



OPEN

## High intragenomic, intergenomic, and phenotypic diversity in pulcherrimin-producing *Metschnikowia* yeasts indicates a special mode of genome evolution

Matthias Sipiczki<sup>1✉</sup>, Kinga Czentye<sup>1</sup> & Zoltán Kállai<sup>1,2</sup>

In molecular systematics, the delimitation of yeast species is based on the notion that the barcode differences are smaller within species than between them. The most widely used barcodes are segments of the chromosomal repeats coding for ribosomal RNAs that are homogenised in yeasts. The analysis of these segments of the type strains of ten species recently merged in *Metschnikowia pulcherrima* and 37 new isolates demonstrated that this is not the case in this species. The intragenomic diversity significantly exceeded the threshold gaps used to differentiate related yeast species. Large segments of the D1/D2 domains were not diverse within the genomes and could therefore be used to determine the taxonomic affiliation of the isolates. The genome structures of the isolates were compared by RAPD and the RFLP of the mitochondrial DNA. Both patterns were highly heterogeneous. The sequence analysis of the *PUL4* gene (a member of the *PUL* gene cluster involved in pulcherrimin production) revealed very high intragenomic differences, suggesting that the genomes may be chimerised. Three phenotypic traits related to the antimicrobial antagonism characteristic of the species were also highly diverse and prone to reversible segregation resembling epigenetic processes (silencing and reactivation of regulators) rather than mutations and back-mutations. These features make *M. pulcherrima* unique among yeasts and indicate that it evolves in a non-standard way.

**Keywords** Barcode diversity, Chimeric genome, Antifungal antagonism, Pulcherrimin, Dimorphic transitions, Invasive growth

Species delimitation in yeasts is based on the combination of barcode DNA sequence analysis and the examination of selected physiological, biological and morphological properties<sup>1</sup>. The most important barcodes are segments of the chromosomal repeats that code for rRNA (primary barcodes) and segments of genes of certain conserved proteins (secondary barcodes). The barcode sequences are identical or highly similar in conspecific organisms and different in allospecific organisms<sup>2,3</sup>. The latter difference is the barcode gap, on which the taxonomic differentiation of species is based.

The application of the barcodes is based on the assumptions that (1) the rRNA cistrons (repeats) containing the primary barcodes are homogenised across the repeat array(s) in the genome of an organism to ensure that all have identical nucleotide sequences<sup>4</sup> and (2) the secondary barcodes are present in single copies or in more but identical copies (no heterozygosity)<sup>1</sup>.

Recent analyses of the barcodes of the type strains of ten pulcherrimin-producing *Metschnikowia* species (grouped in the *pulcherrima* clade) found no clear barcode gaps between them due to very high intra-genomic (intra-strain) diversity of the sequences<sup>5,6</sup>. The intra-genomic sequence diversity is reflected in high numbers of “ambiguous nucleotides” (dimorphic, rarely polymorphic sites) in amplicon sequencing. Individual barcode sequences cloned from the amplified barcode DNA of a strain often differed much more from each other than

<sup>1</sup>Department of Genetics and Applied Microbiology, University of Debrecen, Debrecen, Hungary. <sup>2</sup>Present address: Institute of Horticulture, University of Debrecen, Debrecen, Hungary. ✉email: gecela@post.sk

from sequences cloned from strains of other type strains of the clade. This phenomenon is rather unusual in yeasts and renders the determination of the taxonomic affiliation of strains and their differentiation from each other practically impossible. Because of the lack of clear barcode gaps the ten species of the clade were merged in one species under the taxonomic name *M. pulcherrima*<sup>6</sup>. However, the merged species is highly heterogeneous due to the extremely high inter- and intragenomic primary barcode diversity. The degree of diversity is exceptionally high compared to other yeast species and was hypothesised to be attributable to a birth-and-death mechanism of repeat evolution instead of array homogenisation<sup>7</sup>. The analysis of the secondary barcodes and additional 28 protein-encoding genes of four strains for which complete genome sequences were available revealed heterozygosity for all of them. The analysed genes had at least two copies in the genomes which differed in sequence and phylogenetic history. These copies often differed from each other more than from their counterparts in other strains (higher intra-genomic than inter-genomic differences). Certain genes containing secondary barcode segments were found in more than two copies, all differing in sequence, indicating aneuploidy. From these results it was concluded that the genomes of the type strains of the merged species are chimeras composed of variable copies of genes of different phylogenetic histories<sup>5,6</sup>.

*M. pulcherrima* gained attention due to the antimicrobial activity of its strains which can be exploited in protecting crops from harmful (e.g. rotting) microorganisms (for a review, see<sup>8</sup>). Numerous yeast species are known to exert adverse effects on other microorganisms but *M. pulcherrima* stands out with the extent and strength of its antimicrobial antagonism. Several antagonistic mechanisms have already been revealed and investigated in yeasts (for recent reviews, see<sup>9,10</sup>). One mechanism is related to the production of pulcherrimin, a red complex, a characteristic feature of *M. pulcherrima* and many *Kluyveromyces* strains [e.g.<sup>11–13</sup>]. Pulcherrimin is a complex of pulcherriminic acid and ferric ions<sup>14</sup>. Pulcherriminic acid is a hydroxamic acid derivative of the cyclic dipeptide cyclo(L-leucyl-L-leucyl)<sup>15</sup>. It harbours two hydroxamate groups and forms pulcherrimin upon binding Fe<sub>3</sub><sup>14,16,17</sup>. The reaction can take place within the cytoplasm of the yeast cells and also extracellularly because the pulcherriminic acid can be secreted into the environment. Since pulcherrimin is a coloured substance, its accumulation in the cytoplasm turns the colony dark red, whereas its formation outside the cells produces a pigmented halo in the medium around the colony<sup>16</sup>.

The biosynthesis and secretion of pulcherriminic acid has been investigated in detail in *Bacillus subtilis* and *B. licheniformis* (for a review, see<sup>18</sup>) and a similar pathway is assumed to operate in *K. lactis* and *M. pulcherrima* (for a review, see<sup>8</sup>). In the genomes of the latter organisms, four genes organised in the group called the *PUL* cluster encode proteins participating in the production of pulcherriminic acid and its release in the environment<sup>8,13</sup>.

As the reaction of pulcherriminic acid with ferric ions is irreversible<sup>19</sup> and pulcherrimin is insoluble in water, the process is unlikely to play a role in iron acquisition. The chelated iron in pulcherrimin is inaccessible to the biochemical processes of microorganisms. Since iron is required for the activity of many proteins and cellular processes, its immobilisation by pulcherriminic acid adversely affects the propagation of many microorganisms and can even be lethal to their cells<sup>12</sup>. Due to their antimicrobial activities, the pulcherrimin-producing *M. pulcherrima* strains can be used as biological agents to protect agricultural commodities and food products against pathogenic and destructive microorganisms (for a review, see<sup>8</sup>). The strength of their antagonistic effect correlates with the intensity of the extracellular production of pulcherrimin (e.g.<sup>12,20–22</sup>).

The aim of this study was to investigate how characteristic the high genomic and physiological diversity observed in the type strains of the merged species is for *M. pulcherrima*. For the investigation, 37 “wild” strains were isolated in 8 localities from various fruits which are natural habitats for this species. Their molecular examination revealed high intragenomic (intrastrain), intergenomic (interstrain) molecular diversity and diverse genome structures. All strains had diverse primary barcode sequences and diverse copies of the *PUL4* gene (one of the four genes of the *PUL* cluster) in their genomes. Almost all strains had unique RAPD patterns and significant differences were detected between their mitochondrial genomes. Three phenotypic traits relevant for bioprotection namely pulcherrimin production, antifungal antagonism, and substrate invasion were also diverse and prone to reversible segregation resembling epigenetic processes. However, no clear correlation was observed between the genomic and the phenotypic diversities. Apart from the revealed very high diversity, this study also finds that the *M. pulcherrima* strains can switch to invasive pseudohyphal growth. This property has been overlooked in previous studies, although it may significantly contribute to the biological protective power of *Metschnikowia* strains by sticking their pseudohyphae to injuries on the plant surface.

