FUNGAL BIOLOGY XXX (2015) 1–16

British Mycological Society promoting fungal science

journal homepage: www.elsevier.com/locate/funbio

Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon

Michael MBENOUN[a], *Michael J. WINGFIELD*[a], *Teboho LETSOALO*[a], *Wubetu BIHON*[b,c], *Brenda D. WINGFIELD*[b], *Jolanda ROUX*[a,*]

[a]*Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute (FABI), Private Bag X20 Hatfield, University of Pretoria, Pretoria 0028, South Africa*
[b]*Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), Private Bag X20 Hatfield, University of Pretoria, Pretoria 0028, South Africa*
[c]*Agricultural Research Council-Vegetable and Ornamental Plant Institute (ARC-VOPI), Private Bag X293, Pretoria 0001, South Africa*

ARTICLE INFO

Article history:
Received 26 November 2014
Received in revised form
1 April 2015
Accepted 29 May 2015
Corresponding Editor:
Stephen W. Peterson

Keywords:
Anthropogenic dispersal
Cryptic species
Genealogical concordance
Host-associated differentiation
Species concepts

ABSTRACT

Thielaviopsis ethacetica was recently reinstated as a distinct taxon using DNA phylogenies. It is widespread affecting several crop plants of global economic importance. In this study, microsatellite markers were developed and used in conjunction with sequence data to investigate the genetic diversity and structure of *Th. ethacetica* in Cameroon. A collection of 71 isolates from cacao, oil palm, and pineapple, supplemented with nine isolates from other countries were analysed. Four genetic groups were identified. Two of these were associated with oil palm in Cameroon and showed high genetic diversity, suggesting that they might represent an indigenous population of the pathogen. In contrast, the remaining two groups, associated with cacao and pineapple, had low genetic diversity and, most likely, represent introduced populations. There was no evidence of gene flow between these groups. Phylogenetic analyses based on sequences of the *tef*1-α as well as the combined flanking regions of six microsatellite loci were consistent with population genetic analyses and suggested that *Th. ethacetica* is comprised of two divergent genetic lineages.

Introduction

Thielaviopsis ethacetica Went (Ascomycota, Microascales) was originally described from Java as the causal agent of sett rot (or pineapple disease) of sugarcane (Went 1893). It was subsequently reduced to synonymy with *Thielaviopsis paradoxa* (de Seynes) Höhn., based on morphological similarities of the two fungi (Höhnel 1904). The latter species had been discovered earlier in France associated with black rot of pineapple fruit (de Seynes 1886). Recently, the *Th. paradoxa* complex was reevaluated using the genealogical concordance phylogenetic species recognition (GCPSR) approach (Mbenoun *et al.*

* *Corresponding author.*
E-mail address: jolanda.roux@fabi.up.ac.za (J. Roux).
http://dx.doi.org/10.1016/j.funbio.2015.05.009
1878-6146/

2014). That study provided compelling evidence to support the reinstatement of *Th. ethacetica* as a distinct taxonomic entity, genetically well-differentiated from *Th. paradoxa sensu stricto* and all other related fungi.

Species represent the smallest independent entities of evolutionary lineages (de Queiroz 2007). They evolve through a temporally extended process, which is initiated by the split of an extant species into two reproductively isolated populations (de Queiroz 2007; Giraud *et al.* 2008). At different time points during the course of speciation, divergent populations will develop a number of specific biological and ecological properties pertaining to their unique evolutionary tendencies (de Queiroz 2007; Giraud *et al.* 2008). Some of these properties have been used as operational criteria for alternative species concepts (Sites & Marshall 2004; de Queiroz 2007). For instance, among those most commonly applied to fungi, the morphological species concept (MSC) emphasizes morphological discontinuity, the biological species concept (BSC) emphasizes reproductive incompatibility, the ecological species concept (ESC) emphasizes niche differentiation and the phylogenetic species concept (PSC) emphasizes nucleotide divergence. It is becoming increasingly apparent that the resolution of species boundaries in many fungal groups requires the application of more than one species concept, in an integrated approach (e.g., Cai *et al.* 2009; Samson & Varga 2009; Quaedvlieg *et al.* 2014).

GCPSR is an extension of PSC and relies on the concordance of phylogenies across multiple, unlinked loci to identify independently evolving lineages (Taylor *et al.* 2000). It has been particularly useful in the KINGDOM FUNGI (Voigt & Kirk 2011), where a limited number of robust morphological characters of taxonomic value has resulted in flawed estimations of fungal species numbers and distributions (Bass & Richards 2011). This technique has also been useful in defining supraspecific taxa, as illustrated in the recent restructuring of the Ceratocystidaceae, which has led to the consolidation of the genus *Thielaviopsis* (de Beer *et al.* 2014). However, GCPSR and more generally PSC have limitations with regards to discriminating recently diverged fungal species. This is especially true when applying commonly utilized gene loci in phylogenetic reconstructions, because of the relatively slow rates of temporal change in these genetic markers (Walker *et al.* 2012). This provides a rationale to explore other markers and/or analytical tools in order to unravel the cryptic boundaries between incipient and sister fungal species.

Population genetic analyses using rapidly evolving neutral markers, such as microsatellites (DNA simple sequence repeats), represent a powerful alternative or additional tool to GCPSR for the recognition of cryptic species (e.g., Fisher *et al.* 2002; Pérez *et al.* 2012). These methods are especially suitable when considering outcrossing species in situations where the investigated populations occur in sympatry. Under such conditions, only reproductively isolated lineages would be able to maintain their genetic integrity in the face of potential gene flow (Gladieux *et al.* 2011b), which is the underlying principle of BSC (Mayer 1942). In addition, population genetic studies can provide insights into the genetic structure and diversity, reproductive strategies, as well as ecological differentiation and adaptation of microorganisms. They can also reveal their origins and migration pathways. Such data provide important knowledge necessary for the development of sustainable management strategies for fungal and other pathogens of plants (McDonald & Linde 2002).

A major advance relating to our understanding of microbial biodiversity, and arising from DNA-based analyses, is the recognition that microorganisms are not naturally ubiquitous. In this regard, their genetic structure and distribution can be influenced by geographic isolation and environmental selection (Martiny *et al.* 2006). However, the globalization of trade and travel has resulted in the distribution of many microorganisms beyond their natural range. This has been the case for many fungal plant pathogens that have been introduced into new regions, along with their coevolved hosts or via the trade in plant products, and where some have been associated with the emergence of devastating new diseases in cultivated as well as natural plant systems (Anderson *et al.* 2004; Brasier 2008). In this regard, the increasing number of emerging diseases involving *Th. ethacetica* (reported as *Th. paradoxa* or *Ceratocystis paradoxa*) in different countries (e.g., Tzeng *et al.* 2010; Suwandi *et al.* 2012; Yu *et al.* 2012; Pinho *et al.* 2013) could be interpreted as a consequence of anthropogenic pathogen dispersal (Anderson *et al.* 2004).

Thielaviopsis ethacetica is a heterothallic fungus that requires two compatible mating partners (bearing opposing mating types, MAT 1-1 and MAT 1-2) to produce sexual spores (ascospores), but propagates largely through asexual spores (conidia). It is the most abundant species in *Thielaviopsis*, being a broad generalist pathogen that has been recorded in several countries across five continents (Mbenoun *et al.* 2014). Although there is no information regarding its centre of origin, it could be hypothesized that this pathogen initially had a more restricted distribution before being disseminated and expanding its host range via anthropogenic activities. This view would be supported by its occurrence on very different and widely-grown plant hosts such as cacao (*Theobroma cacao* L.), coconut palm (*Cocos nuciphera* L.), date palm (*Phoenix dactylifera* L.), oil palm (*Elaeis guineensis* Jacq.), pineapple (*Ananas comosus* L.) and sugarcane (*Saccharum officinarum* L.) (Mbenoun *et al.* 2014). Most of these crops occur in Cameroon, especially cacao, oil palm, and pineapple, which are intensively and extensively cultivated in the central and south-western regions of the country. Of these, oil palm is indigenous (Obahiagbon 2012), while the two other crops were introduced from South America (Morton 1987; Motamayor *et al.* 2002). All of these crops are planted in close proximity to each other and are often found intercropped in family gardens and small-holder farms.

In their reevaluation of *Thielaviopsis* species, Mbenoun *et al.* (2014) reported that both the MAT 1-1 and MAT 1-2 mating types of *Th. ethacetica* occur in Cameroon. This would have resulted in sexual recombination under natural conditions and thus a genetically diverse population of *Th. ethacetica* in the country. These authors also noted considerable sequence variation in the translation elongation factor 1 alpha (*tef*1-α) gene for *Th. ethacetica* isolates from different hosts. In the present study, the genetic diversity of a collection of *Th. ethacetica* isolates from cacao, oil palm, and pineapple in Cameroon, supplemented with isolates from various hosts and other regions globally, was analysed using molecular population genetic and phylogenetic inference. The aims were, (i) to develop polymorphic microsatellite markers for *Th. ethacetica* and to use them in a population study in order to, (ii)

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

investigate the influence of hosts on spatial population structure in Cameroon, (iii) to infer the history of the pathogen, and (iv) to assess whether the polymorphisms previously observed in the *tef*1-α gene of *Th. ethacetica* correlates with the genetic structure of the species at the population level.

Materials and methods

Fungal individuals and populations

A total of 80 isolates of *Thielaviopsis ethacetica* were used in this study (Table 1). These included nine isolates from various hosts and origins globally, which were sourced from international fungal culture collections, and 71 isolates from Cameroon. The Cameroonian isolates were collected over a two-year period, from 2009 to 2010, and are preserved in the fungal culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa (Mbenoun *et al.* 2014). The majority (69 isolates) of these isolates were from cacao (on discarded pod husks, except for two isolates obtained from a rotting pod and a wounded stem respectively), oil palm (on stumps and trunks of felled trees, cut basal ends of leaf fronds, and rotting palm-nut bunches), and pineapple (on damaged leaves, fruit peduncles, and rotting fruits), in seven localities, distributed across three regions of the country (Fig 1). Two isolates were collected from a wounded, naturally occurring *Erytrophleum ivorense* A. Chev. tree next to a cacao field.

Detailed descriptions of protocols and procedures used for field sampling and fungal isolation, as well as DNA extraction and molecular identification are outlined in Mbenoun *et al.* (2014). Plant materials were incubated under moist conditions to induce the development of fungal structures. *Thielaviopsis* spp. were mostly isolated as mycelial strands lifted from the surfaces of plant tissues. A few isolates were collected as single ascospore drops at the tips of ascomata on cacao pod husks. Isolates were purified by subculturing from single hyphal tips on 2 % Malt Extract Agar plates, and they were identified based on multilocus DNA phylogenies, using sequence data from the ribosomal internal transcribed spacer (ITS) locus, including the 5.8S rRNA region, the beta tubulin (*β-tub*) gene and the *tef*1-α gene regions.

