Supplementary Material

Snow- and ice-ecosystem cleaning capability of the pucciniomycotinous yeast *Phenoliferia psychrophenolica*

Jade A. Ezzedine; Pierre Guenzi-Tiberi; Gaelle Villain; Riccardo Aiese Cigliano; Yacine Diagne; Enzo Franceschi; Jérôme Forêt; Elodie Drula; Jean-Gabriel Valay; Lenka Procházková; Daniel Remias; Nicolas Terrapon; Alberto Amato; Eric Maréchal

Supplementary Figures

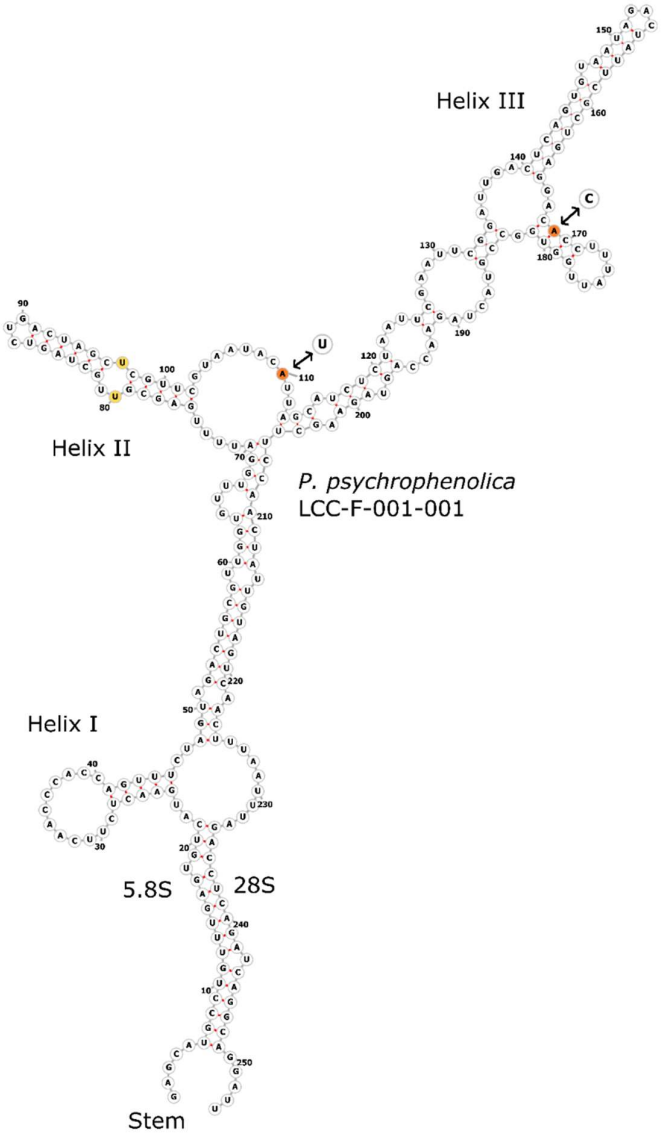


Figure S1: ITS2 secondary structure of *P. psychrophenolica* LCC-F-001-001 compared to the holotype strain *P. psychrophenolica* AG21 (EF151246.1). Three helices can be identified with a complete stem (5.8S – 28S). Structures were folded using RNAfold using Andronescu model. The pyrimidine - pyrimidine mismatch in Helix II was highlighted in yellow. The base differences between the two species are shown in brown.

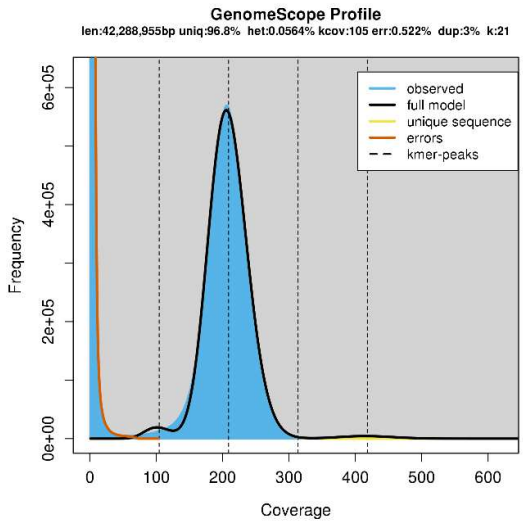
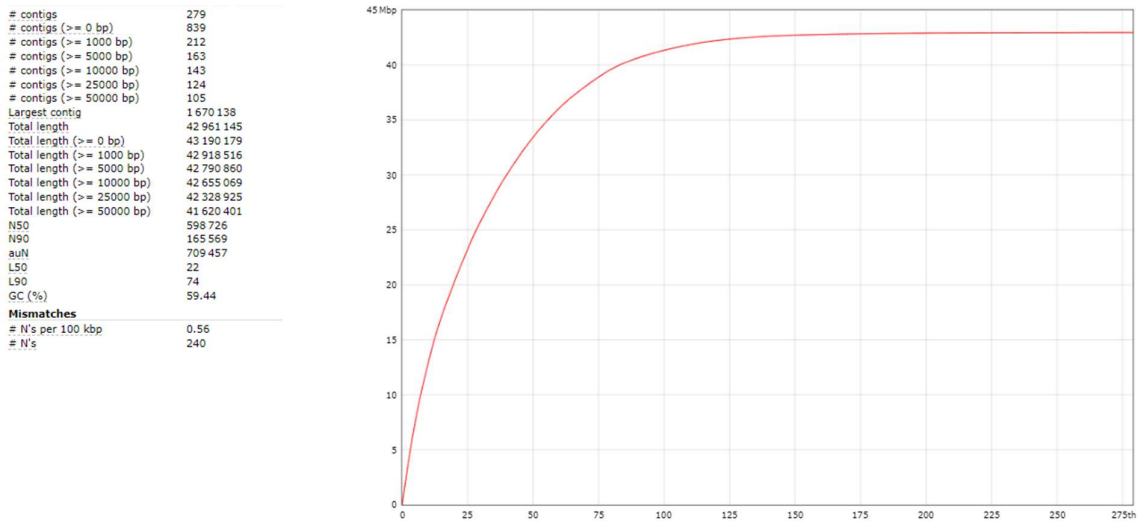


Figure S2: Determination of *P. psychrophenolica* genome size based on GenomeScope. Genomic size following *P. psychrophenolic* sequencing was based on a kmer linear plot after error correction, offering estimations of coverage, genome size and kmer size. In the obtained density plot, the peak gives the sequencing depth used to calculate genome size. The obtained value (42,29 Mb) was corrected as detailed in method, allowing a determination of a genome size of 41.6 Mb. Abbreviations: len, length; uniq, unique kmers; het, heterozygocity; kcov, kmer coverage; err., error; dup., duplications.



# contigs	279
# contigs (>= 0 bp)	839
# contigs (>= 1000 bp)	212
# contigs (>= 5000 bp)	163
# contigs (>= 10000 bp)	143
# contigs (>= 25000 bp)	124
# contigs (>= 50000 bp)	105
Largest contig	1 670 138
Total length	42 961 145
Total length (>= 0 bp)	43 190 179
Total length (>= 1000 bp)	42 918 516
Total length (>= 5000 bp)	42 790 860
Total length (>= 10000 bp)	42 655 069
Total length (>= 25000 bp)	42 328 925
Total length (>= 50000 bp)	41 620 401
N50	598 726
N90	165 569
auN	709 457
L50	22
L90	74
GC (%)	59.44
Mismatches	
# N's per 100 kbp	0.56
# N's	240

Figure S3: Quality assessment of *P. psychrophenlica* genome assembly based on Quast. The cumulative length graph produced by Quast of the *P. psychrophenolica* genome assembly highlights the quality obtained, with 105 contigs ≥ (>=) 50,000 bp, and a largest contig of 1,670,138 bp. Associated statistics are indicated on the left.

