Appl Microbiol Biotechnol (2015) 99:3715–3728
DOI 10.1007/s00253-014-6206-5

BIOENERGY AND BIOFUELS

Biosynthesis and genomic analysis of medium-chain hydrocarbon production by the endophytic fungal isolate *Nigrograna mackinnonii* E5202H

Jeffery J. Shaw · Daniel J. Spakowicz · Rahul S. Dalal · Jared H. Davis · Nina A. Lehr · Brian F. Dunican · Esteban A. Orellana · Alexandra Narváez-Trujillo · Scott A. Strobel

Received: 21 August 2014 / Revised: 29 October 2014 / Accepted: 31 October 2014 / Published online: 12 February 2015

Abstract An endophytic fungus was isolated that produces a series of volatile natural products, including terpenes and odd chain polyenes. Phylogenetic analysis of the isolate using five loci suggests that it is closely related to *Nigrograna mackinnonii* CBS 674.75. The main component of the polyene series was purified and identified as (3E,5E,7E)-nona-1,3,5,7-tetraene (NTE), a novel natural product. Non-oxygenated hydrocarbons of this chain length are uncommon and desirable as gasoline-surrogate biofuels. The biosynthetic pathway for NTE production was explored using metabolic labeling and gas chromatography time of flight mass spectometer (GCMS). Two-carbon incorporation ^{13}C acetate suggests that it is derived from a polyketide synthase (PKS) followed by decarboxylation. There are several known mechanisms for such decarboxylation, though none have been discovered in fungi. Towards identifying the PKS responsible for the production of NTE, the genome of *N. mackinnonii* E5202H (ATCC SD-6839) was sequenced and assembled. Of the 32 PKSs present in the genome, 17 are predicted to contain sufficient domains for the production of NTE. These results exemplify the capacity of endophytic fungi to produce novel natural products that may have many uses, such as biologically derived fuels and commodity chemicals.

Keywords Endophyte · Natural product · Volatile organic compound · Polyene · Medium-chain hydrocarbon · Biofuel · Polyketide synthase

Introduction

Endophytic fungi are a diverse group of organisms that have been shown to reside within every lineage of the roughly 300,000 plant species on Earth (Arnold and Lutzoni 2007). The number and diversity of endophytes in each plant has been shown to increase with decreasing latitude reaching a pinnacle in equatorial rainforests with examples of more than 250 endophyte species isolated from a single species of tropical tree (Arnold et al. 2000). This extraordinary biodiversity has led to increased interest in the diverse natural products that endophytes produce. Many endophytes have been shown to produce molecules that have many uses including as antibiotics, immunosuppressants, and antiparasitics (reviewed in Strobel and Daisy 2003; Aly et al. 2011).

One class of molecules produced by endophytes that are of particular interest is volatile organic compounds (VOCs) (Korpi et al. 2009). These are typically low molecular weight compounds including alcohols, ketones, esters, acids, and hydrocarbons that can be derived from either biosynthetic or degradative pathways. For example, 3-methyl-1-butanol is a common VOC produced from branched-chain amino acid metabolism (Connor et al. 2010) and 1-octen-3-ol, which gives mushrooms their characteristic odor, is derived from oxidative breakdown of linoleic acid by a lipoxygenase and hyperoxide lyase (Tressl et al. 1982; Wurzenberger and

Jeffery J. Shaw and Daniel J. Spakowicz contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00253-014-6206-5) contains supplementary material, which is available to authorized users.

J. J. Shaw · D. J. Spakowicz · R. S. Dalal · J. H. Davis · N. A. Lehr · B. F. Dunican · S. A. Strobel (✉)
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA
e-mail: scott.strobel@yale.edu

E. A. Orellana · A. Narváez-Trujillo
Laboratorio de Biotecnología Vegetal, Pontificia Universidad Católica del Ecuador, Quito, Ecuador

Grosch 1984). The ecological role of many of these VOCs is uncertain, though 1-octen-3-ol is known to be a fungal hormone that affects the development of conidia (Chitarra et al. 2004). Fungal VOCs have attracted interest for a variety of potential applications, including use as characteristic markers of fungal growth in the built environment (e.g., workspaces and residential structures) (Polizzi et al. 2012). VOC production from common, allergenic fungi, especially in the genera *Aspergillus*, *Fusarium*, and *Penicillium,* has been characterized for this purpose (Larsen and Frisvad 1995; Fiedler et al. 2001; Lancker et al. 2008; Wihlborg et al. 2008; Schuchardt and Kruse 2009).

In another application, some fungal VOCs have been used as volatile antibiotics (Wheatley et al. 1997; Strobel et al. 2001; Mitchell et al. 2010; Stoppacher et al. 2010). A number of fungi of the *Muscodor* and *Trichoderma* genera were found to produce VOCs that are toxic to many organisms, including bacteria and fungi. Some of these strains have been used as biocontrol agents to reduce mold growth during fruit transport or promote plant growth in agriculture (Harman et al. 2004; Gabler et al. 2010). In addition, *Muscodor albus* has been used as a selection tool to enrich for the isolation of endophytes that also produce VOCs and are thereby resistant to the *M. albus* gases (Strobel et al. 2008).

Some fungal VOC profiles contain molecules that are also in fuel mixtures, leading to the hypothesis that fungi may be a potential source of biofuels (Strobel et al. 2008). In one example, a sampling of isolates in the genus *Ascocoryne* revealed a series of C8 compounds as well as C6 to C9 alkanes and branched alcohols, many of which are present in gasoline formulations (Strobel et al. 2008; Griffin et al. 2010). Similarly, many fungi have been shown to produce volatile terpene molecules, which are used as both commodity chemicals and biofuels (Gershenzon and Dudareva 2007; Peralta-Yahya et al. 2011). For example, several endophytic *Hypoxylon* spp. have been shown to produce 1,8-cineole, a monoterpene and the main component of eucalyptus oil, which is used as a flavoring and fragrance molecule and has been explored as a gasoline additive (Tomsheck et al. 2010; Tess Mends and Yu 2012; Riyaz-Ul-Hassan et al. 2013). The production of hydrocarbons from these and other pathways has been explored through heterologous expression in yeast and bacteria (Atsumi et al. 2008; Beller et al. 2010; Peralta-Yahya et al. 2011). The optimization of these pathways often utilizes genes from several organisms, highlighting the need for a genetic understanding of the biosynthesis of these molecules (Fortman et al. 2008; Peralta-Yahya et al. 2012).

We set out to isolate new endophytes producing novel VOCs which may be useful in any of these applications. Here, we report the discovery of isolate E5202H, an endophytic *Nigrograna mackinnonii* that produces several secondary metabolites that may have been used as biofuels. We explored the biosynthesis of the most abundant VOC using metabolic labeling and genomic analysis and identify candidate genes for its production.

Materials and methods

E5202H isolation

A 10-×-1-cm stem from a 5-m-tall *Guazuma ulmifolia* tree (Yale catalogue number YU.100464) was collected from the Cerro Blanco Protected Forest near Guayaquil, Ecuador (−02.1752333, −80.0218833). Two weeks after collection, the stem was surface sterilized and plated on dilute potato dextrose agar (2.4 g/L potato dextrose broth (PDB; EMD Millipore) and 15 g/L agar (BD Difco)) in the presence of 3-day-old *M. albus* as described previously (Ezra et al. 2004). Isolate E5202H was observed growing from the stem after 11 days. It has been deposited in the American Type Culture Collection (ATCC) as SD-6839.

Morphology and phylogenetic analysis

Fungal hyphae were examined in water, and pictures were taken with a stereo- and a light microscope (Nikon Diaphot 300). Genomic DNA was isolated from a 9-day-old culture using a Plant DNeasy kit (Qiagen) as described previously (Gianoulis et al. 2012). The internal transcribed spacer (ITS) ribosomal DNA (rDNA), small subunit (SSU) rDNA, large subunit (LSU) rDNA, RNA polymerase II (RPB2) nuclear gene, and translation elongation factor i (TEF1) nuclear gene regions were amplified (primer sequences in Table S1). The PCR amplicons were cleaned and sequenced by the W.M. Keck Foundation as described previously (Griffin et al. 2010) (see Supplemental methods). Trees were constructed by Bayesian and maximum likelihood methods as described previously (organisms in Table S2; see Supplemental methods) (Griffin et al. 2010). The files have been submitted to TreeBase (www.treebase.org).

Culturing and GCMS for VOC analysis

Cultures of E5202H were grown for VOC analysis on a variety of solid and liquid media types, including PDB and potato dextrose agar (PDA; 24 g/L potato dextrose), oatmeal agar (OA; BD Difco), and liquid or agar defined media containing glucose (Glu or Glu-A) (15 g/L, J.T. Baker Chemicals) or cellobiose (CB or CBA; 20 g/L, Acros Organics) (see Supplemental methods). Vials were grown at 23 °C and sampled after 4, 9, and 20 days.

Compounds in the headspace above growing fungal cultures were sampled by solid phase microextraction using a 50/

30-μm divinylbenzene/carboxen/polydimethylsiloxane StableFlex SPME Fiber (Supelco) on a GCT Premier gas chromatography time of flight mass spectometer (GCMS) (Waters) with a ZB-624 column (30 m×0.25 mm ID× 1.40 μm film thickness; Phenomenex) (see Supplemental methods). Electron ionization (EI) spectral data were collected over the mass range 50–650 Da and data were analyzed using the MassLynx Software Suite (Waters). Retention indices were measured by comparison of retention times to those of an alkane mix (Fluka), and potentially interesting compounds were identified by comparison of retention times and mass spectra to pure standards when available (Sigma-Aldrich).