All isolates collected from the same host were treated as belonging to the same population. Thus, three populations were defined, respectively associated with cacao, oil palm, and pineapple (Fig 1). These populations were subdivided into local subpopulations according to sampling locations. Three subpopulations were sampled on cacao, including CO-Bokito and CO-Ngomedzap in the Centre region, and CO-Kumba in the South-West region. The CO-Bokito and CO-Ngomedzap subpopulations each included isolates from two fields, while isolates representing CO-Kumba were from one field. Four subpopulations were sampled on oil palm. Three of these, OP-

Table 1 – Isolates of *Thielaviopsis ethacetica* included in this study, with information of their origin, host, multilocus genotype (MLG), and their genetic grouping based on population analyses.

Isolate[a]	MLG[b]	Genetic group	Host	Sampling field	Site	Country
ATCC28389	FXGIHHB	PINEAPPLE	Sugarcane	NA	Louisiana	USA
CBS 128.32	BFEBMBA	PALM 1	Oil palm	NA	NA	NA
CBS 374.83	BCKBBDA	PALM 1	Date palm	NA	Maspalomas	Spain
CBS 601.70	JGDAJIH	PINEAPPLE	Pineapple	NA	NA	Brazil
CBS 453.66	CCEBABA	PALM 1	Coconut palm	NA	NA	Nigeria
CMW35018	IGAGEFE	CACAO	Cacao	1	Bokito	Cameroon
CMW35019	IGAGEFE	CACAO	Cacao	1	Bokito	Cameroon
CMW35020	IGAGEFE	CACAO	Cacao	1	Bokito	Cameroon
CMW35025	IGAGEFE	CACAO	Cacao	1	Ngomedzap	Cameroon
CMW35026	IGACEFE	CACAO	Cacao	1	Ngomedzap	Cameroon
CMW35027	IGAGEFE	CACAO	Cacao	1	Ngomedzap	Cameroon
CMW35029	DBDBCBA	PALM 1	*E. ivorense*	3	Ngomedzap	Cameroon
CMW35030	DBDBCBA	PALM 1	*E. ivorense*	3	Ngomedzap	Cameroon
CMW36639	CCEBBCD	PALM 1	Oil palm	2	Dibamba	Cameroon
CMW36644	BDIEPDF	PALM 2	Oil palm	2	Dibamba	Cameroon
CMW36645	BAQBGBA	PALM 1	Oil palm	2	Dibamba	Cameroon
CMW36646	EHPBICA	PALM 1	Oil palm	2	Dibamba	Cameroon
CMW36648	BAFBCBA	PALM 1	Oil palm	2	Dibamba	Cameroon
CMW36656	BDGEPCF	PALM 2	Oil palm	1	Dibamba	Cameroon
CMW36659	CCEBBCD	PALM 1	Oil palm	2	Dibamba	Cameroon
CMW36660	BDIEPDF	PALM 2	Oil palm	2	Dibamba	Cameroon
CMW36661	BDIEPDF	PALM 2	Oil palm	2	Dibamba	Cameroon
CMW36662	BEJDQCA	PALM 2	Oil palm	1	Dibamba	Cameroon
CMW36663	ECEBBBA	PALM 1	Oil palm	1	Tiko	Cameroon
CMW36664	BCEBCBA	PALM 1	Oil palm	1	Tiko	Cameroon
CMW36667	ECNBKDA	PALM 1	Oil palm	1	Tiko	Cameroon
CMW36669	BEDBBBA	PALM 1	Oil palm	1	Tiko	Cameroon
CMW36670	BCOBICA	PALM 1	Oil palm	1	Tiko	Cameroon

(*continued on next page*)

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

Table 1 – (*continued*)

Isolate[a]	MLG[b]	Genetic group	Host	Sampling field	Site	Country
CMW36671	BCMBDBA	PALM 1	Oil palm	1	Tiko	Cameroon
CMW36672	BCXXABA	PALM 1	Oil palm	1	Tiko	Cameroon
CMW36687	BDEBBBA	PALM 1	Oil palm	1	Kumba	Cameroon
CMW36688	BFGEOEF	PALM 2	Oil palm	1	Kumba	Cameroon
CMW36691	IGAGJFE	CACAO	Cacao	2	Kumba	Cameroon
CMW36692	IGAAEFE	CACAO	Cacao	2	Kumba	Cameroon
CMW36693	IGACEFE	CACAO	Cacao	2	Kumba	Cameroon
CMW36695	IGAGEFE	CACAO	Cacao	2	Kumba	Cameroon
CMW36698	IGAGEGE	CACAO	Cacao	2	Kumba	Cameroon
CMW36699	IGAGEFE	CACAO	Cacao	2	Kumba	Cameroon
CMW36701	IGACEFE	CACAO	Cacao	2	Kumba	Cameroon
CMW36704	IGAGEFE	CACAO	Cacao	2	Kumba	Cameroon
CMW36707	IGAGEFE	CACAO	Cacao	2	Kumba	Cameroon
CMW36714	IGAGEFE	CACAO	Cacao	2	Kumba	Cameroon
CMW36718	IGAGEFE	CACAO	Cacao	2	Kumba	Cameroon
CMW36720	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36721	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36722	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36723	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36724	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36725	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36727	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36728	EBCBDBA	PALM 1	Pineapple	2	Njombe	Cameroon
CMW36729	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36731	BCPBABA	PALM 1	Pineapple	2	Njombe	Cameroon
CMW36733	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36735	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36736	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36737	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36738	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36739	HJBAJHG	PINEAPPLE	Pineapple	1	Njombe	Cameroon
CMW36740	BCPBABA	PALM 1	Pineapple	2	Njombe	Cameroon
CMW36741	IGAHEFE	CACAO	Cacao	1	Ngomedzap	Cameroon
CMW36745	IGAHEFE	CACAO	Cacao	1	Ngomedzap	Cameroon
CMW36746	AARBGAA	PALM 1	Oil palm	1	Ekona	Cameroon
CMW36747	CCPBCBA	PALM 1	Oil palm	1	Ekona	Cameroon
CMW36748	CCPBCBA	PALM 1	Oil palm	1	Ekona	Cameroon
CMW36749	BCLBABA	PALM 1	Oil palm	1	Ekona	Cameroon
CMW36750	CCGBABA	PALM 1	Oil palm	1	Ekona	Cameroon
CMW36753	BCLBNBA	PALM 1	Oil palm	1	Ekona	Cameroon
CMW36756	BCLBNBA	PALM 1	Oil palm	1	Ekona	Cameroon
CMW36759	BEMBBXC	PALM 1	Oil palm	1	Ekona	Cameroon
CMW36760	BCDBBBA	PALM 1	Oil palm	1	Ekona	Cameroon
CMW36771	EJHBHBA	PALM 1	Sugarcane	NA	NA	South Africa
CMW36817	IGAGEFE	CACAO	Cacao	2	Bokito	Cameroon
CMW37097	IGAGEFE	CACAO	Cacao	2	Ngomedzap	Cameroon
CMW37098	IGAGEFE	CACAO	Cacao	2	Ngomedzap	Cameroon
CMW37099	IGAGEFE	CACAO	Cacao	2	Ngomedzap	Cameroon
CMW37100	IGAGEFE	CACAO	Cacao	2	Ngomedzap	Cameroon
IMI50560	BIBAJHG	PINEAPPLE	Pineapple	NA	NA	Malaysia
IMI3344082	GKDAHEI	PINEAPPLE	Coconut palm	NA	NA	Tanzania
IMI378943	GXDFFBA	PINEAPPLE	Oil palm	NA	NA	Papua New Guinea

Where sequence data were ambiguous, alleles were scored as missing (X).
NA: Data not available.
CBS: Centraalbureau voor Schimmelcultures.
CMW: Culture collection of the Forestry and Agricultural Biotechnology Institute, University of Pretoria.
IMI: International Mycological Institute, now Commonwealth Agricultural Bureaux International Bioscience (CABI).
a Isolates highlighted in bold were included in phylogenetic analyses.
b Only loci included in the analyses.

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

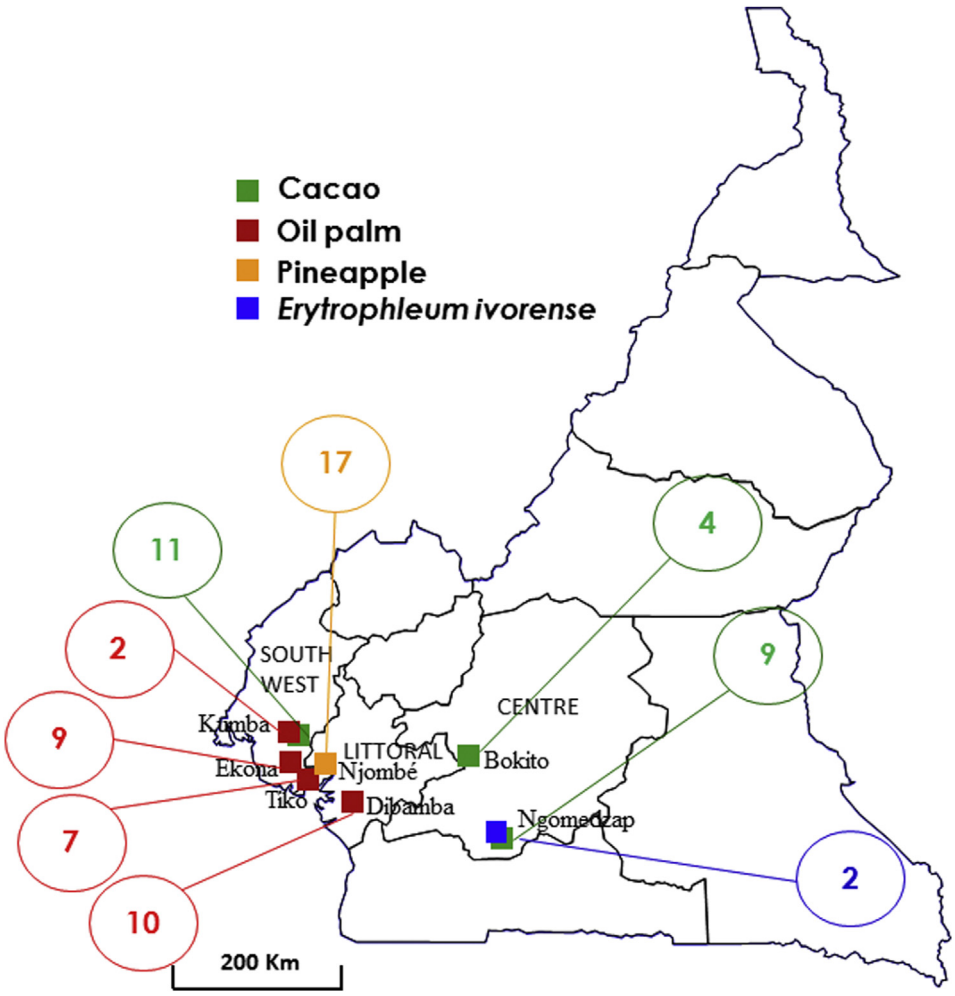

Fig 1 – Map of Cameroon and distribution of host-associated populations of *Thielaviopsis ethacetica* (represented by different colours) used in this study. Circles indicate the number of isolates sampled in each local subpopulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Kumba, OP-Ekona, and OP-Tiko, were each from a single field in the South-West region, while the fourth, OP-Dibamba, was from the Littoral region and included isolates from one field and others from trees occurring in natural vegetation sites. Only one subpopulation, PA-Njombé, was sampled on pineapple and included isolates from two fields. The two isolates obtained from *Erythrophleum ivorense* in the Centre region were also considered to represent a distinct subpopulation, ET-Ngomedzap.

