Ancient duplication and horizontal transfer of a toxin gene cluster reveals novel mechanisms in the cercosporin biosynthesis pathway

Ronnie de Jonge[1,2,3,4,†,*], Malaika K. Ebert[5,6,7 †], Callie R. Huitt-Roehl[8], Paramita Pal[8], Jeffrey C. Suttle[5], Jonathan D. Neubauer[5], Wayne M. Jurick II[9], Gary A. Secor[6], Bart P.H.J. Thomma[7], Yves Van de Peer[1,2,3,10], Craig A. Townsend[8], & Melvin D. Bolton[5,6,*]

[1]Department of Plant Systems Biology, VIB, Ghent, Belgium. [2]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium. [3]Bioinformatics Institute Ghent, Ghent University, B-9052 Gent, Belgium. [4]Plant-Microbe Interactions, Department of Biology, Faculty of Science, Utrecht University, Utrecht, The Netherlands. [5]Northern Crop Science Laboratory, United States Department of Agriculture, Fargo, ND, United States. [6]Department of Plant Pathology, North Dakota State University, Fargo, ND, United States. [7]Laboratory of Phytopathology, Wageningen University, Wageningen, the Netherlands. [8]Department of Chemistry, The Johns Hopkins University, Baltimore, MD, United States. [9]Food Quality Laboratory, United States Department of Agriculture, Beltsville, MD, United States. [10]Department of Genetics, Genomics Research Institute, University of Pretoria, Pretoria, South Africa.

†These authors contributed equally to this work.

*Correspondence and requests for materials should be addressed to R.d.J. (r.dejonge@uu.nl) or M.D.B (Melvin.Bolton@ars.usda.gov).

Abstract

Cercospora species have a global distribution and are best known as the causal agents of leaf spot diseases of many crops. Cercospora leaf spot (CLS) is an economically devastating disease of sugar beet caused by *C. beticola*. The *C. beticola* genome encodes 63 biosynthetic gene clusters, including the cercosporin toxin biosynthesis (*CTB*) cluster. Studies spanning nearly 60 years have shown that cercosporin is photo-activated, critical for disease development, and toxic to most organisms except *Cercospora* spp. themselves, which exhibit cercosporin auto-resistance. We show that the *CTB* gene cluster has experienced an unprecedented number of duplications, losses, and horizontal transfers across a spectrum of plant pathogenic fungi. Although cercosporin biosynthesis has been widely assumed to rely on the eight gene *CTB* cluster, our comparative genomic analysis revealed extensive gene collinearity adjacent to the established cluster in all *CTB* cluster-harboring species. We demonstrate that the *CTB* cluster is larger than previously recognized and includes the extracellular proteins fasciclin and laccase required for cercosporin biosynthesis and the final pathway enzyme that installs the unusual cercosporin methylenedioxy bridge. Additionally, the expanded cluster contains CFP, which contributes to cercosporin auto-resistance in *C. beticola*. Together, our results give new insight on the intricate evolution of the *CTB* cluster.

Cercospora are among the most speciose genera in all Fungi. First described in 1863[1], the genus has sustained a long history largely due to notoriety as the causal agent of leaf spot diseases in a wide range of plants including agriculturally important crops such as sugar beet, soybean, maize and rice that together account for hundreds of millions of dollars in lost revenue annually to growers worldwide. Although *Cercospora* spp. share a number of characteristics associated with pathogenicity, including penetration through natural openings and extracellular growth during the biotrophic stage of infection, most rely on the production of the secondary metabolite (SM) cercosporin to facilitate infection[2,3]. SMs are bioactive molecules that play crucial roles in the establishment of specific ecological niches but, unlike primary metabolites, are not essential for fungal growth, development or reproduction. Cercosporin is a perylenequinone that is nearly universally toxic to a wide array of organisms including bacteria, mammals, plants, and most fungal species with the key exception of cercosporin-producing fungi, which secrete cercosporin in millimolar quantities with no apparent effect on growth and, therefore, exhibit auto-resistance[4]. When exposed to ambient light, cercosporin is a potent producer of reactive oxygen species (ROS) in the presence of oxygen[5] with a quantum efficiency of >80%[6]. This small molecule is lipophilic and can readily penetrate plant leaves leading to indiscriminant cellular damage within seconds of exposure[7]. Therefore, the physiological and genetic mechanisms underlying the ability for *Cercospora* spp. to tolerate cercosporin and associated ROS is an intriguing area of research with potentially important practical applications.

In contrast to the large body of information on cercosporin biology spanning several decades[8,9], the cercosporin toxin biosynthesis (*CTB*) gene cluster was only recently resolved in *C. nicotianae*[10]. The keystone enzyme for cercosporin biosynthesis, CTB1, bears all the hallmarks of an iterative, non-reducing polyketide synthase (NR-PKS)[11]. Using *CTB1* as a point of reference, the complete *C. nicotianae CTB* gene cluster was determined to consist of eight contiguous genes. Six of these genes are believed to be responsible for cercosporin assembly (*CTB1*, *2*, *3*, *5*, *6*, and *7*)[10,12]. The zinc finger transcription factor CTB8

co-regulates expression of the cluster[10], while the major facilitator superfamily (MFS) transporter CTB4 exports the final metabolite[13]. Downstream of the *CTB* cluster are two open reading frames (ORFs) encoding truncated transcription factors while loci designated as *ORF9* and *ORF10* upstream of the *CTB* cluster are not regulated by light and are not hypothesized to encode proteins with metabolic functions[10]. Consequently, the clustering of eight genes with demonstrated co-regulation by light that are flanked by ORFs with no apparent role in cercosporin biosynthesis has suggested that cercosporin production relies on the eight-gene *CTB* cluster[10]. In this study, we used an evolutionary comparative genomics approach to show that the *CTB* gene cluster underwent multiple rounds of duplication and was transferred horizontally across large taxonomic distances. Since these horizontal transfer events included the transfer of genes adjacent to the canonical eight gene *CTB* cluster, we show that the *CTB* cluster includes additional genes in *C. beticola,* including at least one gene that contributes to cercosporin auto-resistance and three previously unrecognized genes involved with biosynthesis.

Results

Secondary metabolite cluster expansion in *Cercospora beticola*.

C. beticola strain 09-40 was sequenced to 100-fold coverage and scaffolded with optical and genome maps, resulting in 96.5% of the 37.06 Mbp assembly being placed in 12 supercontigs. Despite their ubiquitous presence in nature and in many cropping systems, genome sequences of *Cercospora* spp. are not well-represented in public databases. Therefore to aid comparative analysis within the *Cercospora* genus, we also sequenced the genome of *C. berteroae* and reassembled the genome of *C. canescens*[14] (Supplementary Table 1). To identify gene clusters responsible for biosynthesis of aromatic polyketides in *C. beticola*, we mined the genome to identify all SM clusters[15] and compared these with predicted clusters in related Dothideomycete fungi. The *C. beticola* genome possesses a total of 63 SM clusters of several classes (Supplementary Table 2), representing a greatly expanded SM repertoire compared to closely related Dothideomycetes (Supplementary Table 3). In order to identify the *C. beticola* PKS cluster

responsible for cercosporin production, we compared the sequence of the *C. nicotianae CTB* cluster[10] with predicted PKS clusters of *C. beticola.* The *C. beticola* PKS *CBET3_00833* (*CbCTB1*) and flanking genes (*CBET3_00830* – *CBET3_00837*) were ~96% identical to *C. nicotianae CTB1* – *CTB8* and all genes were collinear, strongly suggesting this region houses the *CTB* cluster in *C. beticola*.

Repeated duplication and lateral transfer of the cercosporin biosynthetic cluster.

To study the evolutionary relationships *of C. beticola* PKSs, we conducted large-scale phylogenomic analyses that included various previously characterized PKSs from selected species[16]. Since resolving orthologous relationships among PKSs can predict the type of SM that will be synthesized, we first built a phylogenetic tree of the conserved core ketosynthase (KS) domains of each PKS that resulted in separating PKS enzymes into four major groups (Supplementary Fig. 1). Among the eight *C. beticola* NR-PKSs, phylogenetic analysis revealed significant similarity between *Cb*CTB1 and CBET3_10910-RA, which cluster at the base of the cercosporin clade (Supplementary Fig. 1). Interestingly, *CBET3_10910* flanking genes were also strikingly similar to *CbCTB* cluster genes (Fig. 1). Consequently, we hypothesize that the *CBET3_10910* SM cluster is the result of a *CTB* cluster duplication. Since duplicated SM gene clusters appeared to be relatively rare in fungi[17], we investigated the origin and specificity of the *CTB* cluster and the putative duplication by searching for *Cb*CTB1 homologs against a selected set of 48 published Ascomycete proteomes (Supplementary Table 4) representing a diverse group of fungal orders. We identified *Cb*CTB1 orthologs in *Cercospora* spp. *C. berteroae* and *C. canescens* and confirmed its presence in *Cladosporium fulvum*[16] and *Parastagonospora nodorum*[18]. Surprisingly, seven additional orthologs were identified in Sordariomycete species *Colletotrichum orbiculare*, *C. gloeosporioides*, *C. fioriniae*, *C. graminicola*, *C. higginsianum*, and *Magnaporthe oryzae* as well as one in the Leotiomycete *Sclerotinia sclerotiorum* (Supplementary Fig. 2), representing a diverse taxa harboring *CTB1*. Analysis of sequence similarity showed that intra-species (*Cb*CTB1 - CBET3_10910-RA) sequence similarity (65%) was lower than the inter-species similarity (e.g. *Cb*CTB1 and *C. fulvum* CTB1 (Clafu1_196875) are 73% similar;

Supplementary Table 5), suggesting that the *CTB1* duplication event was ancient and occurred prior to Dothideomycete speciation. Reconciliation[19] of the species tree with the CTB1 protein tree revealed that the predicted evolutionary history of CTB1 can be characterized by four duplications, three transfers and 22 losses (Fig. 2), and further corroborate our hypothesis that the *CTB1* duplication event (D1) occurred prior to Dothideomycete speciation. Reconciliation revealed an ancient transfer in which the lineage leading to *S. sclerotiorum* acquired the duplicated *CTB1* from the last common ancestor of *Cercospora* spp. (T1; Fig. 2A, B). Besides *S. sclerotiorum*, the only species that has retained the duplicated *CTB1* gene are *Cercospora* spp. and *P. nodorum* (Fig. 2B). Duplications 2-4 (D2-4) arose after lateral transfer (T2) of *CTB1* into the last common ancestor of the *Glomerellales*. *CTB1* was then transferred (T3) from a common ancestor in the *Glomerellales* to *Magnaporthe oryzae* (Fig. 2).

