Clonality, inbreeding, and hybridization in two extremotolerant black yeasts

Cene Gostinčar [1,2,*,¶], Xiaohuan Sun [3,¶], Anja Černoša[1], Chao Fang [3], Nina Gunde-Cimerman [1,+] and Zewei Song [3,+]

[1]Department of Biology, Biotechnical Faculty, University of Ljubljana, 1000 Ljubljana, Slovenia
[2]Lars Bolund Institute of Regenerative Medicine, BGI-Qingdao, Qingdao 266555, China
[3]BGI-Shenzhen, Beishan Industrial Zone, Shenzhen 518083, China
*Correspondence address. Cene Gostinčar, Department of Biology, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, SI-1000 Ljubljana, Slovenia.
E-mail: cene.gostincar@bf.uni-lj.si, cgostincar@gmail.com
[¶]These authors contributed equally to this work as first authors.
[+]These authors contributed equally to this work as senior authors.

Abstract

Background: The great diversity of lifestyles and survival strategies observed in fungi is reflected in the many ways in which they reproduce and recombine. Although a complete absence of recombination is rare, it has been reported for some species, among them 2 extremotolerant black yeasts from Dothideomycetes: *Hortaea werneckii* and *Aureobasidium melanogenum*. Therefore, the presence of diploid strains in these species cannot be explained as the product of conventional sexual reproduction.
Results: Genome sequencing revealed that the ratio of diploid to haploid strains in both *H. werneckii* and *A. melanogenum* is about 2:1. Linkage disequilibrium between pairs of polymorphic loci and a high degree of concordance between the phylogenies of different genomic regions confirmed that both species are clonal. Heterozygosity of diploid strains is high, with several hybridizing genome pairs reaching the intergenomic distances typically seen between different fungal species. The origin of diploid strains collected worldwide can be traced to a handful of hybridization events that produced diploids, which were stable over long periods of time and distributed over large geographic areas.
Conclusions: Our results, based on the genomes of over 100 strains of 2 black yeasts, show that although they are clonal, they occasionally form stable and highly heterozygous diploid intraspecific hybrids. The mechanism of these apparently rare hybridization events, which are not followed by meiosis or haploidization, remains unknown. Both extremotolerant yeasts, *H. werneckii* and even more so *A. melanogenum*, a close relative of the intensely recombining and biotechnologically relevant *Aureobasidium pullulans*, provide an attractive model for studying the role of clonality and ploidy in extremotolerant fungi.

Keywords: population genomics, halotolerance, extremotolerance, halophilic fungus, *Hortaea werneckii*, *Aureobasidium melanogenum*, hybridization

Introduction

No single reproductive strategy is optimal for all species and all conditions in which they live. This results in the coexistence of a wide variety of ways in which organisms reproduce and recombine their genetic material. Among the most diverse are fungi, which exhibit a wide range of strategies, from strictly clonal species, which do not recombine at all [1], to species with thousands of mating types [2]. Sexual, parasexual, and clonal reproduction are broad categories that encompass a variety of different phenomena, processes, and mechanisms, some of which are typical of larger groups (e.g., the dikaryon of basidiomycetes), while others differ even among closely related species [3].

Traditionally, up to one-fifth of fungi were thought to be asexual [4]. Subsequent genetic and genomic analyses found at least a rudimentary mating-type locus in nearly every species studied. Population genetics showed that most species previously thought to be asexual are actually recombining [5–7]. However, in most species, asexual reproduction dominates over occasional recombination. To account for the observation that some apparently clonal species can nevertheless recombine at levels low enough not to break the pattern of population structure typical of clonal species, Tibayrenc and Ayala [5] introduced the concept of "restricted recombination." Yet some species appear to be strictly clonal, even by highly sensitive measures of recombination used by population genomics, such as linkage disequilibrium [1, 8]. This study focuses on 2 species of such strictly clonal fungi. The analysis of 115 genomes of haploid and diploid wild strains of these species is used to test the hypothesis that even clonal phylogenetic lineages can generate diversity through hybridization that produces highly heterozygous and stable diploids.

Several reasons for the pervasive clonality in fungi have been proposed, such as hybrid incompatibility or limited opportunities to meet strains of the opposite mating type [6]. Severe bottlenecks (e.g., following the introduction of a small number of strains to a new site) can lead to a skewed balance of mating types. Strains of pathogens with host-to-host transmission may encounter other strains of the species very rarely [6, 9]. Similar may be true for fungi with poor dispersal abilities that are restricted to rare and isolated ecological islands, such as certain extreme environments [1, 10]. However, even without these constraints, sexually meiotic reproductive events are often majorly outweighed by asexual mitotic events, as shown in *Saccharomyces paradoxus* [11].

The absence of recombination carries the risk of accumulation of deleterious mutations, a process known as Muller's ratchet. Sexual reproduction remedies this and efficiently generates diversity, which is a substrate for selection and adaptation to novel conditions. But asexual reproduction has its own advantages. For example, it can conserve energy by eliminating the need to maintain the mating system and form sexual structures [12, 13]. It also allows the organism to faithfully reproduce successful genomic configurations and thus avoid the recombination load, a loss of fitness due to a breakup of advantageous combinations of interacting alleles [13–15]. This may be particularly beneficial in specialists, such as those inhabiting extreme environments [15, 16].

In addition to sexual reproduction, fungi can employ another tool to recombine their genomes: parasexuality [17]. Two cells can fuse to combine their genetic material, providing the opportunity for mitotic recombination. Haploid parents thus produce diploid offspring, but this ploidy change is generally considered unstable. However, it does not revert to the original ploidy of the parental strains through the tightly controlled and high-fidelity process of meiosis. Instead, chromosomes are lost randomly through a series of aneuploid generations. The importance of parasexuality outside of laboratory settings has been questioned [18], but in at least some cases, the process appears to drive adaptation and facilitate survival (e.g., in *Aspergillus fumigatus* in the lungs of patients with cystic fibrosis) [19].

Changes in ploidy itself may be a form of adaptation, either through parasexuality or other processes such as abnormal cell division [20]. Both polyploidy and aneuploidy can be a response to adverse or novel environmental conditions [21–23]. They influence fitness through changes in cell size and shape, changes in the transcriptome (by altering gene dosage) and in the rate of adaptation, but also by providing new options for repairing DNA damage and temporarily masking deleterious mutations [20–22, 24]. A growing body of evidence shows that variation in ploidy is a widespread transient adaptation of fungi to novel conditions (reviewed by Naranjo-Ortiz and Gabaldón [22]). Aneuploidies are more common and more easily reversed than tandem gene duplications, which are an alternative way to increase gene dosage. During cultivation under optimal conditions, such altered ploidies tend to revert to a baseline ploidy of the species, often without clear increases in fitness [21, 25].

Sexual and parasexual recombination can lead to recombinant lineages and interspecific hybrids [26], another process increasingly recognized as an important generator of fungal diversity, including in industrial and clinical settings (reviewed by Naranjo-Ortiz and Gabaldón [22]). With the increasing accessibility of genome sequencing, research on this topic is rapidly expanding to nonmodel species. Some hybrids exhibit higher fitness than their parental strains [27], making hybridization an important driver of adaptation to novel environments [22, 28]. Divergent hybrid genomes can be stabilized by chromosomal aberrations [29, 30], and the outcome of hybridization is often similar to parasexuality. Hybrids of *Cryptococcus neoformans* and *Cryptococcus gattii*, for example, rapidly lose chromosomes and rearrange them [29, 30].

Indications of (intraspecific) hybridization were also reported by Gostinčar et al. [1] in the extremely halotolerant black yeast *Hortaea werneckii* (Capnodiales, Dothideomycetidae, Dothideomycetes, Pezizomycotina, Ascomycota), a globally distributed species specialized for survival in saline environments and able to grow in nearly salt-saturated solutions [31, 32]. Whole-genome sequencing of 12 strains indicated that the species is clonal but also that a majority of the strains are highly heterozygous

diploids. These diploids appeared to be stable enough to spread over considerable distances, with little evidence of haploidization [1, 33]. This explained the ploidy of the reference genome, which was originally interpreted as the result of endoreduplication [34, 35]. Subsequent genome sequencing of 2 additional *H. werneckii* strains provided additional support for the hybridization hypothesis [36]. However, the total number of sequenced *H. werneckii* genomes remained low, limiting the power of the analyses and the interpretation of the results.

A similar pattern of haploid and diploid strains coexisting within an apparently clonal species has since been discovered in *Aureobasidium melanogenum* (Dothideales, Dothideomycetidae, Dothideomycetes, Pezizomycotina, Ascomycota), another black yeast only distantly related to *H. werneckii* [8]. While *A. melanogenum* tolerates less extreme conditions than *H. werneckii*, it is tolerant of a wider range of types of stress and occurs in a variety of environments, from hypersaline waters to various indoor habitats (reviewed by Černoša et al. [8]).

The role of hybridization and ploidy changes are among the overlooked dimensions of fungal genetics [22]. Here we analyze 66 genomes of *H. werneckii* and 49 genomes of *A. melanogenum* to provide insight into the reproductive strategy of these 2 extremotolerant fungi, characterized by coexistence of haploid and highly heterozygous diploid strains that are stable over large geographic and temporal distances.

Results

Whole genomes of 54 strains of the extremely halotolerant black yeast *H. werneckii* were sequenced. Combined with previously sequenced strains [1, 35], this resulted in a data set of 66 whole-genome sequences (Table 1, Supplementary Table S1). A majority of strains (26) were isolated from brine, evaporation-concentrated seawater during salt extraction, followed by strains isolated from bittern (7), a saturated, magnesium-rich solution that remains after precipitation of halite during salt extraction. Nine strains were isolated from a seacoast cave in Atacama, where some of the strains grew on spider webs along with the alga *Dunaliella atacamensis* [37]. Twelve strains were isolated from marine habitats and 4 were clinical isolates. All genomes of *A. melanogenum* (Table 2) were sequenced and described in a previous study [8]. The largest number of strains (19) was isolated from bathroom and kitchen surfaces (including from kitchen appliances), followed by 16 strains from tap water or springs of tap water. In the case of both *H. werneckii* and *A. melanogenum*, the majority of strains were isolated in Slovenia.

Based on previous studies [1, 34, 35], the haploid genomes of both *H. werneckii* and *A. melanogenum* are approximately 25 Mbp in size. Comparing the sizes of genome assembly and the number of predicted genes in each genome, 20 (30%) *H. werneckii* genomes were recognized as haploid, 45 (68%) as diploid, and 1 genome as tetraploid. This was similar to the *A. melanogenum* genomes where 16 (33%) genomes were recognized as haploid and 30 (61%) as diploid, and the ploidy of 3 genomes (2, 18, 38) was unclear [8].