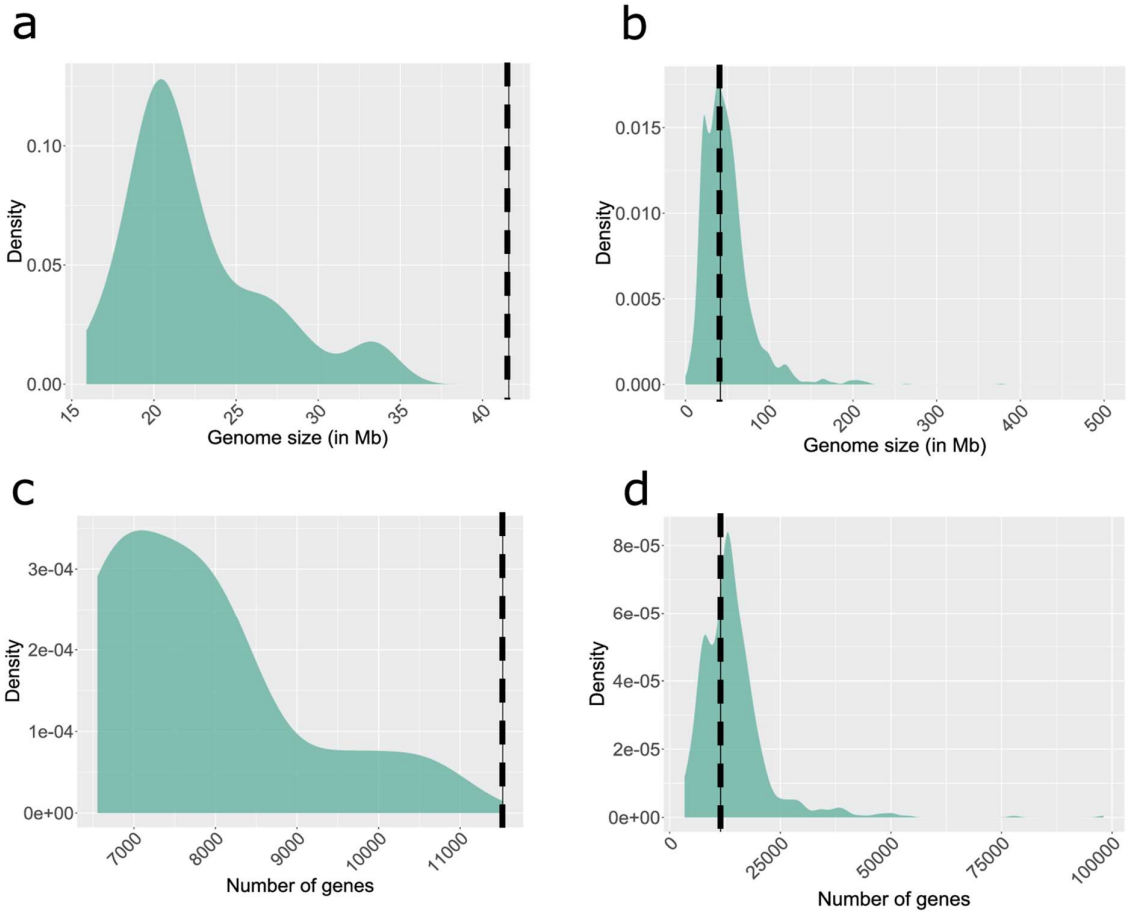


Figure S4: Comparison of the genome size and number of genes of *P. psychrophenolica* LCC-F-001-001 with other sequenced fungi. (**a**) Genome sizes of Microbotryomycetes. (**b**) Genome sizes of Basidiomycota. (**c**) Number of genes of Microbotryomycetes. (**d**) Number of genes of Basidiomycota. The black line represents the value corresponding to *P. psychrophenolica* LCC-F-001-001.

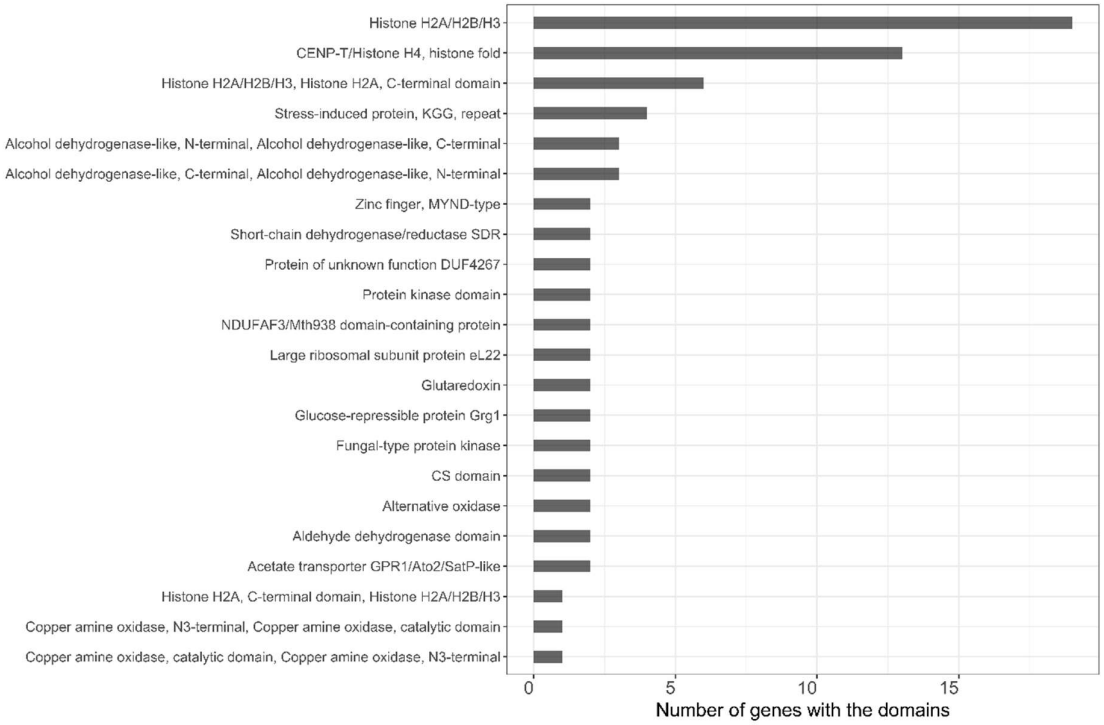


Figure S5: Distribution of the Interpro domain terms associated to the highly-similar duplicated (HSD) genes in *P. psychrophenolica* LCC-F-001-001. The histone H2A/H2B/H3 and the CENP-T/Histone H4 are the most duplicated genes. Alcohol and aldehyde dehydrogenases are involved in multiple metabolic pathways, including carotenoid breakdown.

