# Whole Genome Sequencing and Comparative Genome Analysis of the Halotolerant Deep Sea Black Yeast *Hortaea werneckii*

**Orazio Romeo** [1] **, Alessia Marchetta** [1] **, Domenico Giosa** [2] **, Letterio Giuffrè** [1] **, Clara Urzì** [1] **and Filomena De Leo** [1,*]

[1]  Department of Chemical, Biological, Pharmaceutical and Environmental Sciences, University of Messina, 98166 Messina, Italy; oromeo@unime.it (O.R.); alemarchetta@unime.it (A.M.); lgiuffre@unime.it (L.G.); urzicl@unime.it (C.U.)

[2]  Department of Clinical and Experimental Medicine, University Hospital of Messina, 98125 Messina, Italy; dgiosa@unime.it

*  Correspondence: fdeleo@unime.it; Tel.: +39-0906765201

check for updates

**Abstract:** *Hortaea werneckii*, an extreme halotolerant black yeast in the order of Capnodiales, was recently isolated from different stations and depths in the Mediterranean Sea, where it was shown to be the dominant fungal species. In order to explore the genome characteristics of these Mediterranean isolates, we carried out a de-novo sequencing of the genome of one strain isolated at a depth of 3400 m (MC873) and a re-sequencing of one strain taken from a depth of 2500 m (MC848), whose genome was previously sequenced but was highly fragmented. A comparative phylogenomic analysis with other published *H. werneckii* genomes was also carried out to investigate the evolution of the strains from the deep sea in this environment. A high level of genome completeness was obtained for both genomes, for which genome duplication and an extensive level of heterozygosity (~4.6%) were observed, supporting the recent hypothesis that a genome duplication caused by intraspecific hybridization occurred in most *H. werneckii* strains. Phylogenetic analyses showed environmental and/or geographical specificity, suggesting a possible evolutionary adaptation of marine *H. werneckii* strains to the deep sea environment. We release high-quality genome assemblies from marine *H. werneckii* strains, which provides additional data for further genomics analysis, including niche adaptation, fitness and evolution studies.

**Keywords:** *Hortaea werneckii*; black yeasts; whole-genome sequencing; genome assembly; annotation

## 1. Introduction

*Hortaea werneckii* belongs to the so-called black yeast, a polyphyletic group of pleomorphic and melanised fungi that present in many cases a polyextremotolerant lifestyle [1,2]. Up to date, this species was mainly studied for its remarkable halo-tolerance, being the only known fungus able to grow in a wide range of salinity, from 0 M NaCl to saturation at 5.1 M NaCl [3,4]. For this reason, the fungus is used as a model organism for understanding the osmoadaptation strategies and the molecular mechanisms involved in the tolerance to high salinities of eukaryotic cells [5–8].

*H. werneckii* is assigned to the division of Ascomycota in the family *Teratosphaeriaceae* (Pezizomycotina, Dothideomycetes, Capnodiales) and the species was previously classified as *Exophiala werneckii*, *Cladosporium werneckii* or *Phaeoannellomyces werneckii* [9], until the genus *Hortaea* was established [10]. The genus *Hortaea* currently also includes the species *Hortaea thailandica*, isolated from *Syzygium siamense* leaf spots in Thailand [11]. The identification at species level is complicated by the fact that some housekeeping genes commonly used for taxonomic purposes, such as β-tubulin,

mini-chromosome maintenance protein and translation elongation factor 1, produce ambiguous results in some *H. werneckii* strains [12,13]. So, the sequencing of ribosomal genes together with morphology, biochemical and physiological characteristics is considered the main useful criteria to identify this species [12].

Before 2000, *H. werneckii* was studied mainly for medical interest, as the etiology agent of the human disease "tinea nigra", a non-invasive skin infection of hands and feet with global distribution but with a higher incidence in tropical and subtropical climates [14–17]. The physiological preferences of the species, such as an ability to grow better at higher salt conditions at 37 °C, should facilitate skin infection by the fungus [13]. However, the pathogenetic mechanisms are still not clear.

*H. werneckii* is a cosmopolitan fungal species, well adapted to live in hypersaline environments worldwide, such as brine in eutrophic solar salterns, which are considered its primary ecological niche [18]. In fact, it has been shown to be the dominant fungal species in these habitats at salinities above 20% NaCl, and thus its presence was reported in different salty environments such as seawater, beach soils, saltern microbial mats, immersed wood, fish, corals and salt marsh plants [7,18–20].

Although *H. werneckii* is not listed as a marine species [21], it is stated that it also occurs in marine habitats, and also in the depths of the oceans [22–24].

Recently, *H. werneckii* was also isolated from Mediterranean seawater up to the depth of 3400 m [13,25], where it represented the dominant fungal species. These Mediterranean strains were proven to be highly similar in terms of physiological and genetic characteristics, and significantly deviated from other strains from different sources, being also less halophilic [13]. These findings were explained by the fact that these strains may be less often related to the hypersaline environment of coastal saltpans, suggesting a possible marine origin or evolution in these habitats.

To date, the genomes of 12 *H. werneckii* strains isolated from diverse sources (3 from hypersaline water, 1 from deep-sea water, 1 from rocks, 3 from spider webs, 1 from coast soil and 3 clinical strains) and different countries have been sequenced [26].

The first whole genome sequencing of an *H. werneckii* strain (EXF-2000) isolated from the hypersaline water of salt-pans in Slovenia showed that, differently from related species within the *Dothideomycetes*, the fungus has a diploid genome, characterized by high levels of heterozygosity between the two sub-genomes [27,28]. However, when the genomes of a further 11 *H. werneckii* strains were sequenced, 2 haploid strains (EXF-562 and EXF-2788, isolated from sea coast soil in Namibia and from salterns in Slovenia, respectively) were detected, while the remaining 9 strains, including the strain isolated from the depths of the Mediterranean Sea, were diploid and highly heterozygous, as with the reference genome [26]. Although at least one mating locus was identified in all sequenced strains, the species undergoes clonal reproduction, and the duplication of the genomes seems to be due to intraspecific hybridization events between ancestors [26].

