GigaScience

Clonality, inbreeding, and hybridization in two extremotolerant black yeasts

--Manuscript Draft--

Manuscript Number:	GIGA-D-22-00073R1
Full Title:	Clonality, inbreeding, and hybridization in two extremotolerant black yeasts
Article Type:	Research
Funding Information:	Javna Agencija za Raziskovalno Dejavnost RS (MRIC UL I0-0022) — Not applicable Javna Agencija za Raziskovalno Dejavnost RS (P4-0432) — Assoc Prof Cene Gostinčar Javna Agencija za Raziskovalno Dejavnost RS (P1-0198) — Not applicable Javna Agencija za Raziskovalno Dejavnost RS (J4-2549) — Assoc Prof Cene Gostinčar
Abstract:	The great diversity of lifestyles and survival strategies observed in fungi is reflected in the many ways in which they reproduce and recombine. Although truly asexual fungi are rare, population genomic data support the clonality of two extremotolerant black yeasts from Dothideomycetes: Hortaea werneckii and Aureobasidium melanogenum . Thus, the discovery of a number of diploid strains of these species could not be explained as the product of conventional sexual reproduction. Genome sequencing revealed that the ratio of diploid to haploid strains in both H. werneckii and A. melanogenum is approximately 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. Our results, based on the genomes of over 100 strains of two black yeasts, show that although they are asexual, they occasionally form stable and highly heterozygous diploid hybrids. The mechanism of these apparently rare hybridization events, which are not followed by meiosis or haploidisation, 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.
Corresponding Author:	Cene Gostinčar, Ph.D. University of Ljubljana Biotechnical faculty: Univerza v Ljubljani Biotehniska fakulteta Ljubljana, SLOVENIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Ljubljana Biotechnical faculty: Univerza v Ljubljani Biotehniska fakulteta
Corresponding Author's Secondary Institution:	
First Author:	Cene Gostinčar, Ph.D.
First Author Secondary Information:	
Order of Authors:	Cene Gostinčar, Ph.D. Xiaohuan Sun Anja Černoša

	Chao Fang
	Nina Gunde-Cimerman
	Zewei Song
Order of Authors Secondary Information:	
Response to Reviewers:	We would like to thank the reviewers for their detailed and constructive feedback! Both reviewers requested revisions, which have now been made. The detailed responses to the comments of the reviewers are listed below. We believe this strengthens the manuscript and we hope you can consider it for publication. Reviewer #1 Comment: separate trees were generated for "longest alignments" and I think this approach can hide potential admixture events. It is not reported anywhere what are the average lengths of these alignments, but the point is that if admixture concerns a small part of the chromosome, the alignment of the whole chromosome will not detect those admixture events. I would suggest to split all alignments into portions of equal length or of equal number of informative SNPs, to identify potential admixture events, if any exist. Answer: The range of lengths of the alignments is now reported in the methodology: “lengths of 1364 bp to 5089 bp for H. werneckii and 3400 bp to 13257 bp for A. melanogenum”. The length of these alignments was not overwhelming, therefore the masking of the admixture events was not explicitly addressed in the initial manuscript – they should be able to reflect also admixture events concerning only small parts of the chromosome. If the alignments were split further, the number of phylogenetically informative sites in them would be limited. The divergence of the genomes and likely also the difficult assembly of diploid genomes precluded the recovery of longer alignments with a 100% representation of all haploid genomes. However, as an alternative to the alignment-splitting approach, phylogenetic trees were also constructed from core BUSCOs that were found in all genomes of each species (and in the number of copies corresponding to the ploidy of the genome). Since the number of genomes was considerable and since an unexpected copy number in even a single genome excluded the gene from the analysis, the final dataset was fairly modest. Alignments longer than 200 nucleotides and with an average of at least 15 nucleotide differences between gene pairs were used for phylogeny reconstruction, producing 22 phylogenetic trees in case of H. werneckii and 44 in case of A. melanogenum. While these phylogenies are generally based on shorter alignments and are perhaps less reliable (which was the reason for reporting Sibelia and not BUSCO alignments in the manuscript), they result in almost identical topology and the proportion of trees supporting the major splits are very similar. Please see the here provided phylogeny with percentages to the right of trees marking the proportion of trees supporting the major clusters. Black percentage points are now reported in the revised Fig. 3 (in response to the comment below; calculated from phylogenies based on genomic alignments produced by Sibelia). Red numbers are the percentages of trees supporting the same clusters in trees produced from alignment of BUSCOs (provided here for review purposes only). [please refer to the attached file Answers.to.reviewers.REVIEW.ONLY.docx for the image] Comment: tree discordancies are not quantified in any way and from figure 3 it's hard to judge how much concordance there is. If the species cluster in several groups one could show how many topologies (proportion) of these major clusters are consistent with each other and how many are different Answer: The proportion of phylogenies supporting the major clusters have now been labelled in panels A and B of Fig. 3 as suggested. The figure legend has been amended accordingly. Minor comments: Comment: Table S1 and Table 1: Please explain in the legend what distributions are showing. Also it would be helpful to include a column in the tables with information

about ploidy.
Answer: The legend of Table S1 (which shows the distributions) has been amended to explain the plots in more detail. The column reporting the ploidy has been added as suggested to Table 1 and Table S1. Additionally, the ploidy column has been also added to Table 2 to keep the format of tables consistent.
Comment: line 183: I'm not sure what the authors mean by 'consistent' in this sentence. Wasn't the ploidy decided from genome assembly characteristics? In this case it's expected to be consistent.
Answer: This was an awkwardly written sentence – we did not want to say that the genomic characteristics were consistent with ploidy (which they were, because this is how the ploidy was determined, as correctly pointed out by the reviewer), but that haploid strains were very similar in their characteristics to each other, and the same was true for diploid strains. We rewrote the sentence as follows and hopefully this makes the message clearer: “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).”.
Comment: line 195: I would suggest explaining here in once sentence how SNP calling was made, especially how the reference was constructed, because it's quite important for interpreting the results.
Answer: We added a short description, as suggested: “Single-nucleotide polymorphisms (SNPs) were determined with Genome Analysis Toolkit after mapping the sequencing reads to reference genomes (haploidised genome of diploid strain EXF-2000 in case of H. werneckii, whole genome of haploid strain EXF-3378 in case of A. melanogenum).” We hope this is sufficiently detailed for the Results section, since a more detailed description is provided in the Methods.
Comment: line 202: PCA plots in Fig1: It would be useful to add ploidy information in the plot to see where are those samples located relative to haploids. Are these clusters explained by geography or habitat? Perhaps adding this information would be useful as well. It is also not mentioned anywhere how divergent are these clusters. It would be worth reporting nucleotide divergence between (haploid) genomic groups.
Answer: Diploid strains have now been marked with bold labels in all panels of Fig. 1 – PCA plots as suggested, but also in phylogenetic networks. The sizes and placement of labels have been optimized to increase legibility. The clusters can indeed be partially explained by geography/habitat – these information are visualised in Fig. 4 and discussed there as well. This decision was made after we first attempted to show the geography/habitat data in Fig. 1 (PCA plots) and discovered that the number of genomes and their considerable overlapping make an efficient visualisation all but impossible. Divergence between clusters is now also marked in Fig. 1 with dashed lines marking the groups of haploid strains and distance in millions of SNPs between groups. The legend of the figure has been amended to reflect these changes.
Comment: line 221: Please highlight haploids/diploid on the phylogeny.
Answer: In Fig. 3, ploidy has now been marked on the phylogeny – diploid strain names have been written in bold in panels C and D. In panels A and B diploid (and tetraploid) strains were already marked with letters added to strain names – this has now been explicitly explained in the legend as well.
Comment: line 209: Considering LD decay analysis it looks to me that r2 is very low even between close variants. In general, it is not clear from the figure 2 what is the maximum r2 between adjacent pairs of SNPs (start of the line) and what is the distance over which r2 falls by half. I think the authors should give some quantification of this in the results. This could give a better understanding of the LD.
Answer: The maximum value of LD (drawn as the upper red horizontal line on the LD charts) are now also reported in the Results, as suggested: “Plotting r2 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.”. Regarding the low values, the initial r2 value is not only dependent on the disequilibrium, but also on the allele frequency (i.e. alleles in less than 50% frequency that are in total disequilibrium, will have r2 not 1 but less than 1) and is therefore not unusual.
Comment: line 229: This is a really interesting way to show relationships between hybrids!
Answer: Thank you!
Comment: line 250: Fig 5, What are colours in the legend signifying?
Answer: The sentence explaining the colours was lost in revision and was now again added: “Colours of strain names in the legend mark haploid (blue) and diploid (red)

genomes.” Thank you for alerting us to this!
Comment: line 282: MAT loci: One option to make sure if MAT loci is truly absent would be to look for reads matching the sequence of MAT. This could eliminate the possibility that the quality of an assembly is a source of missing loci.
Answer: The original manuscript already acknowledged the possibility that the apparently missing loci are the consequence of the assembly problems and not of their true absence. We now performed an additional analysis, as suggested. In many cases this was not helpful because the divergence of some loci is so large, that it resulted in gaps in the sequencing coverage even in cases where the putative mating-type loci were found in the assembly. In other cases, especially in A. melanogenum, the coverage did indeed suggest, as the reviewer thought it might, that the loci are actually present in the genome, but not present in the genome assembly. The results of the sequencing coverage analysis have been added to Supplementary Figs. 4 and 5 and the corresponding figure legends amended as appropriate. The corresponding Results section of the manuscript has been changed as follows: “While mapping of sequencing reads of some strains to 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 (Supplemental 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.” Our opinion on this is that the sequencing coverage analysis comes with its own set of problems, but its inclusion does indeed provide a more complete picture of a complex situation. While a more conclusive resolution of this question can only be provided by using a sequencing technology producing longer reads, the results presented in this manuscript, while partial, do provide relevant information tightly linked to the main topic of the manuscript as well as important indications for future research, at the same time acknowledging the partiality of the results.
Comment: line 263: In the figures S4 and S5 one information that is missing is whether the same MAT type is present always on the same genomic background, assuming that these different types are in the same locus. Examples of MAT introgression are common in fungi so it would be nice to check if something like this occurs in these species.
Answer: We agree that this is an interesting question and we did some preliminary investigation into it before preparing the manuscript. However, due to the above discussed suboptimal assembly in the mating-type loci, this was only possible for some strains. Additionally, the precise borders of the mating locus in H. werneckii are not known. Based on all of this we decided not to report highly uncertain data and rather leave this question to be solved by a more targeted experiment, e.g. by using a sequencing technology producing long reads.
Comment: line 619: In the description of processing sequence alignments, please specify what do you mean by "long gaps"? Was there any threshold?
Answer: This was indeed not as precise as it should be. We replaced “long gaps” with “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”.