The distribution of assembly size, number of predicted genes, and other genomic characteristics within both haploid and diploid *H. werneckii* groups was narrow (Table 3). The average genome assembly size was 26.52 Mbp (±1.47 SD) for haploid and 49.30 Mbp (±1.74 SD) for diploid genomes. The average number of predicted genes was 9,519 (±665 SD) for haploids and 20,417 (±1,709 SD) for diploids. As expected, the quality of assembly was much lower in diploid strains, as evidenced by a higher number of contigs and a smaller N50 value, presumably due to regions of high similarity

Table 1: *Hortaea werneckii* strains analyzed in this study

Culture collection strain number	Present study number*	Isolation habitat	Sampling site location	Ploidy
EXF-9	1	brine	Ebre Delta salterns, Spain	1
EXF-12	2	brine	Santa Pola salterns, Spain	1
EXF-15	3	brine	Santa Pola salterns, Spain	1
EXF-20	4	brine	Santa Pola salterns, Spain	2
EXF-152	5	brine	Sečovlje salterns, Slovenia	2
EXF-153, EXF-2781	6	brine	Sečovlje salterns, Slovenia	2
EXF-154	7	brine	Sečovlje salterns, Slovenia	2
EXF-156, CBS 116.90	8	eye infection of aquarium *Spondyliosoma cantharus*	unknown	2
EXF-157, CBS 115.90	9	kidney of *Bufo granulosus*	Brazil	1
EXF-161, EXF-2689, CBS 706.76	10	leaf of *Rhizophora mangle*	Senegal	2
EXF-166, CBS 100496	11	seawater-sprayed marble	Delos, Greece	2
EXF-177, CBS 705.76	12	*Tinea nigra*	France	1
EXF-241	13	brine	Sečovlje salterns, Slovenia	2
EXF-269, EXF-108	14	brine	Santa Pola salterns, Spain	2
EXF-291	15	brine	Sečovlje salterns, Slovenia	2
EXF-561	16	brine	Namibia, salterns at the Atlantic coast	1
EXF-2515	17	brine	salterns, Puerto Rico	1
EXF-2516	18	brine	salterns, Puerto Rico	1
EXF-2683, CBS 117.90	19	salted fish, *Osteoglossum bicirrhosum*	Brazil	2
EXF-2685	20	brine	Sečovlje salterns, Slovenia	1
EXF-2783	21	brine	Sečovlje salterns, Slovenia	1
EXF-2785	22	brine	Sečovlje salterns, Slovenia	2
EXF-3845	23	brine	Candelaria salterns, Puerto Rico	1
EXF-3846	24	brine	Candelaria salterns, Puerto Rico	1
EXF-4716	25	brine bait	Sečovlje salterns, Slovenia	2
EXF-6274	26	brine	Sečovlje salterns, Slovenia	2
EXF-6652	27	spider web in a cave close to the ocean	Atacama, Chile	2
EXF-6663	28	spider web in a cave close to the ocean	Atacama, Chile	1
EXF-8170	29	brine	Sečovlje salterns, Slovenia	2
EXF-8422	30	biofilm from cheese factory brine	Celje, Slovenia	2
EXF-10304	31	brine	Sečovlje salterns, Slovenia	2
EXF-10508	32	seawater, depth 25 m	Italy	2
EXF-10509	33	seawater, depth 200 m	Italy	2
EXF-10510	34	seawater, depth 94 m	Italy	2
EXF-10511	35	seawater, depth 25 m	Italy	2
EXF-10512	36	seawater, depth 25 m	Italy	4
EXF-10816	37	bittern after halite precipitation	Sečovlje salterns, Slovenia	2
EXF-10819	38	bittern after halite precipitation	Sečovlje salterns, Slovenia	2
EXF-10820	39	bittern after halite precipitation	Sečovlje salterns, Slovenia	1
EXF-10843	40	brine	Sečovlje salterns, Slovenia	2
EXF-10904	41	bittern after halite precipitation	Sečovlje salterns, Slovenia	2
EXF-10907	42	bittern after halite precipitation	Sečovlje salterns, Slovenia	2
EXF-10919	43	bittern after halite precipitation	Sečovlje salterns, Slovenia	2
EXF-10958	44	bittern after halite precipitation	Sečovlje salterns, Slovenia	1
EXF-10974	45	brine	Sečovlje salterns, Slovenia	2
EXF-11540	46	sand in a cave close to the ocean	Atacama, Chile	2
EXF-11650	47	sand in a cave close to the ocean	Atacama, Chile	2
EXF-11651	48	sand in a cave close to the ocean	Atacama, Chile	1
EXF-12619	49	coral or deep sea	China	1
EXF-12620	50	coral or deep sea	China	2
EXF-14591, CMF-020	51	plankton tow	Vineyard Sound, USA	2
EXF-14592, AMF 061	52	plankton tow	Vineyard Sound, USA	1
EXF-225**	53	malt extract medium, 25% NaCl (w/v)	long-term experimental evolution	2
EXF-14590, MSW 12–1B	54	marine	List on Sylt, Germany	2
EXF-2000	A***	brine	Sečovlje salterns, Slovenia	2
EXF-120	B	brine	Santa Pola saltpans, Spain	2
EXF-562	C	soil on the sea coast	Namibia	1
EXF-2788	D	brine	Sečovlje salterns, Slovenia	1
EXF-171	E	keratomycosis	Brazil	2
EXF-2682	F	*Trichomycosis nigra*	Italy	2

Table 1: (*Continued*)

Culture collection strain number	Present study number*	Isolation habitat	Sampling site location	Ploidy
EXF-10513	G	deep seawater	Italy	2
EXF-151	H	*Tinea nigra*	Portugal	2
EXF-6651	I	spider web in a cave close to the ocean	Atacama, Chile	2
EXF-6669	J	spider web in a cave close to the ocean	Atacama, Chile	2
EXF-6654	K	spider web in a cave close to the ocean	Atacama, Chile	2
EXF-6656	L	rock wall in a cave close to the ocean	Atacama, Chile	2

∗ Strains 1 to 54 were sequenced in this study; strains A to L were sequenced and named by Gostinčar et al. [1].
∗∗ Strain EXF-225 after 15 years of repeated subcultivation at 25% NaCl (w/v), continuation of experiment described in Gostinčar et al. [46].
∗∗∗ Reference *H. werneckii* genome [35]; naming of strains A to L corresponds to names in Gostinčar et al. [1].

between the 2 haploid subgenomes, a long-standing challenge in sequencing *H. werneckii* genomes [35]. Nevertheless, the assembly and annotation of all strains were of reasonable quality, with only about 3.19% (±0.30 SD) BUSCOs missing completely in the haploid genomes and 6.09% (±4.24 SD) in the diploid genomes.

Single-nucleotide polymorphisms (SNPs) were determined with Genome Analysis Toolkit after mapping the sequencing reads to reference genomes (haploidized genome of diploid strain EXF-2000 for *H. werneckii*, whole genome of haploid strain EXF-3378 for *A. melanogenum*). SNP analysis was performed on all strains except *H. werneckii* strain 36 due to its tetraploid genome. The average density of SNPs in haploid strains of *H. werneckii* was high: 4.04% (±1.11 SD) (Table 3). For diploids, the SNP density was 3.44% (±1.12 SD), of which 71% of the loci were heterozygous. In *A. melanogenum*, the average SNP density was 4.41% (±1.87 SD) in haploids and 3.79% (±0.21 SD) in diploids, with 44% of the latter heterozygous. Based on the SNP data, the genomes of both *H. werneckii* and *A. melanogenum* were clustered into 5 clusters in principal component analysis (PCA), with the first 2 principal components explaining 57.2% of the SNP diversity of *H. werneckii* and 59.3% of *A. melanogenum* (Fig. 1). SNP-based phylogenetic analyses of both species revealed considerable reticulation (Fig. 1). The largest cluster of strains identified by both network analysis and PCA contained 18 strains in *H. werneckii* and 20 strains in *A. melanogenum*.

The squared correlation coefficient (r^2) was calculated for all pairs of biallelic SNP loci present in at least 2 genomes analyzed and within 10 kbp of each other. Plotting r^2 as a function of the distance between pairs of loci showed very little decay of linkage disequilibrium in either species from the maximum initial values of 0.17 for *H. werneckii* and 0.22 for *A. melanogenum*. Linkage disequilibrium remained high above half of the maximum value even for alleles that were 10 kbp apart (Fig. 2). Such strong linkage between loci can be explained by a lack of recombination that would break the linkage, confirming previous reports that *H. werneckii* and *A. melanogenum* are strictly clonal [1, 8].

The phylogenies of the 50 longest alignable genomic regions were also consistent with the presumed lack of recombination within *H. werneckii* and *A. melanogenum*. The phylogenetic trees showed a high degree of concordance (Fig. 3), meeting the "strong phylogenetic signal" criterion for clonality [5]. Sequences representing different haploid subgenomes of diploid strains were positioned in different parts of the phylogeny, corresponding to the high heterozygosity of the strains. An extreme case of this was the tetraploid *H. werneckii* strain 36, which was positioned in 4 different parts of the phylogeny. When all 50 multilabeled trees for each species were collapsed into a consensus supernetwork

(Fig. 3), the result was similar to the phylogenetic network estimated from SNP data (Fig. 1).

The topology of the consensus phylogenies and supernetworks was best explained by a number of intraspecific hybridization events in each species: 9 or 10 events for *H. werneckii* (with an additional event leading to tetraploid strain 36) and 7 events for *A. melanogenum* (Fig. 4). Several phylogenetic lineages resulting from these events appeared to be relatively widespread—more than 1 representative strain was found for most lineages, often in different habitats and geographic locations. However, lineage composition was skewed in favor of specific localities or habitats. This was confirmed by the Fisher exact test, which found significant differences between groups in both the isolation habitat and the geographic location of origin for both species ("*H. werneckii*—habitats": $P < 0.01$; all other: $P < 0.001$). For example, for both species, the 2 largest groups (in both cases marked as group 1 in Fig. 4) contained isolates from Europe, with only 1 exception. In terms of habitats, *H. werneckii* group 1 was isolated mainly from hypersaline habitats and group 9 from seawater; groups 4 and 5 were found mainly in a cave on a desert coast. The largest group of haploid strains was also found mainly in hypersaline habitats. Clinical isolates of *H. werneckii* belonged to different phylogenetic lineages. The tetraploid *H. werneckii* strain 36 was isolated from the deep sea (Figs. 3, 4) and most likely arose by hybridization between diploid hybrids of groups 1 and 9. In the case of *A. melanogenum*, most strains originated in Europe. Some groups showed a preference for particular habitats: *A. melanogenum* group 1 was mostly isolated from tap water and its sources, while the majority of isolates from household surfaces were classified into other groups.

Aneuploidy in the genomes of *H. werneckii* and *A. melanogenum* was investigated by searching for large genome segments with different sequencing coverage from the rest of the genome. Evidence of aneuploidy was found in 23 genomes of *H. werneckii* (35%) and 8 genomes of *A. melanogenum* (16%) (Fig. 5, Supplementary Figs. S1, S2). The majority of these genomes were diploid: 18 (78%) for *H. werneckii* and 8 (100%) for *A. melanogenum*. Some parts of the genome were aneuploid in several strains, with most aneuploid parts representing an increase in ploidy rather than a decrease. The aneuploid strains included 3 of 4 clinical *H. werneckii* isolates and the only clinical isolate of *A. melanogenum* in the study. In some diploid genomes, loss of heterozygosity was observed over large regions (Supplementary Fig. S3). Some of these could be explained by aneuploidy (loss of 1 copy of a chromosome or part of chromosome), while others appeared to be copy neutral, possibly caused by mitotic recombination.