The aim of this study was to sequence the whole genomes of two *H. werneckii* strains isolated from Mediterranean Sea water sampled at 3400 m and 2500 m depths. A de-novo whole genome assembly was carried out for the strain MC873 (3400 m depth), while a re-sequencing of the strain MC848 (2500 m depth) was performed in order to get a better assembly for this strain, as its current draft genome sequence is highly fragmented (5734 contigs; GenBank assembly accession: GCA_003704595.1).

A comparative phylogenomic analysis with other published *H. werneckii* genomes, from different environmental and clinical sources, was also carried out to investigate the evolutionary history and the possible marine origin of the strains from the deep sea.

## 2. Materials and Methods

### 2.1. Fungal Isolates and Genomic DNA Isolation

The genomes of two *H. werneckii* strains, MC848 and MC873, isolated from the stations of Vector and Geostar in the Mediterranean Sea at 2500 and 3400 m depths, respectively [13,25], were examined in this study. Both strains are maintained in the collection of the Department of Chemical, Biological,

Pharmaceutical and Environmental Sciences of University of Messina, Italy. Information relative to the *Hortaea werneckii* strains included in this study for comparative genome analysis is reported in Table 1.

**Table 1.** Strains of *Hortaea werneckii* included in this study.

| Strain | Other Collection | Source | Country | Reference |
|---|---|---|---|---|
| EXF-2000 | CBS 100457 | Hypersaline water of salt-pans | Slovenia | [28] |
| EXF-2788 | | Hypersaline water of salt-pans | Slovenia | [26] |
| EXF-120 | | Hypersaline water of salt-pans | Spain | [26] |
| EXF-562 | | Soil on the sea cost | Namibia | [26] |
| EXF-171 | CBS 111.31 | Human, *Keratomycosis* | Brazil | [26] |
| EXF-2682 | CBS 126.35 | Human, *Trichomycosis nigra* | Italy | [26] |
| EXF-151 | CBS 107.67 | Human, *Tinea nigra* | Portugal | [26] |
| EXF-6651 | | Spider web, Atacama desert | Chile | [26] |
| EXF-6654 | | Spider web, Atacama desert | Chile | [26] |
| EXF-6669 | | Spider web, Atacama desert | Chile | [26] |
| EXF-6656 | | Rock in a cave, Atacama desert | Chile | [26] |
| EXF-10513 | MC848 V2500 b | Seawater of Mediterranean Sea; 2500 m depth, Vector station | Italy | [26]; This study |
| MC873 | Geo f 100 c | Seawater, Mediterranean Sea; 3400 m depth, Geostar station | Italy | This study |

For whole-genome sequencing, genomic DNA was extracted from the two marine *H. werneckii* strains after growth on MEA medium for 7 days, using a mechanical glass-beads disruption method followed by traditional phenol\chloroform extraction and ethanol precipitation. Briefly, fungal cells were transferred to a 2 mL eppendorf tube, containing 500 µL of lysis buffer (1% SDS; 0.1 M Tris, pH 8.0; 50 mM EDTA, pH 8.0), 25 µL of NaCl 5 M and about 400 µL of acid washed glass beads (Ø 0.50 mm). The mixture was vortexed for 2 min and centrifuged for 3 min at 14,000 rpm, then the supernatant was collected in a new tube. A volume of 500 µL of lysis buffer was added to the pellet, then vortexed and centrifuged for 3 min at 14,000 rpm. The supernatant was recovered and added to the one previously collected. After addition of 400 µL phenol, the solution was mixed and centrifuged for 5 min at 12,000 rpm and the supernatant was transferred to a new tube with an equal volume of phenol-chloroform (4:1), then the mixture was mixed and centrifuged for 5 min at 12,000 rpm. The supernatant was collected, and the DNA was precipitated with a volume of 1 mL of absolute ethanol and incubated at −80 °C for 1 h, then centrifuged for 10 min at 12,000 rpm. Subsequently the pellet was washed with cold 70% ethanol, dried and dissolved in 50 µL of ultrapure water. RNase treatment was carried out by adding 2 µL of RNase (20 µg/mL) and incubating for 45 min at 37 °C. Finally, the samples were stored at −20 °C. DNA integrity and purity were evaluated spectrophotometrically and by 1.2% agarose gel electrophoresis. High quality DNA ($A_{260/280} \geq 1.8$) was used for library construction.

*2.2. Library Preparation, Sequencing and Genome Assembly*

The genomes of the *H. werneckii* MC873 and MC848 strains were sequenced using Illumina NextSeq 500 technology (Illumina, Italy). A total of two sequencing libraries, with insert sizes of approximately 350 bp and 550 bp, respectively, were prepared for each strain and sequenced using the paired-end (2 × 150 bp) strategy. After sequencing, raw reads were preprocessed using the fastpv.0.20.0 software (options: sliding window of 5 bp; average base quality-score of 25; minimum read length of 35 bp) [29] to remove adapters and/or other Illumina technical sequences, including low-quality reads/bases. The final data set, used for genome assembling, contained a total of 70,946,302 (MC873 strain) and 97,637,450 (MC848 strain) quality-controlled reads. Using these clean reads, a de-novo genome assembly was performed using SPAdesv.3.13.0 assembler [30] and the resulting contigs were further processed with SSPACE-Standard v.3.0 software [31] to obtain scaffold sequences. The GapFiller v.1-10 program [32] was then used to close the gaps within pre-assembled scaffolds by reducing the number of undetermined bases in the genome.