We extended the search for CTB cluster protein orthologs by scanning the 48 proteomes for homologs of *Cb*CTB2 (CBET3_00830) to *Cb*CTB8 (CBET3_00837) followed by phylogenetic tree construction and subtree selection (Supplementary Fig. 3). This resulted in the identification of orthologs in the same set of species previously listed to contain *CTB1*, with the only exceptions in cases where *CTB* gene homologs were lost in a species. Reconciliation of the subtrees for CTB2, CTB3, CTB4, CTB5 and CTB8 (Supplementary Fig. 4) largely supported the proposed scenario for CTB1 (Fig. 2B). Moreover, *CTB* orthologs were microsyntenic (Fig. 3). The extensive loss of CTB6 and CTB7 orthologs limits reconciliation analysis of these gene families. To further evaluate the significance of cross-species microsynteny, we estimated genome-wide microsynteny between *C. beticola* and *C. gloeosporioides* and between *C. beticola* and *M. oryzae* (Supplementary Fig. 5). Surprisingly, microsynteny of the *CTB* gene cluster is significantly differentiated from the genome-wide average, emphasizing its apparent evolutionary significance. *CTB* cluster GC content from *Cercospora* and *Colletotrichum* spp. was similar to the transcriptome-wide average in *Cercospora* spp. (54%) but different from the GC content of the *Colletotrichum* transcriptome (58%; Supplementary Table 6). Likewise, CTB protein similarity between *C.*

beticola and *Colletotrichum* spp., and to a lesser degree with *M. oryzae*, is significantly higher compared to the genome-wide average (Supplementary Fig. 6). Taken together, we hypothesize that the *CTB* cluster as a whole was transferred multiple times followed by species-specific evolutionary trajectories involving frequent gene loss. However, we cannot rule out the alternative hypothesis where *CTB*-like clusters identified in this study are the result of strong purifying selection of an ancient SM cluster, which was followed by frequent gene loss in nearly all other sequenced fungi. The outcome of reconciliation analyses depend heavily on the relative costs assigned to the various evolutionary events under consideration, namely horizontal gene transfer, gene duplication and gene loss[20]. Nonetheless, either scenario highlights the extraordinary evolution of the *CTB* cluster. Although *C. acutatum* has *CTB* cluster orthologs and is known to secrete a red SM *in vitro*[21], we were unable to identify cercosporin in *C. acutatum* or sister species *C. fioriniae*. However, it is not unlikely that cercosporin, or a comparable SM, is only synthesized under specific experimental conditions, a feature typically observed for SM biosynthesis.

Extension of the predicted cercosporin biosynthetic cluster based on microsynteny.

Besides the microsynteny of *CTB* genes between species, we also observed a striking level of similarity outside of the predicted *CTB* cluster to *C. beticola CBET3_00845* on the 3' end of the cluster (Fig. 1, 3). Considering the observed microsynteny and the significant co-expression of syntenic genes in *C. higginsianum*[22], we hypothesized that the *C. beticola CTB* cluster is significantly larger than previously described[10]. To test this, we first determined the relative expression of all established *C. beticola CTB* genes as well as a number of flanking genes (*CBET3_00828* to *CBET3_00848*) under light (cercosporin-inducing) compared to dark (cercosporin-repressing) conditions, which showed that all candidate *CTB* genes on the 3' flank were induced in the light except *CBET3_00846* and *CBET3_00848* (Supplementary Table 7). Functional annotation of these novel, induced genes revealed one non-conserved phenylalanine ammonia lyase, the major facilitator superfamily transporter CFP[23] (CBET3_00841), a candidate GA_4 desaturase (CBET3_00842), a candidate dehydratase (CBET3_00843), β-ig-h3 fasciclin (CBET3_00844),

laccase (CBET3_00845) and protein phosphatase 2A (CBET3_00847; Supplementary Table 7; 8), which have functions associated with multi-domain enzymes or polyketide biosynthesis in fungi or bacteria[10,24-29]. To confirm individual gene contributions for cercosporin production, we generated single gene deletion mutants of all candidate genes from *CBET3_00840* to *CBET3_00846* and tested their ability to produce cercosporin. These assays showed that cercosporin production in Δ*CBET3_00844* and Δ*CBET3_00845* mutants was completely abolished, while Δ*CBET3_00842* mutants accumulated a red, cercosporin-like metabolite that migrated differently in potato dextrose agar (PDA) culture plates and thin layer chromatography (TLC), and exhibited a different profile obtained via high-performance liquid chromatography (HPLC) (Fig. 4). Complementation of mutants restored the ability to produce cercosporin, validating their role in cercosporin biosynthesis (Fig. 4). All other mutants produced compounds with HPLC profiles similar to cercosporin (Supplementary Fig. 7), suggesting these genes are not involved with cercosporin biosynthesis. Taken together, these results corroborate our hypothesis that the *CTB* cluster extends to *CBET3_00845* at the 3' side and includes at least three novel *CTB* genes as well as *CFP*. Consequently, we propose naming genes *CBET3_00842, CBET3_00844* and *CBET_00845* as *CbCTB9* to *CbCTB11*, respectively (Supplementary Table 7).

Pre-cercosporin isolation and characterization

To characterize the red metabolite that accumulated in the Δ*CTB9* mutant, an ethyl acetate extract of the collected mycelia was analyzed by reverse-phase HPLC. At 280 nm, a single peak was observed and compared to a reference sample of cercosporin produced by *C. nicotianae*. The retention time of this peak was shorter than that of cercosporin (Fig. 5a) suggesting a more polar metabolite. Comparison of the UV-vis spectra (Fig. 5b) of the two compounds revealed nearly identical chromophores, suggesting close structural relation. The exact mass of the metabolite from the Δ*CTB9* mutant was determined (m/z = 537.1762, $[M+H^+]$), consistent with the elemental composition $C_{29}H_{29}O_{10}$. This mass is 2 Da greater than that of cercosporin (+2 hydrogens), which led to a proposed structure for pre-cercosporin (Fig. 5c).

Alternative hydroquinones of cercosporin could be excluded simply on the basis of the UV-vis spectral information and chemical instability. The presence of a free phenol in pre-cercosporin in place of the unusual 7-membered methylenedioxy of cercosporin is consonant with the red shift of the long wavelength λ_{max} and the shorter HPLC retention time.

To firmly support the tentative structure of pre-cercosporin, the crude extract obtained as above was further purified by reverse-phase HPLC and examined by ^{1}H-NMR spectroscopy. Immediately evident in the spectrum was the absence of the methylenedioxy singlet at δ5.74 diagnostic of cercosporin, but the appearance of a new $-OCH_3$ at δ4.28 and a phenol –OH at δ9.25. Consistent with the new asymmetry in pre-cercosporin, two strongly hydrogen-bonded *peri*-hydroxy groups could be seen far downfield at *ca*. 15 ppm and two aryl hydrogens were observed at δ6.92 and δ6.87. That these latter resonances are observed only in pairs as are the two side chain methyl doublets at *ca*. 0.6 ppm and the doubling of other signals imply that pre-cercosporin is formed as a single atropisomer having a helical configuration likely identical to that of cercosporin, although it is conceivable CTB9 also sets the final stereochemistry.

_Cercospora beticola CBET3_00841_ (_CbCFP_) is required for auto-resistance to cercosporin

To date, no *CTB* cluster gene has been shown to be involved with cercosporin auto-resistance[10]. However, targeted disruption of *C. kikuchii CFP* conferred enhanced sensitivity to exogenously-applied cercosporin[23]. To confirm if the *C. beticola* CFP homolog *Cb*CFP also mediates resistance to cercosporin, we compared the growth of the WT versus Δ*CbCFP* and ectopic mutants in cercosporin-amended liquid media assays (Supplementary Fig. 8). In the absence of cercosporin, Δ*CbCFP* exhibited enhanced growth compared to the wild-type, and to a lesser extent at 1.0 µM cercosporin. However, individual knock-out mutants of Δ*CbCFP* grew to ~40% of wild-type in 100 µM cercosporin. Complemented mutants exhibited enhanced growth compared to their progenitor knock-out strains in 100 µM cercosporin. All assays were

repeated with multiple Δ*CbCFP* mutant strains, producing similar results. Thus, the *CTB* cluster gene *CbCFP* is required for cercosporin tolerance in *C. beticola*.