A putative mating-type locus was found in the majority of the genomes. While mapping of sequencing reads of some strains to

Table 2: *Aureobasidium melanogenum* strains analyzed in this study

Culture collection strain number	Present study number*	Isolation habitat	Sampling site location	Ploidy
EXF-924	1	ponds on sea ice	Svalbard, Norway	1
EXF-926	2	surface glacial ice	Svalbard, Norway	2(?)**
EXF-3233	3	deep sea (4,500 m b.s.l.)	Japan	1
EXF-3371	4	soil	Thailand	1
EXF-3378	5	public fountain	Thailand	1
EXF-3397	6	endoperitoneal fluid	Greece	2
EXF-4450	8	Iskra factory	Slovenia	2
EXF-5590	9	dishwasher rubber seal	Slovenia	2
EXF-6171	10	glacial ice	Argentina	2
EXF-7932	11	metal drain on the kitchen sink	Sweden	1
EXF-7946	12	kitchen metal holder for washed dishes	Sweden	1
EXF-8016	13	bathroom faucet and sink contact	Sweden	1
EXF-8022	14	refrigerator inner surface	Sweden	1
EXF-8044	15	kitchen metal holder for washed dishes	Sweden	1
EXF-8258	16	well water	Slovenia	2
EXF-9877	17	tap water	Slovenia	2
EXF-11403	18	refrigerator inner surface	Sweden	2(?)**
EXF-8492	19	well water	Slovenia	2
EXF-8678	20	well water	Slovenia	2
EXF-8689	21	well water	Slovenia	2
EXF-8695	22	well water	Slovenia	2
EXF-8702	23	well water	Slovenia	2
EXF-8986	24	fango mud from Sečovlje salterns	Slovenia	2
EXF-9262	25	rubber on kitchen drain	Slovenia	1
EXF-9470	26	kitchen counter above dishwasher	Slovenia	2
EXF-9272	27	kitchen strainer basket	Slovenia	1
EXF-9298	28	plastic mesh on kitchen drain	Slovenia	2
EXF-9304	29	kitchen strainer basket	Slovenia	2
EXF-9313	30	kitchen sink	Slovenia	2
EXF-9454	31	tap water	Slovenia	2
EXF-9484	32	kitchen counter above dishwasher	Slovenia	2
EXF-9887	33	tap water	Slovenia	2
EXF-9516	34	kitchen sink drain	Slovenia	2
EXF-9539	35	kitchen strainer basket	Slovenia	1
EXF-9540	36	dishwasher door	Slovenia	2
EXF-10064	37	tap water	Slovenia	2
EXF-11060	38	ceiling surface	Slovenia	2(?)**
EXF-9875	39	tap water	Slovenia	2
EXF-9906	40	*Arthrocnemum* sp. plant surface from Sečovlje saltern	Slovenia	1
EXF-9911	41	kitchen sink drain	Slovenia	2
EXF-9937	42	kitchen sink drain	Slovenia	2
EXF-10061	43	tap water	Slovenia	2
EXF-10062	44	tap water	Slovenia	2
EXF-10066	45	tap water	Slovenia	2
EXF-10333	46	tap water	Slovenia	2
EXF-10372	47	air in the National Gallery restoration center	Slovenia	1
EXF-10726	48	integument of a male alate ant of *Atta sexdens rubropilosa*	Brazil	1
EXF-11028	49	water from the aquarium with *Proteus anguinus*	Slovenia	2

* Same numbering as in Černoša et al. [8]. ** Ploidy unclear, see Černoša et al. [8].

the reference mating-type locus of each species contained gaps due to high divergence of the locus, in some other strains, the sequencing coverage indicated the presence of the locus despite its absence in the whole-genome assembly (Supplementary Figs. S4, S5). The poor assembly of the locus was particularly problematic in diploid genomes and even more so in the tetraploid genome of *H. werneckii*, precluding a conclusive analysis in all strains. The absence of the locus in some strains should at this point in time not be seen as a conclusive result. However, where entire genes MAT1-1 and MAT1-2 were assembled, they showed a substantial diversity, clustering into distinct phylogenetic groups (Supplementary Figs. S4, S5). For *H. werneckii*, "blastx" searches against the nonre-

Table 3: Statistics for the *H. werneckii* genomes sequenced in this study (strains 1–54)

	Haploid strains			Diploid strains			Tetraploid strain
	Average	Median	SD	Average	Median	SD	/
Coverage	730	619	464	469	485	177	276
Genome assembly size (Mbp)	26.52	26.19	1.47	49.30	49.22	1.74	94.67
Number of contigs	796	638	421	6,885	3,806	4,457	30,312
Contig N50 (kbp)	136	138	28	22	26	14	5
GC content	53.22%	53.33%	0.33%	53.40%	53.40%	0.19%	53.40%
CDS total length (Mbp)	14.56	14.39	0.80	27.02	27.87	1.45	49.02
CDS total length (% of genome)	54.94%	55.27%	1.67%	54.80%	55.59%	2.15%	51.78%
Gene models (*n*)	9,519	9,344	665	20,417	19,240	1,709	46,596
Exons per gene (average)	2.34	2.34	0.06	2.10	2.20	0.14	1.87
Intron average length (bp)	93.17	93.00	2.53	94.11	94.00	4.73	92.00
Complete BUSCOs	95.99%	96.00%	0.34%	86.86%	93.40%	10.09%	89.60%
Complete and single-copy BUSCOs	95.83%	95.90%	0.35%	21.33%	16.30%	10.50%	33.20%
Complete and duplicated BUSCOs	0.16%	0.20%	0.06%	65.53%	77.30%	19.20%	56.40%
Fragmented BUSCOs	0.82%	0.75%	0.14%	7.05%	3.20%	5.87%	5.10%
Missing BUSCOs	3.19%	3.10%	0.30%	6.09%	3.30%	4.24%	5.30%
Total SNP density (SNPs per total genome size)	4.04%	4.54%	1.11%	3.44%	3.56%	1.12%	/
Heterozygous SNP density (SNPs per total genome size)	0.01%	0.01%	0.01%	2.46%	2.58%	0.74%	/

dundant GenBank protein database showed that 1 large phylogenetic group of each locus was highly similar to homologues from other fungi ("true" MAT1-1 and MAT1-2), while the other groups only produced matches with these "true" *H. werneckii* homologues, but not with homologues from other fungi, especially in case of MAT1-2, a result of an intense diversification. For *A. melanogenum*, all putative MAT1-1 and MAT1-2 could be matched to homologues from other fungi. Phylogenetic groups of MAT1-1 and MAT1-2 in both species generally corresponded to hybrid and haploid groups of strains but with numerous exceptions (Supplementary Figs. S4, S5). For example, for *H. werneckii*, hybrid groups generally contained 2 similar copies of MAT1-1, but groups 1a, 3, and 4 contained very different copies and group 9 consistently lacked 1 copy of MAT1-1 altogether (possibly due to poor assembly of the locus). In contrast, 2 different copies of MAT1-2 were found in group 1, but 2 similar copies in group 9. Three distinct copies of the mating-type locus were found in strain 10. In *A. melanogenum* hybrid, group 1 contained only a single well-assembled homologue of MAT1-1 and 2 homologues of MAT1-2, but otherwise, the diversity of the mating-type locus in this species was generally lower and mostly corresponded to groups of hybrids and haploid strains.

The search for proteins of the Pfam families HET and Het-C, typical of heterokaryon incompatibility proteins, identified a large number of such proteins in the predicted proteomes of *A. melanogenum* (on average, 4.87 HET and 3.74 Het-C proteins per strain) and even more in *H. werneckii* (on average, 27.82 HET and 3.5 Het-C proteins per strain) (Supplementary Table S2). Phylogenetic analysis of these proteins showed that they formed several clusters, some of which contained representatives only from specific groups of hybrid strains (as identified in Fig. 4). For example, in *H. werneckii*, the hybrid genomes of groups 9 and 10 (and in 1 case genome 23) were the only ones to contain HET genes belonging to phylogenetic clusters 6 and 24 (Supplementary Table S2). HET proteins from cluster 12 were found only in hybrid groups 1 to 3 and their closely related haploid strains (and tetraploid strain 36). In *A. melanogenum*, hybrid groups 5 and 6 were the only ones to contain HET proteins belonging to a small phylogenetic cluster

5 (Supplementary Table S2), and similarly, hybrid group 2 was the only one to contain HET proteins from phylogenetic group 6.

Discussion

Genome sequencing of 66 strains of the black yeast *H. werneckii* and 49 strains of the black yeast *A. melanogenum* revealed some unexpected similarities between these species, which belong to different orders of the subclass Dothideomycetidae. Approximately one-third of the sequenced strains of each species were haploid and approximately two-thirds were diploid. Principal component analysis of single-nucleotide polymorphisms identified several clusters of strains in each species. In both cases, the first 2 principal components explained nearly 60% of the observed diversity—much more than, for example, in the homogeneous and recombining species *A. pullulans*, where the first 2 principal components together explained less than 15% of the diversity [38]. The clustering of strains was consistent with the previous reports that both *H. werneckii* and *A. melanogenum* are strictly clonal [1, 8]. Despite the presence of a mating-type locus in the reference genomes of both species [1, 39], the clonality of the species was confirmed here by a high degree of concordance between the phylogenetic histories of different genomic regions and by a lack of decay in linkage disequilibrium, an established measure of recombination often expressed as the distance over which linkage disequilibrium falls to half its maximum value [6].

The existence of highly heterozygous intraspecific hybrids, first observed in *H. werneckii* [1], is confirmed here on a much larger genomic data set of *H. werneckii* and also *A. melanogenum*. The mechanism of hybridization is unknown and could range from vegetative hyphal fusion between different haploid strains to plasmogamy and karyogamy of gametes. Regardless of the mechanism, at least some diploid hybrids appear to be stable over long periods of time, allowing them to disperse over long distances and constitute a large proportion of the species in some habitats. Hybrid strains have previously been reported in many other fungal species, including *Saccharomyces cerevisiae* [40], *Candida tropicalis* [41], and *Cryptococcus neoformans* [42], but none of these species
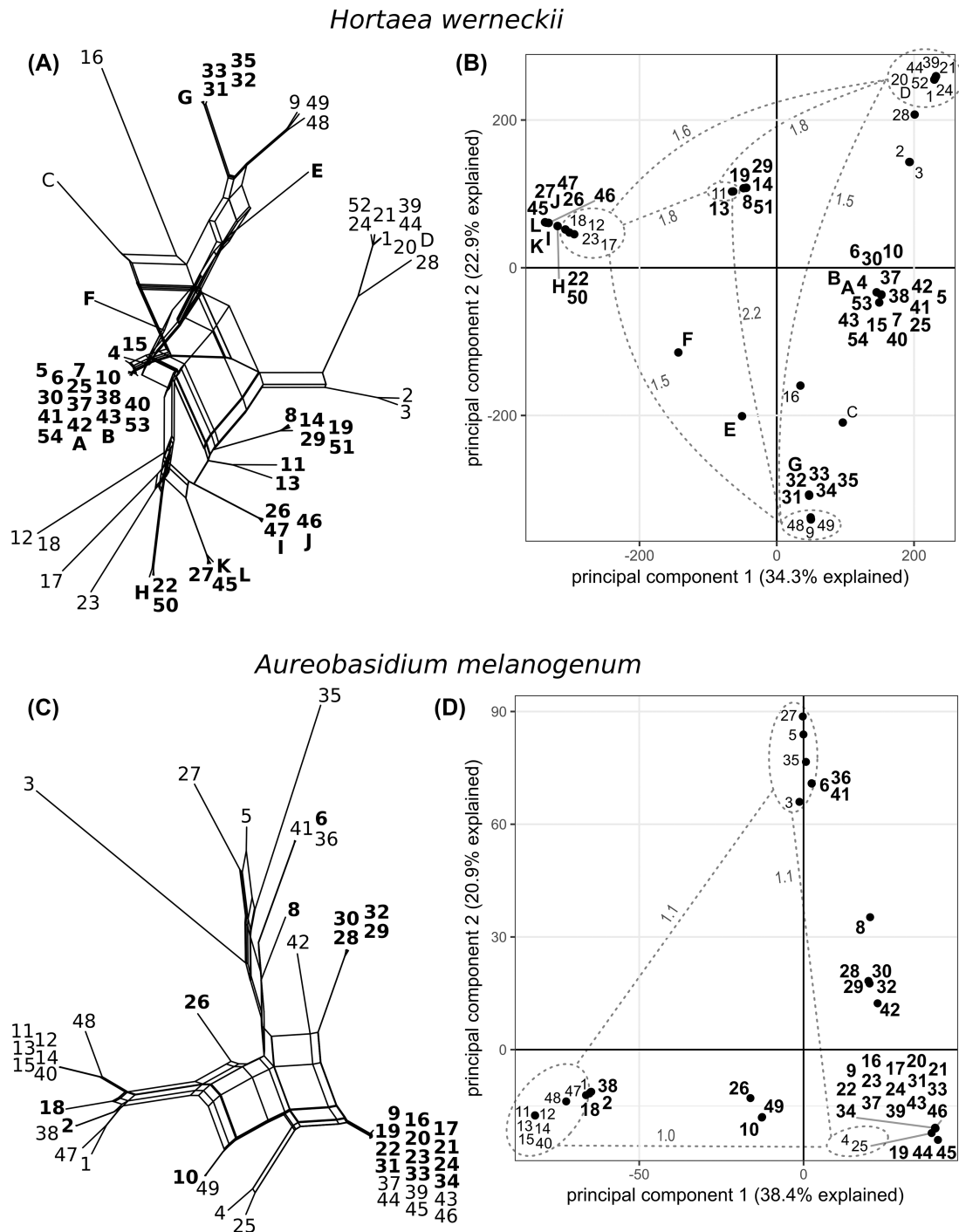
Figure 1: Single-nucleotide polymorphism (SNP) diversity of *Hortaea werneckii* (A, B) and *Aureobasidium melanogenum* (C, D). Names of diploid strains are written in bold. (A, C) Phylogenetic networks reconstructed with a Neighbor-Net algorithm from a dissimilarity distance matrix calculated from SNP data. (B, D) Principal component analysis of SNPs. The genomes are represented by circles. The average divergence between groups of haploid genomes (dashed lines) is expressed as millions of SNPs (numbers next to dashed lines).