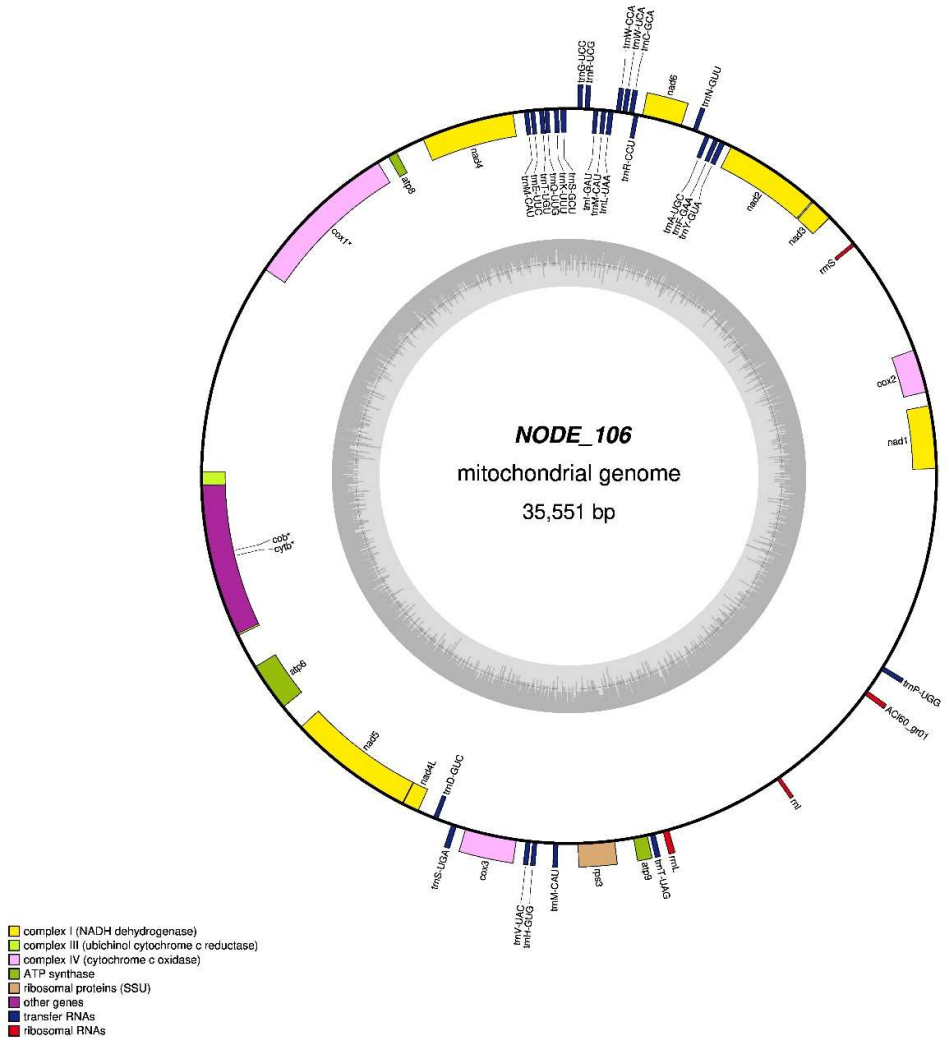


Figure S6: Annotation map of the *P. psychropheolica* LCC-F-001-001 mitochondrial genome.

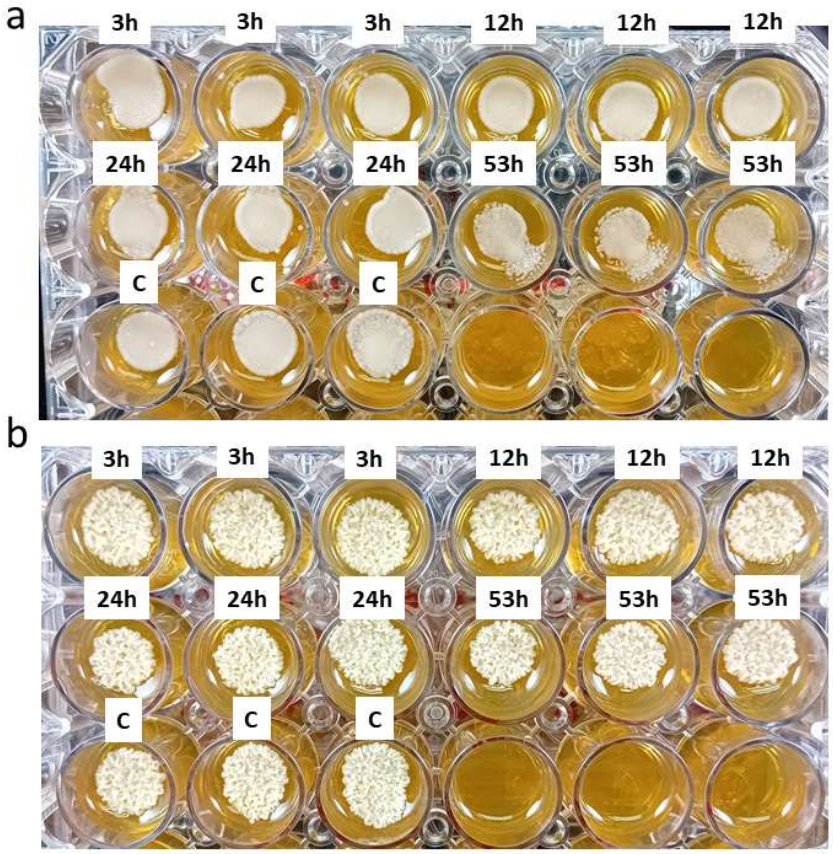


Figure S7: Freezing resistance of *P. psychrophenolica* LCC-F-001-001 acclimated at 5°C and 20°C, and kept frozen at -20°C for 3, 12, 24, and 53 hours. After freezing, fungal cells were incubated at their corresponding acclimation temperatures, 5°C (**a)** and 20°C (**b**). The observation of colonies was achieved 10 days after freezing initiation.

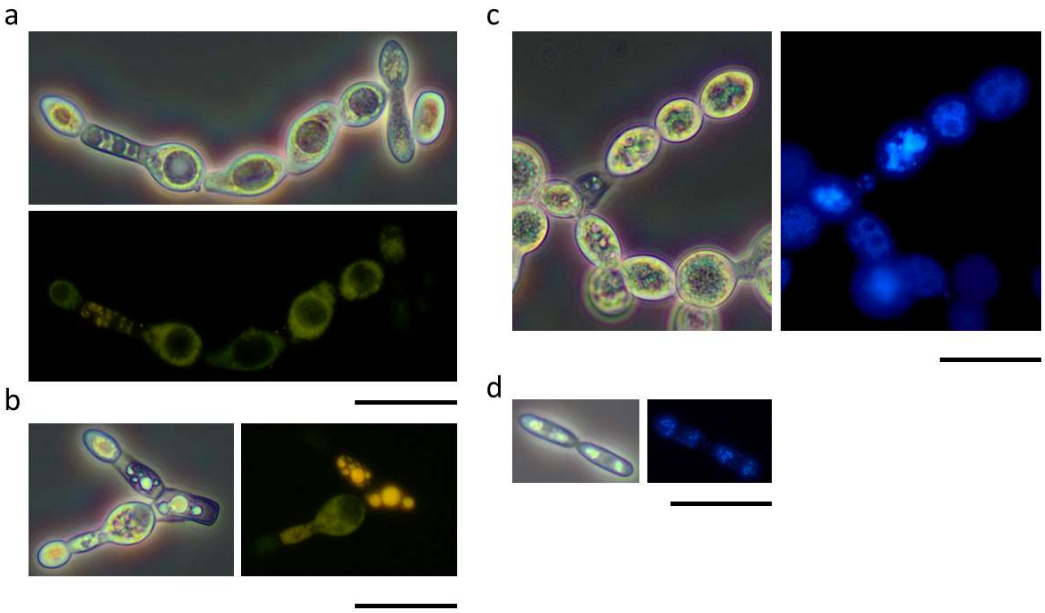


Figure S8: Detection of cytosolic lipid droplets involved in triacylglycerol carbon storage inside *P. psychrophenolica* cells. Fluorescence staining of lipid droplets using either Nile Red at 5°C (**a**) and 20°C (**b**) or Autodot (Monodansylpentane, MDH) fluorescent probes at 5°C (**c**) and 20°C (**d**) (scale bar = 20 µm).

