Bioinformatics, 31(7), 2015, 981–985
doi: 10.1093/bioinformatics/btu753
Advance Access Publication Date: 19 November 2014
Discovery Note

OXFORD

Genome analysis

Ustiloxins, fungal cyclic peptides, are ribosomally synthesized in *Ustilaginoidea virens*

Takahiro Tsukui[1,†], Nozomi Nagano[1,†], Myco Umemura[2,*], Toshitaka Kumagai[3], Goro Terai[4], Masayuki Machida[2] and Kiyoshi Asai[1,5,*]

[1]Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064, Japan, [2]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Hokkaido 062-8517, Japan, [3]Fermlab Inc., Tokyo 135-0021, Japan, [4]INTEC Inc., Tokyo 136-8637, Japan, and [5]Department of Computational Biology, Graduate School of Frontier Sciences, the University of Tokyo, Chiba 277-8561, Japan

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Associate Editor: Alfonso Valencia

Received and revised on September 4, 2014; accepted on November 8, 2014

Abstract

Motivation: Ustiloxins A and B are toxic cyclic tetrapeptides, Tyr-Val/Ala-Ile-Gly (Y-V/A-I-G), that were originally identified from *Ustilaginoidea virens*, a pathogenic fungus affecting rice plants. Contrary to our report that ustiloxin B is ribosomally synthesized in *Aspergillus flavus*, a recent report suggested that ustiloxins are synthesized by a non-ribosomal peptide synthetase in *U.virens*. Thus, we analyzed the *U.virens* genome, to identify the responsible gene cluster.
Results: The biosynthetic gene cluster was identified from the genome of *U.virens* based on homologies to the ribosomal peptide biosynthetic gene cluster for ustiloxin B identified from *A.flavus*. It contains a gene encoding precursor protein having five Tyr-Val-Ile-Gly and three Tyr-Ala-Ile-Gly motifs for ustiloxins A and B, respectively, strongly indicating that ustiloxins A and B from *U.virens* are ribosomally synthesized.
Availability and implementation: Accession codes of the *U.virens* and *A.flavus* gene clusters in NCBI are BR001221 and BR001206, respectively.
Contact: umemura-m@aist.go.jp or asai@k.u-tokyo.ac.jp
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Ustiloxins are toxic cyclic peptides that were originally identified from *Ustilaginoidea virens*, a pathogenic fungus affecting rice plant (Koiso *et al.*, 1992, 1994, 1998). The circular moieties of these peptides are composed of Tyr-Val/Ala-Ile-Gly tetrapeptides that are circularized at the side-chains of Tyr and Ile (Koiso *et al.*, 1992, 1994, 1998). Ustiloxin A, a derivative of ustiloxin peptides, contains a circularized tetrapeptide with the composition of Tyr-Val-Ile-Gly (YVIG), whereas ustiloxin B, another derivative of ustiloxin peptides, contains the circularized peptide of Tyr-Ala-Ile-Gly (YAIG). The aromatic rings of the Tyr residues in both ustiloxins A and B are modified by norvaline, a non-protein-coding amino acid, through a sulfoxide bond (Koiso *et al.*, 1994, 1998) (Fig. 1).

Recently, the ustiloxin B biosynthetic gene cluster was experimentally identified from *Aspergillus flavus* using a novel method to predict gene clusters from transcriptome data, MIDDAS-M, and the subsequent validation by liquid chromatography-mass spectrometry analysis of the gene deletion mutants (Umemura *et al.*, 2013, 2014). The ustiloxin B gene cluster contains 15 genes including those encoding a fungal type C6 transcription factor, a major facilitator

