Ochratoxin A production by *Penicillium thymicola*

Hai D.T. Nguyen^{a,b,*}, David R. McMullin^c, Ekaterina Ponomareva^b, Robert Riley^d, Kyle R. Pomraning^e, Scott E. Baker^e, Keith A. Seifert^{a,b}

^aUniversity of Ottawa, Department of Biology, 30 Marie-Curie Private, Ottawa, ON, K1N 6N5, Canada

^bAgriculture and Agri-Food Canada, Ottawa Research and Development Centre, 960 Carling Avenue, Ottawa, ON, K1A 0C6, Canada

^cCarleton University, Department of Chemistry, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6, Canada

^dUS Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

^ePacific Northwest National Laboratory, Environmental Molecular Sciences Laboratory, Earth and Biological Sciences Directorate, 3335 Innovation Boulevard, Richland, WA 99354, USA

* Corresponding author. Tel.: +1 (613) 769 3756; fax: +1 (613) 759 1701.

E-mail address: hai.nguyen.1984@gmail.com (H.D.T. Nguyen).

Abstract: Ochratoxin A (OTA) is a mycotoxin produced by some *Aspergillus* and *Penicillium* species that grow on economically important agricultural crops and food products. OTA is classified as Group 2B carcinogen and is potently nephrotoxic, which is the basis for its regulation in some jurisdictions. Using high resolution mass spectroscopy, OTA and ochratoxin B (OTB) were detected in liquid culture extracts of *P. thymicola* DAOMC 180753 isolated from Canadian cheddar cheese. The genome of this strain was sequenced, assembled and annotated to probe for putative genes involved in OTA biosynthesis. Known OTA biosynthetic genes from *P. verrucosum* or *P. nordicum*, two related *Penicillium* species that produce OTA, were not found in *P. thymicola*. However, a gene cluster containing a polyketide synthase (PKS) and PKS-nonribosomal peptide synthase (NRPS) hybrid encoding genes were located in the *P. thymicola* genome that showed a high degree of similarity to OTA biosynthetic enzymes of *Aspergillus carbonarius* and *A. ochraceus*. This is the first report of ochratoxin from *P. thymicola* and a new record of the species in Canada.

Key words: mycotoxin; fumiquinazoline F; polyketide synthase (PKS); non-ribosomal peptide synthase (NRPS)

1. Introduction

The mycotoxin ochratoxin A (OTA) is a natural agricultural contaminant produced by several *Penicillium* and *Aspergillus* species (Varga et al. 2001). OTA is potentially nephrotoxic in rodents, pigs and poultry and exposure results in renal tumours in a variety of animal species (Barlow et al. 2008; Bondy et al. 2015). OTA is classified as a possible human carcinogen (Group 2B) by the International Agency of Research on Cancer (IARC 1993). The European Commission has several regulations on OTA in various commodities (van Egmond et al. 2007). These regulations directly affect the trade of raw materials and finished food products globally. This can affect the entire food production chain all the way to farmers and food producers (Miller et al. 2014).

Two sister species of *Penicillium* sect. *Fasiculata* (Frisvad & Samson 2004) are important OTA producers in food; *P. verrucosum* on grains (Larsen et al. 2001a; Lund & Frisvad 2003) and *P. nordicum* on dried, salted meats (Cabañes et al. 2010; Sonjak et al. 2011). The growth of *P. verrucosum* on damp grains in storage is a major source of OTA and results in significant human exposure, either by direct consumption of contaminated products or indirectly via products made from animals fed contaminated feed (Duarte et al. 2010). *Penicillium nordicum* produces more OTA than *P. verrucosum* and is a serious concern in some European countries, where cured meats are consumed regularly. *Penicillium thymicola* is a recently described species, isolated from dried thyme, but also from soil and sorghum grain (Frisvad & Samson 2004). Not previously known to produce OTA, although European strains synthesize the amino acid derived compounds alantrypinone and fumiquinazoline F, and the alpha-pyrone PC-2 (Larsen et al. 1998; Frisvad & Samson 2004).

Thirteen species of *Aspergillus* produce OTA, mainly in sections *Circumdati* (Visagie et al. 2014a) and *Nigri* (Samson et al. 2004). Although the occurrence of OTA in beer is a

consequence of *P. verrucosum* growing on barley, the occurrence of OTA in grapes, wine, cocoa and coffee is probably attributable to contaminating *Aspergillus* species (Barlow et al. 2008; Kuiper-Goodman et al. 2010; Nguyen & Ryu 2014; Lee & Ryu 2015), such as *A. carbonarius* or some strains of *A. niger* (Varga et al. 2011).

The chemical structure of OTA consists of a polyketide-derived dihydroisocoumarin moiety linked via the carboxyl group of the amino acid phenylalanine (Harris & Mantle 2001). Despite the negative health effects caused by OTA, little is currently known about the genetics of its biosyntheses. It is possible that a polyketide synthase (PKS) catalyzes the formation of a dihydroisocoumarin pentaketide intermediate that is subsequently chlorinated by a chloroperoxidase or halogenase and the amino acid phenylalanine is then linked to the polyketide intermediate by a non-ribosomal peptide synthase (NRPS) (Harris & Mantle 2001; el Khoury & Atoui 2010). Fungal PKS's consist of multiple catalytic domains with ketosynthase (KS), acetyltransferase (AT) and acyl carrier protein (ACP) domains. Additional processing domains, including ketoreductase (KR), dehydratase (DH), enoyl reductase (ER), cyclase, thioesterase and methyltransferase functions, are also often present within a single PKS complex (Keller et al. 2005). A typical NRPS consists of adenylation (A), thiolation (T), condensation (C) and terminal release domains (Schwarzer et al. 2003). Functional characterizations of some PKS and NRPS genes have been performed by gene inactivation and expression experiments in *P. nordicum* (Karolewicz & Geisen 2005; Geisen et al. 2006), *P. verrucosum* (Schmidt-Heydt et al. 2007; Abbas et al. 2013), *A. carbonarius* (Gallo et al. 2009; Gallo et al. 2012; Gallo et al. 2014), *A. ochraceus* (O'Callaghan et al. 2003) and *A. westerdijkiae* (Bacha et al. 2009).

