Glu Asn Leu Ser Glu Ser Ala Arg Ser
1105 1110 1115 1120

Leu Val Phe His Glu Pro Lys Ile Leu Leu Glu Cys Cys Thr Leu Glu
1125 1130 1135

Glu Leu Asn His Val Gly Pro Asp His Leu Ala Arg His Thr Arg Glu
1140 1145 1150

Pro Val Tyr Phe Tyr His Ala Val Arg Arg Leu Glu Gln Arg Leu Gly
1155 1160 1165

Thr Cys Leu Trp Leu Glu Ala Gly Phe Asp Ser Pro Ile Ile Pro Met

1170 1175 1180
 Thr Lys Arg Ala Val Glu Phe Pro Glu Arg His Thr Phe Leu Asp Met
 1185 1190 1195 1200
 Lys Thr Pro Ser Gly Thr Asn Pro Thr Lys Met Leu Thr Thr Ala Thr
 1205 1210 1215
 Ile Asn Leu Trp Gln Asn Gly Ala Thr Ser Ser Phe Trp Gly Phe His
 1220 1225 1230
 Pro Ile Glu Thr Thr Asn Ile Lys Gln Val Trp Leu Pro Pro Tyr Gln
 1235 1240 1245

 Phe Asp Arg Thr Ser His Trp Met Pro Tyr Thr Asp His Ala Leu Glu
 1250 1255 1260
 Met Ser Lys Ile Gln Ala Val Ile Ser Asn Ser Glu Pro Leu Val Glu
 1265 1270 1275 1280
 Leu Ser Thr Lys Pro Pro Arg Leu Val Glu Pro Arg Thr Lys Pro Ser
 1285 1290 1295
 Glu Lys Gly Glu Phe Ser Met Asn Thr Gln Ala Arg Arg Tyr Thr Glu
 1300 1305 1310
 Ile Val Ser Gly His Ala Val Leu Ser Arg Pro Leu Cys Pro Ala Ala

 1315 1320 1325
 Met Tyr Met Glu Cys Ala Ile Met Ala Ala Gln Leu Ser Ile Gly Asn
 1330 1335 1340
 Ile Val Gly Gln Ala Pro Trp Phe Glu Asn Leu Thr Phe Glu Ala Pro
 1345 1350 1355 1360
 Leu Gly Met Asp Pro Asp Asn Asp Thr Thr Val Val Leu Lys Asp Asp
 1365 1370 1375
 Gly Ser Lys Ser Arg Trp Ser Phe Val Ala Arg Ser Thr Ser Arg Ser
 1380 1385 1390

 Asn Pro Lys Arg Lys Pro Val Leu His Ala Lys Gly Asp Phe Gly Phe
 1395 1400 1405
 Thr Thr Gln Thr Gln Val His Arg Tyr Glu Arg Leu Val Thr Asp Arg
 1410 1415 1420
 Met Arg His Leu Gln His Ser Lys Ser Glu Thr Leu Lys Ser Lys Arg

1425 1430 1435 1440
 Ala Tyr Gly Leu Phe Ser Arg Ile Val Arg Tyr Ala Glu Leu Leu Lys
 1445 1450 1455
 Gly Ile Ser Ser Ile Thr Leu Gly Asp Ser Glu Ala Ser Ala Ile Ile

 1460 1465 1470
 Asp Val Pro Leu Gly Ala Ser Thr Glu Asp Ser Ser Ala Thr Gly Leu
 1475 1480 1485
 Cys Asp Cys Val Ala Leu Asp Ala Phe Ile Gln Val Val Gly Leu Leu
 1490 1495 1500
 Ile Asn Ser Gly Asp Asp Cys Ala Glu Asp Glu Val Phe Val Ala Thr
 1505 1510 1515 1520
 Gly Val Glu Asn Phe Ser Met Ser Leu Ala Cys Asp Phe Asp Arg Cys
 1525 1530 1535