NTE purification and structural elucidation

Cultures of E5202H were grown in 1 L PDB in 2 L Erlenmeyer flasks shaking at 150 rpm at 30 °C for 10 days, filtered through cheesecloth, extracted with 1 L methylene chloride (Fisher Scientific), and rotary-evaporated at 10 °C to 0.5 mL. Concentrated extracts were separated by high-performance liquid chromatography (HPLC) on a Gilson preparative C-18 column with a gradient of 10–100 % acetonitrile (J.T. Baker Chemicals) in water over 20 min, holding at 100 % for 5 min, with a flow rate of 20 mL/min. Fractions were dried on a V10 Evaporator (Biotage) and resuspended in deuterated chloroform (Sigma). Accurate-mass measurements were performed on a GCT Premier GC TOF mass spectrometer (Waters). NMR studies were performed on a 500-MHz spectrometer with a 5-mm HCN probe (Bruker).

An authentic standard of (3E,5E,7E)-nona-1,3,5,7-tetraene (NTE) was synthesized by Richman Chemical, Inc. (Lower Gwynedd, PA) to 68 % EEE and 32 % ZEE isomers as previously reported (Spangler et al. 1986). NMR of this compound was performed on a 400-MHz spectrometer with a 5-mm HCN probe (Agilent). Ultraviolet–visible (UV–Vis) spectroscopy was performed on a Cary 3E Spectrophotometer. The synthesized molecule was hydrogenated to nonane in ethyl acetate in the presence of hydrogen and 10:1 (*v*/*w*) 10 % Pd/C catalyst (Supplemental methods).

(3E,5E,7E)-nona-1,3,5,7-tetraene

Purified: 1H NMR (500 MHz, $CDCl_3$) δ 6.36–6.25 (m, 1H), 6.25–6.16 (m, 2H), 6.15 (s, 1H), 6.15–6.09 (m, 2H), 6.07 (d, *J*=10.2 Hz, 1H), 6.02 (d, *J*=11.9 Hz, 1H), 5.67 (dq, *J*=13.9, 6.9 Hz, 1H), 5.12 (d, *J*=16.8 Hz, 1H), 4.99 (d, *J*=10.0 Hz, 1H), and 1.72 (d, *J*=6.8 Hz, 3H). Synthesized: 1H NMR (400 MHz, $CDCl_3$), TMS δ 0, δ 6.43–6.25 (m, 1H), 6.20 (td, *J*=12.9, 12.4, 7.0 Hz, 2H), 6.15–6.08 (m, 1H), 6.08–5.92 (m, 1H), 5.73 (tt, *J*=13.8, 6.6 Hz, 1H), 5.20 (s, 0H), 5.19–5.09 (m, 1H), 5.05 (d, *J*=10.2 Hz, 1H), 1.77 (d, *J*= 6.6 Hz, 3H). 13C NMR (400 MHz, $CDCl_3$) δ 137.12, 133.74, 133.51, 132.48, 131.76, 130.35, 130.05, 116.61, and 22.63. λ_{max}: 306, 293, 281, 269, and 258sh.

Metabolic labeling of nonatetraene

Isolate E5202H was inoculated into GCMS vials containing glucose media with 5 mM unlabeled glucose to allow for fungal growth. On day 2, before detectable production of nonatetraene, media was supplemented with 5 or 10 mM ^{13}C-labeled glucose (Cambridge Isotope Laboratories) (see Supplemental methods for additional media conditions). The headspace was sampled for label incorporation on day 3, observed in the EI spectral data as a mass shift compared with unlabeled control. A similar procedure was carried out for labeling with 1,2-^{13}C acetate, with the culture growth in glucose media supplemented on day 2 with 10, 5, or 2.5 mM acetate. The 5-mM concentration was chosen for differential labeling with 1-^{13}C- or 2-^{13}C-labeled acetate (Cambridge Isotope Laboratories).

Genome sequencing, assembly, and annotation

Genomic DNA was prepared in two libraries, 180 bp fragments and 3 kB mate-pairs following the company's specifications (Illumina). The two libraries were barcoded and pooled into a single lane on the Illumina HiSeq 2000 and sequenced in paired-end mode. Reads were assembled using ALLPATHS-LG v44034 (Gnerre et al. 2010). The genome sequence has been deposited in GenBank as accession number JGVQ00000000 and reads deposited in the SRA as accession number SRP040662. Gene models were identified with the self-training algorithm GeneMark-ES v2 (Ter-Hovhannisyan et al. 2008) and conserved domains with PFAM v27.0 (Punta et al. 2012). Gene clusters were identified with SMURF v1.0 (Khaldi et al. 2010), and BLAST searches were performed with BLAST+ v2.2.29 (Altschul et al. 1997). Assembly accuracy was verified by aligning all available reads to the scaffolds using Bowtie2 v2.2.4 (Langmead and Salzberg 2012) and analyzing with SAMtools v1.1 (Li et al. 2009).

Results

Isolation and phylogenetic assignment of fungal strain E5202H

The fungus E5202H was isolated from a 5-m-tall *G. ulmifolia* malvaceous tree in a secondary growth forest in the Cerro Blanco Protected Forest near Guayaquil, Ecuador. No visible fungal pathology was noted on the tree. The stem was surface-sterilized before plating, suggesting that E5202H may have existed within the plant tissue as an endophyte.

The isolate produces fluffy, aerial mycelia that exude a brown pigment after 3 days of growth on potato dextrose agar, and the pigment persists for months (Fig. 1a). By light microscopy, the hyphae appear hyaline and septate (Fig. 1b). No spores or fruiting bodies were observed on tens of different media and growth conditions. Molecular phylogeny using five nuclear loci showed E5202H to be closely related to *N. mackinnonii* CBS 674.75 ((Borelli) de Gruyter et al. 2013) with strong support from both Bayesian and maximum likelihood methods (Fig. 1c) (Borelli 1976; de Gruyter et al. 2013). The phylogenetic distance from *N. mackinnonii* CBS 674.75 is consistent with the distance between two other known isolates of the species, suggesting that E5202H is an isolate of *N. mackinnonii*. However, the clade is not resolved from *Biatriosproa marina* (Hyde and Borse 1986). Distinction between these monotypic genera can be observed by the aseptate conidia of *N. mackinnonii* as compared with the septate ascospores of *B. marina*. As E5202H has been observed to produce neither, we rely on the molecular systematics and submit E5202H as an isolate of *N. mackinnonii* until such time that E5202H reproductive structures are observed or the discovery of more *Biatriospora/Nigrograna* isolates further refines the molecular phylogeny of the clade.

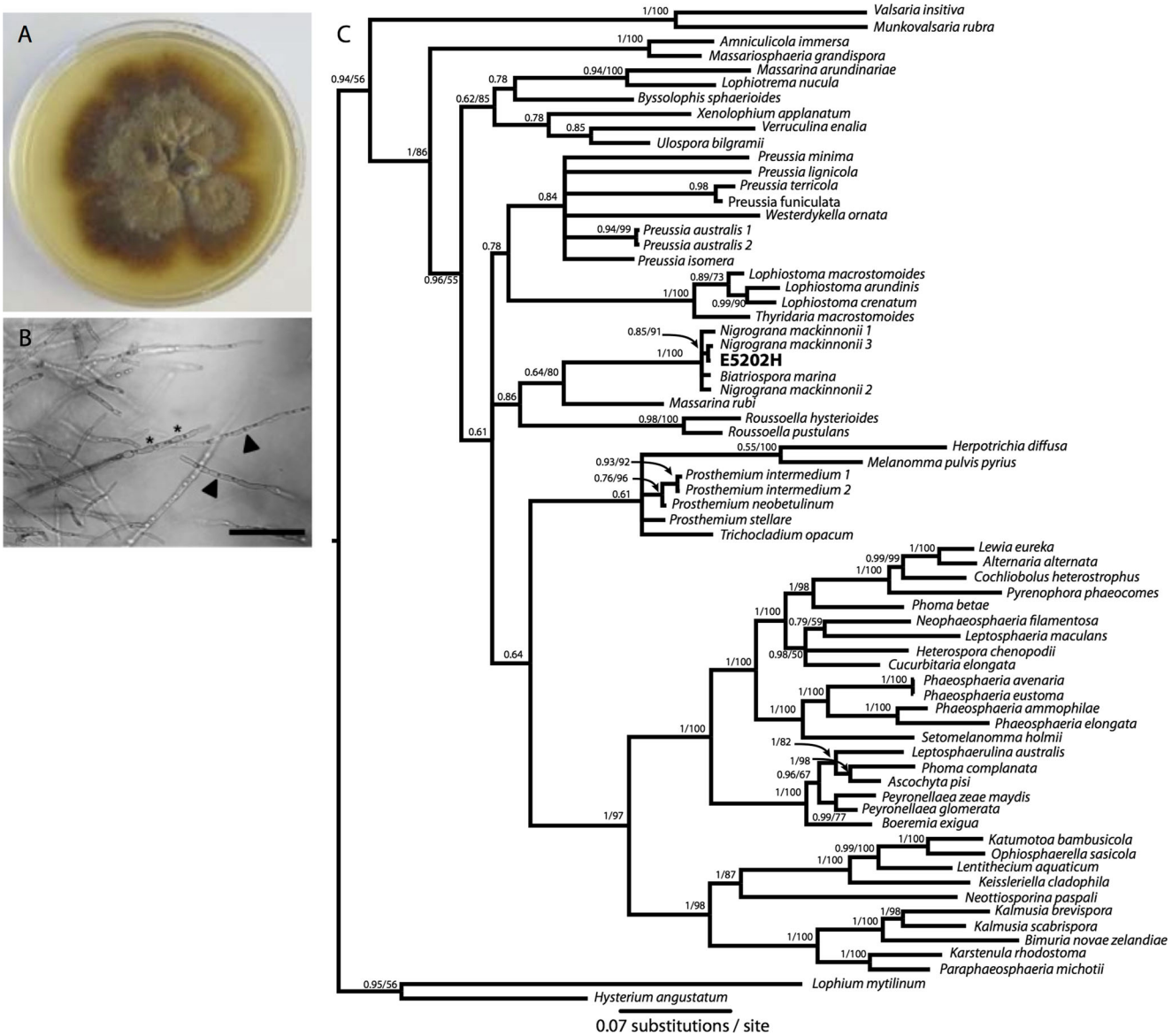


Fig. 1 Morphology and phylogenetics of E5202H. **a** Four-month-old culture grown on PDA at room temperature. **b** The magnification of aerial hyphae shows septa (*asterisk*) and vacuoles (*arrow heads*). *Bar*=50 μm. **c** Interleaved SSU, LSU, RPB2 TEF1, and ITS phylogram of isolate E5202H in the context of members of the order *Pleosporales*. Nodes are labeled with Bayesian posterior probabilities (*top*) and maximum likelihood bootstrap values (*bottom*)