Reviewer #2

Major:
Comment: The methods are missing a description of how ploidy was estimated, the results of which are integral to the conclusions.
Answer: Thank you for pointing this out. To correct this omission, the methods have been amended with the following paragraph: “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 labelled as “unclear” (Table 2).”.

	Comment: The authors should be clearer about how they use and what they mean by clonality and asexuality particularly with respect to recombination. Answer: We carefully re-read the manuscript to make the use of terminology unambiguous. We use the term “asexuality” sparsely and only in the introduction, when referencing to the works of other authors, which wrote about asexuality in the sense of the absence of conventional sexual reproduction in fungi. In other parts of the manuscript we use the term “clonality” – reproduction, which produces clonal offspring. While the absence of sexual reproduction still allows for recombination (e.g. through parasexuality), the term clonality implies the absence of recombination altogether – something that is supported by genomic data for both H. werneckii and A. melanogenum (with the exception of intraspecific hybrids). We also added the short definition of clonality in the beginning of the introduction to indicate how the term is used in the manuscript: “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].” We hope this provides the requested clarity. Minor: Comment: The use of the term 'hybrid' should be preceded by the descriptor 'intraspecific' for clarity's sake Answer: We have added the descriptor “interspecific” to the manuscript, as suggested. Comment: A clearer description of the question/hypotheses being addressed earlier in the introduction would go a long way to improve readability. Answer: As suggested, the hypothesis has now been briefly introduced at the end of the second paragraph of the Introduction: “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 two species of such strictly clonal fungi. The analysis of 115 genomes of haploid and diploid wild strains is used to test the hypothesis that even clonal phylogenetic lineages can generate diversity through hybridization that produces highly heterozygous and stable diploids.”.
Additional Information:	
Question	**Response**
Are you submitting this manuscript to a special series or article collection?	No
Experimental design and statistics Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the data presented should be made available in the figure legends. Have you included all the information requested in your manuscript?	Yes
Resources A description of all resources used, including antibodies, cell lines, animals and software tools, with enough	Yes

information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible. Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?	
Availability of data and materials All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript. Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?	Yes

Clonality, inbreeding, and hybridization in two extremotolerant black yeasts

Cene Gostinčar [a,b,¶,*], Xiaohuan Sun [c,¶], Anja Černoša [a], Chao Fang [c], Nina Gunde-Cimerman [a,+], Zewei Song [c,+]

[a] Department of Biology, Biotechnical Faculty, University of Ljubljana, 1000 Ljubljana, Slovenia

[b] Lars Bolund Institute of Regenerative Medicine, BGI-Qingdao, Qingdao 266555, China

[c] BGI-Shenzhen, Beishan Industrial Zone, Shenzhen 518083, China

¶ *These authors contributed equally to this work as first authors.*

+ *These authors contributed equally to this work as senior authors.+*

ORCID iDs: Cene Gostinčar [0000-0002-0149-3674]; Xiaohuan Sun [0000-0002-3411-5995]; Anja Černoša []; Chao Fang [0000-0002-6830-3337]; Nina Gunde-Cimerman [0000-0002-9464-3263]; Zewei Song [0000-0002-0606-6946];

*** Corresponding author: Cene Gostinčar**
Department of Biology, Biotechnical Faculty
University of Ljubljana
Jamnikarjeva 101
SI-1000 Ljubljana
Slovenia
Tel.: +386-1-320 3392
Fax: +386-1-2573390
E-mail: cene.gostincar@bf.uni-lj.si, cgostincar@gmail.com

Xiaohuan Sun: sunxiaohuan@genomics.cn
Anja Černoša: anja.cernosa@bf.uni-lj.si
Chao Fang: fangchao@genomics.cn
Nina Gunde-Cimerman: nina.Gunde-Cimerman@bf.uni-lj.si
Zewei Song: songzewei@genomics.cn

Running Title

Hybridization in extremotolerant yeasts

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 two 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 two 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 haploidisation, 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 two 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 break up 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 recognised 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 non-model 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 stabilised 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 specialised for survival in saline environments and able to grow in nearly salt-saturated solutions [31,32]. Whole genome sequencing of twelve 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 haploidisation [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 two 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 analyse 66 genomes of *H. werneckii* and 49 genomes of

A. melanogenum to provide insight into the reproductive strategy of these two extremotolerant fungi, characterised 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, Supplemental 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 four 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) were 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 recognised as haploid, 45 (68%) as diploid, and one genome as tetraploid. This was similar to the *A. melanogenum* genomes where 16 (33%) genomes were recognised as haploid, 30 (61%) as diploid, and the ploidy of three 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 9519 (±665 SD) for haploids and 20417 (±1709 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 between the two 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) Benchmarking Universal Single-Copy Orthologs 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 (haploidised genome of diploid strain EXF-2000 in case of *H. werneckii*, whole genome of haploid strain EXF-3378 in case of *A. melanogenum*). SNP analysis was performed on all strains except *H. werneckii* strain 36 due to its tetraploid genome. The average density of single nucleotide polymorphisms (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, with the first two 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 two genomes analysed 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 fifty 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 four different parts of the phylogeny. When all 50 multi-labelled 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 in the case of *H. werneckii* (with an additional event leading to tetraploid strain 36) and 7 events in the case of *A. melanogenum* (Fig. 4). Several phylogenetic lineages resulting from these events appeared to be relatively widespread – more than one representative strain was found for most lineages, often in different habitats and geographic locations. However, lineage composition was skewed in favour of specific localities or habitats. This was confirmed by Fisher's 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 two largest groups (groups 1 in Fig. 4) contained isolates from Europe, with only one 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 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 than 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, Supplemental Figs. S1, S2). The majority of these genomes were diploid: 18 (78%) in the case of *H. werneckii* and 8 (100%) in the case of *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 three of four 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 (Supplemental Fig. S3). Some of these could be explained by aneuploidy (loss of one 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 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 (Supplemental 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 (Supplemental Figs. S4, S5). In case of *H. werneckii* 'blastx' searches against the non-redundant GenBank protein database showed that one 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. In the case of *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 (Supplemental Figs. S4, S5). For example, in case of *H. werneckii* hybrid groups generally contained two similar copies

of MAT1-1, but groups 1a, 3 and 4 contained very different copies and group 9 consistently lacked one copy of MAT1-1 altogether (possibly due to poor assembly of the locus). In contrast, two different copies of MAT1-2 were found in group 1, but two 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 two 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) (Supplemental 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 one case genome 23) were the only ones to contain HET genes belonging to phylogenetic clusters 6 and 24 (Supplemental Table S2). HET proteins from cluster 12 were found only in hybrid groups 1-3 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 (Supplemental 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 *Hortaea werneckii* and 49 strains of the black yeast *Aureobasidium 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 two principal components explained nearly 60% of the observed diversity – much more than, for example, in the homogenous and

recombining species *Aureobasidium pullulans*, where the first two 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 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 four 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 recognised 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 two or more genomic 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 either 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, and their possible role in the formation of hybrids in *H. werneckii* and *A. melanogenum*.

While hybridization to diploids and the absence of meiosis is 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 haploidisation – 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 haploidisation. Either such haploidisation 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 haploidisation. 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 hybridisation 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 recognised as a common tool for genome shaping and stabilisation 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 two 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 three 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 two 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 does not only have shortcomings, but also considerable advantages in adapting to stress [6,13]. The same is true for parasexuality, which not only alters ploidy but 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

427 certain drugs) that promotes parasexual recombination [19]. In *Candida albicans*,
428 which can undergo regular sexual recombination, stress additionally promotes the
429 parasexual cycle, which generates a high degree of diversity, including aneuploidy
430 [53].

431 There are also several examples of at least occasionally recombining extremotolerant
432 and extremophilic species. For example, the polyextremotolerant yeast *Aureobasidium*
433 *pullulans*, a close relative of *A. melanogenum*, has one of the highest rates of
434 recombination demonstrated in fungi by population genomics [38]. Two salt-adapted
435 basidiomycetes, the halotolerant *Wallemia mellicola* and the halophilic *Wallemia*
436 *ichthyophaga*, also appear to recombine, albeit much less frequently than *A. pullulans*,
437 even though *W. ichthyophaga* is exceptionally rare and limited to highly fragmented
438 environments [10,54]. If clonality or hybridization are indeed beneficial for adaptation
439 to extreme conditions – and more data are needed to test this hypothesis –
440 recombination appears to be compatible with extremotolerant lifestyle as well.

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

450 Some species are able to thrive in a wide range of different environments without
451 adapting to any of them at the genomic level – the ubiquitous and polyextremotolerant
452 *A. pullulans* is one such example [38]. But while *A. pullulans* is an exceptionally
453 generalistic species, the two species analysed here are less so: *A. melanogenum* is
454 mostly restricted to aquatic and indoor environments, while *H. werneckii* is mostly
455 found in marine and hypersaline environments and has a much higher upper salinity
456 growth limit than *A. pullulans*. The preference of some *H. werneckii* and *A.*
457 *melanogenum* strain groups for specific habitats (Fig. 4) possibly indicates an ongoing
458 clonality-driven specialisation of these groups. This would be in line with the suggestion
459 of Romeo et al. [36]. However, based on the here studied dataset the observed habitat

preferences might be an artefact 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 analysed 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 four out of five analysed 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 specialisation of either of the two species [56].