are strictly clonal. However, diploid hybrids that cannot reproduce sexually have also been reported in some species [43]. The majority of *Candida orthopsilosis* are hybrids with 5% divergence between their haploid genomes at the nucleotide level, arising from at least 4 independent hybridization events [44]. *Candida metapsilosis* is also a species originating from hybridization, with a similar divergence between haploid genomes [45]. It is possible that *H. werneckii* and *A. melanogenum* follow a similar reproductive strategy as the *Candida parapsilosis/C. orthopsilosis/C. metapsilosis* group,

but in the case of both black yeasts, the hybrids are not recognized as separate species (for reasons discussed below).

In the reticulate history of *H. werneckii* and *A. melanogenum*, the unit of genetic exchange are whole haploid genomes. This allowed us to trace phylogenies of haploid genomes, for example, in the case of *H. werneckii* (Fig. 4A) of the "green haploid" genome (strains 2 and 3) and the "red haploid" genome (strains 4 and C), as well as their "red and green diploid" hybrids (group 1). The absence of haploid strains with genomes combined from 2 or more genomic
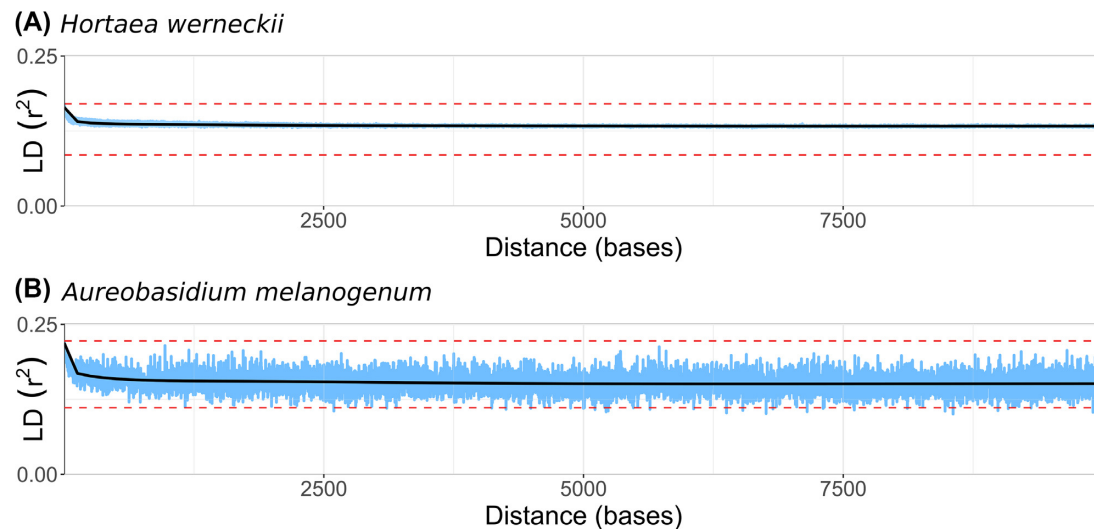
(A) *Hortaea werneckii*

(B) *Aureobasidium melanogenum*

Figure 2: Linkage disequilibrium (LD) decay in *Hortaea werneckii* (A) and *Aureobasidium melanogenum* (B). Squared correlation coefficient (r^2) was calculated for all pairs of nonsingleton biallelic loci within the distance of 10 kbp or less and plotted as a function of the distance between the loci (blue line). The maximum observed value and its half value are marked with red horizontal dashed lines. A generalized additive model curve was fitted to the data (black line).

lineages implies that—as is often seen in hybrids—diploids do not revert to haploid state but are stuck at the F1 stage. They are unable to undergo meiosis, or the progeny of such meiosis has sufficiently low viability or fitness to evade sampling. The presence of a putative homothallic mating-type locus in the majority of sequenced strains and the distribution of diversified lineages of MAT genes in groups of hybrids and haploids raise the possibility that these loci play a role in the formation of hybrid lineages. At the same time, a substantial number of strains containing unexpected combinations of MAT lineages, the fast diversification of the MAT genes (particularly in *H. werneckii*), and the methodological limitations of the analysis (suboptimal assembly of the putative mating-type locus) demand further research before providing conclusive evidence on the presence, type, and functionality of mating-type genes, as well as their possible role in the formation of hybrids in *H. werneckii* and *A. melanogenum*.

While hybridization to diploids and the absence of meiosis are reminiscent of parasexuality, *H. werneckii* and *A. melanogenum* do not conform to this mode of reproduction either. In parasexual reproduction, diploids typically revert to haploids through haploidization—a random loss of chromosomes with the end result similar to meiosis [18]. In addition to the absence of recombinant haploid strains, there is also little evidence for the existence of intermediate aneuploid states characteristic of haploidization. Either such haploidization does not occur, or its products are not viable due to incompatibility of parental chromosomes. The aneuploidy observed in both *H. werneckii* and *A. melanogenum* mostly involves increases in ploidy above the diploid state rather than decreases below it, as would be expected with haploidization. Thus, this aneuploidy is almost certainly not part of a parasexual cycle but might be an adaptive evolutionary response to adverse or novel conditions—a common adaptive response of fungi [21–23]. This explanation is also supported by the observations of large-scale duplications of specific genomic regions in a diploid strain of *H. werneckii* subjected to long-term experimental evolution at extreme salinity [46]. Interestingly, the high level of heterozygosity of diploid strains of *H. werneckii* and *A. melanogenum* resulting from hybridization is mostly preserved not only by the rarity of large-scale deletions but also by the relative scarcity of large-scale

copy-neutral loss of heterozygosity. In other hybrids, loss of heterozygosity has been recognized as a common tool for genome shaping and stabilization after hybridization [47, 48] but appears to be largely avoided by both *H. werneckii* and *A. melanogenum*.

An integral part of fungal parasexuality is the heterokaryon—a cell with 2 genetically distinct nuclei that sometimes undergo karyogamy [49]. In *H. werneckii*, both haploid and diploid cells contain a single nucleus per cell [50]. The viability and stability of heterokaryons are controlled by heterokaryon incompatibility loci. At least 3 genes for proteins with domains characteristic of such loci were found in all *A. melanogenum* genomes and at least 17 in *H. werneckii* (with up to 43 in diploid genomes, although this number may be an overestimate due to fragmented genome assembly). While it has been shown in some species that even differences in heterokaryon incompatibility loci as small as a single amino acid can be sufficient to trigger incompatibility [51], the diversity of these loci in *H. werneckii* and *A. melanogenum* was much higher than that. The distribution of certain types of these loci is consistent with the hybrid groups described above. On the one hand, this could simply reflect the phylogenetic distance between the strains. On the other hand, it opens the possibility that heterokaryon incompatibility loci might be involved in the successful formation of diploid hybrids in *H. werneckii* and *A. melanogenum*.

Hybrid fungal genomes have so far been described mostly in pathogens of animals or plants. This is the first time we document the formation of stable and highly heterozygous diploids in wild populations of 2 extremotolerant clonal species. Five aspects of this phenomenon are discussed below.

1. Is clonality related to the extremotolerance of *H. werneckii* and *A. melanogenum*? It has long been speculated that avoiding energetically costly sexual reproduction may be advantageous in extreme environments, allowing the fixation of beneficial alleles and genomic configurations in small populations that have managed to adapt to extreme conditions at the ecological edge of the species [15, 16]. One of the mechanisms that can promote adaptation at the margin of species distribution is hybridization [52].

Of course, sexual reproduction has not only shortcomings but also considerable advantages in adapting to stress [6, 13]. The same is true for parasexuality, which not only alters ploidy but
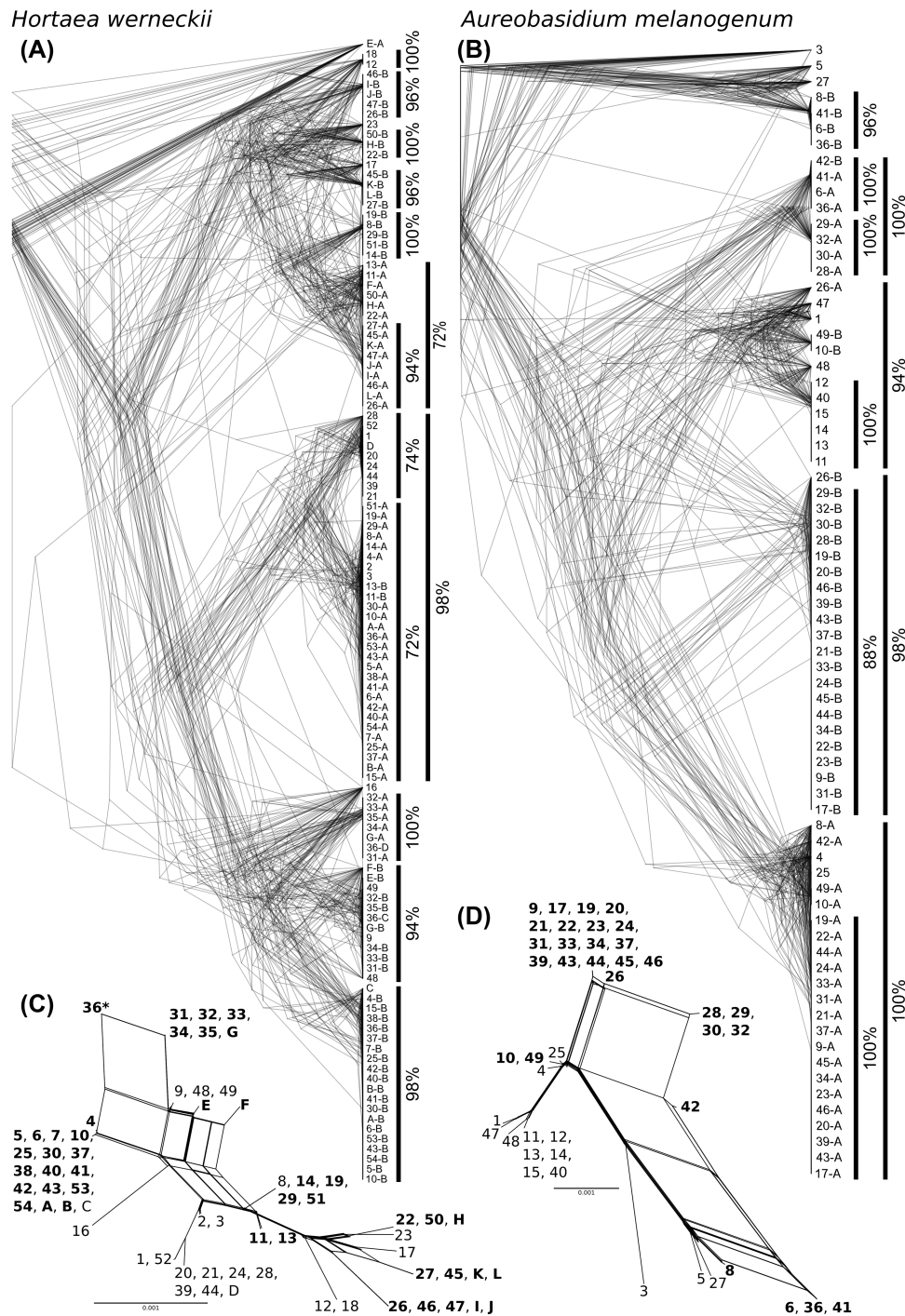
Figure 3: Phylogenies of 50 longest alignable genomic regions of *Hortaea werneckii* (A, C) and *Aureobasidium melanogenum* (B, D). The alignable regions were extracted from the genomes and aligned with SibeliaZ, optimized with Gblocks, manually inspected, and used for phylogeny reconstruction with IQ-TREE and standard model selection. (A, B) Overlay of 50 phylogenies for each species. Numbers on leaf nodes represent genomes, and different sequences from the same genomes (for genomes with ploidy >1) are distinguished with letters added to the genome numbers. Vertical lines mark major clusters and the proportion of trees that supported them. (C, D) Consensus supernetworks calculated from 50 phylogenies for each species in SplitsTree. Names of diploid (and tetraploid) strains are written in bold, and tetraploid strain is additionally marked with an asterisk.