Production of VOCs by E5202H

To assess VOC production, E5202H was grown in sealed vials on a variety of solid and liquid media types. The culture headspace was sampled after 4, 9, and 24 days of growth to account for varying growth and production rates on the various media. The widest variety of compounds were produced when grown on PDA or OA, with consistent production also observed on PDB and a defined medium with either glucose or cellobiose as the carbon source (Table 1; Fig. 2a, b). Fewer molecules were observed with defined glucose or cellobiose agar media. Co-elution with available standards was used to identify the molecules when available; otherwise, the chemical formula is reported from an accurate-mass measurement of the predicted molecular ion. Among the VOCs produced are a series of sesquiterpenes ($C_{15}H_{24}$) and monoterpenes, identified as α- and β-pinene, D-limonene, p-cymene, and terpinolene. Two unidentified compounds with molecular formula $C_{10}H_{14}O_2$ were widely produced under the tested conditions. These were the only volatile compounds observed on solid cellobiose media.

Most interesting was a series of compounds with the formula C_9H_{12}, and perhaps related C_7H_{10} and C_9H_{10} molecules. Hydrocarbons with these formulas are not commonly reported fungal VOCs (Combet et al. 2006), nor is their biosynthetic route immediately obvious. They are highly unsaturated and non-oxygenated, which is unlike typical fatty acid or alpha-keto acid elongation metabolites, and they are shorter and less

Table 1 VOC production by E5202H

Peak no.[a]	Compound ID	RT	RI	ID method[b]	Solid media				Liquid media		
					PDA	OA	CBA	Glu-A	PDB	Glu	CB
	Benzene	10.42	687	MS	X	X					
	C_7H_{10}	12.81	744	MS	X	X		X		X	
1	1,3,5-Heptatriene	15.21	798	MS	X	X		X		X	X
2	C_9H_{12}	19.27	900	MS	X	X			X	X	X
	α-Pinene	20.89	946	STD	X	X					
3	C_9H_{12}	21.29	950	MS	X	X			X		X
4	C_9H_{12}	21.48	955	MS	X	X			X		
	C_9H_{12} (Propylbenzene)	22.18	977	STD	X	X					
5	C_9H_{12}	22.68	990	MS	X	X			X		
	β-Pinene	22.71	995	STD	X						
	Benzaldehyde	23.91	1018	MS		X					
	C_9H_{10}	23.98	1020	MS	X					X	
6	D-Limonene	24.47	1046	STD	X	X					
	p-Cymene	24.65	1050	STD	X	X					
	C_9H_{10}	25.13	1062	MS	X	X				X	
	C_9H_{10}	25.30	1068	MS	X	X				X	
7	C_9H_{12} (1,3,5,7-Nonatetraene)	25.47	1074	NMR/STD	X	X		X	X	X	X
8	C_9H_{12} (NTE-isomer)	25.59	1078	STD	X	X			X	X	X
9	C_9H_{12} (NTE-isomer)	25.70	1081	STD	X	X			X	X	X
10	C_9H_{12} (NTE-isomer)	25.80	1084	STD	X	X			X	X	X
11	C_9H_{12}	25.91	1087	MS	X	X				X	
12	C_9H_{12}	26.04	1091	MS	X	X				X	
13	Terpinolene	26.46	1105	STD	X	X					
14	$C_{10}H_{14}O_2$	32.51	1301	MS	X	X	X	X			
15	$C_{10}H_{14}O_2$	33.61	1343	MS	X	X	X	X	X	X	X
	$C_{15}H_{24}$	37.61	1518	MS	X					X	X
	$C_{15}H_{24}$	37.76	1525	MS	X						
	$C_{15}H_{24}$	37.94	1533	MS						X	
	$C_{15}H_{24}$	38.11	1540	MS	X						

RT retention time, *RI* retention index, *X* compound observed at least three times in this condition, *MS* mass spectral database, *NMR* structure by NMR, *STD* authentic standard

[a] Peak indicated in Fig. 2

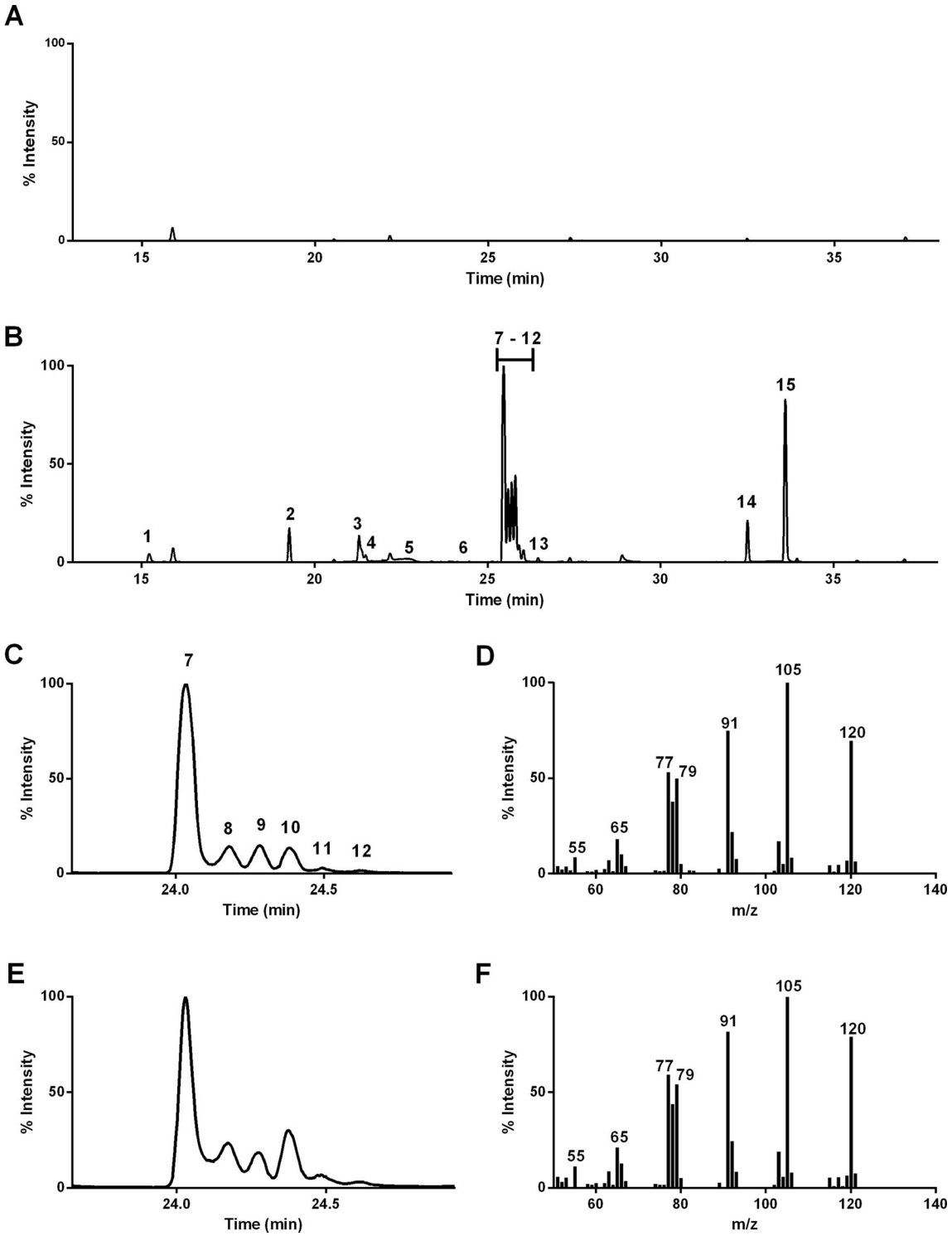


Fig. 2 Representative GC chromatogram of E5202H volatiles. GC chromatogram of volatiles in the headspace of a PDA media blank **a** and a E5202H sample **b** grown on PDA for 9 days. Peak numbers correspond to compounds in Table 1. Zoom of 24–24.5 min of chromatograms containing NTE peaks and associated fragmentation spectra (with indicated masses) from the E5202H sample (**c**, **d**) and synthesized NTE (**e**, **f**)

saturated than terpenes (Felnagle et al. 2012; Peralta-Yahya et al. 2012). The C_9H_{12} peak eluting at 25.47 min (RI 1074, Table 1) was generally the most abundant molecule in the spectrum and eluted with a series of five other peaks (Fig. 2c, peaks 7–12) with similar fragmentation patterns and a molecular ion of 120 Da. Chemical Ionization GCMS showed enrichment of the *m*/*z* 120 fragment suggesting that this is indeed the molecular ion. Accurate-mass GCMS of the *m*/*z* 120 fragment in electron impact ionization mode expanded the measured precision of the mass to 120.192, which is most

consistent with the formula C_9H_{12}. The fragmentation pattern of these molecules included *m/z* 105 Da, the base peak most likely derived from loss of a methyl radical, *m/z* 91, characteristic of rearrangement to a tropylium ion common in unsaturated hydrocarbons, and *m/z* 79 and 77, which are all consistent with unsaturated six-carbon linear and cyclized fragments, respectively (Fig. 2d). While the spectra of all of the C_9H_{12} peaks contained the same mass fragments, the relative abundance was slightly different for those peaks eluting between 21 and 24 min. The base peak in these spectra was *m/z* 91, similar to the spectrum of the known compound propylbenzene. The C_7H_{10} compound fragmentation pattern contained a molecular ion of *m/z* 94 and base peak of *m/z* 79 from loss of a methyl radical. Other mass fragments include the peaks at *m/z* 91 and 77, which were also observed in the C_9H_{12} spectra and are likely due to tropylium and phenyl ions respectively.