3. If the formation of diploids in *H. werneckii* and *A. melanogenum* is irreversible, what drives the co-existence 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 co-existence 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 one parameter tested in a laboratory setting. The possibility of different adaptation value of haploids vs. 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

493 randomly selected environmental strain is about twice as likely to be diploid as haploid,
494 a single tetraploid strain of *H. werneckii* is the only evidence that these diploids can
495 hybridize further. No triploid strains of either species have been found. Either the fusion
496 of diploid cells is prevented by some as yet unknown mechanism, or the resulting
497 strains do not persist in the environment long enough to be detected.

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

521 5. How common is hybridisation in clonal fungi? Clonality itself appears to be much
522 rarer in fungi than once believed, but population genomic studies of *Neurospora* spp.
523 showed that even closely related species can differ substantially in their reproductive
524 strategies [3]. This is also the case here: while *A. melanogenum* is clonal, the closely
525 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 five fungal species from extreme environments that we have previously studied using population genomics, two 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 two 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, nor by haploidisation through random chromosome loss, as is typical of parasexuality. Hybrid lineages avoid the loss of heterozygosity even over timeframes allowing 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 haploidisation, 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 *Hortaea 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). 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, USA), preceded by homogenisation with a pestle and mortar in liquid nitrogen and 1 min in Retsch Mixer Mill 301 (ThermoFisher Scientific, USA) at 20 Hz. After the RNAse A treatment (ThermoFisher Scientific, USA), the isolated DNA was evaluated using agarose electrophoresis and by fluorometry (Qubit; ThermoFisher Scientific, USA).

Genome sequencing

The genome sequencing was performed using the platform BGISEQ-500, with 2×150-bp libraries, prepared as described previously [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 (CNSA) of China National GeneBank DataBase (CNGBdb) 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 haploidised 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 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 single-nucleotide polymorphism 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 optimised with training using the scripts provided with the program with (i) the RNAseq data from Sinha et al. [35] deposited at GenBank Sequence Read Archive with the accession 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 benchmarking universal single-copy orthologs (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 labelled 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=15000. 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 optimised 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 1364 bp to 5089 bp for *H. werneckii* and 3400 bp to 13257 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 1000 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 visualisations 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 analysed in R using Fisher's 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 (Table 1,2) 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 1000 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 1000 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 visualised with ‘ggplot2’ [66,71] for 50 longest reference genome contigs in the case of *H. werneckii* and for 35 longest reference genome contigs in the case of *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 in case of *H. werneckii*, 35 in case of *A. melanogenum*) and plotted as 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.

Acknowledgements

The authors would like to thank Yonglun Luo (Lars Bolund Institute of Regenerative Medicine, Qingdao-Europe Advanced Institute for Life Sciences, BGI-Qingdao; BGI-Shenzhen; Department of Biomedicine, Aarhus University) for supporting the work on genomics of fungi from extreme environments; Yuchong Tang (China National GeneBank, BGI-Shenzhen) for his invaluable help in project management, organization of scientific visits and facilitation of collaboration between the project partners; and Toni Gabaldón (Biomedical Research Institute (IRB), Barcelona Supercomputing Centre (BSC)) for inspiring and constructive discussions on hybridization in fungi. The authors thank Rafael R. Montalvo-Rodriguez (University of Puerto Rico) for the strains EXF-2515 and EXF-2516; Filomena de Leo (University of Messina, Italy) for the strains EXF-10508 to EXF-10512; Armando Azua Bustos (Pontificia Universidad Católica de Chile) for the strains EXF-11540, EXF-11650 and EXF-11651; Zhu-Hua Luo (Third Institute of Oceanography, China) for the strains EXF-12619 and EXF-12620; Amy Gladfelter (University of North Carolina) for the strains EXF-14591 and EXF-14592; and Christine Beemelmanns (Hans Knöll Institute, Germany) for the strain EXF-14590. We acknowledge the China National Gene Bank for the support of sequencing library preparation and shotgun sequencing. This study was supported by funding from the Slovenian Research Agency to Infrastructural Centre Mycosmo (MRIC UL I0-0022), programmes P4-0432, P1-0198, and project J4-2549.

Authors contributions

Conceptualization of the study: NGC, ZS and CG; experimental work: XS, AČ, and CF; bioinformatic analyses: CG; data curation: CG; resources: ZS, NGC; preparation of the manuscript and visualisations: CG; review and editing of the manuscript: CG, NGC, AČ, XS, CF and ZS; supervision: ZS and NGC; funding acquisition: NGC, ZS.

Declarations

Funding

The authors acknowledge the China National Gene Bank for the support of sequencing library preparation and shotgun sequencing. This study was supported by funding from the Slovenian Research Agency to Infrastructural Centre Mycosmo (MRIC UL), programmes P1-0170, P1-0198, project J4-2549, and young researcher grant to Anja Černoša.

Conflicts of interest/Competing interests

Not applicable.

Ethics approval

Not applicable.

Availability of data and material

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].

References

1. Gostinčar C, Stajich JE, Zupančič J, Zalar P, Gunde-Cimerman N. Genomic evidence for intraspecific hybridization in a clonal and extremely halotolerant yeast. *BMC Genomics*. 2018; doi: 10.1186/s12864-018-4751-5.

2. Brown A. Mating in mushrooms: increasing the chances but prolonging the affair. *Trends Genet*. 2001; doi: 10.1016/S0168-9525(01)02343-5.

3. Gladieux P, De Bellis F, Hann-Soden C, Svedberg J, Johannesson H, Taylor JW. Neurospora from natural populations: Population genomics insights into the life history of a model microbial eukaryote. In: Y. DJ, editor. *Stat Popul Genomics*. New york: Humana;

4. Hawksworth DL, Kirk PM, Sutton BC, Pegler DN. Ainsworth & Bisby's Dictionary of the Fungi. 8th ed. Oxford University Press;

5. Tibayrenc M, Ayala FJ. Reproductive clonality of pathogens: A perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. *Proc Natl Acad Sci*. 2012; doi: 10.1073/pnas.1212452109.

6. Taylor JW, Hann-Soden C, Branco S, Sylvain I, Ellison CE. Clonal reproduction in fungi. *Proc Natl Acad Sci*. 2015; doi: 10.1073/pnas.1503159112.

7. Carreté L, Ksiezopolska E, Pegueroles C, Gómez-Molero E, Saus E, Iraola-Guzmán S, 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; doi: 10.1016/j.cub.2017.11.027.

8. Černoša A, Sun X, Gostinčar C, Fang C, Gunde-Cimerman N, Song Z. Virulence traits and population genomics of the black yeast *Aureobasidium melanogenum*. *J Fungi*. 2021; doi: 10.3390/jof7080665.

9. Kasuga T, White TJ, Koenig G, Mcewen J, Restrepo A, Castañeda E, et al.. Phylogeography of the fungal pathogen *Histoplasma capsulatum*. *Mol Ecol*. 2003; doi: 10.1046/j.1365-294X.2003.01995.x.

10. Gostinčar C, Sun X, Zajc J, Fang C, Hou Y, Luo Y, et al.. Population genomics of an obligately halophilic basidiomycete *Wallemia ichthyophaga*. *Front Microbiol*. 2019; doi: 10.3389/fmicb.2019.02019.

11. Tsai IJ, Bensasson D, Burt A, Koufopanou V. Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proc Natl Acad Sci*. 2008; doi: 10.1073/pnas.0707314105.

12. Xu J. Cost of interacting with sexual partners in a facultative sexual microbe. *Genetics*. 2005; doi: 10.1534/genetics.105.045302.
13. Sun S, Heitman J. Is sex necessary? *BMC Biol*. 2011; doi: Artn 56 Doi 10.1186/1741-7007-9-56.
14. Otto SP. The evolutionary enigma of sex. *Am Nat*. 2009; doi: 10.1086/599084.
15. Gostinčar C, Grube M, De Hoog S, Zalar P, Gunde-Cimerman N. Extremotolerance in fungi: evolution on the edge. *FEMS Microbiol Ecol*. 2010; doi: 10.1111/j.1574-6941.2009.00794.x.
16. Gostinčar C, Gunde-Cimerman N, Grube M. Polyextremotolerance as the fungal answer to changing environments. In: Bakermans C, editor. *Microb Evol under Extrem Cond*. Berlin, München, Boston: De Gruyter;
17. Pontecorvo G. The parasexual cycle in fungi. *Annu Rev Microbiol*. 1956; doi: 10.1146/annurev.mi.10.100156.002141.
18. Clutterbuck AJ. Parasexual recombination in fungi. *J Genet*. 1996; doi: 10.1007/BF02966308.
19. Engel T, Verweij PE, van den Heuvel J, Wangmo D, Zhang J, Debets AJM, et al.. Parasexual recombination enables *Aspergillus fumigatus* to persist in cystic fibrosis. *ERJ Open Res*. 2020; doi: 10.1183/23120541.00020-2020.
20. Todd RT, Forche A, Selmecki A. Ploidy variation in fungi: Polyploidy, aneuploidy, and genome evolution. *Microbiol Spectr*. 2017; doi: 10.1128/microbiolspec.FUNK-0051-2016.
21. Gerstein AC, Sharp NP. The population genetics of ploidy change in unicellular fungi. *FEMS Microbiol Rev*. 2021; doi: 10.1093/femsre/fuab006.
22. Naranjo–Ortiz MA, Gabaldón T. Fungal evolution: cellular, genomic and metabolic complexity. *Biol Rev*. 2020; doi: 10.1111/brv.12605.
23. Hill R, Leitch IJ, Gaya E. Targeting Ascomycota genomes: what and how big? *Fungal Biol Rev*. 2021; doi: 10.1016/j.fbr.2021.03.003.
24. Chen G, Rubinstein B, Li R. Whole chromosome aneuploidy: Big mutations drive adaptation by phenotypic leap. *BioEssays*. 2012; doi: 10.1002/bies.201200069.
25. Gerstein AC, Lim H, Berman J, Hickman MA. Ploidy tug-of-war: Evolutionary and genetic environments influence the rate of ploidy drive in a human fungal pathogen. *Evolution (N Y)*. 2017; doi: 10.1111/evo.13205.
26. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, et al.. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*. 2018; doi:

