AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Black Yeast Genomes Assembled from Plastic Fabric Metagenomes Reveal an Abundance of Hydrocarbon Degradation Genes

Osman Radwan,[a] Oscar N. Ruiz[b]

[a]Environmental Microbiology Group, University of Dayton Research Institute, Dayton, Ohio, USA
[b]Fuels and Energy Branch, Aerospace Systems Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio, USA

ABSTRACT We report the assembly and annotation of 10 different black yeast genomes from microbiome metagenomic data derived from biofouled plastic fabrics. The draft genomes are estimated to be 9 to 33.2 Mb, with 357 to 5,108 contigs and G+C contents of 43.9% to 57.4%, and they harbor multiple genes for hydrocarbon adaptation and degradation.

Black yeast is a diverse group of fungi, belonging to the *Chaetothyriomycetidae* and *Dothideomycetidae* subclasses of ascomycetes, with the ability to degrade hydrocarbons and plasticizers (1–3). In this study, the metagenome-assembled genomes (MAGs) of 10 black yeast species were assembled and annotated from a shotgun metagenomic library of microbiomes associated with biofouled plastic fabric materials (4).

As described by Radwan et al., total genomic DNA was extracted from plasticized fabrics exposed to the Panama jungle for 14 months, and the DNA libraries were sequenced using a shotgun metagenomic approach (4). To maximize DNA extraction efficiency, fabrics were cut into 0.5-cm^2 pieces, and DNA was extracted using the DNeasy UltraClean kit (catalogue number 12224-250; Qiagen) (4). DNA libraries were constructed using the PrepX DNA library kit and the Apollo 324 next-generation sequencing (NGS) automatic library preparation system (WaferGen, Fremont, CA). The TruSeq paired-end DNA libraries were sequenced using an Illumina HiSeq 2000 system, generating 161,537,275,209 bp of raw reads of 100-bp length (4). Trimmomatic v0.36 with the RepBase-20170 library (5) was used for quality control, removing raw reads shorter than 50 bp and those with an average quality score below 15. Multiple bioinformatics programs (6) were used with default parameters except where otherwise noted. After sorting of paired-end reads for compatibility and normalization using BBtools (https://jgi.doe.gov/data-and-tools/bbtools), reads were assembled by the MEGAHIT assembler (7) with options of minimum contig length of 200 bp and meta-sensitive. Fasta contigs produced by MEGAHIT with coverage and abundance files for each fabric sample were used by MaxBin v2.7.7 (8) to bin individual genomes with options of minimum contig length of 2000 and depth of 2.

DNA sequences of 10 different black yeast genomes were selected for assembly improvement and sequence gap filling using SSPACE v3.0 (9) and GapFiller v1.10 (10), respectively (Table 1). After masking of the repetitive sequences with RepeatMasker (11), the genes of each genome were predicted by AUGUSTUS v2.5.5 (12) with an option set for *Yarrowia lipolytica*. CEGMA (13) was used to calculate genome completeness, which ranged from 13.64% for *Hortaea werneckii* to 94.00% for *Exophiala oligosperma*. According to the MAG completeness criteria (14), two MAGs were high quality, two were medium quality, and six were low-quality drafts reflecting partial genomes. ABySS v2.0.2 (15) calculated genome sizes ranging from 9 to 33.2 Mb, with L_{50} values of 22 to 1,258 contigs and G+C contents of 43.9% to 57.4% (Table 1). HMMER v3.3 (16) with E values of $1e^{-3}$ was used to search the Pfam database for functional annotation of proteins involved in hydrocarbon degradation (i.e.,

Citation Radwan O, Ruiz ON. 2021. Black yeast genomes assembled from plastic fabric metagenomes reveal an abundance of hydrocarbon degradation genes. Microbiol Resour Announc 10:e01459-20. https://doi.org/10.1128/MRA.01459-20.

Editor Antonis Rokas, Vanderbilt University

Address correspondence to Oscar N. Ruiz, oscar.ruiz@us.af.mil.

Received 20 December 2020
Accepted 14 March 2021
Published 8 April 2021

TABLE 1 Accession numbers, general statistics of black yeast MAGs, and numbers of genes involved in hydrocarbon degradation, polymer hydrolysis, and efflux pumps for each genome[a]

Code[b]	Genome	GenBank accession no.	Size (bp)	No. of contigs	L_{50} (contigs)	N_{50} (bp)	G+C content (%)	Completeness (%)[c]	No. of degradation genes	No. of hydrolase genes	No. of MFS[d]	No. of ABC[e]
A04	*Exophiala oligosperma*	JADCRK000000000	33,192,209	357	22	451,697	51.6	94.00	137	6	143	85
B02	*Rhinocladiella mackenziei*	JADCRN000000000	26,458,337	925	101	73,418	43.9	93.95	19	0	50	16
B03	*Baudoinia panamericana*	JADCRE000000000	27,253,305	5,108	1,198	6,330	49.7	68.55	112	4	125	65
B04	*Cyphellophora europaea*	JADCRG000000000	18,212,429	4,098	1,132	4,855	54.4	55.65	77	3	216	63
B06	*Exophiala* sp.	JADCRL000000000	20,651,700	4,626	1,258	4,953	45.8	47.58	26	0	47	20
C02	*Cyphellophora europaea*	JADCRH000000000	18,567,089	4,594	1,232	4,116	56.9	42.74	74	3	178	31
A11	*Baudoinia panamericana*	JADCRF000000000	13,998,733	3,358	935	4,505	49.9	39.52	28	3	53	13
F22	*Cyphellophora europaea*	JADCRI000000000	11,716,283	3,003	806	4,045	55.3	24.19	23	3	131	67
B07	*Cyphellophora europaea*	JADCRJ000000000	9,001,416	2,896	1,036	3,049	57.4	18.95	80	4	166	14
A17	*Hortaea werneckii*	JADCRM000000000	9,816,071	2,164	625	5,040	52.6	13.64	1	0	1	11

[a] The Pfam database with E values of $1e^{-3}$ was used for functional annotation. High-quality (A04 and B02), medium-quality (B03 and B04), and low-quality (C02, A11, F22, B06, B07, and A17) draft genomes are shown. The taxonomic identity of each species was determined using Kaiju (19).