We attempted to match these compounds to a variety of commercially available standards of the same molecular formula. Included among these authentic standards were the C_7H_{10} molecules 1-methyl-1,4-cyclohexadiene, bicyclohept-2-ene, and 1,3-cycloheptadiene and the C_9H_{12} compounds propylbenzene, 1,2,4-trimethylbenzene, and 1-ethyl-3-methylbenzene. While a small peak in the production spectrum matched the retention time and fragmentation pattern of propylbenzene (RT 22.18, Table 1), none of the commercially available C_9H_{12} standards matched the retention time of the most abundant 25.47 peaks. Because of the unique and unidentified nature of these compounds, we pursued further characterization of their structure and biosynthesis.

Structural determination of NTE

The 25.47-min C_9H_{12} peak was observed to partition into the methylene chloride phase of an extraction of liquid potato dextrose cultures, which could then be concentrated by cold rotary evaporation. The concentrated extraction was separated by HPLC and the fraction containing a single C_9H_{12} peak was confirmed by GCMS. The GC chromatogram of this fraction continued to show the series of five peaks, as opposed to one peak in the HPLC, with the first peak at 25.47 being 95 % of the total area (data not shown). Such patterns may be due to incomplete separation by HPLC or thermal rearrangements within the GCMS inlet (Frankel et al. 1981).

The purified C_9H_{12} compound was analyzed by NMR. The observed proton chemical shifts and patterns did not show the presence of non-carbon bonding or a benzene ring, consistent with the GCMS elemental composition and standards analyses. Observed was a single methyl group with a chemical shift of 1.72, as well as a terminal double-bonded carbon with a doublet of doublets at 4.99 and 5.13, which restrict the chemical space to polyene hydrocarbons. The rest of the proton spectrum was consistent with the molecule being (3E,5E,7E)-nona-1,3,5,7-tetraene (NTE) (Table 2; Fig. S1a). Though NTE has not been observed as a natural product, its chemical synthesis has been published, and the 1H-NMR data are consistent with published spectra (Spangler and Little 1982; Spangler et al. 1986; Block et al. 1986; Keitel et al. 1990; Pohnert and Boland 1994).

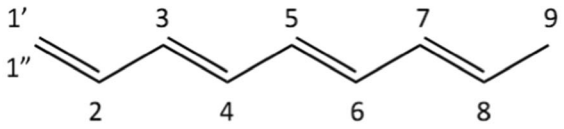
As an additional method to confirm the identity of the C_9H_{12} compound, an authentic NTE standard was synthesized. The proton NMR and UV–Vis spectra of the synthesized NTE matched that of the fungal-purified molecule and reported values (Spangler and Little 1982) (Figs. S1a, b and S2). Furthermore, hydrogenation of the compound yielded nonane, as confirmed with a nonane GCMS standard, which is the expected product for a straight-chain, nine carbon polyene but not for a branched or cyclized isomers. The synthesized NTE was used as an authentic GCMS standard for comparison to the fungal VOCs and was an exact match with the major fungal product by both retention time and mass spectral fragmentation pattern (Fig. 2c–f). In addition, the five other C_9H_{12} peaks in the series were also present in the synthesized sample, indicating that they are likely related isomers and supporting the idea that they may arise during GCMS sampling.

Stable isotope labeling of NTE suggests polyketide-like biosynthesis

To our knowledge, NTE has not been previously observed as a natural product. To decipher its biosynthetic pathway, we monitored the incorporation of ^{13}C-labeled precursors into the molecule by mass spectrometry. Isolate E5202H was grown in GCMS vials with defined media containing 5 mM unlabeled glucose and supplemented with 5 or 10 mM labeled glucose on day 2. Sampling of the headspace on day 3 showed partial label incorporation into NTE, observed as a shift in the fragment ions to heavier masses. Efficient incorporation of ^{13}C from glucose into NTE demonstrated that production is likely the result of biosynthesis from the organism, rather than catabolism of a compound in the media. Furthermore, the rate of production and incorporation would argue against biosynthetic production of a larger molecule followed by non-enzymatic breakdown.

We went on to test incorporation of ^{13}C-labeled acetate into NTE. Under conditions of 5 mM unlabeled glucose and 2.5 mM universally labeled acetate, we observed a nine unit mass shift in the molecular ion peak (*m/z* 120 to 129), indicating that NTE was fully labeled (Fig. 3a, b). That acetate was able to out-compete glucose for incorporation into NTE suggests that it is a more immediate precursor in the biosynthetic pathway and provides further evidence for a biosynthetic rather than a catabolic synthetic route.

Table 2 Structure and chemical shifts of NTE



Position	Shift	H's	Integral	Class	J coupling
1′	4.99	1	1.05	dd	1.47, 10.18
1″	5.13	1	0.89	dd	1.21, 17.17
2	6.31	1	0.92	dt	9.96, 9.96, 16.91
3, 7	6.05	2	1.67	m	
4, 5, 6	6.15	3	3.27	m	
8	5.67	1	1.15	dq	6.73, 6.73, 6.57, 13.67
9	1.72	3	3.05	d	6.93

We next tested the pattern of incorporation with differentially labeled acetate, ^{13}C labeled at either the 1-C or 2-C position. Molecules derived from head-to-head condensation of acetate (i.e., fatty acid or polyketide synthesis) incorporate label from both carbon positions, while pathways such as alpha-keto elongation only incorporate the 2-C position (Kroumova et al. 1994). We observed a shift in the mass of the molecular ion to *m*/*z* 124 in the case of the 1-C label and *m*/*z* 125 in the case of the 2-C label. This indicates that the molecule contains four atoms derived from the carboxyl carbon of acetate and five atoms from the methyl carbon (Fig. 3c, d). Furthermore, the masses of the fragmentation ions produced by differentially labeled NTE reveal a sequential pattern of carbon incorporation into the molecule. The base peak of *m*/*z* 105 in the unlabeled molecule is shifted to *m*/*z* 109 in both the 1-C- and 2-C-labeled spectra, demonstrating that the methyl ion lost during fragmentation is derived from the methyl carbon of acetate. When two carbons are lost, they are derived from both 1-C and 2-C positions of the acetate precursor, giving peaks of *m*/*z* 94 and 95, respectively. A similar alternating pattern of incorporation was also observed in the C_7H_{10} compound peaks. These results reveal that these molecules are derived from head-to-tail condensation of the acetate precursors followed by a decarboxylation, most likely arising from the polyketide biosynthetic pathway.

De novo genome assembly and PKS analysis

As a first step towards understanding the genetic basis of NTE production, the genome of E502H was sequenced and assembled de novo. Fragment and mate-pair reads were assembled into 42 scaffolds with 272X coverage of the estimated 52.4 MB genome (Table 3). Over 16,000 gene models were identified and analyzed for polyketide synthases by the presence of the highly conserved keto-synthase (KS) module (Castoe et al. 2007). Of the 32 putative polyketide synthases identified, 17 are predicted to contain sufficient modifying domains (ketoreductase and dehydratase) to produce an unsaturated polyene (Shen 2003).

Discussion

An endophytic fungus was isolated from Ecuador using *M. albus* volatile selection. Testing this isolate for VOC production revealed a novel set of molecules, one of which was purified and shown to be NTE, a hydrocarbon of a chain-length compatible with addition to gasoline fuels. Metabolic labeling suggested that the molecule is synthesized via a polyketide synthase (PKS)-like mechanism, and genomic analysis revealed 17 PKS candidates that may be responsible for its production.