807 10.1038/s41586-018-0030-5.
808 27. Depotter JR, Seidl MF, Wood TA, Thomma BP. Interspecific hybridization impacts
809 host range and pathogenicity of filamentous microbes. *Curr Opin Microbiol*. 2016; doi:
810 10.1016/j.mib.2016.04.005.
811 28. Samarasinghe H, You M, Jenkinson TS, Xu J, James TY. Hybridization facilitates
812 adaptive evolution in two major fungal pathogens. *Genes (Basel)*. 2020; doi:
813 10.3390/genes11010101.
814 29. Morrow CA, Fraser JA. Ploidy variation as an adaptive mechanism in human
815 pathogenic fungi. *Semin Cell Dev Biol*. 2013; doi: 10.1016/j.semcdb.2013.01.008.
816 30. Forche A. Large-scale chromosomal changes and associated fitness
817 consequences in pathogenic fungi. *Curr Fungal Infect Rep*. 2014; doi:
818 10.1007/s12281-014-0181-2.
819 31. Gunde-Cimerman N, Zalar P, Hoog S, Plemenitaš A. Hypersaline waters in
820 salterns - natural ecological niches for halophilic black yeasts. *FEMS Microbiol Ecol*.
821 2000; doi: 10.1111/j.1574-6941.2000.tb00716.x.
822 32. Plemenitaš A, Lenassi M, Konte T, Kejžar A, Zajc J, Gostinčar C, et al.. Adaptation
823 to high salt concentrations in halotolerant/halophilic fungi: a molecular perspective.
824 *Front Microbiol*. 2014; doi: 10.3389/fmicb.2014.00199.
825 33. Zalar P, Zupančič J, Gostinčar C, Zajc J, de Hoog GSS, De Leo F, et al.. The
826 extremely halotolerant black yeast *Hortaea werneckii* - a model for intraspecific
827 hybridization in clonal fungi. *IMA Fungus*. 2019; doi: 10.1186/s43008-019-0007-5.
828 34. Lenassi M, Gostinčar C, Jackman S, Turk M, Sadowski I, Nislow C, et al.. Whole
829 genome duplication and enrichment of metal cation transporters revealed by *de novo*
830 genome sequencing of extremely halotolerant black yeast *Hortaea werneckii*. *PLoS*
831 *One*. 2013; doi: 10.1371/journal.pone.0071328.
832 35. Sinha S, Flibotte S, Neira M, Formby S, Plemenitaš A, Gunde-Cimerman N, et al..
833 Insight into the recent genome duplication of the halophilic yeast *Hortaea werneckii*:
834 Combining an improved genome with gene expression and chromatin structure. *G3-*
835 *Genes Genomes Genet*. 2017; doi: 10.1534/g3.117.040691.
836 36. Romeo O, Marchetta A, Giosa D, Giuffrè L, Urzì C, De Leo F. Whole genome
837 sequencing and comparative genome analysis of the halotolerant deep sea black yeast
838 *Hortaea werneckii*. *Life*. 2020; doi: 10.3390/life10100229.
839 37. Azua-Bustos A, Gonzalez-Silva C, Salas L, Palma RE, Vicuna R. A novel subaerial
840 *Dunaliella* species growing on cave spiderwebs in the Atacama Desert. *Extremophiles*.

841 2010; doi: 10.1007/s00792-010-0322-7.
842 38. Gostinčar C, Turk M, Zajc J, Gunde-Cimerman N. Fifty *Aureobasidium pullulans*
843 genomes reveal a recombining polyextremotolerant generalist. *Environ Microbiol*.
844 2019; doi: 10.1111/1462-2920.14693.
845 39. Gostinčar C, Ohm RA, Kogej T, Sonjak S, Turk M, Zajc J, et al.. Genome
846 sequencing of four *Aureobasidium pullulans* varieties: biotechnological potential,
847 stress tolerance, and description of new species. *BMC Genomics*. 2014; doi:
848 10.1186/1471-2164-15-549.
849 40. Zhu YO, Sherlock G, Petrov DA. Whole genome analysis of 132 clinical
850 *Saccharomyces cerevisiae* strains reveals extensive ploidy variation. *G3*
851 *Genes|Genomes|Genetics*. 2016; doi: 10.1534/g3.116.029397.
852 41. O'Brien CE, Oliveira-Pacheco J, Ó Cinnéide E, Haase MAB, Hittinger CT, Rogers
853 TR, et al.. Population genomics of the pathogenic yeast *Candida tropicalis* identifies
854 hybrid isolates in environmental samples. Vylkova S, editor. *PLOS Pathog*. 2021; doi:
855 10.1371/journal.ppat.1009138.
856 42. Rhodes J, Desjardins CA, Sykes SM, Beale MA, Vanhove M, Sakthikumar S, et
857 al.. Tracing genetic exchange and biogeography of *Cryptococcus neoformans* var.
858 *grubii* at the global population level. *Genetics*. 2017; doi:
859 10.1534/genetics.117.203836.
860 43. Gabaldón T. Hybridization and the origin of new yeast lineages. *FEMS Yeast Res*.
861 2020; doi: 10.1093/femsyr/foaa040.
862 44. Pryszcz LP, Németh T, Gácser A, Gabaldón T. Genome comparison of *Candida*
863 *orthopsilosis* clinical strains reveals the existence of hybrids between two distinct
864 subspecies. *Genome Biol Evol*. 2014; doi: 10.1093/gbe/evu082.
865 45. Pryszcz LP, Németh T, Saus E, Ksiezopolska E, Hegedűsová E, Nosek J, et al..
866 The genomic aftermath of hybridization in the opportunistic pathogen *Candida*
867 *metapsilosis*. Zhang J, editor. *PLOS Genet*. 2015; doi: 10.1371/journal.pgen.1005626.
868 46. Gostinčar C, Stajich JE, Kejžar A, Sinha S, Nislow C, Lenassi M, et al.. Seven
869 years at high salinity — experimental evolution of the extremely halotolerant black
870 yeast *Hortaea werneckii*. *J Fungi*. 2021; doi: 10.3390/jof7090723.
871 47. Boekhout T, Aime MC, Begerow D, Gabaldón T, Heitman J, Kemler M, et al.. The
872 evolving species concepts used for yeasts: from phenotypes and genomes to
873 speciation networks. *Fungal Divers*. 2021; doi: 10.1007/s13225-021-00475-9.
874 48. Mixão V, Gabaldón T. Hybridization and emergence of virulence in opportunistic

875 human yeast pathogens. *Yeast*. 2018; doi: 10.1002/yea.3242.
876 49. Strom NB, Bushley KE. Two genomes are better than one: history, genetics, and
877 biotechnological applications of fungal heterokaryons. *Fungal Biol Biotechnol*. 2016;
878 doi: 10.1186/s40694-016-0022-x.
879 50. Mitchison-Field LMY, Vargas-Muñiz JM, Stormo BM, Vogt EJD, Van Dierdonck S,
880 Pelletier JF, et al.. Unconventional cell division cycles from marine-derived yeasts. *Curr*
881 *Biol*. 2019; doi: 10.1016/j.cub.2019.08.050.
882 51. Saupe SJ. Molecular genetics of heterokaryon incompatibility in filamentous
883 Ascomycetes. *Microbiol Mol Biol Rev*. 2000; doi: 10.1128/MMBR.64.3.489-502.2000.
884 52. Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, Vienne DM, et al.. Fungal
885 evolutionary genomics provides insight into the mechanisms of adaptive divergence in
886 eukaryotes. *Mol Ecol*. 2014; doi: 10.1111/mec.12631.
887 53. Berman J, Hadany L. Does stress induce (para)sex? Implications for *Candida*
888 *albicans* evolution. *Trends Genet*. 2012; doi: 10.1016/j.tig.2012.01.004.
889 54. Sun X, Gostinčar C, Fang C, Zajc J, Hou Y, Song Z, et al.. Genomic Evidence of
890 Recombination in the Basidiomycete *Wallemia mellicola*. *Genes (Basel)*. 2019; doi:
891 10.3390/genes10060427.
892 55. Drenth A, McTaggart AR, Wingfield BD. Fungal clones win the battle, but
893 recombination wins the war. *IMA Fungus*. 2019; doi: 10.1186/s43008-019-0020-8.
894 56. Gostinčar C, Zajc J, Lenassi M, Plemenitaš A, de Hoog S, Al-Hatmi AMS, et al..
895 Fungi between extremotolerance and opportunistic pathogenicity on humans. *Fungal*
896 *Divers*. 2018; doi: 10.1007/s13225-018-0414-8.
897 57. Zajc J, Gostinčar C, Černoša A, Gunde-Cimerman N. Stress-tolerant yeasts:
898 Opportunistic pathogenicity versus biocontrol potential. *Genes (Basel)*. 2019; doi:
899 10.3390/genes10010042.
900 58. Gostinčar C. Towards genomic criteria for delineating fungal species. *J Fungi*.
901 2020; doi: 10.3390/jof6040246.
902 59. Fogelqvist J, Tzelepis G, Bejai S, Ilbäck J, Schwelm A, Dixelius C. Analysis of the
903 hybrid genomes of two field isolates of the soil-borne fungal species *Verticillium*
904 *longisporum*. *BMC Genomics*. 2018; doi: 10.1186/s12864-017-4407-x.
905 60. Michelotti LA, Sun S, Heitman J, James TY. Clonal evolution in serially passaged
906 *Cryptococcus neoformans* × *deneoformans* hybrids reveals a heterogenous landscape
907 of genomic change. Stajich J, editor. *Genetics*. 2022; doi: 10.1093/genetics/iyab142.
908 61. Steenwyk JL, Lind AL, Ries LNA, dos Reis TF, Silva LP, Almeida F, et al..