Multilocus genotyping of isolates

Development of microsatellite markers

A set of ten microsatellite markers (Table 2) was developed for population biology studies of *Thielaviopsis ethacetica* following the procedure described by Santana *et al.* (2009). Pooled total genomic DNA of two *Th. ethacetica* isolates (CMW36661 and CMW36722) from oil palm and pineapple in Cameroon respectively, provided the template for microsatellite enrichment, using intersimple sequence repeat PCR protocols (ISSR-PCR) and previously reported primers (Santana *et al.* 2009). A microsatellite-enriched genomic DNA library of *Th. ethacetica* was prepared, by mixing standardized aliquots of all ISSR-PCR purified products, and analysed on a 454 GS-FLX (Roche Applied Science, Penzburg, Germany) platform, at Inqaba Biotech (Pretoria, Gauteng, South Africa). Sequence reads were assembled using the ContigExpress tool of Vector NTI software (Invitrogen, Carlsbad, CA, USA). All contigs (and singleton reads) of $\geq$150 bp were screened for simple sequence repeats (SSRs) online, using the GRAMENE SSRIT tool (Temnykh *et al.* 2001; http://archive.gramene.org/db/markers/ssrtool). Primers were designed for all identified SSR loci with suitable flanking regions using the programme Primer3 (Untergrasser *et al.* 2012; http://primer3.ut.ee/).

Microsatellite genotyping

The newly identified microsatellite markers were initially tested for polymorphisms on 12 *Thielaviopsis ethacetica* isolates from different hosts and locations in Cameroon. This was achieved by deriving and analyzing allele numbers for each locus from aligned full sequence data of all 12 selected isolates.

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

Table 2 – Characteristics of microsatellite markers developed for population studies of *Thielaviopsis ethacetica*, including sequence variation in the flanking regions of simple sequence repeats. The size range is for the genomic region amplified by the two primers.

Locus	GenBank accession	Primer sequence (5′–3′)	Repeat motif	T_a (°C)	Size range (bp)	# Allele	# INDEL	# SNP	# PIC
CESSR06	KM853151	F: CACCACCAGCAATTAGACGA R: GGACGAGCGGGACTATTATACA	$(AGC)_{4-5}+(TAC)_{7-24}$	56	106–161	10	1	5	1
CESSR10	KM853152	F: AGTGATGCTGGAAGGGATTG R: TGGAACCCGAGACTTATGGT	$(TAG)_{3-28}$	52	88–160	12	2	0	0
CESSR17	KM853153	F: AAAGAGGCCAAGGGAAGG R: CATTGCTCACTTCTGTTTCC	$(CA)_{8-30}$	54	138–182	19	0	1	1
CESSR19	KM853154	F: GTTAGCACCAGAACAGAGCTA R: ACCTGGTCAATACCAAACAGGA	$(TC)_{10-20}$	52–56	ND	10	0	2	1
CESSR20	KM853155	F: GAATGTCGTGATGGGTTGG R: CGTTTCTGCCTGTAGTATCT	$(TAC)_{12-42}$	52–56	168–258	16	1	7	4
CESSR21	KM853156	F: CGTTGCCATCACAAATGAAG R: TCTCTCCGTGTCAATGTCTTG	$(GTT)_{3-10}+(AGT)_{10-19}$	56	188–214	10	1	6	6
CESSR24[a]	KM853157	F: GCCCGTTTCTGCCTGTAGTA R: AGTCGGCGTTAACAAAGTGC	$(TAG)_{6-40}$	56	102–192	17	2	1	1
CESSR29[a,b]	KM853158	F: GGGTAAGGGTGAAGGGTAT R: ACCACACGACAAGACACGAC	$(CTA)_{11-15}$	56	96–108	3	0	3	3
CESSR38[a,b]	KM853159	F: TGGAGTGCAGGCTGTAGTTG R: CCTCATACTCAAGGCCAAGC	$(TGC)_{6-8}+(TGT)_{4-6}$	56	139–145	4	0	2	2
CESSR39	KM853160	F: CCTCATACTCAAGGCCAAGC R: TGGAGTGCAGGCTGTAGTTG	$(CAA)_{3-6}(CAG)_{7-8}(CAA)_{3-7}$	56	122–128	7	0	4	4

INDEL: insertion/deletion mutation.
SNP: single nucleotide polymorphism.
PIC: parsimony-informative character.
ND: not determined.
T_a: annealing temperature.
a Based on 12 isolates as opposed to 80 isolates for the other markers.
b Markers not used in this study.

Additionally, 68 isolates were sequenced for the eight most polymorphic microsatellite markers and genotyped following the same procedure. PCR reactions were prepared in 25 μL total volumes, including 0.5 μL (2.5 U) MyTaq™ DNA polymerase (Bioline), 5 μL 5× MyTaq reaction buffer (supplied with the enzyme), 0.5 μL of each primer (10 mM), 1 μL DNA template (75 ηg μL^{-1}), and 17.5 μL sterile distilled water (SABAX water, Adcock Ingram, Bryanston). Reactions were carried out using an iCycler thermocycler (BIO-RAD, Hercules, CA, USA). The thermal cycling conditions comprised an initial denaturation step at 96 °C for 2 min followed by 35 cycles of 30 s at 94 °C, 60 s at 52–56 °C (depending on the locus) and 90 s at 72 °C, and a final extension step at 72 °C for 10 min. To check for successful amplification, PCR products were stained with GelRed™ (Biotium, USA) nucleic acid dye, electrophoresed on 2 % agarose gels along with a 100 bp molecular weight marker (Fermentas O′ Gene Ruler™) and, thereafter, visualized under UV illumination. Amplified PCR products were purified by gel filtration through 6 % G-50 sephadex, in sephadex columns (Sigma, Steinheim, Germany) following the manufacturer's instructions.

Forward and reverse sequencing reactions were set up in 96-well MicroAmp® plates (Applied BioSystems, Foster City, CA). Each reaction mix contained 2.5 μL sequencing buffer, 0.5 μL Big Dye ready reaction mixture with Amplitaq DNA polymerase (Perkin-Emmer, Warrington, UK), 1 μL of specific forward or reverse primer (10 mM), 3 μL purified PCR product and sterile Sabax® water in 12 μL final volumes. The thermal cycling conditions comprised 25 cycles of 10 s at 96 °C, 5 s at 50 °C and 4 min at 60 °C. Sequencing products were purified by sodium acetate precipitation and, thereafter, run on an ABI PRISM™ 3100 DNA Analyzer (Applied BioSystems, Foster City, CA).

Bidirectional sequencing reads were edited and assembled into contigs using MEGA version 5 (Tamura *et al.* 2011). Multiple sequence alignments were constructed using MUSCLE (Edgar 2004), as implemented in MEGA, and edited manually. Alignments were trimmed at their ends so as to include the entire DNA region circumscribed by the two primers. Alleles were determined by combining the number of SSRs and information pertaining to the presence of single point or insertion/deletion (INDEL) mutations in the flanking regions of SSRs. Individual alleles were assigned a distinct haplotype (represented by an upper case alpha character), thus, defining a data matrix for analysis including eight multistate characters by which each isolate could be identified by a single multilocus genotype (MLG), e.g., IGAGEFEE for isolate CMW35018 (Table 1). One microsatellite marker (CESSR24) was ultimately excluded from the analyses because it was found to be linked to another locus (CESSR20), while two markers (CESSR29 and CESSR38), developed after population analyses had been completed, were not used in the study.

Genetic diversity

Genetic diversity was analysed for all host-associated populations and locally sampled subpopulations. The total number

of polymorphic loci, alleles as well as private alleles was determined using the software POPGENE version 1.31 (Yeh *et al.* 1999). Nei's (1973) gene diversity (*H*) was also estimated using POPGENE, with and without clone-correction (removal of replicated MLGs) of the data. Genotypic diversity (*G*) was estimated according to Stoddart & Taylor (1988). Because *G* is limited by the sample size (*N*), and in order to enable comparison among populations with different sample sizes, normalized percentage of genotypic diversity (G^*) was calculated (Chen *et al.* 1994). Furthermore, genotypic diversity was scaled by the expected number of MLGs ($E(g_n)$) for the smallest sample size being compared (Grünwald *et al.* 2003). $E(g_n)$ was estimated using rarefaction curves with R statistical software (R Development Core Team 2011; http://www.R-project.org/), based on a sample size of 17 isolates at the population level and two isolates at the subpopulation level.

Population differentiation

Genetic differentiation analyses were undertaken for *Thielaviopsis ethacetica* in Cameroon based on the entire as well as the clone-corrected data sets. The proportion of genetic variability explained by host plants was estimated from the partition of total genetic variance into two components (among populations and within populations). This was achieved using analyses of molecular variance (AMOVA) with GenAlEx 6.5 (Peakall & Smouse 2012). This programme uses an ANOVA framework based on an analogue (Φ) of the index of fixation, F_{ST} (Excoffier *et al.* 1992). The significance of each variance component was tested using 999 random permutation replicates. Φ-Statistics were also used to assess differentiation between pairs of populations.

Genotype grouping and genetic composition of host-associated populations

In order to identify putative genetic groups (or gene pools) within *Thielaviopsis ethacetica*, the clone-corrected data set for the Cameroonian isolates was combined with that of isolates from other countries (international collection). This combined data set was analysed using both Bayesian Clustering (BC) and Principal Component Analysis (PCoA). BC analyses were performed to provide a first estimate of the number of genetic clusters (*K*) represented in the data set. This was achieved using Markov Chain Monte Carlo (MCMC) algorithms implemented in the program STRUCTURE v2.3.4 (Pritchard *et al.* 2000). Ten simulations were performed, respectively assuming values for *K* between one and ten. Each simulation included 100 000 MCMC runs after a burn-in period of 10 000 runs and was reiterated ten times. Prior parameters were set to a nonadmixture model of ancestry and independent allele frequencies. The most likely value for *K* was inferred using the *L(K)* (Pritchard *et al.* 2000) and *ΔK* (Evanno *et al.* 2005) maximum likelihood-based methods implemented in STRUCTURE HARVESTER web v0.6.93 (Earl & vonHoldt 2012; http://taylor0.biology.ucla.edu/structureHarvester/). For the selected values of *K*, results of 200 000 MCMC runs (burn-in, 100 000) reiterated 5 times were collated in CLUMPP v1.1.2 (Jakobsson & Rosenberg 2007), and graphically visualized in DISTRUCT v1.1 (Rosenberg 2004). A 90 % identity cut-off was used to assign individual MLGs to clusters. MLGs having a membership proportion of less than 90 % in a single cluster were considered admixed.