The lack of diagnostic mophological characters led to a molecular phylogeny using five nuclear loci. The E5202H isolate was found to be within the order *Pleosporales*, one of the largest fungal Orders containing over 4700 described species (Kirk and Ainsworth 2008). Members of this order are known to inhabit every fungal niche, including as endophytes. Many familial circumscriptions within this Order have recently been modified by molecular techniques as several basal characters have been shown to be paraphyletic (Schoch

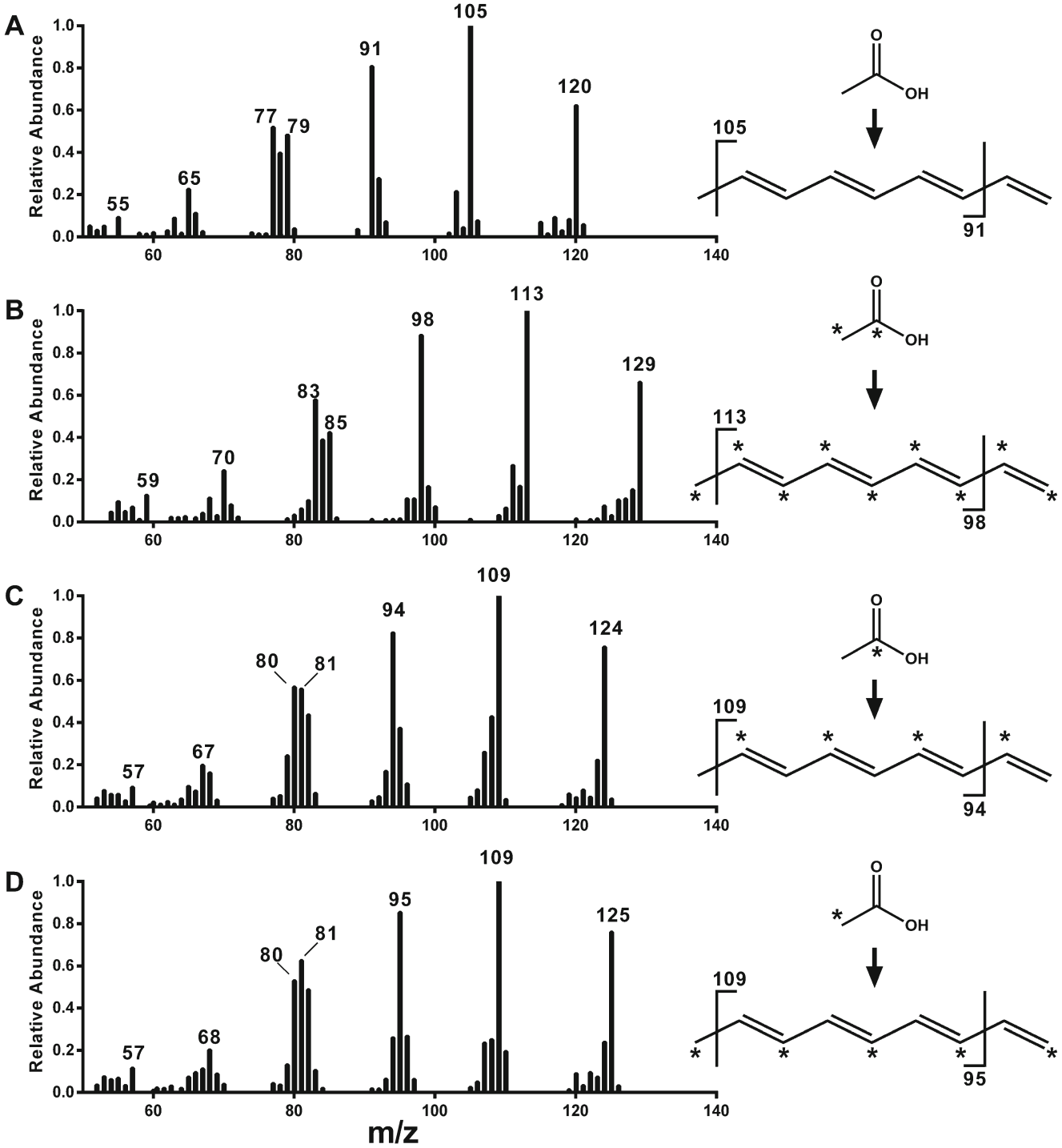


Fig. 3 Mass spectra of nonatetraene and ^{13}C-labeled acetate incorporation. Mass spectral fragmentation of the major nonatetraene peak eluting at 25.42 min. E5202H was grown on defined media with glucose, supplemented with 5 mM sodium acetate either **a** unlabeled, **b** universally ^{13}C labeled, **c** 1-^{13}C labeled, or **d** 2-^{13}C labeled. Each spectrum is depicted next to the corresponding acetate precursor and NTE molecule with the labeling patter indicated (*asterisk*). The fragments giving major *m/z* peaks are indicated on each NTE molecule

et al. 2009; Zhang et al. 2009). This lends support to our strictly molecular method of identification.

Isolate E5202H closely clustered with *N. mackinnonii*, a genus established in 2012 upon discovery of a distinct clade within anamorphs that have a *Phoma*-like morphotype (de Gruyter et al. 2013). The basionym of this species is *Pyrenochaeta mackinnonii,* a known human pathogen that causes mycetomas (Borelli 1976). While strong node support closely ties E5202H to *N. mackinnonii* CBS 675.74, a polytomy with *B. marina* and other *N. mackinnonii* isolates was unable to

Table 3 Genome assembly statistics and PKS analysis

Estimated genome size	52.4 MB
Est coverage by fragment reads	272
Number scaffolds (>1.5 KB)	42
N50 scaffolds	2.3 MB
Number of contigs (>1 KB)	395
% genome estimated to be repetitive	5
% fragment reads assembled	45
% 3 kB jump reads assembled	14
Gene models	16,773
Predicted PKS genes	32
PKS containing at least KS, AT, KR, and DH domains	17
—with potential release mechanism in cluster	2

be resolved; further analyses will be necessary to resolve the circumscriptions of the *Nigrograna* and *Biatriospora* lineages. To date, no VOC analysis from this clade has been reported.

The VOC profile of E5202H consists mainly of hydrocarbon compounds that fall into two groups; terpenes (C10 and C15) and odd carbon-number polyenes not likely derived from isoprene (C7 and C9). The VOC's produced by the terpene pathway in E5202H include both mono- and sesquiterpenes. We were unable to positively identify the sesquiterpene compounds because the large number of potential isoforms makes screening authentic standards difficult. The monoterpenes, including pinene, limonene, terpinolene, and p-cymene, were identified by comparison to commercially available standards and are known fungal metabolites (Korpi et al. 2009; Stoppacher et al. 2010; Ul-Hassan et al. 2012). The $C_{10}H_{14}O_2$ compound that was produced in all of the tested conditions may also belong to the terpene class, though we were unable to match it to a commercially available standard. Known natural products with this molecular formula have derived from monoterpene oxygenation or iridoid biosynthesis (McElvain et al. 1941; Geu-Flores et al. 2012).

Terpene compounds, including those produced by E5202H, are commonly used as commodity chemicals in a variety of industries including as fragrances, flavorings, and cleaning products. D-limonene (RT 24.47) is a major component of citrus oil and has a strong orange scent. It is conventionally produced by extraction of citrus rind, but more recently, production has been engineered into *Escherichia coli* (Pourbafrani et al. 2010; Alonso-Gutierrez et al. 2013). Cymene (RT 24.65) is used as an intermediate in the chemical synthesis of a variety of products. It is commonly produced by alkylation of petroleum derivatives benzene or toluene, although production from limonene has been explored recently as a more renewable source (Fernandes et al. 2007). Both mono- and sesquiterpenes have also been explored as advanced biofuels or biofuel precursors (Peralta-Yahya and Keasling 2010; Peralta-Yahya et al. 2012). Pinene (RT 20.89, 24.65), the main component of turpentine, is of particular interest as a renewable fuel. It can be dimerized through a simple chemical process, and the resulting compound has density and volumetric heating value similar to tactical fuel JP-10 (Harvey et al. 2010). Optimization of the terpene biosynthetic pathway in heterologous hosts such as *Saccharomyces cerevisiae* and *E. coli* is being developed as a a renewable source of these chemicals, and identifying genes from new fungal sources such as E5202H may contribute to these efforts (Redding-Johanson et al. 2011; Peralta-Yahya et al. 2011).

The odd carbon-number polyene hydrocarbons produced by E5202H, the major component of which was NTE, appear to be unique. Compounds with the same formulas (C_9H_{12}, C_9H_{10}, and C_7H_{10}) have been observed to be produced by a *Hypoxylon* sp. isolate; however, in each tested case, authentic standards of these reported compounds did not match the products from our organism (Tomsheck et al. 2010). While NTE has been chemically synthesized previously, to our knowledge it has not been observed as a natural product (Spangler and Little 1982; Spangler et al. 1986; Block et al. 1986; Keitel et al. 1990; Pohnert and Boland 1994). Eleven and nine carbon polyenes have been found to be produced by brown algae including *Cutleria multifida* and *Ectocarpus siliculosus*, where the compounds ectocarpene and 7-methylcycloocta-1,3,5-triene were shown to act as pheromones (Müller et al. 1971; Keitel et al. 1990). An NTE isomer (3Z,5Z,7E)-nona-1,3,5,7-tetraene was proposed as an unstable intermediate in the biosynthesis of the C9 compounds but was not observed as a product. While 7-methylcycloocta-11,3,5-triene was not available as a GCMS standard for detection in the E5202H headspace, it was clearly not the dominant C9 product of this fungus.

Other polyunsaturated, all-E polyenes have been biologically produced, such as the enediyne precursors (Ahlert et al. 2002; Liu et al. 2002, 2005; Van Lanen et al. 2007; Zhang et al. 2008). However, no such molecules with fewer than 14 carbons have been observed (Horsman et al. 2009). Other conjugated, unbranched tetraene molecules such as this have been observed as part of larger macromolecules such as the antifungal polyene macrolides, including nystatin, amphotericin B and natamycin (reviewed in Aparicio et al. 2004). To date, all of the polyene antibiotics have been synthesized by *Streptomyces* bacteria. In the context of the full macrolide, these molecules have been shown to interact with sterols in the plasma membrane to form pores, where the carboxy and amino groups form a hydrogen bond network (Bolard 1986). Though we cannot rule out the possibility that NTE is a precursor molecule that is later attached to such a macrolide, no such molecules were observed in the extractions of the fungus.