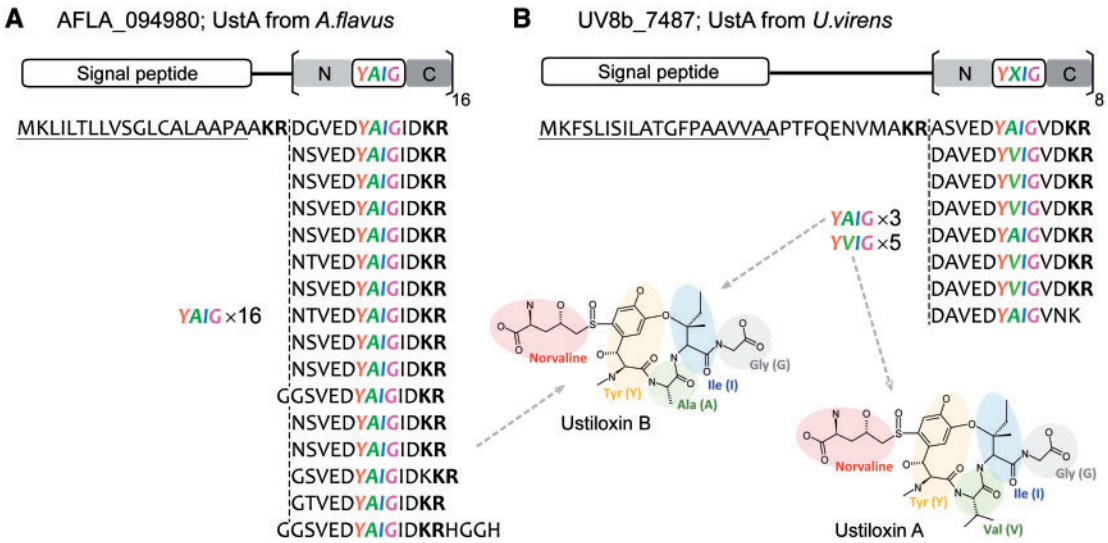

Fig. 1. The amino acid sequence of the precursor protein for ustiloxins, UstA, and the chemical structures of the corresponding ustiloxins. (**A**) *A.flavus*, which contains 16 repeats of the core peptide YAIG (Umemura *et al.*, 2014). (**B**) *U.virens*, which possesses 3 repeats of YAIG and 5 repeats of YVIG. The repeated core peptide sequences of *A.flavus* and *U.virens* completely correspond to the structures of the ustiloxins produced by the two fungi: *A.flavus* produces only ustiloxin B, whereas *U.virens* produces both ustiloxins A and B.

superfamily transporter and cytochrome P450. Until recently, this cyclic peptide was believed to be biosynthesized by the so-called non-ribosomal peptide synthetase (NRPS) because it contains a non-proteinogenic amino acid, norvaline, in its chemical structure. However, bioinformatic analyses of amino acid sequence indicated that no NRPS domains were observed in the gene cluster (Umemura *et al.*, 2013). Instead, a function-unknown gene, *ustA*, with a translated product containing 16 repeats of the tetrapeptide YAIG, which is the exact amino acid sequence of the cyclic moiety of ustiloxin B, was identified in the gene cluster (Umemura *et al.*, 2014) (Fig. 1A). Consequently, ustiloxin B was shown to be the first case of a ribosomally synthesized peptide reported from filamentous fungi (Umemura *et al.*, 2014).

Very recently, Zhang *et al.* (2014) reported the whole genome sequence of *U.virens*, and proposed that an NRPS gene cluster is responsible for biosynthesis of ustiloxins in *U.virens*. However, our analysis strongly suggests that an entirely different gene cluster is responsible for the ustiloxin biosynthesis as ribosomal peptides in *U.virens*, based on homology to the ustiloxin B biosynthetic gene cluster from *A.flavus*.

2 Methods

2.1 Reannotation of coding sequences in the ustiloxin B gene cluster from *A.flavus* by RNA-seq analysis

To reannotate the gene coding sequences (CDSs) in the ustiloxin B biosynthetic gene cluster, we performed RNA-seq analysis using 15 types of transcriptome data for *A.flavus* obtained under different conditions in the NCBI SRA database (http://www.ncbi.nlm.nih.gov/SRA) (accession codes. SRR283857, SRR283858, SRR495778, SRR495942, SRR495943, SRR544871, SRR544872, SRR544873, SRR610531, SRR610532, SRR610533, SRR610534, SRR610535, SRR610537 and SRR610538). We mapped the sequence reads >30 bp after trimming the bases with average quality values of <20 over 7 bases, without unidentified bases, on the genome sequence of *A.flavus*, GenBank EQ963472-EQ96623211, using the Bowtie (Langmead *et al.*, 2009) and Tophat (Trapnell *et al.*, 2009) softwares on a DELL server with Ubuntu 13.10 (CPU, Intel Xeon E52643; memory, 256 GB; hard disk, 40 TB). We visually observed the mapping results in the region of the ustiloxin cluster, AFLA_094940–AFLA_095110 (Umemura *et al.*, 2013, 2014), using the IGV software (Robinson *et al.*, 2011) and determined the region without inconsistencies in the start and stop codons as the gene coding regions.