## Results

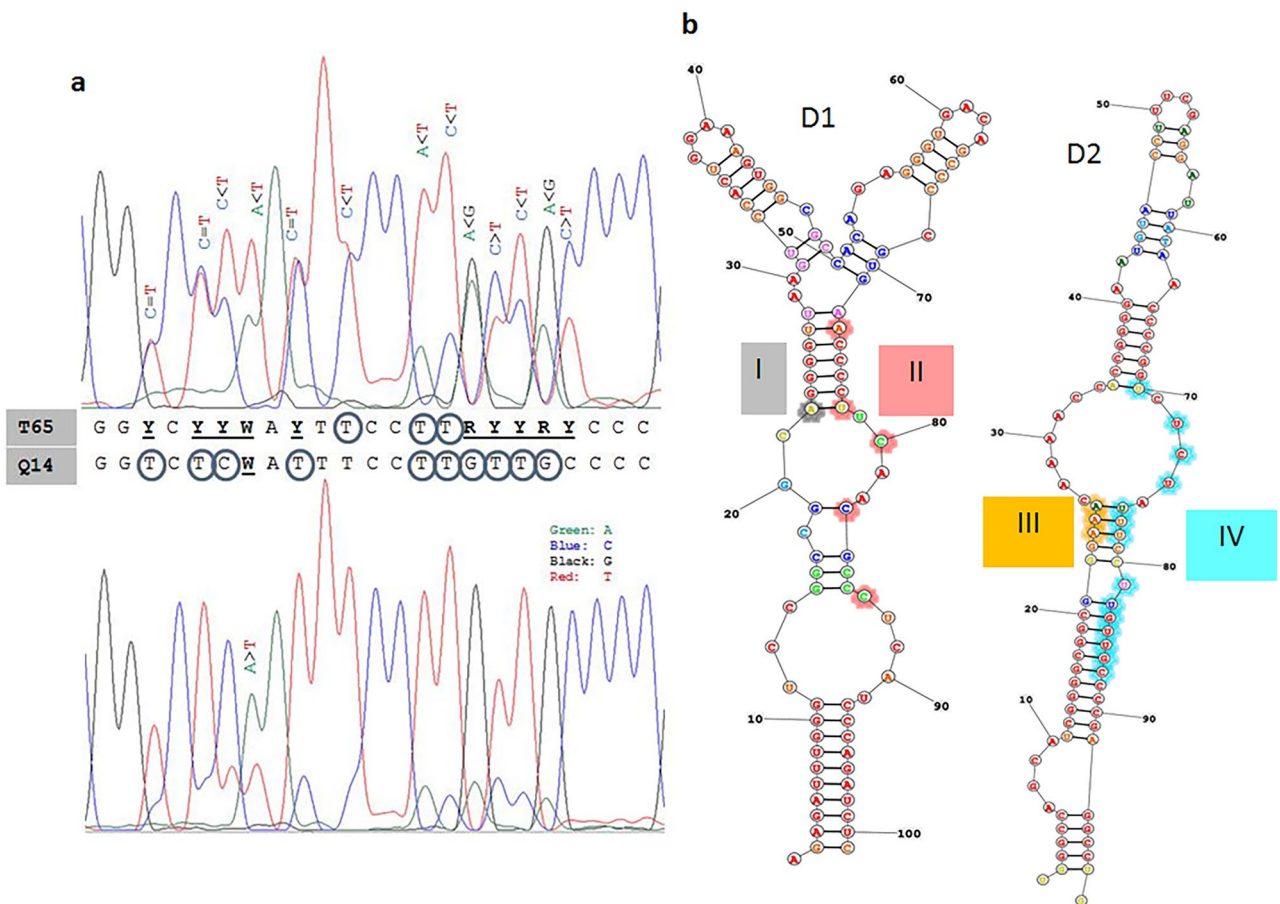
### Strain isolation

To isolate pulcherrimin-producing *Metschnikowia* strains, fruits were collected in the years 2022 and 2023. Supplementary Table S1 lists the fruit samples from which pigmented colonies were obtained. One colony was isolated for each fruit sample if the colonies were homogeneous in colour intensity on the FeCl<sub>3</sub>-containing plates and two colonies were isolated when the colour intensity was heterogeneous. Such pairs are marked with “a” and “b” in the table. 37 isolates were retained for the diversity analysis.

### Genetic-genomic diversity

*The barcode sequences of the chromosomal repeats coding for rRNA are diverse both within the genomes and in comparison with other genomes (isolates)*

The D1/D2 domains of the genes coding for the large subunit rRNA and the ITS1-5S-ITS2 segments were amplified and sequenced with the Sanger method. All sequences contained ambiguous nucleotides at numerous positions and double peaks in even more positions of the chromatogrammes (Fig. 1a). In the latter cases, the peaks were different in height, and the nucleotides corresponding to the higher peaks were shown in the final



**Figure 1.** Intragenomic diversity of primary barcodes. **(a)** Double peaks in D1/D2 chromatograms. The variable blocks IV (see Table 1) of the isolates Z65 and Q14 are shown. Ambiguous nucleotides are underlined. Note that ambiguous (dimorphic) symbol is shown when both peaks are high. When one peak is much higher than the other, only the nucleotide that corresponds to the higher peak is shown. This nucleotide occurs more frequently but not exclusively in the amplified DNA molecules. The majority nucleotides are marked with circles. **(b)** Locations of the blocks of nucleotide variability in the predicted secondary structures of D1 and D2 hairpin-stem loops. The most variable positions (see Table 1) are marked with coloured background.

sequence. All variable positions were SNDs (single nucleotide dimorphisms), mostly transitions Y (C/T) and R (A/G). No SNP (single nucleotide polymorphism) was detected.

In the D1/D2 sequences, the variable positions were concentrated in four blocks (I, II, III and IV in Table 1). The sequence of the block IV of an isolate with high number of ambiguous nucleotides and the sequence of the isolate with the lowest number of ambiguous nucleotides are compared in Fig. 1a. All blocks are located in hairpin-stem loops of the secondary structures of the transcripts (Fig. 1b).

The intragenomic diversity of the ITS segments was much higher. Very few positions had identical nucleotides in all ITS1 sequences and indels occurred frequently in ITS2. In contrast to the dimorphism of the variable positions of the D1/D2 domains, three or four nucleotides alternated in certain positions of the ITS1 segments (SNT or SNP sites). The ITS2 segments contained much fewer ambiguous nucleotides but much more indels (Table 2). Because of their high intragenomic diversity, the ITS sequences were not deposited in databases.

Blast searches with the D1/D2 sequences in INSDC databases identified the sequences of *M. pulcherrima* CBS 5833<sup>T</sup> (NRRL Y-7111<sup>T</sup>) as the most similar type-strain sequences. The identity varied between 93 and 99%, depending on the numbers of ambiguous positions in the sequence of the isolate and in the database sequence. Searches with ITS sequences lead to ambiguous results in certain isolates: low coverage, low percentage of identity with the *M. pulcherrima* type-strain sequences or higher identity with type strains of related species.

#### *The genomes of most isolates have unique RAPD patterns*

To compare the genomes of the isolates, PCR reactions were performed with two primers widely used for diversity studies. Both primers generated diverse banding patterns but the number of the fragments was higher when RAPD24 was used. Almost all strains had unique RAPD24 patterns. Figure 2 shows a dendrogramme inferred from the presence of 20 bands. All but three pairs of strains could be differentiated by comparing their banding patterns. One pair was isolated from the same sample suggesting that they might have been members of the same clone. Interestingly, their barcode sequences differed both in the D1/D2 (Table 1) and in the ITS segments (Table 2).

Isolate	Variable positions in sequence blocks			
	I	II	III	IV
A2	GCRGG	AACCCCYTYAACGCCYT	GAARC	GTCYYWAYTYCCTTRYRYRC
A59	GCRGG	AACCCCYTYAACGCCYT	GAAAC	GTCTCWATTTCCCTGYYGCC
A60	GCAGG	ARCCCCYTYAAMGCCYT	GAARC	GYCYYWAYTYCCTTRYYGCC
B1	GCAGG	AACCCCCTCAACGCCYT	GAAAC	GTCTCAATTTCCCTGYYGCC
B19a	GCRGG	AACCCCYTYAACGCCCT	GRARC	GYCYYWAYTYCCTRYRYRCC
B19b	GCAGG	AACCCCYTYAACGCCYT	GRARC	GYCYYWAYTYCCTRYRYRYC
B20a	GCAGG	AACCCCYTYAAMGCCYT	GRARC	GYCYYWAYTYCCTTRCYGCC
B20b	GCRGG	AACCCCTTYAAMGCCCT	GRARC	GTCYYWAYTYCCTRYRYRCC
B21	GCRGG	AACCCCYTYAACGCCYT	GAARC	GYCYYWAYTTCCCTGYYGCC
B22	GCAGG	AACCCCTTYAAMGCCYT	GRARC	GYCYYWAYTYCCWRYRYRCC
B35a	GCRGG	AACCCCTTYAACGCCYT	GRARC	GYCCTTAYTYCCWRYRYRCC
B35b	GCRGG	ARCCCCYTYAAMGCCYT	GAAAC	GTCTCWATTTCCCTGYTGCC
B36a	GCRGG	AACCCCYTYCAAMGCCCT	GAARC	GYCTYWAYTYCCTTGYGCC
B36b	GCRGG	AACCCCTTYAAMGCCYT	GRARC	GYCTYWAYTYCCATRYYGCC
B52	GCAGG	AACCCCTTYAAMGCCCT	GRARC	GCCYYTAYTYCCWTRYRYRCC
B53	GCAGG	AACCCCTTYAAMGCCCT	GAAAC	GTCTCWATTTCCCTGYTGCC
B54	GCAGG	AACCCCYTYAAMGCCYT	GAARC	GYCYYWAYTYCCTRYRYRCC
B57	GCRGG	AACCCCYTYAACGCCCT	GRARC	GTCYYWAYTYCCTRYCYRYC
B58	GCRGG	AACCCCCTYAAMGCCYT	GRARC	GYCYYWAYTYCCWTRYRYRCC
T4a	GCAGG	AACCCCTTTAAMGCCCT	GGAGC	GCCCTTACTCCCAYACCACC
T4b	GCAGG	AACCCCYTYAAMGCCCT	GGAGC	GCCYTTAYTYCCWRYCYRCC
T5	GCAGG	AACCCCTTYAAMGCCCT	GAARC	GYCYYWAYTTCCWTRYTGCC
T12a	GCAGG	AACCCCYTYAACGCCCT	GRAGC	GCCYTTACTCCCWTRCYRCC
T12b	GCRGG	AACCCCYTYAACGCCCT	GRARC	GCCYYWAYTYCCWRYCYRCC
T15	GCAGG	AACCCCTTYAAMGCCCT	GAAAC	GTCTCWATTTCCCTGYTGCC
T16a	GCAGG	AACCCCTTYAAMGCCCT	GAAAC	GTCTCAATTTCCCTGYTGCC
T16b	GCRGG	AACCCCTTCAACGCCYT	GRARC	GYCYYWAYTYCCTRYRYGCC
T40	GCRGG	AACCCCYTYAACGCCYT	GRARC	GYCYTTAYTYCCWTRYYGCC
T62	GCAGG	AACCCCYTYAACGCCCT	GRARC	GTCYTTAYTYCCTRYCYRYC
T63	GCRGG	ARCCCTTYAAMGCYYT	GAARC	GYCCTWAYTTCCCTGYTRYC
T64	GCRGG	AACCCCTTYAAMGCCYT	GAARC	GTCTYWATTTCCCTGYRYRCC
T65	GCRGG	AACCCCTTYAACGCCYT	GRARC	GYCYYWAYTTCCCTTRYRYRC
T66	GCRGG	ARCCCTTYAAMGCCYT	GRAAC	GTCTYWATTTCCCTGTGYCYC
T67	GCRGG	AACCCCTTYAAMGCCYT	GRARC	GYCYYWAYTYCCTRYRYRYC
Q14	GCAGG	AACCCCTTYAACGCCCT	GAAAC	GTCTCWATTTCCCTGTTGCC
Z17	GCRGG	AACCCCTTYAACGCCYT	GRARC	GTCYYWAYTYCCTRYCYRYC
Z18	GCRGG	AACCCCTTYAAMGCCYT	GRAGC	GYCCTTACTCCCWTACCACC