Discussion

Several hypotheses exist for the maintenance of SM biosynthetic genes as clusters. In one, unlinked SM pathway genes are at a greater risk for dissociation during meiotic recombination[30] or chromosomal rearrangements[31]. Additionally, clustering may facilitate strict coordination of gene expression, which may be particularly important during the biosynthesis of SMs that have potentially toxic intermediates[32]. Besides the maintenance of genome and cellular integrity, clustering may promote or be a consequence of horizontal transfer. Horizontal gene transfer of entire biosynthetic gene clusters has been reported[33-36], but not to the extent and frequency observed in this study. The unprecedented number of horizontal transfers of the ancient *CTB* cluster specifically among plant pathogens suggests that it was critical for disease development in diverse pathosystems. The *CTB*-like clusters in *C. higginsianum* and *C. graminicola* were reported as one of the few SM clusters between these species that are microsyntenic[22]. Moreover, O'Connell detected specific upregulation of the *CTB*-like cluster in *C. higginsianum* during colonization of Arabidopsis[22]. Indeed, nine of 14 *C. higginsianum CTB*-like genes were among the top 100 most highly expressed genes *in planta*. Recent analysis of natural selection processes in *C. graminicola* identified orthologs of *CTB* genes *CTB1* and *CFP* as one of ~80 genes undergoing significant positive selection[37], further suggesting a role in pathogenicity. We were not able to detect cercosporin production in *C. acutatum* or *C. fioriniae in vitro*, and thus future research is directed to examine *in planta* biosynthesis as well as investigate the potential role of this cluster in *Colletotrichum* pathogenicity.

The extensive microsynteny outside of the established *CTB* cluster prompted us to test whether the flanking genes in *C. beticola* are also required for cercosporin biosynthesis. Notably, we observed that these flanking genes, similar to the established *CTB* genes, were up-regulated under cercosporin-inducing conditions. Furthermore, targeted gene replacement of *CbCTB10* and *CbCTB11* completely abolished

cercosporin biosynthesis while replacement of *CbCTB9* resulted in the accumulation of a new, red metabolite. We thus conclude that the *CTB* cluster is significantly larger than previously described[10]. Our phylogenomics approach further revealed the presence of a duplicated, *CTB*-like cluster encompassing a similar aggregation of genes. It is noteworthy that the presence of this duplicated cluster in *C. beticola* and *C. berteroae*, and its absence in *C. zeae-maydis* (JGI MycoCosm), coincides with the ability to produce beticolin, another non-host specific toxin which causes plant cell membrane disruption[38,39]. Future research will be directed to determine the function of this gene cluster in the *C. beticola*/sugar beet interaction.

The isolation and characterization of a new intermediate in the cercosporin biosynthetic pathway, pre-cercosporin, strongly suggests that formation of the unique 7-membered methylenedioxy bridge in the final product is the result of a two-step process. First, one of two precursor aryl methoxyl groups is oxidatively removed (possibly by CTB7, a flavin-dependent oxidoreductase), followed by the action of CTB9, an apparent α-ketoglutarate-dependent dioxygenase annotated as similar to GA4 desaturase. In contrast, a single cytochrome P450 is known to convert two aryl *ortho*-methoxyl groups into the relatively more common 5-membered methylenedioxy group in alkaloid biosynthesis[40]. A tentative cercosporin biosynthesis scheme was recently proposed[12] without knowledge of the expanded *CTB* cluster. However, in light of the new CTB9 intermediate and the potential functions of the other newly discovered *CTB* genes, the previously reported biosynthesis pathway[12] will have to be revised. These investigations will be reported in due course.

Considering the universal toxicity of perylenequinones and the similar involvement of photo-activated perylenequinones in disease development by several plant pathogens, research on cercosporin auto-resistance mechanisms may serve as a universal model towards better understanding of cellular resistance to oxidative stress[41]. Moreover, the incorporation of cercosporin auto-resistance genes into crop germplasm may provide durable resistance to *Cercospora* diseases in several important crop hosts

and reduce the reliance on fungicides for disease management. Unlike resistance proteins that recognize pathogen protein effectors that can evade recognition with as little as a single amino acid exchange[42], resistance to SMs is likely to be more durable because pathogen populations cannot easily alter SM structure to avoid recognition and cercosporin mutants of most *Cercospora* spp. are significantly less virulent[23,43-45]. Consequently, the search for SM auto-resistance genes within SM clusters has significant potential to lead to new developments for engineering disease resistance in crop plants against diverse fungal pathogens.

Methods

Data access

The *C. beticola* genome sequence, assembly and annotation as well as transcriptome reads are deposited at the National Center for Biotechnology Information (NCBI) SRA as BioProject PRJNA270309. NCBI Locus Tag Prefix *CB0940* replaces the *CBET3* Locus Tag Prefix that is used throughout this manuscript. A browser of this genome assembly is available at http://bioinformatics.psb.ugent.be/orcae/overview/Cerbe/. Custom Perl and R codes are available through a public GitHub repository: https://github.com/rdejonge/genomics, and all supplementary files are available through *figshare*; doi: 10.6084/m9.figshare.4056522.

Cercospora spp. genome sequencing

Genomic DNA of *C. beticola* strain 09-40 was isolated using the CTAB method from mycelia scraped from the surface of V8 juice agar Petri plates[46]. Library preparation of three genomic libraries with increasing insert size (500 bp, 5 Kbp and 10 Kbp) and subsequent paired-end (PE) and mate-pair (MP) genome sequencing was performed by BGI Americas Corporation (BGI) via the Illumina platform. A total of ~11,100,000 high-quality filtered sequence reads with an average length of 100 bp were generated for the 500 bp PE library, ~23,500,000 reads of 50 bp length were derived from the 5 Kbp MP library, and a total of ~16,800,000 reads of 90 bp length were obtained from the 10 Kbp MP library. Total sequence output was ~3.8 Gbp, corresponding to an estimated 100-fold coverage. In the first stage of the genome assembly, SOAPdenovo (version 2.223) was used with the following command: "*SOAPdenovo-63mer all —s config.txt —K 51 —R —o soapfinal.fa —p 8 —L 200*" to assemble contigs and scaffolds incorporating all three libraries as input. The libraries were ranked according to insert size, using the smallest insert library as ranked first. Only the smallest insert library was used for all stages of the assembly with option "*asm_flags=3*", the other two libraries were only used for scaffold assembly with option "*asm_flags=2*".

For the second stage of the genome assembly, optical mapping was used to scaffold sequence reads. Optical maps were generated on the Argus System by BGI and sequences were placed using MapSolver (version 3.2.2; Opgen). BioNano Genomics optical maps (genome maps) were generated on the Irys System (version 9; BioNano Genomics) at the Nucleomics Core facility (Vlaams Instituut voor Biotechnologie), and sequences were placed using the IrysView software. Recommendations by the software packages were followed for scaffold placement. The resulting alignment maps were subsequently compared visually, and accordingly an AGP-like (A Golden Path) file was constructed manually, applying a majority-consensus ruling wherever necessary that details the placement (position and orientation) of all sequences (Supplementary File 1). Using the custom Perl script *parse_agp-like_to_fasta.pl*, the AGP-like file was used to construct 12 supercontigs (possibly chromosomes) ranging in size from ~227 Kbp to ~6.2 Mbp. We then applied Pilon (version 1.7; http://www.broadinstitute.org/software/pilon) for automated assembly improvement. To this end, the sequencing reads from all three libraries (MP libraries were reverse complemented prior to mapping) were aligned to the genome assembly using Bowtie[47] (version 2.1.0) implementing default parameters for end-to-end mapping (i.e. "*—end-to-end*, *--sensitive*", with the exception of insert length, that was adjusted accordingly for each library). Pilon was run subsequently with default parameters, incorporating the mapped PE library as "frags" and the mapped MP libraries as "jumps."

Repetitive sequences were identified by RepeatMasker (version open-4.0.3; http://www.repeatmasker.com) using two distinct libraries: 1) the RepBase repeat library (http://www.girinst.org/server/RepBase/index.php, obtained on October 23, 2013) and 2) a *C. beticola* specific repeat library constructed using the RepeatModeler (version 1.0.7; http://www.repeatmasker.org/RepeatModeler.html). A total of 45 repeat families were found by RepeatModeler, of which 13 could be classified (Supplementary File 2).

Cercospora berteroae strain CBS 538.71 was obtained from Centraal Bureau voor Schimmelcultures (CBS) and cultivated on Petri plates containing potato dextrose agar (PDA; Difco). High-quality DNA was extracted using the CTAB method[48]. Library preparation (500 bp) and subsequent paired-end (PE) genome sequencing was performed by BGI via the Illumina platform. A total of 31 million high-quality filtered sequence reads with an average length of 100 bp were generated. A draft genome assembly was constructed using SOAPdenovo (version 2.04), applying default parameters and K-mer length 51.

Transcriptome sequencing

To aid in gene prediction and discover gene expression patterns under specific conditions, three RNA samples were prepared for RNA sequencing; one from *in vitro* cultured *C. beticola* tissue and two from *C. beticola* infected sugar beet seedlings at four and seven days post inoculation (DPI). Total RNA was extracted from flash-frozen fungal and/or plant material using a Qiagen total RNA extraction kit as described[49]. Library preparation and sequencing were performed at BGI, which resulted in the generation of ~12,800,000 44,500,000 and 44,800,000 PE reads of length 90 bp and fragment size of 200 bp for the *in vitro*, the 4DPI *in planta* and the 7 DPI *in planta* libraries respectively.

Genome-guided Trinity assembly[50] of sequencing reads was subsequently performed as described by the software manual, with minimal intron length of 20 bp and the GSNAP aligner. Transcript assemblies were subsequently generated using PASA2 (version r20140417). We applied another transcript assembly approach simultaneously by mapping the RNA-Seq reads to the reference genome using Tophat (version 2.0.8b; http://ccb.jhu.edu/software/tophat/index.shtml), converting these to raw transcript assemblies using Cufflinks (version 2.1.1; http://cufflinks.cbcb.umd.edu/) and final selection of the best assemblies using TransDecoder (version 2014-07-04; http://transdecoder.sourceforge.net/).