The number of genetic groups was optimized by superimposing the clusters identified with STRUCTURE and those delineated by the first three principal components of the PCoA. The latter was performed with GenAlEx 6.5, based on the standardized covariance method and haploid distance matrix. Once the genetic groups had been identified, gene and genotypic diversities were estimated for each genetic group in Cameroun and the structure of host-associated populations was determined based on the frequencies of the different genetic groups.

Phylogenetic analyses

Sequencing with coding gene loci

All isolates included in this study were shown to be identical at both the ITS and *β-tub* loci (Mbenoun *et al.* 2014). Polymorphism was, however, found in the *tef*1-*α* gene for a subset of isolates used in multilocus phylogenetic reconstructions, delineating three phylogenetic clusters (Mbenoun *et al.* 2014). Therefore, additional genotypic information was generated in this study by sequencing a portion of the *tef*1-*α* gene for 40 isolates, representing unique MLGs. Sequences were generated with the oligonucleotide primer pair EF1F (5′-TGCGGTGGTATCGACAAGCGT-3′)/EF2R (5′-AGCATGTTGTCGCCGTTGAAG-3′) (Jacobs *et al.* 2004). The PCR cycle included 4 min initial denaturation at 96 °C, ten cycles of 40 s at 94 °C, 40 s at 55 °C (annealing step), and 45 s at 72 °C (extension), followed by 30 additional cycles of the same sequence, with a 5 s increase in the annealing step per cycle. Reactions were completed by a final extension step at 72 °C for 10 min. All other PCR and sequencing procedures were the same as those used for SSR markers.

Phylogenetic reconstructions

Phylogenetic reconstructions were conducted based on microsatellite and *tef*1-*α* loci, using 40 isolates, representing the same clone-corrected population that was used for genetic clustering, minus two Cameroonian MLGs (CMW36672 and CMW36759) for which sequence data at two microsatellite loci were ambiguous. These analyses were performed using maximum parsimony (MP) phylogenetic inferences with PAUP version 4.0b10* (Swofford 2002). One microsatellite locus (CESSR10) that did not include any parsimony-informative characters was excluded from the analyses. The remaining six microsatellite loci were considered, separately and in combination. Gaps were treated as missing data in the microsatellite analyses to minimize the effect of differences in SSR repeats among isolates. For the *tef*1-*α* gene, gaps were considered as a fifth character state. Parsimony-uninformative characters were excluded and all informative characters were unordered and had the same weight. Phylogenetic trees were generated with a full heuristic tree search of 1000 random stepwise addition replicates and tree-bisection-reconstruction branch-swapping. All equally most parsimonious trees (MPTs) were saved and Tree Length (TL), Consistency index (CI), Retention Index (RI), and Rescaled retention index (RC) were estimated. Statistical support values for branch

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

nodes of MPTs were assessed by performing 1000 bootstrap analyses replicates.

To investigate possible recombination events between genetic groups and network analyses were undertaken based on a concatenation of sequence alignment data sets including the *tef*1-α and the six microsatellite loci used in tree reconstructions. These analyses were performed using the algorithms implemented in the SplitsTree program version 4.13.1 (Huson & Bryant 2006), including a general time reversible (GTR) character transformation, NeighborNet distance transformation and EqualAngle split transformation methods. Statistical support was estimated for each split based on 1000 bootstrap replicates.

Random mating

In order to assess whether random mating occurs between individuals related to different genetic groups of *Thielaviopsis ethacetica*, linkage disequilibrium among loci was estimated for increasingly complex schemes of Cameroonian populations, including (i) the most abundant genetic cluster on oil palm, (ii) the entire oil palm-associated population, as well as (iii) the entire Cameroonian population. This was achieved by computing the index of association (I_A) (Brown *et al.* 1980) for each of these populations. The observed values of I_A were compared with the distribution of this index from 1000 hypothetical data sets generated under the null hypothesis of random association of alleles. When the observed values fell outside the 95 % confidence range (mean ± 1.96 sd), the null hypothesis that alleles are randomly associated (i.e., all genotypes within the investigated population were freely interbreeding) was rejected. Analyses were performed with the programme Multilocus 1.3 (Agapow & Burt 2000), using clone-corrected data sets.

Results

Genetic diversity

All ten microsatellite markers developed in this study were polymorphic and could be amplified for all *Thielaviopsis ethacetica* isolates tested. The seven markers used for analyses produced a total of 84 alleles from the 80 isolates. Of these, 69 alleles were recovered from the 71 Cameroonian isolates. The largest number of alleles recovered per subpopulation sampled was 10 (average: 8.66) on cacao (CO-Kumba), 30 (average: 22.75) on oil palm (OP-Dibamba), while the only subpopulation sampled on pineapple (PA-Njombé) included 18 alleles (Table 3). Oil palm-associated subpopulations included from three to seven private alleles while no cacao-associated subpopulation included more than two private alleles (Table 3). When comparing host-associated populations across all sampling locations, the number of alleles recovered was 12 for cacao, 18 for pineapple, and 49 for oil palm. There were seven private alleles associated with pineapple, 10 with cacao and 40 with oil palm. Nine alleles were shared between oil palm and pineapple, two alleles between pineapple and cacao, and no allele between cacao and oil palm. All seven loci were polymorphic for both the oil palm- and pineapple-associated populations, while only three loci were polymorphic for the cacao-associated population.

The level of gene and genotypic diversity varied considerably among subpopulations (Table 3). The two indices were the lowest for the Bokito cacao-associated subpopulation [$H = 0.000$; $E(g_n) = 1$], for which all isolates represented the same MLG. In contrast, the maximum genotypic diversity [$E(g_n) = 2$] was measured for OP-Tiko and OP-Kumba, two subpopulations associated with oil palm and for which each

Table 3 – Gene and genotypic diversity estimated for host-associated populations and local subpopulations of *Thielaviopsis ethacetica* in Cameroon.

Subpopulation/population	Location	Host	# Isolates	Genetic diversity								
				# PLC	# AL	# PAL	H	H_{corr}	# MLG	G	G^*	$E(g_n)$
CO-Bokito	Bokito	Cacao	4	0	7	0	0.000	0.000	1	1.00	25.00	1.00
CO-Kumba	Kumba	Cacao	11	3	10	2	0.130	0.206	5	2.81	21.64	1.71
CO-Ngomedzap	Ngomedzap	Cacao	9	2	9	1	0.071	0.095	3	1.98	17.96	1.56
OP-Dibamba	Dibamba	Oil palm	10	7	30	9	0.660	0.659	7	5.88	58.82	1.91
OP-Ekona	Ekona	Oil palm	9	6	26	9	0.444	0.478	7	6.23	69.23	1.94
OP-Kumba	Kumba	Oil palm	2	6	13	3	0.429	0.429	2	2.00	100	2.00
OP-Tiko	Tiko	Oil palm	7	6	22	5	0.426	0.426	7	7.00	100	2.00
PA-Njombé	Njombé	Pineapple	17	7	18	7	0.299	0.571	3	1.44	8.45	1.32
ET-Ngomedzap	Ngomedzap	Tali	2	0	7	0	0.000	0.000	1	1.00	50.00	1.00
Cacao	–	–	24	3	12	10	0.093	0.206	6	2.11	8.82	5.03
Oil palm	–	–	28	7	46	37	0.633	0.614	23	19.60	70.00	15.05
Pineapple	–	–	17	7	18	7	0.299	0.571	3	1.44	8.45	3.00

CO ≡ Cacao, OP ≡ Oil palm, PA ≡ Pineapple, ET ≡ *Erythrophleum ivorense*.
PLC: Polymorphic loci, AL: Alleles, PAL: Private alleles.
H: Gene diversity (Nei 1973), H_{corr}: Gene diversity with clone-corrected data sets.
MLG: multilocus genotypes, G: Genotypic diversity (Stoddart & Taylor 1988), G^*: percentage of maximum genotypic diversity (Chen *et al.* 1994).
$E(g_n)$: scaled genetic diversity with rarefaction (Grünwald *et al.* 2003) = Expected number of genotypes for the smallest population size being compared (17 isolates at the population level and two isolates at the subpopulation level).

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

Table 4 – Pair wise differentiation (Φ_{PT}) among host associated populations of *Thielaviopsis ethacetica* in Cameroon based on entire (values above diagonal) and clone-corrected (values below diagonal) data sets.

	Cacao	Oil palm	Pineapple
Cacao	–	0.637***	0.781***
Oil palm	0.520***	–	0.0479***
Pineapple	0.541*	0.028^{ns}	–

Stars indicate the level of statistical significance (*$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, and ^{ns}nonsignificant).

isolate represented a unique MLG. The highest level of gene diversity was estimated in an oil palm-associated subpopulation, OP-Dibamba ($H = 0.660$; $H_{corr} = 0.659$), which included seven MLGs from ten isolates (Table 2). Overall, gene and genotypic diversity indices were substantially greater for oil palm [$H = 0.633$; $H_{corr} = 0.614$; $E(g_n) = 15$] as compared to cacao [$H = 0.093$; $H_{corr} = 0.206$; $E(g_n) = 5$] and pineapple [$H = 0.299$; $H_{corr} = 0.571$; $E(g_n) = 3$].

Population differentiation

Results of AMOVA showed that a substantial and highly significant proportion of the genetic variance of *Thielaviopsis ethacetica* in Cameroon was due to variation between host-associated populations (i.e., between hosts), regardless of whether the entire (62 %; $\Phi_{CT} = 0.625$, $p < 0.001$) or clone-corrected (42 %; $\Phi_{CT} = 0.417$; $p < 0.001$) data sets were used. In pair-wise comparisons, all host-associated populations emerged as well-differentiated from one another ($\Phi_{PT} > 0.500$; $p < 0.05$), except for the pair pineapple/oil palm, especially when using clone-corrected data sets and, to a lesser extent, when using the entire data sets (Table 4).

Genotype grouping and genetic composition of host-associated populations

The most likely number of genetic clusters inferred from Bayesian analyses using the ΔK method was $K = 2$, while with the $L(K)$ method $K = 3$ clusters were delineated (Supplementary material — Fig S1). Assuming $K = 2$, most of the MLGs were assigned to either of the two clusters (Green and Red) with maximum identity, and only a few admixed individuals could be identified (Fig 2). However, cluster assignment and estimation of admixture proportions were not always consistent with the allelic composition of the MLGs. For example, the dominant MLG from pineapple in Cameroon and two international MLGs from the same host shared one to two alleles and were clustered with cacao MLGs in the Green cluster. On the other hand, four admixed MLGs from oil palm, sharing up to three alleles with other oil palm MLGs assigned to the Red cluster with maximum identity, were estimated to have a greater proportion of ancestry from the Green cluster, although they shared no alleles with cacao or pineapple MLGs. The output of STRUCTURE analyses assuming $K = 3$ (Fig 2) suffered less from these inconsistencies by delineating an additional cluster (Blue) to accommodate three of four admixed MLGs from oil palm. It also allowed for the identification of a greater number of admixed genotypes, including the three MLGs from pineapple mentioned above. Most of this admixture involved the Green and Blue clusters.