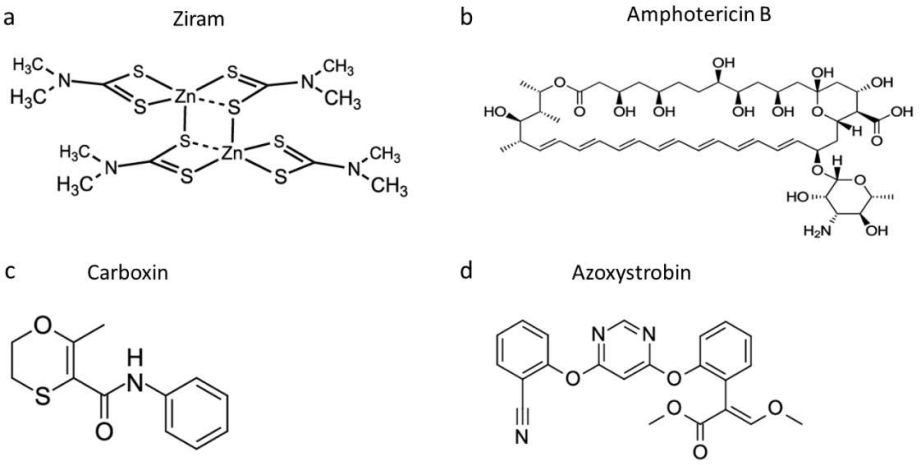


Figure S9: Fungicide chemical structures. **(a)** Ziram, **(b)** Amphotericin B, **(c)** Carboxin and **(d)** Azoxystrobin.

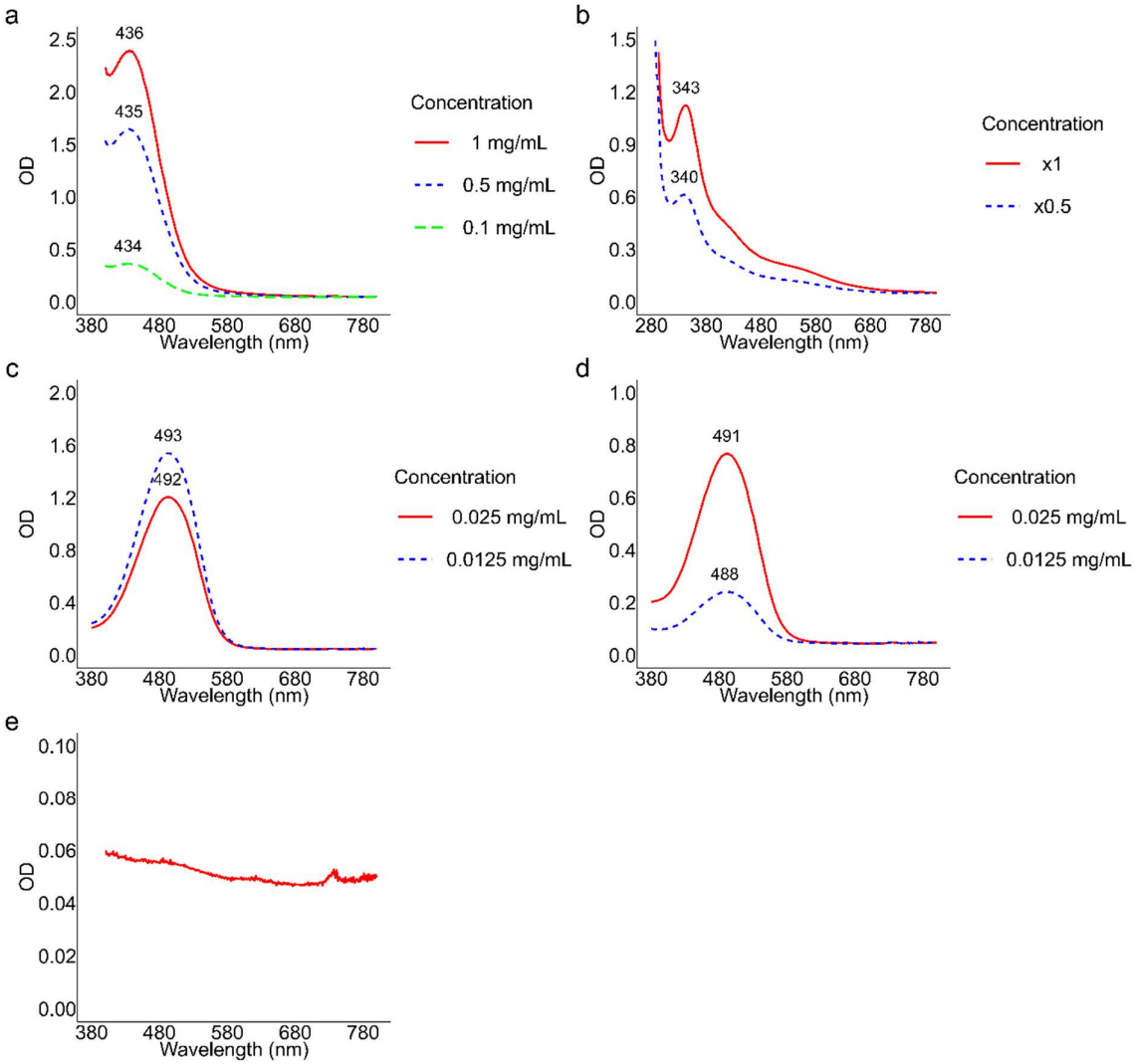


Figure S10: Absorbance spectra of algal pigments. (a) Pure purpurogallin solubilized in DMSO and diluted to 1, 0.5 and 0.1 mg.mL^{-1}. **(b)** Purpurogallin carboxylic acid-6-O-β-D-glucopyranoside extracted from *Ancylonema alaskana*, solubilized and diluted to half in water. **(c)** Pure astaxanthin solubilized in DMSO and diluted to 0.025 and 0.0125 mg.mL^{-1}. **(d)** Astaxanthin acyl ster solubilized in DMSO and diluted to 0.025 and 0.0125 mg.mL^{-1}. **(e)** Pure DSMO.