During a large study screening Canadian isolates of *P. verrucosum* and related species for OTA, OTA and OTB were unexpectedly identified by high-resolution mass

spectrometry (HRMS) from a culture filtrate extract of *P. thymicola* DAOMC 180753, originally isolated from cheddar cheese. Because of the close evolutionary relationship of *P. thymicola* to *P. nordicum* and *P. verrucosum*, their metabolite profiles were further compared. A genomic analysis of *P. thymicola* was undertaken to locate putative genes required for OTA biosynthesis. Here, we report the production of OTA and OTB, locate putative OTA biosynthetic genes using genomic data and document the occurrence of *P. thymicola* in Canada.

2. Materials and methods

2.1 Chemical detection of OTA

An agar slant of *Penicillium thymicola* DAOMC 180753 growing on 2 % Blakeslee's malt extract agar (Visagie et al. 2014b) was macerated in sterile distilled deionized water and a 5 % (v/v) aliquot was used to inoculate nine 250 mL Erlenmeyer flasks containing 100 mL of yeast extract sucrose media (30 g/L yeast extract (Sigma, USA) and 150 g/L sucrose (BioShop, Canada)). Flasks were incubated on a rotary shaker (100 rpm at 25 °C) for 5 d.

After the incubation period, culture filtrates of three individual flasks were combined, resulting in three replicates for subsequent analysis. From these, the mycelia were separated from the filtrate by vacuum filtration through Whatman #4 filter paper (Whatman GE Healthcare, UK) and the resulting volume of media recorded. Mycelia were subsequently frozen at -20 °C, lyophilized and weighed. Each culture filtrate was saturated with NaCl and extracted with three equal volumes of ethyl acetate. The organic layer was filtered through a Whatman #1 (Whatman GE Healthcare, UK) and anhydrous sodium sulfate prior to drying by rotary evaporation. The crude extract was dissolved in a minimal amount of HPLC grade methanol, filtered through 13 mm PTFE, 0.2 µm syringe filter (Tisch Scientific, USA) and dried

under a gentle stream of nitrogen gas. Extracts were stored dry in amber vials at -20 °C. Additionally, extracts of *P. nordicum* DAOMC 158653 and *P. verrucosum* DAOMC 242724 were prepared as reported above.

Crude filtrate extracts were dissolved in 300 µL of HPLC grade methanol and analyzed for OTA by liquid chromatography tandem mass spectrometry (LC-MSMS) in multiple reaction monitoring (MRM) mode. This was achieved using a Waters 2795 separations module and MicroMass Quatro Ultima triple quadrupole mass spectrometer (Waters Corporation, USA). Extracts were separated with a Kinetix C18 (100 × 4.60 mm, 2.6 µm) column (Phenomenex, USA) and a mobile phase consisting of acetonitrile-deionized water with 0.1 % formic acid (v/v). The solvent gradient was a linear program that increased from 5 to 100 % acetonitrile over 13 minutes with a flow rate of 1.0 mL/min. Positive ESI conditions included: capillary voltage 3.50 kV, cone voltage 20 V, source temperature 80 °C, desolvation temperature 180 °C, cone gas flow (N₂) 90 L/h, desolvation gas flow (N₂) 540 L/h, collision gas (Ar) pressure 2.3×10^{-4} mbar and multiplier voltage 650 V. LC-MSMS analysis of OTA used two transitions, m/z 404→358 and m/z 404→341, with the same collision energy, 20 eV. The m/z 358 transition was the quantifier whereas m/z 341 was the qualifier. A standard curve was constructed with an authentic OTA standard (Sigma, USA) ranging from 1 to 10,000 ng on column with an R² value of 0.99 over the entire concentration range.

HRMS and tandem high resolution mass spectrometry (HRMS²) of *Penicillium* extracts were acquired with a Q-Exactive Quadrupole Orbitrap Mass Spectrometer (Thermo Scientific, USA), coupled to a 1290 HPLC system (Agilent, USA) in ESI positive mode (scan range 100-1500 m/z). Metabolites were separated with a Zorbax Eclipse Plus RRHD C18 (2.1 × 50 mm, 1.8 µm) column (Agilent, USA) maintained at 35 °C and a mobile phase consisting of acetonitrile-deionized water with 0.1% formic acid (v/v). Acetonitrile was held at 0 % for 0.5

min, before increasing to 100 % over 3.5 min and held at 100 % for 1.5 min before returning to 0 % over 0.5 min at a flow rate of 0.3 mL/min. Authentic standards of OTA (Sigma, USA), citrinin (Sigma, USA) and phenylalanine (Sigma, USA) were utilized in conjunction with high resolution mass data to identify secondary metabolites.

2.2 Growth, DNA extraction and genome sequencing

Penicillium thymicola DAOMC 180753 was grown on Blakeslee's malt extract agar for 7 d at 25 °C (Visagie et al. 2014b). Colonies were flooded with 5 mL of sterile distilled water to make a spore suspension. One mL was inoculated in 100 mL of Blakeslee's malt extract broth and was left on a rotary at 300 rpm at 25 °C for 6 d. To isolate fungal culture pellets for DNA extraction, the broth was removed by filtration. DNA was extracted with the OmniPrep kit for fungi (G-Biosciences, St. Louis, Missouri, USA) following manufacturer's instructions. DNA extracts were combined and whole-genome sequencing (paired-end with 101 base pairs (bp)) was performed on an Illumina HiSeq 2500 with TrueSeq V3 chemistry at the National Research Council Canada in Saskatoon, Saskatchewan, Canada.