*2.3. Evaluation of the Quality and Completeness of the Genome Assembly and Determination of the Level of Heterozygosity*

The quality of genome assemblies was assessed using the QUAST v.5.0.2 program [33] whereas their completeness was quantitatively evaluated with Benchmarking Universal Single-Copy Orthologs (BUSCO) v.3.1.0 [34] by searching for universal single-copy orthologs in different lineage-specific datasets (eukaryota_odb9, fungi_odb9 and ascomycota_odb9) [34]. Finally, the KAT tool v. 2.4.2 [35] and GenomeScope [36] were employed to perform the K-mer frequency (k = 27) analysis and to estimate the level of heterozygosity and duplication of each genome.

*2.4. Gene Prediction, Functional Genome Annotation and Phylogenomic Analysis*

Gene models were predicted separately for each *H. werneckii* genome using the MAKER pipeline (v.3.00.0) [37] combined with the ab-initio gene predictors SNAP (v.2.39) [38] and AUGUSTUS (v. 3.3.3) [39], and further integrated with a full set of reference proteins and expressed sequence tags (ESTs) from Capnodiales (NCBI: txid134362) downloaded from the NCBI Protein and Nucleotide databases, respectively. However, before using these datasets, sequences sharing more than 90% similarity were removed by CD-HIT software [40] in order to reduce the redundancy.

Automatic functional annotation of predicted proteins was performed using PANNZER2 program [41] and resulting proteomes were compared by OrthoFinder (v.2.3.11) (https://github.com/davidemms/OrthoFinder) in order to evaluate the content of orthologous genes and the abundance of distinct orthogroups among the two sequenced strains. An orthogroup was defined as a set of genes derived from a single gene in the last common ancestor of different species. Orthogroups, containing a different number of genes, were subsequently submitted to ShinyGO v0.61 (http://bioinformatics.sdstate.edu/go) to determine if any particular metabolic pathway was enriched in a strain-specific manner.

Finally, the two genome assemblies were also screened for transposable elements (TE) and repetitive DNA sequence content by using Tandem Repeat Finder v.4.09 [42] and RepeatMasker v.4.0.9 (www.repeatmasker.org) softwares. tRNAs were also predicted using the tRNAscan-SE program v.1.3.1 [43].

Phylogenomic analysis was conducted using our two sequenced genomes and additional publicly available *H. werneckii* genome assemblies retrieved from the NCBI BioProject database (accession number: PRJNA428320) and GenBank (assembly accession: GCA_002127715.1) (Table 1). Phylogenetic distance was estimated by the Mash 2.2 toolkit [44] (sketch size, $s = 1000$; k-mer size, $k = 21$) while cluster analysis was performed with the pvclustR package [45], using the UPGMA algorithm with a bootstrap analysis of 1000 replicates. The resulting tree was visualized and edited using iTOL (https://itol.embl.de).

## 3. Results

Basic statistics of the *H. werneckii* genome assemblies generated in this study are shown in Table 2. The MC848 and MC873 genomes were sequenced with an average coverage of ~276× and ~207×, respectively, and their predicted size was quite similar, consisting of approximately ~51 Mbp each, with a total GC content of 53.4% (Table 2). The raw reads have been deposited in the Sequence Reads Archive (SRA) database and are available under BioProject identifier (ID) PRJNA641248 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA641248).

**Table 2.** Overall assembly statistics and gene content of the two *H. werneckii* genomes sequenced in this study.

| *H. werneckii* Genome Features | MC873 Strain | MC848 Strain |
|---|---|---|
| Total number of sequenced bases | 10,481,489,013 | 14,036,248,558 |
| Total number of reads with Q-score ≥25 | 70,784,397 | 98,305,008 |
| Total number of mapped reads (%) | 70,711,190 (99.9%) | 95,135,691 (96.8%) |
| Unmapped reads | 73,207 (0.1%) | 3,169,317 (3.2%) |
| Number of total contigs | 925 | 1218 |
| Number of contigs >1 Kbp | 761 | 612 |
| Largest contig (bp) | 604,832 | 576,474 |
| Genome size (bp) | 50,705,820 | 51,030,830 |
| Average coverage depth | ~207× | ~276× |
| GC content (%) | 53.4 | 53.4 |
| Total number of predicted genes | 15,542 | 15,565 |
| Protein-coding genes | 15,397 | 15,410 |
| tRNAs | 126 | 132 |
| snRNAs | 5 | 5 |
| Total gene length (bp) | 26,021,394 | 26,184,209 |
| Longest gene (bp) | 15,052 | 19,877 |
| Number of exons | 32,681 | 32,716 |
| Longest exon (bp) | 12,966 | 12,966 |
| Exon average length (bp) | ~742 | ~746 |
| Gene average length (bp) | ~1690 | ~1699 |
| Tandem Repeat Number | 13,188 | 13,315 |
| Simple repeats | 11,183 | 11,282 |
| Low-complexity repeats | 2005 | 2033 |

The draft genome sequences contained a total of 1218 (MC848 strain) and 925 (MC873 strain) assembled contigs (minimum length: ≥200 bp), and have been submitted to the GenBank database under the following accession numbers: JACSRB000000000 (MC848) and JACSRC000000000 (MC873). The comparison of k-mer frequencies to the final assemblies revealed that most of the 27-mers in the reads were represented once in the respective assemblies, consistent with a high-quality genome assembly (estimated assembly completeness: >99.8%) of a predominantly heterozygous microorganism (Figure 1a). This genome-level variability was also confirmed by GenomeScope analysis (k-mer: 27), which showed an extensive level of heterozygosity (average ~4.6%) for both *H. werneckii* genomes.

To examine the phylogenetic relationships among our *H. werneckii* strains and other previously sequenced strains, we generated a Mash-based tree using the whole genomes currently available in GenBank (Figure 1b).