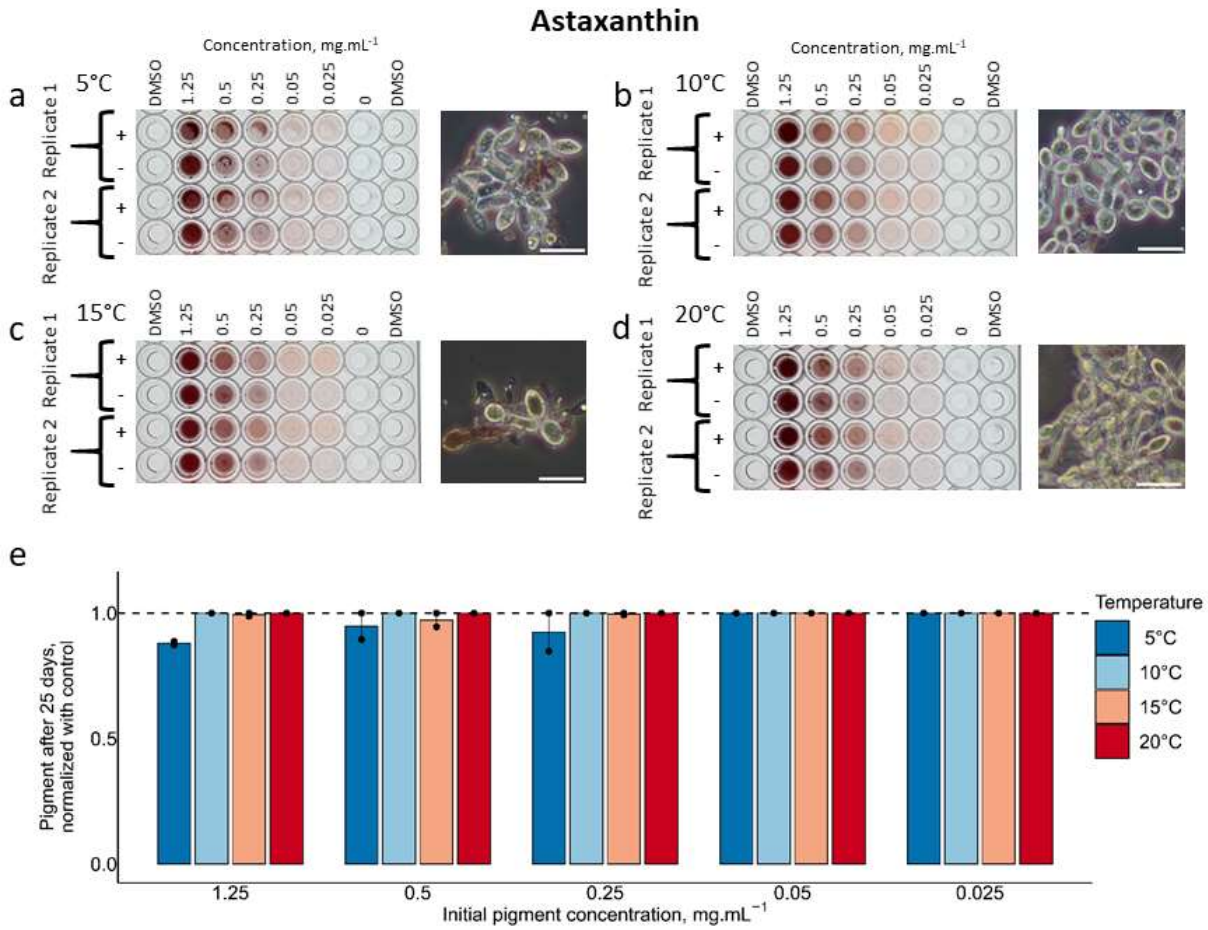


Figure S11: Analysis of *P. psychrophenolica* LCC-F-001-001 capacity to hydrolyze pure astaxanthin. *P. pyschrophenolica* LCC-F-001-001 cell cultures were acclimated at 5, 10, 15 and 20°C and incubated in presence of pigments for 25 days. Pigment concentrations were set initially at 1.25, 0.5, 0.25, 0.05, 0.025 and 0 mg.mL^{-1}, as indicated. DMSO was used as control. A + and a – sign indicate presence or absence of cells, respectively. Microscopic images shown on the right of plates are representative of each temperature condition (scale bar = 20 µm). **(a – d)** Astaxanthin detection after a 25-day incubation. Pigment detection after 25 days is not visibly impacted by the presence of fungal cells. **(e)** OD_{490} decrease in presence of cells.

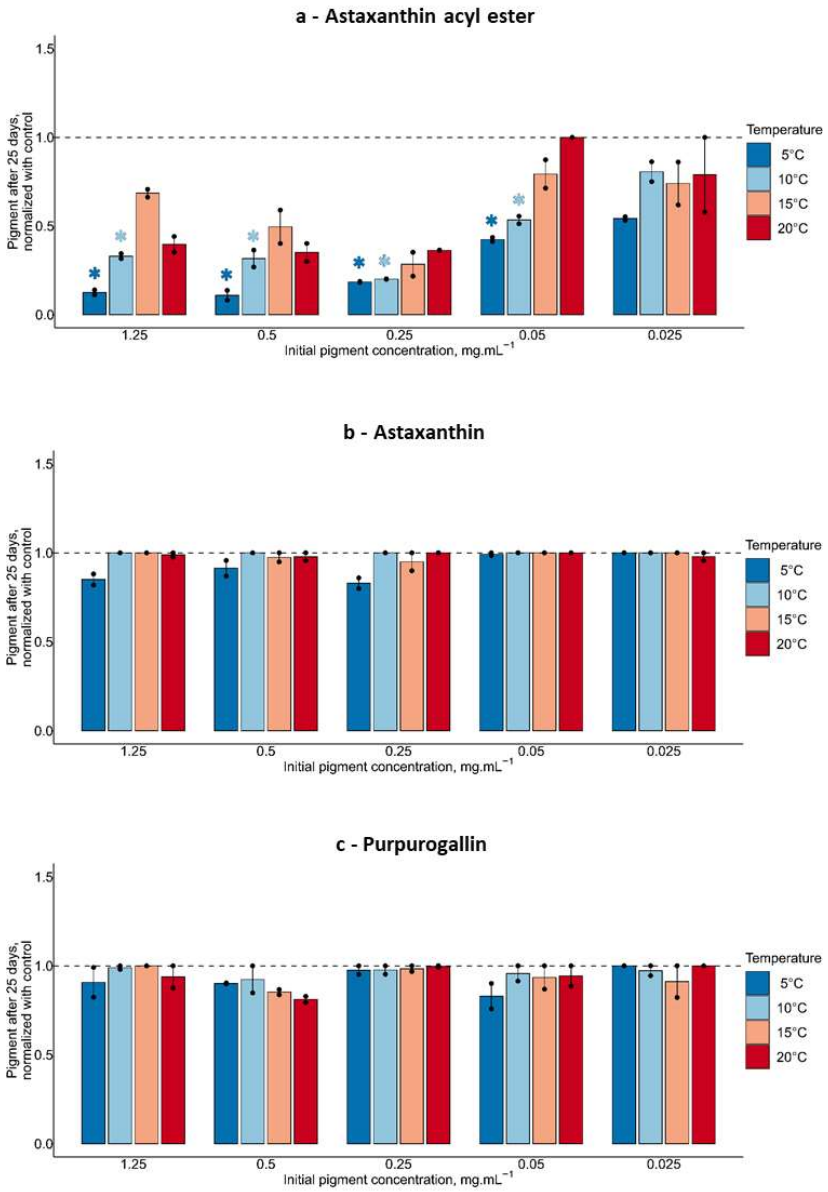


Figure S12: Analysis of *P. psychrophenolica* LCC-F-001-001 capacity to hydrolyze snow and ice algal pigments over 35 days. Experimental conditions are identical to those described in Fig. 4 and Fig S11. OD was measured after 35 days of incubation. Histograms depict the variations of substrates in comparison to the control (substrate with no cells). **(a)** Astaxanthin acyl ester, **(b)** Astaxanthin, and **(c)** Purpurogallin. Asterisks correspond to p-values < 0.1 with t-test.