The ustiloxin B biosynthetic gene cluster from *A.flavus* contained 15 genes; two pairs of two genes, AFLA_095000–AFLA_095010 and AFLA_095080–AFLA_095090, were concatenated to single genes; AFLA_094970 was omitted because it did not express; the CDSs for eight genes, AFLA_094950, AFLA_094960, AFLA_094990, AFLA_095020, AFLA_095050, AFLA_095070, AFLA_095100 and AFLA_095110, were corrected. The other five genes in the ustiloxin cluster, AFLA_094940, AFLA_094980, AFLA_095030, AFLA_095040 and AFLA_095060, were identical to those in the NCBI database after RNA-seq analysis. The functions of the reannotated genes were analyzed by Blastp search (Altschul *et al.*, 1997) against UniProtKB (Magrane and Consortium, 2011) (Table 1).

2.2 Identification of the ustiloxin gene cluster from the *U.virens* genome

A tBLASTx search (Altschul *et al.*, 1997) was performed against the *U.virens* genome sequence of UV-8b (NCBI accession nos. JHTR01000001–JHTR01000449) (Zhang *et al.*, 2014) with an *E*-value cutoff of <1e-4 using all 15 gene sequences of the ustiloxin B gene cluster from *A.flavus* identified by the RNA-seq analysis as queries. The region detected by the tBLASTx search in the *U.virens* genome was reannotated by the pipeline used for the *Aspergillus oryzae* and *Umbelopsis isabellina* genomes (Machida *et al.*, 2005; Takeda *et al.*, 2014). Briefly, the pipeline uses two different types of gene finding methods, ALN (Gotoh, 2000) and GlimmerHMM (Majoros *et al.*, 2004). The ALN method predicts gene structures based on homologies to known genes. In this case, the 15 ustiloxin biosynthetic genes from *A.flavus*, genes encoded in the *A. oryzae* genome and genes in the UniProtKB database (Magrane and Consortium, 2011) were used as the known genes. The prediction of

Table 1. The *U.virens* gene cluster homologous to the ustiloxin biosynthetic cluster from *A.flavus*

Gene ID by us[a]	Locus in *U.virens* scaffold JHTR01000063.1	Gene ID in *U.virens* genome sequence by Zhang *et al.*(2014)[b]	Homologue in *A.flavus* ustiloxin cluster[c]	*E*-value by Blastp search	Functional annotation based on a Blastp search against the UniProtKB database
uv_ustT	complement(40534..42355)	—	*ustT*	1.0e-130	MFS multidrug transporter
uv_ustS	complement(44807..45532)	UV8b_7484	*ustS*	1.0e-106	Glutathione S-transferase
uv_ustM	46841..47779	UV8b_7485	*ustM*	4.0e-97	SAM-dependent methyltransferase
uv_ustR	complement(47837..49772)	UV8b_7486	*ustR*	1.0e-113	C6 transcription factor
uv_ustC	50823..52666	—	*ustC*	0.0e+00	Cytochrome P450
uv_ustA	53637..54093	UV8b_7487	*ustA*	4.0e-49	Ustiloxin precursor protein with signal peptide
uv_ustYa	complement(55577..56554)	UV8b_7488	*ustYa*	4.0e-85	Hypothetical protein
uv_ustYb	complement(57163..58117)	—	*ustYb*	4.0e-78	Hypothetical protein
uv_ustQ	58561..59854	UV8b_7490	*ustQ*	1.0e-157	Tyrosinase
uv_ustF1	complement(59909..61608)	UV8b_7491	*ustF1*	1.0e-166	Flavin-containing monooxygenase
uv_ustF2	complement(61980..63659)	UV8b_7492	*ustF2*	0.0e+00	Flavin-containing monooxygenase
uv_ustD	64167..65343	UV8b_7493	*ustD*	0.0e+00	PLP-dependent cysteine desulfurase
uv_ustO	complement(65730..66547)	UV8b_7494	*ustO*	1.0e-103	Pyridoxamine 5′-phosphate oxidase-like enzyme
—[d]	—	—	*ustP*	—	Peptidase S41 family protein
—[d]	—	—	*ustH*	—	Gamma-glutamyltranspeptidase