 Arg Thr Trp Leu Val Phe Ala Met Phe Thr Pro Ser Gly Asn Gly Lys
 1540 1545 1550
 Ala Met Gly Asp Val Phe Ile Leu Thr Arg Asp Asn Val Leu Val Met
 1555 1560 1565
 Thr Ile Met Gly Val Gln Phe Thr Lys Leu Pro Ile Thr Arg Leu Glu
 1570 1575 1580
 Lys Leu Leu Asp Ser Ala Asn Pro Lys Ala His Asn Thr Pro Ile Leu
 1585 1590 1595 1600
 Lys Ser Ser Gln Gln Asp Ser Ile Val Ser Ala Ser Ser Ser Ser Ser

 1605 1610 1615
 Thr Glu His Ser Asp Asp Asp Ser Glu Asp Asp Gly Ser Arg Ser Pro
 1620 1625 1630
 Ser His Ser Asp Thr Ser Val Asp Ser Glu Ser Glu Ala Pro Ala Asp
 1635 1640 1645
 Asn Gly Ala Ala Lys Lys Leu Lys Ser Leu Ile Ala Ser Tyr Val Gly
 1650 1655 1660
 Ile Ala Glu Asp Ala Ile Ser Asp Asp Ala Asn Ile Ala Asp Leu Gly
 1665 1670 1675 1680

Val Asp Ser Leu Ala Ala Thr Glu Leu Ala Asp Glu Ile Ser Asn Asp
1685 1690 1695

Phe Ala Lys Glu Ile Asp Gly Gly Glu Leu Pro Met Met Thr Phe Gly
1700 1705 1710

Glu Leu Cys Arg Ile Val Ala Pro Glu Met Ala Ala Lys Pro Ala Lys
1715 1720 1725

Ala Lys Lys Lys Ile Pro Tyr Lys Gly Lys Asp Glu Ala Thr Val Val
1730 1735 1740

Glu Ser His Pro Gly Lys Ser Gln Ser Glu Ile Lys Asp Leu Lys Ala
1745 1750 1755 1760

Val Val Glu Pro Leu Pro Arg Ser Thr Pro Met Leu Ser Asp Thr Thr
1765 1770 1775

Val Val Arg Ser Asp Pro Thr Gln Val Leu Arg Gln Ile Asp Thr Met
1780 1785 1790

Phe Gln Ser Ser Ala Asp Thr Phe Gly Phe Thr Asp Tyr Trp Thr Ala
1795 1800 1805

Val Ala Pro Lys Gln Asn Lys Leu Val Leu Ala Tyr Ile Gly Glu Glu
1810 1815 1820

Phe Arg Lys Leu Gly Leu Asp Leu Trp Ala Val Gln Pro Gly Ala Thr
1825 1830 1835 1840

Leu Pro His Ile Glu Tyr Leu Pro Lys His Glu Lys Val Val Gln Arg
1845 1850 1855

Leu Trp Asp Ile Leu Ala Asp His Gly Ile Val Tyr Asn Tyr Asp Ala
1860 1865 1870

Ala Lys Val Arg Ser Ser Lys Pro Leu Pro Asp Ala Pro Ser Thr Ala
1875 1880 1885

Leu Leu Asn Glu Leu Asn Ala Leu Phe Pro Asn Phe Ala Asn Glu Asn
1890 1895 1900

Arg Leu Met Ser Val Thr Ala Pro His Phe Ala Asp Gly Leu Arg Gly
1905 1910 1915 1920

Lys Thr Asp His Ile Ser Leu Leu Phe Gly Ser Gln Arg Gly Gln Glu
1925 1930 1935

Cys Leu Asn Asp Phe Tyr Asn Asn Ser Pro Gln Leu Ala Val Met Thr
 1940 1945 1950
 Asp His Leu Leu Thr Phe Phe Lys Gln Leu Leu Lys Glu Ala Pro Leu
 1955 1960 1965

 Glu Gly Ser Leu Arg Ile Leu Glu Val Gly Gly Gly Phe Gly Gly Thr
 1970 1975 1980
 Thr Lys Arg Leu Ala Glu Met Leu Glu Ala Leu Gly Gln Pro Val Glu
 1985 1990 1995 2000
 Tyr Thr Phe Thr Asp Val Ser Ser Met Leu Val Lys Glu Ala Arg Lys
 2005 2010 2015
 Lys Phe Ser Lys His Ser Trp Met Asp Phe Gln Ser Leu Asn Leu Glu
 2020 2025 2030
 Lys Asp Pro Pro Ala Ser Leu Gln Arg Thr Tyr Asp Ile Val Ile Gly