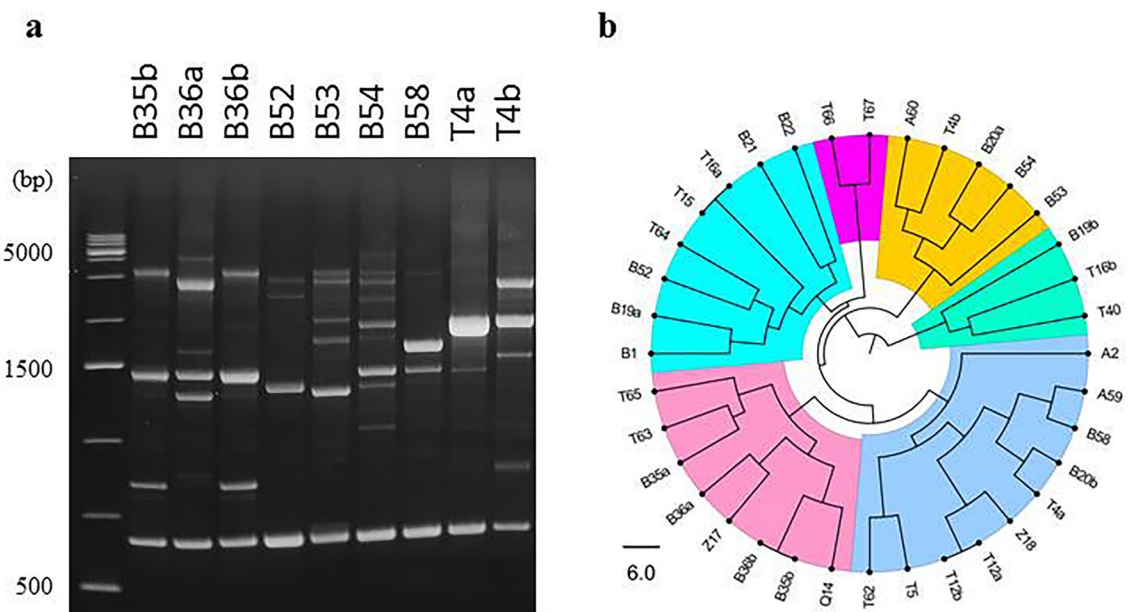
**Table 1.** Intragenomic variability in D1/D2 barcode sequences. Ambiguous nucleotides (SND sites) are marked with blue background. SND symbols: M: A or C (transversion), R: A or G (transition), Y: C or T (transition), W: A or T (transversion).

*The RFLP of mitochondrial DNA shows lower diversity among the isolates*

The digestion of the mitochondrial DNA of the isolates differentiated 11 patterns among the bands larger than 1 kb (Supplementary Fig. S1). The isolates that had identical RAPD patterns did not differ in the mitochondrial patterns either. Overall, the mtDNA RFLP diversity was lower than the RAPD diversity.

Isolate	Variable positions in ITS1	ITS2
A2	TTAAAAATTTTACACCCCTTTAGCMAAAMTC-TRAATCTTWWYCTTCAATAACCAATWA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGYCTTTTTC
A59	TTAAAAATTTTATACCCCTTTGGAAAAACCTYVWRCATYTWATTCTCAAWACACAATWA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGYCTCT--CC
A60	TTAAAAATKTTTACCCCTTTAGGAAAAA-CATCTRAACCTTWWYKTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGYCTCT--CC
B1	TTAAAAWRTTKWTACACCCCTTTAGGAAAAAMTCCTRAATCTTWWYKTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGYCTCT--CC
B19a	TTAAAAATATTATTACCCCTTTAGGAAAAAM-TCTRAATCTTWWYKTCAAATAACWYATYA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCA
B19b	TTAAAAATATTATTACCCCTTTAGGAAAAAMCTTTRAATCTTAACTTCAAATAACCAATWA	CMTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCATC---TGTC
B20a	TTAAAAATTTTATTACCCCTTTAGGAAAAAACTYTA-ATYTTATTCTCAAATAACCAATYA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGCCTTT-TCC
B20b	TTAAAAATTTTATTACACCTTTAGGAAAAAACTY-ATAATYTTATTCTCAAATAACCTWATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGYCTCT--CC
B21	TTAAWATGTTTACACCCCTTTAGGAAAAAMCTC-TAAATCTTWWCTCAAATAACWATWA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATCTTTTTC
B22	TTAAAAATATTATTACCCCTTTGGAAAAACMMWYTGACACWYTTTTTCAAWACACAATTA	CMTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGYCTCT--CC
B35a	TTAAWATGTTTACACCCCTTTAGGAAAAAACTYTTAAATYTTAAYCTCAAATAACCTWATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGCCTCT--CC
B35b	TTAAWATGTTTACACCCCTTTAGGAAAAAMCTC-TAAATCTTWWYKTCAAATAACWATWA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATCTTTTTC
B36a	TTAAATATTGTTTACACCCCTTTAGGAAAAAACTTTAAATTTAACCCTCAAATAACWYATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCA-XTCT--CC
B36b	TTAAWATGTTTACACCCCTTTAGGAAAAAMCTC-TAAATCTTWWYKTCAAATAACWATWA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATCTTTTTC
B52	TTAAAAATTTTACACCCCTTTAGGAAAAAACTYTTAAATYTTATTCTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGYCTCT--CC
B53	TTAAWATGTTTACACCCCTTTAGGAAAAAACT-ATAATCTTAACCGTCAAATACTTTTAA	CMTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCACTTTTTC
B54	TTAAAAATTTTACCCCTTTAGGAAAAAMMTTTRAATCTTWTYTTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATYCTCT--CC
B57		
B58	TTAAATTGATATTACACCCCTTTAGGAAAAAACTY-TAAATYTTATTCTCAAATAACCAATTA	CMTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGYCTCT--CC
T4a	TTAAAAATAATAACCCCTTTAGGAAAAAACTWYCATCSWYTTTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTAAGAATAAAA-GTTTTAA-GCCCATCTCT--CC
T4b	TTAAAAATTTTTTACCCCTTTAGGAAAAAACTWYTRAWCCTTTTTCAAATAACCAATTA	CMTAA-TATCAACGGCGCTA-GAATA----AGTTTTAAGCCCCAGCCTCT--CC
T5	TTAAAAWATTTTACACCCCTTTAGGAAAAAMTC-TRAATCTTWWYKTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATCTCTYTTCC
T12a	TTAAWATGTTTACACCCCTTTAGGAAAAAACTCTAAATCTTAACCTCAAATAACCAATWA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCA-TCTCT--CC
T12b	TTAAWATGTTTACACCCCTTTAGGAAAAAACTC-TAAATCTTAACTCAAATAACCAATWA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATCTTTTTC
T15	TTAAWATGTTTACACCCCTTTAGGAAAAAMTC-TRAATCTTWWYKTCAAATAACCAATWA	CCTAA-TATCAACGGCGCTA-GAAAAATAAAAGTTTTAAGCCCCATYCTCT--CC
T16a	TTAAWATGTTTACACCCCTTTAGGAAAAAACTC-TAAATCTTAACTCAAATAACCAATWA	CCTAAWATCAACGGCGCTA-GRKARK--AGTTTAA-GCCCATYCTCT--CC
T16b	TTAAATWATTTATACCCCTTTAGGAAAAAACTY-ATAATCTTAACCTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATCT
T40	AAAAATRTTATACACCTTTAGGAAAAAAACATYTGAACTTTTTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCACTCT--CC
T62	AAAAATRTTATACACCTTTAGGAAAAAACTYTTCAATYRAATCTTCAAATAACCAATYA	CCTAA-TATCAACGGCGCTAGAAATA----AGTTTTA-GCCCATCTCT-TCC
T63	TTAAWATGTTTACACCCCTTTAGGAAAAAACTC-TAAATCTTAACCGTCAAATACTYATYA	CMTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATCTTTTTC
T64	TTAAAAWSTTTTACACCCCTTTAGGAAAAAACTY-TAAATCTTAACCTCAAATAACCAATWA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATCTCTYTTCC
T65	TTAAWATGTTTACACCCCTTTAGGAAAAAACTC-TAAATCTTAACCGTCAAATACTYATYA	CCTAAATATCAACGGCGCTA-GAATA---AGTTTTA-GCCCA-XTCT--CC
T66	TTAAAAATGTTTATACACTTTA-GGCATAAACCTTWCATTAACTCTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATCT-CT--CC
T67	TTAAAAATTTTACACCCCTTTAGGAAAAAACTC-TAAATCTTAACTCAAATAACCAATYA	CCTAAATATCAACGGCGCTA-GAATA----AGTTTTA-GCCCATYCTTTTTC
Q14	TTAAAAATATTATTACCCCTTTAGGAAAAAMMTCTAAATCTTWWYKTCAAATAACWYATWA	CCTAA-TATCAACGGCGCTA-AGAAAAATAAGTTTTAAGCCCCA-XTCT--CC
Z17	TTAAAAATTTATTTCCTTTAGGAAAAAACTC-TAAATCTTWWYKTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCA-XTCT--CC
Z18	TTAAAAATATTATTACCCCTTTAGGAAAAAMCTCTRAATCTTWTYKTCAAATAACCAATTA	CCTAA-TATCAACGGCGCTA-GAATA----AGTTTTA-GCCCCAGYCTCT--CC

