Fungal Genetics and Biology 86 (2016) 71–80

Contents lists available at ScienceDirect

Fungal Genetics and Biology

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

Horizontal acquisition of toxic alkaloid synthesis in a clade of plant associated fungi

Marina Marcet-Houben [a,b], Toni Gabaldón [a,b,c,*]

[a] Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Dr. Aiguader, 88, 08003 Barcelona, Spain
[b] Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain
[c] Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain

ARTICLE INFO

Article history:
Received 24 August 2015
Revised 3 December 2015
Accepted 9 December 2015
Available online 10 December 2015

Keywords:
Clavicipitaceae
HGT
Loline
Ergot alkaloids
Phylogenetics
Phylome

ABSTRACT

Clavicipitaceae is a fungal group that comprises species that closely interact with plants as pathogens, parasites or symbionts. A key factor in these interactions is the ability of these fungi to synthesize toxic alkaloid compounds that contribute to the protection of the plant host against herbivores. Some of these compounds such as ergot alkaloids are toxic to humans and have caused important epidemics throughout history. The gene clusters encoding the proteins responsible for the synthesis of ergot alkaloids and lolines in Clavicipitaceae have been elucidated. Notably, homologs to these gene clusters can be found in distantly related species such as *Aspergillus fumigatus* and *Penicillium expansum*, which diverged from Clavicipitaceae more than 400 million years ago. We here use a phylogenetic approach to analyze the evolution of these gene clusters. We found that the gene clusters conferring the ability to synthesize ergot alkaloids and loline emerged first in Eurotiomycetes and were then likely transferred horizontally to Clavicipitaceae. Horizontal gene transfer is known to play a role in shaping the distribution of secondary metabolism clusters across distantly related fungal species. We propose that HGT events have played an important role in the capability of Clavicipitaceae to produce two key secondary metabolites that have enhanced the ability of these species to protect their plant hosts, therefore favoring their interactions.

1. Introduction

Clavicipitaceae is a fungal clade within the Sordariomycetes. Species belonging to this clade are known to interact mainly with plants and insects. Interactions with plants range from parasitic or pathogenic, to symbiotic relationships (Schardl et al., 2013c). Their potential to interact with plants has been associated to the ability of these fungi to produce certain secondary metabolites, among which several types of alkaloids are considered of key importance. Indeed the production of alkaloids, which are toxic to mammals and insects, has been proposed to provide plant protection against herbivores (Schardl et al., 2004; Wäli et al., 2013). Alkaloids produced by Clavicipitaceae species include, among others, ergot alkaloids, lolines and indole-terpenes (Schardl et al., 2013a). Different species, and even different strains, of Clavicipitaceae are able to synthesize different combinations of these compounds though they rarely produce all of them (Schardl et al., 2013b).

Probably the best known alkaloid class produced by Clavicipitaceae are the ergot alkaloids. These compounds are toxic to humans and livestock and have been the cause of many epidemics during human history (Lee, 2009). However, the incidence of ergotism in humans is currently very low and usually related to overdose of drugs derived from ergot alkaloids (Strickland et al., 2011). Contrary to the situation in humans, livestock is still often exposed to toxic alkaloids produced by Clavicipitaceae. Indeed pasture grasses are often colonized by endophytic fungi that are able to synthesize ergot alkaloids. In the United States annual losses in cattle production due to ergot alkaloid intoxication are estimated to be of the order of 1 billion dollars (Strickland et al., 2011). In contrast, loline alkaloids have rarely been linked to intoxications in mammals. Instead, they are broad spectrum insecticides (Schardl et al., 2007).

The gene clusters coding for the enzymes for the synthesis of ergot alkaloids and loline were discovered in *Claviceps purpurea* (Haarmann et al., 2005; Tudzynski et al., 1999) and in *Epichloë uncinata* (Spiering et al., 2005), respectively. Subsequent sequencing of numerous additional Clavicipitaceae genomes has shown that there is a tight relationship between the presence of the cluster and the production of these metabolites (Schardl et al., 2013b). Cluster presence and absence is very variable across species, and even across different strains of the same species. For instance, of

* Corresponding author at: C. Doctor Aiguader, 88, 08003 Barcelona, Spain.
E-mail address: tgabaldon@crg.es (T. Gabaldón).

http://dx.doi.org/10.1016/j.fgb.2015.12.006
1087-1845/

the two sequenced *Epichloë festucae* genomes, only one contains the loline cluster and is able to produce loline, while the other contains a cluster to synthesize indole-diterpenes (Schardl et al., 2013b). Differences in gene order and gene content can also be observed within a cluster. This high variability notwithstanding, there is generally a conserved core set of genes that are necessary to form the first stable compound of the pathway. This core set of genes is present in all species able to synthesize the compound. In addition to these core genes, there can be a variable number of accessory genes that participate in forming derivative compounds from the first stable compound. It is the variation in these accessory genes which provides the bulk of variability among the compounds synthesized by the clusters (Schardl et al., 2013b).