909 Pathogenic allodiploid hybrids of *Aspergillus* fungi. *Curr Biol*. 2020; doi:
910 10.1016/j.cub.2020.04.071.
911 62. Fang C, Zhong H, Lin Y, Chen B, Han M, Ren H, et al.. Assessment of the cPAS-
912 based BGISEQ-500 platform for metagenomic sequencing. *Gigascience*. 2018; doi:
913 10.1093/gigascience/gix133.
914 63. Huang S, Kang M, Xu A. HaploMerger2: rebuilding both haploid sub-assemblies
915 from high-heterozygosity diploid genome assembly. Berger B, editor. *Bioinformatics*.
916 2017; doi: 10.1093/bioinformatics/btx220.
917 64. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al.. The Sequence
918 Alignment/Map format and SAMtools. *Bioinformatics*. 2009; doi:
919 10.1093/bioinformatics/btp352.
920 65. Alkan C, Coe BP, Eichler EE. GATK toolkit. *Nat Rev Genet*. 2011; doi:
921 10.1038/nrg2958.
922 66. R Development Core Team. R: A language and environment for statistical
923 computing. Vienna, Austria: R Foundation for Statistical Computing;
924 67. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide
925 SNP data. *Bioinformatics*. 2011; doi: 10.1093/bioinformatics/btr521.
926 68. Schliep K, Potts AJ, Morrison DA, Grimm GW. Intertwining phylogenetic trees and
927 networks. *Methods Ecol Evol*. 2017; doi: 10.1111/2041-210X.12760.
928 69. Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide
929 population genetic data with emphasis on clonality. *Front Genet*. 2015; doi:
930 10.3389/fgene.2015.00208.
931 70. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al.. The
932 variant call format and VCFtools. *Bioinformatics*. 2011; doi:
933 10.1093/bioinformatics/btr330.
934 71. Wickham H. ggplot2. Cham: Springer International Publishing;
935 72. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: A *de novo* assembler for single-
936 cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*.
937 2012; doi: 10.1093/bioinformatics/bts174.
938 73. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in
939 eukaryotes that allows user-defined constraints. *Nucleic Acids Res*. 2005; doi:
940 10.1093/nar/gki458.
941 74. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. BUSCO:
942 Assessing genome assembly and annotation completeness with single-copy

orthologs. *Bioinformatics*. 2015; doi: 10.1093/bioinformatics/btv351.
75. Kriventseva E V., Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, Pozdnyakov IA, et al.. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res*. 2015; doi: 10.1093/nar/gku1220.
76. Geib SM, Hall B, Derego T, Bremer FT, Cannoles K, Sim SB. Genome Annotation Generator: a simple tool for generating and correcting WGS annotation tables for NCBI submission. *Gigascience*. 2018; doi: 10.1093/gigascience/giy018.
77. Minkin I, Medvedev P. Scalable multiple whole-genome alignment and locally collinear block construction with SibeliaZ. *Nat Commun*. 2020; doi: 10.1038/s41467-020-19777-8.
78. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*. 2007; doi: 10.1080/10635150701472164.
79. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al.. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Teeling E, editor. *Mol Biol Evol*. 2020; doi: 10.1093/molbev/msaa015.
80. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2006; doi: 10.1093/molbev/msj030.
81. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 2012; doi: 10.1111/j.2041-210X.2011.00169.x.
82. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al.. Pfam: The protein families database in 2021. *Nucleic Acids Res*. 2021; doi: 10.1093/nar/gkaa913.
83. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; doi: 10.1093/molbev/mst010.
84. Altschul S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; doi: 10.1093/nar/25.17.3389.
85. Gostinčar C, Sun X, Černoša A, Fang C, Gunde-Cimerman N, Song Z. Supporting data for "Clonality, inbreeding, and hybridization in two extremotolerant black yeasts" GigaScience Database 2022. http://dx.doi.org/10.5524/102257

976 **Table 1.** *Hortaea werneckii* strains analysed 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	sea water-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	sea water, depth 25 m	Italy	2
EXF-10509	33	sea water, depth 200 m	Italy	2
EXF-10510	34	sea water, depth 94 m	Italy	2
EXF-10511	35	sea water, depth 25 m	Italy	2
EXF-10512	36	sea water, 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
EXF-10513	G	deep sea water	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 1-54 were sequenced in this study; strains A-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-L corresponds to names in Gostinčar et al. [1]

985 **Table 2.** *Aureobasidium melanogenum* strains analysed 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 (4500 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 centre	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]

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	6885	3806	4457	30312
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)	9519	9344	665	20417	19240	1709	46596
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%	/

* BUSCOs, Benchmarking Universal Single-Copy Orthologues.

Figure legends

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) are expressed as millions of SNPs (numbers next to dashed lines).

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 non-singleton 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).

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, 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, tetraploid strain is additionally marked with an asterisk.

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 coloured 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 coloured line in the outermost edge of the tree, diploid genomes are represented by a double coloured line, the only tetraploid genome is represented by four coloured lines. Around the tree, coloured 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.

Figure 5. Aneuploid regions in *Hortaea werneckii* (**A**) and *Aureobasidium melanogenum* (**B**) genomes. Per-nucleotide sequencing depth of regions corresponding to 50 and 35 longest contigs of *H. werneckii* and *A. melanogenum* were 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 one putatively aneuploid region are plotted in colour. Other genomes are plotted in light grey. Colours of strain names in the legend mark haploid (blue) and diploid (red) genomes.

Supplementary Information

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

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

Supplemental Figure S1. Aneuploid regions in *Hortaea werneckii* genomes. Per-nucleotide sequencing depth of regions corresponding to 50 longest contigs of *H. werneckii* and *A. melanogenum* were 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 are marked with red rectangles.

Supplemental Figure S2. Aneuploid regions in *Aureobasidium melanogenum* genomes. Per-nucleotide sequencing depth of regions corresponding to 35 longest contigs of *A. melanogenum* were 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 are marked with red rectangles.

Supplemental Figure S3. Heterozygosity in diploid *Hortaea werneckii* and *Aureobasidium 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 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.

Supplemental Figure S4. Putative mating type loci of *Hortaea werneckii*. Visualisation 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.

Supplemental Figure S5. Putative mating type loci of *Aureobasidium melanogenum*. Visualisation 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.

Assoc Prof Dr Cene Gostinčar
Department of Biology, Biotechnical Faculty, University of Ljubljana
Jamnikarjeva 101
SI-1000 Ljubljana
Slovenia

Ljubljana, 29. 7. 2022

Dear Editors,

We are submitting a revised version of the manuscript of an article entitled "**Clonality, inbreeding, and hybridization in two extremotolerant black yeasts**".

The opinions of the reviewers were generally positive. One found the studied system very "intriguing" and the other saw it as an "interesting and novel result that could spark important research into evolutionary losses of sex, speciation, and recombination rates". However, the reviewers also requested several revisions, mostly in the form of additional analyses of the data. The revised manuscript was prepared in accordance to these requests and comments. We performed all requested analyses, amended the text, figures and tables accordingly and made (mostly minor) changes to the text, where requested by the reviewers.

The detailed list of changes and replies to the reviewers is submitted separately. Besides the revised manuscript, we are also submitting a version of the manuscript in which all the changes are clearly marked with the "track changes" option for your reference.

The manuscript is prepared as a MS Word file, with the figures submitted as separate files.

We thank the reviewers for their detailed and constructive reviews, which helped us to improve and strengthen the text. We thus hope you can re-consider the manuscript for publication.

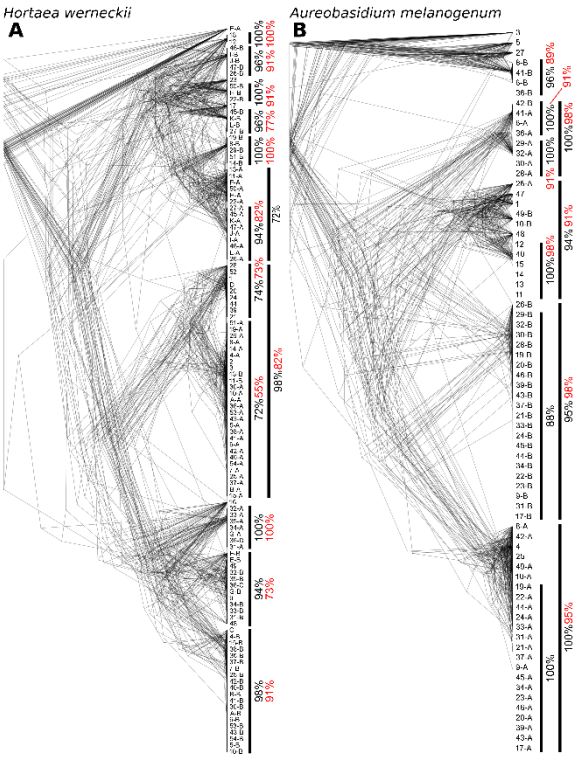
Sincerely yours,
Assoc Prof Dr Cene Gostinčar

We would like to thank the reviewers for their detailed and constructive feedback! Both reviewers requested revisions, which have now been made. The detailed responses to the comments of the reviewers are listed below. We believe this strengthens the manuscript and we hope you can consider it for publication.