**Table 2.** Intragenomic variability of ITS barcode sequences. Ambiguous nucleotides are marked with blue background. SND symbols: SND symbols: K: G or T (transversion), M: A or C (transversion), R: A or G (transition), S: C or G (transversion), Y: C or T (transition), W: A or T (transversion).

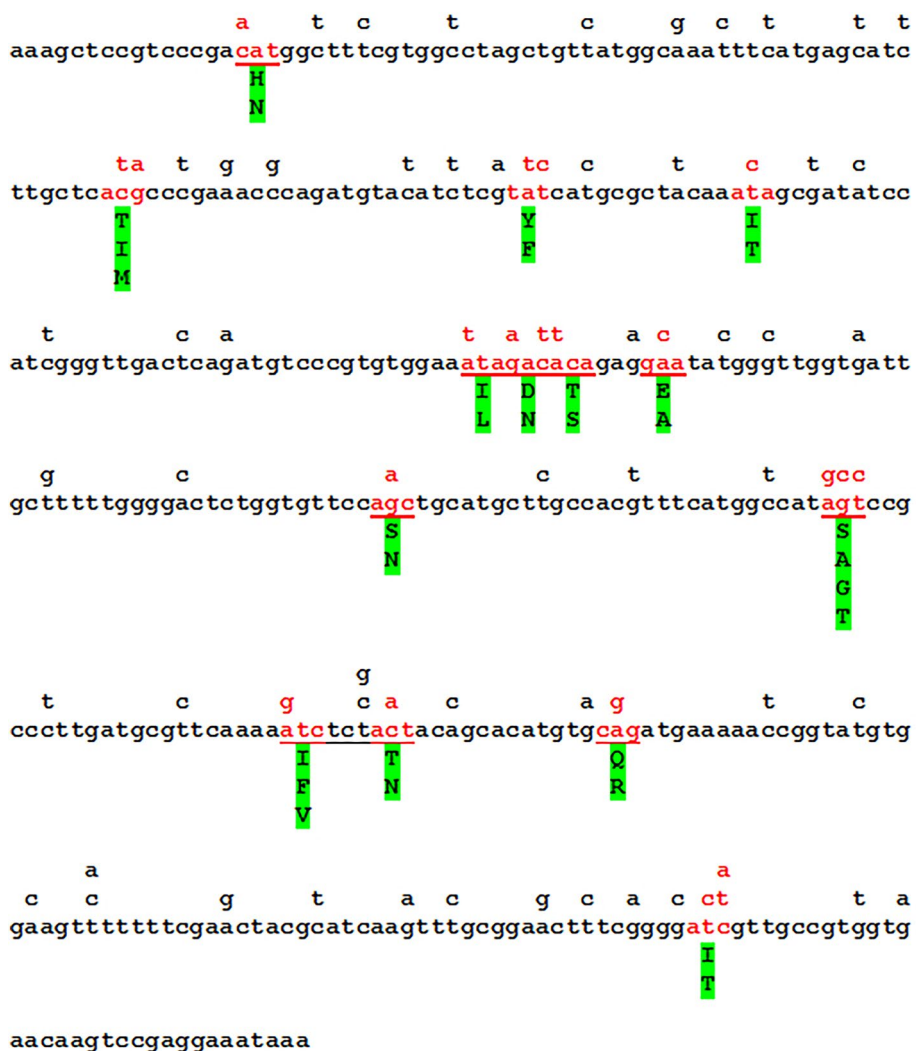


**Figure 2.** RAPD patterns of isolates. (a) Example band patterns. (b) UPGMA dendrogram of the patterns of all isolates. The grouping of the isolates (coloured clades) does not correlate with their origin.

*The PULA genes of the isolates also show both intra- and intergenomic diversity*

To find a segment suitable for the examination of the diversity of *PULA*, the sequences of the *PULA* genes of the *M. pulcherrima* strains for which genome sequences (GCA\_000317355.2; GCA\_003017285.1; GCA\_003123635.1; GCA\_004217705.1; GCA\_009746055.1; GCA\_009932455.1; GCA\_014905795.1; GCA\_029606385.1; GCA\_030583425.1; GCA\_937857175.1) are available in the NCBI Genome database and those cloned from other strains<sup>5</sup> of the species were aligned. In the MAFFT alignment, a centrally located 380 nt-long segment of the genes proved to be highly variable. This sequence was amplified and sequenced from each isolate. All isolates had ambiguous nucleotides in the sequences, but they differed in the number and location of these nucleotides (Supplementary Table S2; Fig. 3). When all sequences were considered, ambiguous nucleotides were detected in a total of 70 sites (18% of all sites). In all variable positions two nucleotides alternated (SND sites). From the lack of SNP sites it can be inferred that each strain only had two different *PULA* alleles and thus at least two copies of the *PULA* gene.

The majority of the dimorphic sites were in third positions of codons in which the nucleotide substitutions did not cause amino-acid differences between the translated sequences (Fig. 3). Translation of all sequences identified altogether 14 codons (12% of all codons) in which the nucleotide substitutions caused amino acid substitutions. Within these codons the first or the second nucleotide position or both positions were variable.



**Figure 3.** Location of nucleotide substitutions in *PULA* and their effect on the translated amino acid sequence. The entire sequence of the amplified region extracted from the genome sequences of the type strain of *M. pulcherrima* is shown. It starts at the 754th nucleotide of the *PULA* ORF. At each variable position, the nucleotide(s) differing from that of this sequence in certain isolates are shown above it. Where more than one type of SND was found at the site, the number of these nucleotides can be higher than one. All codons in which the nucleotide substitutions caused amino acid substitutions are marked with red character colour. The amino acids that can substitute each other at these sites are shown under the *PULA* sequence on green background. More specific data are available in Supplementary Table S2.

The isolate pairs T15-T16a and B35b-B36b that could not be differentiated by RAPD and mtDNA RFLP analyses were also indistinguishable in *PUL1* heterozygosity but T12a and T12b were differently heterozygous.