2.3 Genome assembly and annotation

Quality of the reads was checked with the program FastQC v0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Using fastx_trimmer (part of the FASTX-Toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/)), 10 bases from the 5' end were trimmed to yield 91 bp reads of higher quality. Afterwards, adaptor sequences were removed with Trimmomatic v0.33 (Bolger et al. 2014). The optimal k parameter was predicted with KmerGenie v1.6950 (Chikhi & Medvedev 2014). *De novo* genome assembly was performed with SPAdes v3.5.0 (Bankevich et al. 2012) and error correction using

BayesHammer (Nikolenko et al. 2013) with mismatch correction enabled. Scaffolds shorter than 1000 bp were discarded. Assembly statistics were generated with QUAST v2.3 (Gurevich et al. 2013). The assembly was checked by alignment of the corrected reads onto the scaffolds using Bowtie2 v2.0.0 (Langmead & Salzberg 2012). Alignments produced by Bowtie2 in SAM format were converted to sorted BAM format by SAMtools v0.1.19 (Li et al. 2009) and statistics for coverage were generated with Qualimap v2.1 (Garcia-Alcalde et al. 2012). To assess the completeness of our genome assembly, CEGMA v2.5 (Parra et al. 2007) was run on the scaffolds to detect the percentage of conserved eukaryotic genes (CEG's).

Gene annotation was performed using the JGI Annotation pipeline (Grigoriev et al. 2006). Subsequently, the antibiotics & Secondary Metabolite Analysis Shell (antiSMASH) v2.1.1 (Blin et al. 2013) was used to predict gene clusters involved in secondary metabolite production with annotations generated from the JGI annotation pipeline. All data are available for download on the JGI MycoCosm portal (Grigoriev et al. 2014) (jgi.doe.gov/fungi) and raw reads are available on NCBI SRA (ncbi.nlm.nih.gov/sra) under accession number SRR1909079.

3. Results

3.1 Secondary metabolite profiles of P. thymicola, P. verrucosum and P. nordicum

The secondary metabolite profiles of *P. thymicola* DAOMC 180753, *P. verrucosum* DAOMC 242724 and *P. nordicum* DAOMC 158653 were investigated by HRMS approaches and these data are presented in Table 1. Each of the three *Penicillium* sister species produced OTA as determined by comparisons of high resolution product ion spectra to the literature (Nielsen & Smedsgaard 2003; Sulyok et al. 2006; Varga et al. 2013) and an

authentic standard. HRMS analysis of the *P. thymicola* extract revealed an $[M+H]^+$ OTA precursor ion at m/z 404.08945 (1.6 ppm mass error) and product ions at m/z 358.08438, 341.05795, 257.02130 and 239.01070 (Fig 1). The retention index and mass spectra of OTA from the crude extract were compared to the authentic standard and found to be identical, unambiguously identifying the mycotoxin OTA within the culture filtrate extract of *P. thymicola*. Each *Penicillium* species culture filtrate extract was also analyzed by LC-MRM mass spectrometry. The OTA precursor ion at m/z 404.0 $[M+H]^+$ was selected and quantitation was achieved using the m/z 358.0 product ion, which is in agreement with the results of Nielsen & Smedsgaard (2003). Using the reported LC-MRM conditions, *P. thymicola* DAOMC 180753 produced 7.5 ± 1.5 ng OTA/mL/g dry cells, *P. verrucosum* DAOMC 242724 39.1 ± 1.5 ng OTA/mL/g dry cells and *P. nordicum* DAOMC 158653 276.3 ± 43.1 ng OTA/mL/g dry cells. OTB, the dechlorinated form of OTA, and phenylalanine were also synthesized by each *Penicillium* species (Table 1). From the *P. thymicola* extract, an OTB precursor $[M+H]^+$ ion at m/z 370.12860 (1.3 ppm mass error) and product ions at m/z 307.09845, 223.06009 and 205.04970 were observed, characteristic of this metabolite (Nielsen & Smedsgaard, 2003). Fumiquinazoline F (m/z 359.15024 $[M+H]^+$; 1.6 ppm mass error) and alantrypinone (m/z 373.12967 $[M+H]^+$; 1.1 ppm mass error; $C_{21}H_{16}N_4O_3$) were produced as major metabolites by this strain. The alpha-pyrone PC-2 was identified based on high-resolution mass (m/z 213.11217 $[M+H]^+$; 0.23 ppm mass error; $C_{11}H_{17}O_4$) from *P. thymicola*. At the same retention index, this metabolite was also detected in the two other *Penicillium* species extracts. The polyketide citrinin was only detected from *P. verrucosum*. *Penicillium nordicum* is suspected to produce sclerotigenin based on a $[M+H]^+$ ion at m/z 278.09286 (0.36 ppm mass error; $C_{16}H_{11}N_3O_2$) and appropriate retention index. A product ion of this metabolite at m/z 249.06618 and a large proportion of the precursor ion were observed in its HRMS² spectra.

Neither lumpidin nor verrucines A and B were detected in any of the culture filtrate extracts examined based on interpretation of HRMS data.

3.2 Genome sequencing and characterization

The genome of *P. thymicola* DAOMC 180753 was sequenced using a whole genome shotgun-sequencing strategy to locate putative OTA biosynthetic gene cluster(s).

Approximately 28 million reads were assembled into 2167 scaffolds. The genome size was 33.9 Mb with a GC content of 47.9 %. The N50 value was 45.4 kb and the longest scaffold was approximately 215 kb. The median nucleotide coverage across the whole assembly was 53.7x. The assembled genome had a CEGMA score of 94.8% calculated from the complete gene set and 97.2 % when calculated from both partial and complete gene sets. A total of 12467 protein-coding genes were predicted. Of all predicted gene models, 11240 were complete (90.2 %), but 1227 gene models lacked a start codon, stop codon or both (9.8 %). The mean gene length was 1582 bp, mean exon length was 462 bp and mean intron length was 90 bp.