Reviewer #1

Comment: separate trees were generated for "longest alignments" and I think this approach can hide potential admixture events. It is not reported anywhere what are the average lengths of these alignments, but the point is that if admixture concerns a small part of the chromosome, the alignment of the whole chromosome will not detect those admixture events. I would suggest to split all alignments into portions of equal length or of equal number of informative SNPs, to identify potential admixture events, if any exist.

Answer: The range of lengths of the alignments is now reported in the methodology: “lengths of 1364 bp to 5089 bp for *H. werneckii* and 3400 bp to 13257 bp for *A. melanogenum*”. The length of these alignments was not overwhelming, therefore the masking of the admixture events was not explicitly addressed in the initial manuscript – they should be able to reflect also admixture events concerning only small parts of the chromosome. If the alignments were split further, the number of phylogenetically informative sites in them would be limited. The divergence of the genomes and likely also the difficult assembly of diploid genomes precluded the recovery of longer alignments with a 100% representation of all haploid genomes. However, as an alternative to the alignment-splitting approach, phylogenetic trees were also constructed from core BUSCOs that were found in all genomes of each species (and in the number of copies corresponding to the ploidy of the genome). Since the number of genomes was considerable and since an unexpected copy number in even a single genome excluded the gene from the analysis, the final dataset was fairly modest. Alignments longer than 200 nucleotides and with an average of at least 15 nucleotide differences between gene pairs were used for phylogeny reconstruction, producing 22 phylogenetic trees in case of H. werneckii and 44 in case of A. melanogenum. While these phylogenies are generally based on shorter alignments and are perhaps less reliable (which was the reason for reporting Sibelia and not BUSCO alignments in the manuscript), they result in almost identical topology and the proportion of trees supporting the major splits are very similar. Please see the here provided phylogeny with percentages to the right of trees marking the proportion of trees supporting the major clusters. Black percentage points are now reported in the revised Fig. 3 (in response to the comment below; calculated from phylogenies based on genomic alignments produced by Sibelia). Red numbers are the percentages of trees supporting the same clusters in trees produced from alignment of BUSCOs (provided here for review purposes only).



[*please refer to the attached file Answers.to.reviewers.REVIEW.ONLY.docx for the image*]

Comment: tree discordancies are not quantified in any way and from figure 3 it's hard to judge how much concordance there is. If the species cluster in several groups one could show how many topologies (proportion) of these major clusters are consistent with each other and how many are different
Answer: The proportion of phylogenies supporting the major clusters have now been labelled in panels A and B of Fig. 3 as suggested. The figure legend has been amended accordingly.

Minor comments:
Comment: Table S1 and Table 1: Please explain in the legend what distributions are showing. Also it would be helpful to include a column in the tables with information about ploidy.
Answer: The legend of Table S1 (which shows the distributions) has been amended to explain the plots in more detail. The column reporting the ploidy has been added as suggested to Table 1 and Table S1. Additionally, the ploidy column has been also added to Table 2 to keep the format of tables consistent.
Comment: line 183: I'm not sure what the authors mean by 'consistent' in this sentence. Wasn't the ploidy decided from genome assembly characteristics? In this case it's expected to be consistent.
Answer: This was an awkwardly written sentence – we did not want to say that the genomic characteristics were consistent with ploidy (which they were, because this is how the ploidy was determined, as correctly pointed out by the reviewer), but that haploid strains were very similar in their characteristics to each other, and the same was true for diploid strains. We rewrote the sentence as follows and hopefully this makes the message clearer: “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).”.
Comment: line 195: I would suggest explaining here in once sentence how SNP calling was made, especially how the reference was constructed, because it's quite important for interpreting the results.
Answer: We added a short description, as suggested: “Single-nucleotide polymorphisms (SNPs) were determined with Genome Analysis Toolkit after mapping the sequencing reads to reference genomes (haploidised genome of diploid strain EXF-2000 in case of *H. werneckii*, whole genome of haploid strain EXF-3378 in case of *A. melanogenum*).” We hope this is sufficiently detailed for the Results section, since a more detailed description is provided in the Methods.
Comment: line 202: PCA plots in Fig1: It would be useful to add ploidy information in the plot to see where are those samples located relative to haploids. Are these clusters explained by geography or habitat? Perhaps adding this information would be useful as well. It is also not mentioned anywhere how divergent are these clusters. It would be worth reporting nucleotide divergence between (haploid) genomic groups.
Answer: Diploid strains have now been marked with bold labels in all panels of Fig. 1 – PCA plots as suggested, but also in phylogenetic networks. The sizes and placement of labels have been optimized to increase legibility. The clusters can indeed be partially explained by geography/habitat – these information are visualised in Fig. 4 and discussed there as well. This decision was made after we first attempted to show the geography/habitat data in Fig. 1 (PCA plots) and discovered that the number of genomes and their considerable overlapping make an efficient visualisation all but impossible. Divergence between clusters is now also marked in Fig. 1 with dashed lines marking the groups of haploid strains and distance in millions of SNPs between groups. The legend of the figure has been amended to reflect these changes.
Comment: line 221: Please highlight haploids/diploid on the phylogeny.

Answer: In Fig. 3, ploidy has now been marked on the phylogeny – diploid strain names have been written in bold in panels C and D. In panels A and B diploid (and tetraploid) strains were already marked with letters added to strain names – this has now been explicitly explained in the legend as well.
Comment: line 209: Considering LD decay analysis it looks to me that r2 is very low even between close variants. In general, it is not clear from the figure 2 what is the maximum r2 between adjacent pairs of SNPs (start of the line) and what is the distance over which r2 falls by half. I think the authors should give some quantification of this in the results. This could give a better understanding of the LD.
Answer: The maximum value of LD (drawn as the upper red horizontal line on the LD charts) are now also reported in the Results, as suggested: "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*.". Regarding the low values, the initial r2 value is not only dependent on the disequilibrium, but also on the allele frequency (i.e. alleles in less than 50% frequency that are in total disequilibrium, will have r2 not 1 but less than 1) and is therefore not unusual.
Comment: line 229: This is a really interesting way to show relationships between hybrids!
Answer: Thank you!
Comment: line 250: Fig 5, What are colours in the legend signifying?
Answer: The sentence explaining the colours was lost in revision and was now again added: "Colours of strain names in the legend mark haploid (blue) and diploid (red) genomes." Thank you for alerting us to this!
Comment: line 282: MAT loci: One option to make sure if MAT loci is truly absent would be to look for reads matching the sequence of MAT. This could eliminate the possibility that the quality of an assembly is a source of missing loci.
Answer: The original manuscript already acknowledged the possibility that the apparently missing loci are the consequence of the assembly problems and not of their true absence. We now performed an additional analysis, as suggested. In many cases this was not helpful because the divergence of some loci is so large, that it resulted in gaps in the sequencing coverage even in cases where the putative mating-type loci were found in the assembly. In other cases, especially in A. melanogenum, the coverage did indeed suggest, as the reviewer thought it might, that the loci are actually present in the genome, but not present in the genome assembly. The results of the sequencing coverage analysis have been added to Supplementary Figs. 4 and 5 and the corresponding figure legends amended as appropriate. The corresponding Results section of the manuscript has been changed as follows: "While mapping of sequencing reads of some strains to 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 (Supplemental 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."
Our opinion on this is that the sequencing coverage analysis comes with its own set of problems, but its inclusion does indeed provide a more complete picture of a complex situation. While a more conclusive resolution of this question can only be provided by using a sequencing technology producing longer reads, the results presented in this manuscript, while partial, do provide relevant information tightly linked to the main topic of the manuscript as well as important indications for future research, at the same time acknowledging the partiality of the results.
Comment: line 263: In the figures S4 and S5 one information that is missing is whether the same MAT type is present always on the same genomic background, assuming that these different types are in the same locus. Examples of MAT introgression are common in fungi so it would be nice to check if something like this occurs in these species.

Answer: We agree that this is an interesting question and we did some preliminary investigation into it before preparing the manuscript. However, due to the above discussed suboptimal assembly in the mating-type loci, this was only possible for some strains. Additionally, the precise borders of the mating locus in H. werneckii are not known. Based on all of this we decided not to report highly uncertain data and rather leave this question to be solved by a more targeted experiment, e.g. by using a sequencing technology producing long reads.
Comment: line 619: In the description of processing sequence alignments, please specify what do you mean by "long gaps"? Was there any threshold?
Answer: This was indeed not as precise as it should be. We replaced “long gaps” with “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”.

Reviewer #2

Major:
Comment: The methods are missing a description of how ploidy was estimated, the results of which are integral to the conclusions.
Answer: Thank you for pointing this out. To correct this omission, the methods have been amended with the following paragraph: “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 labelled as “unclear” (Table 2).”.

Comment: The authors should be clearer about how they use and what they mean by clonality and asexuality particularly with respect to recombination.
Answer: We carefully re-read the manuscript to make the use of terminology unambiguous. We use the term “asexuality” sparsely and only in the introduction, when referencing to the works of other authors, which wrote about asexuality in the sense of the absence of conventional sexual reproduction in fungi. In other parts of the manuscript we use the term “clonality” – reproduction, which produces clonal offspring. While the absence of sexual reproduction still allows for recombination (e.g. through parasexuality), the term clonality implies the absence of recombination altogether – something that is supported by genomic data for both *H. werneckii* and *A. melanogenum* (with the exception of intraspecific hybrids). We also added the short definition of clonality in the beginning of the introduction to indicate how the term is used in the manuscript: “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].” We hope this provides the requested clarity.

Minor:
Comment: The use of the term 'hybrid' should be preceded by the descriptor 'intraspecific' for clarity's sake
Answer: We have added the descriptor “interspecific” to the manuscript, as suggested.
Comment: A clearer description of the question/hypotheses being addressed earlier in the introduction would go a long way to improve readability.

