Draft Genome Sequence of an Aflatoxigenic Aspergillus Species, A. bombycis

Geromy G. Moore*, Brian M. Mack, Shannon B. Beltz, and Matthew K. Gilbert
Southern Regional Research Center, Agricultural Research Service, United States Department of Agriculture

*Corresponding author: E-mail: geromy.moore@ars.usda.gov.

Accepted: September 19, 2016

Data deposition: The draft genome sequence has been deposited at DDBJ/ENA/GenBank under the accession LYCR00000000. The version described in this article is version LYCR01000000.

Abstract

Aspergillus bombycis was first isolated from silkworm frass in Japan. It has been reportedly misidentified as A. nomius due to their macro-morphological and chemotype similarities. We sequenced the genome of the A. bombycis Type strain and found it to be comparable in size (37 Mb), as well as in numbers of predicted genes (12,266), to other sequenced Aspergilli. The aflatoxin gene cluster in this strain is similar in size and the genes are oriented the same as other B- + G-aflatoxin producing species, and this strain contains a complete but nonfunctional gene cluster for the production of cyclopiazonic acid. Our findings also showed that the A. bombycis Type strain contains a single MAT1-2 gene indicating that this species is likely heterothallic (self-infertile). This draft genome will contribute to our understanding of the genes and pathways necessary for aflatoxin synthesis as well as the evolutionary relationships of aflatoxigenic fungi.

Key words: Aspergillus bombycis, aflatoxins, genome sequence, phylogenomics, mating-type locus.

Introduction

Several Aspergillus species are capable of producing two important secondary metabolites: B and G aflatoxins. B aflatoxins are generally considered more potent mycotoxins than G aflatoxins, but both are considered serious carcinogenic compounds and their purpose or function in nature has yet to be determined (Santini and Ritieni, 2013). Aspergillus bombycis (NRRL 26010) was characterized and associated with the genus' Section Flavi in 2001 (Peterson et al., 2001). Although this fungus likely existed for many years prior to 2001, it may have been subject to repeated misidentifications as A. nomius because it was sampled in the same environment, exhibited similar macro-morphological characters, and produced similar toxic secondary metabolites (Peterson et al., 2001). Aspergillus bombycis has not been reported as a pathogen of animals, and reports of its pathogenesis on agricultural commodities are rare (Ehrlich et al. 2003), but it does appear more often to be associated with insect species since it was first found in the excreta of silkworms in two different Asian countries (Peterson et al., 2001). It is unclear whether the preferred food source for silkworms, the leaves of the white mulberry plant (Morus alba), or the silkworms themselves are the intended host for A. bombycis. By sequencing the genome of the A. bombycis Type strain, it may be possible to better understand how aflatoxin production has evolved. Additionally, we may better understand the biology of this organism that appears to be so often misidentified due to its similarities to multiple Aspergilli.

Comparative Analysis against Sequenced and Annotated Aspergillus Genomes

The draft genome assembly for A. bombycis is 37.5 Mb, comprised of 451 contigs with an N50 of 44 and N50 length of 243,233 bp, a maximum contig size of 1,192,132 bp, and 48.7% G + C content. Additional sequencing quality statistics and predicted genomic information for this Type strain are shown in table 1.

The genome of A. bombycis is comparable in size to other aflatoxigenic species in Section Flavi. Of its 12,226 predicted genes, orthology analysis revealed 917 as unique to A. bombycis. The number of secondary metabolite (SM) clusters within A. bombycis is inferred to be 66 (SMURF) and 202
(antiSMASH), whereas closely-related A. nomius is inferred to contain 62 and 153 SM clusters by SMURF and antiSMASH, respectively. The comparative SM data is shown in Table 2. The reason for the discrepancies between SM counts is because the antiSMASH algorithm is designed to predict 43 types of gene clusters (e.g., Type 1–3 PKS, Nrps, and terpenes), thus it generally provides a more comprehensive list of cluster predictions than SMURF. In contrast to antiSMASH, SMURF conducts cluster predictions for five general SM cluster categories (Khalidi et al., 2010).