### Phenotypic diversity

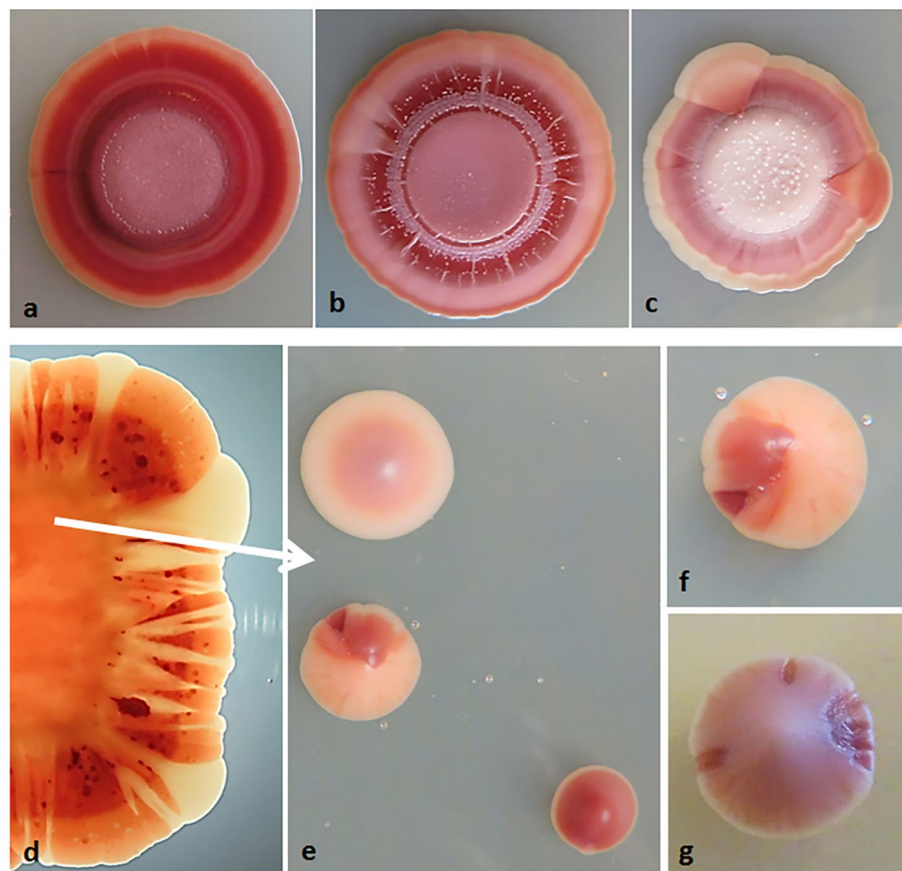
#### *The ability to produce pulcherrimin is variable and prone to reversible segregation*

The isolated colonies were maintained on YEA plates. Since the medium was not supplemented with  $\text{FeCl}_3$ , all isolates formed faintly pigmented colonies. To compare their capacity to produce pulcherrimin, the isolates were re-inoculated onto plates supplemented with various concentrations of  $\text{FeCl}_3$ . Both the intensity of colony colour and the formation of pigmented halo in the agar medium around the colonies were compared. The isolates were diverse in both traits, but the colony colour increased and the size of pigmented halo decreased with the increase of the iron content in the medium (Supplementary Table S3). The halos were only faintly visible on YEA plates without  $\text{FeCl}_3$  supplementation, so their width could not be measured.

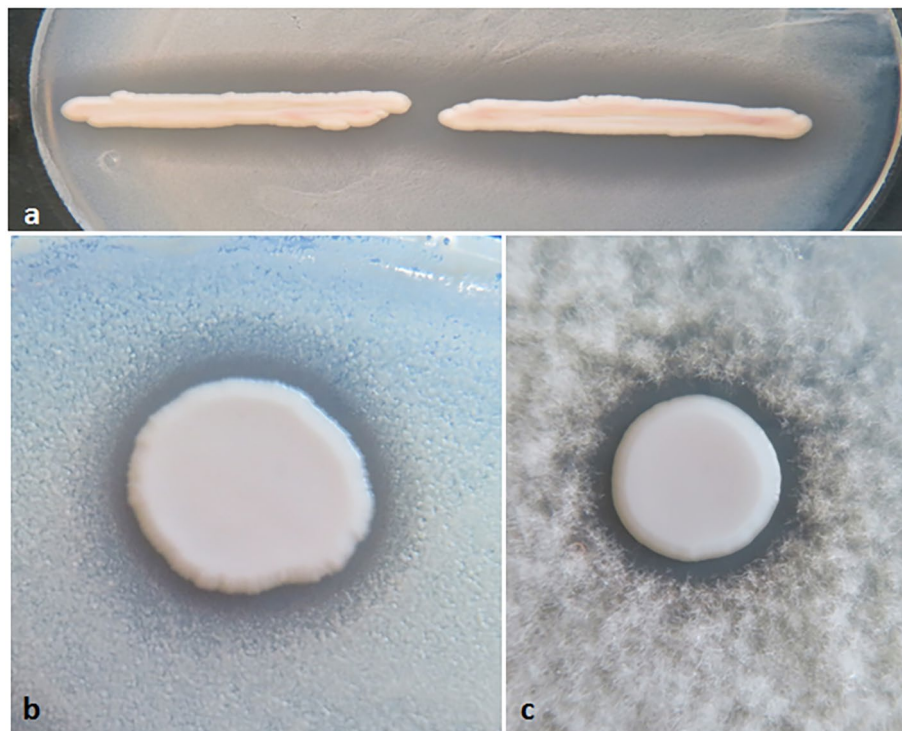
Upon longer periods of incubation (3 weeks or more), sectors of various colour intensities appeared in the colony edges of certain isolates (Fig. 4, Supplementary Table S3). This observation hints at instability (segregation) in pulcherrimin production. When diluted samples of suspensions of these colonies were spread onto plates with iron supplementation, mixed populations of single-cell colonies of different colour intensities were formed. However, not all colonies were homogeneously coloured or white. There were also colonies containing darker or fainter sectors, indicating that the segregation was reversible.

#### *The strength of antagonism is also diverse*

As the ability to produce pulcherrimin correlates with the ability to antagonise other microorganisms<sup>12</sup>, the isolates were also tested for antifungal activity with three “indicator” organisms on two types of media that were not supplemented with  $\text{FeCl}_3$ . All isolates proved to be antagonistic against *Zygosaccharomyces* and *Botrytis* and most of them could also inhibit the growth of *Debaryomyces* but less efficiently (Supplementary Table S3; Fig. 5). Remarkably, no clear correlation was seen between the presence of pigmented halos on media supplemented with iron and the antagonistic efficiency on media with no iron supplementation.



**Figure 4.** Pulcherrimin production. Colony morphology on YEA supplemented with 0.04 mg/ml  $\text{FeCl}_3$ . Colonies formed from dropped suspensions of the isolates B19b (a), T16a (b) and T12a (c). A colony formed from cells of B1 smeared on the medium (d). Colonies formed by cells of D spread on the medium (e and f). A segregating colony of B19 (g).



**Figure 5.** Antifungal antagonism. Inhibition zones around four isolates in the YOI (Yeast on indicator) test. **(a)** Isolates Q14 (left) and T67 (right) on *Zygosaccharomyces* lawn. **(b)** Isolate B1 on *Debaryomyces* lawn. **(c)** Isolate B19b on a lawn of *Botrytis* conidia.

*All isolates can switch to invasive pseudohyphal growth but vary in the efficiency, morphology and stability of this trait*

As substrate colonisation through the formation of invasive pseudohyphae was observed in certain *M. pulcherrima* strains in previous studies<sup>8</sup>, the isolates were also tested for the ability to grow into the agar medium. All isolates were able to penetrate the medium but with highly diverse efficiency and morphology (Supplementary Table 3S; Fig. 6; Supplementary Fig. S2). Invasive growth was mainly visible under the edges of the colonies and under the circumferential “extension belt” formed by cell propagation at the colony edges (Fig. 6a). Under the edges usually dense spindle-like bodies were formed from which pseudohyphae emanated (Fig. 6c). Under the extension belts, detached patches of looser pseudomycelium were formed (Fig. 6d). The isolates varied in the size, density and morphology of both types of intrusion. Morphological variation was frequently seen even within the colonies. Different patches formed sectors (Fig. 6a, Supplementary Fig. S2) similar to the differently coloured sectors observed in the pulcherrimin tests. However, the differently coloured sectors did not correlate with the differently invasive sectors (Fig. 6f and g). The structure of the invasive pseudomycelium could not be examined microscopically because of the blurring effect of the medium. For microscopic examination the sandwich method was used. The formation and extension of the pseudohyphae was monitored in thin agar layers sandwiched between microscopic slides and cover slips, where the oxygen supply was limited as in the medium during invasive growth. These pseudohyphae exhibited the standard morphology of rudimentary pseudohyphae: elongated unipolarly budding cells connected by uncleaved septa between the mother and daughter cells (Fig. 6d and e).

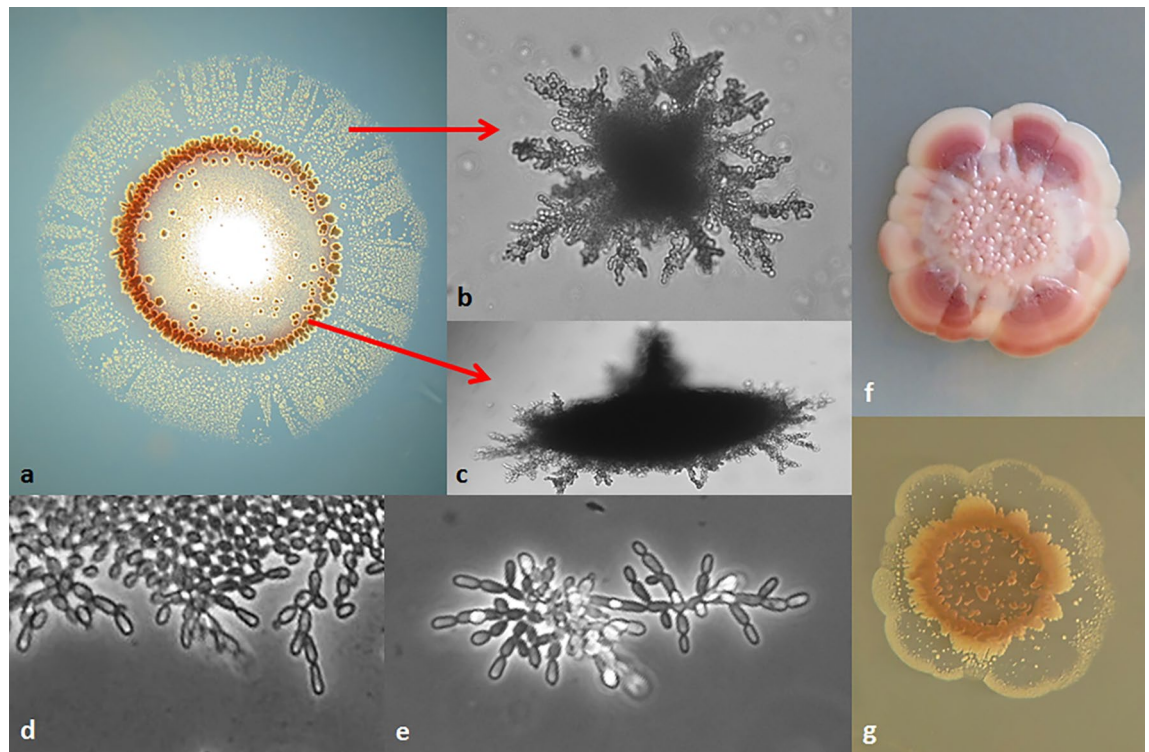
*Chlamydospore formation is less diverse*