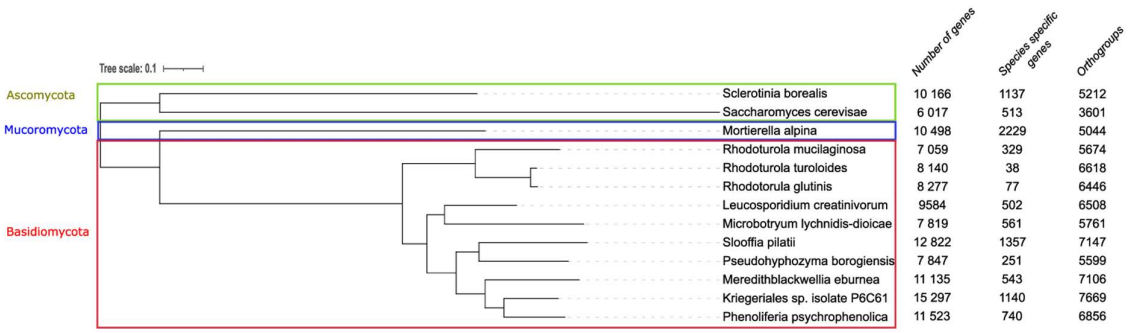


Figure S13: Phylogenomic tree of different pucciniomycetous yeast, with available complete genome data, and Orthofinder statistics. Pucciniomycotina belong to Basidiomycota, and include *Phenoliferia psychrophenolica*, *Kriegeriales* sp. P6C61 (*P. himalayensis*), *Meredithblackwellia eburnean*, *Pseudohyphozyma bogoriensis*, *Slooffia pilatii*, *Microbotryum lichnis-dioicae, Leucosporidium creatinivorum* and *Rhodotorula* spp, used for this phylogenomic tree. *Mortieralla alpina* was used as distant fungus from the Mucoromycota, also occurring in cold environments, whereas *Sclerotinia borealis* and *Saccharomyces cerevisiae* are two Ascomycota.

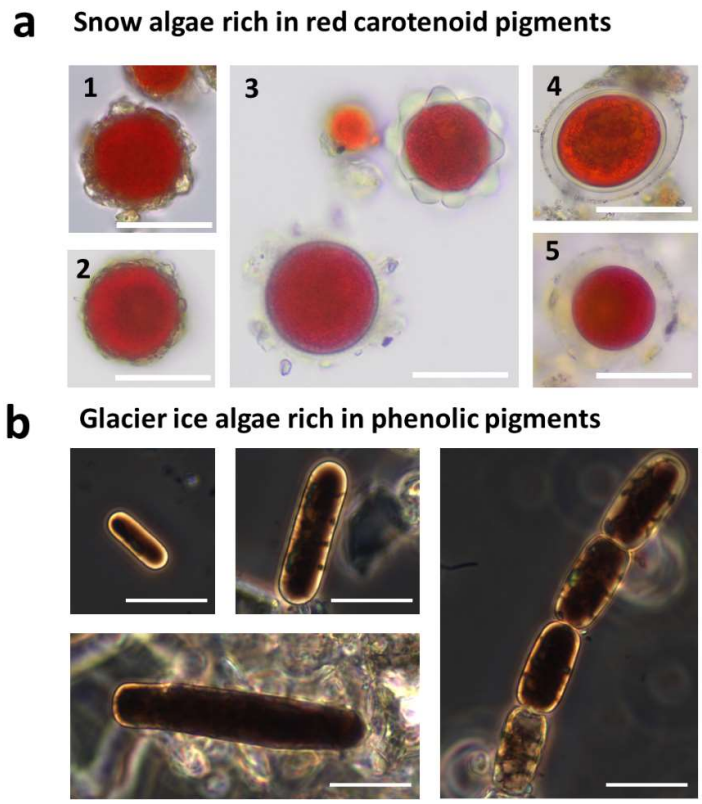


Figure S14: Diversity of snow algae and ice algae producing high levels of secondary carotenoid and phenolic pigments, possibly used as a source of carbon for pucciniomycotinous yeasts. (a) Examples of microalgae from red snow blooms, occurring worldwide and producing astaxanthin in cytosolic lipid droplets. *Sanguina nivaloides*, collected in the French Alps (1) and in the Scoresby Fjord, Greenland (2); *Rosetta* spp. (3), *Chlainomonas* spp. (4) and still uncharacterized species (5), collected in metling snowfields in the Scoresby Fjord, Greenland. *Phenoliferia psychrophenolica* occurring in melting snowfields, where these species thrive has the capability to feed from pigments released from dead cells during bloom decline. (**b)** Examples of microalgae from glacier ice. *Ancylonema* spp. cells collected at the surface of Milneland glacier, Greenland. *Phenoliferia psychrophenolica* is known as a snow yeast, however, we also detect it in the microbiome of glacier algal blooms. Its capability to degrade high levels of phenol may relate to an adaptation to glacier ice habitats as well, where *Ancylonema* species thrive and produce high levels of glycosylated purpurogallin pigments. Scale bar = 20 µm.

Supplementary Tables

Table S1: Genome statistics generated by BUSCO for assembly of *P. psychrophenolica* LCC-F-001-001 genome. Benchmarking of Universal Single-Copy Orthologs (BUSCO): C, complete; S, single, D, duplicated; F, fragmented; M, missing.

Statistics	***Phenoliferia psychrophenolica* LCC-F-001-001**
Size	41.6 Mb
Number of contigs	105
N50	621 014
Longest contig	1.37 Mb
% GC	59.48
Busco % (n = 758)	C:93.5% [S:91.8%, D:1.7%], F:0.7%, M:5.8%
Size mitochondrion genome	35,551
BUSCO for genome (nt)	C:93.5% [S:91.8%, D:1.7%], F:0.7%, M:5.8%, n:758
BUSCO for coding sequences predicted (nt)	C:92.7% [S:91.0%, D:1.7%], F:1.6%, M:5.7%, n:758
BUSCO for predicted proteins (aa)	C:92.0% [S:90.4%, D:1.6%], F:1.5%, M:6.5%, n:758

Table S2: High-similar duplicate (HSD) proteins in *P. psychrophenolica* LCC-F-001-001. HSD proteins were detected as described in Methods. Functional annotations were predicted based on the detection of Pfam and Interpro domains.

See Table S2.xlsx

Table S3: HSD proteins in *Meredithblackwellia eburnea*. HSD proteins were detected as described in Methods. Functional annotations were predicted based on the detection of Pfam and Interpro domains.

See Table S3.xlsx

Table S4: HSD proteins in Kriegeriales sp. P6C61. HSD proteins were detected as described in Methods. Functional annotations were predicted based on the detection of Pfam and Interpro domains.