Our feeding experiments with ^{13}C-labeled acetate revealed a sequential pattern of carbon incorporation into the NTE molecule from acetate with the loss of one carboxyl carbon. Based on the structure of NTE and our labeling data, the most likely biosynthetic mechanism of NTE production is via head-to-tail condensation of acetate molecules followed by alkene-generating decarboxylation (Fig. 4). The condensation

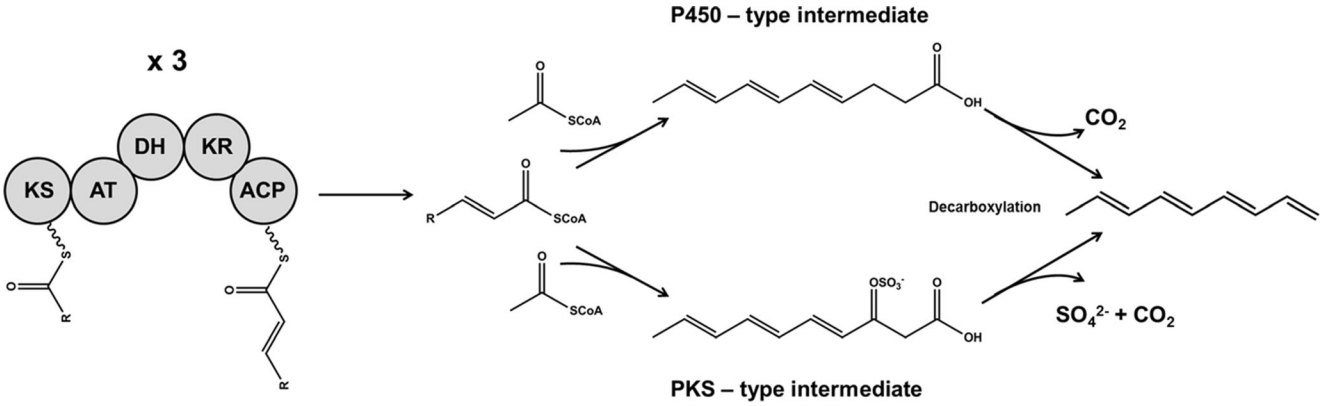


Fig. 4 A proposed biosynthetic path to NTE. The carbon chain of the molecule is produced by three rounds of malonyl-CoA condensation followed by reduction of the β-keto group to a double bond. A fourth condensation produces a ten-carbon chain intermediate that undergoes decarboxylation to form the terminal double bond of the nine-carbon NTE. The exact structure of the intermediate depends on the decarboxylation mechanism that may be similar to the known P450 (P450-type) or PKS thioesterase domain (PKS-type) activities

reaction is most likely catalyzed by a partially reducing PKS with ketoreductase and dehydratase activities to generate the conjugated double bond structure.

NTE elutes from the GC column immediately before five other C_9H_{12} peaks, which have the same fragmentation patterns and isotope-labeling patterns (Fig. 2, peaks 7–12). These compounds are likely stereoisomers of NTE, and their presence in the synthesized molecule sample suggests they arise from thermal isomerization in the GCMS. In addition, several C_9H_{12} molecules elute earlier in the spectrum, one of which matches the standard propylbenzene. This series of molecules (Fig. 2, peaks 3–5) are likely cyclized derivatives of NTE, which could be formed either enzymatically or non-enzymatically. The C_7H_{10} molecules may also be structurally and biosynthetically related to NTE. Each showed the same alternating pattern of acetate incorporation in the labeling experiment and the best match to their fragmentation patterns in the Wiley Compound Database was 1,3,5-heptatriene. This molecule was previously observed as a fungal metabolite and could be synthesized by the same PKS and decarboxylation mechanism, but incorporating one fewer malonyl-CoA extender units (Larsen and Frisvad 1995; Fiedler et al. 2001).

The genetic basis for the production of hydrocarbons such as NTE has not previously been explored in fungi. Towards that end, we sequenced and analyzed the genome of E5202H. To our knowledge, this is the largest genome in the Order *Pleoporales* at 52 MB (average size, 34.4±5.4 MB from 27 other genomes) with over 3000 more predicted genes than the next closest genome (*Bipolaris maydis*, PRJNA42739, 13,456 predicted genes). Of the 16,670 predicted gene models, 32 were identified as putative PKSs, with 17 predicted to have sufficient domains to produce NTE (Table S3). All but three of the PKSs were predicted to reside within biosynthetic clusters, including 16 of the 17 of those predicted to have sufficient domains for NTE production. The assembly accuracy over each of these 17 genes was verified by re-aligning the sequence data to the assembly (Table S4).

Homology modeling was insufficient to determine which of these 17 PKSs is likely to be responsible for the production of NTE. The coding sequences for several enediyne genes are known and their domain organization is well conserved (KS-AT-ACP-KR-DH). However, none of the E5202H PKSs matched this pattern (Table S3).

Although it is fairly straightforward to imagine how the precursor could be biosynthesized from acetate precursors, the mechanism of release for NTE remains unclear. At present, there are three known mechanisms for the decarboxylation reaction that generate terminal double bonds. The first is catalyzed by a PKS module within the coding sequence of the protein, as in the CurM and Ols enzymes that were recently identified in cyanobacteria (Gu et al. 2009; Mendez-Perez et al. 2011). CurM has been shown to form a terminal alkene through sulfonation of β-hydroxy-acyl-ACP by a sulfotransferase domain, followed by decarboxylation and sulfate elimination by a thioesterase domain. The coding sequence of Ols also contains homology to a sulfotransferase domain, leading to the prediction that it follows the same mechanism. Second is a thioesterase activity that acts in trans, such as the SgcE10 and NcsE10 proteins responsible for enediyne precursor release in Actinomycetes (Zhang et al. 2008). A third mechanism involves the activity of a decarboxylating cytochrome P450 enzyme such as the OleT enzyme discovered in *Jeotgalicoccus* sp. (Rude et al. 2011). There are mechanisms other than decarboxylation for making hydrocarbons, including decarbonylation of a fatty aldehyde (Schirmer et al. 2010; Qiu et al. 2012) and head-to-head condensation (Beller et al. 2010), though these produce longer carbon-chain molecules (more than 15 and 23 carbons respectively) and have not been observed to generate a terminal alkene. Pathways for generating hydrocarbons such as the alkene series produced by E5202H have not been explored in fungi.

None of the 17 PKSs encode a sulfo-transferase domain as observed for CurM nor a thioesterase domain as in OleT (Table S3). However, for both cases, the release was catalyzed by an enzyme within the biosynthetic cluster that acted in trans (Zhang et al. 2008; Rude et al. 2011). All but one of the PKSs resides within a biosynthetic cluster (PKS 9, nm9236 does not); however, none of the associated genes show significant homology (e<10) to the enediyne thioesterases SgcE10 or NcsE10. One of the clusters contains a gene (nm8442) with domain homology to the thioesterase-like superfamily (PFAM ID 4HBT_2), thereby presenting a possible release mechanism for PKS 1 (Table S5). Moreover, the E5202H genome contains 126 protein-coding genes homologous to OleT, of which two reside within PKS clusters (clusters of PKSs 8 and 10, genes nm8450 and nm10052, Table S5). Interestingly, the cluster containing PKS 8 contains two potential release mechanisms: both a thioesterase-family protein and an OleT homolog. Moreover, PKS 10 is predicted to contain a methyl transferase (ME) domain, which would not be predicted to be active for the production of NTE. Sequence alignment suggests that the PKS 10 ME domain would be active (Fig. S3). This leads PKS 8 to be the top candidate for the biosynthesis of NTE, but experiments are required to test this prediction.

The metabolites produced by E5202H have the potential for use in a variety of applications, including commodity chemicals and biofuels. Like the terpene molecules discussed earlier, volatile hydrocarbons such as NTE have been recently studied as biofuels or biofuel precursors (Connor and Liao 2009). The chain length of nine carbons is similar to the compounds that make up current gasoline formulations, and would give NTE greater energy density and hydrophobicity compared with ethanol. While the number of unsaturations in NTE may be problematic for biofuel applications, chemical hydrogenation to nonane is predicted to improve some of its fuel properties, such as the flash point. A similar approach was used to produce a bisabolane biofuel from the bisabolene precursor (Peralta-Yahya et al. 2011). Aside from biofuel applications, NTE may have broader uses in the petrochemical industry. C9 petroleum resins, typically aromatic hydrocarbons, are widely used as "tackifying" agents in paints, ink, adhesives, rubber, and a variety of other uses. If future efforts successfully overproduce biologically derived NTE, either through optimization and modification of this organism or the expression of the biosynthetic genes in heterologous hosts, it could provide an additional, renewable source of chemical precursors for this industry.

Acknowledgments This research was performed with a collecting and research permit provided to SAS by the Ministerio del Ambiente of Ecuador. The authors would like to thank Percy Vargas Nunez for help with collection and identification of the host plant, Gary Strobel for the *M. albus* isolate used for the selection, Joseph Wolenski for providing the microscope facilities, Nicholas J. Carriero, and Robert D. Bjornson in the Yale University Biomedical High Performance Computing Center funded by NIH grants RR19895 and RR029676-01, and the Office of Assistant Secretary of Defense for Research and Engineering NSSEFF grant N00244-09-1-0070 awarded to SAS. DJS and BFD were supported in part by the NIH Cell and Molecular Biology training grant number T32 GM007223.