[a]Accession no. of the *U.virens* gene cluster in NCBI is BR001221, with the information of gene-coding regions.
[b]Gene-coding regions are not identical to those annotated by us.
[c]Accession no. of the *A.flavus* gene cluster in NCBI is BR001206, with the information of gene-coding regions.
[d]Genes identified only from *A.flavus*.

gene structure by GlimmerHMM depends on the statistical features of genes, such as frequencies of codons in exons and nucleotides around splice sites. Regarding the precursor gene, uv_*ustA*, UV8b_7487 was adopted because it retains the signal peptide. Because the sequence of UV8b_7487 is largely different from the sequence of *A.flavus ustA* in the repeat count of the core peptide (Fig. 1), it is difficult to correctly align the *A.flavus ustA* sequence especially at the 5′ terminal against *U.virens* genome. Using the amino acid sequences of the reannotated genes in the sequence region of the *U.virens* genome as queries, Blastp search (Altschul *et al.*, 1997) was performed against the 15 *A.flavus* ustiloxin genes after the RNA-seq analysis to evaluate homologies between the two gene clusters from *U.virens* and *A.flavus*, and against the UniProtKB (Magrane and Consortium, 2011) database to annotate their function (Table 1).

2.3 NRPS analyses

The NRPS gene clusters in the *U.virens* scaffold of JHTR01000005 (Zhang *et al.*, 2014), which contains the gene cluster Zhang *et al.* (2014) believed to be the ustiloxin biosynthetic cluster, were analyzed using the antiSMASH program (Blin *et al.*, 2013; Medema *et al.*, 2011). One NRPS gene cluster from UV8b_1483 to UV8b_1497 was detected, including two and one NRPS A domains in UV8b_1490 and UV8b_1491, respectively. The amino acid sequence of UV8b_1490 was aligned with the sequence of the destruxins (Dtxs) NRPS gene from *Metarhizium anisopliae* (UniProtKB; E9FCP4) using the MAFFT program (Katoh *et al.*, 2005).

3 Results and discussion

3.1 Identification of the ustiloxin biosynthetic gene cluster in the *U.virens* genome

Our sequence analysis revealed that the *U.virens* genome sequence (Zhang *et al.*, 2014) contains a gene cluster homologous to the ustiloxin B biosynthetic gene cluster in *A.flavus* (Fig. 2A). The *U.virens* gene cluster is composed of 13 genes, out of which 10 genes of UV8b_7484–UV8b_7494 were already annotated (Zhang *et al.*, 2014) but three other genes newly annotated by us, all of which are highly homologous to the genes in the ustiloxin B biosynthetic gene cluster from *A.flavus* (Table 1). The order and direction of the genes in *U.virens* are different from those in *A.flavus*, but the components of the cluster are the same in both genomes, except the former lacks two peptidase genes, *ustP* and *ustH*.

In *U.virens*, UstA, the precursor protein for ustiloxins, has five YVIG motifs and three YAIG motifs for ustiloxin A and ustiloxin B, respectively, whereas 16 YAIG motifs have been identified in UstA from *A.flavus* (Umemura *et al.*, 2014) (Fig. 1). The sequence characteristics are fairly consistent; *U.virens* produces both ustiloxin A and B (Koiso *et al.*, 1992, 1994), whereas *A.flavus* produces only ustiloxin B (Umemura *et al.*, 2013, 2014). This report indicates that two types of ustiloxins are synthesized from a single precursor protein, UstA, which is the first identification of the precursor protein for the ribosomal peptide ustiloxin A. Our results strongly suggest that the cyclic peptide ustiloxins are not synthesized by NRPS but are instead ribosomally synthesized in *U.virens*, in the same manner as in *A.flavus*.