 2035 2040 2045
 Thr Asn Val Val His Ala Thr Ser Asn Ile Val Asn Ser Thr Thr Arg
 2050 2055 2060
 Met Arg Ser Leu Leu Arg Lys Gly Gly Phe Ile Val Leu Ser Glu Val
 2065 2070 2075 2080
 Thr Arg Ile Val Asp Trp Tyr Asp Leu Val Tyr Gly Leu Leu Asp Gly
 2085 2090 2095
 Trp Trp Ala Phe Lys Asp Ser Arg Thr Tyr Pro Leu Gln Pro Ala Asp
 2100 2105 2110

 Asp Trp Val Arg Asp Leu Met Lys Ala Gly Phe Glu Thr Ala Ser Tyr
 2115 2120 2125
 Ser Arg Gly Asp Ser Glu Glu Ser Asn Thr Gln Gln Leu Ile Ile Gly
 2130 2135 2140
 Ser Thr Arg Pro Ser Lys Val Ala Ser Thr Ser Gly Leu Ser Glu Ala
 2145 2150 2155 2160
 Arg Leu Ser Lys Ser Tyr Arg Ile Glu Thr Met Pro Tyr Lys Ile Ile
 2165 2170 2175
 Asp Asp Thr Glu Ile Leu Ala Asp Val Phe Phe Pro Glu His Glu Val

2180	2185	2190	
Ala Ser Glu Ala Met Pro Ile Ala Leu Met Ile His Gly Gly Gly Phe			
2195	2200	2205	
Met Thr Leu Ser Lys Thr Ala Ile Arg Pro Tyr Gln Thr Gln Phe Leu			
2210	2215	2220	
Val Glu Asn Gly Tyr Leu Pro Ile Ser Ile Asp Tyr Arg Leu Cys Pro			
2225	2230	2235	2240
Glu Ile Asp Leu Ile Ala Gly Pro Met Thr Asp Val Arg Asp Ala Leu			
2245	2250	2255	
Thr Trp Val Arg Lys Gln Leu Pro Ala Ile Ala Arg Thr Arg Gly Ile			
2260	2265	2270	
Asn Val Asp Pro Thr Lys Val Val Val Ile Gly Trp Ser Thr Gly Gly			
2275	2280	2285	
His Leu Ala Leu Thr Thr Ala Trp Thr Cys Glu Asp Ile Gly Glu Glu			
2290	2295	2300	
Pro Pro Val Ala Val Leu Ser Phe Tyr Gly Pro Thr Asn Phe Glu Ser			
2305	2310	2315	2320
Glu Asp Ile Asp Arg Arg Arg Ala Glu Gln Tyr Pro Glu Arg Thr Met			
2325	2330	2335	
Ser Phe Asp Arg Ile Arg Lys Ser Leu Ser Thr Lys Pro Ile Thr Ser			
2340	2345	2350	
Tyr Asp Cys Pro Thr Gly Thr Asp Ser Thr Gly Leu Gly Trp Val Arg			
2355	2360	2365	
Pro Gly Asp Pro Arg Ser Glu Leu Val Leu Ser Leu Phe Lys Glu Ser			
2370	2375	2380	
His Gly Leu Gln Val Met Leu Asn Gly Leu Ser Ala Ala Asp Leu Ala			
2385	2390	2395	2400
Lys Pro Pro Pro Leu Ala Lys Ile Gln Ala Ile Ser Pro Met Ala Gln			
2405	2410	2415	
Leu Lys Ala Gly Arg Tyr Asn Val Pro Thr Phe Val Ile His Ser Asp			
2420	2425	2430	
Cys Asp Glu Ile Ala Pro Phe Arg Asp Ser Glu Ala Phe Val Glu Glu			