Gene prediction and curation

EvidenceModeler[51] was used to predict protein-coding genes by integrating protein-coding evidence from multiple sources according to specified weights. First, Augustus[52] parameters (version 2.5.5) were trained locally with the *autoAug.pl* script included in the Augustus software package, using the assembled Trinity transcripts sequences as training input as well as hints in the final gene prediction step. Secondly, GeneMarkES[53] (version 2.3c) was used in self training modus, optimizing coding parameters for the full genome sequence. Protein similarity was detected by the alignment of proteins from the related fungal species *Z. tritici*, *P. nodorum*, *L. maculans*, *C. heterostrophus* and *C. zeae-maydis* against the genome sequence using the Analysis and Annotation Tool (AAT) package[54] (version r03052011). Protein sequences from these related fungi were derived from the MycoCosm platform at the Joint Genome Institute (JGI; http://genome.jgi-psf.org/programs/fungi/index.jsf). For integration of the various protein-coding sources with the EvidenceModeler, we ranked all sources according to their expected accuracy and importance. From high to low the ranking was: genome-aligned Trinity transcripts (with PASA), Cufflinks-derived TransDecoder transcripts, trained Augustus *ab initio* predictions, aligned protein sequences and self-trained GeneMarkES *ab initio* predictions. These initial EvidenceModeler transcripts were substantially further improved by manual curation on the WebApollo[55] (version 2013-05-16) platform. Approximately 10,500 gene models were visually analyzed and updated according to the available evidence as needed. The updates included adjustments to splice sites (e.g. alternative donor or acceptor sites), exon usage, five prime and three prime UTRs (untranslated regions) and the addition of alternative transcripts where supported by RNA-Seq evidence. A majority-consensus ruling was used for the updates, considering all evidence tracks. For transfer of the final gene predictions to newer versions of the assembly, the map2assembly algorithm from Maker was used routinely (version 2.28; http://www.yandell-lab.org/software/maker.html).

To investigate the predicted gene models in more detail, we calculated a number of statistics based on the gene structures. For this purpose, all gene models were stored in GFF3 format according to

the specifications at The Sequence Ontology Project (http://www.sequenceontology.org/gff3.shtml). The *gff3_gene_prediction_file_validator.pl* Perl script, part of the EvidenceModeler package was routinely used for verification of GFF3 file format. GFF3 gene statistics, e.g. the mean/median gene length, the number of exons, CDSs or UTRs, were calculated using the custom Perl script, *GFF3stats.pl*.

Protein function characterization

For functional characterization of the predicted protein sequences, hardware-accelerated BLASTp on a DeCypher machine (TimeLogic; Carlsbad, USA) was used to identify homologous proteins in the non-redundant (nr) protein database obtained at the NCBI. InterProScan (version 5.44; http://www.ebi.ac.uk/Tools/pfa/iprscan5/) was used to identify conserved protein domains. The results of both analyses where imported into Blast2GO[56] and used to generate single, uniquely functional annotations for each protein as well as a list of all associated gene ontology (GO) terms.

Secondary metabolite cluster identification, characterization and visualization

SM clusters were identified in the genome sequence of *C. beticola* and that of related fungi using antiSMASH2[15] (version 2.1.0; https://bitbucket.org/antismash/antismash2/). To generate antiSMASH2-required EMBL formatted genome files, the GFF3 gene features files in combination with the respective genome sequences were converted to the EMBL sequence format using the custom Perl script *GFF3_2_EMBL.pl*. Subsequently, antiSMASH2 was run with default parameters, allowing for the identification of PKS SM clusters, NRPS SM clusters, Hybrid PKS-NRPS SM clusters, terpene cyclase (TC) SM clusters, siderophores SM clusters and lantipeptide SM clusters. SM clusters that showed similarity to a mixture of these clusters or only a minimal set of homologous protein domains were depicted as "other." In addition, DMAT, for dimethylallyl tryptophan synthase clusters were identified by screening the InterProScan results for Pfam domain PF11991.

Secondary metabolite phylogenetic analyses

For phylogenetic analyses of the type I polyketide enzymes we used Mafft (version 7.187), applying global alignment (--globalpair) and a 1000 cycles of iterative refinement (--maxiterate 1000), to align full-length sequences as well as selected domains of all PKS enzymes that were identified by antismash2 in the genome sequences of the six Dothideomycetes: *C. beticola*, *D. septosporum*, *Z. tritici*, *L. maculans*, *P. tritici-repentis* and *P. nodorum*, and one Eurotiomycete: *Aspergillus nidulans*. In addition, previously characterized polyketide synthases (Supplementary Table 3) were included for reference. Prior to phylogenetic tree reconstruction, the alignments were trimmed with TrimAl[57] (version 1.2). Maximum likelihood phylogenetic trees were determined with RaxML (version 8.1.3), applying rapid bootstrapping (-# 100) and automated protein model selection (-m PROTGAMMAAUTO). Final trees were prepared online using EvolView[58]. Species tree topologies were built with Cvtree[59] webserver by uploading the predicted proteomes of 48 published Ascomycete fungi (Supplementary Table 4).

For phylogenetic tree reconciliation analyses of the protein and species trees, the protein trees were pre-processed with treefixDTL[60] (version 1.1.10) to minimize errors introduced during tree reconstruction. TreefixDTL is able to correct phylogenetic trees in the presence of horizontal gene transfer. Reconciliation analyses as well as rooting were conducted in NOTUNG[19] according to the instructions (version 2.8 beta).

Secondary metabolite cluster alignment visualization

For comparative analyses of the secondary metabolite clusters, we used the R-package *genoPlotR*[61] (version 0.8.2; http://genoplotr.r-forge.r-project.org/). To this end, individual clusters were extracted from the genome sequence using the Perl script *get_seq_by_id.pl*, using the start or stop positions of the flanking genes in the cluster as extremities. The resulting sequences were then aligned using the bl2seq (BLAST two sequences) algorithm, part of the BLAST toolkit (version 2.2.26). Also, transcripts locations (CDS start and stop coordinates specifically) were extracted from the GFF3 files using *gff3CDS_2_genoplot.pl*, and adjusted according to the start and end coordinates of the extracted cluster

sequence, by means of the Unix command line tool Awk. We then used custom R code (example in *genoPlotR.R*) to parse these input files and generate cluster alignment figures.

Genome-wide gene cluster microsynteny and protein identity analysis

Genome-wide gene-by-gene cluster analyses were performed using the custom Perl script *calClusterSimilarity.pl*, and plotted using ggplot2 in R using *synteny.R*. As input, this pipeline takes the typical output of an orthoMCL analysis, reformatted by *analyseOrthoMCL.pl*. In short, it requires each proteinId to have an associated clusterId. Furthermore, it requires properly formatted GFF3 files for each genome that are used to associate location of protein-coding genes and their flanks. Last but not least, the number of flanking genes to be used can be chosen freely, but must be set ODD. For the analyses presented in Supplementary Figure 5, a cluster size of 30 was set. Genome-wide protein-by-protein best-BLAST percent identities were derived from the similarities table prepared during orthoMCL analyses and subsequently plotted in R using *pairwise_pident_boxplots.R*.

Gene expression analysis

To investigate the expression of cercosporin cluster genes, *C. beticola* was grown in a 250 ml Erlenmeyer flask containing 100 ml potato dextrose broth (PDB; Difco) in the light and dark, which are conditions that promote and repress cercosporin production, respectively. Total RNA was isolated using TRIzol (ThermoFisher) following the manufacturer's instructions followed by an on-column Dnase treatment (Qiagen). Total RNA was used for cDNA synthesis using an oligo-(dT) primer and the SuperScript II reverse transcriptase kit (Invitrogen) following manufacturer's instructions. The resulting cDNA was used as a template for quantitative polymerase chain reaction (qPCR). Selected genes were queried for expression using the Power SYBR Green PCR Master Mix (Applied Biosytems) using a PTC-2000 thermal cycler (MJ Research) outfitted with a Chromo4 Real-Time PCR Detector (Bio-Rad) and MJ Opticon Monitor analysis software version 3.1 (Bio-Rad). Primers for gene expression are listed in Supplementary Table 9.

Transformation and disruption of target genes

Split-marker PCR constructs for targeted gene replacement were prepared as described[46] using genomic DNA of 10-73-4 and 09-40 wild-type *C. beticola* and pDAN as PCR templates. Selected mutants were complemented using pFBT005, which encodes resistance to nourseothricin and allowed us to clone our gene of interest between the ToxA promoter and TrpC terminator using PacI and NotI (Promega) restriction sites. Primers for split-marker and complementation constructs are listed in Supplementary Table 9.

A 5 mm plug was taken from the actively growing zone of *C. beticola* on PDA. Liquid cultures were initiated by grinding the plug with 500 µL Fries media[46], which was subsequently transferred to a 125 mL flask containing 50 ml Fries media. Flasks were wrapped in aluminum foil and shaken at 150 rpm at 21 °C for four days. Cultures were then transferred to a sterile blender cup, ground for 10 s, and transferred to a 500 mL flask containing 200 mL Fries media. Cultures were grown as described above for an additional 24 h after which mycelium was harvested with two layers of Miracloth (Calbiochem) using a Büchner funnel. The mycelium was rinsed with mycelium wash solution (0.7M KCl, 10mM $CaCl_2 \cdot 2H_2O$), broken into small pieces with a sterile spatula, transferred to a deep Petri dish containing 40 mL of osmoticum (20 g L^{-1} lysing enzymes from *T. harzianum* (Sigma-Aldrich) and 12.5 g L^{-1} Driselase (Sigma-Aldrich), and incubated at 30 °C for 6 to 14 h while shaking at 50 rpm. Protoplasts were harvested by filtering the osmoticum solution through two layers of Miracloth and collected by centrifugation (2000 x *g* for 5 min). After a washing step with 15 mL STC (1 M Sorbitol, 10 mM Tris-HCl pH 7.5, 10 mM $CaCl_2$), protoplasts were adjusted to 10^8 mL^{-1} with STC:PEG (4:1) and divided into 200-µL aliquots. Subsequently, PEG-based transformation was performed essentially as described[46] using 70 µg of each construct or 70 µg of complementation plasmid linearized with ApaI (Promega) per 200 µL protoplast aliquot.