3.2 Previously reported NRPS as ustiloxin biosynthetic gene

In the recent article on the *U.virens* genome sequence, Zhang *et al.* (2014) suggested that ustiloxins are synthesized by an NRPS and proposed that an NRPS gene cluster was responsible for ustiloxin biosynthesis (Fig. 1B). They selected the UV8b_1490 gene as the NRPS for ustiloxin biosynthesis because it is similar to the NRPS for Dtxs biosynthesis in *Metarhizium robertsii*, an insect pathogenic fungus (Wang *et al.*, 2012; Zhang *et al.*, 2014). Although Zhang *et al.* (2014) insisted that ustiloxin compounds are similar to Dtxs, they are composed of an α-hydroxyisocaproic acid and the five residues of L-proline (Pro) or pipecolic acid (Pip), L-isoleucine (Ile) or L-valine (Val), *N*-methyl-L-valine (MeVal), *N*-methyl-L-alanine

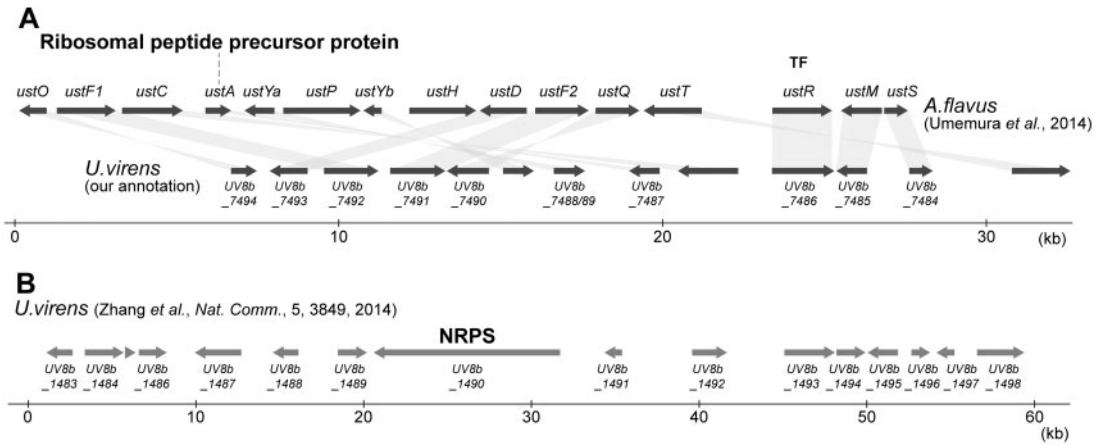

Fig. 2. The ustiloxin biosynthetic gene cluster in *A.flavus* and *U.virens*. (**A**) The cluster experimentally identified in *A.flavus* and its homologous region in *U.virens*. (**B**) The putative NRPS cluster for ustiloxin biosynthesis suggested by Zhang *et al.*

(MeAla) and *β*-alanine (*β*-Ala) (Wang *et al.*, 2012), indicating an entirely different amino acid composition from that of ustiloxins (Supplementary Figure S1). Furthermore, the NRPS gene (Uv8b_1490) suggested by Zhang *et al.* (2014) contains only two NRPS adenylation (A) domains according to an antiSMASH analysis (Blin *et al.*, 2013; Medema *et al.*, 2011). The amino acid sequence of UV8b_1490 only shared 16% similarity with the NRPS for Dtxs (UniProtKB; E9FCP4), which contains six NRPS A domains, according to the MAFFT software (Katoh *et al.*, 2005). Even including the neighbouring gene UV8b_1491, which contains an A domain, only three A domains of NRPS in total can be identified in the gene cluster (Supplementary Figure S1 (B)). With only three A domains, the NRPS does not synthesize a tetrapeptide, nevertheless the main cyclic component of both ustiloxins A and B is composed of a tetrapeptide. Instead of NRPS domains, we identified a gene corresponding to the precursor protein UstA, which contains the cyclic moiety of ustiloxins A and B (YVIG and YAIG, respectively), in the gene cluster from the *U.virens* genome. Therefore, we conclude that ustiloxins are ribosomally synthesized in *U.virens*, in the same manner as in *A.flavus*.