Comparatively, the genes within the AF cluster of A. bombycis have the same orientation as in other species in Section Flavi. Its AF cluster also spans approximately the same genomic distance (68.1 kb) as other sequenced B + G producing species such as A. parasiticus (68.3 kb) and A. nomius (68.4 kb). Other SM clusters that are found in common between A. nomius and A. bombycis, as predicted by antiSMASH, include the cluster responsible for the biosynthesis of the neurotoxic and tremorgenic mycotoxin aflatrem (Valdes et al., 1992), the antilisterial/anticancer insectan compound aflavarin (TePaske et al., 2009), and clusters characterized in other Penicillia as well as various Aspergilli (Chang et al., 2009). Whether this gene cluster was inherited through horizontal gene transfer between these two genera, or whether it was inherited from their most recent common ancestor, is unclear. The CPA biosynthesis cluster has been characterized (Chang et al., 2009) for several A. flavus strains and reportedly contains three genes that are responsible for its production: a monoamine oxidase gene (maoA),

The availability of whole-genome data could help to elucidate the origins, or the evolution, of AF-producing fungi. Phylogenetic comparison of A. bombycis with other Aspergillus species, and the outgroup taxa Penicillium chrysogenum, indicate that this species shares a most recent common ancestor with A. nomius (fig. 1). Aspergillus bombycis and A. nomius divergence was earlier than the AFB-producing morphotypes of A. flavus. Sequencing the genomes of more aflatoxigenic species will reveal and refine our understanding of the steps in the evolution of the AF cluster, and offer insights regarding the potential impacts of recombination on this, and other, SM clusters within mycotoxigenic Aspergillus species.

Another toxic secondary metabolite that has been associated with aflatoxin producing species is CPA (Chang and Ehrlich 2011). This compound was first discovered in Penicillium cyclopium (Holzapfel, 1968) and has since been found in other Penicillia as well as various Aspergilli (Chang et al., 2009). Whether this gene cluster was inherited through horizontal gene transfer between these two genera, or whether it was inherited from their most recent common ancestor, is unclear. The CPA biosynthesis cluster has been characterized (Chang et al., 2009) for several A. flavus strains and reportedly contains three genes that are responsible for its production: a monoamine oxidase gene (maoA),

### Table 1

| Genome characteristic | Value |
|-----------------------|-------|
| Assembly size (bp)    | 37,476,653 |
| N50                   | 44 |
| N50 length (bp)       | 243,233 |
| CEGMA % completeness  | 95.16 |
| Average depth         | 53 |
| G+C (%)               | 48.7 |
| Protein coding genes  | 12,266 |
| Protein coding genes >100 amino acids | 12,013 |
| Predicted protein coding sequences >100 amino acids | |
| Gene density (1 gene every n bp) | 3,119 |
| Median gene length (bp) | 1,440 |
| Mean gene length (bp) | 1,703 |
| Average number of exons per gene | 3.27 |

### Table 2

| A. bombycis | A. nomius |
|-------------|-----------|
| Number of identified secondary metabolite clusters | |
| SMURF       | 66        |
| Anti-SMASH  | 202       |
| BIOSYNTHETIC TYPES (antiSMASH) | |
| Polyketide Synthase Type 1 | 21        |
| Polyketide Synthase Type 3 | 3         |
| Nonribosomal Peptide Synthetases (Nrps) | 18       |
| Indole      | 4         |
| Terpene     | 13        |
| Siderophore | 1         |
| PKS Type 1-Type 3 hybrid | 0       |
| Indole-Nrps | 1         |
| Nrps-PKS Type 1 | 6      |
| Nrps-Type 1 PKS-Indole | 1      |
| Other       | 15        |
| ClusterFinder algorithm | |
| Putative    | 112       |
| Saccharide  | 1         |
| Fatty acid  | 5         |
| Fatty acid—Saccharide | 0      |
| Fatty acid—T1 PKS | 0 |

The availability of whole-genome data could help to elucidate the origins, or the evolution, of AF-producing fungi. Phylogenetic comparison of A. bombycis with other Aspergillus species, and the outgroup taxa Penicillium chrysogenum, indicate that this species shares a most recent common ancestor with A. nomius (fig. 1). Aspergillus bombycis and A. nomius divergence was earlier than the AFB-producing morphotypes of A. flavus. Sequencing the genomes of more aflatoxigenic species will reveal and refine our understanding of the steps in the evolution of the AF cluster, and offer insights regarding the potential impacts of recombination on this, and other, SM clusters within mycotoxigenic Aspergillus species.