Cercosporin production

To screen for the ability to produce cercosporin, 10 mm plugs were transferred from the growing edge of *C. beticola* 10-73-4 wild-type, mutant and complemented mutant colonies to 100 mm x 15 mm Petri plates (Falcon; Oxnard, USA) that contained 10 ml PDA. After an incubation period of 6 days at 21 °C with a natural light-dark cycle, five 10 mm (dia) plugs of PDA containing *C. beticola* mycelium were collected and extracted with ethyl acetate for 18 h in the dark at 4 °C. The extracts were filtered and dried under a stream of nitrogen (30 °C). The dried extracts were re-dissolved in 25% acetonitrile in 1% aqueous acetic acid and fractionated by reverse-phase HPLC on a Waters 600 HPLC system fitted with a Waters radial pak column (8 mm x 120 mm, 8 µm). Injections were 200 µl. Solvent A was 1% (v/v) aqueous acetic acid and solvent B was acetonitrile. Starting conditions were 25% A 75% B, hold for 1 min, a linear gradient to 100% B in 19 min, hold for 5 min, and return to initial conditions in 5 min (Fig. 4b). Under these conditions, cercosporin had a retention time of ca. 15.4 min. Where needed, extracts were also subjected to thin-layer chromatography (TLC) on Silica HF Plates pre-coated with calcium phosphate and reactivated as described Balis and Payne[62]. Plates were developed in hexane: 2-propanol (8:2, v/v) and visualized by UV fluorescence. For cercosporin isolation, multiple mycelial mats were collected and extracted two times with ethyl acetate (4 °C, dark). After filtration, the extracts were partitioned against an equal volume of reagent-grade water, residual water was removed by passage through anhydrous sodium sulfate, and the extracts were dried under a stream of nitrogen (30 °C). Cercosporin was purified by TLC as described above and quantified by spectrophotometry using published extinction coefficients[63].

Pre-cercosporin isolation and characterization

Mycelial plugs of *C. beticola ΔCTB9* were placed on top of eight "thin" PDA (Difco) plates (3.0 mL PDA per 50 mm Petri plate). Cultures were incubated at 22 °C for one week under continuous light. Three separate methods were attempted to prepare crude secondary metabolite extractions. 1) PDA and mycelia were extracted with ethyl acetate for 4 min. The resulting supernatant was collected and frozen for further analysis. 2) PDA and mycelia were placed into a GenElute Maxiprep binding column (Sigma Aldrich) and

centrifuged at 3500 x *g* for 10 min. The flow through was collected and frozen for further analysis. 3) PDA and mycelia were placed into a GenElute Maxiprep binding column along with 15 mL ethyl acetate. After a 30 s incubation, the column was centrifuged at 2400 x *g* for 10 min. The flow through was collected and frozen for further analysis. To obtain sufficient pre-cercosporin for isolation and NMR analysis, primary extracts from all three methods were combined.

The combined extracts were re-suspended with water and acidified with conc. HCl. Pre-cercosporin was extracted quickly from this aqueous solution by partitioning thrice with ethyl acetate in dark, wrapping the glassware with aluminum foil. The combined ethyl acetate fractions were washed with brine, dried over anhydrous sodium sulfate and evaporated under vacuum at 30 °C. The reddish-brown residue was resuspended in methanol and filtered through 0.2 µm PTFE filters. The methanol extracts were initially analyzed by reverse phase HPLC (Fig. 5a) on an Agilent model 1200 fitted with a Kinetex XB-C18 column (4.6 mm x 75 mm, 2.6 µm, Phenomenex). Injections of 1 µl were run at 1.25 mL/min on a linear gradient of 5% solvent C/95% solvent D to 95% solvent C/5% solvent D over 10.8 min, where solvent C is water + 0.1% formic acid and solvent D is acetonitrile + 0.1% formic acid. Chromatograms were monitored at 436, 280, and 210 nm, and UV-vis spectra were recorded over a range of 210-800 nm (Fig. 5b). High-resolution mass data were obtained from a Waters Acquity/Xevo-G2 UPLC-ESI-MS in positive ion mode.

To isolate sufficient pre-cercosporin for ^{1}H-NMR analysis, the filtered methanol extract prepared above was purified by reverse-phase HPLC on an Agilent model 1100 fitted with a Kinetex XB-C18 semi-prep column (10 mm x 250 mm, 5 µm, Phenomenex). The crude extract (10 mg/mL in methanol) was injected (generally 500 µl) and run at 4 mL/min using the following method: 20% solvent C/80% solvent D for 3 min, 20-70% solvent C over 17 min, 70-95% solvent D over 5 min, where solvent C and D were as above. Chromatograms were recorded at 436, 280, and 210 nm. The metabolite of interest was collected

from multiple injections, combined, and lyophilized to dryness. The purified pre-cercosporin was analyzed by UPLC-ESI-MS as described above and ^{1}H-NMR.

4,6,9-trihydroxy-1,12-bis(2-hydroxypropyl)-2,7,11-trimethoxyperylene-3,10-dione (pre-cercosporin)

Data obtained at 5 °C on a Bruker AVANCE spectrometer. ^{1}H NMR (400 MHz, $CDCl_3$, δ): 15.24 (s, 1H), 14.93 (s, 1H), 9.25 (s, 1H), 6.92 (s, 1H), 6.87 (s, 1H), 4.28 (s, 3H), 4.19, (s, 3H), 4.18 (s, 3H), 3.57 – 3.51 (m, 2H), 3.42 – 3.36 (sym 5-line overlapping signal, 2H), 2.86 – 2.74 (m, 2H), 0.64 (d, *J* = 6.1 Hz, 3H), 0.60 (d, *J* = 6.1 Hz, 3H). UPLC-ESI-HRMS: calculated for $C_{29}H_{29}O_{10}$ [M+H$^+$]: 537.1761, found [M+H$^+$]: 537.1762. ^{1}H-NMR spectra (400 MHz in $CDCl_3$) of pre-cercosporin and cercosporin for comparison are available as Supplementary Figures 9 and 10.

Cercosporin auto-resistance

To identify mutants that were sensitive to cercosporin, approximately 10^3 spores collected from selected strains were transferred to individual wells of a Nunc 96 well flat-bottom plate (Fisher Scientific) using Fries media as the growth medium and sealed with SealPlate (Excel Scientific). Within four hours, each well was amended with cercosporin extracted from *C. beticola* strain 10-73-4 as described above to a final concentration of 0.0, 1.0, 10.0, or 100.0 µM. Cultures were incubated at 21°C under 24 hour light conditions. Growth was quantified daily by collecting absorbance values from each well generated from a GENios plate reader (Tecan) using XLUOR4 software (V 4.51) using a measurement wavelength of 595 nm and 2x2 reads per well after vortexing the plate briefly. The absorbance value and associated standard deviation and error for each strain was determined by taking the average of 16 wells (replications) per timepoint. Each experimental run consisted of five timepoints, which were taken every 24 h starting with day 0, which were the measurements taken immediately after cercosporin treatments were added to each well, up to and including day 4. Fungal growth was calculated by subtracting the absorbance value

of day 0 from each timepoint. Each run was repeated at least three times with each of three individual knock-out mutant strains.

Colletotrichum spp. cercosporin assay

To determine whether *Colletotrichum* species were able to produce cercosporin, seven monoconidial isolates of *C. acutatum* (isolates CA1, HC50, HC72, HC 75, HC81, HC89, and HC91) and *one C. fiorinae* (3HN12) were grown on 9 cm Petri plates containing 15 ml of PDA as described above to replicate conditions that were conducive for cercosporin production *in vitro*. Seven day old cultures of each isolate were grown in a temperature controlled incubator at 25 °C with natural light. A pinkish to dark red color was visible in the media for all isolates except HC75 which had a yellow-colored pigment. Using a #2 cork borer, three plugs were removed from each isolate from the edge, middle and center of each colony and placed in small screw cap glass vials. Three plugs were also removed from an uncolonized PDA plate and included as a negative control. Cercosporin (Sigma-Aldrich) was dissolved in acetone to 100 mM and used as a positive control. There were 10 samples, seven total for each of the *Colletotrichum spp*. isolates, one containing cercosporin toxin as a positive control, and one blank plug that served as a negative control. 5N KOH was added to each vial to cover the surface of the plugs and incubated on a shaking incubator at room temperature for 4 hours. Supernatants were examined for cercosporin spectrophotometrically.

Acknowledgements This work was supported by USDA-ARS project 3060-22000-047-00D. R. J. was supported by an EMBO Long Term fellowship and is currently supported by a FWO postdoctoral fellowship. C.R.H.-R., P.P and C.A.T. were supported by NIH grant RO1 ES001670. We thank W. Underwood and T. L. Friesen (USDA – ARS) for review of the manuscript, A. G. Newman for helpful discussions and a reference sample of cercosporin and N. Metz (USDA – ARS) for excellent technical assistance. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Author Contributions Experimental work was conducted by M.K.E., C.R.H.-R., P.P, J.C.S., and W.M.J. Bioinformatics was conducted by R.d.J. All experiments were overseen by R.d.J., G.A.S., B.T., Y.P., C.A.T. and M.D.B. The manuscript was written by R.d.J., M.D.B. and C.A.T. with help from other authors.