Phylogenomic analysis showed the existence of two main clusters (Figure 1b). The strains from the Mediterranean Sea clustered together, and they were very close to one clinical strain from Brazil (EXF-171) and two environmental isolates—one from salterns of Spain (EXF-120) and one haploid strain from Namibia (EXF-562). The second cluster grouped eight strains: four from the Atacama desert in Chile (EXF-6651, EXF-6654, EXF-6669 and EXF-6656), two clinical strains from Portugal (EXF-151) and from Italy (EXF-2682), and two strains from salterns of Slovenia (respectively, EXF-2000 and the haploid strain EXF-2788).
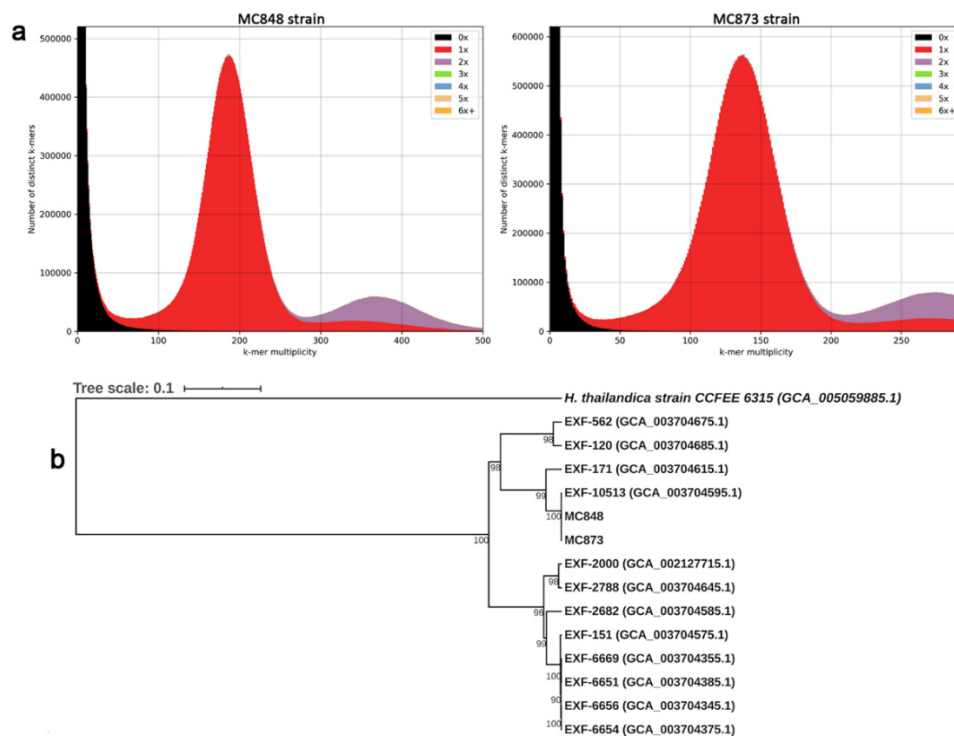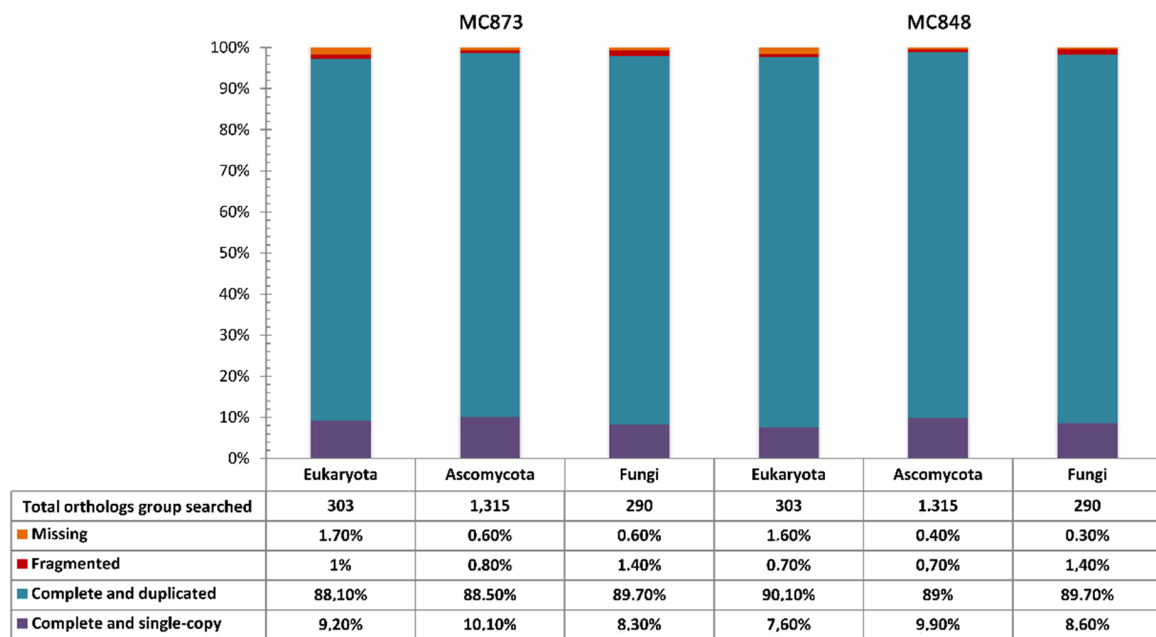
**Figure 1.** Genome assembly analysis and phylogenomic relationships of *H. werneckii* strains. (**a**) k-mer frequency analysis generated by the KAT tool; plots show KAT spectra of the 27-mer multiplicity of the read set (x axis) against their respective assembly (y axis) for MC873 and MC848 strains. Read content in black represents the 27-mers present in the reads but absent in the assemblies and suggests a fairly complete genome assembly. Red indicates 27-mers are present once in the assembly while purple 27-mers are present twice, indicating that genome duplication events likely occurred; (**b**) Mash-based phylogenetic tree generated using our two *H. werneckii* genome assemblies along with whole-genomes currently available in GenBank database. The code of each strain used is shown to the right of each branch tip and its GenBank accession number. Numbers at the nodes represent percent bootstrap support values based on 1000 replicates.