[b] A04, A11, and A17 were derived from physical biosample A (BioSample accession number SAMN16111753; SRA accession number SRS7586490). B02, B03, B04, B06, and B07 were derived from physical biosample B (BioSample accession number SAMN16111857; SRA accession number SRS7586491). C02 was derived from physical biosample C (BioSample accession number SAMN16111912; SRA accession number SRS7586492). F22 was derived from physical biosample F (BioSample accession number SAMN16112156; SRA accession number SRS7586495).

[c] Genome completeness was estimated with CEGMA software (13).

[d] MFS, major facilitator superfamily efflux pumps.

[e] ABC, ABC transporters.

cytochrome P450 and aromatic ring-opening dioxygenases), hydrolases (esterases and lipases), and efflux pumps potentially associated with hydrocarbon resistance (Table 1). The number of protein-coding genes identified for each pathway depended on the assembled genome completeness and species. For example, *Exophiala oligosperma*, with a genome completeness of 94.00%, contained the highest number of proteins belonging to different pathways, while *Hortaea werneckii*, with only 13.64% completeness, contained the lowest number of proteins. The results of this study support the ability of black yeasts such as *Exophiala oligosperma* and *Cyphellophora europaea* to degrade hydrocarbons (17, 18) and plasticizers (1, 2).

Data availability. The raw sequence reads and MAGs were deposited in DDBJ/ENA/GenBank under BioProject accession number PRJNA656291 with BioSample accession numbers SAMN15776667 to SAMN15776676 and SRA accession numbers SRX9364069 to SRX9364074. Individual accession numbers for MAGs are provided in Table 1. HMMER results can be retrieved from Mendeley Data at https://data.mendeley.com/datasets/kh6x88k83y/1.

ACKNOWLEDGMENTS

This material is based on research sponsored by the Air Force Research Laboratory, Aerospace Systems Directorate (AFRL/RQTF), under agreement FA8650-16-2-2605.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFRL/RQTF or the U.S. Government.

REFERENCES

1. Webb JS, Nixon M, Eastwood IM, Greenhalgh M, Robson GD, Handley PS. 2000. Fungal colonization and biodeterioration of plasticized polyvinyl chloride. Appl Environ Microbiol 66:3194–3200. https://doi.org/10.1128/aem.66.8.3194-3200.2000.
2. Gostinčar C, Ohm RA, Kogej T, Sonjak S, Turk M, Zajc J, Zalar P, Grube M, Sun H, Han J, Sharma A, Chiniquy J, Ngan CY, Lipzen A, Barry K, Grigoriev IV, Gunde-Cimerman N. 2014. Genome sequencing of four *Aureobasidium pullulans* varieties: biotechnological potential, stress tolerance, and description of new species. BMC Genomics 15:549. https://doi.org/10.1186/1471-2164-15-549.
3. Isola D, Selbmann L, de Hoog GS, Fenice M, Onofri S, Prenafeta-Boldu FX, Zucconi L. 2013. Isolation and screening of black fungi as degraders of volatile aromatic hydrocarbons. Mycopathologia 175:369–379. https://doi.org/10.1007/s11046-013-9635-2.
4. Radwan O, Lee JS, Stote R, Kuehn K, Ruiz ON. 2020. Metagenomic characterization of microbial communities on plasticized fabric materials exposed to harsh tropical environments. Int Biodeterior Biodegrad 154:105061. https://doi.org/10.1016/j.ibiod.2020.105061.
5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
6. Radwan O, Ruiz ON. 2020. Shotgun metagenomic data of microbiomes on plastic fabrics exposed to harsh tropical environments. Data Brief 32:106226. https://doi.org/10.1016/j.dib.2020.106226.
7. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.
8. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2:26. https://doi.org/10.1186/2049-2618-2-26.
9. Boetzer M, Henke CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578–579. https://doi.org/10.1093/bioinformatics/btq683.
10. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. Genome Biol 13:R56. https://doi.org/10.1186/gb-2012-13-6-r56.
11. Smit AFA, Hubley R, Green P. 2010. RepeatMasker Open-3.0. http://www.repeatmasker.org.
12. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res 32:W309–W312. https://doi.org/10.1093/nar/gkh379.
13. Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067. https://doi.org/10.1093/bioinformatics/btm071.
14. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Genome Standards Consortium, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. https://doi.org/10.1038/nbt.3893.
15. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123. https://doi.org/10.1101/gr.089532.108.
16. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Res 46:W200–W204. https://doi.org/10.1093/nar/gky448.
17. Rustler S, Chmura A, Sheldon RA, Stolz A. 2008. Characterization of the substrate specificity of the nitrile hydrolyzing system of the acidotolerant black yeast *Exophiala oligosperma* R1. Stud Mycol 61:165–174. https://doi.org/10.3114/sim.2008.61.17.
18. Prenafeta-Boldú FX, Summerbell R, Sybren de Hoog G. 2006. Fungi growing on aromatic hydrocarbons: biotechnology's unexpected encounter with biohazard. FEMS Microbiol Rev 30:109–130. https://doi.org/10.1111/j.1574-6976.2005.00007.x.
19. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7:11257. https://doi.org/10.1038/ncomms11257.