PCoA analysis supported the results from BC assuming $K = 3$. Three clusters were identified when MLGs were plotted against the first and second principal coordinates (Fig 3A). The one cluster included all and only MLGs recovered on cacao, corresponding to the Green cluster delineated by STRUCTURE assuming $K = 3$. This cluster is designated here as the CACAO genetic group. The second cluster, designated here as PAML 1 genetic group, included all MLGs assigned with maximum

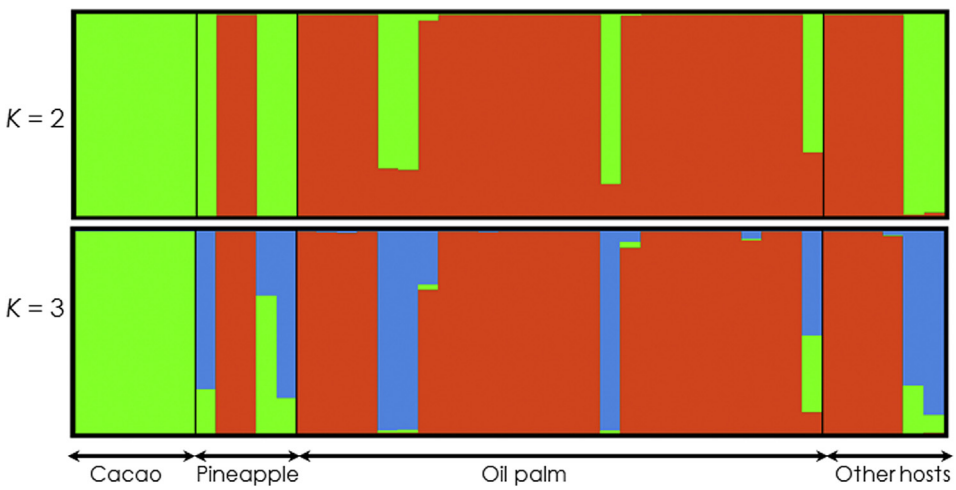

Fig 2 – Genetic structure of *Thielaviopsis ethacetica* inferred using Bayesian clustering analyses based on seven microsatellite markers with STRUCTURE v2.3.4 (Pritchard *et al.* 2000), assuming $K = 2$ and $K = 3$ genetic clusters (indicated by different colours). Individual multilocus genotypes (MLGs) are represented by discrete vertical bars and their affiliation expressed in terms of height proportion of each colour within the vertical bar. MLGs showing more than one colour are admixed. Vertical lines separate isolates from different hosts, as indicated by the labels below the figures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

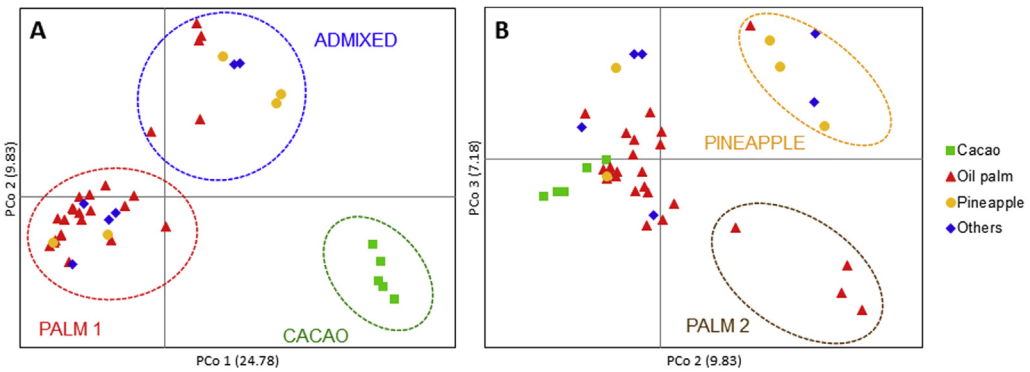

Fig 3 – Genotype grouping based on seven microsatellite markers of unique multilocus genotypes (MLGs) from the available collection of *Thielaviopsis ethacetica* using Principal Coordinates Analysis (PCoA) with GenAlEx 6.5 (Peakall & Smouse 2012). Graph A, based on the first and second ordination coordinate, enables discrimination of three different (CACAO, PALM 1, and ADMIXED) genetic clusters. Graph B, based on the second and third ordination coordinates, enables to split the ADMIXED cluster into the PALM 2 and PINEAPPLE clusters. The three coordinates explain 42 % of observed genetic variation. Labels indicate the hosts from which isolates were collected.

identity to the STRUCTURE Red cluster, assuming $K = 3$. The third cluster (designated as ADMIXED genetic group) included all admixed MLGs identified by STRUCTURE analyses either assuming $K = 2$ or $K = 3$ (including the three MLGs assigned to the Blue cluster). The first principal coordinate of the PCoA alone discriminated the PALM 1 group from the CACAO and ADMIXED groups, and the second component separated the ADMIXED from the PALM 1 and CACAO groups. When considering the third coordinate, it was possible to split the ADMIXED group into two genetic groups (Fig 3B). The one group (designated as PALM 2) corresponded to the Blue cluster defined by STRUCTURE, assuming $K = 3$, plus one MLG. The remainder of admixed MLGs formed another group, designated as PINEAPPLE genetic group.

The PALM 1 was the most abundant and diverse genetic group from our collection, including 31 isolates and 26 MLGs (see also Table 1). Of these, 27 isolates were from Cameroon ($H = 0.517$; $H_{corr} = 0.522$; $G^* = 73$ %), including 22 isolates and 19 MLGs from oil palm, three isolates and two genotypes from pineapple, and two isolates and one genotype from *Erytrophleum ivorense*. Three international MLGs from oil palm in Nigeria, date palm in Spain, and sugarcane in South Africa also belonged to this group. The PALM 2 included six isolates and four MLGs, all collected from oil palm in Cameroon ($H = 0.397$; $H_{corr} = 0.464$; $G^* = 50$ %). The CACAO group included 24 isolates and 6 MLGs, all collected from cacao in Cameroon ($H = 0.093$; $H_{corr} = 0.206$; $G^* = 9$ %). The PINEAPPLE group included 19 isolates and six MLGs. Of these, 14 isolates representing the same MLG were collected from pineapple in Cameroon ($H = H_{corr} = 0$; $G^* = 11$ %). This group also included two international isolates from pineapple in Brazil and Indonesia respectively, one from coconut palm in Tanzania, one from oil palm in Papua New Guinea and one from sugarcane in the USA. Overall, the distribution of these groups in Cameroon reflected a high degree of host specialization. For example, all isolates collected from cacao belonged to the CACAO group; all isolates collected from oil palm belonged to the PALM (PALM 1 + PALM 2) group, while the majority (>80 %) of isolates collected from pineapple belonged to the PINEAPPLE group. Only three cases of cross host infection were recorded, including two PALM 1-related MLGs recovered from pineapple and one from *E. ivorense*.

Phylogenetic reconstructions

The six microsatellite loci (CESSR06, CESSR17, CESSR19, CESSR20, CESSR21, and CESSR39) retained for phylogenetic analyses included 25 SNPs (Table 2). Seventeen of these were parsimony-informative. Most parsimonious trees (MPTs) based on single SSR loci produced different schemes of genetic relatedness among MLGs (Fig 4). None could discriminate any of the four genetic groups identified from population genetic analyses with significant statistical support. However, two loci (CESSR17 and CESSR20) consistently delineated two clades with strong bootstrap support. The one clade included the PALM (PALM 1 & PALM 2) group and the second clade included the CACAO and PINEAPPLE groups. The same structure was obtained with locus CESSR21, with the exception of isolate IMI378943 from the PINEAPPLE group, which was grouped with the PALM group, emerging as a putative hybrid between the two clades. Another two-clade phylogenetic structure, with strong bootstrap support, was produced by locus CESSR19. But here, the PALM 2 group clustered with the CACAO and PINEAPPLE groups, while the PALM 1 formed the second clade.

When combined, the six SSR loci produced a total of 35 equal MPTs (TL: 25, Fig 5A). All four genetic groups were well-delineated, although goodness of character fit indices were weak (TL: 26; CI: 0.654; RI: 0.029; RC: 0.346) and no individual genetic group was statistically supported. In contrast, a very strong bootstrap value supported the clade formed by the PALM 1 and PALM 2 groups as distinct from another clade including the CACAO and PINEAPPLE group. Isolate IMI378943 occupied an intermediary position between the PALM and CACAO/PINEAPPLE clades.

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

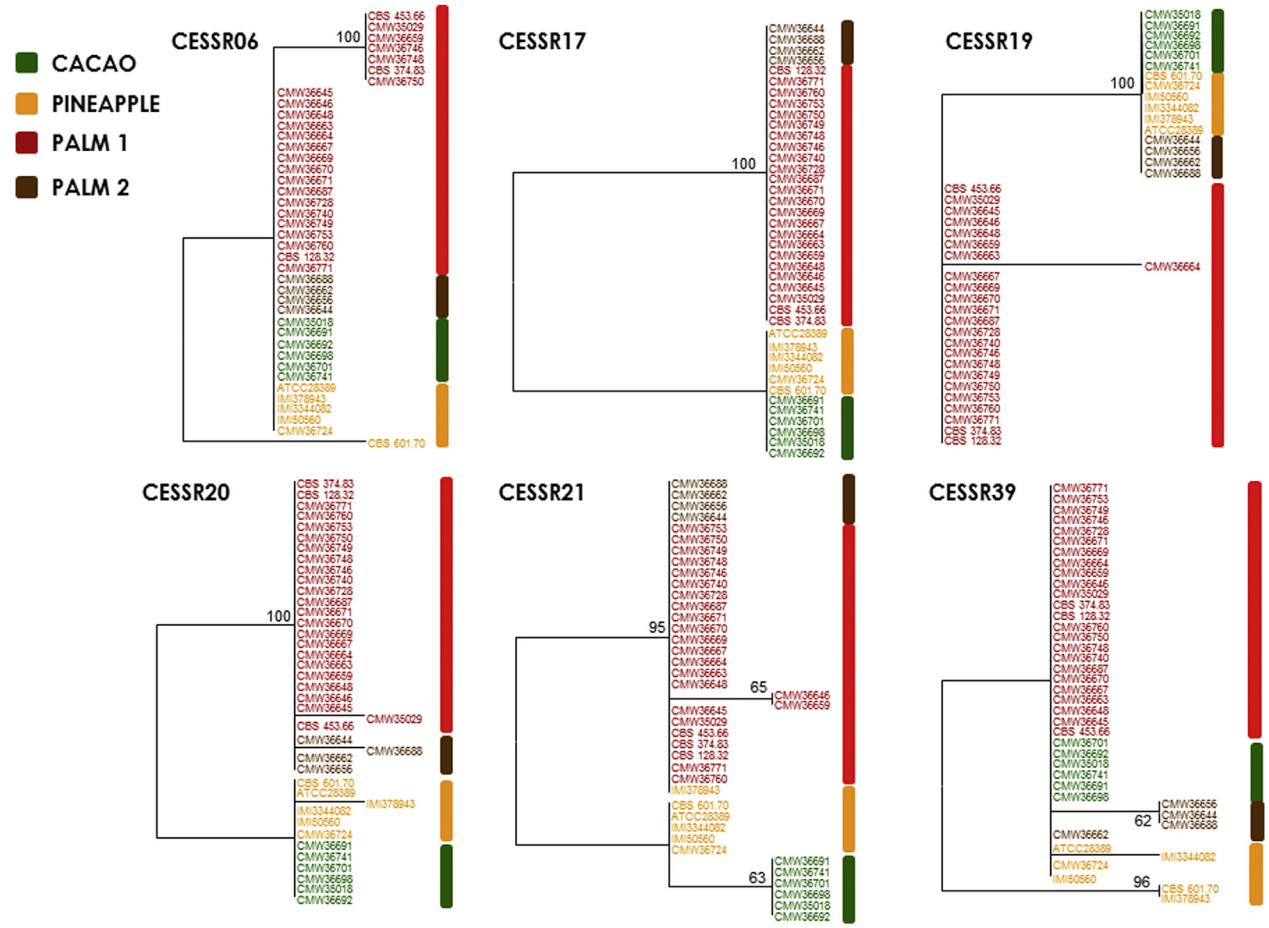

Fig 4 – Most parsimonious trees based on flanking regions at six single simple sequence repeat (SSR) loci including unique multilocus genotypes (MLGs) of *Thielaviopsis ethacetica*.