For both MC848 and MC873 genomes, we observed evidence of genome duplication by k-mer frequency analysis (Figure 1a), which was further corroborated byBUSCOanalysis (Figure 2).

Notably, for both *H. werneckii* genomes, we detected a high level of complete and duplicated BUSCO genes (>88%; Figure 2), which is consistent with the level of genome duplication observed. Overall, BUSCO results also confirmed a high level of genome completeness, with over 97% of eukaryotic and fungal genes found to be complete in our assemblies (Figure 2).

Totals of 15,542 and 15,565 genes were predicted in the *H. werneckii* MC873 and MC848 genomes, respectively. There was only a slight difference in protein-coding gene content between the two sequenced genomes (Table 2). Further small variations were also observed for the transposon content (23 extra TE found only in the MC848 strain), especially for the LTR-retrotransposon gipsy and non-LTR LINE-like retrotransposons, including some DNA TE that were exclusively detected in the MC848 strain (Table 3).

| | Eukaryota | Ascomycota | Fungi | Eukaryota | Ascomycota | Fungi |
|---|---|---|---|---|---|---|
| Total orthologs group searched | 303 | 1,315 | 290 | 303 | 1.315 | 290 |
| Missing | 1.70% | 0.60% | 0.60% | 1.60% | 0.40% | 0.30% |
| Fragmented | 1% | 0.80% | 1.40% | 0.70% | 0.70% | 1.40% |
| Complete and duplicated | 88,10% | 88.50% | 89.70% | 90,10% | 89% | 89.70% |
| Complete and single-copy | 9,20% | 10.10% | 8,30% | 7,60% | 9,90% | 8,60% |

**Figure 2.** (BUSCO) results showing the number of complete (single and duplicated), missing and fragmented orthologs obtained by searching the Eukaryota, Ascomycota and fungal lineage datasets.

**Table 3.** Classes of transposable elements (TE) and their relative numbers detected in the two sequenced marine *H. werneckii* genomes.

| TE Class | MC848 | MC873 | TE Class | MC848 | MC873 |
|---|---|---|---|---|---|
| DNA (notclassified) | 2 | 3 | LINE I-Jockey | 1 | 1 |
| DNA CMC-EnSpm | 2 | 3 | LINE L1 | 4 | 3 |
| DNA Dada | 10 | 8 | LINE L1-Tx1 | 4 | 3 |
| DNA Ginger-1 | 2 | 2 | LINE L2 | 1 | 0 |
| DNA hAT | 1 | 2 | LINE RTE | 1 | 0 |
| DNA hAT-Ac | 18 | 17 | LINE RTE-BovB | 3 | 5 |
| DNA hAT-Charlie | 2 | 2 | LTR (notclassified) | 3 | 3 |
| DNA Kolobok-T2 | 1 | 0 | LTR Copia | 8 | 1 |
| DNA Merlin | 1 | 0 | LTR ERV1 | 7 | 6 |
| DNA-Maverick | 0 | 1 | LTR ERVK | 6 | 8 |
| DNA Mule-MuDR | 1 | 1 | LTR Gypsy | 165 | 151 |
| DNA P | 1 | 1 | LTR Ngaro | 12 | 12 |
| DNA TcMar-ISRm11 | 1 | 1 | LTR Pao | 18 | 17 |
| DNA-TcMar-Tc1 | 0 | 1 | RC Helitron | 1 | 2 |
| LINE CR1 | 1 | 0 | SINE B2 | 1 | 1 |
| LINE I | 1 | 0 | SINE-tRNA-RTE | 0 | 1 |
| Total | | | | 279 | 256 |

Regarding functional analysis of the gene set annotated in this study, we observed small differences among the two *Hortaea* genomes. In fact, over 99% of these genes were assigned to 10,911 unique orthogroups, of which 6052 (55.5%) contained single-copy orthologs, whereas the remaining 4859 (44.5%) included at least 2 genes from a single strain (Supplementary Material Table S1). Interestingly, 1047 orthogroups (1047/10,911;9.6%) showed strain-level differences in the number of genes included in each group. More specifically, of the 1047 orthogroups, 528 (50.4%) were enriched in the MC848 strain and 497 (47.5%) were enriched in the MC873 strain, while the remaining 22 (2.1%) orthogroups included strain-specific genes. In particular, the MC848 strain displayed 10 specific orthogroups containing a total of 20 genes, while the remaining 12 MC873-specific orthogroups included a total

of 24 genes (Supplementary Material Table S1). However, most of the strain-specific genes encoded proteins with unknown or still uncharacterized functions (Supplementary Material Table S1).

## 4. Discussion

The recent finding that *H. werneckii* has a ubiquitous distribution in the seawater of the Mediterranean Sea, being the dominant fungal species up to a depth of 3400 m [13,25], has led us to deepen the genomic characterization of these strains to better understand their origin and evolution.

The presence of the fungus was demonstrated also in other deep-sea environments, such as deep-sea hydrothermal ecosystems in the Pacific Ocean [22,24] and sediments at 5000 m depths in the Central Indian Basin [23]. However, the species is not currently considered a marine fungus [21].

Comparative analyses of Mediterranean isolates with other *H. werneckii* strains recovered from different sources evidenced the peculiar genetic and physiological characteristics of the seawater strains compared to the others [13], but these were not sufficient to assign them to another taxonomic group [12].

In this study we carried out a de-novo sequencing of the genome of one strain isolated at a 3,400 m depth (MC873), and a re-sequencing of the genome derived from a 2500 m depth (MC848). The latter strain MC848 was previously sequenced and assembled in 5734 contigs, which were highly fragmented and presumably had an extensive number of errors, especially in terms of total gene number predicted [46].