also increases diversity through cycles of regular and double ploidy. A high frequency of diploids in *Aspergillus fumigatus* has been reported in cystic fibrosis, presumably due to local stress (e.g., nitrogen deficiency or the presence of certain drugs) that promotes parasexual recombination [19]. In *Candida albicans*, which can undergo regular sexual recombination, stress additionally

promotes the parasexual cycle, which generates a high degree of diversity, including aneuploidy [53].

There are also several examples of at least occasionally recombining extremotolerant and extremophilic species. For example, the polyextremotolerant yeast *A. pullulans*, a close relative of *A. melanogenum*, has one of the highest rates of recombi-
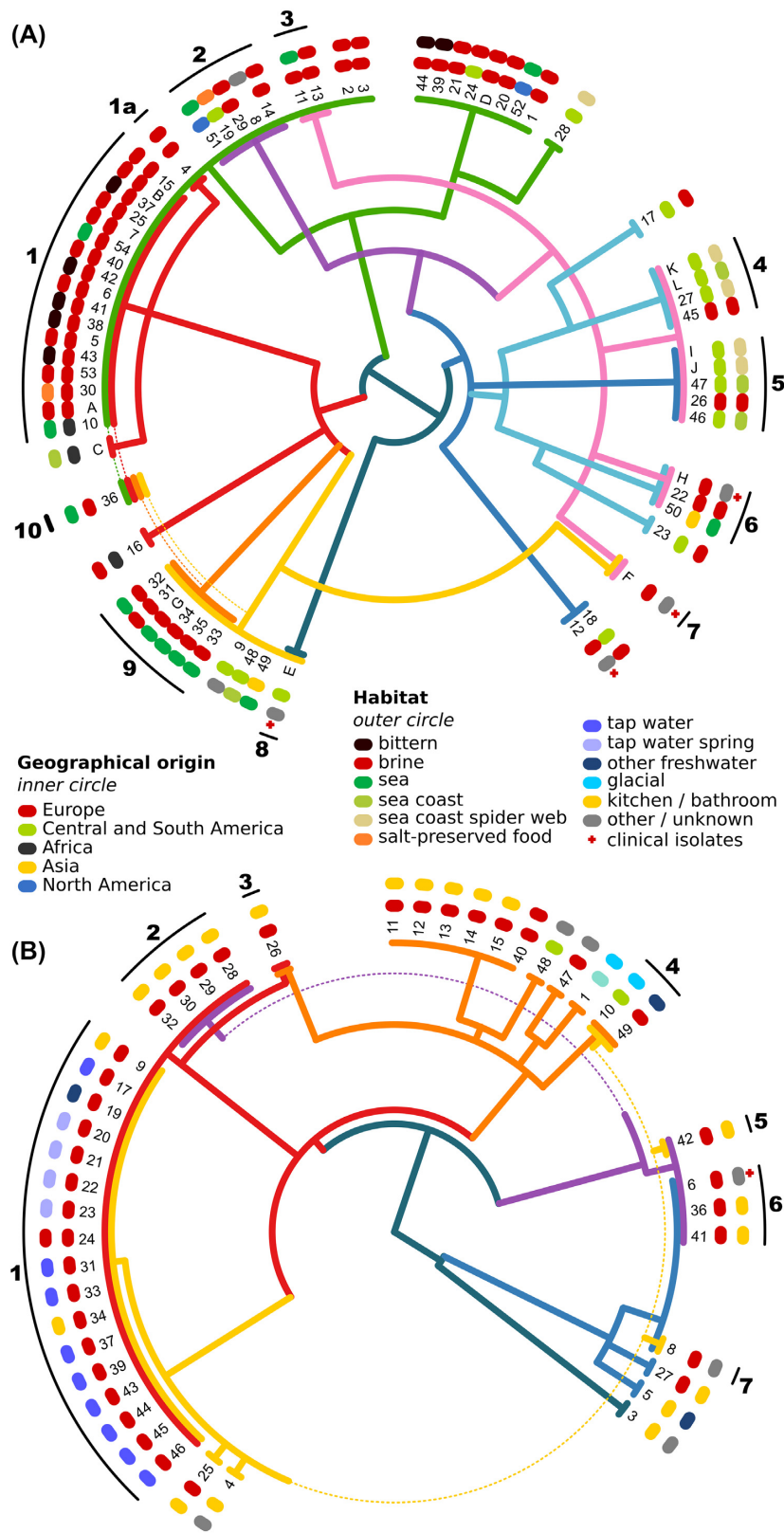
Figure 4: Hypothesis of the genome evolution and hybridization in *Hortaea werneckii* (A) and *Aureobasidium melanogenum* (B). The hypothesis is based on the majority consensus phylogeny of 50 longest alignable regions per species. Each colored line in the central tree represents a haploid genome. The distances between the nodes of the tree correspond to the distances in an ultrametric majority consensus phylogeny. Haploid genomes are represented by a single colored line in the outermost edge of the tree, diploid genomes are represented by a double colored line, and the only tetraploid genome is represented by 4 colored lines. Around the tree, colored symbols mark the continent (inner circle) and habitat (outer circle) from which the strains have been originally isolated. Black lines and numbers in the outermost circle mark the genome/strain groups presumably originating from the same hybridization event.

(A) *Hortaea werneckii*

(B) *Aureobasidium melanogenum*
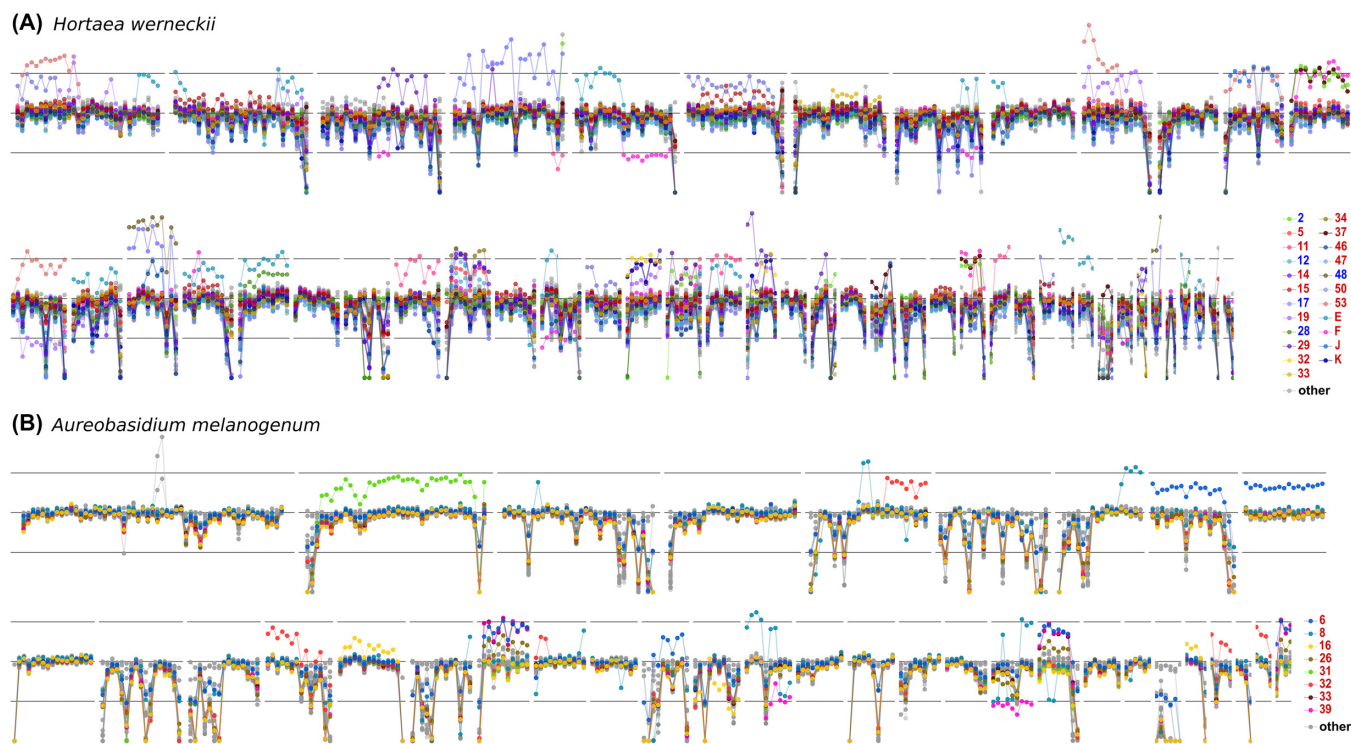
Figure 5: Aneuploid regions in *Hortaea werneckii* (A) and *Aureobasidium melanogenum* (B) genomes. Per-nucleotide sequencing depth of regions corresponding to the 50 and 35 longest contigs of *H. werneckii* and *A. melanogenum* was converted into proportion of the median sequencing depth of each individual genome. Circles represent an average of this depth in 30-kbp windows. The central horizontal line marks the median sequencing depth of the genome. Upper and lower horizontal lines mark the expected depth for haploid and triploid regions in an otherwise diploid genome. Genomes with at least 1 putatively aneuploid region are plotted in color. Other genomes are plotted in light gray. Colors of strain names in the legend mark haploid (blue) and diploid (red) genomes.

nation demonstrated in fungi by population genomics [38]. Two salt-adapted basidiomycetes, the halotolerant *Wallemia mellicola* and the halophilic *Wallemia ichthyophaga*, also appear to recombine, albeit much less frequently than *A. pullulans*, even though *W. ichthyophaga* is exceptionally rare and limited to highly fragmented environments [10, 54]. If clonality or hybridization is indeed beneficial for adaptation to extreme conditions—and more data are needed to test this hypothesis—recombination appears to be compatible with extremotolerant lifestyle as well.

