Title
The genome sequence of four isolates from the family Lichtheimiaceae

Journal
Pathogens and Disease, 73(5)

ISSN
0928-8244

Authors
Chibucos, Marcus C
Etienne, Kizee A
Orvis, Joshua
et al.

Publication Date
2015-07-01

DOI
10.1093/femspd/ftv024

Peer reviewed

FEMS Pathogens and Disease, 73, 2015, ftv024

doi: 10.1093/femspd/ftv024
Advance Access Publication Date: 9 April 2015
Shortomics

SHORTOMICS

The genome sequence of four isolates from the family *Lichtheimiaceae*

Marcus C. Chibucos[1,2], Kizee A. Etienne[3], Joshua Orvis[1], Hongkyu Lee[4], Sean Daugherty[1], Shawn R. Lockhart[3], Ashraf S. Ibrahim[4,5] and Vincent M. Bruno[1,2,*]

[1]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA, [2]Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201, USA, [3]Centers for Disease Control and Prevention, Atlanta, GA 30333, USA, [4]The Division of Infectious Diseases, Los Angeles Biomedical Research Institute at Harbor-University of California Los Angeles (UCLA) Medical Center, Torrance, CA 90502, USA and [5]David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA

*Corresponding author: Institute for Genomes Sciences, University of Maryland School of Medicine, 801 W. Baltimore Street, Baltimore, MD 21201, USA. Tel: +1-410-706-0812; Fax: +1-410-706-1482; E-mail: vbruno@umaryland.edu

One sentence summary: We present four genome sequences of isolates belonging to the family *Lichtheimiaceae*, which cause devastating fungal infections.

Editor: David Rasko

ABSTRACT

This study reports the release of draft genome sequences of two isolates of *Lichtheimia corymbifera* and two isolates of *L. ramosa*. Phylogenetic analyses indicate that the two *L. corymbifera* strains (CDC-B2541 and 008-049) are closely related to the previously sequenced *L. corymbifera* isolate (FSU 9682) while our two *L. ramosa* strains CDC-B5399 and CDC-B5792 cluster apart from them. These genome sequences will further the understanding of intraspecies and interspecies genetic variation within the *Mucoraceae* family of pathogenic fungi.

Keywords: *Lichtheimia*; Mucorales; mucormycosis; *Rhizopus*; fungal phylogeny; fungal genomics

THE GENUS LICHTHEIMIA

Lichtheimia (formerly *Absidia*) is a genus of saprotrophic zygomycetous fungi known to cause mucormycosis in human hosts. Although less prevalent than infections caused by *Aspergillus* or *Candida*, there has been an increase in reports of *Lichtheimia corymbifera* infections among immunocompromised patients (Schwartze and Jacobsen 2014). *Lichtheimia* species are the second and third most isolated organisms from patients with mucormycosis in Europe and worldwide, respectively (Roden *et al.* 2005; Alvarez *et al.* 2009; Skiada *et al.* 2011; Lanternier *et al.* 2012). Various studies have examined the underlying reasons behind the differences in clinical representation among *Lichtheimia* strains. For example, Schwartze *et al.* (2012) evaluated the virulence potential of 46 *Lichtheimia* isolates, representing all five species, in a chicken embryo model of infection. *Lichtheimia ramosa* has also been shown to be a primary infective agent in a burn victim, although treatment with amphotericin B was effective (Kaur *et al.* 2014). While *Lichtheimia* species tend to be morphologically and genetically distinct, they often share very similar antifungal drug susceptibilities. Recent studies continue to implicate *Lichtheimia* species in cutaneous (Cateau *et al.* 2013;

Received: 24 February 2015; **Accepted:** 23 March 2015

Table 1. General genome and predicted proteome statistics for two *L. corymbifera* and two *L. ramosa* strains sequenced by IGS, with *L corymbifera* FSU 9682 shown for comparison.