In this study, a high-quality genome assembly (>99.8%) was obtained for both MC848 and MC873 genomes, which proved to be diploid. In fact, k-mer frequency analysis (Figure 1a) and BUSCO analysis (Figure 2) showed the existence of genome duplication in both genomes. In addition, a high level of heterozygosity (~4.6%) was also detected for both *H. werneckii* genomes. This is a typical hallmark of diploid species and/or eukaryotic lineages that have undergone whole-genome duplication (WGD) by occasional intraspecific hybridization events between two haploid progenitors. This finding was suggested also by Gostinčar et al. [26], who, after analyzing 12 genomes of *H. werneckii*, found that only 2 (EXF-562 and EXF-2788) were haploid, while the remaining strains were diploid and highly heterozygous. Our results support this hypothesis.

Phylogenetic analyses carried out in this study showed the existence of two separate clusters. Interestingly, our strains, recovered from the deep sea at 3400 and 2500 m of depth, but in different stations of the Mediterranean Sea, clustered together, suggesting a degree of environmental specificity for these genotypes (Figure 1b). The placement of our strains in a single well-supported group, together with the previously sequenced EXF-10513 strain (BioSample accession n°: SAMN08295408) (Figure 1b), is consistent with our expectation, as this latter strain and MC848 are the same strain, although strain codes may vary in the publications [26]. However, from the re-sequencing of the genome of this strain (MC848=EXF-10513), we obtained a better assembly with fewer contigs (1218 vs. 5734 contigs), and consequently an improved genome annotation according to other studies [46]. In fact, we detected a lower number of predicted genes (15,410 vs. 17,094) and exons (32,716 vs. 37,837) [26] as expected for less fragmented genome assemblies [46].

Environmental and/or geographical specificity was also observed for other *H. werneckii* genomes in the phylogenetic tree, such as EXF-2000 and EXF-2788 strains, isolated from hypersaline water, or EXF-6651 and EXF-6669 strains, recovered from spider webs in the Atacama desert, Chile [26]. These findings may suggest an evolutionary adaptation of *H. werneckii* strains to such environments.

The sizes of the MC848 and MC873 genomes sequenced in this study were 51,030,830 bp and 50,705,820 bp, respectively, confirming that the DNA content of marine strains may be slightly larger than that of other *H. werneckii* genomes sequenced so far, whose sizes range from 25.2 Mbp to 49.9 Mbp [26].

Genome sizes show dramatic variation among the *Dothideomycetes* [47], and after *Pseudocercospora* (*Mycosphaerella*) *fijiensis* (genome size ~74.14 Mbp) [47], *H. werneckii* possesses the largest genome among members of the order of Capnodiales [28] (www.zbi.ee/fungal-genomesize). However, additional large

genomes are frequently described in other dothideomycetous black fungi, such as *Friedmanniomyces endolithicus* (genome size 46.75 Mbp), isolated from the extreme environment of the Antarctic [48].

Large scale genome duplication in fungi seems to be associated with selective advantage in terms of stressful environmental conditions. In *Rhizopus oryzae* and *Phycomyces blakeesleeanus* species, WGD contributes to pathogenicity and the expansion of signal transduction and light sensing [49,50], while increases in genome size were also observed in experimental evolution studies in *Saccharomyces cerevisiae*, when the yeast is exposed to stressful concentrations of salt and UV radiations [51,52]. In fact, genome duplication provides the raw material for further evolution processes. In the case of *H. werneckii*, the large genome also contains expanded gene families encoding metal cation transporters [27] that are supposed to confer a selective advantage in hypersaline environments [28].

The results obtained in this study support the recent hypothesis that most *H. werneckii* strains are likely derivatives from intraspecific hybridization and, due the phylogenomic differences observed between strains from different sources, we could suppose that marine strains are evolving in this environment where they are well adapted.

The release of high-quality genome assemblies from marine *H. werneckii* strains provides additional data enabling further genomics analysis, including niche adaptation, fitness and evolution studies for investigating the diversification of *Hortaea* species and their specific associations with stressful and harsh environments.

**Supplementary Materials:** The following are available online at http://www.mdpi.com/2075-1729/10/10/229/s1, Table S1: Overall statistics of all orthogroups detected in the two *Hortaea werneckii* strains.

**Data Availability:** The Illumina raw reads have also been submitted into the Sequence Read Archive (SRA) database and are associated with the BioProject ID: PRJNA641248. The draft whole-genome sequences of strains MC848 and MC873 have also been deposited at DDBJ/ENA/GenBank under accession numbers JACSRB000000000 andJACSRC000000000, respectively. The versions described in this article are the first versions JACSRB000000000 and JACSRC000000000.

## References

1. Gostinčar, C.; Muggia, L.; Grube, M. Polyextremotolerant black fungi: Oligotrophism, adaptive potential, and a link to lichen symbioses. *Front. Microbiol.* **2012**, *3*, 390. [CrossRef]

2. Moreno, L.F.; Vicente, V.A.; de Hoog, S. Black yeasts in the omics era: Achievements and challenges. *Med. Mycol.* **2018**, *56*, S32–S41. [CrossRef]

3. Butinar, L.; Sonjak, S.; Zalar, P.; Plemenitaš, A.; Gunde-Cimerman, N. Melanized halophilic fungi are eukaryotic members of microbial communities in hypersaline waters of solar salterns. *Bot. Mar.* **2005**, *48*, 73–79. [CrossRef]

4. Gunde-Cimerman, N.; Zalar, P.; de Hoog, G.S.; Plemenitaš, A. Hypersaline waters in salterns- Natural ecological niches for halophilic black yeasts. *FEMS Microbiol. Ecol.* **2000**, *32*, 235–240. [CrossRef]

5. Petrovič, U.; Gunde-Cimerman, N.; Plemenitaš, A. Cellular responses to environmental salinity in the halophilic black yeast *Hortaea werneckii*. *Mol. Microbiol.* **2002**, *45*, 665–672. [CrossRef]