2435 2440 2445
 Leu Ala Arg Arg Gly Val Lys Thr Gly Leu Gly Arg Val Arg Gly Lys
 2450 2455 2460
 Lys His Ile His Asp Leu Ala Leu Lys Pro Glu Lys Asp Gly Trp Val

 2465 2470 2475 2480
 Asp Gly Ala Gly Val Gly Tyr Glu Phe Ile Phe Asp Val Val Gly Arg
 2485 2490 2495
 Gly Val Arg Gly
 2500
 <210> 3
 <211> 1761
 <212> DNA
 <213> Artificial Sequence
 <220><223> Stereocaulon alpinum atr2 gene(MZ277878)
 <400> 3
 atggctctcc tagacacaat tgaactgttc tccaacttct cccttagtgg agtggtcgt 60
 ggcttagttc tagcttctct cctcgtaagt tcaaccactg ttacgtcgtc aaaatgtgcc 120

 agccttcctt ttgcgagaac tgacacataa gcaactagac cacaacttac tgcatttgga 180
 acatcttcta caatatttac cttcatcctt tgaaggtta ccctggccca aaattcctca 240
 caacttctcg gcttccatat ctcaaatgga tgttctcagg aacccttggtg cccaatttcc 300
 aaaggctcca tgagcagtac ggtcctgttg ttctgtgtgc gcccaacgag ctctcgtaca 360
 tcaaccggga agccttgaaa accatctacg gccaccgtca accaggcgag gggtttcgca 420
 agaatccagc attcttccag ccagctacga atggcggtta ttccatcttg acctccgaag 480
 gtgacgtcga tagtagtgtc cgccggaaga tctgccccgc tttctcgat aaggcgcttg 540

 cggaacagca ggacatcctc cagcatttca cagatctgtt gatccggaaa ttgcgcgaga 600
 gagtcgaagc atccaaaagc tcagaacctg ttgacatgtt cgaatggtac atctggacta 660
 ccttcgactt gattggagat ctgcatttgc gcgagccgtt caactgcctc gaagcagcta 720
 gctttacaga atgggtcgcc cttgtgttca acgcgtttta gacttttgcg ttcataca 780
 tcaggtaagc aaggtagact caccgcgcaa cgaataaaag aaacgcgctt acgaaaccga 840
 actagcaaac agctagctcc ccttgacaag ctagtacgac tgatgatccc aaagtccatg 900
 aaagcccggc aagacaaagt cttctccctg aacgtcgcca aagtcgaccg ccgatttgct 960

tccaaagcag accgtccaga ctctctatcg tacattatta agggcaagga cggggtagca 1020
atggcacttc cagaactcta cgcaaattcc acgtctcttg tcctcgccgg tagcgagagc 1080
acagcatctg gcttggctgg cattaccttt gagctactca agcaccgaga ggctcagaag 1140
aaggccgtgg aagagatacg cagcgctttc aaaaccgaag atgagattgt acctgagagc 1200
gtcaagcgac ttccatacct tgcggccatg gtttccgaag ggctccgaat gtatccgccc 1260
ttcccagagg gacttctctg tcttacacca cggcaagggg ctacagatctg tggccaatgg 1320
gtccctggag gagtaagcta tgaccccat cctccgggac tagctaattct tctcatatag 1380

acatatgtgc aatttagcac ccacgccgt caccgcgcca gcgccaattt tacagaccgc 1440
aatgtcttcg cccctgagcg ctggcttggc gacacgaaat tcgctcgga tatcaaagaa 1500
gcttcgcagc ccttctcgat aggtccacgt agctgcattg gcaggaagtg agtactttgt 1560
tcgctcatt tgcacgaaga gtactaatta tacgggaaac ccagtcttgc gtaccttgaa 1620
atgcgcttga ttctagcgcg tatgttgtgg agcttcgaca tgcaactcac tcccgaatgc 1680
gaagactggg acgaccagaa tagctggatc cagtgggaca agaaaccttt gatggtgaag 1740
ttgtcgcttg tcaaaagata g 1761

<210> 4
<211> 506
<212> PRT
<213> Artificial Sequence
<220><223> Stereocaulon alpinum atr2(QXF68952)
<400> 4