Strain information Strain name	CDC-B2541	008–049	FSU 9682	CDC-B5399	CDC-B5792
Genus and species	*L. corymbifera*	*L. corymbifera*	*L. corymbifera*	*L. ramosa*	*L. ramosa*
Sequencing & assembly					
Genome coverage	86.17×	97.12×	–	47.23x	41.57x
Assembly statistics					
Contig length (MB)	36.62	36.58	33.53	45.57	42.42
No. of reads (mate-pair)	35 688 332	31 601 382	–	29 414 032	21 456 838
No. of reads (paired-end)	33 695 768	37 099 364	–	44 712 938	18 208 190
Contigs (no.)	1176	1626	589	3968	3694
N50 contig length (nt)	130 684	118 022	66 718	33 650	35 903
Scaffold length (MB)	36.65	36.66	33.6	45.66	42.52
Scaffolds (no.)	935	1306	209	3191	2782
N50 scaffold length (nt)	207 011	176 654	367 562	68 280	90 825
G+C content	43.5%	43.4%	43.4%	41.1%	41.0%
Genomic features					
CEGMA percent completeness[a]	98.4%/98.8%	98.0%/98.8%	–	97.6%/99.2%	96.0%/99.6%
Known fungal repeats detected	2.58%	2.94%	–	1.90%	1.72%
Predicted protein-coding genes					
Predicted genes	9607	10 800	12 379	14 426	13 483
Gene length (mean)	1648.1	1773.4	N/A	1455.6	1516.6
Average coding sequence size (nt)	1294.5	1450.3	1287	1173.9	1211.6
Exons per mRNA (mean)	5.4	5.6	–	5.0	5.1
Total introns	41 816	49 599	48 663	57 423	54 924
Introns per gene (mean)	4.4	4.6	3.9	4.0	4.1
Predicted non-coding RNA genes					
Predicted tRNA	164	193	174	278	259
Predicted rRNA	2	6	*	3	4

[a]Complete/partial core eukaryotic genes.
*rRNA reported amongst other types of non-coding RNA.

Poirier *et al.* 2013) and other infections (Bellanger *et al.* 2010; Irtan *et al.* 2013; Kutlu *et al.* 2014). Two *Lichtheimia* genomes have recently been published (Linde *et al.* 2014; Schwartze *et al.* 2014). As more is learned about the physiological and molecular mechanisms of pathogenesis in *Lichtheimia*, it is important to have a deeper understanding of the underlying genetics and genomics of this important group of opportunistic pathogens.

GENOME SEQUENCING

In this study, we have sequenced two *L. corymbifera* isolates (CDC-B2541 and 008-049) and two *L. ramosa* isolates (CDC-B5399 and CDC-B5792). *Lichtheimia corymbifera* CDC-B2541 was isolated as a plate contaminant in 1977 in Wisconsin, USA, while isolate 008-049 was isolated from a human in a 2008 Deferasirox-AmBisome Therapy for Mucormycosis (DEFEAT) study (Spellberg *et al.* 2012). *Lichtheimia ramosa* CDC-B5792 was isolated from human sputum in 1997 in New Mexico, USA, whereas isolate CDC-B5399 was isolated as a gluteal abscess from India in 1993. DNA was extracted from fungi grown on Sabouraud's Dextrose agar using the GeneRite Kit (Carlsbarg, CA) or the OmniPrep Kit (GBiosciences). The genome sequence of each isolate was generated at the Institute for Genome Sciences (IGS) Genomics Resource Center (Baltimore, MD) using a combination of paired-end libraries (average insert size of 459 bp) and mate-pair (3 kb) libraries on the Illumina HiSeq 2000. We generated an average of 33.4 million sequence reads from each of the paired-end libraries and 29.5 million sequence reads from each of the mate-pair libraries (Table 1). The draft genome data were assembled using the MaSuRCA v.1.9.2 genome assembler (Zimin *et al.* 2013). The relevant statistics from the genome assemblies and annotations are summarized in Table 1. The resulting *L. corymbifera* genome assemblies contained an average of 1401 contigs per genome. The *L. ramosa* genome assemblies contained 3831 contigs on average. The average estimated coverage was 91.6 × for *L. corymbifera* and 44.4 × for *L. ramosa*.

STRUCTURAL & FUNCTIONAL ANNOTATION

Structural and functional annotation were performed with the IGS Eukaryotic Annotation Pipeline protocol 1.0 at the IGS Informatics Resource Center (Baltimore, MD). We generated 439 million RNA-seq reads from isolate 008–049 grown in the presence of epithelial cell line (A549 adenocarcinomic human alveolar basal cells), human umbilical vein endothelial cells or in mammalian tissue culture media alone. RNA-seq reads were pooled and RNA-seq assemblies, both *de novo* and genome-guided against 008–049 genomic scaffolds, were generated with Trinity (Grabherr, Haas *et al.* 2011). Both types of assemblies were mapped to the 008–049 genome using PASA (Haas *et al.* 2003), and *de novo* assemblies only were mapped to other *Lichtheimia* genomes with Genomic Mapping and Alignment Program (GMAP) (Wu and Watanabe 2005). Genomic repeat regions were annotated and masked using RepeatModeler (Smit and