6. Kogej, T.; Stein, M.; Volkmann, M.; Gorbushina, A.A.; Galinski, E.A.; Gunde-Cimerman, N. Osmotic adaptation of the halophilic fungus *Hortaea werneckii*: Role of osmolytes and melanization. *Microbiology* **2007**, *153*, 4261–4273. [CrossRef]

7. Plemenitaš, A.; Lenassi, M.; Konte, T.; Kejžar, A.; Zajc, J.; Gostinčar, C.; Gunde-Cimerman, N. Adaptation to high concentrations in halotolerant/halophilic fungi: A molecular prospective. *Front. Microbiol.* **2014**, *5*, 199. [CrossRef]

8.  Gunde-Cimerman, N.; Plemenitaš, A.; Oren, A. Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiol. Rev.* **2018**, *42*, 353–375. [CrossRef]

9.  Gunde-Cimerman, N.; Plemenitaš, A. Ecology and molecular adaptations of the halophilic black yeast *Hortaeawerneckii. Rev. Environ. Sci. Biotechnol.* **2006**, *5*, 323–331. [CrossRef]

10. Nishimura, K.; Miyaji, M. Hortaea, a new genus to accommodate *Cladosporium werneckii. Jpn. J. Med. Mycol.* **1984**, *25*, 139–146. [CrossRef]

11. Crous, P.W.; Schoch, C.L.; Hyde, K.D.; Wood, A.R.; Gueidan, C.; de Hoog, G.S.; Groenewald, J.Z. Phylogenetic lineages in the *Capnodiales. Stud. Mycol.* **2009**, *64*, 17–47. [CrossRef]

12. Zalar, P.; Zupančič, J.; Gostinčar, C.; Zajc, J.; de Hoog, G.S.; De Leo, F.; Azua-Bustos, A.; Gunde-Cimerman, N. The extremely halotolerant black yeast *Hortaea werneckii*–a model for intraspecific hybridization in clonal fungi. *IMA Fungus* **2019**, *10*, 10. [CrossRef]

13. Marchetta, A.; van den Ende, G.B.; Al-Hatmi, A.M.S.; Hagen, F.; Zalar, P.; Sudhadham, M.; Gunde-Cimerman, N.; Urzì, C.; de Hoog, S.; De Leo, F. Global molecular diversity of the halotolerant fungus *Hortaea werneckii. Life* **2018**, *8*, 31. [CrossRef]

14. Göttlich, E.; de Hoog, G.S.; Yoshida, S.; Takeo, K.; Nishimura, K.; Miyaji, M. Cell-surface hydrophobicity and lipolysis as essential factors in human tinea nigra. *Mycoses* **1995**, *38*, 489–494. [CrossRef]

15. Perez, C.; Colella, M.T.; Olaizola, C.; de Capriles, C.H.; Magaldi, S.; Mata-Essayag, S. Tineanigra: Report of twelvecases in Venezuela. *Mycopathologia* **2005**, *160*, 235–238, ds11046–s005. [CrossRef]

16. Bonifaz, A.; Badali, H.; de Hoog, G.S.; Cruz, M.; Araiza, J.; Cruz, M.A.; Fierro, L.; Ponce, R.M. Tinea nigra by *Hortaea werneckii*, a report of 22 cases from Mexico. *Stud. Mycol.* **2008**, *61*, 77–82. [CrossRef]

17. Giordano, M.C.; De la Fuente, A.; Lorca, M.B.; Kramer, D. Tinea nigra: Report of three pediatrics cases. *Rev. Chil. Pediatr.* **2018**, *89*, 506–510. [CrossRef]

18. Gunde-Cimerman, N.; Zalar, P. Extremely halotolerant and halophilic fungi inhabit brine in solar salterns around the globe. *Food Technol. Biotech.* **2014**, *52*, 170–179.

19. Xu, W.; Pang, K.-L.; Luo, Z.-H. High fungal diversity and abundance recovered in the deep-sea sediments of the Pacific Ocean. *Microb. Ecol.* **2014**, *68*, 688–698. [CrossRef]

20. Formoso, A.; Heidrich, D.; Felix, C.R.; Tenório, A.C.; Leite, B.R.; Pagani, D.M.; Ortiz-Monsalve, S.; Ramírez-Castrillón, M.; FontesLandell, M.; Scroferneker, M.L.; et al. Enzymatic activity and susceptibility to antifungal agents of brazilian environmental isolates of *Hortaeawerneckii. Mycopathologia* **2015**, *180*, 345–352, ds11046–s015. [CrossRef]

21. Jones, E.B.G.; Suetrong, S.; Bahkali, A.H.; Abdel-Wahab, M.A.; Boekhout, T.; Pang, K.-L. Classification of marine Ascomycota, Basidiomycota, Blastocladiomycota and Chytridiomycota. *Fungal Divers.* **2015**, *3*, 1–72. [CrossRef]

22. Le Calvez, T.; Burgaud, G.; Mahé, S.; Barbier, G.; Vandenkoornhuyse, P. Fungal diversity in deep-sea hydrothermal ecosystems. *Appl. Environ. Microbiol.* **2009**, *75*, 6415–6421. [CrossRef] [PubMed]

23. Singh, P.; Raghukumar, C.; Meena, R.M.; Verma, P.; Shouche, Y. Fungal diversity in deep-sea sediments revealed by culture-dependent and culture-independent approaches. *Fungal Ecol.* **2012**, *5*, 543–555. [CrossRef]

24. Pang, K.-L.; Guo, S.-Y.; Chen, I.-A.; Burgaud, G.; Luo, Z.-H.; Dahms, H.U.; Hwang, J.-S.; Lin, Y.-L.; Huang, J.-S.; Ho, T.-W.; et al. Insights into fungal diversity of a shallow-water hydrothermal vent field at Kueishan Island, Taiwan by culture-based and metabarcoding analyses. *PLoS ONE* **2019**, *14*, e0226616. [CrossRef]

