Supplementary Methods

Evolution, structure and function of the biosynthetic gene cluster of myriocin, a potent inhibitory sphingolipid

Ben Rutter[1], Michael Herrera[2], Gustavo Perez Ortiz[2], Claudio Greco[3], Ben Ashley[2], Hai Deng[4], Abigail Hay[1,5], Cameron Bedford[6], Marius Wenzel[7,8], Stephen Mondo[9], Zachary Konkel[10], Jason C Slot[10], Rhys A. Farrer[6], Dominic Campopiano*[2] & Alexandra C Brand*‡[1,6,11]

[1] School of Medicine, Medical Sciences & Nutrition, University of Aberdeen, Aberdeen, UK

[2] School of Chemistry, University of Edinburgh, Edinburgh, UK

[3] Department of Biosciences, Swansea University, Swansea, UK

[4] School of Chemistry, University of Aberdeen, Aberdeen, UK

[5]The Pirbright Institute, Woking, UK

[6] Medical Research Council Centre for Medical Mycology at the University of Exeter, Exeter, UK

[7] Centre for Genome-Enabled Biology & Medicine, University of Aberdeen, Aberdeen, UK

[8]School of Biological Sciences, University of Aberdeen, Aberdeen, UK

[9]Joint Genome Institute, Walnut Creek, CA 94598, USA.

[10]Department of Plant Pathology, The Ohio State University, USA

[11]Living Systems Institute, University of Exeter, Exeter, UK

*Co-senior authors

‡Corresponding author: Alexandra C. Brand: a.brand@exeter.ac.uk

Genome sequencing: gDNA (500 ng) was cleaned using a QIAGEN QIAamp kit. Quantification and quality was checked using a Qubit dsDNA HS Assay on Qubit Fluorometer 2.0 and Agilent gDNA ScreenTape on an Agilent 2200, TapeStation respectively.

High molecular weight (hmw) gDNA was fragmented to ~ 350 bp using the Diagenode Bioruptor Pico. *M. sterilia* lmw gDNA was not fragmented. Fragmentation was confirmed with HSD1000 ScreenTape on the Agilent 2200 TapeStation. Sequencing libraries were prepared using the Illumina TruSeq Nano DNA Library Preparation Kit. The libraries were quantified using the D1000 ScreenTape on the Agilent 2200 TapeStation in conjunction with qPCR and the Kapa Biosciences Complete kit for Illumina library quantification (Universal) on an Illumina Eco qPCR machine. The libraries were diluted and loaded at a ratio of 50 % *I. sinclairii* and 50 % *M. sterilia* material on the Illumina MiSeq and sequenced with the Illumina MiSeq v3 600 cycle kit. A single Oxford NanoPore MinION library was prepared from fresh independent DNA extracts, using the native barcoding (EXP-NBD103) and ligation kits (SQK-LSK108), and sequenced on a single MinION R9.4 flowcell (FLO-MIN106) for 48 h.

Genome assembly The raw Illumina data were examined in *FASTQC* 0.11.3[1] and processed with *TRIM_GALORE* 0.4.0[2] to remove Illumina sequence adapters and trim poor-quality bases with a phred score below 30. The trimmed Illumina data were used to estimate genome size, levels of heterozygosity and duplication using the k-mer coverage method. K-mers of 21 bp length were counted using *KMC* 3.0.0[3] and the counts were analysed in *GENOMESCOPE*[4]. The raw data were basecalled, demultiplexed and quality-filtered in *ALBACORE* 2.3.1 (Oxford NanoPore Technologies), followed by adapter trimming and another round of demultiplexing in *PORECHOP* 0.2.3 (https://github.com/rrwick/Porechop). These filtering steps retained 1,126,913 reads for *I. sinclairii* (1.106 Gbp) and 2,159,418 reads for *M. sterilia* (2.19 Gbp).