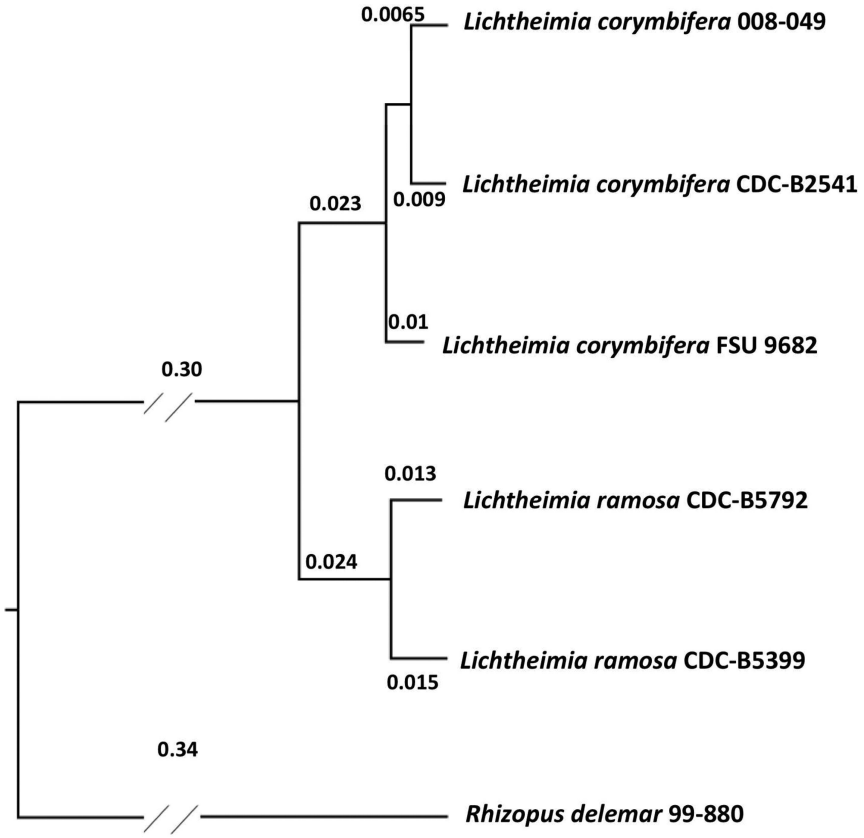

Figure 1. Mid-point rooted phylogenetic tree showing relationships among sequenced *Lichtheimia* genomes, along with *L. corymbifera* FSU 9682 and *R. delemar* 99–880. Tree was generated from a phylogenetic analysis that used over two thousand concatenated protein sequences. Branches show amino acid substitutions per site.

Hubley 2008–2010) and RepeatMasker (Smit *et al.* 1996–2010). Protein-coding genes were predicted *ab initio* with CEGMA (Parra *et al.* 2007), GeneMark-ES (Ter-Hovhannisyan *et al.* 2008), Augustus (Stanke *et al.* 2006), SNAP (Korf 2004), GlimmerHMM (Majoros, Pertea and Salzberg 2004) and GeneID (Blanco *et al.* 2007). Augustus, SNAP and GlimmerHMM used CEGMA predictions for parameter training, and GeneID used a parameter file generated by CEGMA. Raw RNA-seq reads were used to augment Augustus training for *L. corymbifera* 008–049. Spliced alignments of SwissProt proteins against each genome were generated with AAT (Huang *et al.* 1997) using cutoffs of 80% similarity and 1500 bp max intron length. To generate a consensus gene model set, all intrinsic and extrinsic predictions were combined with Evidence Modeler (Haas *et al.* 2008) using the following evidence weights: CEGMA 4, Augustus 4, GeneMark-ES 2, GlimmerHMM 2, SNAP 2, GeneID 2 and AAT alignments 2. Assembled RNA-seq transcript alignments were weighted 10 for alignment to self (e.g. *L. corymbifera* 008–049 transcripts aligned with PASA to *L. corymbifera* 008–049 genome), but weighted 1 when aligned to other (e.g. *L. corymbifera* 008–049 transcripts aligned with GMAP to *L. corymbifera* CDC-B2541). Non-coding RNAs were predicted with tRNAScan-SE and RNAmmer. Predicted proteins were compared to UniProt with BLAST and against TIGRFAMs/PFAMs with HMM searches to generate functional assignments including Gene Ontology terms and Enzyme Commission numbers. A summary of our structural annotation of each of the four genomes can be found in Table 1. Genome completeness, as assessed by detecting complete conserved eukaryotic genes with CEGMA (Parra *et al.* 2007), for each of the genomes was estimated to range from 96–98% complete (Table 1).