2. Do clonality and hybridization allow for greater specialization? While recombination generates potentially useful diversity and thus provides substrate for natural selection, it can also break successful genomic configurations—a shortcoming known as recombination load [14–16]. In well-adapted subpopulations, clonality prevents beneficial adaptations from being diluted by gene flow from other environments, which promote different adaptations. Clonal lineages may thus be more successful in the short term but may collapse due to reduced adaptability or Muller's ratchet and are replaced by the next successful clone, which can be generated by sexual or parasexual recombination—or hybridization [55].

Some species are able to thrive in a wide range of different environments without adapting to any of them at the genomic level—the ubiquitous and polyextremotolerant *A. pullulans* is one such example [38]. But while *A. pullulans* is an exceptionally generalistic species, the 2 species analyzed here are less so: *A. melanogenum* is mostly restricted to aquatic and indoor environments, while *H. werneckii* is mostly found in marine and hypersaline environments and has a much higher upper salinity growth limit than *A. pullulans*. The preference of some *H. werneckii* and *A. melanogenum*

strain groups for specific habitats (Fig. 4) possibly indicates an ongoing clonality-driven specialization of these groups. This would be in line with the suggestion of Romeo et al. [36]. However, based on the data set studied here, the observed habitat preferences might be an artifact of skewed geographic distribution due to limited dispersal and unequal habitat sampling in different locations.

Interestingly, the clinical isolates of *H. werneckii* belong to different strain groups within the species (Fig. 4). Although the number of clinical isolates analyzed here was too small to draw reliable conclusions, this could mean that no lineage within the species is better able to cause infections in humans than others. Aneuploidy was observed in 4 of 5 analyzed clinical strains of both species. As discussed above, aneuploidy can be a signature of adaptation to novel environments [21–23]. Both *H. werneckii* and *A. melanogenum* are opportunistic pathogens that rarely cause infections. The conditions they encounter in the human body are almost certainly suboptimal for their survival [56, 57], resulting in high selection pressure and possibly in aneuploidies. However, due to their rarity, such infections most likely do not contribute meaningfully to the evolution and specialization of either of the 2 species [56].

3. If the formation of diploids in *H. werneckii* and *A. melanogenum* is irreversible, what drives the coexistence of diploid and haploid strains? A study of 12 *H. werneckii* genomes reported 7 successful intraspecific hybridization events, and expansion of the data set to 66 genomes uncovered only 2 or 3 additional hybridizations in the history of the species. While isolation of strains from novel environments or geographic locations might lead to the discovery of new hybrid lineages, their number is likely to remain limited. This might indicate that intraspecific hybridization events are relatively uncommon or that only a small number of them result in

offspring with sufficient fitness to persist in the environment. The coexistence of haploid and diploid strains may be supported by their divergent performance in different conditions. A preliminary comparison of halotolerance between diploid and haploid strains shows slightly higher halotolerance of diploid *A. melanogenum* but no such difference in *H. werneckii* (our unpublished data), but this comparison was limited to only 1 parameter tested in a laboratory setting. The possibility of different adaptation value of haploids versus diploids should be more carefully addressed in the future, for example, by competition experiments.

Whatever the mechanism of hybridization, it appears to operate almost exclusively between haploid strains, and compared to many other fungal species [21], the range of observable ploidies in *H. werneckii* and *A. melanogenum* is modest. Although a randomly selected environmental strain is about twice as likely to be diploid as haploid, a single tetraploid strain of *H. werneckii* is the only evidence that these diploids can hybridize further. No triploid strains of either species have been found. Either the fusion of diploid cells is prevented by some as yet unknown mechanism, or the resulting strains do not persist in the environment long enough to be detected.

4. How should the hybrids of clonal species be treated in taxonomy? In such situation, even the definition of otherwise well-established terminology is not trivial. For example, Naranjo-Ortiz and Gabaldón [22] defined hybrids as lineages emerging from ancestors, which differ from each other more than the most distant strains of well-recognized species. According to Boekhout et al. [47], lineages of interspecific hybrids can be recognized as separate species, while intraspecific hybrids better fit in the concept of varieties. This returns us to the problem of species delineation in clonal taxa, which may involve arbitrary decisions. On the one hand, diversity in *H. werneckii* is high, and distances between genomes of some strains are substantially greater than what is typical for fungal species [58]. On the other hand, Fig. 4 clearly illustrates why a more fragmented taxonomy of *Hortaea* would result in unpractical taxonomic inflation. If hybrids and the remaining monophyletic groups of haploids were treated as species, *H. werneckii* could easily be split into 15 or more new species. In several of these new species, 2 different sequences of standard taxonomic markers carried by a single diploid strain [33] would in many cases belong to different species—a decidedly untenable situation. Similarly, if all haploid strains were treated as one species and all diploids as another, such species would be polyphyletic. Since clonality precludes the application of the biological species concept to *H. werneckii*, we suggest the dense reticulation of its phylogeny can be pragmatically interpreted as an analogue of interbreeding. Thus, a single *H. werneckii* species is maintained despite its high diversity, as suggested also by a recent in-depth phylogenetic study of the taxon [33]. Although a comparably detailed taxonomic revision of *A. melanogenum* is still pending, it is expected to lead to a similar outcome.

5. How common is hybridization in clonal fungi? Clonality itself appears to be much rarer in fungi than once believed, but population genomic studies of *Neurospora* spp. showed that even closely related species can differ substantially in their reproductive strategies [3]. This is also the case here: while *A. melanogenum* is clonal, the closely related species *A. pullulans* has exceptionally high recombination rates [38]. Genome sequencing of another species of the genus, *Aureobasidium subglaciale*, revealed a small number of apparently clonal diploid strains that may belong to a new species [8], the reproductive strategy of which should be investigated if more such strains can be isolated and sequenced.

Of 5 fungal species from extreme environments that we have previously studied using population genomics, 2 were strictly clonal, and both contained stable diploid intraspecific hybrids. This situation is at least similar to the one described in the *Candida parapsilosis/C. orthopsilosis/C. metapsilosis* group [48]. Other examples may include the plant pathogen *Verticillium longisporum* [59] and some clinically relevant hybrids of *Cryptococcus* spp. [60] and *Aspergillus* spp. [61]. Such reproductive strategy may thus be more common than currently known, especially since it can be easily overlooked without performing careful population genomic studies. Polyploid strains often produce highly fragmented but otherwise inconspicuous assemblies, and even after genome sequencing, hybrids, polyploids, and aneuploids can easily go undetected [22]. Any study that discovers genomes of different ploidy in clonal fungal species should investigate hybridization as a possible explanation for the data.

Conclusions

Genome sequencing of 2 black yeasts from extreme environments, *H. werneckii* and *A. melanogenum*, revealed that both species are strictly clonal. Their populations consist of both haploid and diploid strains, and diploid strains were formed by a handful of intraspecific hybridization events between haploids. These hybridizations were not followed by meiosis as part of sexual reproduction, or by haploidization through random chromosome loss, as is typical of parasexuality. Hybrid lineages avoid the loss of heterozygosity even over time frames that are long enough to allow them to disperse over large geographic distances. Such "stable parasexuality," a process of forming stable and highly heterozygous diploids in a clonal species without evidence of subsequent meiosis or haploidization, is an unusual reproductive strategy, which merits further study. This is the first time it has been documented in wild populations of extremotolerant fungi. The increasing use of population genomics in fungi will show whether this reproductive strategy is more widespread than is currently known, and careful comparative studies should investigate its potential role in adaptation to extreme (and other) environments.

Materials and Methods
Cultivation and DNA isolation

Strains of the extremely halotolerant *H. werneckii* (Table 1) were obtained from the Ex Culture Collection of the Department of Biology, Biotechnical Faculty, University of Ljubljana (IC Mycosmo, MRIC UL, Slovenia). The cultivation and DNA isolation were performed as described previously [1], using the standard chemically defined Yeast Nitrogen Base medium (Qbiogene, Carlsbad, California, USA). Biomass harvested from liquid cultures was frozen in liquid nitrogen and kept at −80°C until DNA isolation, performed as described previously [1], using the UltraClean Microbial DNA isolation kit (MO BIO Laboratories, Carlsbad, California, USA), preceded by homogenization with a pestle and mortar in liquid nitrogen and 1 minute in Retsch Mixer Mill 301 (ThermoFisher Scientific, Waltham, Massachusetts, USA) at 20 Hz. After the RNAse A treatment (ThermoFisher Scientific), the isolated DNA was evaluated using agarose electrophoresis and by fluorometry (Qubit; ThermoFisher Scientific).

Genome sequencing

The genome sequencing was performed using the platform BGISEQ-500, with 2 × 150-bp libraries, prepared as described pre-

viously [62], with multiplexed sequencing barcodes. The resulting output was demultiplexed, the quality was checked with FastQC, and the reads were trimmed for adaptors and quality (removal of bases with Q <20) using the "bbduk" script (https://jgi.doe.gov/data-and-tools/bbtools/).

The raw sequencing reads have been deposited in China National GeneBank Sequence Archive of China National GeneBank DataBase with accession number CNP0001993. Sequencing reads, together with assembly and annotation data, have been deposited in GenBank under BioProject PRJNA428320. Genome sequences of previously sequenced *H. werneckii* strains [1] are deposited in GenBank under the same BioProject (PRJNA428320). Genome sequences of previously sequenced *A. melanogenum* strains [8] are deposited in GenBank under the BioProject PRJNA721240 and listed in Table 2. Genome 7 from the study by Černoša et al. [8] was excluded from this study due to the large phylogenetic distance from other *A. melanogenum* strains, while strains 2, 18, and 38 were excluded from phylogenetic analyses based on alignments produced by SibeliaZ (described below) due to their unclear ploidy.

Variant calling

Sequencing reads of *H. werneckii* genomes were mapped to the reference genome of *H. werneckii* EXF-2000 (GenBank MUNK00000000.1 [35]), which was first haploidized with HaploMerger2 [63]. Mapping was performed by "bwa mem," using the default parameters. The reads were sorted with Samtools 1.6 [64], deduplicating with Picard 2.10.2, and then used for variant calling with the Genome Analysis Toolkit 4.1 [65], following the "Genome Analysis Toolkit (GATK) Best Practices" workflow in diploid mode but using "hard filtering" with the expression "QD < 2.0 || FS > 20.0 || SOR > 3.0 || MQ < 50.0." Strain 36 was excluded from the analysis due to its tetraploid genome. Variants of *A. melanogenum* genomes were determined by Černoša et al. [8].

Variant-based analysis

Variant-based analyses for both *H. werneckii* and *A. melanogenum* were performed in R [66], except the calculation of the linkage disequilibrium squared correlation coefficient (r^2; described below). Genomes were clustered based on the SNP data using the principal component analysis with the "glPca" function of the "adgenet" package in R [67]. The phylogenetic networks estimated from SNP data were reconstructed with the Neighbor-Net algorithm of the package "phangorn" [68] based on a dissimilarity distance matrix calculated with the package "poppr" [69].

Linkage disequilibrium squared correlation coefficient (r^2) was calculated for all pairs of biallelic SNP loci within 10,000 nucleotides of each other with "vcftools" [70]. Then r^2 was plotted as a function of distance between pairs of loci using "ggplot2" [71]. A generalized additive model ("gam") curve was fitted to the data.