3.3 Comparison of the two ustiloxin gene clusters from *U.virens* and *A.flavus*

The ustiloxin gene cluster from *U.virens* exhibits some differences from the *A.flavus* cluster. The primary difference is that the two genes encoding peptidases, *ustH* and *ustP*, were not identified within the ustiloxin gene cluster in *U.virens*. The *ustH* and *ustP* genes encode a glutathione hydrolase and peptidase S41 family protein, respectively, both of which are major enzymes in fungi. Thus, the functions of UstH and UstP essential for the ustiloxin biosynthesis in *A.flavus* were considered to be compensated by the gene products involved in basic metabolism in the case of *U.virens*.

Another significant difference is the order of the genes in the cluster between *U.virens* and *A.flavus* (Fig. 2). Only a set of three genes in the *U.virens* cluster, UV8b_7486, UV8b_7485 and UV8b_7484, are conserved in the same order as in the *A.flavus* cluster. Both clusters include two pairs of homologous genes, *ustF1-ustF2* and *ustYa-ustYb*; however, the genes in each pair are located adjacent to each homologue in the *U.virens* cluster, whereas they are located separately in the *A.flavus* cluster (Fig. 2). These homologues may represent a duplication event in *U.virens* and may have separated during the evolution of *A.flavus*. Comparing the ustiloxin gene clusters from *U.virens* and *A.flavus* will lead to further characterization of the ustiloxin biosynthetic pathway.

Acknowledgement

We thank American Journal Experts for proofing the manuscript.

Funding

This work was supported by the commission for Development of Artificial Gene Synthesis Technology for Creating Innovative Biomaterial from the Ministry of Economy, Trade and Industry (METI), Japan.

Conflict of Interest: none declared.

References

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Blin,K. *et al.* (2013) antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.*, **41**, W204–W212.

Gotoh,O. (2000) Homology-based gene structure prediction: simplified matching algorithm using a translated codon (tron) and improved accuracy by allowing for long gaps. *Bioinformatics*, **16**, 190–202.

Katoh,K. *et al.* (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.*, **33**, 511–518.

Koiso,Y. *et al.* (1992) Ustiloxin: a phytotoxin and a mycotoxin from false smut balls on rice panicles. *Tetrahedron Lett.*, **33**, 4157–4160.

Koiso,Y. *et al.* (1994) Ustiloxins, antimitotic cyclic peptides from false smut balls on rice panicles caused by *Ustilaginoidea virens*. *J. Antibiot. (Tokyo)*, **47**, 765–773.

Koiso,Y. *et al.* (1998) Isolation and structure of an antimitotic cyclic peptide, ustiloxin F: chemical interrelation with a homologous peptide, ustiloxin B. *J. Antibiot. (Tokyo)*, **51**, 418–422.

Langmead,B. *et al.* (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

Machida,M. *et al.* (2005) Genome sequencing and analysis of *Aspergillus oryzae*. *Nature*, **438**, 1157–1161.

Magrane,M. and Consortium,U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database*, **2011**, bar009.

Majoros,W.H. *et al.* (2004) TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics*, **20**, 2878–2879.

Medema,M.H. *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.

Robinson,J.T. *et al*. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.

Takeda,I. *et al*. (2014) Genome sequence of the mucoromycotina fungus *Umbelopsis isabellina*, an effective producer of lipids. *Genome Announc.*, **2**, e00071–e00014.

Trapnell,C. *et al*. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.

Umemura,M. *et al*. (2013) MIDDAS-M: motif-independent de novo detection of secondary metabolite gene clusters through the integration of genome sequencing and transcriptome data. *PLoS One*, **8**, e84028.

Umemura,M. *et al*. (2014) Characterization of the biosynthetic gene cluster for the ribosomally synthesized cyclic peptide ustiloxin B in *Aspergillus flavus*. *Fungal Genet. Biol.*, **68**, 23–30.

Wang,B. *et al*. (2012) Unveiling the biosynthetic puzzle of destruxins in *Metarhizium* species. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 1287–1292.

Zhang,Y. *et al*. (2014) Specific adaptation of *Ustilaginoidea virens* in occupying host florets revealed by comparative and functional genomics. *Nat. Commun.*, **5**, 3849.