FUNGAL PHYLOGENY

We probed the phylogenetic relationship between our isolates and with two Mucorales isolates whose genomes have been sequenced and annotated (*Rhizopus delemar* 99–880 and *L. corymbifera* FSU 9682). To accomplish this, ortholog pairs were detected among Mucorales genomes using InParanoid 4.1 (Remm, Storm and Sonnhammer 2001) with *Umbelopsis isabellina* (CDC-B7317) as an out group, using the two-pass BLAST strategy, bootstrapping and all other algorithm parameters set to default. MultiParanoid (Alexeyenko *et al.* 2006) was run on InParanoid output files to detect ortholog groups common to all isolates. Protein sequences from each ortholog group were aligned using Muscle v.3.7 (Edgar 2004) and gapped regions were removed with Gblocks_0.91b with default settings (Talavera and Castresana 2007). Conserved block alignments were concatenated, and phylogenetic analysis was performed with Phyml 3.0 (Guindon *et al.* 2010) with 100 bootstrap replicates, BioNJ starting tree, nearest neighbor interchange (NNI) tree topology search, and LG amino acid substitution model. The resulting tree was visualized in FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/, 9 July 2014, date last accessed). For the genome sequence of *L. corymbifera* FSU 9682, the previously published annotation was used (Schwartze *et al.* 2014). All *Lichtheimia* isolates, including *L. ramosa*, were very closely related based on a phylogenetic tree generated using over 2000 highly conserved orthologous genes (Fig. 1). For perspective, *R. delemar* 99–880, a better-characterized Mucorales genome, was used as an outgroup for the tree. Our phylogenetic analysis indicates that the two *L. corymbifera* isolates (CDC-B2541 and CDC-008–049) are closely related to the previously sequenced *L. corymbifera* isolate (FSU 9682), while the

two *L. ramosa* isolates (CDC-B5399 and CDC-B5792) form a separate clade.

SUMMARY

The genome sequence data from these *Lichtheimia* species provide a valuable resource for comparative genome analyses to determine interspecies and intraspecies genomic variation which will, in turn, further our understanding of the genetic elements that govern virulence, tropism and antifungal resistance of this genus.

FUNGAL GENOMIC ACCESSIONS

Nucleotide sequence accession numbers: these Whole Genome Shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accessions JNEU00000000, JNEP00000000, JNEE00000000, JNDO00000000 corresponding to strains CDC-B2541, CDC-B5792, 008–049and CDC-B5399, respectively. The versions described in this paper are the first versions: JNEU00000000.1, JNEP00000000.1, JNEE00000000.1 and JNDO00000000.1.

ACKNOWLEDGEMENTS

We thank Anastasia Litvintseva for project guidance and for critical review of the manuscript. We thank Kerstin Voigt for making the data for *L. corymbifera* strain FSU 9682 available before publication.

FUNDING

This project has been funded in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract number HHSN272200900009C. VMB and ASI were supported by U19AI110820. ASI was also supported by R01 AI063503. The findings and conclusions of this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Conflict of interest. None declared.

REFERENCES

Alexeyenko A, Tamas I, Liu G, *et al.* Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* 2006;**22**:e9–15.

Alvarez E, Sutton DA, Cano J, *et al.* Spectrum of zygomycete species identified in clinically significant specimens in the United States. *J Clin Microbiol* 2009;**47**:1650–6.

Bellanger AP, Reboux G, Botterel F, *et al.* New evidence of the involvement of *Lichtheimia corymbifera* in farmer's lung disease. *Med Mycol* 2010;**48**:981–7.

Blanco E, Parra G, Guigo R. Using geneid to identify genes. In: Baxevanis AD, *et al.* (eds). *Curr Protoc Bioinformatics*. Hoboken, NJ: John Wiley & Sons, Inc. 2007, Chapter 4: Unit 4.3.