A hybrid-assembly strategy was used to combine the benefits of short-read and long-read sequencing technologies. The Illumina data were assembled to short high-quality contigs using the *PLATANUS* 1.2.4 assembler[5] with initial k-mer size of 32. Hybrid assemblies were then generated by extending and joining these contigs with the MinION reads using the *DBG2OLC* hybrid assembler[6] followed by consensus generation with two rounds of *RACON* 1.3.1[7] and consensus polishing with five rounds of *PILON* 1.20[8]. In parallel, MinION-only assemblies were generated in *CANU* 1.7[9] and polished with *PILON* for five rounds. MinION-only assemblies were merged with the hybrid assemblies using *QUICKMERGE*[10], followed by a final round of contig extension and repeat-resolution with MinION reads using *FINISHERSC*[11]. The final merged assemblies were then polished a final time with two rounds of *PILON*.

Assembly quality statistics were obtained using *QUAST*[12], *QUALIMAP* 2.1.1[13] and *BUSCO*[14] with the OrthoDB v8 fungal reference set of 290 single-copy genes. Assembly redundancy/duplications with minimum 90 % sequence identity across at least 70 % contig length were quantified using REDUNDANS 0.14a[15].

Myriocin detection by LC-MS - analytical instrument and method:

Ultra-high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UHPLC–QTOFMS) was performed on an Agilent Infinity 1290 UHPLC system (Agilent Technologies, Santa Clara, CA, USA). Separation was achieved on a 2.1 x 50 mm, 2.5 µm, XBridge® BEH C18 column (Waters) fitted with 2.1 x 5 mm, 2.5 µm, XBridge® BEH C18 V-Gd cartridge (Waters) held at 25 °C. The sample (1 µL) was eluted at a flow rate of 0.4 mL/min using a linear gradient from 30% acetonitrile (LC-MS grade) in Milli-Q water buffered with 0.1% formic acid (gradient steps on

Table S1). Mass spectrometry (MS) detection was performed on an Agilent 6545 LC/Q-TOF equipped with Agilent Dual Jet Stream electrospray ion source (ESI) with a drying gas temperature of 320 °C, a gas flow of 8 L/min, sheath gas temperature of 350 °C and flow of 11 L/min. Capillary voltage was set to 3500 V and nozzle voltage to 1000 V in positive mode. MS spectra were recorded as centroid data, at an *m/z* of 100–1700, with the scan rate set to 1 spectra/sec. Data were handled using Agilent MassHunter Qualitative Analysis software (Agilent Technologies, Santa Clara, CA).

Table S1: Gradient steps for 30% acetonitrile LC-Method. A = MeCN and B = H_2O + 0.1% formic acid.

Time [min]	A [%]	B [%]	Flow [mL/min]
0.00	70.	30	0.4
1.00	70	30	0.4
8.00	2	98	0.4
9.50	2	98	0.4
9.60	70	30	0.4
10.00	70	30	0.4

References

1 Andrews, S. FastQC: a quality control tool for high throughput sequence data. *(No Title)* (2010).

2 Krueger, F. Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. *Babraham Institute* (2015).

3 Kokot, M., Długosz, M. & Deorowicz, S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* **33**, 2759-2761 (2017).

4 Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202-2204 (2017).

5 Kajitani, R. *et al.* Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome research* **24**, 1384-1395 (2014).

6 Ye, C., Hill, C., Wu, S., Ruan, J. & Ma, Z. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci Rep 6: 31900. *North Pacific individual showing demographic changes from* **2** (2016).

7 Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome research* **27**, 737-746 (2017).

8 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* **9**, e112963 (2014).

9 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* **27**, 722-736 (2017).

10 Chakraborty, M., Baldwin-Brown, J. G., Long, A. D. & Emerson, J. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic acids research* **44**, e147-e147 (2016).

11 Lam, K.-K., LaButti, K., Khalak, A. & Tse, D. FinisherSC: a repeat-aware tool for upgrading de novo assembly using long reads. *Bioinformatics* **31**, 3207-3209 (2015).

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

13 Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292-294 (2016).

14 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).

15 Pryszcz, L. P. & Gabaldón, T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic acids research* **44**, e113-e113 (2016).