Assembly and annotation

Reference-guided genome assembly was performed for all here sequenced *H. werneckii* genomes with IDBA-Hybrid 1.1.3 [72] using the same reference as for variant calling. The maximum *k* value selected was 120, the minimum support in each iteration was 2, the similarity threshold for alignment was 0.95, seed kmer was 20, maximum allowed gap in the reference was 100, and the minimum size of contigs included in the final assembly was 500. Genomes were annotated with Augustus 3.4 [73]. Augustus parameters were optimized with training using the scripts provided with the program with (i) the RNAseq data from Sinha et al. [35] deposited at the GenBank Sequence Read Archive with the ac-

cession number SRP094740 and (ii) all predicted proteins of *H. werneckii* EXF-2000 (GenBank MUNK00000000.1). These hints were also used for the final annotation.

Predicted proteomes were benchmarked with the BUSCO 4.1.1 [74] using the default parameter values and the data set of BUSCOs for Dothideomycetes [75].

The ploidy of the genomes was determined based on the following criteria for both species: haploids had a genome size smaller than 31 Mbp, number of predicted genes smaller than 13,000, and the average copy number of core BUSCOs (those present in all strains of the species) lower than 1.1. Diploid strains had a genome size larger than 46 Mbp, number of predicted genes greater than 18,000, and the average copy number of core BUSCOs greater than 1.5. The ploidy of genomes with any of the criteria between the above thresholds was labeled as "unclear" (Table 2).

The files for submission to GenBank were prepared with the Genome Annotation Generator (GAG) 2.0.1 [76]. Gene models with short coding regions (<150 bp) and/or introns (<10 bp) were removed before the submission.

Phylogenetic analyses

SibeliaZ 1.2.2 [77] was used to align parts of the genomes of both *H. werneckii* and *A. melanogenum* into multiple sequence alignments. The parameters used were $k = 21$, $a = 150$, $b = 15,000$. Alignments were then filtered to keep only those in which the number of sequences from each genome exactly matched the ploidy of the genome. Alignments were optimized with Gblocks 0.91 [78], using the options "-b3 = 10 -b4 = 3 -b5 = n," and then inspected manually to trim the ends to the shortest sequence in the alignment and remove any alignments with more than 15% gaps over the whole alignment length in any of the sequences of *H. werneckii* or 25% in case of *A. melanogenum*. Fifty longest alignments (lengths of 1,364 to 5,089 bp for *H. werneckii* and 3,400 to 13,257 bp for *A. melanogenum*) were selected for each species, and each alignment was used for the estimation of the phylogenetic tree with IQ-TREE 2.0.3 using standard model selection and 1,000 replicates for the SH approximate likelihood ratio test [79]. The resulting collection of 50 phylogenetic trees for each species was visualized as an overlay using "densiTree()" from the "phangorn" package in R [68] and as a consensus supernetwork using SplitsTree 4.16.2 [80]. These visualizations and a majority rule consensus tree calculated with the "consensus.edges" from the package "phytools" in R [81] were used to draw a schematic representation of phylogenetic histories of genomes in the open-source vector graphics software Inkscape 1.1 (http://inkscape.org). Enrichment of phylogenetic clusters of strains for certain geographic origin or habitat was analyzed in R using the Fisher exact test with simulated *P* value [66].

Genes with HET (PF06085) and Het-C (PF07217) domains were identified in predicted proteomes of all strains investigated in this study (Tables 1 and 2) with "hmmsearch" 3.3.1 and HMM profiles with default parameters from the Pfam-A.hmm database version 34.0 [82]. The identified proteins were aligned with Mafft 7.475 [83], and the alignments were used for reconstruction of phylogenies with IQ-TREE 2.0.3 using standard model selection and 1,000 replicates for the SH approximate likelihood ratio test [79].

Putative mating-type loci were identified in the genomes with stand-alone BLAST 2.9.0+ [84], aligned with Mafft 7.475 [83], and annotated based on previously published annotations of mating-type loci in *H. werneckii* [1] and *A. melanogenum* [39]. Phylogeny of putative MAT1-1 and MAT1-2 homologues was estimated with IQ-TREE 2.0.3 using standard model selection and 1,000 replicates for the SH approximate likelihood ratio test [79] after first excluding

all putative homologues truncated to less than 80% of expected length due to suboptimal genome assembly.

Detection of aneuploidies and loss of heterozygosity

Per-nucleotide sequencing depth of reads mapped to the reference genome as described above was calculated with Samtools 1.6 [64]. For each sequenced genome, the median values of per-nucleotide depths in 30-kbp windows were plotted as proportion of the median depth of the whole genome. These values were calculated in R and visualized with "ggplot2" [66, 71] for the 50 longest reference genome contigs for *H. werneckii* and for the 35 longest reference genome contigs for *A. melanogenum*.

Evidence for loss of heterozygosity in diploid genomes was searched for by counting the number of heterozygous SNPs in 25-kbp windows along the longest reference genome contigs (50 for *H. werneckii*, 35 for *A. melanogenum*) and plotted as a proportion of the median heterozygosity of each genome with "ggplot2" [66, 71]. Depth of sequencing was plotted as described above, but in 25-kbp windows, to distinguish between copy-neutral loss of heterozygosity and loss of heterozygosity caused by aneuploidy.

Additional Files

Supplementary Table S1. Statistics of *H. werneckii* genomes sequenced in this study. Violin plots show the distribution of values in corresponding columns below the plots.

Supplementary Table S2. Putative HET and HET-C proteins in different strains of *H. werneckii* and *A. melanogenum*.

Supplementary Figure S1. Aneuploid regions in *H. werneckii* genomes. Per-nucleotide sequencing depth of regions corresponding to the 50 longest contigs of *H. werneckii* and *A. melanogenum* was converted into proportion of the median sequencing depth of each individual genome and plotted in 50-kbp rolling median windows (black line). Upper and lower horizontal lines mark the expected depth for haploid and triploid regions in an otherwise diploid genome. Putatively aneuploid region of increased ploidy is marked with red rectangles.

Supplementary Figure S2. Aneuploid regions in *A. melanogenum* genomes. Per-nucleotide sequencing depth of regions corresponding to the 35 longest contigs of *A. melanogenum* was converted into proportion of the median sequencing depth of each individual genome and plotted in 50-kbp rolling median windows (black line). Upper and lower horizontal lines mark the expected depth for haploid and triploid regions in an otherwise diploid genome. Putatively aneuploid region of increased ploidy is marked with red rectangles.

Supplementary Figure S3. Heterozygosity in diploid *H. werneckii* and *A. melanogenum* genomes. Levels of heterozygosity (black lines) and sequencing depth (purple lines) were expressed as proportions of median heterozygosity and sequencing depth of each individual genome. The values were plotted in 25-kbp windows across regions corresponding to the 50 and 35 longest contigs of *H. werneckii* and *A. melanogenum*, respectively. Diploid regions (i.e., with sequencing depth similar to the rest of the diploid genome) with extensive loss of heterozygosity are marked with red rectangles.

Supplementary Figure S4. Putative mating-type loci of *H. werneckii*. Visualization of phylogenies, presence/absence, and sequencing depth of MAT1-1 and MAT1-2 homologues, as well as an annotated alignment of the whole putative mating locus and its flanking regions.

Supplementary Figure S5. Putative mating-type loci of *A. melanogenum*. Visualization of phylogenies, presence/absence, and sequencing depth of MAT1-1 and MAT1-2 homologues, as well as an annotated alignment of the whole putative mating locus and its flanking regions.

Abbreviations

BUSCO: Benchmarking Universal Single-Copy Orthologs; kbp: kilobase pair; PCA: principal component analysis; SNP: single-nucleotide polymorphism.

Funding

Data Availability

All data used in the study are available via GenBank (BioProject: PRJNA721240 and BioProject: PRJNA428320). Supporting data, including alignments, SNPs, annotations, and phylogenetic tree files, are available via the *GigaScience* database GigaDB [85].

Competing interests

The authors declare that they have no competing interests.

Authors' Contributions

Conceptualization of the study: N.G.C., Z.S., and C.G.; experimental work: X.S., A.Č., and C.F.; bioinformatic analyses: C.G.; data curation: C.G.; resources: Z.S. and N.G.C.; preparation of the manuscript and visualizations: C.G.; review and editing of the manuscript: C.G., N.G.C., A.Č., X.S., C.F., and Z.S.; supervision: Z.S. and N.G.C.; funding acquisition: N.G.C. and Z.S.

Acknowledgments

References

1. Gostinčar, C, Stajich, JE, Zupančič, J, *et al.* Genomic evidence for intraspecific hybridization in a clonal and extremely halotolerant yeast. *BMC Genomics* 2018;**19** (364):1–12.

2. Brown, A. Mating in mushrooms: increasing the chances but prolonging the affair. *Trends Genet* 2001;**17** (7):393–400.

3. Gladieux, P, De Bellis, F, Hann-Soden, C, *et al.* Neurospora from natural populations: population genomics insights into the life history of a model microbial eukaryote. *Methods Mol Biol* 2020;**2090**:313–36.

4. Hawksworth, DL, Kirk, PM, Sutton, BC, *et al.* *Ainsworth & Bisby's Dictionary of the Fungi.* 8th ed. Oxford, UK: Oxford University Press, 1996

5. Tibayrenc, M, Ayala, FJ. Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. *Proc Natl Acad Sci* 2012;**109**(48):3305–3313.

6. Taylor, JW, Hann-Soden, C, Branco, S, *et al.* Clonal reproduction in fungi. *Proc Natl Acad Sci* 2015;**112** (29):8901–8.

7. Carreté, L, Ksiezopolska, E, Pegueroles, C, *et al.* Patterns of genomic variation in the opportunistic pathogen *Candida glabrata* suggest the existence of mating and a secondary association with humans. *Curr Biol* 2018;**28** (1):15–27.e7.

8. Zajc, Z, Černoša, A, Sun, X, *et al.* From glaciers to refrigerators: the population genomics and biocontrol potential of the black yeast Aureobasidium subglaciale. In TR O'Meara, editor. *Microbiol Spectr*. 2022

9. Kasuga, T, White, TJ, Koenig, G, *et al.* Phylogeography of the fungal pathogen *Histoplasma capsulatum*. *Mol Ecol* 2003;**12** (12): 3383–401.

10. Gostinčar, C, Sun, X, Zajc, J, *et al.* Population genomics of an obligately halophilic basidiomycete *Wallemia ichthyophaga*. *Front Microbiol* 2019:1–12.

11. Tsai, IJ, Bensasson, D, Burt, A, *et al.* Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. *Proc Natl Acad Sci* 2008;**105** (12):4957–62.

12. Xu, J. Cost of interacting with sexual partners in a facultative sexual microbe. *Genetics* 2005;**171** (4):1597–604.

13. Sun, S, Heitman, J. Is sex necessary? *BMC Biol* 2011;**9** (1): 1–4.

14. Otto, SP. The evolutionary enigma of sex. *Am Nat* 2009;**174** (Suppl 1):S1–S14.

15. Gostinčar, C, Grube, M, De Hoog, S, *et al.* Extremotolerance in fungi: evolution on the edge. *FEMS Microbiol Ecol* 2010;**71** (1):2–11.

16. Gostinčar, C, Gunde-Cimerman, N, Grube, M. Polyextremotolerance as the fungal answer to changing environments. In: C Bakermans, editor *Microbial Evolution under Extreme Conditions*. Berlin, Germany: De Gruyter, 2015

17. Pontecorvo, G. The parasexual cycle in fungi. *Annu Rev Microbiol* 1956;**10** (1):393–400.

18. Clutterbuck, AJ. Parasexual recombination in fungi. *J Genet* 1996;**75** (3):281–6.

19. Engel, T, Verweij, PE, van den Heuvel, J. *et al.* Parasexual recombination enables *Aspergillus fumigatus* to persist in cystic fibrosis. *ERJ Open Research* 2020;**6** (4):1–10.