Cateau E, Randriamalala E, Elsendoorn A, *et al.* Fatal-mixed cutaneous zygomycosis-aspergillosis: a case report. *Mycopathologia* 2013;**176**:423–7.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;**32**:1792–7.

Grabherr MGB, Haas J, Yassour M, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;**29**:644–52.

Guindon S, Dufayard JF, *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;**59**:307–21.

Haas BJ, Delcher AL, Mount SM, *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* 2003;**31**:5654–66.

Haas BJS, Salzberg L, Zhu W, *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol* 2008;**9**:R7.

Huang X, Adams MD, Zhau H, *et al.* A tool for analyzing and annotating genomic sequences. *Genomics* 1997;**46**:37–45.

Irtan S, Lamerain M, Lesage F, *et al.* Mucormycosis as a rare cause of severe gastrointestinal bleeding after multivisceral transplantation. *Transpl Infect Dis* 2013;**15**:E235–8.

Kaur R, Bala K, Ahuja RB, *et al.* Primary cutaneous mucormycosis in a patient with burn wounds due to *Lichtheimia ramosa*. *Mycopathologia* 2014;**178**:291–5.

Korf I. Gene finding in novel genomes. *BMC Bioinformatics* 2004;**5**:59.

Kutlu M, Ergin C, Bir F, *et al.* Pulmonary mucormycosis due to *Lichtheimia ramosa* in a patient with HIV infection. *Mycopathologia* 2014;**178**:111–5.

Lanternier F, Dannaoui E, Morizot G, *et al.* A global analysis of mucormycosis in France: the RetroZygo study (2005–2007). *Clin Infect Dis* 2012;**54**:S35–43.

Linde J, Schwartze V, Binder U, *et al.* De novo whole-genome sequence and genome annotation of *Lichtheimia ramosa*. *Genome Announc* 2014;**5**:e00888–14.

Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 2004;**20**:2878–9.

Parra G, Bradnam K, Korf I, *et al.* CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 2007;**23**:1061–7.

Poirier P, Nourrisson C, Gibold L, *et al.* Three cases of cutaneous mucormycosis with Lichtheimia spp. (ex Absidia/Mycocladus) in ICU. Possible cross-transmission in an intensive care unit between 2 cases. *J Mycol Med* 2013;**23**:265–9.

Remm M, Storm CE, Sonnhammer EL. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* 2001;**314**:1041–52.

Roden MMT, Zaoutis E, Buchanan WL, *et al.* Epidemiology and outcome of zygomycosis: a review of 929 reported cases. *Clin Infect Dis* 2005;**41**:634–53.

Schwartze VU, Hoffmann K, Nyilasi I, *et al.* Lichtheimia species exhibit differences in virulence potential. *PLoS One* 2012;**7**:e40908.

Schwartze VU, Jacobsen ID. Mucormycoses caused by Lichtheimia species. *Mycoses* 2014;**57**:73–8.

Schwartze VU, Winter S, Shelest E, *et al.* Gene expansion shapes genome architecture in the human pathogen *Lichtheimia corymbifera*: an evolutionary genomics analysis in the ancient terrestrial mucorales (Mucoromycotina). *PLoS Genet* 2014;**10**:e1004496.

Skiada A, Pagano L, Groll A, *et al.* Zygomycosis in Europe: analysis of 230 cases accrued by the registry of the European Confederation of Medical Mycology (ECMM) working group on Zygomycosis between 2005 and 2007. *Clin Microbiol Infec* 2011;**17**:1859–67.

Smit AFA, Hubley R. RepeatModeler Open-1.0. 2008–2015. http://www.repeatmasker.org (30 April 2013, date last accessed).

Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. http://www.repeatmasker.org (30 April 2013, date last accessed).

Spellberg B, Ibrahim AS, Chin-Hong PV, *et al*. The Deferasirox-AmBisome therapy for Mucormycosis (DEFEAT Mucor) study: a randomized, double-blinded, placebo-controlled trial. *J Antimicrob Chemoth* 2012;**67**:715–22.

Stanke M, Schoffmann O, Morgenstern B, *et al*. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 2006;**7**:62.

Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007;**56**:564–77.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, *et al*. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res* 2008;**18**:1979–90.

Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 2005;**21**:1859–75.

Zimin AV, Marcais G, Puiu D, *et al*. The MaSuRCA genome assembler. *Bioinformatics* 2013;**29**:2669–77.