Another toxic secondary metabolite that has been associated with aflatoxin producing species is CPA (Chang and Ehrlich 2011). This compound was first discovered in Penicillium cyclopium (Holzapfel, 1968) and has since been found in other Penicillia as well as various Aspergilli (Chang et al., 2009). Whether this gene cluster was inherited through horizontal gene transfer between these two genera, or whether it was inherited from their most recent common ancestor, is unclear. The CPA biosynthesis cluster has been characterized (Chang et al., 2009) for several A. flavus strains and reportedly contains three genes that are responsible for its production: a monoamine oxidase gene (maoA), a
dimethylallyl tryptophan synthase gene (dmaT), and a hybrid polyketide synthase and nonribosomal peptide synthetase gene (pks-nrps). In *A. flavus*, the CPA gene cluster is immediately adjacent to the AF cluster. Although there are no reports of *A. bombycis* producing CPA, BLAST of both the nucleotide and protein sequences for *A. flavus* maoA, dmaT and pks-nrps genes yielded sequence identities within the *A. bombycis* genome. Closer examination of these genes, compared with those from a functional CPA cluster, revealed a deletion mutation within the 11.7-kb pks-nrps gene. This single nucleotide deletion, found at position 954, introduces a frameshift stop codon at position 1096, truncating 3541 amino acids. One CPA-negative *A. flavus* strain, a candidate biocontrol strain known as K49, has a substitution mutation in its pks-nrps gene at amino acid 703 that changes a serine (TCA) to a stop codon (TGA) and truncates 3202 amino acids (Chang et al., 2012). The proximity of the CPA gene cluster in *A. flavus* is within 9000 nucleotide bases of the aflatoxin gene cluster on Chromosome III (Chang et al., 2009). However, in *A. bombycis* the genomic distance could not be determined because each gene cluster was located on separate contigs with no overlap. Whether this is because in *A. bombycis* it is much farther between these gene clusters, or because the CPA and AF gene clusters reside on separate chromosomes, or because of a data quality issue, is unclear.

Previous research reported a possible heterothallic existence for most of the species in Section *Flavi*, with each species containing a single mating-type idiomorph (Ramirez-Prado et al., 2008). Our findings from sequencing the genome indicate that the *A. bombycis* Type strain contains a single mating-type (MAT1-2) gene. The ability of this species to outcross has not yet been reported. The other heterothallic *Aspergillus* species, such as *A. flavus* and *A. parasiticus*, have a mating-type gene flanked by two conserved genes in close proximity: one for DNA lyase (APN) and one for cytoskeleton assembly control (SLA). These two genes are consistently found to flank the MAT idiomorph in heterothallic fungi, although the genomic distances between them vary. For example, a previous report determined the relative distances of these genes to the MAT1-2 idiomorph in *A. flavus* and *A. parasiticus* are ~2500 and 2000 bp, respectively (Ramirez-Prado et al., 2008). For the MAT1-2 gene in *A. bombycis*, the distances are 634 and 3030 bp, respectively. The chromosomal location of the mating-type locus in *A. flavus* and *A. parasiticus* is reported to be Chromosome VI (Ramirez-Prado et al., 2008), but this has not yet been determined for *A. bombycis*.