The *tef*1-α sequence alignment data set included 20 variable characters from a total of 637 characters. Of these, 12 were parsimony-informative. The phylogenetic reconstruction based on *tef*1-α generated a single MPT (TL: 12, Fig 5B), with maximum goodness of character fit indices (CI: 1.000; RI: 1.000; RC: 1.000). The general structure of the phylogram supported the combined SSR phylogeny, including two major and statistically supported clades. All the genetic groups from population genetic analyses were also recovered here and their clades were supported by strong bootstrap values (>75 %). The only discrepancies found between the *tef*1-α and combined SSR phylogenies was the grouping of isolate IMI3344082 from the PINEAPPLE genetic group within the CACAO clade. This isolate emerged as a putative hybrid between the PINEAPPLE and the CACAO group.

Results from network analyses with combined SSR and *tef*1-α sequence revealed a major split between the PALM 1 and PALM 2 groups on the one side and the CACAO and PINEAPPLE groups on the other side (Fig 6). Another statistically supported shift was observed between the PALM 1 and PALM 2 groups, though with residual reticulation. Substantial reticulation occurred between the CACAO and PINEAPPLE groups as well as within each genetic group considered individually.

Random mating

Analyses of index of association amongst loci were undertaken for three Cameroonian populations: (i) PALM 1, (ii) PALM (PALM 1 + PALM), and (iii) CAMEROON (PALM + CACAO + PINEAPPLE). The results showed increasing departure from a scenario of free, random mating as more complex populations were considered (Supplementary material — Fig S2). The observed index of association fell within the 95 % confidence range of random mating for the PALM 1 population ($I_A = 0.199$; $p = 0.108$) and outside the 95 % confidence range for the PALM population ($I_A = 0.739$; $p < 0.002$) and the CAMEROON population ($I_A = 1.799$; $p < 0.002$). This highlighted the existence of potential linkage disequilibrium among loci within the latter two populations and, therefore, the likely lack of ongoing interbreeding between individuals representing the different genetic groups of *Thielaviopsis ethacetica* in Cameroon.

Discussion

A combination of phylogenetic and population genetic analyses provided evidence to suggest that the cosmopolitan and

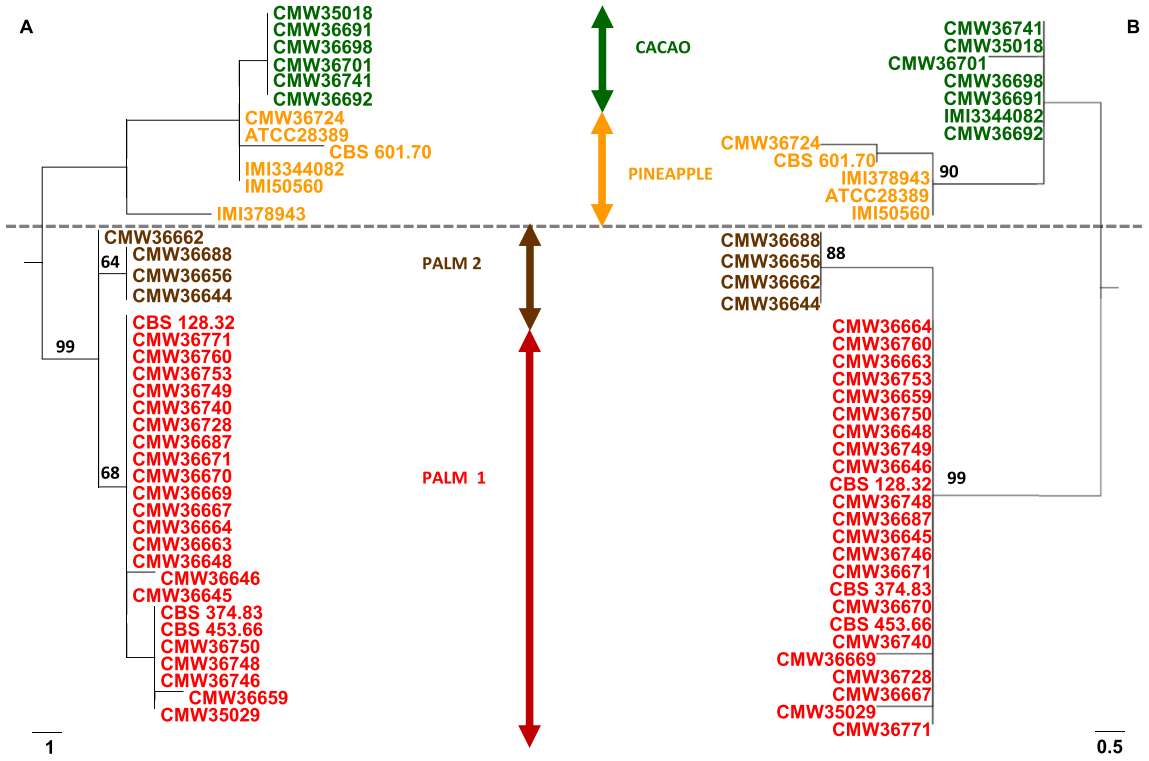

Fig 5 – Most parsimonious phylogenetic trees including unique multilocus genotypes (MLGs) of *Thielaviopsis ethacetica* based on combined flanking regions of six simple sequence repeat (SSR) loci (A) and the translation elongation factor 1-alpha (*tef*1-α) gene (B). The genetic groups identified from population genetic analyses are highlighted, as well as the demarcation between two putative discreet evolutionary lineages (dashed line). Trees were generated using a full heuristic search of 1000 replicates with PAUP version 4.0b10* (Swofford 2002). Values from 1000 bootstrap replicates are indicated above branches.

broad host range plant pathogen, *Thielaviopsis ethacetica*, includes at least four well-differentiated genetic groups. All of these occurred sympatrically in Cameroon where the greatest number of isolates originated. Yet, there was no evidence for ongoing gene flow between them, as seen in the very limited number of alleles shared among groups. This could imply that these groups are reproductively isolated and, thus, represent discrete evolutionary entities. Two of these groups were prevalent on oil palm (a native plant) in Cameroon, where a high genetic diversity of the pathogen suggested that they may be indigenous to the country, or have been present for an extended period of time. In contrast, the remaining two groups, found exclusively on nonnative cacao and pineapple respectively, had low genetic diversity and have most likely been introduced. These groups have evolved multiple mutations on the *tef*1-α gene and the flanking regions of SSR loci, enabling a statistically supported delineation of two phylogenetic taxa within *Th. ethacetica*.

We initially defined populations of *Th. ethacetica* in Cameroon based on the hosts from which isolates had been collected, and genetic diversity indices were estimated on these host-associated populations. The results showed that the genetic diversity of *Th. ethacetica* on oil palm was particularly high. The value of gene diversity estimated for this population ($H > 0.600$) was substantially greater than those of all purported introduced populations of other species of Ceratocystidaceae where SSR markers have been applied (Engelbrecht *et al.* 2004; van Wyk *et al.* 2006; Kamgan Nkuekam *et al.* 2009). It was also greater or at least comparable with those for native populations of these fungi (Roux *et al.* 2001; Engelbrecht *et al.* 2004; 2007; Barnes *et al.* 2005; Ferreira *et al.* 2010). Gene diversity was also substantially greater when compared to the values obtained by Alvarez *et al.* (2012) for populations of *Th. ethacetica* (treated as *Thielaviopsis paradoxa*) on oil palm trees in Colombia, Ecuador, and Brazil, using random amplified polymorphic DNA (RAPD) markers. Therefore, our results strongly suggest that the population of *Th. ethacetica* associated with oil palm in Cameroon may be native to this country, although ancient introduction events cannot be completely dismissed until studies can be conducted in other regions. This population would have coevolved with oil palm, the native range of which includes the humid forests of west and south-western Africa, extending from Gambia to Angola, including southern Cameroon (Obahiagbon 2012).

In contrast to the results obtained for *Th. ethacetica* on oil palm, isolates collected from cacao in Cameroon displayed little genetic diversity, reflecting a possible genetic bottleneck or founder effect, which is a characteristic of introduced populations (Nei *et al.* 1975). Despite the observation of ascomata, and thus evidence of sexual reproduction, on cacao pod husks in the field, and the fact that some of the isolates obtained from this host had been isolated as ascospore droplets (Mbenoun *et al.* 2014), the gene diversity was very low on cacao ($H < 0.1$, without clone-correction). Irrespective of sampling