3.3 Identification of secondary metabolite gene clusters

After the genome of *P. thymicola* DAOMC 180753 was annotated, the bioinformatics program antiSMASH predicted a total of 99 secondary metabolite gene clusters, including 23 PKS, 17 NRPS, 6 PKS-NRPS hybrid, 13 terpene, 1 siderophore, and 39 labelled as putative or other.

To identify genes involved in OTA biosynthesis from *P. thymicola*, protein sequences from published OTA biosynthetic gene knockout and expression studies of *Penicillium* and *Aspergillus* species were queried. These sequences were compared to the predicted proteins

from *P. thymicola* DAOMC 180753 with the BLASTp algorithm (Suppl. Table 1). Proteins with percent identity > 70% and an e-value < $1e^{-100}$ were considered strong matches. These were additionally verified to be arranged in secondary metabolite gene clusters. Protein sequences of three genes (*gene_4875*, *gene_11104* and *gene_3514*) showed strong homology to proteins identified as involved in OTA biosynthesis in *Penicillium* (Karolewicz & Geisen 2005) and *Aspergillus* species (O'Callaghan et al. 2003; Bacha et al. 2009; Gallo et al. 2009; Gallo et al. 2012; Gallo et al. 2014) (Table 2).

Two of these genes making proteins homologous to known OTA biosynthetic enzymes were not organized within recognized gene clusters, but were located near the middle of genomic scaffolds. The *gene_11104* was located on scaffold 785 (9430 bp long, positions 5059-6399). Its protein (jgi|Penth1|176637) showed homology to an alkaline serine protease from *P. nordicum* (*aspPN*). The *gene_3514* was located on scaffold 64 (88623 bp long, positions 46890-58790) and encoded a 3916 amino acid protein (jgi|Penth1|168734) similar to the PKS enzyme from gene *ACpks* (GenBank CAQ16344.1) from *A. carbonarius* (Gallo et al. 2009). This *P. thymicola* protein was identified as producing a hybrid PKS-NRPS and had the following domains at various amino acid positions: KS from 44-467, AT from 575-890, DH from 967-1144, ER from 1649-1966, KR from 1992-2167, a C domain from 2388-2689, an A domain from 2881-3291, and a PCP domain from 3422-3489.

The *gene_4875* belongs to a type I PKS gene cluster called cluster 39 (Fig 2). Cluster 39 comprises 25 genes spanning 68.4 kb on scaffold 112 (Table 3). Within cluster 39, *gene_4875* encodes a 848 amino acid protein (jgi|Penth1|212294), that has an AT domain from position 280-581, *gene_4876* encodes a 332 amino acid protein (jgi|Penth1|170133) with an ER domain from position 6-260 and *gene_4877* encodes a 166 amino acid protein (jgi|Penth1|78053) with an ACP domain from position 104-151. The translated proteins of

remaining genes within cluster 39 had poor matches to other proteins in the manually curated UniProt/Swiss-Prot protein database (Table 3).

Additionally, within the *P. thymicola* genome, there is a single halogenase (jgi|Penth1|178787) and two chloroperoxidase (jgi|Penth1|237379 and 238064) that could potentially catalyze the chlorination step in OTA biosynthesis.

4. Discussion

Here, we report the production of OTA and OTB from a Canadian strain of *P. thymicola* isolated from cheddar cheese. Because of the close relationship of *P. thymicola* to well-documented ochratoxin-producing species of *Penicillium* (*P. nordicum* and *P. verrucosum*), a tight clade within *Penicillium* sect. *Fasciculata* (Frisvad & Samson 2004), their chemical profiles were compared (Table 1). Each of the three species produced OTA and its dechlorinated form, OTB, based on comparisons of experimental HRMS and HRMS² spectra to the literature and an authentic standard for OTA (Nielsen & Smedsgaard 2003; Sulyok et al. 2006; Varga et al. 2013). *Penicillium nordicum* produced the most OTA of the strains studied, whereas *P. verrucosum* produced slightly more OTA compared to *P. thymicola*.

To date, OTA biosynthetic studies have only examined the inactivation of one gene and/or measure the expression of other genes encoded within an OTA gene cluster to investigate a correlation with its production. No study has been conducted where each gene in an identified OTA gene cluster was knocked out. It is not yet clear which genes, or protein domains, found within an OTA biosynthetic gene cluster are essential for its production.

The genome sequencing of *P. thymicola* DAOMC 180753 was originally undertaken as a negative control for comparative genomics of OTA gene clusters in *P. verrucosum* and *P. nordicum*. However, when OTA and OTB production were unexpectedly detected by HRMS,

the genome was probed for putative OTA biosynthetic genes using protein sequences from known OTA genes of *Penicillium* and *Aspergillus* species. Secondary metabolite biosynthetic genes are often organized in clusters that encode key biosynthetic enzymes such as polyketide synthases (PKS's), non-ribosomal peptide synthases (NRPS's) and PKS-NRPS hybrid complexes, as well as both pathway-specific and global regulatory elements (Keller et al. 2005; Yu & Keller 2005). Initially, we thought that *P. thymicola* would have an OTA gene cluster similar to that of *P. nordicum* (Karolewicz & Geisen 2005; Geisen et al. 2006) or *P. verrucosum* (Schmidt-Heydt et al. 2007; Abbas et al. 2013). However, no proteins were identified from *P. thymicola* that were homologous with *P. verrucosum* (GenBank ABH05672.1) biosynthetic enzymes. One *P. thymicola* protein (jgi|Penth1|176637 from *gene_11104*) was homologous to an alkaline serine protease (*aspPN*) (GenBank AAT65816), previously found 6 kb distant from the *otapksPN* gene in *P. nordicum*, perhaps part of the same OTA gene cluster. However, *aspPN* was not inactivated by Karolewicz & Geisen (2005), leaving its impact on OTA biosynthesis ambiguous.