**Materials and Methods**

**Genome Sequence and Annotation**

We sequenced the genome of the *A. bombycis* Type strain using a Personal Genome Machine (PGM) from Life Technologies (Grand Island, New York). Template preparation and sequencing was conducted according to previously reported protocols (Moore et al., 2015). Totals of 6.27 M reads were obtained for this strain. The genome assembly and annotation was performed as reported in Moore et al. (2015) with slight modifications. For example, we used CEGMA (Parra et al., 2007) to identify conserved genes which were used to train Augustus. Maker was then used to integrate ab initio gene predictions with protein homology evidence from the UniRef50 protein database (http://ftp.ncbi.nlm.nih.gov/RefSeq/).
Genomic Comparisons to Various Aspergillus Species

The Antibiotics-Secondary Metabolite Analysis Shell (antiSMASH) and the Secondary Metabolite Unique Regions Finder (SMURF) programs were used to predict SM clusters in A. bombycis (Medema et al., 2011; Khaldi et al. 2010). Default parameters were used except for the incorporation of the ClusterFinder algorithm (Cimermanci et al. 2014). The phylogenomic analysis was performed by detecting orthologous proteins within other fungi using Proteinortho (version 3.8.31; Edgar 2004), and concatenating them into a 2.2-Mb amino acid alignment using GBLOCKS (version 0.91; Castresana 2000). The phylogenetic tree was inferred using RAxML-HPC (version 8.1.17; Stamatakis 2006) with the rREV (Castresana 2000). The phylogenetic tree was inferred using RAxML-HPC (version 8.1.17; Stamatakis 2006) with the rREV (Castresana 2000).

Acknowledgments

This work was supported by research funding from the Agricultural Research Service, an agency within the United States Department of Agriculture.

Literature Cited

Brakhage AA, Browne P, Turner G. 1992. Regulation of Aspergillus nidulans penicillin biosynthesis and penicillin biosynthesis genes acvA and ipnA by glucose. J Bacteriol. 174:3789–3799.

Cimermanci P, et al. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158:412–421.

Davison J, et al. 2012. Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. Proc Natl Acad Sci U S A. 109:7642–7647.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Ehrlich KC, Montalbano BG, Cotty PJ. 2003. Sequence comparison of aflR from different Aspergillus species provides evidence for variability in regulation of aflatoxin production. Fungal Genet Biol. 38:63–74.

Holzapfel CW. 1968. The isolation and structure of cyclopiazonic acid, a toxic metabolite of Penicillium cyclopium Westling. Tetrahedron 24:2101–2119.

Khaldi N, et al. 2010. SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol. 47:736–741.

Lechner M, et al. 2011. Proteinortho: detection of (Co-) orthologs in large-scale analysis. BMC Bioinformatics 12:124.

Mapari SA, Thrane U, Meyer AS. 2010. Fungal polyketide azaphilone pigments as future natural food colorants?. Trends Biotechnol. 28:300–307.

Martinez-Luis S, et al. 2012. Antiparasitic and anticancer constituents of the endophytic fungus Aspergillus sp. strain F1544. Nat Prod Commun. 7:165–168.

Medema MH, et al. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39:W339–W346.

Moore GG, Mack BM, Beltz SB. 2015. Genomic sequence of the aflatoxinogenic filamentous fungus Aspergillus nomius. BMC Genomics 16:551.

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

Peterson SW, Ito Y, Horn BW, Goto T. 2001. Aspergillus bombycis, a new aflatoxinogenic species and genetic variation in its sibling species, A. nomius. Mycologia 93:689–703.

Ramirez-Prado JH, Moore GG, Horn BW, Carbone I. 2008. Characterization and population analysis of the mating-type genes in Aspergillus flavus and Aspergillus parasiticus. Fungal Genet Biol. 45:1292–1299.

Santini A, Riteni A. 2013. Aflatoxins: risk, exposure and remediation. In: Razzaghi-Abyaneh M, editor. Aflatoxins – recent advances and future prospects. InTech, Croatia: Rijeka. DOI: 10.5772/52866.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

TePaske MR, Glover JB, Wicklow DT, Dowd PF. 1992. Aflavirin and β-Aflatrem: new anti-insectan metabolites from the sclerotia of Aspergillus flavus. J Nat Prod. 55:1080–1086.

Valdes JJ, Cameron JE, Cole RJ. 1985. Aflatrem: a tremorgenic mycotoxin with acute neurotoxic effects. Environ Health Perspect. 62:459–463.

Associate editor: Howard Ochman

Advance Access publication September 23, 2016

3300 Genome Biol. Evol. 8(11):3297–3300. doi:10.1093/gbe/eww238