References

1. Fuckel, K. Fungi Rhenani exsiccati, Fasc. I-IV. *Hedwigia* **2**, 132-136 (1863).
2. Stergiopoulos, I., Collemare, J., Mehrabi, R. & De Wit, P. J. G. M. Phytotoxic secondary metabolites and peptides produced by plant pathogenic *Dothideomycete* fungi. *FEMS Microbiol. Rev.* **37**, 67-93 (2013).
3. Goodwin, S. B. & Dunkle, L. D. in *Cercospora leaf spot of sugar beet and related species* (eds Robert T. Lartey *et al.*) Ch. 9, 97-108 (The American Phytopathological Society, 2010).
4. Daub, M. E. & Ehrenshaft, M. The photoactivated *Cercospora* toxin cercosporin: contributions to plant disease and fundamental biology. *Annu. Rev. Phytopathol.* **38**, 461-490 (2000).
5. Daub, M. E. & Hangarter, R. P. Light-induced production of singlet oxygen and superoxide by the fungal toxin, cercosporin. *Plant Physiol.* **73**, 855-857 (1983).
6. Dobrowolski, D. C. & Foote, C. S. Cercosporin, a singlet oxygen generator. *Angew. Chem., Int. Ed. Engl.* **22**, 720-721 (1983).
7. Steinkamp, M. P., Martin, S. S., Hoefert, L. L. & Ruppel, E. G. Ultrastructure of lesions produced in leaves of *Beta vulgaris* by cercosporin, a toxin from *Cercospora beticola*. *Phytopathol.* **71**, 1272-1281 (1981).
8. Daub, M. E. Destruction of tobacco cell-membranes by the photosensitizing toxin, cercosporin. *Phytopathol.* **71**, 869-869 (1981).
9. Daub, M. E. Resistance of fungi to the photosensitizing toxin, cercosporin. *Phytopathol.* **77**, 1515-1520 (1987).
10. Chen, H. Q., Lee, M. H., Daub, M. E. & Chung, K. R. Molecular analysis of the cercosporin biosynthetic gene cluster in *Cercospora nicotianae*. *Mol. Microbiol.* **64**, 755-770 (2007).
11. Newman, A. G., Vagstad, A. L., Belecki, K., Scheerer, J. R. & Townsend, C. A. Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem. Commun.* **48**, 11772-11774 (2012).
12. Newman, A. G. & Townsend, C. A. Molecular characterization of the cercosporin biosynthetic pathway in the fungal plant pathogen *Cercospora nicotianae*. *J. Am. Chem. Soc.* **138**, 4219-4228 (2016).
13. Choquer, M., Lee, M. H., Bau, H. J. & Chung, K. R. Deletion of a MFS transporter-like gene in *Cercospora nicotianae* reduces cercosporin toxin accumulation and fungal virulence. *FEBS Lett.* **581**, 489-494 (2007).
14. Chand, R. *et al.* Draft genome sequence of *Cercospora canescens*: a leaf spot causing pathogen. *Curr. Sci.* **109**, 2103-2110 (2015).
15. Blin, K. *et al.* antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **41**, W204-W212 (2013).
16. Collemare, J. *et al.* Secondary metabolism and biotrophic lifestyle in the tomato pathogen *Cladosporium fulvum*. *PLoS ONE* **9**, e85877 (2014).
17. Medema, M. H., Cimermancic, P., Sali, A., Takano, E. & Fischbach, M. A. A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLoS Computational Biology* **10**, e1004016 (2014).
18. Chooi, Y.-H., Muria-Gonzalez, M. J. & Solomon, P. S. A genome-wide survey of the secondary metabolite biosynthesis genes in the wheat pathogen *Parastagonospora nodorum*. *Mycology* **5**, 192-206 (2014).
19. Stolzer, M. *et al.* Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* **28**, i409-i415 (2012).

20 Wisecaver, J. H. & Rokas, A. Fungal metabolic gene clusters—caravans traveling across genomes and environments. *Frontiers in Microbiology* **6** (2015).

21 Lardner, R., Johnston, P. R., Plummer, K. M. & Pearson, M. N. Morphological and molecular analysis of *Colletotrichum acutatum sensu lato*. *Mycol. Res.* **103**, 275-285 (1999).

22 O'Connell, R. J. *et al.* Lifestyle transitions in plant pathogenic *Colletotrichum fungi* deciphered by genome and transcriptome analyses. *Nat. Genet.* **44**, 1060-1065 (2012).

23 Callahan, T. M., Rose, M. S., Meade, M. J., Ehrenshaft, M. & Upchurch, R. G. CFP, the putative cercosporin transporter of *Cercospora kikuchii*, is required for wild type cercosporin production, resistance, and virulence on soybean. *Mol. Plant-Microbe Interact.* **12**, 901-910 (1999).

24 Tudzynski, B. *et al.* Characterization of the final two genes of the gibberellin biosynthesis gene cluster of *Gibberella fujikuroi*: *des* and *P450-3* encode GA4 desaturase and the 13-hydroxylase, respectively. *J. Biol. Chem.* **278**, 28635-28643 (2003).

25 Kim, J.-E. *et al.* Putative polyketide synthase and laccase genes for biosynthesis of aurofusarin in *Gibberella zeae*. *Appl. Environ. Microbiol.* **71**, 1701-1708 (2005).

26 Williams, J. S., Thomas, M. & Clarke, D. J. The gene *stlA* encodes a phenylalanine ammonia-lyase that is involved in the production of a stilbene antibiotic in *Photorhabdus luminescens* TT01. *Microbiol.* **151**, 2543-2550 (2005).

27 Choquer, M., Lee, M.-H., Bau, H.-J. & Chung, K.-R. Deletion of a MFS transporter-like gene in *Cercospora nicotianae* reduces cercosporin toxin accumulation and fungal virulence. *FEBS Lett.* **581**, 489-494 (2007).

28 Frandsen, R. J. N. *et al.* Two novel classes of enzymes are required for the biosynthesis of aurofusarin in *Fusarium graminearum*. *J. Biol. Chem.* **286**, 10419-10428 (2011).

29 Gao, Q. *et al.* Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genet.* **7**, e1001264 (2011).

30 Galazka, J. M. & Freitag, M. Variability of chromosome structure in pathogenic fungi - of 'ends and odds'. *Curr. Opin. Microbiol.* **20**, 19-26 (2014).

31 de Jonge, R. *et al.* Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res.* **23**, 1271-1282 (2013).

32 McGary, K. L., Slot, J. C. & Rokas, A. Physical linkage of metabolic genes in fungi is an adaptation against the accumulation of toxic intermediate compounds. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11481-11486 (2013).

33 Slot, J. C. & Rokas, A. Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. *Curr. Biol.* **21**, 134-139 (2011).

34 Khaldi, N., Collemare, J., Lebrun, M.-H. & Wolfe, K. H. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome Biol.* **9**, R18 (2008).

35 Slot, J. C. & Hibbett, D. S. Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. *PLoS ONE* **2**, e1097 (2007).

36 Dhillon, B. *et al.* Horizontal gene transfer and gene dosage drives adaptation to wood colonization in a tree pathogen. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3451-3456 (2015).

37 Rech, G. E., Sanz-Martín, J. M., Anisimova, M., Sukno, S. A. & Thon, M. R. Natural selection on coding and noncoding DNA sequences is associated with virulence genes in a plant pathogenic fungus. *Genome Biol. Evol.* **6**, 2368-2379 (2014).

38 Assante, G., Locci, R., Camarda, L., Merlini, L. & Nasini, G. Screening of the genus *Cercospora* for secondary metabolites. *Phytochem.* **16**, 243-247 (1977).

39 Goudet, C., Milat, M.-L., Sentenac, H. & Thibaud, J.-B. Beticolins, nonpeptidic, polycyclic molecules produced by the phytopathogenic fungus *Cercospora beticola*, as a new family of ion channel-forming toxins. *Mol. Plant-Microbe Interact.* **13**, 203-209 (2000).

40 Díaz Chávez, M. L., Rolf, M., Gesell, A. & Kutchan, T. M. Characterization of two methylenedioxy bridge-forming cytochrome P450-dependent enzymes of alkaloid formation in the Mexican prickly poppy *Argemone mexicana*. *Arch. Biochem. Biophys.* **507**, 186-193 (2011).

41 Daub, M. E., Herrero, S. & Chung, K.-R. Reactive oxygen species in plant pathogenesis: the role of perylenequinone photosensitizers. *Antioxid. Redox Signal.* **19**, 970-989 (2013).

42 Joosten, M. H. A. J., Cozijnsen, T. J. & De Wit, P. J. G. M. Host resistance to a fungal tomato pathogen lost by a single base-pair change in an avirulence gene. *Nature* **367**, 384-386 (1994).

43 Shim, W. B. & Dunkle, L. D. Identification of genes expressed during cercosporin biosynthesis in *Cercospora zeae-maydis*. *Physiol. Mol. Plant Pathol.* **61**, 237-248 (2002).

44 Choquer, M. *et al.* The *CTB1* gene encoding a fungal polyketide synthase is required for cercosporin biosynthesis and fungal virulence of *Cercospora nicotianae*. *Mol. Plant-Microbe Interact.* **18**, 468-476 (2005).

45 Staerkel, C. *et al.* CbCTB2, an O-methyltransferase is essential for biosynthesis of the phytotoxin cercosporin and infection of sugar beet by *Cercospora beticola*. *BMC Plant Biol.* **13**, 50 (2013).

46 Bolton, M. D. *et al.* RNA-sequencing of *Cercospora beticola* DMI-sensitive and -resistant isolates after treatment with tetraconazole identifies common and contrasting pathway induction. *Fungal Genet. Biol.* **92**, 1-13 (2016).

47 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* **9**, 357-359 (2012).

48 Stewart, C. & Via, L. E. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *BioTechniques* **14**, 748-749 (1993).

49 Bolton, M. D. *et al.* Evaluation of the potential for sexual reproduction in field populations of *Cercospora beticola* from USA. *Fungal Biol.* **116**, 511-521 (2012).

50 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652 (2011).

51 Haas, B. *et al.* Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).

52 Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24**, 637-644 (2008).

53 Borodovsky, M. & Lomsadze, A. in *Current Protocols in Bioinformatics* (John Wiley & Sons, Inc., 2002).

54 Huang, X., Adams, M. D., Zhou, H. & Kerlavage, A. R. A tool for analyzing and annotating genomic sequences. *Genomics* **46**, 37-45 (1997).

55 Lee, E. *et al.* Web Apollo: a web-based genomic annotation editing platform. *Genome Biol.* **14**, R93 (2013).

56 Conesa, A. & Götz, S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* **2008**, 12 (2008).

57 Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).