Considering that *P. thymicola* protein sequences did not have meaningful homologies with any *P. nordicum* or *P. verrucosum* OTA biosynthetic genes, genes from *Aspergillus* species were queried instead. As a result, we suggest that genes encoded within cluster 39 (named herein), which covers the entirety of scaffold_112, and *gene_3514* may be involved with OTA production by *P. thymicola* DAOMC 180753. Together, they encode the typical protein domains found in PKS and NRPS enzyme complexes and are homologous to enzymes critical for OTA biosynthesis by *Aspergillus* species. The protein from *P. thymicola* for *gene_4875* (jgi|Penth1|212294), part of cluster 39, showed >70% homology to an *A. ochraceus* PKS protein (GenBank AAP32477.1) (O'Callaghan et al. 2003) and the *AcOTApks* protein (jgi|Aspca3|173482) from *A. carbonarius* (Gallo et al. 2014). Both these *Aspergillus*

genes are directly implicated in OTA biosynthesis, as demonstrated by gene inactivation and expression experiments (Gallo et al. 2009; Gallo et al. 2012; Gallo et al. 2014). The alignment length of the *P. thymicola* and *Aspergillus* genes was short, therefore it did not occur as one of the top BLASTp hits (Suppl. Table 1). However, the portion that aligned coincided with the AT domain of the *AcOTApks* protein, an integral component of any PKS enzyme complex. Aside from biosynthetic genes, cluster 39 includes one regulatory (*gene_4856*) and one transporter (*gene_4860*) gene (Table 3). This transporter gene is weakly homologous to a major facilitator superfamily (MFS) transporter. Geisen et al. (2006) reported that the *P. nordicum* gene *otatraPN* has homology with a MFS transporter. A BLAST search querying the UniProt/Swiss-Prot database did not reveal a halogenase, chloroperoxidase or peptide synthetase genes within cluster 39. Elsewhere in the genome, genes for a potential halogenase and two chloroperoxidases were identified, but they were not located in a cluster with other putative OTA biosynthesis related genes.

Gene_3514 (jgi|Penth1|168734) is a PKS-NRPS hybrid and not part of a gene cluster, although it was located in the middle of a genomic scaffold. This protein was homologous to a *ACpks* protein from *A. carbonarius* (Gallo et al. 2009) that was differentially expressed only when OTA was synthesized, although no knockouts were investigated by Gallo et al. (2009). Using gene inactivation, Gallo et al. (2012) demonstrated that the NRPS gene *AcOTAnrps* is implicated in OTA biosynthesis by *A. carbonarius*. However, the protein of *gene_3514* in *P. thymicola* did not have significant homology with the *AcOTAnrps* protein (Suppl. Table 1).

While *gene_3514* encodes a protein with all the typical NRPS domains, it lacks an ACP domain in the PKS portion; however, *gene_4877* in cluster 39 has this domain. Conversely, genes within cluster 39 do not have the KS domain of a typical PKS while the PKS part of *gene_3514* has a KS domain. Therefore, OTA biosynthesis in *P. thymicola* may

fall into one of these occurrences of secondary metabolite genes that are not clustered in a single locus. Unclustered secondary metabolite biosynthetic pathways are rare but do occur. For example, the DHN melanin biosynthesis pathway is not always clustered (Baker 2008; Chiang et al. 2011) and the T-toxin cluster in *Cochliobolus heterostrophus* is divided among two clusters (Kodoma et al. 1999; Turgeon & Baker 2007). In *Fusarium* spp., trichothecene biosynthetic genes are found in multiple loci, with some genes clustered and others existing as isolated genes (reviewed in Kimura et al. 2007).

Our analysis of *P. thymicola* DAOMC 180753 genomic data indicates that it possess OTA biosynthetic machinery more similar to that of *Aspergillus* species compared to its sibling species *P. verrucosum* and *P. nordicum*. The major metabolites produced by this species were fumiquinazoline F and alantrypinone, previously identified from the type strain of *P. thymicola* CBS 111225. They are derived from the amino acids alanine and tryptophan (Larsen et al. 1998). The amino acid phenylalanine, the non-ribosomal peptide structural component of ochratoxins (Harris & Mantle 2001), was also detected from each of the three *Penicillium* species examined. Fumiquinazoline F was originally reported from an *Aspergillus fumigatus* strain (Takahashi et al. 1995). Because *P. thymicola* harbors secondary metabolite biosynthetic capacity for fumiquinazoline F, and this metabolite is absent from *P. verrucosum* or *P. nordicum*, there is some precedence for the notion that some of its secondary metabolite biosynthesis is more similar to *Aspergillus*. Next generation sequencing does have biases and it is possible that an OTA gene cluster in *P. thymicola* occurs in a region that is difficult to sequence, where there is lower coverage or repetitive regions, and therefore remained unassembled. Gene knockout, gene expression and chemical tracing experiments should verify whether the putative genes identified here actually do play a role in OTA biosynthesis in *P. thymicola*. Although a major undertaking, it is only when the intermediate products are

detected after inhibition or deletion of a particular gene that functional OTA biosynthesis can be accurately assessed. The detection of citrinin from the *P. verrucosum* culture filtrate extract and not from *P. thymicola* or *P. nordicum* is consistent with the results of Larsen et al. (2001a). The alpha pyrone PC-2 was identified from each of the *Penicillium* extracts based on HRMS. *Penicillium thymicola*, *P. nordicum* and *P. verrucosum* produce this metabolite (Frisvad & Samson 2004), which has been suggested as a biomarker for known OTA producing *Penicillium* species (Rahbaek et al. 2003). An attempt was made to investigate the production of verrucines A and B (*P. verrucosum* metabolites; Larsen et al. 1999) and the diketopiperazine lumpidin (*P. nordicum* metabolite; Larsen et al. 2001b). Their high resolution precursor ions as well as verrucines product ions reported by Nielsen & Smedsgaard (2003) were queried. None of these compounds were confirmed in the studied extracts. Apart from the production of OTA and OTB by *P. thymicola*, the secondary metabolite profiles reported here agree with the results of Frisvad & Samson (2004).