20. Todd, RT, Forche, A, Selmecki, A. Ploidy variation in fungi: Polyploidy, aneuploidy, and genome evolution. *Microbiol Spectrum* 2017;**5** (4):1–31.

21. Gerstein, AC, Sharp, NP. The population genetics of ploidy change in unicellular fungi. *FEMS Microbiol Rev* 2021;**45** (5):1–20.

22. Naranjo-Ortiz, MA, Gabaldón, T. Fungal evolution: cellular, genomic and metabolic complexity. *Biol Rev* 2020;**95** (5):1198–232.

23. Hill, R, Leitch, IJ, Gaya, E. Targeting Ascomycota genomes: what and how big? *Fungal Biol Rev* 2021;**36**:52–9.

24. Chen, G, Rubinstein, B, Li, R. Whole chromosome aneuploidy: big mutations drive adaptation by phenotypic leap. *Bioessays* 2012;**34** (10):893–900.

25. Gerstein, AC, Lim, H, Berman, J, *et al.* Ploidy tug-of-war: evolutionary and genetic environments influence the rate of ploidy drive in a human fungal pathogen. *Evolution* 2017;**71** (4):1025–38.

26. Peter, J, De Chiara, M, Friedrich, A, *et al.* Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 2018;**556** (7701):339–44.

27. Depotter, JR, Seidl, MF, Wood, TA, *et al.* Intersp ecific hybridization impacts host range and pathogenicity of filamentous microbes. *Curr Opin Microbiol* 2016;**32**:7–13.

28. Samarasinghe, H, You, M, Jenkinson, TS, *et al.* Hybridi zation facilitates adaptive evolution in two major fungal pathogens. *Genes* 2020;**11** (1):101.

29. Morrow, CA, Fraser, JA. Ploidy variation as an adaptive mechanism in human pathogenic fungi. *Semin Cell Dev Biol* 2013;**24** (4):339–46.

30. Forche, A. Large-scale chromosomal changes and associated fitness consequences in pathogenic fungi. *Curr Fungal Infect Rep* 2014;**8** (2):163–70.

31. Gunde-Cimerman, N, Zalar, P, Hoog, S, *et al.* Hypersaline waters in salterns—natural ecological niches for halophilic black yeasts. *FEMS Microbiol Ecol* 2000;**32** (3):235–40.

32. Plemenitaš, A, Lenassi, M, Konte, T, *et al.* Adaptation to high salt concentrations in halotolerant/halophilic fungi: a molecular perspective. *Front Microbiol* 2014;**5**:1–12.

33. Zalar, P, Zupančič, J, Gostinčar, C, *et al.* The extremely halotolerant black yeast *Hortaea werneckii*—a model for intraspecific hybridization in clonal fungi. *IMA Fungus* 2019;**10**(1):1–27.

34. Lenassi, M, Gostinčar, C, Jackman, S, *et al.* Whole genome duplication and enrichment of metal cation transporters revealed by *de novo* genome sequencing of extremely halotolerant black yeast *Hortaea werneckii*. *PLoS One* 2013;**8** (8): e71328.

35. Sinha, S, Flibotte, S, Neira, M, *et al.* Insight into the recent genome duplication of the halophilic yeast *Hortaea werneckii*: combining an improved genome with gene expression and chromatin structure. *G3 (Bethesda)* 2017;**7** (7):2015–22.

36. Romeo, O, Marchetta, A, Giosa, D, *et al.* Whole genom sequencing and comparative genome analysis of the halotolerant deep sea black yeast *Hortaea werneckii*. *Life* 2020;**10** (10):229.

37. Azua-Bustos, A, Gonzalez-Silva, C, Salas, L, *et al.* A novel subaerial *Dunaliella* species growing on cave spiderwebs in the Atacama Desert. *Extremophiles* 2010;**14** (5):443–52.

38. Gostinčar, C, Turk, M, Zajc, J, *et al.* Fifty A ureobasidium pullulans genomes reveal a recombining polyextremotolerant generalist. *Environ Microbiol* 2019;**21** (10):3638–52.

39. Gostinčar, C, Ohm, RA, Kogej, T, *et al.* Genome sequencing of four *Aureobasidium pullulans* varieties: biotechnological potential, stress tolerance, and description of new species. *BMC Genomics* 2014;**15** (1):549.

40. Zhu, YO, Sherlock, G, Petrov, DA. Whole genome analysis of 132 clinical *Saccharomyces cerevisiae* strains reveals extensive ploidy variation. *G3 (Bethesda)* 2016;**6** (8):2421–34.

41. O'Brien, CE, Oliveira-Pacheco, J, Ó Cinnéide, E, *et al.* Population genomics of the pathogenic yeast *Candida tropicalis* identifies hybrid isolates in environmental samples. *PLoS Pathog* 2021;**17** (3):e1009138.

42. Rhodes, J, Desjardins, CA, Sykes, SM, *et al.* Tracing genetic exchange and biogeography of *Cryptococcus neoformans* var. *grubii* at the global population level. *Genetics* 2017;**207** (1):327–46.

43. Gabaldón, T. Hybridization and the origin of new yeast lineages. *FEMS Yeast Res* 2020;**20** (5):1–8.

44. Pryszcz, LP, Németh, T, Gácser, A, *et al.* Genome comparison of *Candida orthopsilosis* clinical strains reveals the existence of hybrids between two distinct subspecies. *Genome Biol Evol* 2014;**6** (5):1069–78.

45. Pryszcz, LP, Németh, T, Saus, E, *et al.* The genomic aftermath of hybridization in the opportunistic pathogen *Candida metapsilosis*. *PLos Genet* 2015;**11** (10):e1005626.

46. Gostinčar, C, Stajich, JE, Kejžar, A, *et al.* Seven years at high salinity—experimental evolution of the extremely halotolerant black yeast *Hortaea werneckii*. *J Fungi* 2021;**7** (9):723.

47. Boekhout, T, Aime, MC, Begerow, D, *et al.* The evolving species concepts used for yeasts: from phenotypes and genomes to speciation networks. *Fungal Diversity* 2021;**109** (1):27–55.

48. Mixão, V, Gabaldón, T. Hybridization and emergence of virulence in opportunistic human yeast pathogens. *Yeast* 2018;**35** (1):5–20.

49. Strom, NB, Bushley, KE. Two genomes are better than one: history, genetics, and biotechnological applications of fungal heterokaryons. *Fungal Biol Biotechnol* 2016;**3** (1):1–14.

50. Mitchison-Field, LMY, Vargas-Muñiz, JM, Stormo, BM, *et al.* Unconventional cell division cycles from marine-derived yeasts. *Curr Biol* 2019;**29** (20):3439–3456.e5.

51. Saupe, SJ. Molecular genetics of heterokaryon incompatibility in filamentous Ascomycetes. *Microbiol Mol Biol Rev* 2000;**64** (3):489–502.

52. Gladieux, P, Ropars, J, Badouin, H, *et al.* Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Mol Ecol* 2014;**23** (4):753–73.

53. Berman, J, Hadany, L. Does stress induce (para)sex? Implications for *Candida albicans* evolution. *Trends Genet* 2012;**28** (5): 197–203.

54. Sun, X, Gostinčar, C, Fang, C, *et al.* Genomic evidence of recombination in the basidiomycete *Wallemia mellicola*. *Genes* 2019;**10** (6) 427.

55. Drenth, A, McTaggart, AR, Wingfield, BD. Fungal clones win the battle, but recombination wins the war. *IMA Fungus* 2019;**10**(18):1–6.

56. Gostinčar, C, Zajc, J, Lenassi, M, *et al.* Fungi between extremotolerance and opportunistic pathogenicity on humans. *Fungal Diversity* 2018;**93** (1):195–213.

57. Zajc, J, Gostinčar, C, Černoša, A, *et al.* Stress-tolerant yeasts: Opportunistic pathogenicity versus biocontrol potential. *Genes* 2019;**10** (1):42.

58. Gostinčar, C. Towards genomic criteria for delineating fungal species. *J Fungi* 2020;**6** (4):246.

59. Fogelqvist, J, Tzelepis, G, Bejai, S, *et al.* Analysis of the hybrid genomes of two field isolates of the soil-borne fungal species *Verticillium longisporum*. *BMC Genomics* 2018;**19**(14):1–12.

60. Michelotti, LA, Sun, S, Heitman, J, *et al.* Clonal evolution in serially passaged *Cryptococcus neoformans × deneoformans* hybrids reveals a heterogenous landscape of genomic change. *Genetics* 2022;**220** (1):1–13.

61. Steenwyk, JL, Lind, AL, Ries, LNA, *et al.* Pathogenic allodiploid hybrids of *Aspergillus* fungi. *Curr Biol* 2020;**30** (13):2495–2507.e7.

62. Fang, C, Zhong, H, Lin, Y, *et al.* Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. *Gigascience* 2018;**7** (3):1–8.

63. Huang, S, Kang, M, Xu, A. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. Berger B, editor. *Bioinformatics* 2017;**33** (16):2577–9.

64. Li, H, Handsaker, B, Wysoker, A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25** (16):2078–9.

65. Alkan, C, Coe, BP, Eichler, EE. GATK toolkit. *Nat Rev Genet* 2011;**12** (5):363–76.

66. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2021

67. Jombart, T, Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 2011;**27** (21):3070–1.

68. Schliep, K, Potts, AJ, Morrison, DA, *et al.* Intertwining phylogenetic trees and networks. *Methods Ecol Evol* 2017;**8** (10):1212–20.

69. Kamvar, ZN, Brooks, JC, Grünwald, NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front Genet* 2015;**6**:1–10.

70. Danecek, P, Auton, A, Abecasis, G, *et al.* The variant call format and VCFtools. *Bioinformatics* 2011;**27** (15):2156–8.

71. Wickham, H. *ggplot2*. Cham, Switzerland: Springer International Publishing, 2016.

72. Peng, Y, Leung, HCM, Yiu, SM, *et al.* IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012;**28** (11):1420–8.

73. Stanke, M, Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 2005;**33** (Web Server):W465–7

74. Simão, FA, Waterhouse, RM, Ioannidis, P, *et al.* BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;**31** (19):3210–2.

75. Kriventseva, EV., Tegenfeldt, F, Petty, TJ, *et al.* OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res* 2015;**43** (D1):D250–6.

76. Geib, SM, Hall, B, Derego, T, *et al.* Genome Annotation Generator: a simple tool for generating and correcting WGS annotation tables for NCBI submission. *Gigascience* 2018;**7** (4):1–5.

77. Minkin, I, Medvedev, P. Scalable multiple whole-genome alignment and locally collinear block construction with SibeliaZ. *Nat Commun* 2020;**11** (1):1–11.

78. Talavera, G, Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007;**56** (4):564–77.

79. Minh, BQ, Schmidt, HA, Chernomor, O. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;**37** (5):1530–4.

80. Huson, DH, Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 2006;**23** (2):254–67.

81. Revell, LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 2012;**3** (2):217–23.

82. Mistry, J, Chuguransky, S, Williams, L, *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* 2021;**49** (D1):D412–9.

83. Katoh, K, Standley, DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30** (4):772–80.

84. Altschul, S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25** (17):3389–402.

85. Gostinčar, C, Sun, X, Černoša, A, *et al.* Supporting data for "Clonality, inbreeding, and hybridization in two extremotolerant black yeasts." *GigaScience Database* 2022.