58 Zhang, H., Gao, S., Lercher, M. J., Hu, S. & Chen, W.-H. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* **40**, W569-W572 (2012).

59 Qi, J., Luo, H. & Hao, B. CVTree: a phylogenetic tree reconstruction tool based on whole genomes. *Nucleic Acids Res.* **32**, W45-W47 (2004).

60 Bansal, M. S., Wu, Y.-C., Alm, E. J. & Kellis, M. Improved gene tree error correction in the presence of horizontal gene transfer. *Bioinformatics* **31**, 1211-1218 (2014).

61 Guy, L., Roat Kultima, J. & Andersson, S. G. E. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**, 2334-2335 (2010).

62 Balis, C. & Payne, M. G. Triglycerides and cercosporin from *Cercospora beticola*: fungal growth and cercosporin production. *Phytopathol.* **61**, 1477-1484 (1971).

63 Milat, M.-L. & Blein, J.-P. Cercospora beticola toxins III. Purification, thin-layer and high performance liquid chromatographic analyses. *J. Chromatogr. A* **699**, 277-283 (1995).

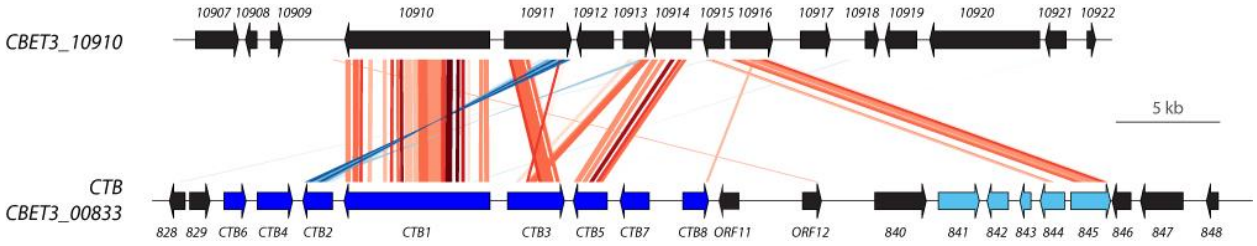

Figure 1: The cercosporin biosynthetic cluster is duplicated and maintained in *C. beticola*. *CBET3_10910* and flanking genes are syntenic with the *CTB* cluster (*CBET3_00833* and flanking genes) in *C. beticola*. Alignment lines correspond to DNA fragments exhibiting significant similarity when the genomic regions comprising the gene clusters are compared with tBLASTx. Direct hits are displayed in red, whereas complementary hits are in blue. The hue of the alignments represents the percentage similarity ranging from 23 to 100 percent.

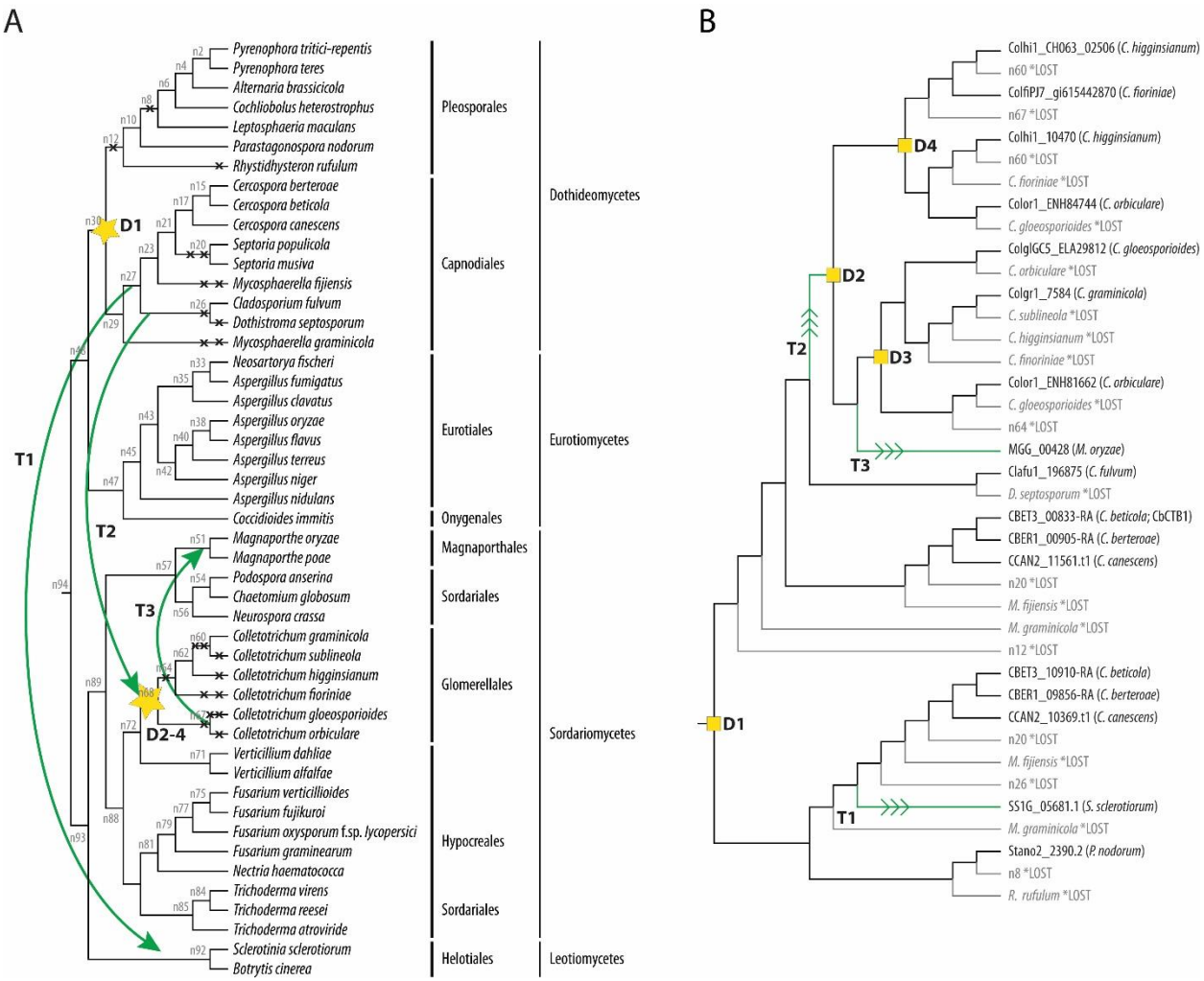

Figure 2: Phylogeny of *Cercospora* spp. and related Ascomycete fungi and reconciliation of CTB1 orthologs. (**a**) Cladogram showing the phylogenetic relationship of *Cercospora* spp. and 45 other sequenced fungi. The unscaled tree was constructed using CVTree[59]. (**b**) Tree reconciliation of CTB1. The phylogenetic tree of CTB1 determined by RaxML, corrected using the treefixDTL algorithm, and rooted and reconciled by Notung with the species tree in A to infer the number of duplications, losses and transfers. Duplication nodes are marked with yellow squares, losses are in grey and transfers are highlighted by green edges and arrow markings. Descendants of each duplication are colored according to the cross marks at the duplication nodes.

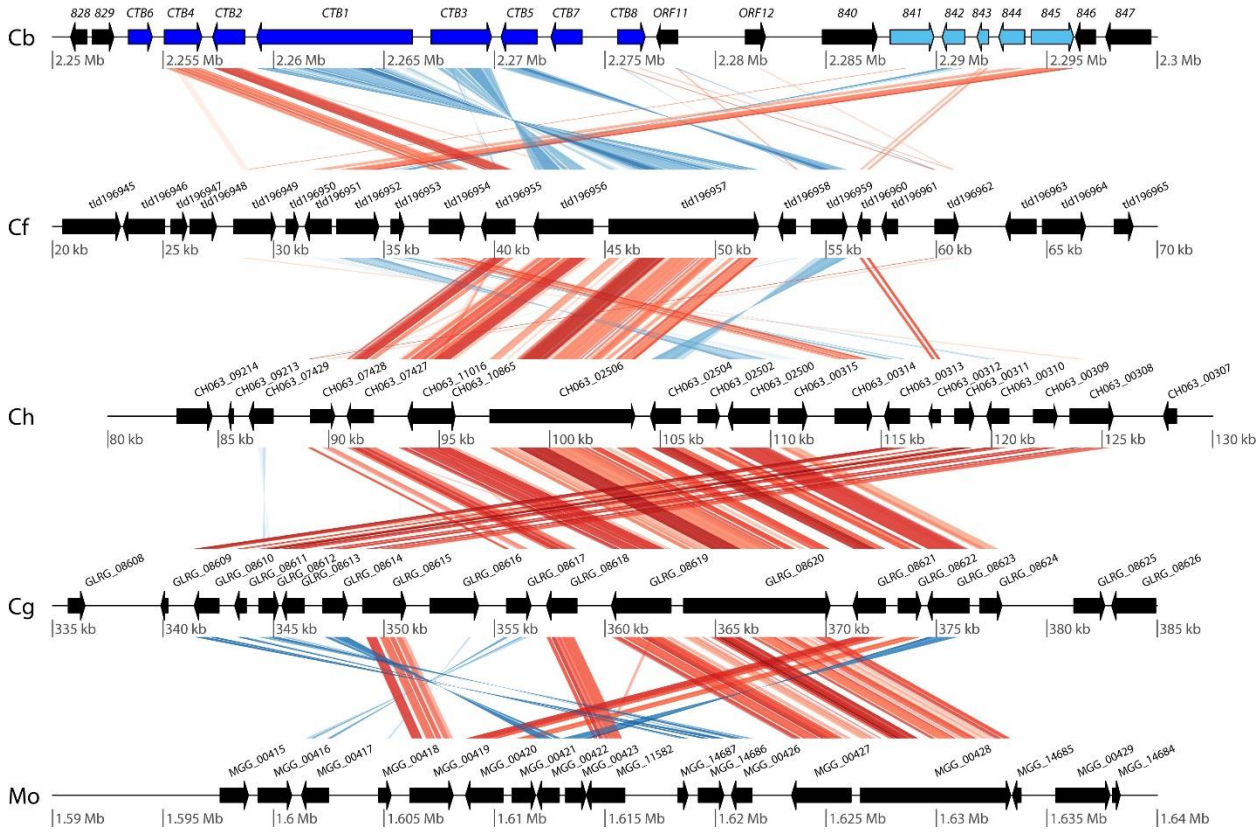

Figure 3: Synteny and rearrangements of the conserved *C. beticola* cercosporin biosynthetic cluster. The cercosporin biosynthetic cluster in *C. beticola* (Cb), marked in blue, and flanking genes are conserved in *C. fulvum* (Cf), *C. higginsianum* (Ch), *C. graminicola* (Cg) and *M. oryzae* (Mo). For all species the displayed identifiers are transcript IDs and the corresponding sequences can be retrieved from JGI MycoCosm or ORCAE. Alignment lines correspond to DNA fragments exhibiting significant similarity when the genomic regions comprising the gene clusters are compared with tBLASTx. Direct hits are displayed in red, whereas complementary hits are in blue. The hue of the alignments represents the percentage similarity ranging from 21 to 100 percent. Novel *C. beticola CTB* genes, identified in this study, are marked in light blue.