5. Acknowledgements

Genome sequencing of *P. thymicola* was funded by Growing Forward 2 funding from Agriculture & Agri-Food Canada, with additional support from Canadian Safety and Security Programme grant CRTI 09-462RD/CSSP 30vv01. We thank: J. David Miller (Carleton University) for assistance with fermentations and helpful suggestions with the manuscript. Mark Sumarah and Justin Renaud (AAFC, London, Ontario, Canada) for acquisition of HRMS data. Kyle R. Pomraning and Scott E. Baker are scientists at EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory.

6. References

- Abbas A, Coghlan A, O'Callaghan J, Garcia-Estrada C, Martin JF, Dobson AD, 2013. Functional characterization of the polyketide synthase gene required for ochratoxin A biosynthesis in *Penicillium verrucosum*. *International Journal of Food Microbiology* **161**: 172–181.
- Bacha N, Atoui A, Mathieu F, Liboz T, Lebrihi A, 2009. *Aspergillus westerdijkiae* polyketide synthase gene "aoks1" is involved in the biosynthesis of ochratoxin A. *Fungal Genetics and Biology* **46**: 77–84.
- Baker SE, 2008. *Aspergillus* genomics and DHN-melanin conidial pigmentation. In: Varga J, Samson R (eds), *Aspergillus in the Genomics Era*. Wageningen Academic Publishers, Wageningen, pp. 73–85.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijbelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA, 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.
- Barlow S, Bolger M, Pitt JI, Verger P, 2008. Ochratoxin A (addendum). In: *WHO Food Additives Series 59. Safety evaluation of certain contaminants in food*. World Health Organization, Geneva, pp. 357–429.

Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T, 2013. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research* **41**(Web Server issue): W204–212.

Bolger AM, Lohse M, Usadel B, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bondy GS, Caldwell DS, Aziz SA, Coady LC, Armstrong CL, Curran IH, Koffman RL, Kapal K, Lefebvre DE, Mehta R, 2015. Effects of chronic Ochratoxin A exposure on p53 heterozygous and p53 homozygous mice. *Toxicologic Pathology* **43**:715–729.

Cabañes FJ, Bragulat MR, Castellá G, 2010. Ochratoxin A producing species in the genus *Penicillium*. *Toxins* **2**:1111–1120.

Chiang YM, Meyer KM, Praseuth M, Baker SE, Bruno KS, Wang CCC, 2011.

Characterization of a polyketide synthase in *Aspergillus niger* whose product is a precursor for both dihydroxynaphthalene (DHN) melanin and naphtho- γ -pyrone. *Fungal Genetics and Biology* **48**: 430–437.

Chikhi R, Medvedev P, 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**:31–37.

Duarte SC, Pena A, Lino CM, 2010. A review on ochratoxin A occurrence and effects of processing of cereal and cereal derived food products. *Food Microbiology* **27**: 187-198.

el Khoury A, Atoui A, 2010. Ochratoxin a: general overview and actual molecular status. *Toxins* **2**: 461–493.

Frisvad JC, Samson RA, 2004. Polyphasic taxonomy of *Penicillium* subgenus *Penicillium*: a guide to identification of food and airborne terverticillate *Penicillia* and their mycotoxins. *Studies in Mycology* **49**:1–174.

Gallo A, Perrone G, Solfrizzo M, Epifani F, Abbas A, Dobson AD, Mulè G, 2009. Characterisation of a pks gene which is expressed during ochratoxin A production by *Aspergillus carbonarius*. *International Journal of Food Microbiology* **129**: 8-15.

Gallo A, Bruno KS, Solfrizzo M, Perrone G, Mulè G, Visconti A, Baker SE, 2012. New insight into the ochratoxin A biosynthetic pathway through deletion of a nonribosomal peptide synthetase gene in *Aspergillus carbonarius*. *Applied and Environmental Microbiology* **78**: 8208–8218.

Gallo A, Knox BP, Bruno KS, Solfrizzo M, Baker SE, Perrone G, 2014. Identification and characterization of the polyketide synthase involved in ochratoxin A biosynthesis in *Aspergillus carbonarius*. *International Journal of Food Microbiology* **179**: 10–17.

Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, Dopazo J, Meyer TF, Conesa A, 2012. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**: 2678–2679.

Geisen R, Schmidt-Heydt M, Karolewicz A, 2006. A gene cluster of the ochratoxin A biosynthetic genes in *Penicillium*. *Mycotoxin Research* **22**: 134–141.

Grigoriev IV, Martinez DA, Salamov AA, 2006. Fungal genomic annotation. In: Aurora, DK, Berka, RM, Singh, GB (eds), *Applied Mycology and Biotechnology Volume 6*, Elsevier, Amsterdam, pp. 123–142.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I, 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**(Database issue): D699–704.

Gurevich A, Saveliev V, Vyahhi N, Tesler G, 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**:1072–1075.

Harris JP, Mantle PG, 2001. Biosynthesis of ochratoxins by *Aspergillus ochraceus*. *Phytochemistry* **58**: 709–716.