Met Ala Leu Leu Asp Thr Ile Glu Leu Phe Ser Asn Phe Ser Leu Ser
1 5 10 15
Gly Val Phe Ala Gly Leu Val Leu Ala Ser Leu Leu Thr Thr Thr Tyr
20 25 30
Cys Ile Trp Asn Ile Phe Tyr Asn Ile Tyr Leu His Pro Leu Lys Gly
35 40 45
Tyr Pro Gly Pro Lys Phe Leu Thr Thr Ser Arg Leu Pro Tyr Leu Lys
50 55 60
Trp Met Phe Ser Gly Thr Leu Val Pro Asn Phe Gln Arg Leu His Glu
65 70 75 80
Gln Tyr Gly Pro Val Val Arg Val Ala Pro Asn Glu Leu Ser Tyr Ile
85 90 95

Asn Pro Glu Ala Leu Lys Thr Ile Tyr Gly His Arg Gln Pro Gly Glu
 100 105 110
 Gly Phe Arg Lys Asn Pro Ala Phe Phe Gln Pro Ala Thr Asn Gly Val
 115 120 125

 His Ser Ile Leu Thr Ser Glu Gly Asp Ala His Ser Ser Val Arg Arg
 130 135 140
 Lys Ile Leu Pro Ala Phe Ser Asp Lys Ala Leu Ala Glu Gln Gln Asp
 145 150 155 160
 Ile Leu Gln His Phe Thr Asp Leu Leu Ile Arg Lys Leu Arg Glu Arg
 165 170 175
 Val Glu Ala Ser Lys Ser Ser Glu Pro Val Asp Met Phe Glu Trp Tyr
 180 185 190
 Ile Trp Thr Thr Phe Asp Leu Ile Gly Asp Leu Ala Phe Gly Glu Pro

 195 200 205
 Phe Asn Cys Leu Glu Ala Ala Ser Phe Thr Glu Trp Val Ala Leu Val
 210 215 220
 Phe Asn Ala Phe Lys Thr Phe Ala Phe Ile Asn Ile Ser Lys Gln Leu
 225 230 235 240
 Ala Pro Leu Asp Lys Leu Val Arg Leu Met Ile Pro Lys Ser Met Lys
 245 250 255
 Ala Arg Gln Asp Lys Val Phe Ser Leu Asn Val Ala Lys Val Asp Arg
 260 265 270

 Arg Ile Ala Ser Lys Ala Asp Arg Pro Asp Phe Leu Ser Tyr Ile Ile
 275 280 285
 Lys Gly Lys Asp Gly Val Ala Met Ala Leu Pro Glu Leu Tyr Ala Asn
 290 295 300
 Ser Thr Leu Leu Val Leu Ala Gly Ser Glu Ser Thr Ala Ser Gly Leu
 305 310 315 320
 Ala Gly Ile Thr Phe Glu Leu Leu Lys His Arg Glu Ala Gln Lys Lys
 325 330 335
 Ala Val Glu Glu Ile Arg Ser Ala Phe Lys Thr Glu Asp Glu Ile Val

340 345 350
 Pro Glu Ser Val Lys Arg Leu Pro Tyr Leu Ala Ala Met Val Ser Glu
 355 360 365
 Gly Leu Arg Met Tyr Pro Pro Phe Pro Glu Gly Leu Pro Arg Leu Thr
 370 375 380
 Pro Arg Gln Gly Ala Gln Ile Cys Gly Gln Trp Val Pro Gly Gly Thr
 385 390 395 400
 Tyr Val Gln Phe Ser Thr His Ala Ala His Arg Ala Ser Ala Asn Phe
 405 410 415

 Thr Asp Pro Asn Val Phe Ala Pro Glu Arg Trp Leu Gly Asp Thr Lys
 420 425 430
 Phe Ala Ser Asp Ile Lys Glu Ala Ser Gln Pro Phe Ser Ile Gly Pro
 435 440 445
 Arg Ser Cys Ile Gly Arg Asn Leu Ala Tyr Leu Glu Met Arg Leu Ile
 450 455 460
 Leu Ala Arg Met Leu Trp Ser Phe Asp Met Gln Leu Thr Pro Glu Cys
 465 470 475 480
 Glu Asp Trp Asp Asp Gln Asn Ser Trp Ile Gln Trp Asp Lys Lys Pro