25. De Leo, F.; Lo Giudice, A.; Alaimo, C.; De Carlo, G.; Rappazzo, A.C.; Graziano, M.; De Domenico, E.; Urzì, C. Occurrence of the blackyeast *Hortaeawerneckii* in the Mediterranean Sea. *Extremophiles* **2019**, *23*, 9–17. [CrossRef]

26. Gostinčar, C.; Stajich, J.E.; Zupančič, J.; Zalar, P.; Gunde-Cimerman, N. Genomic evidence for intraspecific hybridization in a clonal and extremely halotolerant yeast. *BMC Genom.* **2018**, *19*, 364. [CrossRef]

27. Lenassi, M.; Gostinčar, C.; Jackman, S.; Turk, M.; Sadowski, I.; Nislow, C.; Jones, S.; Birol, I.; Gunde-Cimerman, N.; Plemenitaš, A. Whole genome duplication and enrichment of metal cation transporters revealed by *de-novo* genome sequencing of extremely halotolerant black yeast *Hortaeawerneckii. PLoS ONE* **2013**, *8*, e71328. [CrossRef]

28. Sinha, S.; Flibotte, S.; Neira, M.; Formby, S.; Plemenitaš, A.; Gunde-Cimerman, N.; Lenassi, M.; Gostinčar, C.; Stajich, J.E.; Nislow, C. Insight into the recent genome duplication of the halophilic yeast *Hortaea werneckii*: Combining an improved genome with gene expression and chromatin structure. *G3* **2017**, *7*, 2015–2022. [CrossRef]

29. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]

30. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 4554–4577. [CrossRef]

31. Boetzer, M.; Henkel, C.V.; Jansen, H.J.; Butler, D.; Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **2011**, *27*, 578–579. [CrossRef]

32. Boetzer, M.; Pirovano, W. Toward almost closed genomes with GapFiller. *Genome Biol.* **2012**, *13*, R56. [CrossRef] [PubMed]

33. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef] [PubMed]

34. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness. In *Gene Prediction. Methods in Molecular Biology*; Kollmar, M., Ed.; Humana: New York, NY, USA, 2019; Volume 1962, pp. 2272–2345. [CrossRef]

35. Mapleson, D.; Garcia Accinelli, G.; Kettleborough, G.; Wright, J.; Clavijo, B.J. KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **2017**, *33*, 574–576. [CrossRef] [PubMed]

36. Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, *33*, 2202–2204. [CrossRef]

37. Campbell, M.S.; Holt, C.; Moore, B.; Yandell, M. Genome Annotation and Curation Using MAKER and MAKER-P. *Curr. Protoc. Bioinform.* **2014**, *48*, 4–11. [CrossRef]

38. Korf, I. Gene finding in novel genomes. *BMC Bioinform.* **2004**, *5*, 59. [CrossRef]

39. Stanke, M.; Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **2005**, *33*, W465–W467. [CrossRef]

40. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef]

41. Törönen, P.; Medlar, A.; Holm, L. PANNZER2: A rapid functional annotation web server. *Nucleic Acids Res.* **2018**, *46*, W84–W88. [CrossRef]

42. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [CrossRef] [PubMed]

43. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. In *Gene Prediction. Methods in Molecular Biology*; Kollmar, M., Ed.; Humana: New York, NY, USA, 2019; Volume 1962, pp. 11–14. [CrossRef]

44. Ondov, B.D.; Treangen, T.J.; Melsted, P.; Mallonee, A.B.; Bergman, N.H.; Koren, S.; Phillippy, A.M. Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **2016**, *17*, 132. [CrossRef] [PubMed]

45. Suzuki, R.; Shimodaira, H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **2006**, *22*, 15401–15542. [CrossRef] [PubMed]

46. Denton, J.F.; Lugo-Martinez, J.; Tucker, A.E.; Schrider, D.R.; Warren, W.C.; Hahn, M.W. Extensive error in the number of genes inferred from draft genome assemblies. *Ploscomput. Biol.* **2014**, *10*, e1003998. [CrossRef]

47. Ohm, R.A.; Feau, N.; Henrissat, B.; Schoch, C.L.; Horwitz, B.A.; Barry, K.W.; Condon, B.J.; Copeland, A.C.; Dhillon, B.; Glaser, F.; et al. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. *PLoS Pathog.* **2012**, *8*, e1003037. [CrossRef]

48. Coleine, C.; Masonjones, S.; Sterflinger, K.; Onofri, S.; Selbmann, L.; Stajich, J.E. Peculiar genomic traits in the stress-adapted cryptoendolithic Antarctic fungus *Friedmanniomyces Endolithicus*. *Fungal Biol.* **2020**, *124*, 458–467. [CrossRef]

49. Ma, L.-J.; Ibrahim, A.S.; Skory, C.; Grabherr, M.G.; Burger, G.; Butler, M.; Elias, M.; Idnurm, A.; Lang, B.F.; Sone, T.; et al. Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet.* **2009**, *5*, e1000549. [CrossRef]

50. Corrochano, L.M.; Kuo, A.; Marcet-Houben, M.; Polaino, S.; Salamov, A.; Villalobos-Escobedo, J.M.; Grimwood, J.; Álvarez, M.I.; Avalos, J.; Bauer, D.; et al. Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication. *Curr. Biol.* **2016**, *26*, 1577–1584. [CrossRef]

51. Lidzbarsky, G.A.; Shkolnik, T.; Nevo, E. Adaptive response to DNA-damaging agents in natural *Saccharomyces cerevisiae* populations from Evolution Canyon, Mt. Carmel, Israel. *PLoS ONE* **2009**, *4*, e5914. [CrossRef]

52. Dhar, R.; Sägesser, R.; Weikert, C.; Yuan, J.; Wagner, A. Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution. *J. Evol. Biol.* **2011**, *24*, 1135–1153. [CrossRef]