IARC (International Agency for Research on Cancer), 1993. Ochratoxin A. In: *IARC monographs on the evaluation of carcinogenic risks to humans volume 56*. IARC Press, Lyon, pp. 489–521.

Karolewicz A, Geisen R, 2005. Cloning a part of the ochratoxin A biosynthetic gene cluster of *Penicillium nordicum* and characterization of the ochratoxin polyketide synthase gene. *Systematic and Applied Microbiology* **28**: 588–595.

Keller NP, Turner G, Bennett JW, 2005. Fungal secondary metabolism - from biochemistry to genomics. *Nature Reviews Microbiology* **3**: 937–947.

Kimura M, Tokai T, Takahashi-Ando N, Ohsato S, Fujimura M, 2007. Molecular and genetic studies of *Fusarium* trichothecene biosynthesis: pathways, genes, and evolution. *Bioscience, Biotechnology, and Biochemistry* **71**: 2105–2123.

Kuiper-Goodman T, Hilts C, Billiard SM, Kiparissis Y, Richard ID, Hayward S, 2010. Health risk assessment of ochratoxin A for all age-sex strata in a market economy. *Food additives & contaminants. Part A, Chemistry, analysis, control, exposure & risk assessment* **27**: 212–240.

Larsen TO, Frydenvang K, Frisvad JC, Christophersen C, 1998. UV-guided isolation of alantrypinone, a novel *Penicillium* alkaloid. *Journal of Natural Products* **61**: 1154–1157.

Larsen TO, Franzyk H, Jensen SR, 1999. UV-guided isolation of verrucines A and B, novel quinazolines from *Penicillium verrucosum* structurally related to anacine from *Penicillium aurantiogriseum*. *Journal of Natural Products* **62**: 1578–1580.

Larsen TO, Svendsen A, Smedsgaard J, 2001a. Biochemical characterization of ochratoxin A-producing strains of the genus *Penicillium*. *Applied and Environmental Microbiology* **67**: 3630–3635.

Larsen TO, Petersen BO, Duus JO, 2001b. Lumpidin, a novel biomarker of some ochratoxin A producing *Penicillia*. *Journal of Agricultural and Food Chemistry* **49**: 5081–5084.

Lee HJ, Ryu D, 2015. Significance of Ochratoxin A in Breakfast Cereals from the United States. *Journal of Agricultural and Food Chemistry* **63**: 9404–9409.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Lund F, Frisvad JC, 2003. *Penicillium verrucosum* in wheat and barley indicates presence of ochratoxin A. *Journal of Applied Microbiology* **95**: 1117–1123.

Miller JD, Schaafsma A, Bhatnagar D, Bondy G, Carbone I, Harris L, Harrison G, Munkvold G, Oswald I, Pestka J, Sharpe L, Sumarah M, Tittlemier S, Zhou T, 2014. Mycotoxins that affect the North American agri-food sector: state of the art and directions for the future. *World Mycotoxin Journal* **7**: 63–82.

Nguyen KTN, Ryu D, 2014. Concentration of ochratoxin A in breakfast cereals and snacks consumed in the United States. *Food Control* **40**: 140-144.

Nielsen KF, Smedsgaard J, 2003. Fungal metabolite screening: database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography–UV–mass spectrometry methodology. *Journal of Chromatography A* **1002**: 111–136.

Nikolenko SI, Korobeynikov AI, Alekseyev MA, 2013. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* **14** Suppl 1:S7.

O'Callaghan J, Caddick MX, Dobson AD, 2003. A polyketide synthase gene required for ochratoxin A biosynthesis in *Aspergillus ochraceus*. *Microbiology* **149**: 3485–3491.

Parra G, Bradnam K, Korf I, 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**: 1061–1067.

Rahbaek L, Sperry S, Frisvad JC, Larsen TO, 2003. PC-2, LL-P888gamma and some novel analogue alpha-pyrones from *Penicillium nordicum*, *P. verrucosum* and *P. olsonii*. *Biochemical Systematics and Ecology* **31**: 313–317.

Samson RA, Houbraken J, Kuijpers A, Frank JM, Frisvad JC, 2004. New ochratoxin A or sclerotium producing species in *Aspergillus* section Nigri. *Studies in Mycology* **50**: 45–61.

Schmidt-Heydt M, Baxter E, Geisen R, Magan N, 2007. Physiological relationship between food preservatives, environmental factors, ochratoxin and *otapksPV* gene expression by *Penicillium verrucosum*. *International Journal of Food Microbiology* **119**: 277–283.

Schwarzer D, Finking R, Marahiel MA, 2003. Nonribosomal peptides: from genes to products. *Natural Product Reports* **20**: 275–287.

Sonjak S, Ličen M, Frisvad JC, Gunde-Cimerman N, 2011. Salting of dry-cured meat—A potential cause of contamination with the ochratoxin A-producing species *Penicillium nordicum*. *Food Microbiology* **28**: 1111–1116.

Sulyok M, Berthiller F, Krska R, Schuhmacher R, 2006. Development and validation of a liquid chromatography/tandem mass spectrometric method for the determination of 39 mycotoxins in wheat and maize. *Rapid Communications in Mass Spectrometry* **20**: 2649–2659.

Takahashi C, Matsushita T, Doi M, Minoura K, Shingu T, Kumeda Y, Numata A, 1995. Fumiquinazolines A-G, novel metabolites of fungus separated from a *Pseudolabrus* marine fish. *Journal of the Chemical Society Perkin Transactions 1*: 2345–2353.

Turgeon BG, Baker SE, 2007. Genetic and genomic dissection of the *Cochliobolus heterostrophus* *Tox1* locus controlling biosynthesis of the polyketide virulence factor T-toxin. *Advances in Genetics* **57**: 219–261.

van Egmond HP, Schothorst RC, Jonker MA, 2007. Regulations relating to mycotoxins in food. *Analytical and Bioanalytical Chemistry* **389**: 147-157.