A characteristic feature of *M. pulcherrima* is the production of chlamydospores, large globose cells containing large lipid globules<sup>8,22,23</sup>. A recent study observed a positive correlation between the relative number of chlamydospores in the culture and the intensity of pulcherrimin production<sup>24</sup>. Prompted by this observation, the efficiency of chlamydospore formation and the colour intensity of the colonies of the isolates were compared. The percentage of chlamydospores varied in the range of 0–25% (Supplementary Table S3, Supplementary Fig. S3), but correlated neither with pigmentation intensity nor with antagonistic efficacy.

## Discussion

Amplicon sequencing identified very high numbers of (predominantly dimorphic) ambiguous nucleotide positions in the primary barcode segments (D1/D2 and ITS1-5.8S-ITS2) of all new isolates. Previous studies revealed similar nucleotide ambiguity in the type strains of the pulcherrimin-producing species recently merged in *M. pulcherrima*<sup>5,6</sup>. In addition to their internal diversity the new isolates and the type strains of the former species also differed from each other in the number and location of the ambiguities. These results demonstrate





**Figure 6.** Substrate invasion. (a) Transilluminated image of the invasive growth (intrusions) of a colony of the isolate A59 after washing off the cells from the surface of the agar medium. (b and c) Two morphological types of the intrusions. (d) Formation of pseudohyphae at the edge of a culture growing in a sandwich culture. (e) Structure of a pseudohypha formed in a sandwich culture. (f) Sectors of various pigmentation intensity on the surface and (g) invasive growth into the medium in a colony of the isolate B58.

that very high intragenomic and intergenomic variability is characteristic of the repeats coding for rRNA in *M. pulcherrima*, much higher than in any other yeast species for which data are available.

The distribution of the ambiguous sites differed in the two barcodes. Whereas in D1/D2 the SND sites occurred in four short segments, they were randomly scattered across ITS1. This difference can be attributed to different roles of the transcripts in the generation of rRNA for the ribosomes. The former is part of the large subunit rRNA whose correct folding is essential for the ribosome. Consistent with previous observations made in the type strains of the former species<sup>5</sup>, SNDs were found in the D1 and D2 hairpin loops (stem loops) of the reconstructed secondary RNA structures. From these specific locations, it can be inferred that nucleotide substitutions are not allowed in segments important for correct RNA folding but can be tolerated in structurally less critical (double-helical) segments. The double helical segments are less sensitive to substitutions because of the neutralising effect of wobble pairing with the nucleotides with the other strand of the loop<sup>5</sup>. Because of the concentrated occurrence of the SND sites, large parts of the D1/D2 barcodes are free of ambiguity and thus suitable for the determination of the isolates' taxonomic affiliation. The transcripts of the ITS segments are not so structure-sensitive because they are not incorporated in the ribosomes. The extremely high numbers of di- and polymorphic sites and the deletions of various lengths in ITS2 make this barcode unsuitable for taxonomic identification of isolates in this group of organisms. This is another peculiarity of *M. pulcherrima* because it is the ITS region that was proposed as the universal barcode marker for Fungi<sup>25</sup>. It should be noted that the analysis of ITS sequences did not give clear results in numerous other yeast groups either, therefore the D1/D domain has become preferentially used in yeasts<sup>26</sup>. The results of this study are in line with this practice.

Remarkably, the intragenomic barcode variance (the proportion of di- and polymorphic sites) was much greater in all isolates than the gaps separating many ascomycetous species and even certain related genera<sup>27</sup>, as if these isolates had accumulated barcodes from several species. This phenomenon may cause false overestimation of taxon diversity in metabarcoding and metagenomic analysis of environmental samples, in which the delimitation of OTUs (species-like units) is based on (arbitrarily) set similarity thresholds (e.g. 97%). When a strain has barcode segments that are less similar than the threshold, the analysis may consider them representatives of different OTUs (species). The high number of ambiguous nucleotides in barcode sequences obtained by direct amplicon sequencing may cause difficulties in patenting of biotechnological strains. *M. pulcherrima* strains are often employed as biological control agents [BCAs] in the management of crop pathogens or as lipogenic/oleaginous yeasts<sup>4</sup>.

Obviously, the *M. pulcherrima* repeats coding for rRNAs evolve in a non-standard way. In a previous study the intragenomic diversity of the primary barcodes of a *M. fructicola* (synonym of *M. pulcherrima*) strain was proposed to be due to the evolution of the repeats by a birth-and-death-type mechanism instead

of homogenisation operating in other yeasts<sup>7</sup>. In yeasts, a similar mode of repeat evolution was only observed in *Cyberlindnera* but with much lower intragenomic repeat diversity<sup>28</sup>. In an earlier study<sup>29</sup>, different D1/D2 sequences were detected in certain *Clavispora lusitanae* strains. Their presence within the same genome may be due to birth-and-death evolution, but this possibility was not considered.

The analysis of 33 protein-encoding nuclear genes (including secondary barcode segments) performed in recent studies<sup>3,6</sup> revealed chimeric genome structures in the type strains of the merged species. The analysed genes usually had two or more copies that differed in sequence and phylogenetic history. The four genes of the *PUL* cluster were among them. Most strains involved in that analysis had two sets of *PUL* genes. The sets often differed from each other more than from one or the other set of a different strain (higher intragenomic than intergenomic differences). Here *PUL4*, as the representative of the *PUL* cluster was used to get an insight into the genome structures of the isolates. The sequences of the amplified *PUL4* segments had multiple dimorphic sites in all isolates, demonstrating that these isolates also had at least two different *PUL4* genes. Thus, the intragenomic sequence diversity is not confined to the primary barcodes. Most SNPs occurred in the third positions of codons and had no effect on the amino acid sequences of the proteins. The nucleotide substitutions in the first and second codon positions caused amino acid substitutions in altogether 12% of the codons. It can be inferred from these results that the different *PUL4* genes within a genome are orthologues of different origin that were brought together by hybridisation. The *PUL4* sequences (including the ambiguous sites) also showed considerable interstrain (intergenomic) variations. The interstrain differences indicate that the strains of the species may have chimeric genome structures. The RAPD patterns covering entire genomes and the RFLP patterns of the mitochondrial genomes also revealed high interstrain differences which also proves that *M. pulcherrima* is a very heterogeneous species.

To examine whether the genomic/genetic differences are reflected in the phenotypes of the isolates, three properties relevant for the application of *M. pulcherrima* strains as bioprotective agents were examined: iron-chelating efficiency (intensity of pulcherrimin production), antifungal antagonism (inhibition of tester indicator organisms) and substrate invasion (transition to invasive pseudohyphal growth). The former two properties are assumed to be linked because the synthesis of pulcherriminic acid is involved in both. The higher the pulcherrimin production, the stronger the antimicrobial activity<sup>12</sup>. Considerable diversity was found in both of them, but the results of the antagonism tests carried out on different culture media and with different indicator organisms were not completely congruent. The lower inhibition efficiency against *Debaryomyces* can be attributed to the higher growth vigour (faster growth) of this yeast. The more efficient inhibition of *Botrytis* can be due to the higher vulnerability of the germination of conidia compared to that of the budding of yeast cells.

The efficiency of the yeast-to-pseudohypha transition and the morphology of substrate invasion by the pseudomycelium also greatly varied among the isolates. Switching from the planktonic yeast growth to the formation of chains of non-separated cells (pseudohyphae) may play an important role in the protection of the plant surface because the pseudomycelium formed by the invasive pseudohyphae can form biofilms on the lesions which are also gateways for the invasion by destructive microorganisms. Previous antagonism studies tested the ability of *Metschnikowia* (planktonic) yeast cells for adhesion to non-plant surfaces such as glass and polypropylene<sup>21</sup>. The ability of the tested strains to form pseudohyphae was not examined, although a pseudomycelium is a stronger structure than a layer of planktonic yeast cells, especially if its pseudohyphae establish a strong bond with the damaged plant tissue by penetrating it.