485 490 495
 Leu Met Val Lys Leu Ser Leu Val Lys Arg

500 505

<210> 5

<211> 1254

<212> DNA

<213> Artificial Sequence

<220><223> Stereocaulon alpinum atr3 gene(MZ277877)

<400> 5

atgacttccg ttgatacaat gcctccaccg atggttcgcc tcgagagtca gcctgacgac 60

ttgatgggct ctctgatgt cgcgacgtt tcagaccttc ttccaggcca tacaatgga 120

ctcgaggatg aagtgaatat tccagcaacc aatggactca agagccatcc agtggttacg 180

acaggaacag aaaaaaccgg tgtgatgccg ccccttcagc ccgaaagcaa gaagaacaac 240

aaaggcgttc cgtgtaaga cctggacata caccgctctt tttgtcaagc catgactgtt 300

acgacttcct gttgttttct atcattttct tgccagtagt gctgaggtag acaggtatca 360
 cgctcaccg aacgatatcg atcccgtagc tcgcggtctg ctcgaaaatt acagcaaat 420
 cccctcggat caagtacagc agcacgttat cgcaattgta agtttccaaa ggttactcct 480
 tgatgttctg caaccggatt gcagttggcg cactagatgg ctctctccct agacaagcat 540
 cgctaaaacc gtctttttct cccaaaacag cgcgaaaaag cctggggacgt ctaccctac 600

ccctgcatcg gccatttctt ctctctgaac ctaaccataa acctctcccc ctactaccg 660
 tccctcgtct cccgcctccg cgatcaaat caaacactcc tcgacctcgg ctgctgcttc 720
 gcccagacg tccgcaaat cgtctctgat ggcgccccga gtcaaaacat ctacggcgcc 780
 gacctctatg gcgagttcat ggatctaggc tttgaactct ttcgcgaccg aaagacgctc 840
 aaaagcacgt ttttcccgac tgatatacta aacgagcggg atctgctgtt gaaggccttg 900
 gatggcgaga tggatgttgt ttatttgggg ctgtttttgc atcattttga ttttgagact 960
 tgtgtcaaag tttgtacgag cgttactagg ctctgaagc ccaagccggg gagtctggtt 1020

atgggggtcc aggtggggag tttggtgggt gatactaaac ccatttctat tcctagtgg 1080
 gggattttgt ggagcatga tattgcgagt ctgaaaggg tttgggagga ggttgagcgc 1140
 ctgacgggga cgaagtggaa ggttgaggct aggttgagga ggggcaaagg gttcggggag 1200
 aagtggcagc ttgaggggac gaggagggtt gggtttgagg tttataggct ttga 1254

<210> 6

<211> 346

<212> PRT

<213> Artificial Sequence

<220><223> Stereocaulon alpinum atr3(QXF68951)

<400> 6

Met Thr Ser Val Asp Thr Met Pro Pro Pro Met Val Arg Leu Glu Ser

1	5	10	15
Gln Pro Asp Asp Leu Met Gly Ser Ser Asp Val Ala Asp Val Ser Asp			
20	25	30	
Leu Leu Pro Gly His Thr Asn Gly Leu Glu Asp Glu Val Lys Ile Pro			
35	40	45	
Ala Thr Asn Gly Leu Lys Ser His Pro Val Val Thr Thr Gly Thr Glu			
50	55	60	
Lys Thr Gly Val Met Pro Pro Leu Gln Pro Glu Ser Lys Lys Asn Asn			