Varga J, Rigo K, Teren J, Mesterhazy A, 2001. Recent advances in ochratoxin research. I. Production, detection and occurrence of ochratoxins. *Cereal Research Communications* **29**: 85–92.

Varga J, Frisvad JC, Kocsubé S, Brankovics B, Tóth B, Szigeti G, Samson RA, 2011. New and revisited species in *Aspergillus* section Nigri. *Studies in Mycology* **69**: 1-17.

Varga E, Glauner T, Berthiller F, Krska R, Schuhmacher R, Sulyok M, 2013. Development and validation of a (semi-)quantitative UHPLC-MS/ MS method for the determination of 191 mycotoxins and other fungal metabolites in almonds, hazelnuts, peanuts and pistachios. *Analytical and Bioanalytical Chemistry* **405**: 5087–5104.

Visagie CM, Varga J, Houbraken J, Meijer M, Kocsubé S, Yilmaz N, Fotedar R, Seifert KA, Frisvad JC, Samson RA, 2014a. Ochratoxin production and taxonomy of the yellow aspergilli (*Aspergillus* section Circumdati). *Studies in Mycology* **78**: 1–61.

Visagie CM, Houbraken J, Frisvad JC, Hong SB, Klaassen CHW, Perrone G, Seifert KA, Varga J, Yaguchi T, Samson RA, 2014b. Identification and nomenclature of the genus *Penicillium*. *Studies in Mycology* **78**: 343–371.

Yu JH, Keller N, 2005. Regulation of secondary metabolism in filamentous fungi. *Annual Review of Phytopathology* **43**: 437–458.

Tables

Table 1. Secondary metabolite production by strains of *P. thymicola*, *P. nordicum* and *P. verrucosum*. OTA, citrinin and phenylalanine were verified by an authentic standard (marked by an asterisk). The + indicates the compound was detected while the - indicates the compound was not detected. It should be noted that phenylalanine is a primary metabolite important for OTA biosynthesis and not a secondary metabolite.

Table 2. Accession numbers of reference proteins involved in OTA biosynthesis and its putative homologue in *P. thymicola* DAOMC 180753.

Table 3. Gene accession numbers, protein accession numbers and top UnitProt/SwissProt hit for genes in cluster 39.

Figure Captions

Fig 1. Extracted ion chromatogram for OTA (A; 3.54 min; chemical structure illustrated), HRMS spectra in positive mode where the chlorine isotope profile is visualized (B) and HRMS² fragmentation ions for OTA (C) produced by *P. thymicola* DAOMC 180753.

Fig 2. Type I PKS gene cluster in *P. thymicola* DAOMC 180753 putatively involved in OTA biosynthesis.

Table 1.

Secondary Metabolite	<i>P. thymicola</i> DAOMC 180753	<i>P. nordicum</i> DAOMC 158653	<i>P. verrucosum</i> DAOMC 242724
OTA*	+	+	+
OTB	+	+	+
Citrinin*	-	-	+
Alantrypinone	+	-	-
Fumiquinazoline F	+	-	-
Phenylalanine*	+	+	+
PC-2	+	+	+
Sclerotigenin	-	+	-
Lumpidin	-	-	-

Table 2.

Reference protein accession	Species	Study	<i>P. thymicola</i> protein accession	<i>P. thymicola</i> gene accession	% ID	E-value	Reference protein length
gi 81176227 gb AAT65816.2	<i>Penicillium nordicum</i>	Karolewicz & Geisen 2005	jgi Penth1 176637	gene_11104	96.98	0	398
gi 30525891 gb AAP32477.1	<i>Aspergillus ochraceus</i>	O'Callaghan et al. 2003	jgi Penth1 212294	gene_4875	71.38	3.00E-141	502
jgi Aspca3 173482	<i>Aspergillus carbonarius</i>	Gallo et al. 2014	jgi Penth1 212294	gene_4875	70.59	5.00E-173	2541
gi 193811752 emb CAQ16344.1	<i>Aspergillus carbonarius</i>	Gallo et al. 2009	jgi Penth1 168734	gene_3514	70.4	0	729

Table 3.

<i>P. thymicola</i> gene accession	<i>P. thymicola</i> protein accession	antiSMASH gene type
gene_4856	jgi Penth1 153991	Regulatory gene
gene_4857	jgi Penth1 249954	Other
gene_4858	jgi Penth1 153997	Other
gene_4859	jgi Penth1 2433	Biosynthetic gene
gene_4860	jgi Penth1 108108	Transport related gene
gene_4861	jgi Penth1 170113	Biosynthetic gene
gene_4862	jgi Penth1 108125	Biosynthetic gene
gene_4863	jgi Penth1 212277	Other
gene_4864	jgi Penth1 108137	Biosynthetic gene
gene_4865	jgi Penth1 108097	Other
gene_4866	jgi Penth1 108116	Biosynthetic gene
gene_4867	jgi Penth1 108123	Biosynthetic gene
gene_4868	jgi Penth1 182132	Biosynthetic gene
gene_4869	jgi Penth1 108077	Biosynthetic gene
gene_4870	jgi Penth1 212279	Other
gene_4871	jgi Penth1 108111	Other
gene_4872	jgi Penth1 108098	Other
gene_4873	jgi Penth1 182137	Other
gene_4874	jgi Penth1 212286	Other
gene_4875	jgi Penth1 212294	Biosynthetic gene
gene_4876	jgi Penth1 170133	Biosynthetic gene
gene_4877	jgi Penth1 78053	Biosynthetic gene
gene_4878	jgi Penth1 249956	Other
gene_4879	jgi Penth1 249957	Other
gene_4880	jgi Penth1 212313	Other

Figure 1 EPS

Figure 2 EPS colour