Answer: As suggested, the hypothesis has now been briefly introduced at the end of the second paragraph of the Introduction: “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 two species of such strictly clonal fungi. The analysis of 115 genomes of haploid and diploid wild strains is used to test the hypothesis that even clonal phylogenetic lineages can generate diversity through hybridization that produces highly heterozygous and stable diploids.”.

Figure 1

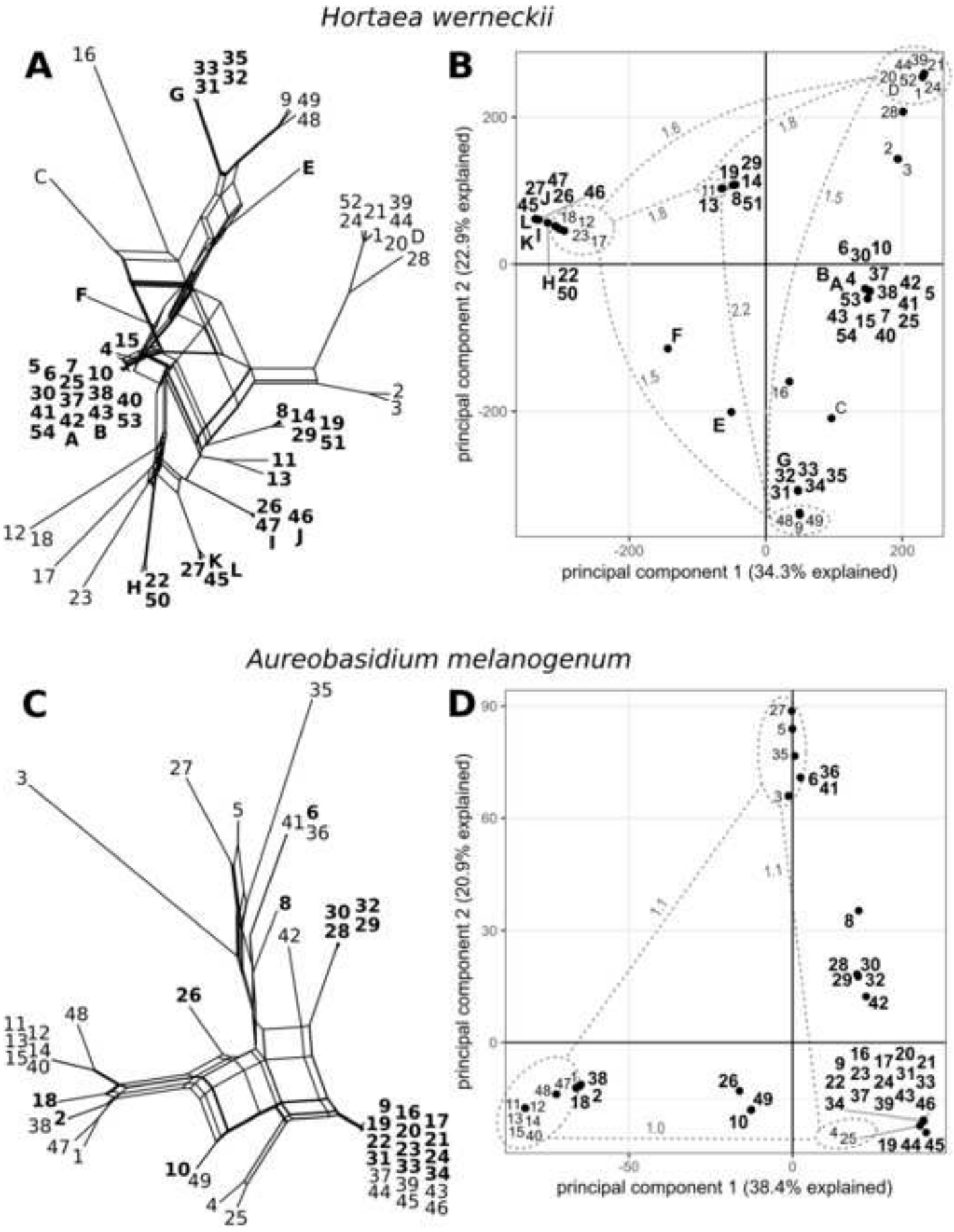


