

## Diverse signatures of convergent evolution in cacti-associated yeasts

Carla Gonçalves<sup>1,2,3,4</sup>, Marie-Claire Harrison<sup>1,2</sup>, Jacob L. Steenwyk<sup>1,2,5</sup>, Dana A. Opulente<sup>6,7</sup>, Abigail L. LaBella<sup>1,2,8</sup>, John F. Wolters<sup>6</sup>, Xiaofan Zhou<sup>1,2,9</sup>, Xing-Xing Shen<sup>1,2,10</sup>, Marizeth Groenewald<sup>11</sup>, Chris Todd Hittinger<sup>6</sup> and Antonis Rokas<sup>1,2\*</sup>

1 Vanderbilt University, Department of Biological Sciences, VU Station B #35-1634, Nashville, TN 37235, United States of America

2 Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN 37235, USA

3 Present address: Associate Laboratory i4HB—Institute for Health and Bioeconomy and UCIBIO—Applied Molecular Biosciences Unit, Department of Life Sciences, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal

4 Present address: UCIBIO-i4HB, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

5 Howards Hughes Medical Institute and the Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA

6 Laboratory of Genetics, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI 53726, USA

7 Biology Department, Villanova University, Villanova, PA 19085, USA

8 Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte NC 28223

9 Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Center, South China Agricultural University, Guangzhou 510642, China

10 College of Agriculture and Biotechnology and Centre for Evolutionary & Organismal Biology, Zhejiang University, Hangzhou 310058, China

11 Westerdijk Fungal Biodiversity Institute, 3584 Utrecht, The Netherlands

**Keywords:** convergent evolution, adaptive evolution, yeast ecology, cactophily, machine learning, pathogenicity, thermotolerance

**Running title:** Convergent evolution of yeast cactophily

\* **Correspondence:** antonis.rokas@vanderbilt.edu

### **Conflicts of interest**

JLS is a scientific advisor for WittGen Biotechnologies. JLS is an advisor for ForensisGroup Inc. AR is a scientific consultant for LifeMine Therapeutics, Inc.

### **Acknowledgements**

We thank members of the Rokas Lab, Hittinger Lab, and members of the Y1000+ project (<http://y1000plus.org>) for helpful discussions. This work was performed using resources contained within the Advanced Computing Center for research and Education at Vanderbilt University in Nashville, TN.

### **Funding**

National Science Foundation Grant DEB-2110403 (CTH)

National Science Foundation Grant DEB-2110404 (AR)

DOE Great Lakes Bioenergy Research Center, funded by BER Office of Science Grant DE-SC0018409 (CTH)

USDA National Institute of Food and Agriculture Hatch Project 1020204 (CTH)

H. I. Romnes Faculty Fellow, supported by the Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation (CTH)

National Institutes of Health/National Institute of Allergy and Infectious Diseases Grant R01 AI153356 (AR)

Burroughs Wellcome Fund (AR)

Research supported by the National Key R&D Program of China Grant 2022YFD1401600 (XXS)

National Science Foundation for Distinguished Young Scholars of Zhejiang Province Grant LR23C140001 (XXS)

Fundamental Research Funds for the Central Universities Grant 226-2023-00021 (XXS)

National Institutes of Health Grant T32 HG002760-16 (JFW)

National Science Foundation Grant Postdoctoral Research Fellowship in Biology 1907278 (JFW)

JLS is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation.

## Abstract

Many distantly related organisms have convergently evolved traits and lifestyles that enable them to live in similar ecological environments. However, the extent of phenotypic convergence evolving through the same or distinct genetic trajectories remains an open question. Here, we leverage a comprehensive dataset of genomic and phenotypic data from 1,049 yeast species in the subphylum Saccharomycotina (Kingdom Fungi, Phylum Ascomycota) to explore signatures of convergent evolution in cactophilic yeasts, ecological specialists associated with cacti. We inferred that the ecological association of yeasts with cacti arose independently ~17 times. Using machine-learning, we further found that cactophily can be predicted with 76% accuracy from functional genomic and phenotypic data. The most informative feature for predicting cactophily was thermotolerance, which is likely associated with duplication and altered evolutionary rates of genes impacting the cell envelope in several cactophilic lineages. We also identified horizontal gene transfer and duplication events of plant cell wall-degrading enzymes in distantly related cactophilic clades, suggesting that putatively adaptive traits evolved through disparate molecular mechanisms. Remarkably, multiple cactophilic lineages and their close relatives are emerging human opportunistic pathogens, suggesting that the cactophilic lifestyle—and perhaps more generally lifestyles favoring thermotolerance—may preadapt yeasts to cause human disease. This work underscores the potential of a multifaceted approach involving high throughput genomic and phenotypic data to shed light onto ecological adaptation and highlights how convergent evolution to wild environments could facilitate the transition to human pathogenicity.

## Introduction

Convergent evolution, the repeated evolution of similar traits among distantly related taxa, is ubiquitous in nature and has been documented across all domains of life<sup>1-3</sup>. Convergence typically arises when organisms occupy similar ecological niches or encounter similar conditions and selective pressures; facing similar selective pressures, organisms from distinct lineages often evolve similar adaptations.