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

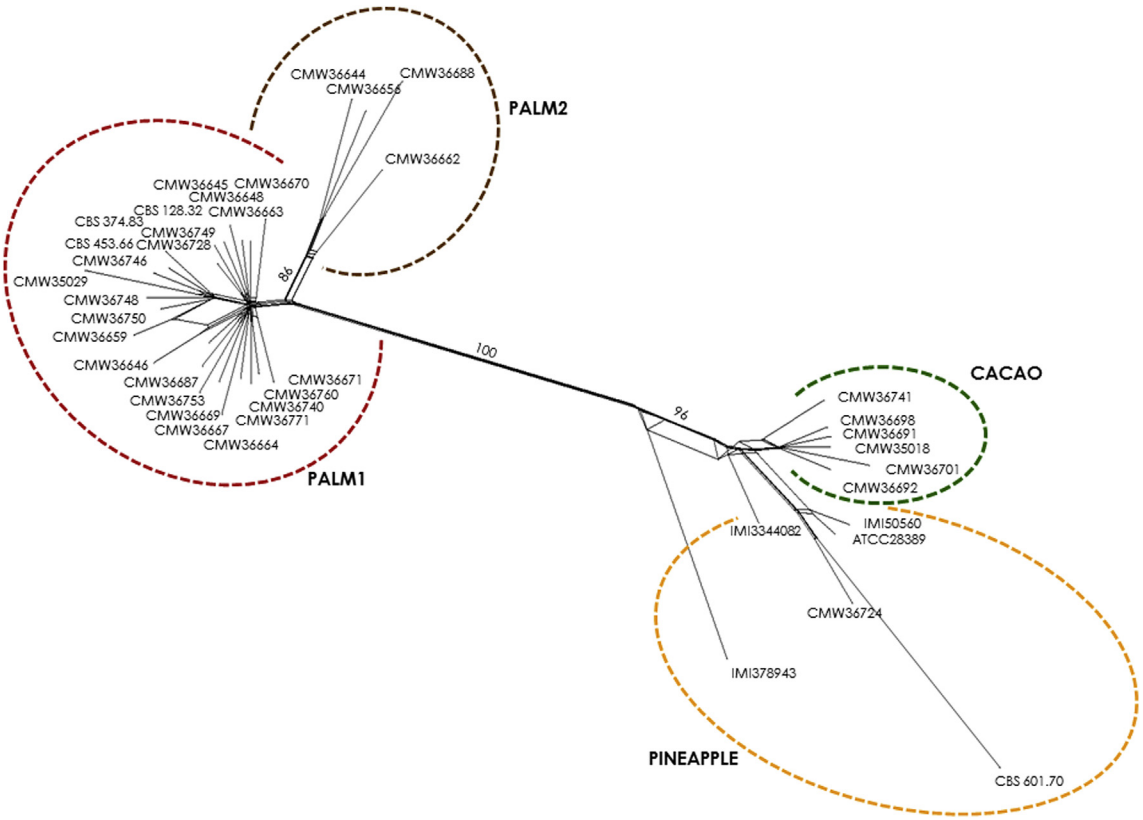

Fig 6 – Reticulate network of unique multilocus genotypes (MLGs) of *Thielaviopsis ethacetica* based on a combination of sequence data from six simple sequence repeat (SSR) loci and the translation elongation factor 1-alpha (*tef*1-α) gene. The network was constructed using the program SplitsTree, applying the GTR character transformation, NeighborNet distance transformation, and EqualAngle split transformation settings. Bootstrap support values (≥70 %) from 1000 replications are indicated next to split branches.

sites, the population was dominated by one MLG (IGAGEFE), representing 67 % of all isolates obtained from cacao. Five of the six MLGs from cacao differed at only one locus. These results support the view that the cacao population of *Th. ethacetica* in Cameroon arose from a small number of individuals. This was probably introduced and in view of the lack of any evidence for a possible host shift from oil palm to cacao, probably spread throughout the country with its host.

Low genetic diversity of *Th. ethacetica* was generally observed for isolates from pineapple in Cameroon. These isolates were collected from two fields at one location. In the first field, 14 isolates were obtained from damaged leaves following harvesting. They represented the same MLG (HJBAJHG), highlighting the absence of genetic diversity in that field. Greater diversity was observed in the second field, where three isolates representing two MLGs were collected from rotten fruits and fruit peduncles. An introduction similar to that suggested for the cacao isolates, could apply for pineapple when the first sampling field is considered. From the similarity in allele composition with MLGs associated with oil palm, we suggest that the two MLGs recovered from the second field may have originally been from oil palm. The limited and restricted sampling on pineapple does not allow for robust conclusions regarding the origin of the pathogen genotypes infecting this host in Cameroon. But, the identification of two MLGs related to HJBAJHG from pineapple in Brazil and Indonesia respectively, supports the hypothesis that a genetic group of *Th. ethacetica* may have spread around the world along with pineapple cultivation.

Genetic clustering analyses identified four genetic groups (or gene pools) for the collection of *Th. ethacetica* used in this study, and these were designated as CACAO, PALM 1, PALM 2, and PINEAPPLE, based on their host associations. These same four groups were also recovered in two separate MP-based phylogenies derived from sequences of the *tef*1-α gene and combined flanking regions of six SSR loci respectively. In the two phylogenies, the PALM 1 and PALM 2 groups emerged as descendants of the same ancestral population and formed a strong statistically supported clade, while the CACAO and PINEAPPLE groups descended from another ancestral population and formed another clade. This illustrated a phylogenetic structure with two main internal lineages for *Th. ethacetica*, which was also supported by single gene phylogenies of two SSR loci (CESSR17 and CESSR20). Following the principles of the genealogical concordance PSC (Taylor *et al.* 2000), the congruence between independent gene loci, as represented by the CESSR17, CESSR20, and *tef*1-α, suggests that the two main lineages of *Th. ethacetica* represent two discrete

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

species. Additional support for the two species hypothesis was provided by network analyses, which revealed a major split between lineages and substantial reticulation within each lineage.

There is good evidence to suggest that the two discrete lineages identified for *Th. ethacetica* have diverged relatively recently. First, the differentiation observed between them at the *tef*1-α gene and flanking regions of some SSR loci is not reflected at the beta tubulin (β-*tub*) and ribosomal ITS loci (Mbenoun *et al.* 2014). These are two markers with strong resolution and are commonly used for species phylogenetic reconstructions in the Ceratocystidaceae (de Beer *et al.* 2014). Also, the identification of a putative hybrid (isolate IMI378943) supports the notion that these groups have not reached the stage of complete reproductive isolation. Indeed, representative isolates from all genetic groups of *Th. ethacetica* were crossed *in vitro* (Mbenoun *et al.* 2014), resulting in the formation of ascomata for all possible combinations; and the progenies of these crosses were found to be viable in laboratory conditions.

The apparent absence of intrinsic reproductive barriers between the different genetic groups of *Th. ethacetica* as suggested by *in vitro* mating tests (Mbenoun *et al.* 2014) was not supported by analyses of I_A among loci of *Th. ethacetica* isolates from Cameroon, where all four groups occur sympatrically. In line with the identification of isolates representing the two mating types in the Cameroonian PALM 1 population (Mbenoun *et al.* 2014), the observed value of I_A supported the existence of ongoing natural random mating in this population. However, when more complex combinations of populations, PALM (PALM1 + PALM 2) and CAMEROON (PALM + CACAO + PINEAPPLE), were tested, the estimated I_A departed increasingly and significantly from a zero value that would be expected where interbreeding is occurring freely. This would suggest a reduced gene flow and the improbability that natural random mating is occurring between individuals representing different genetic groups of *Th. ethacetica* in Cameroon.

The low level of occurrence of shared alleles provides an additional argument to discount recurrent intergroup recombination within *Th. ethacetica* in Cameroon. In total, only six of 69 alleles were shared between groups, including two of 17 between CACAO and PINEAPPLE and four of 46 between PALM 1 and PALM 2, while no alleles were shared between the CACAO and PALM or between the PINEAPPLE and PALM groups (i.e., between the two lineages of the fungus). In comparison, Begoude Boyogueno *et al.* (2012), using five polymorphic SSR markers, found five shared alleles out of 22 detected between the two sibling species, *Lasiodiplodia theobromae* (Pat.) Griffon & Maubl and *Lasiodiplodia pseudotheobromae* A.J.L. Phyllips, A. Alves & Crous, coinfecting cacao and *Terminalia* trees in Cameroon. Likewise, Pérez *et al.* (2012) identified three cryptic species associated with the eucalypt leaf pathogen, *Teratosphaeria nubilosa*, (Cooke) Crous & U. Braun, in Australia, which shared seven out of 92 alleles, based on eight polymorphic SSR loci. Thus, the level of shared alleles between the different genetic groups representing *Th. ethacetica* in Cameroon conforms to expectations, when comparing sympatric populations representing closely related but distinct species under the BSC (Mayer 1942).

A number of ecological factors could explain the maintenance in sympatry of divergent populations but that lack complete intrinsic, reproductive isolation (Giraud *et al.* 2008; Gladieux *et al.* 2011a,b; and references therein). These factors include premating and postmating reproductive barriers. Among the premating barriers, host specificity, which results in reduced viability of immigrants, is one that could be considered for *Th. ethacetica* in Cameroon. This is because a high level of host specificity was noted in the distribution of the different genetic groups of this fungus in Cameroon. All isolates collected from cacao belonged to the CACAO groups, all isolates collected from oil palm belonged to PALM group and more than 80 % of isolates collected from pineapple represented the same PINEAPPLE genotype. This translated into high genetic differentiation measured between host-associated populations as well as a substantial proportion of genetic variance explained by variation between hosts. However, the cases of cross host infection (as incidental as they may be) and the broader range of hosts from which international isolates were collected, highlights the need for further investigation into the nature and stringency of host specificity within *Th. ethacetica*.

Another potential ecological premating barrier to gene flow in *Th. ethacetica* that deserves attention is that concerning the insect vectors associated with this fungus on different hosts. Nitidulid beetles (Nitidulidae, Coleoptera), which are well-known vectors of the Ceratocystidaceae fungi (e.g., Moller & DeVay 1968; Chang & Jensen 1973; Heath *et al.* 2009), were commonly found in association with *Th. ethacetica* on cacao and pineapple in Cameroon, but they were never found on oil palm on which a different community of insects was found (unpublished data). In either case, premating ecological barriers can hardly explain the divergence between the PALM 1 and PALM 2 genetic groups, both of which occurred on the same host in Cameroon. A more plausible ecological reproductive barrier in this case would be postmating, linked to the lack of fitness of hybrids in natural environments, although they may behave differently under laboratory conditions (Harrington & Rizzo 1999; Giraud *et al.* 2008).

Conclusions

Collectively, the results of this study and previous biological and morphological investigations (Mbenoun *et al.* 2014) suggest that *Thielaviopsis ethacetica* represents a taxonomic complex. This would include at least two morphologically cryptic, but genetically well-differentiated emerging species based on the BSC and GCPSC species concepts. One of these putative incipient species could have evolved in Africa, in association with and within the native range of oil palm. The origin of the other species could not be inferred because of limited sampling. Its prevalence on two crop plants (cacao and pineapple), originating in South America, however, suggests that the latter species may have evolved in that region, although the complexity of the dispersion histories of these crops allows for alternative hypotheses. Isolates of both species have been moved extensively through anthropogenic agricultural activities, expanding their host ranges and facilitating secondary contact where hitherto extrinsic geographic barriers existed. This, together with the maintenance of an interbreeding potential establishes opportunities for

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009

hybridisation and introgression in the nascent *Th. ethacetica* species complex.

Acknowledgements

We thank Thierry Hoareau and Emilie Boissin for enlightening and helpful discussions during the preparation of this manuscript. We thank the anonymous reviewers whose constructive comments and suggestions helped improve the manuscript. Funding was provided by the Department of Science and Technology (DST)/National Research Foundation (NRF), Centre of Excellence in Tree Health Biotechnology (CTHB) of South Africa (Grant number UID 40945), the Forestry and Agricultural Biotechnology Institute (FABI), and the Department of Corporate International Relations of the University of Pretoria. The Institute of Agricultural Research for Development (IRAD) provided logistical assistance during field collections in Cameroon.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.funbio.2015.05.009.

REFERENCES

Agapow PM, Burt A, 2000. *Multilocus 1.2*. Department of Biology, Imperial College, Silwood Park, Ascot, UK.