References

Ahlert J, Shepard E, Lomovskaya N, Zazopoulos E, Staffa A, Bachmann BO, Huang K, Fonstein L, Czisny A, Whitwam RE, Farnet CM, Thorson JS (2002) The calicheamicin gene cluster and its iterative type I enediyne PKS. Science 297:1173–1176. doi:10.1126/science.1072105

Alonso-Gutierrez J, Chan R, Batth TS, Adams PD, Keasling JD, Petzold CJ, Lee TS (2013) Metabolic engineering of *Escherichia coli* for limonene and perillyl alcohol production. Metab Eng 19C:33–41. doi:10.1016/j.ymben.2013.05.004

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402

Aly AH, Debbab A, Proksch P (2011) Fungal endophytes: unique plant inhabitants with great promises. Appl Microbiol Biotechnol 90: 1829–1845. doi:10.1007/s00253-011-3270-y

Aparicio JF, Mendes MV, Antón N, Recio E, Martín JF (2004) Polyene macrolide antibiotic biosynthesis. Curr Med Chem 11:1645–1656

Arnold AE, Lutzoni F (2007) Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? Ecology 88: 541–549

Arnold A e, Maynard Z, Gilbert G s, Coley P d, Kursar T a (2000) Are tropical fungal endophytes hyperdiverse? Ecol Lett 3:267–274. doi: 10.1046/j.1461-0248.2000.00159.x

Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451:86–89. doi:10.1038/nature06450

Beller HR, Goh E-B, Keasling JD (2010) Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. Appl Environ Microbiol 76:1212–1223. doi:10.1128/AEM. 02312-09

Block E, Aslam M, Eswarakrishnan V, Gebreyes K, Hutchinson J, Iyer R, Laffitte JA, Wall A (1986) Alpha.-Haloalkanesulfonyl bromides in organic synthesis. 5. Versatile reagents for the synthesis of conjugated polyenes, enones, and 1,3-oxathiole 1,1-dioxides. J Am Chem Soc 108:4568–4580. doi:10.1021/ja00275a051

Bolard J (1986) How do the polyene macrolide antibiotics affect the cellular membrane properties? Biochim Biophys Acta 864:257–304

Borelli D (1976) *Pyrenochaeta mackinnonii* nova species agente de micetoma. Castellania 4:227–234

Castoe TA, Stephens T, Noonan BP, Calestani C (2007) A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. Gene 392:47–58. doi:10.1016/j.gene.2006.11.005

Chitarra GS, Abee T, Rombouts FM, Posthumus MA, Dijksterhuis J (2004) Germination of *Penicillium paneum* Conidia is regulated by 1-Octen-3-ol, a volatile self-inhibitor. Appl Environ Microbiol 70:2823–2829. doi:10.1128/AEM. 70.5.2823-2829.2004

Combet E, Henderson J, Eastwood DC, Burton KS (2006) Eight-carbon volatiles in mushrooms and fungi: properties, analysis, and biosynthesis. Mycoscience 47:317–326. doi:10.1007/s10267-006-0318-4

Connor MR, Liao JC (2009) Microbial production of advanced transportation fuels in non-natural hosts. Curr Opin Biotechnol 20:307–315. doi:10.1016/j.copbio.2009.04.002

Connor MR, Cann AF, Liao JC (2010) 3-Methyl-1-butanol production in *Escherichia coli*: random mutagenesis and two-phase fermentation. Appl Microbiol Biotechnol 86:1155–1164. doi:10.1007/s00253-009-2401-1

De Gruyter J, Woudenberg JHC, Aveskamp MM, Verkley GJM, Groenewald JZ, Crous PW (2013) Redisposition of phoma-like anamorphs in *Pleosporales*. Stud Mycol 75:1–36. doi:10.3114/sim0004

Ezra D, Hess WM, Strobel GA (2004) New endophytic isolates of *Muscodor albus*, a volatile-antibiotic-producing fungus. Microbiology 150:4023–4031. doi:10.1099/mic. 0.27334-0

Felnagle EA, Chaubey A, Noey EL, Houk KN, Liao JC (2012) Engineering synthetic recursive pathways to generate non-natural small molecules. Nat Chem Biol 8:518–526. doi:10.1038/nchembio.959

Fernandes C, Catrinescu C, Castilho P, Russo PA, Carrott MR, Breen C (2007) Catalytic conversion of limonene over acid activated Serra de Dentro (SD) bentonite. Appl Catal Gen 318:108–120. doi:10.1016/j.apcata.2006.10.048

Fiedler K, Schütz E, Geh S (2001) Detection of microbial volatile organic compounds (MVOCs) produced by moulds on various materials. Int J Hyg Environ Health 204:111–121. doi:10.1078/1438-4639-00094

Fortman JL, Chhabra S, Mukhopadhyay A, Chou H, Lee TS, Steen E, Keasling JD (2008) Biofuel alternatives to ethanol: pumping the microbial well. Trends Biotechnol 26:375–381. doi:10.1016/j.tibtech.2008.03.008

Frankel EN, Neff WE, Selke E (1981) Analysis of autoxidized fats by gas chromatography-mass spectrometry: VII. Volatile thermal decomposition products of pure hydroperoxides from autoxidized and photosensitized oxidized methyl oleate, linoleate and linolenate. Lipids 16:279–285. doi:10.1007/BF02534950

Gabler FM, Mercier J, Jiménez JI, Smilanick JL (2010) Integration of continuous biofumigation with *Muscodor albus* with pre-cooling fumigation with ozone or sulfur dioxide to control postharvest gray mold of table grapes. Postharvest Biol Technol 55:78–84. doi:10.1016/j.postharvbio.2009.07.012

Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol 3:408–414. doi:10.1038/nchembio.2007.5

Geu-Flores F, Sherden NH, Courdavault V, Burlat V, Glenn WS, Wu C, Nims E, Cui Y, O'Connor SE (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. Nature 492:138–142. doi:10.1038/nature11692

Gianoulis TA, Griffin MA, Spakowicz DJ, Dunican BF, Alpha CJ, Sboner A, Sismour AM, Kodira C, Egholm M, Church GM, Gerstein MB, Strobel SA (2012) Genomic analysis of the hydrocarbon-producing, cellulolytic, endophytic fungus *Ascocoryne sarcoides*. PLoS Genet 8:e1002558. doi:10.1371/journal.pgen.1002558

Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB (2010) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci 201017351. doi:10.1073/pnas.1017351108

Griffin MA, Spakowicz DJ, Gianoulis TA, Strobel SA (2010) Volatile organic compound production by organisms in the genus *Ascocoryne* and a re-evaluation of myco-diesel production by NRRL 50072. Microbiology 156:3814–3829. doi:10.1099/mic. 0.041327-0

Gu L, Wang B, Kulkarni A, Gehret JJ, Lloyd KR, Gerwick L, Gerwick WH, Wipf P, Håkansson K, Smith JL, Sherman DH (2009) Polyketide decarboxylative chain termination preceded by o-sulfonation in curacin a biosynthesis. J Am Chem Soc 131:16033–16035. doi:10.1021/ja9071578

Harman GE, Howell CR, Viterbo A, Chet I, Lorito M (2004) Trichoderma species—opportunistic, avirulent plant symbionts. Nat Rev Microbiol 2:43–56. doi:10.1038/nrmicro797

Harvey BG, Wright ME, Quintana RL (2010) High-density renewable fuels based on the selective dimerization of Pinenes. Energy Fuel 24:267–273. doi:10.1021/ef900799c

Horsman GP, Van Lanen SG, Shen B (2009) Chapter 5 iterative type I polyketide synthases for enediyne core biosynthesis. In: Hopwood DA (ed) Methods Enzymol. Academic Press, pp 97–112

Hyde KD, Borse BD (1986) Marine fungi from Seychelles: 5 *Biatriospora marina* gen. and sp. nov. from mangrove wood

Keitel J, Fischer-Lui I, Boland W, Müller DG (1990) Novel C9 and C11 hydrocarbons from the Brown Alga *Cutleria multifida*; sigmatropic and electrocyclic reactions in nature. Part VI. Helv Chim Acta 73:2101–2112. doi:10.1002/hlca.19900730806

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol FGB 47:736–741. doi:10.1016/j.fgb.2010.06.003

Kirk PM, Ainsworth GC (2008) Ainsworth & Bisby's dictionary of the fungi. CABI

Korpi A, Järnberg J, Pasanen A-L (2009) Microbial volatile organic compounds. Crit Rev Toxicol 39:139–193. doi:10.1080/10408440802291497

Kroumova AB, Xie Z, Wagner GJ (1994) A pathway for the biosynthesis of straight and branched, odd- and even-length, medium-chain fatty acids in plants. Proc Natl Acad Sci 91:11437–11441

Lancker FV, Adams A, Delmulle B, Saeger SD, Moretti A, Peteghem CV, Kimpe ND (2008) Use of headspace SPME-GC-MS for the analysis of the volatiles produced by indoor molds grown on different substrates. J Environ Monit 10:1127–1133. doi:10.1039/B808608G

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. doi:10.1038/nmeth.1923

Larsen TO, Frisvad JC (1995) Characterization of volatile metabolites from 47 *Penicillium* taxa. Mycol Res 99:1153–1166. doi:10.1016/S0953-7562(09)80271-2

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. Bioinforma Oxford Engl 25:2078–2079. doi:10.1093/bioinformatics/btp352

Liu W, Christenson SD, Standage S, Shen B (2002) Biosynthesis of the enediyne antitumor antibiotic C-1027. Science 297:1170–1173. doi:10.1126/science.1072110

Liu W, Nonaka K, Nie L, Zhang J, Christenson SD, Bae J, Van Lanen SG, Zazopoulos E, Farnet CM, Yang CF, Shen B (2005) The neocarzinostatin biosynthetic gene cluster from *Streptomyces carzinostaticus* ATCC 15944 involving two iterative type I polyketide synthases. Chem Biol 12:293–302. doi:10.1016/j.chembiol.2004.12.013