In addition to these three traits, chlamydospore formation, a trait not known to have relevance to antagonism, was also examined. The isolates also varied in the efficiency of chlamydospore formation but the overall diversity of this trait was lower.

When the molecular differences were compared with the phenotypic differences, no clear correlation between the examined genetic/genomic diversities and the phenotypic diversities was noticed. Thus, none of the molecular tests employed in this study can be used for differentiation of strains showing different phenotypes.

Interestingly, pulcherrimin production and pseudohyphal growth also varied within clones (single-cell colonies) of numerous isolates. Intraclonal changes (segregation) of the intensity of pulcherrimin production in *M. pulcherrima* were already reported in previous studies and found reversible<sup>6</sup>. The cultures of numerous new isolates described in this study also formed sectors differing in colour intensity and mixtures of differently coloured single-cell colonies. The colonies often had sectors differing in colour from the rest of the colony. Thus, the changes (segregation) of the intensity of pulcherrimin production were reversible in these isolates as well. These changes may be due to epigenetic processes (silencing and reactivation of regulators) rather than to mutations and backmutations. The intraclonal heterogeneity of pseudohyphal growth may be attributed to similar processes. However, it has to be mentioned that in *Saccharomyces cerevisiae*, whose epigenetic processes have been most deeply explored among yeasts, certain important processes such as DNA methylation, repressive histone methylation, and siRNAs may be missing<sup>30</sup>. Experiments are already underway to learn more about the reversible transitions in *M. pulcherrima* strains and explore the underlying mechanisms.

## Methods

### Sample collection and yeast isolation

Overripe and fallen fruits were collected from abandoned roadside and orchard trees or purchased from greengrocers. Decaying tissue blocks were cut from them for yeast isolation. The blocks were homogenised in sterile water and loopful amounts of the homogenised material were spread onto YEA (Yeast Extract Agar: 0.5% yeast extract, 2% glucose, 2% agar) plates supplemented with 0.04 mg/ml FeCl<sub>3</sub>. Colonies producing pulcherrimin turn red on this medium. After 7 days of incubation at 25 °C, 1 to 2 red-brownish colonies were isolated from each sample (Supplementary Table S1) and inoculated onto YEA plates for maintenance.

### Barcode amplification and sequencing

DNA was isolated from overnight cultures grown in YEA broth (YEA without agar) and the isolated DNA was used for amplification of chromosomal segments. The D1/D2 domains and the ITS1-5.8S-ITS2 segments of the repeats encoding rRNA were amplified with the primer pairs NL1-NL4<sup>31</sup> and ITS1-ITS4<sup>32</sup>, respectively. The amplified fragments were sequenced in both directions using the same primer pairs. The D1-D2 sequences were deposited in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) under accession numbers listed in Supplementary Table S1. The ITS sequences could not be deposited in the database because of the very high number of ambiguous nucleotides.

### Amplification and sequencing of *PUL4*

*PUL4* is one of the four genes of the *PUL* cluster. It codes for a putative transcription factor belonging to a subclass of zinc finger proteins found exclusively in fungi. The Pul4 proteins contain a Fungal\_TF\_MHR region, a fungal transcription factor regulatory middle homology region<sup>33</sup>. A 380 nt-long segment of this region was amplified from the DNA prepared for barcode sequencing with the primer pair PUL4F (742–767: CCCAGCATGTTGAAA GCTCCGTCCTCG) and PUL4R (1115–1144: AGTAATTGGTAGTTTATTTCTCGGACTTG). The amplicons were sequenced with the same primers and the sequences were deposited in GenBank. The accession numbers are listed in Supplementary Table S1.

### Sequence analysis

To identify variable positions, the nucleotide sequences were aligned with the MAFFT version 7 multiple alignment algorithm available at <https://www.ebi.ac.uk/jdispatcher/msa/mafft><sup>34</sup>. The *PUL4* sequences were also translated using the NCBI ORF finder (<https://www.ncbi.nlm.nih.gov/orffinder/>) and the amino acid sequences of the Pul4 proteins of the isolates and those extracted from the whole-genome sequences of *M. pulcherrima* strains available in the NCBI genome database (<https://www.ncbi.nlm.nih.gov/genome/?term=>) were aligned with the same MAFFT tool. Minimum free energy secondary structures were generated for the D1 and D2 loops of the transcripts of the D1/D2 domains with Predict, a secondary structure webserver (<http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>) using default settings.

### RAPD-PCR

Samples of the solutions of genomic DNA (50 ng DNA) were used for RAPD-PCR reactions with the primers RAPD24 (5'-GCGTGACTTG-3')<sup>35</sup> and RAPD1283 (5'-GCG ATC CCC A-3')<sup>36</sup> and GoTaq DNA polymerase. The PCR reactions were performed with the following programme: 94 °C for 5 min, 30x (94 °C for 50 s, 38 °C for 50 s, 72 °C for 50c), 72 °C for 5 min. The amplified fragments were separated in 1.2% (w/v) agarose gels (1× TBE buffer) stained with ethidium-bromide. The band patterns were analysed using the GelAnalyzer 2010 software<sup>37</sup> with manual adjusting and the results were converted to binary sets of data. The data sets were subjected to cluster analysis by the UPGMA (unweighted pair group method with arithmetic means) algorithm using the version available at <http://genomes.urv.es/UPGMA><sup>38</sup>.

### Mitochondrial DNA extraction and restriction analysis

Mitochondrial DNA was prepared from exponential-phase YEA cultures with the method described in<sup>39</sup> and digested with *Mbo*I. The fragments were separated by electrophoresis in 0.7% agarose, 0.5× TBE. The band patterns were analysed as described for RAPD-PCR.

### Examination of pulcherrimin production

Pulcherrimin production was examined on YEA plates supplemented with 0.02 mg/ml and 0.04 mg/ml FeCl<sub>3</sub>. Colonies producing pulcherrimin turn red on these media. The higher the concentration of FeCl<sub>3</sub>, the more intense the pigmentation. When pulcherrimin is also produced extracellularly, a coloured halo is formed around the pigmented colony. The isolates were inoculated on these plates in two ways, either by streaking (smearing) cells with a loop or by dropping 10 µl of dense suspension of cells (10<sup>7</sup> cells/ml). Growth and pigment formation were monitored at 25 °C for four weeks. With this method the stability of pulcherrimin production could also be tested because the colonies of the unstable isolates formed sectors of different colour intensities at their edges. Stability was also tested by spreading cells on an iron-containing medium and monitoring the formation of sectors within colonies growing from single cells.

### Examination of chlamydospore formation

This trait was examined in cultures grown on vegetable juice medium prepared from diluted (1:20) BIO Gemüsesaft (Josef Pölz, Bio-Produkte, 84,518 Garching an der Alz, Germany) and supplemented with 2% agar<sup>7</sup>. 10-µl aliquots of dense suspensions of cells (10<sup>7</sup> cells/ml) were dropped onto the medium and incubated at 14 °C. After 2 weeks of incubation, samples of the cultures were examined microscopically.

### Invasive growth and formation of pseudohyphae

To test the isolates for the ability to penetrate the agar medium, 10-µl samples of dense cell suspensions (10<sup>7</sup> cells/ml) were dropped onto YEA. After two weeks of incubation at 25 °C, the colonies were washed off from the plate with water to visualise the intrusions into the agar medium beneath them. The morphological structure of the intrusions was examined microscopically at low magnification. As the microscopic images were blur because of the distorting effect of the agar medium, the transition from the yeast growth to the invasive filamentous growth

was also examined in thin films (0.5–1 mm) of YEA sandwiched between microscope slides and coverslips as described previously<sup>40</sup>.

### Antagonism tests

The isolates were tested for antimicrobial antagonistic activity by the “Yeast on indicator (YOI)” method<sup>41</sup>. Briefly, the agar test plate was flooded with a dense suspension ( $10^7$  cells/ml) of cells or conidia of the indicator organism to obtain a homogeneous lawn on the surface of the medium. After removing the rest of the suspension and drying the surface of the plate, loopful amounts of the *Metschnikowia* isolates were smeared on the lawn to form spots of ~5 mm in diameter. The growth of the spots and the lawn around them was monitored at 20 °C for 7 days. If the isolate inhibited or reduced the growth of the lawn, a clear (or only turbid) inhibition zone appeared in the lawn around its colony. The tests were carried out on YEA and EMMA (Edinburgh Minimal Medium,<sup>42</sup>) plates and the indicator organisms were *Debaryomyces* aff. *hanseni* Vm<sup>743</sup>, *Zygosaccharomyces* aff. *siamensis* 11-2106 and *Botrytis cinerea* 3318<sup>12</sup>.

### Data availability

The datasets analysed during the current study are included in this published article [and its supplementary information files]. The D1/D2 and *PUL4* sequences are available in GenBank. The GenBank accession numbers of are listed in Supplementary Table S1.