See Table S4.xlsx

Table S5: Distribution of enzymes in the different CAZY subclasses for Microbotryomycetes. Quantitative comparison of CAZy domain families in Microbotryomycetes. Annotations were extracted for all Microbotryomycetes genomes available in CAZy (https://www.cazy.org) as of 02-Jan-2025. (*) species also identified as *P. himalayensis*.

See Table S5.xlsx

Table S6: **List of candidate genes involved in the breakdown of phenolic compounds, carotenoid acyl esters, and lipids (lipases).** InterPro (https://www.ebi.ac.uk/interpro) domains corresponding to ortho-cleavage pathways detected in *P. psychrophenolica* predicted proteome were used to define a list of 144 gene locuses.

See Table S6.xlsx

Table S7: Number of proteins and domains involved in cold adaptation in *P. psychrophenolica* LCC-F-001-001 and other fungi used for phylogenomic analyses. Abbreviation: nb, number.

See Table 7.xlsx

Table S8: Comparison of the distribution of enzymes in the different CAZY classes in the translated genome of *P. psychrophenolica* LCC-F-001-001 and all microbotryomycetes present in CAZY database. Analysis was performed with the CAZy database as of 2024, Apr. 06.

See Table S8.xlsx

Table S9: Function of genes acquired via HGT of bacterial origin in *P. psychrophenolica* LCC-F-001-001 and their presence in *M. eburnea* and *Kregeriales* sp.

Gene ID	Gene associated	Function of gene	contigs where the gene is present	Shared with
g5660	WP_242474272.1	family 1 glycosylhydrolase [*Tomitella cavernea*]	NODE_19	
g10829	MCU1645103.1	glycoside hydrolase family 1 [*Nocardia* sp.]	NODE_74	
g7503	WP_143907069.1	family 1 glycosylhydrolase [*Tomitella fengzijianii*]<>QDQ96912.1 glycoside hydrolase family 1 protein [*Tomitella fengzijianii*]	NODE_31	
g5058	MBV1903330.1	MmgE/PrpD family protein [*Marinosulfonomonas* sp.]	NODE_16	
g9239	NLU84323.1	glycoside hydrolase family 1 protein [*Rhodococcus* sp. HNM0569]	NODE_47	
g11002	MBX3093935.1	glucose 1-dehydrogenase [*Cryobacterium* sp.]	NODE_79	*Kregeriales* sp. 2 genes
g9039	WP_206038255.1	family 1 glycosylhydrolase [*Rhodococcus* sp. HNM0569]	NODE_45	
g5089	WP_242089926.1	glucose 1-dehydrogenase [*Microbacterium lacticum*]	NODE_16	*Kregeriales* sp. 2 genes
g11250	OGK85510.1	MAG: cyclopentanol dehydrogenase [*Candidatus* Rokubacteria bacterium GWC2_70_16]	NODE_88	*M. eburnea* 1 gene
g8224	WP_103888932.1	hypothetical protein [*Actinacidiphila yanglinensis*]<>SEG84565.1 Predicted glycosyl hydrolase, GH43/DUF377 family [*Actinacidiphila yanglinensis*]	NODE_37	
g7874	WP_040863559.1	phosphotransferase [*Nocardia niigatensis*]	NODE_34	
g6689; g7236; g7252; g7254; g7328; g7329; g7373; g8236; g8237; g8842; g8851; g9013; g9479; g9482; g9483; g9714; g2230; g2463; g10135; g2869; g10734; g3205; g11194; g11426		Ice binding proteins		

Table S10: Fungal genome used in phylogenomic analyses. Phylogenetic position, (Species, Phylum, Class), capacity to grow in the snow, genome statistics (Genome size, Number of proteins, Busco completion) and references (NCBI or JGI IDs and Link to genomic data) are presented.

See Table S10.xlsx

Table S11: Orthofinder statistics per fungal species used in phylogenomic analyses.

See Table S11.xlsx

Table S12: Detection of *P. psychrophenolica* in metagenomics sequences from snow blooms dominated by *S. nivaloides* and in glacier ice blooms dominated by *Ancylonema* spp. Mapping of reads was achieved using *P. phsychrophenolica* IT2 as a reference, as described in Methods. Presence was assessed based on 100% identity. Mapping plots are visualized using IGV.

See Table S12.xlsx

Supplementary Methods

Analysis of ITS2 secondary structure. To extract ITS2 sequence from *P. psychrophenolica* LCC-F-001-001 sequences and from the type species *P. psychrophenolica* AG21 (accession number: EF151246)[1]. ITSx 1.1.3[2] was used with the option --save_regions all. At both extremities of each ITS2 sequence, 25 nucleotides corresponding to the stem sequence of the 5.8S and 28S were added to avoid analyzing pseudogenes[3]. The ITS2 secondary structure of each sample was folded using the online tool RNAfold WebServer[4,5] with default parameter, except for energy parameter where Andronescu model was selected. The minimum free energy (MFE) structure was chosen and viewed using the online tool Forna[6]. The image of each folded ITS2 structure was downloaded in SVG format and modified using Inkscape 0.92.4[7]. The ITS2 secondary structure of *P. psychrophenolica* LCC-F-001-001 was compared to the type strain *P. psychrophenolica* AG21 using the Compensatory Base Changes (CBC) approach, visually and using the tool 4SALE 1.7.1[8]. Variant emerging from the alignment and the folded structure, and the RNA processing sites, pyrimidine/pyrimidine mismatch in Helix II were highlighted using Inkscape 0.92.4[7].

Growth curve on medium supplemented with pigments. *P. psychrophenolica* LCC-F-001-001 acclimated at 5°C was tested for growth on dimethylsulfoxide (DMSO), astaxanthin acyl esters, and purpurogallin extract from *Ancylonema alaskanum* as indicated. The culture was diluted to an OD of 0.07 and washed three times in minimal medium to remove YPD. In a 96-well plate, 200 µL final of *P. psychrophenolica* mixed with DMSO, or astaxanthin acyl ester solubilized with DMSO or purpurogallin carboxylic acid-6-O-β-D-glucopyranoside extract from *A. alaskana*. Controls with astaxanthin acyl esters, DMSO, and purpurogallin carboxylic acid-6-O-β-D-glucopyranoside extract were prepared without fungal culture. Experiments were performed in triplicate. During 25 days, cell counts were performed twice per week using a Malassez counting chamber with a 10 µL aliquot fraction. For Malassez counts, 10 squares were counted following the manufacturer's instructions, except for the experiment in presence of purpurogallin carboxylic acid-6-O-β-D-glucopyranoside extract, where 5 squares were counted due to the high cell density. OD was measured using a Tecan Infinite M1000 Pro. OD values were normalized to the initial measure at day 1. Graphs were generated using the R package ggplot2 and edited using Inkscape 0.92.4.

Resistance to freezing. *P. psychrophenolica* LCC-F-001-001 cultures were acclimated at 5°C and 20°C for 15 days and then diluted to an OD of 0.07. For freezing resistance evaluation, 500 µL of culture at 5°C and 20°C in YPD liquid media were incubated in a freezer at -20°C for four different durations: 3, 12, 24, and 53 hours. After each freezing period, 10 µL of the culture were plated in triplicate into 24-well plates containing YPD solid agar medium. The plates corresponding to cultures acclimated at 5°C and 20°C were then transferred to their respective temperatures (5°C and 20°C). Growth was checked after 10 days.