Figure 2

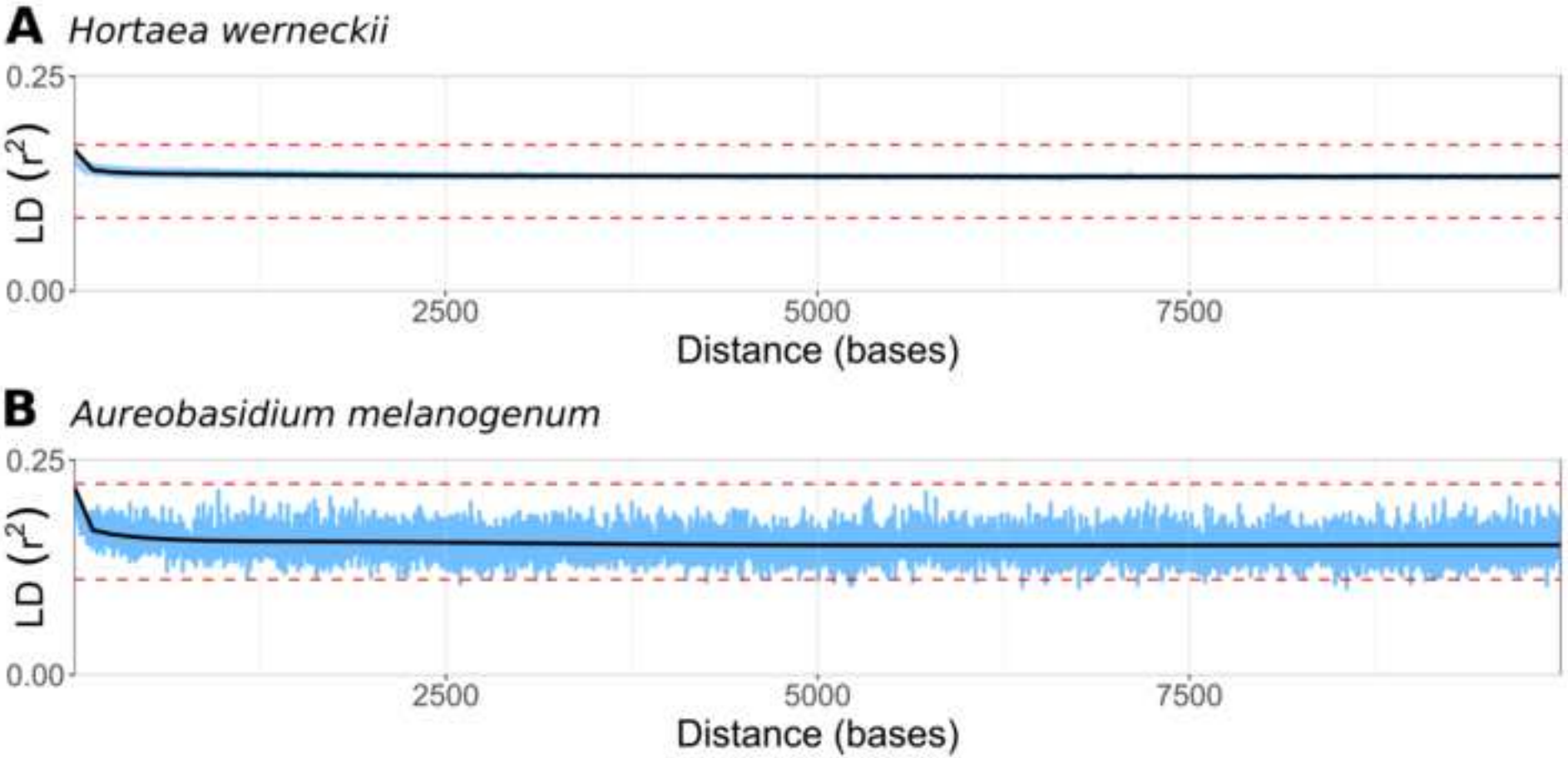


Figure 3

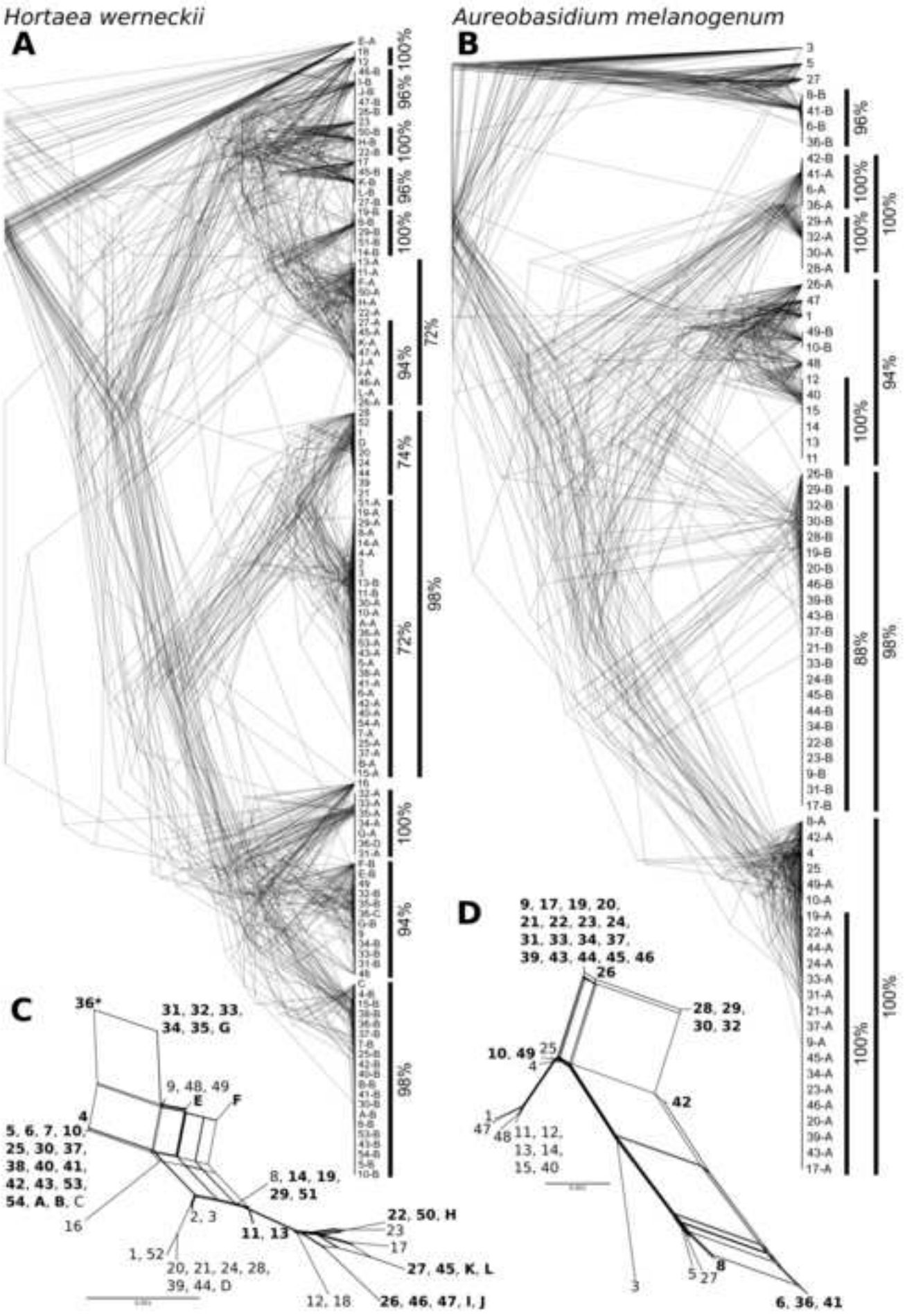


Figure 4

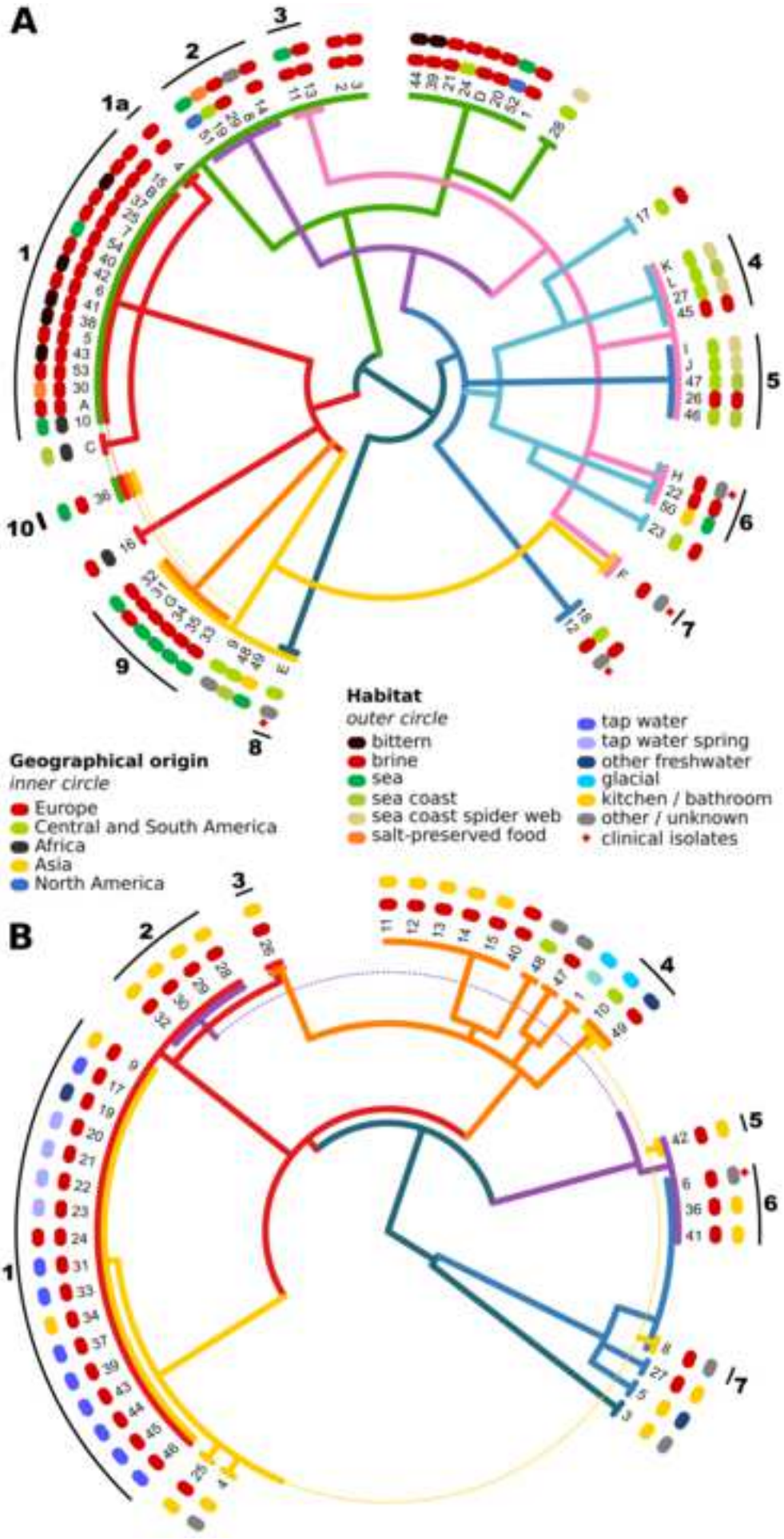
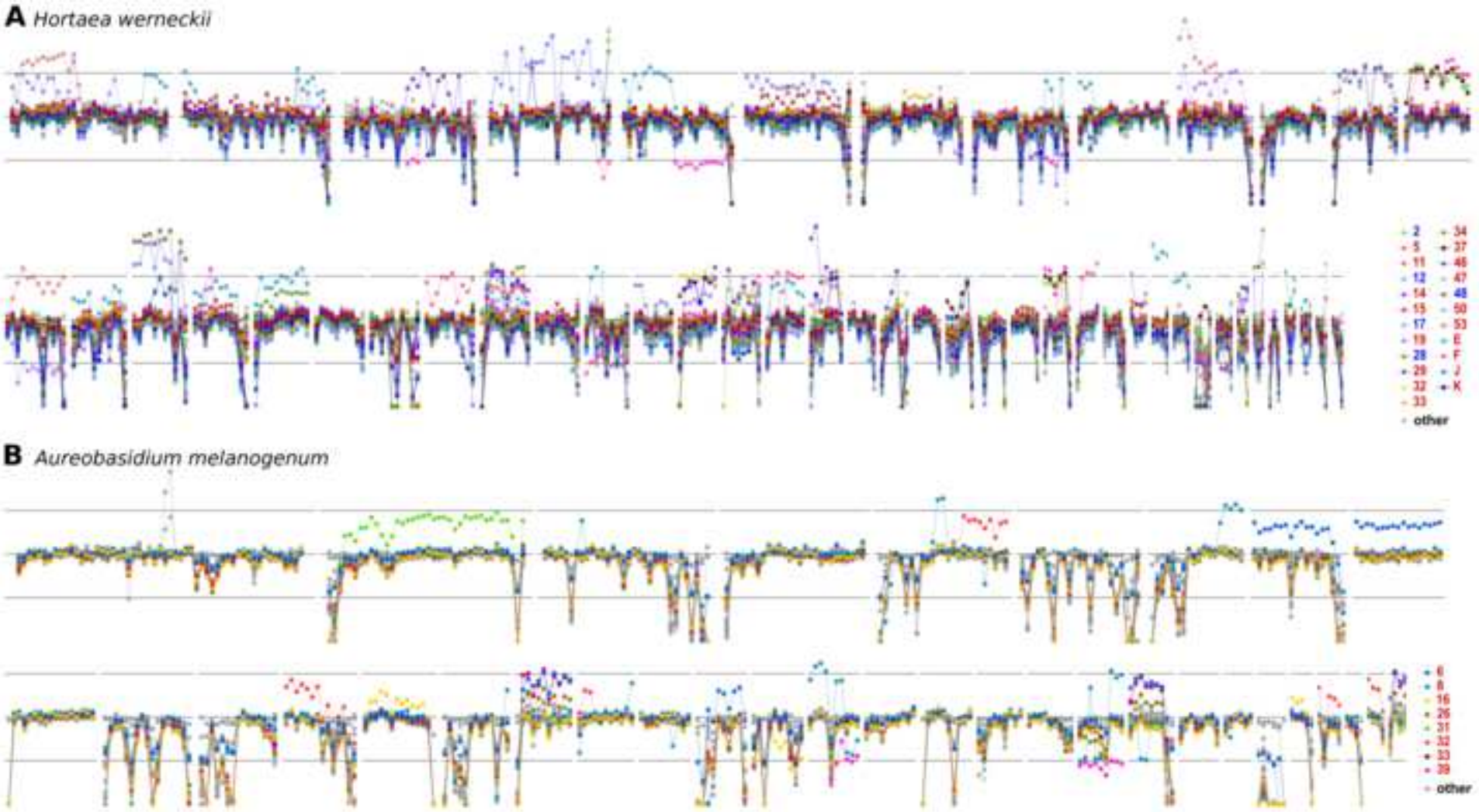


Figure 5

Click here to access/download;Figure;Fig5.png



Supplemental Table S1

Supplemental Table S2

Supplemental Figure S1

Supplemental Figure S2

Click here to access/download
Supplementary Material
FigS2.png

Supplemental Figure S3

Supplemental Figure S4

Supplemental Figure S5