McElvain SM, Bright RD, Johnson PR (1941) The constituents of the volatile oil of catnip. I. Nepetalic acid, nepetalactone and related compounds. J Am Chem Soc 63:1558–1563. doi:10.1021/ja01851a019

Mendez-Perez D, Begemann MB, Pfleger BF (2011) Modular synthase-encoding gene involved in α-olefin biosynthesis in *Synechococcus* sp. strain PCC 7002. Appl Environ Microbiol 77:4264–4267. doi:10.1128/AEM. 00467-11

Mitchell AM, Strobel GA, Moore E, Robison R, Sears J (2010) Volatile antimicrobials from *Muscodor crispans*, a novel endophytic fungus. Microbiol Read Engl 156:270–277. doi:10.1099/mic. 0.032540-0

Müller DG, Jaenicke L, Donike M, Akintobi T (1971) Sex attractant in a brown alga: chemical structure. Science 171:815–817. doi:10.1126/science.171.3973.815

Peralta-Yahya PP, Keasling JD (2010) Advanced biofuel production in microbes. Biotechnol J 5:147–162. doi:10.1002/biot.200900220

Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD, Lee TS (2011) Identification and microbial production of a terpene-based advanced biofuel. Nat Commun 2:483. doi:10.1038/ncomms1494

Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD (2012) Microbial engineering for the production of advanced biofuels. Nature 488:320–328. doi:10.1038/nature11478

Pohnert G, Boland W (1994) Pericyclic reactions in nature: evidence for a spontaneous [1.7]-hydrogen shift and an 8πe electrocyclic ring closure in the biosynthesis of olefinic hydrocarbons from marine brown algae (*Phaeophyceae*). Tetrahedron 50:10235–10244. doi:10.1016/S0040-4020(01)81756-7

Polizzi V, Adams A, Malysheva SV, De Saeger S, Van Peteghem C, Moretti A, Picco AM, De Kimpe N (2012) Identification of volatile markers for indoor fungal growth and chemotaxonomic classification of *Aspergillus* species. Fungal Biol 116:941–953. doi:10.1016/j.funbio.2012.06.001

Pourbafrani M, Forgács G, Horváth IS, Niklasson C, Taherzadeh MJ (2010) Production of biofuels, limonene and pectin from citrus wastes. Bioresour Technol 101:4246–4250. doi:10.1016/j.biortech.2010.01.077

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD (2012) The PFAM protein families database. Nucleic Acids Res 40:D290–D301. doi:10.1093/nar/gkr1065

Qiu Y, Tittiger C, Wicker-Thomas C, Goff GL, Young S, Wajnberg E, Fricaux T, Taquet N, Blomquist GJ, Feyereisen R (2012) An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. Proc Natl Acad Sci 109:14858–14863. doi:10.1073/pnas.1208650109

Redding-Johanson AM, Batth TS, Chan R, Krupa R, Szmidt HL, Adams PD, Keasling JD, Soon Lee T, Mukhopadhyay A, Petzold CJ (2011) Targeted proteomics for metabolic pathway optimization: application to terpene production. Metab Eng 13:194–203. doi:10.1016/j.ymben.2010.12.005

Riyaz-Ul-Hassan S, Strobel G, Geary B, Sears J (2013) An endophytic *Nodulisporium* sp. from Central America producing volatile organic compounds with both biological and fuel potential. J Microbiol Biotechnol 23:29–35

Rude MA, Baron TS, Brubaker S, Alibhai M, Del Cardayre SB, Schirmer A (2011) Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from *Jeotgalicoccus* species. Appl Environ Microbiol 77:1718–1727. doi:10.1128/AEM. 02580-10

Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. Science 329:559–562. doi:10.1126/science.1187936

Schoch CL, Crous PW, Groenewald JZ, Boehm EWA, Burgess TI, De Gruyter J, De Hoog GS, Dixon LJ, Grube M, Gueidan C, Harada Y, Hatakeyama S, Hirayama K, Hosoya T, Huhndorf SM, Hyde KD, Jones EBG, Kohlmeyer J, Kruys A, Li YM, Lücking R, Lumbsch HT, Marvanová L, Mbatchou JS, McVay AH, Miller AN, Mugambi GK, Muggia L, Nelsen MP, Nelson P, Owensby CA, Phillips AJL, Phongpaichit S, Pointing SB, Pujade-Renaud V, Raja HA, Plata ER, Robbertse B, Ruibal C, Sakayaroj J, Sano T, Selbmann L, Shearer CA, Shirouzu T, Slippers B, Suetrong S, Tanaka K, Volkmann-Kohlmeyer B, Wingfield MJ, Wood AR, Woudenberg JHC, Yonezawa H, Zhang Y, Spatafora JW (2009) A class-wide phylogenetic assessment of *Dothideomycetes*. Stud Mycol 64, 1–15–S10. doi:10.3114/sim.2009.64.01

Schuchardt S, Kruse H (2009) Quantitative volatile metabolite profiling of common indoor fungi: relevancy for indoor air analysis. J Basic Microbiol 49:350–362. doi:10.1002/jobm.200800152

Shen B (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol 7:285–295

Spangler CW, Little DA (1982) Synthesis and characterization of representative octa-1,3,5,7-tetraenes and deca-1,3,5,7,9-pentaenes. J Chem Soc [Perkin] 1:2379–2385. doi:10.1039/P19820002379

Spangler CW, McCoy RK, Karavakis AA (1986) 3-Alkoxypropenals as precursors in the synthesis of conjugated and semiconjugated polyenes: methyl-substituted octa- and nona-tetraenes. J Chem Soc [Perkin] 1:1203–1207. doi:10.1039/P19860001203

Stoppacher N, Kluger B, Zeilinger S, Krska R, Schuhmacher R (2010) Identification and profiling of volatile metabolites of the biocontrol fungus *Trichoderma atroviride* by HS-SPME-GC-MS. J Microbiol Methods 81:187–193. doi:10.1016/j.mimet.2010.03.011

Strobel G, Daisy B (2003) Bioprospecting for microbial endophytes and their natural products. Microbiol Mol Biol Rev 67:491–502. doi:10.1128/MMBR. 67.4.491-502.2003

Strobel GA, Dirkse E, Sears J, Markworth C (2001) Volatile antimicrobials from *Muscodor albus*, a novel endophytic fungus. Microbiol Read Engl 147:2943–2950

Strobel GA, Knighton B, Kluck K, Ren Y, Livinghouse T, Griffin M, Spakowicz D, Sears J (2008) The production of myco-diesel hydrocarbons and their derivatives by the endophytic fungus *Gliocladium roseum* (NRRL 50072). Microbiology 154:3319–3328. doi:10.1099/mic. 0.2008/022186-0

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res 18:1979–1990. doi:10.1101/gr.081612.108

Tess Mends M, Yu E (2012) An endophytic *Nodulisporium* sp. producing volatile organic compounds having bioactivity and fuel potential. J Pet Environ Biotechnol. doi:10.4172/2157-7463.1000117

Tomsheck AR, Strobel GA, Booth E, Geary B, Spakowicz D, Knighton B, Floerchinger C, Sears J, Liarzi O, Ezra D (2010) *Hypoxylon* sp., an endophyte of *Persea indica*, producing 1,8-cineole and other bioactive volatiles with fuel potential. Microb Ecol 60:903–914. doi:10.1007/s00248-010-9759-6

Tressl R, Bahri D, Engel KH (1982) Formation of eight-carbon and ten-carbon components in mushrooms (*Agaricus campestris*). J Agric Food Chem 30:89–93

Ul-Hassan SR, Strobel GA, Booth E, Knighton B, Floerchinger C, Sears J (2012) Modulation of volatile organic compound formation in the Mycodiesel-producing endophyte *Hypoxylon* sp. CI-4. Microbiol Read Engl 158:465–473. doi:10.1099/mic. 0.054643-0

Van Lanen SG, Oh T, Liu W, Wendt-Pienkowski E, Shen B (2007) Characterization of the maduropeptin biosynthetic gene cluster from *Actinomadura madurae* ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis. J Am Chem Soc 129:13082–13094. doi:10.1021/ja073275o

Wheatley R, Hackett C, Bruce A, Kundzewicz A (1997) Effect of substrate composition on production of volatile organic compounds from *Trichoderma* spp. Inhibitory to wood decay fungi. Int Biodeterior Biodegrad 39:199–205. doi:10.1016/S0964-8305(97)00015-2

Wihlborg R, Pippitt D, Marsili R (2008) Headspace sorptive extraction and GC-TOFMS for the identification of volatile fungal metabolites. J Microbiol Methods 75:244–250. doi:10.1016/j.mimet.2008.06.011

Wurzenberger M, Grosch W (1984) The formation of 1-octen-3-ol from the 10-hydroperoxide isomer of linoleic acid by a hydroperoxide lyase in mushrooms (*Psalliota bispora*). Biochim Biophys Acta BBA Lipids Lipid Metab 794:25–30

Zhang J, Lanen SGV, Ju J, Liu W, Dorrestein PC, Li W, Kelleher NL, Shen B (2008) A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis. Proc Natl Acad Sci 105:1460–1465. doi:10.1073/pnas.0711625105

Zhang Y, Schoch CL, Fournier J, Crous PW, De Gruyter J, Woudenberg JH, Hirayama K, Tanaka K, Pointing SB, Spatafora JW, Hyde KD (2009) Multi-locus phylogeny of *Pleosporales*: a taxonomic, ecological and evolutionary re-evaluation. Stud Mycol 64, 85–102–S5. doi:10.3114/sim.2009.64.04