Received: 20 March 2024; Accepted: 4 May 2024

Published online: 08 May 2024

### References

- Kurtzman, C. P., Fell, J. W. & Boekhout, T. Definition, classification and nomenclature of the yeasts. In *The Yeasts—A Taxonomic Study* (eds Kurtzman, C. P. et al.) 3–5 (Elsevier, 2011).
- Herbert, P. D., Cywinska, A., Ball, S. L. & de Waard, J. R. Biological identifications through DNA barcodes. *Proc. Biol. Sci.* **270**, 313–321 (2003).
- Kurtzman, C. P. & Robnett, C. J. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie Van Leeuwenhoek* **73**, 331–371 (1998).
- Elder, J. F. & Turner, B. J. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* **70**, 297–320 (1995).
- Sipiczki, M. When barcoding fails: Genome chimerization (admixing) and reticulation obscure phylogenetic and taxonomic relationships. *Mol. Ecol. Resour.* **22**, 1762–1785 (2022).
- Sipiczki, M. Taxonomic revision of the pulcherrima clade of *Metschnikowia* (Fungi): Merger of species. *Taxonomy* **2**, 107–123 (2022).
- Sipiczki, M., Horvath, E. & Pfliegler, W. P. Birth-and-death evolution and reticulation of ITS segments of *Metschnikowia andauensis* and *Metschnikowia fructicola* rDNA repeats. *Front. Microbiol.* **9**, 1193 (2018).
- Sipiczki, M. *Metschnikowia pulcherrima* and related pulcherrimin-producing yeasts: Fuzzy species boundaries and complex antimicrobial antagonism. *Microorganisms* **8**, 1029 (2020).
- Gil-Rodríguez, A. M. & Garcia-Gutierrez, E. Antimicrobial mechanisms and applications of yeasts. *Adv. Appl. Microbiol.* **114**, 37–72 (2021).
- Ma, Y., Wu, M., Qin, X., Dong, Q. & Li, Z. Antimicrobial function of yeast against pathogenic and spoilage microorganisms via either antagonism or encapsulation: A review. *Food Microbiol.* **112**, 104242 (2023).
- Lachance, M. A. Chapter 35—*Kluyveromyces van der Walt* (1971). In *The Yeasts—A Taxonomic Study* (eds Kurtzman, C. P. et al.) 471–481 (Elsevier, 2011).
- Sipiczki, M. *Metschnikowia* strains isolated from botrytized grapes antagonize fungal and bacterial growth by iron depletion. *Appl. Environ. Microbiol.* **72**, 6716–6724 (2006).
- Krause, D. J. et al. Functional and evolutionary characterization of a secondary metabolite gene cluster in budding yeasts. *Proc. Natl. Acad. Sci. USA* **115**, 11030–11035 (2018).
- Cook, A. H. & Slater, C. A. The structure of pulcherrimin. *J. Chem. Soc.* 4133–4135 (1956).
- MacDonald, J. C. Biosynthesis of pulcherriminic acid. *Biochem. J.* **96**, 533–538 (1965).
- Kluyver, A. J., van der Walt, J. P. & van Triet, A. J. Pulcherrimin, the pigment of *Candida pulcherrima*. *Proc. Natl. Acad. Sci. USA* **39**, 583–593 (1953).
- MacDonald, J. C. The structure of pulcherriminic acid. *Can. J. Chem.* **41**, 165–172 (1963).
- Yuan, S., Yong, X., Zhao, T., Li, Y. & Liu, J. Research progress of the biosynthesis of natural bio-antibacterial agent pulcherriminic acid in *Bacillus*. *Molecules* **25**, 5611 (2020).
- Charron-Lamoureux, V. et al. Pulcherriminic acid modulates iron availability and protects against oxidative stress during microbial interactions. *Nat. Commun.* **14**, 2536 (2023).
- Türkel, S. & Ener, B. Isolation and characterization of new *Metschnikowia pulcherrima* strains as producers of the antimicrobial pigment pulcherrimin. *Z. Naturforsch. C. J. Biosci.* **64**, 405–410 (2009).
- Pawlikowska, E., James, S. A., Breierova, E., Antolak, H. & Kregiel, D. Biocontrol capability of local *Metschnikowia* sp. isolates. *Antonie Van Leeuwenhoek* **112**, 1425–1445 (2019).
- Tatay-Núñez, J. et al. Isolation of local strains of the yeast *Metschnikowia* for biocontrol and lipid production purposes. *World J. Microbiol. Biotechnol.* **40**, 88 (2024).
- Lachance, M. A. *Metschnikowia Kamienski*. In *The Yeasts—A Taxonomic Study* (eds Kurtzman, C. P. et al.) 575–620 (Elsevier, 2011).
- Melvydas, V., Svediene, J., Skridlaite, G., Vaiciuniene, J. & Garjonyte, R. In vitro inhibition of *Saccharomyces cerevisiae* growth by *Metschnikowia* spp. triggered by fast removal of iron via two ways. *Braz. J. Microbiol.* **51**, 1953–1964 (2020).
- Schoch, C. L. et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. USA* **109**, 6241–6246 (2012).
- Kurtzman, C. P., Fell, J. W. & Boekhout, T. Gene sequence analyses and other DNA-based methods for yeast species recognition. In *The Yeasts—A Taxonomic Study* (eds Kurtzman, C. P. et al.) 137–144 (Elsevier, 2011).
- Vu, D. et al. DNA barcoding analysis of more than 9000 yeast isolates contributes to quantitative thresholds for yeast species and genera delimitation. *Stud. Mycol.* **85**, 91–105 (2016).
- Bizarria, R., Pietrobon, T. C., Ferreira, H. & Rodrigues, A. Polymorphisms of rDNA genes in *Cyberlindnera* yeast suggest birth-and-death evolution events. *FEMS Yeast Res.* **23**, foad032 (2023).

29. Lachance, M. A. *et al.* The D1/D2 domain of the large-subunit rDNA of the yeast species *Clavispora lusitaniae* is unusually polymorphic. *FEMS Yeast Res.* **4**, 253–258 (2003).
30. Fuchs, S. M. & Quasem, I. Budding yeast as a model to study epigenetics. *Drug Discov. Today Dis. Models* **12**, 1–6 (2014).
31. O'Donnell, K. Fusarium and its near relatives. In *The Fungal Holomorph: Mitotic, Meiotic and Pleomorphic Speciation in Fungal Systematics* (eds Reynolds, D. R. & Taylor, J. W.) 225–233 (CAB International, 1993).
32. White, T. J., Bruns, T., Lee, S. & Taylor, J. Amplification and sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols A Guide to Methods and Applications* (eds Innis, M. A. *et al.*) 315–322 (Academic Press, 1990).
33. Todd, R. B. & Andrianopoulos, A. Evolution of a fungal regulatory gene family: The Zn(II)2Cys6 binuclear cluster DNA binding motif. *Fungal Genet. Biol.* **21**, 388–405 (1997).
34. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**(W1), W276–W279 (2022).
35. Baleiras Couto, M. M., Eijmsa, B., Hofstra, H., Huis in't Veld, J. H. & Van der Vossen, J. M. Evaluation of molecular typing techniques to assign genetic diversity among *Saccharomyces cerevisiae* strains. *Appl. Environ. Microbiol.* **62**, 41–46 (1996).
36. Akopyanz, N., Bukanov, N. O., Westblom, T. U., Kresovich, S. & Berg, D. E. DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. *Nucleic Acids Res.* **20**, 5137–5142 (1992).
37. Lazar, I. & Lazar, I. Gel Analyzer 2010a: Freeware 1D Gel Electrophoresis Image Analysis Software. <http://www.gelalyzer.com> (2010).
38. Garcia-Vallve, S., Palau, J. & Romeu, A. Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. *Mol. Biol. Evol.* **16**, 1125–1134 (1999).
39. Nguyen, H. V., Lepingle, A. & Gaillardin, C. A. Molecular typing demonstrates homogeneity of *Saccharomyces uvarum* strains and reveals the existence of hybrids between *S. uvarum* and *S. cerevisiae*, including the *S. bayanus* type strain CBS 380. *Syst. Appl. Microbiol.* **23**, 71–85 (2000).
40. Sipiczki, M. Dimorphic cycle in *Candida citri* sp. nov., a novel yeast species isolated from rotting fruit in Borneo. *FEMS Yeast Res.* **11**, 202–208 (2011).
41. Sipiczki, M. Identification of antagonistic yeasts as potential biocontrol agents: Diverse criteria and strategies. *Int. J. Food Microbiol.* **406**, 110360 (2023).
42. Mitchison, J. M. Physiological and cytological methods for *Schizosaccharomyces pombe*. *Methods Cell Physiol.* **4**, 131–165 (1970).
43. Sipiczki, M. & Hrabovszki, V. *Galactomyces candidus* diversity in the complex mycobiota of cow-milk bryndza cheese comprising antagonistic and sensitive strains. *Int. J. Food Microbiol.* **388**, 110088 (2023).

## Acknowledgements

The authors are grateful to Ilona Kiss for expert technical assistance.

## Author contributions

MS designed the research; KC conducted the RAPD and RFLP experiments and *PUL4* sequencing; MS sequenced the primary barcodes and analysed all experimental data; ZK participated in yeast isolation; MS wrote the paper.

## Funding

Open access funding provided by University of Debrecen. This work was supported by the National Research, Development and Innovation Fund Office, Hungary (2020-1.1.2-PIACI-KFI-2020-00130) and by the University of Debrecen Programme for Scientific Publication.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-61335-5>.

**Correspondence** and requests for materials should be addressed to M.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024