65						70						75						80	
Lys	Gly	Val	Pro	Trp	Tyr	His	Ala	Ser	Pro	Asn	Asp	Ile	Asp	Pro	Val				
					85					90					95				
Thr	Arg	Gly	Leu	Leu	Glu	Asn	Tyr	Ser	Lys	Ile	Pro	Ser	Asp	Gln	Val				
					100					105					110				
Gln	Gln	His	Val	Ile	Ala	Ile	Arg	Glu	Lys	Ala	Trp	Asp	Val	Tyr	Pro				
					115					120					125				
Tyr	Pro	Cys	Ile	Gly	Gln	Phe	Leu	Phe	Leu	Asn	Leu	Thr	Ile	Asn	Leu				
					130					135					140				
Ser	Pro	Tyr	Tyr	Pro	Ser	Leu	Val	Ser	Arg	Leu	Arg	Asp	Gln	Asn	Gln				
145					150					155					160				
Thr	Leu	Leu	Asp	Leu	Gly	Cys	Cys	Phe	Ala	Gln	Asp	Val	Arg	Lys	Leu				
					165					170					175				
Val	Ser	Asp	Gly	Ala	Pro	Ser	Gln	Asn	Ile	Tyr	Gly	Ala	Asp	Leu	Tyr				
					180					185					190				
Gly	Glu	Phe	Met	Asp	Leu	Gly	Phe	Glu	Leu	Phe	Arg	Asp	Arg	Lys	Thr				
					195					200					205				
Leu	Lys	Ser	Thr	Phe	Phe	Pro	Thr	Asp	Ile	Leu	Asn	Glu	Arg	Asp	Leu				
					210					215					220				
Leu	Leu	Lys	Gly	Leu	Asp	Gly	Glu	Met	Asp	Val	Val	Tyr	Leu	Gly	Leu				
225					230					235					240				
Phe	Leu	His	His	Phe	Asp	Phe	Glu	Thr	Cys	Val	Lys	Val	Cys	Thr	Arg				
					245					250					255				
Val	Thr	Arg	Leu	Leu	Lys	Pro	Lys	Pro	Gly	Ser	Leu	Val	Met	Gly	Val				
					260					265					270				
Gln	Val	Gly	Ser	Leu	Val	Gly	Asp	Thr	Lys	Pro	Ile	Pro	Ile	Pro	Ser				
					275					280					285				
Gly	Gly	Ile	Leu	Trp	Arg	His	Asp	Ile	Ala	Ser	Leu	Glu	Arg	Val	Trp				
290					295					300									
Glu	Glu	Val	Gly	Ala	Leu	Thr	Gly	Thr	Lys	Trp	Lys	Val	Glu	Ala	Arg				
305					310					315					320				

Leu Glu Arg Gly Lys Gly Phe Gly Glu Lys Trp Gln Leu Glu Gly Thr

325 330 335

Arg Arg Leu Gly Phe Glu Val Tyr Arg Leu

340 345

<210> 7

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Sol1PpDS33_fwd

<400> 7

agatctcggc acgttttgat cgcaaag 27

<210> 8

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Sol1PpDS33_rev

<400> 8

aagcttagta atatctggcg atgaagcgtc ag 32

<210> 9

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Inf_Sta046440_fwd

<400> 9

cagatattac taagctcatg gcgtctaacc atgaggag 38

<210> 10

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Inf_Sta046440_rev

<400> 10

gtaacgttaa gtggatcctt acaactaaag ctctcaatca accc 44

<210> 11

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Sta046440_check1F
 <400> 11
 ggacgatgag aacgtcctag 20
 <210> 12
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> Sta046440_check2F
 <400> 12
 cgagtcaggt gatccagtg 19
 <210> 13
 <211> 19
 <212>
 > DNA
 <213> Artificial Sequence
 <220><223> Sta046440_check3F
 <400> 13
 ccaatcatac tgagcgctg 19
 <210> 14
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> Sta046440_check4F
 <400> 14
 cacatgtctg tggcttgag 19
 <210> 15
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220><223> Sta046440_check5F
 <400> 15

caactccggt gatgattg	18
<210> 16	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Sta046440_check6F	
<400> 16	
caacccttcc acatattgag	20
<210> 17	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> pII99_sol5_fwd	
<400> 17	
gaattcagag cgacagtga agc	23
<210> 18	
<211> 37	
<212> DNA	
<213> Artificial Sequence	
<220><223> pII99_sol5_rev	
<400> 18	
agatctagct agcttggcgt gatgtcagtc aatgctg	37
<210> 19	
<211> 27	
<212> DNA	
<213> Artificial Sequence	
<220><223> pII99_tefla_fwd	
<400> 19	
gaattccagc cgagacagca gaatcac	27
<210> 20	
<211> 41	
<212> DNA	
<213> Artificial Sequence	