The synthesis of alkaloid compounds is not limited to Clavicipitaceae species. Ergot alkaloid production, for instance, has been detected in the distantly related *Aspergillus fumigatus* (Li and Unsöld, 2006; Panaccione and Coyle, 2005). A cluster containing 14 genes of which 8 were homologous to the genes found in the *C. purpurea* gene cluster was found in *A. fumigatus* (Coyle and Panaccione, 2005). As suggested by the differences in the specific gene content of the clusters, *A. fumigatus* and Clavicipitaceae species synthesize only a few common ergot alkaloid compounds, while most compounds are specific for each group of species. Despite this, it is thought that the two groups of species use orthologous genes to catalyze the first steps in the pathway and that differences affect later steps. More recently, an homologous loline cluster was found in the genome of *Penicillium expansum* (Ballester et al., 2015). Horizontal gene transfer (HGT) is known to play a role in the evolution of genetic clusters involved in the production of secondary metabolites (Wisecaver and Rokas, 2015). This process could also explain the appearance of similar gene clusters across these two distant groups of species. In order to assess whether HGT played a role in the evolutionary history of the gene clusters responsible to synthesize loline and ergot alkaloids, we performed a comprehensive phylogenetic analysis.

2. Materials and methods

2.1. Fungal genomes included

We downloaded 15 Clavicipitaceae genomes from NCBI (http://www.ncbi.nlm.nih.gov/). The latest version of their proteomes was downloaded from the University of Kentucky (http://www.endophyte.uky.edu/) (Schardl et al., 2013b). Two additional Clavicipitaceae were included from Uniprot. In addition 18 other fungal genomes were selected for comparative purposes. Among these genomes there were four additional Sordariomycetes, eight Eurotiomycetes, including *P. expansum* and *A. fumigatus*, two Leotiomycetes, two Dothideomycetes and one outgroup species. Details are found in Supplementary Table S1.

2.2. Phylogenetic analysis of alkaloid gene clusters

The proteins encoded in the gene clusters that synthesize ergot alkaloids and loline were downloaded from UniProt. The list of genes can be found in Supplementary Table S2. For each gene in these two clusters, a blast search against a database including the complete UniProt (UniProt Consortium, 2015) database and the proteomes of 15 Clavicipitaceae species was performed. The sequences of the first 150 hits were downloaded. In the case of lpsA, lpsB and lpsC proteins with a sequence length over 5000 were discarded from the analysis. For each group of homologous sequences a maximum likelihood tree was reconstructed. This was done using the same pipeline described for phylome reconstruction in Huerta-Cepas et al. (2011). Briefly, the homologous sequences were aligned using three different alignment programs (MUSCLE v3.8 (Edgar, 2004), MAFFT v6.712b (Katoh et al., 2005) and Kalign (Lassmann and Sonnhammer, 2005)). Alignments were done in forward and reverse (Landan and Graur, 2007). The six resulting alignments were then used to create a consensus alignment with M-COFFEE (Wallace et al., 2006). The alignment was trimmed using trimAl v1.4. (Capella-Gutiérrez et al., 2009) (consistency-score cut-off 0.1667, gap-score cut-off 0.9). The resulting alignment was then used to reconstruct a maximum likelihood tree. First the evolutionary model best fitting the data was chosen by reconstructing neighbor joining trees as implemented in BiONJ (Gascuel, 1997) and assessing the likelihood using seven different models. The best model according to the AIC criterion (Akaike, 1973) was selected and used to reconstruct a maximum likelihood tree as implemented in PhyML v3.0 (Guindon et al., 2010). In all trees four rate categories were used and invariant positions were inferred from the data. Bootstrap supports were calculated for each tree. Trees were rooted preferentially at a species that did not belong to the Pezizomycotina and was far related to the event of interest. When that was not possible due to a lack of suitable homolog a Pezizomycotina species belonging preferentially to Dothideomycetes or Leotiomycetes was chosen, always selecting leaves as far related as possible to the seed sequence. Trees were then analyzed manually to assess the consistency between the topology and the known species topology. Trees can be found in Supplementary Figs. S1–S23.

2.3. Species tree reconstruction

In order to reconstruct the species tree we reconstructed a phylome so that we could obtain the genes that had one to one orthologs in all the species considered. For each gene encoded in the genome of *E. festucae* E2368 a homology search was performed against a database that contained 35 fungal species (see Supplementary Table S1). Results were filtered according to an e-value and an overlap threshold (e-value < 1e−05 and overlap > 0.5). A maximum of 150 homologous sequences was taken. Then, for each group of homologs a maximum likelihood tree was reconstructed using the same methodology detailed above. Data produced during phylome reconstruction was stored at phylomeDB (Huerta-Cepas et al., 2014) (http://phylomedb.org/) with phyID code 125. Trees were then scanned using ETE v2.2 (Huerta-Cepas et al., 2010) to search for trees that were single copies in the 35 species. 906 such trees were selected and their alignments, as reconstructed in the phylome, were concatenated into a single multiple sequence alignment that contained 657,273 amino acids. Then RAxML v8.0.3 (Stamatakis et al., 2005) was used to reconstruct the species tree. Rapid bootstrap, as implemented in RAxML was used to calculate branch support. In addition, the phylome was also used to infer a supertree using duptree (Wehe et al., 2008). This algorithm looks for the species tree that infers the least number of duplication events when reconciling it to the gene trees generated in the phylome. Both methods produced identical trees.

A larger species tree was reconstructed in order to more accurately calculate the number of gene loss events. 268 fully sequenced genomes with their proteome predictions were downloaded from NCBI. A super tree approach was used to obtain the most likely topology for the species tree. Firstly, random groups of 15 species were selected, a blast search was performed between the species and best bidirectional hits were selected. Genes that had one hit per species were chosen to build a concatenated gene tree. At most 100 genes were selected. The species tree was reconstructed using the same methodology explained above. This was repeated over 50,000 times obtaining a total of 52,527 trees. A super tree was reconstructed using duptree and including all the small species trees.