DNA extraction for genomic sequencing. For genomic DNA extraction one *P. psychrophenolica* LCC-F-001-001 colony was cultured in 250 mL YPD liquid medium for 30 days. After culture centrifugation at maximum speed, cell pellets were resuspended in 800 µL of CTAB buffer composed of cetyltrimethylammonium bromide (CTAB) 10%, NaCl 2 M, Tris-HCl 1 M pH 7.5, EDTA 0.5 M pH 8 and 10 µL of 2-Mercaptoethanol. Resuspended pellets were then broken using a Precellys Evolution (Bertin, France) homogenizer with glass beads of

425-600 µm (Sigma, USA). The lysis program comprised 3 homogenization cycles of 30 seconds each at 10,000 rpm, with a 20-second pause between cycles. The lysate was then incubated 30 minutes at 60°C in a water bath. Following incubation, the sample was centrifuged for 15 minutes at 16,000 x g at room temperature. The supernatant (500 µL) was mixed with 500 µL of $CHCl_3$-isoamyl alcohol (24:1), vortexed for 10 seconds, and centrifuged for 10 minutes at 16,000 x g at room temperature. The upper phase (400 µL) was then mixed with 400 µL of cold isopropanol and incubated overnight at -20°C. After incubation, the sample was centrifuged for 15 minutes at 16,000 xg at 4°C. The resulting pellet was washed with 1 mL of 70% (v/v) ethanol, and the supernatant was discarded. The pellet was dried for 10 minutes at 37°C and then resuspended in 30 µL of sterile water DNAse free. DNA was measured using a Qubit 4 fluorometer (Invitrogen, USA) and Nanodrop 2000 spectrophotometer (Thermo Scientific, USA) according to the manufacture's instruction. A quantity of ~8 µg of DNA at ~ 88 ng/µL with 260/280 and 260/230 ratio higher than 2 was sequenced with Illumina Miseq and Novaseq 6000 at Macrogen (South Korea) after RNase treatment. Data were co-analyzed with Sequentia Biotech (Spain).

Horizontal gene transfer (HGT) detection and highly similar duplicate (HSD) genes predictions. Horizontal gene transfer (HGT) detection was performed on the *P.psychrophenolica* LCC-F-001-001 predicted protein dataset. First, DIAMOND v2.1.8 BLASTP (ultra-sensitive mode, E-value < 1×10^{-5}) sequence alignments were conducted on the non-redundant (NR) database (April 2023). The top 800 hits were retained to ensure hits inside the same kingdom (in this case, the Fungi kingdom). Then, the Alien index [9] was calculated for each proteins with a custom script (https://github.com/pguenzi-tiberi/Alien_index_report_generator) using the following formula:

$$AI = \log\left(\left(Best_{e\ value\ for\ Fungi}\right) + e^{-200}\right) - \log\left(\left(Best_{evalue\ for\ Fungi}\right) + e^{-200}\right)$$

Proteins with an $AI > 45$ were considered as HGT candidates. To verify the presence of HGT in the other species, the Unix grep function was used in the DIAMOND v2.1.8 BLASTP output of Kriegeriales sp. and *Meredithblackwellia eburnea*. For HSD predictions, the predicted proteins were aligned against themselves with DIAMOND v2.1.8 BLASTP (ultra-sensitive mode, E-value < 1×10^{-5}). Pfam domains were annotated using the Interproscan suite/software to assign function to the genes. HSDfinder[10] was then used to detect highly similar duplicate genes in the BLAST results. The parameters set for HSDfinder were a minimum of 90% similarity and a minimum hit length of 10 amino acids. In this table, we considered that two genetic loci are tandemly duplicated provided they exhibited co-localization within the confines of the same genomic contig, with an intergenic distance not exceeding ten neighboring genes[11].

Mitochondrial genome annotation. The mitochondrial genome was identified from the initial assembly by mapping the cox1 gene of *Rhodotorula mucilaginosa* (NC_036340.1, position: 11605-15597). *De novo* assembly was obtained with same result using NovoPlasty v4.3.1[12]. Gene annotation was generated using CHLOROBOX[13] and Geseq v2.03[14]. *R. mucilaginosa, R. toruloides* and *Saccharomyces cerevisiae* were used as reference sequences. Annotation of tRNA was performed with ARAGORN63 v1.2.38[15] (Genetic code, Yeast mitochondrion; Max intron length = 3000; Fix intron). Graphical visualization was conducted using OGDRAW v1.3.1[16].

Phylogenomics. Phylogenomic analysis of *P. psychrophenolica* LCC-F-001-001 was conducted using Orthofinder v2.5.4[17]. Orthofinder is a comparative genomics tool, allowing the identification of orthogroups and orthologs, the reconstruction of gene and species trees, and the detection of gene duplication events. It requires protein sequences in FASTA format for each species. For this analysis, Orthofinder was run with the following settings: MSA (https://www.ebi.ac.uk/jdispatcher/msa) for gene tree inference, diamond_ultra_sens[18] for sequence search, and IQtree[19] for tree inference. Outputs included summary statistics, species-specific statistics, orthogroup overlaps, and orthologous gene pair statistics. In addition to the species used for CAZy protein comparison, *Mortierella alpina*, a snow fungus without CAZy annotation, was added, to identify potential cold-adapted proteins. The resulting species tree in Newick format was visualized using iTol[20].

Supplementary References

1. Margesin, R., Fonteyne, P. A., Schinner, F. & Sampaio, J. P. Rhodotorula psychrophila sp. nov., Rhodotorula psychrophenolica sp. nov. and Rhodotorula glacialis sp. nov., novel psychrophilic basidiomycetous yeast species isolated from alpine environments. *Int. J. Syst. Evol. Microbiol.* **57**, 2179–2184 (2007).

2. Bengtsson-Palme, J. *et al.* Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* **4**, 914–919 (2013).

3. Tesson, S. V. M. & Pröschold, T. Description of Limnomonas gen. nov., L. gaiensis sp. nov. and L. spitsbergensis sp. nov. (Chlamydomonadales, Chlorophyta). *Diversity* **14**, (2022).

4. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, 70–74 (2008).

5. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 1–14 (2011).

6. Gendron, P., Lemieux, S. & Major, F. Quantitative analysis of nucleic acid three-dimensional structures. *J. Mol. Biol.* **308**, 919–936 (2001).

7. Inkscape Project. Inkscape. (2020).

8. Seibel, P. N., Müller, T., Dandekar, T., Schultz, J. & Wolf, M. 4SALE - A tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics* **7**, 1–7 (2006).

9. Rancurel, C., Legrand, L. & Danchin, E. G. J. Alienness: Rapid detection of candidate horizontal gene transfers across the tree of life. *Genes (Basel).* **8**, (2017).

10. Zhang, X., Hu, Y. & Smith, D. R. HSDFinder: A BLAST-Based Strategy for Identifying Highly Similar Duplicated Genes in Eukaryotic Genomes. *Front. Bioinforma.* **1**, 1–12 (2021).

11. Noel, B. *et al.* Pervasive tandem duplications and convergent evolution shape coral genomes. *Genome Biol.* **24**, 123 (2023).

12. Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **45**, e18–e18 (2016).

13. of Molecular Plant Physiology, M. P. I. ChloroBox. (2023).

14. Tillich, M. *et al.* GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11 (2017).

15. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11–16 (2004).

16. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, W59–W64 (2019).

17. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

18. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale

using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).

19. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2014).

20. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).