<220><223> pII99_tefla_rev
 <400> 20
 agatctagct agcgtgttat gttttgtgga atataaaagg g 41
 <210> 21
 <211> 40
 <212> DNA

<213> Artificial Sequence
 <220><223> Inf95_Sta04642_fwd
 <400> 21
 atcacgcaa gctagcacga tgacttccgt tgatacaatg 40
 <210> 22
 <211> 41
 <212> DNA

<213> Artificial Sequence
 <220><223> Inf98_Sta04642_fwd
 <400> 22
 aaacataaca cgctagcacg atgacttccg ttgatacaat g 41
 <210> 23
 <211> 39
 <212> DNA

<213> Artificial Sequence
 <220><223> Inf99_Sta04642_rev
 <400> 23
 ggagaccggc agatcccgtc acatgtccgt cataataac 39

<210> 24
 <211> 42
 <212> DNA

<213> Artificial Sequence
 <220><223> Inf98_Sta04643_fwd
 <400> 24
 aaacataaca cgctagaacc atggetctcc tagacacaat tg 42
 <210> 25
 <211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Inf95_Sta04643_fwd

<400> 25

atcacgccaa gctagaacca tggctctcct agacacaatt g 41

<210> 26

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Inf99_Sta04643_rev

<400> 26

ggagaccggc agatcgagga gcaaaggaac aaggaacag 39

<210> 27

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> pII99_MK5

<400> 27

cgtcaagaga cctacgagac tg 22

<210> 28

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> pII99_MK3

<400> 28

gctatacttc taggtcttgg aagag 25

<210> 29

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Inf_BLE_pII99_fwd

<400> 29

gtaggtctct tgacgcatta agacctcagc gctagtgg 38

<210> 30
 <211> 39
 <212> DNA
 <213> Artificial Sequence
 <220><223> Inf_BLE_pII99_rev
 <400> 30
 gacctagaag tatagcgggtg ttacggagca ttcactagg 39
 <210> 31
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> L5
 <400> 31
 gaggtcttca caccacaagt tcg 23

 <210> 32
 <211> 52
 <212> DNA
 <213> Artificial Sequence
 <220><223> L3
 <400> 32
 ggcaaaggaa tagagtagat gccgaccggg ctcttgcgta gtggtacaga tg 52
 <210> 33
 <211> 51
 <212> DNA
 <213> Artificial Sequence
 <220><223> R5
 <400> 33
 gttgacctcc actagctcca gccaagcgaa taggtaacaa agccagccga g 51
 <210> 34
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> R3
 <400> 34

gagaattgcg gcgcaggatg ttc	23
<210> 35	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> N5	
<400> 35	
ctgactttct tcttcagacc ctgc	24
<210> 36	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> N3	
<400> 36	
ctaacgatgt cgtagtcgc cttg	24
<210> 37	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> HYG-F	
<400> 37	
gcttggtgg agctagtga g	21
<210> 38	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> HYG-R	
<400> 38	
cggtcggcat ctactctatt cctt	24
<210> 39	
<211> 25	
<212> DNA	
<213> Artificial Sequence	

<220><223> YG-F

<400> 39

cgatgtagga gggcgtggat atgtc 25

<210> 40

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> HY-R

<400> 40

tgtagtgtat tgaccgattc cttgcg 26

<210> 41

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> P1_fwd

<400> 41

ggacgatgag aacgtcctag 20

<210> 42

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> P1_rev

<400> 42

gagagttcct catgctcgtc 20

<210> 43

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> P2_fwd

<400> 43

catcgacgac acggagatac tg 22

<210> 44

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> P2_rev

<400> 44

gaccgtgaga ttctttgaag agg 23

<210> 45

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Actin1_fwd

<400> 45

caatggttcg ggtatgtgca ag 22

<210> 46

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Actin1_rev

<400> 46

gaagagcgaa accctcgtag at 22