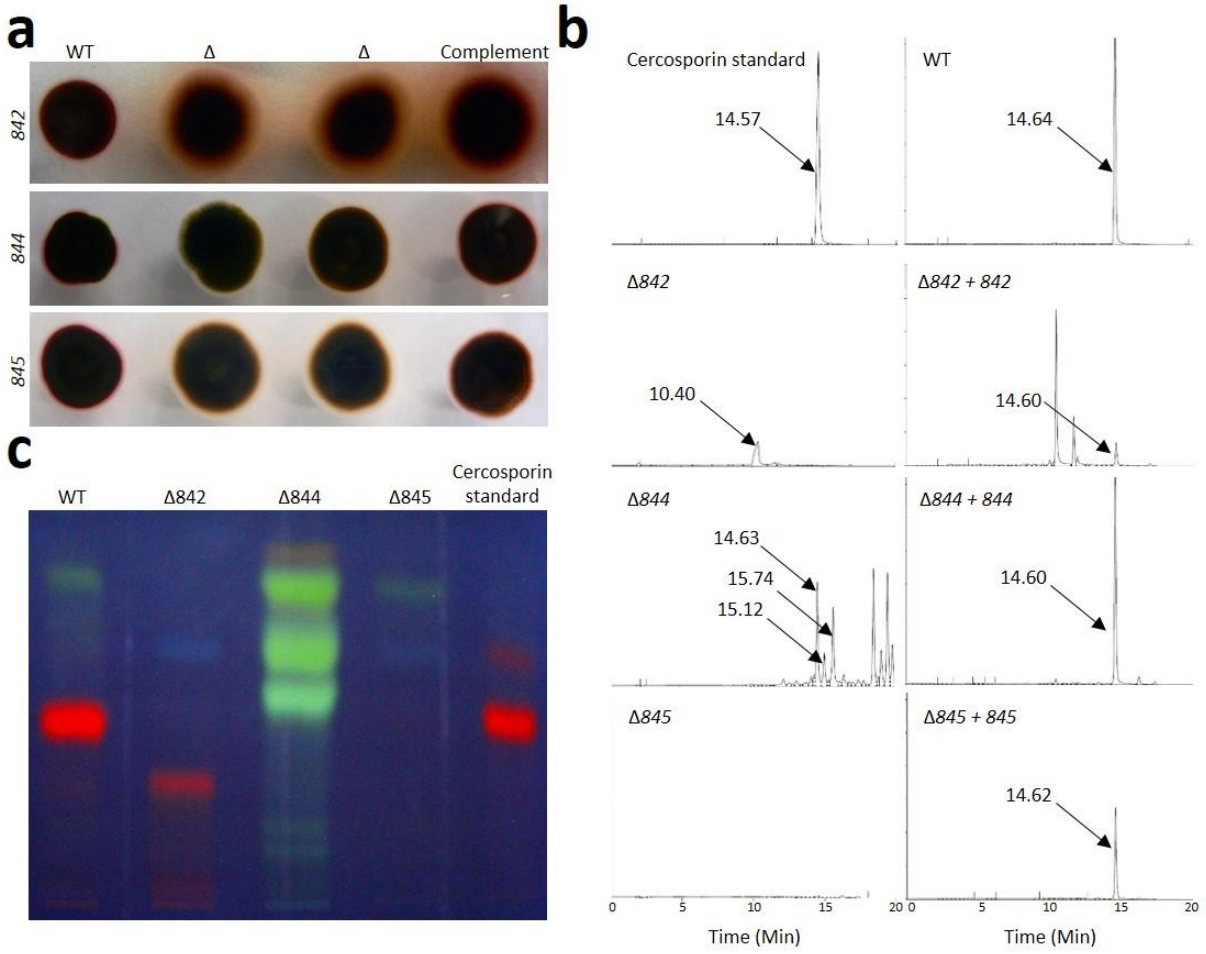

Figure 4: Analysis of cercosporin production in *CTB* mutants of *C. beticola*. Site-directed knock-out mutants in genes *CBET3_00842* (*CbCTB9*), *CBET3_00844* (*CbCTB10*), and *CBET3_00845* (*CbCTB11*) and their associated complements were assayed for cercosporin production (**a**) visually by growth on Petri plates containing thin PDA where the red pigment around cultures is indicative of cercosporin, (**b**) comparing HPLC retention times recorded at 280 nm from extracts of *C. beticola* and *C. beticola* mutants, and (**c**) TLC. A commercial cercosporin standard and cercosporin extracted from *C. beticola* strain 10-73-4 (WT) were used as controls in HPLC and TLC assays. Complemented mutants in HPLC assays are indicated as Δ842+842, Δ844+844, and Δ845+845.

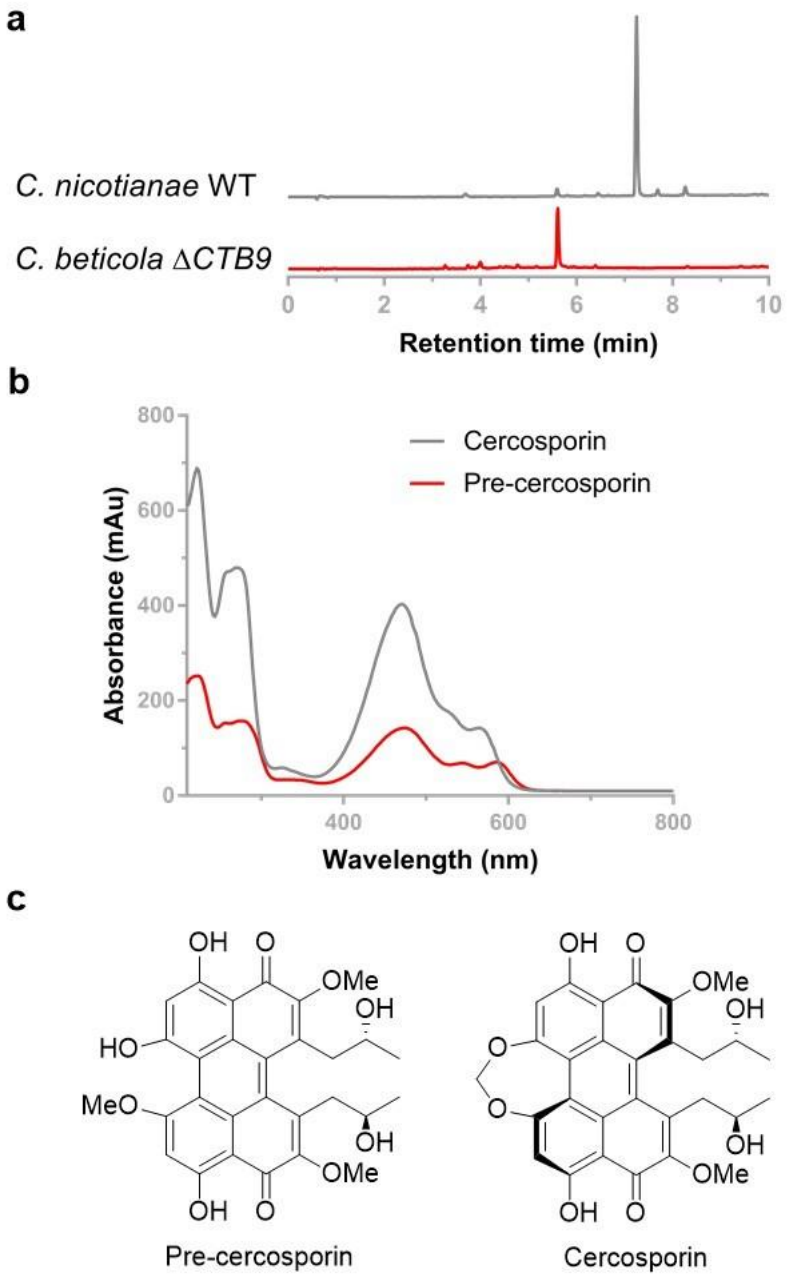

Figure 5: Characterization of the metabolite isolated from *C. beticola ΔCTB9*. (a) HPLC retention time comparison of cercosporin (gray) extracted from *C. nicotianae* and pre-cercosporin (red) extracted from Δ*CTB9*, recorded at 280 nm. (**b**) Overlay of the UV-vis spectra (210-800 nm) of cercosporin (gray) and pre-cercosporin (red). (**c**) Structures of pre-cercosporin and cercosporin (for ^{1}H-NMR spectra, see Methods).