Alvarez E, Llano GA, Loke JB, Chacon MI, 2012. Characterization of *Thielaviopsis paradoxa* Isolates from oil palms in Colombia, Ecuador and Brazil. *Journal of Phytopathology* **160**: 690–700.

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P, 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution* **19**: 535–544.

Barnes I, Nakabonge G, Roux J, Wingfield BD, Wingfield MJ, 2005. Comparison of populations of the wilt pathogen *Ceratocystis albifundus* in South Africa and Uganda. *Plant Pathology* **54**: 189–195.

Bass D, Richards TA, 2011. Three reasons to re-evaluate fungal diversity "on Earth and in The Ocean". *Fungal Biology Reviews* **25**: 159–164.

Begoude Boyogueno AD, Slippers B, Perez G, Wingfield MJ, Roux J, 2012. High gene flow and outcrossing within populations of two cryptic fungal pathogens on a native and non-native host in Cameroon. *Fungal Biology* **116**: 343–353.

Brasier CM, 2008. The biosecurity threat to the UK and global environment from international trade in plants. *Plant Pathology* **57**: 792–808.

Brown AHD, Feldman MW, Nevo E, 1980. Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* **96**: 523–536.

Cai L, Hyde KD, Taylor PWJ, Weir BS, Waller J, Abang MM, Zhang JZ, Yang YL, Phoulivong S, Liu ZY, Prihastuti H, Shivas RG, McKenzie EHC, Johnston PR, 2009. A polyphasic approach for studying *Colletotrichum*. *Fungal Diversity* **39**: 183–204.

Chang VCS, Jensen L, 1973. Transmission of the pineapple disease organism of sugarcane. *Journal of Economical Entomology* **67**: 190–192.

Chen RS, Boeger JM, McDonald BA, 1994. Genetic stability in a population of a plant pathogenic fungus over time. *Molecular Ecology* **3**: 209–218.

de Queiroz K, 2007. Species concepts and species delimitation. *Systematic Biology* **56**: 879–886.

de Seynes J, 1886. *Recherches pour servir à l'histoire naturelle des végétaux inférieurs III*. Masson, Paris.

de Beer ZW, Duong TA, Barnes I, Wingfield BD, Wingfield MJ, 2014. Redefining *Ceratocystis* and allied genera. *Studies in Mycology* **79**: 187–219.

Edgar RC, 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113. http://dx.doi.org/10.1186/1471-2105-5-113.

Engelbrecht CJB, Harrington TC, Steimel J, Capretti P, 2004. Genetic variation in eastern North American and putatively introduced populations of *Ceratocystis fimbriata* f. *platani*. *Molecular Ecology* **13**: 2995–3005.

Engelbrecht CJB, Harrington TC, Alfenas AC, Suarez C, 2007. Genetic variation in populations of the cacao wilt pathogen, *Ceratocystis cacaofunesta*. *Plant Pathology* **56**: 923–933.

Evanno G, Regnaut S, Goudet J, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Excoffier L, Smouse PE, Quattro JM, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes – application to human mitochondrial-DNA restriction data. *Genetics* **131**: 479–491.

Ferreira EM, Harrington TC, Thorpe DJ, Alfenas AC, 2010. Genetic diversity and interfertility among highly differentiated populations of *Ceratocystis fimbriata* in Brazil. *Plant Pathology* **59**: 721–735.

Fisher MC, Koenig GL, White TJ, Taylor JW, 2002. Molecular and phenotypic description of *Coccidioides posadasii* sp. nov., previously recognized as the non-California population of *Coccidioides immitis*. *Mycologia* **94**: 73–84.

Giraud T, Refrégier G, Le Gac M, de Vienne DM, Hood ME, 2008. Speciation in fungi. *Fungal Genetics and Biology* **45**: 791–802.

Gladieux P, Guerin F, Giraud T, Caffier V, Lemaire C, Parisi L, Didelot F, Le Cam B, 2011a. Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Molecular Ecology* **20**: 4521–4532.

Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, Couloux A, Giraud T, 2011b. Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Molecular Biology and Evolution* **28**: 459–471.

Grünwald NJ, Goodwin SB, Milgroom MG, Fry WE, 2003. Analysis of genotypic diversity data for populations of microorganisms. *Phytopathology* **93**: 738–746.

Harrington TC, Rizzo DM, 1999. Defining species in the fungi. In: Worrall JJ (ed.), *Structure and Dynamics of Fungal Populations*. Kluwer Press, Dordrecht, pp. 43–71.

Heath RN, Wingfield MJ, van Wyk M, Roux J, 2009. Insect associates of *Ceratocystis albifundus* and patterns of association in a native savanna ecosystem in South Africa. *Environmental Entomology* **38**: 356–364.

Höhnel FXR, 1904. Zur kentniss einige fadenpilze. *Hedwigia* **43**: 295–299.

Huson DH, Bryant D, 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.

Jacobs K, Bergdahl DR, Wingfield MJ, Halik S, Seifert KA, Bright DE, Wingfield BD, 2004. *Leptographium wingfieldii* introduced into North America and found associated with exotic *Tomicus piniperda* and native bark beetles. *Mycological Research* **108**: 411–418.

Jakobsson M, Rosenberg NA, 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and

multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.

Kamgan Nkuekam G, Barnes I, Wingfield MJ, Roux J, 2009. Distribution and population diversity of *Ceratocystis pirilliformis* in South Africa. *Mycologia* **101**: 17–25.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Øvreås L, Reysenbach AL, Smith VH, Staley JT, 2006. Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology* **4**: 102–112.

Mayer E, 1942. *Systematics and the Origin of Species*. Columbia University Press, Irwington.

Mbenoun M, de Beer ZW, Wingfield MJ, Wingfield BD, Roux J, 2014. Reconsidering species boundaries in the *Ceratocystis paradoxa* complex, including a new species from oil palm and cacao in Cameroon. *Mycologia* **106**: 757–784.

McDonald BA, Linde C, 2002. The population genetics of plant pathogens and breeding strategies for durable resistance. *Euphytica* **124**: 459–471.

Moller WJ, DeVay JE, 1968. Insect transmission of *Ceratocystis fimbriata* in deciduous fruit orchards. *Phytopathology* **58**: 1499–1508.

Morton J, 1987. Pineapple. In: Morton JF, Miami FL (eds), *Fruits of Warm Climates*, pp. 18–28.

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C, 2002. Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* **89**: 380–386.

Nei M, 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **70**: 3321–3323.

Nei M, Maruyama T, Chakraborty R, 1975. The bottleneck effect and genetic variability in populations. *Evolution* **29**: 1–10.

Obahiagbon FI, 2012. Aspect of the African oil palm (*Elaeis guineensis* jacq.) and the implications of its bioactives in human health. *American Journal of Biochemistry and Molecular Biology*. http://dx.doi.org/10.3923/ajbmb.

Peakall R, Smouse PE, 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* **28**: 2537–2539.

Pérez G, Slippers B, Wingfield MJ, Wingfield BD, Carnegie AJ, Burgess TI, 2012. Cryptic species, native populations and biological invasions by a eucalypt forest pathogen. *Molecular Ecology* **21**: 4452–4471.

Pinho DB, Dutra DC, Pereira OL, 2013. Notes on *Ceratocystis paradoxa* causing internal post-harvest rot disease on immature coconut in Brazil. *Tropical Plant Pathology* **38**: 152–157.

Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Quaedvlieg W, Binder M, Groenewald JZ, Summerell BA, Carnegie AJ, Burgess TI, Crous PW, 2014. Introducing the Consolidated Species Concept to resolve species in the Teratosphaeriaceae. *Persoonia* **33**: 1–40.

Roux J, Harrington TC, Steimel JP, Wingfield MJ, 2001. Genetic variation in the wattle wilt pathogen *Ceratocystis albofundus*. *Mycoscience* **42**: 327–332.

Rosenberg NA, 2004. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**: 137–138.

Samson RA, Varga J, 2009. What is a species in *Aspergillus*? *Medical Mycology* **47** (Supplement I): S13–S20.

Santana Q, Coetzee M, Steenkamp E, Mlonyeni O, Hammond G, Wingfield MJ, Wingfield BD, 2009. Microsatellite discovery by deep sequencing of enriched genomic libraries. *BioTechniques* **46**: 217–223.

Sites Jr JW, Marshall JC, 2004. Operational criteria for delimiting species. *Annual Review of Ecology, Evolution and Systematics* **35**: 199–227.

Stoddart JA, Taylor JF, 1988. Genotypic diversity: estimation and prediction in samples. *Genetics* **118**: 705–711.

Suwandi, Akino S, Kondo N, 2012. Common spear rot of oil palm in Indonesia. *Plant Disease* **96**: 537–543.

Swofford DL, 2002. *Phylogenetic Analysis using Parsimony (*and Other Methods), Version 4*. Sinauer Associates, Sunderland, Massachusetts.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731–2739.

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC, 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology* **31**: 21–32.

Temnykh S, Lukashova A, Cartinhour S, DeClerk G, Lipovich L, McCouch S, 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* **11**: 1441–1452.

Tzeng SJ, Sun EJ, Hsiao WW, 2010. First report of fruit basal rot by *Ceratocystis paradoxa* on coconut in Taiwan. *Plant Disease* **94**: 487.

Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG, 2012. Primer3-new capabilities and interfaces. *Nucleic Acids Research* **40**: e115.

van Wyk M, Van Der Merwe NA, Roux J, Wingfield BD, Kamgan Nkuekam G, Wingfield MJ, 2006. Population genetic analyses suggest that the *Eucalyptus* fungal pathogen *Ceratocystis fimbriata* has been introduced into South Africa. *South African Journal of Botany* **102**: 259–263.

Voigt K, Kirk PM, 2011. Recent developments in the taxonomic affiliation and phylogenetic positioning of fungi: impact in applied microbiology and environmental biotechnology. *Applied Microbiology and Biotechnology* **90**: 41–57.

VonHoldt BM, 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361.

Walker DM, Castlebury LA, Rossman AY, White JF, 2012. New molecular markers for fungal phylogenetics: two genes for species-level systematics in the Sordariomycetes (Ascomycota). *Molecular Phylogenetics and Evolution* **64**: 500–512.

Went FAFC, 1893. De ananasziekte van het suikerriet. *Mededeelingen van het Profestation West-Java* **5**: 1–8.

Yeh FC, Yang R-C, Boyle T, 1999. *POPGENE Version 1.31. Microsoft Windows Based Freeware for Population Genetic Analysis*. University of Alberta, Alberta.

Yu F-Y, Niu X-Q, Tang Q-H, Zhu H, Song W-W, Qin W-Q, Lin C-H, 2012. First report of stem bleeding in coconut caused by *Ceratocystis paradoxa in* Taiwan, China. *Plant Disease* **96**: 290.

Please cite this article in press as: Mbenoun M, et al., Independent origins and incipient speciation among host-associated populations of *Thielaviopsis ethacetica* in Cameroon, Fungal Biology (2015), http://dx.doi.org/10.1016/j.funbio.2015.05.009