Independently evolved phenotypes often share the same genetic underpinnings (parallel molecular evolution)<sup>4-7</sup> such as similar mutations in specific genes<sup>4,5,8</sup>, but can also arise through distinct molecular paths and by distinct evolutionary mechanisms<sup>9,10</sup>, such as copy number variation<sup>11-13</sup>, gene losses<sup>14-16</sup>, or gene gains (e.g., horizontal gene transfer or HGT)<sup>17,18</sup>. Molecular signatures of convergence can also be inferred from independent shifts in overall evolutionary rates<sup>19</sup> and examined at higher hierarchical levels of molecular organization, such as functions or pathways<sup>20</sup>. For instance, comparing rates of evolution across distantly related animal lineages could pinpoint convergent slowly evolving genes involved in adaptive functions<sup>21</sup> or convergent rapidly evolving genes indicating parallel relaxed constraints acting on dispensable functions<sup>22</sup>. Parallel molecular changes are common across all domains of life<sup>9</sup>, but their occurrence can be reduced by mutational epistasis or the polygenic nature of some phenotypic traits<sup>10,23,24</sup>, particularly when studying convergence in distantly related organisms.

Fungi exhibit very high levels of evolutionary sequence divergence<sup>25</sup>; the amino acid sequence divergence between the baker's yeast *Saccharomyces cerevisiae* and the human commensal and opportunistic pathogen *Candida albicans*, both members of subphylum Saccharomycotina (one of the three subphyla in Ascomycota, which is one of the more than one dozen fungal phyla) is comparable to the divergence between humans and sponges<sup>26</sup>. Due to their very diverse genetic makeups, convergent phenotypes arising in fungi might involve distinct genetic determinants and/or mechanisms, including HGT<sup>27,28</sup>, a far less common mechanism among animals (but see<sup>29,30</sup>).

Saccharomycotina yeasts are ecologically diverse, occupy diverse ecosystems<sup>31</sup>, and vary considerably in their degree of ecological specialisation ranging from cosmopolitan generalists to ecological specialists. For instance, *Sugiyamaella* yeasts are mostly isolated from insects<sup>32</sup> and most *Tortispora* species have been almost exclusively found in association with cacti plants<sup>33</sup>. The cactus environment accommodates numerous yeast species rarely found in other niches<sup>34-37</sup>. Moreover, cactophilic yeasts are part of a model ecological system involving the tripartite relationship between cactus, yeast, and *Drosophila*<sup>34,36,38,39</sup>. Cactophilic yeasts use necrotic tissues of cacti as substrates for growth<sup>40</sup> while serving as a food source to cactophilic *Drosophila*. Cactophilic flies (and other insects) play, in turn, a crucial role in the yeast's life cycle by acting as vectors<sup>34</sup>.

In *Drosophila*, the adoption of cacti as breeding and feeding sites evolved ~16-21 million years ago (Mya) and is considered one of the most extensive and successful ecological transitions within the genus<sup>41</sup>. Cactophilic *Drosophila* thrive across a wide range of cacti species that differ in the profiles of toxic metabolites they produce – *Opuntia* species, commonly called prickly pear cactus, generally contain fewer toxic metabolites than columnar cacti species<sup>40</sup> and are likely the ancestral hosts<sup>41</sup>. The distinctive characteristics of the cacti environment<sup>42</sup> seemingly selected for adaptive traits across cactophilic *Drosophila*, such as high resistance to heat, desiccation, and toxic alkaloid compounds produced by certain types of cacti<sup>43-48</sup>. These traits are likely associated with several genomic signatures (e.g., positive selection, gene duplications, gene gains) impacting multiple functions, such as water preservation or detoxification<sup>45-50</sup>.

Contrasting with *Drosophila*, where cactophily is largely found within the monophyletic *D. repleta* group<sup>41</sup>, molecular phylogenetic analyses revealed that cactophilic yeasts belong to phylogenetically distinct clades, indicating that association with cacti evolved multiple times independently in the Saccharomycotina<sup>37</sup>. While relevant ecological and physiological information of cacti-associated yeasts is available<sup>35-37,39</sup>, the genetic changes that facilitated the convergent evolution of multiple yeast lineages to the cacti environment are unknown.

Benefiting from the wealth of genomic and phenotypic data available for nearly all known yeast species described in the subphylum Saccharomycotina<sup>51</sup> and cross-referencing with the ecological data available from the cactus-yeast-*Drosophila* system<sup>35-37,52-60</sup>, we employed a high throughput framework to detect signatures of convergent evolution in 17 independently evolved lineages of cactophilic yeasts. Using a machine learning algorithm, we uncovered distinctive phenotypic traits enriched among cacti-associated yeasts, including the ability to grow at high ( $\geq 37^{\circ}\text{C}$ ) temperatures. We found that thermotolerance might be related to duplication and distinctive rates of evolution in genes impacting the integrity of the cell envelope, some of which are under positive selection in distantly related cactophilic clades. Gene family evolution analyses identified gene duplications and HGT events involving plant cell wall-degrading enzymes in distinct clades, suggesting adaptations associated with feeding on plant material. These results reveal that convergence to cactophily by distinct lineages of Saccharomycotina yeasts was accomplished through diverse evolutionary mechanisms acting on distinct genes, some of which might be associated with similar biological functions. Interestingly, we found that several cacti-associated yeasts and close relatives are emerging opportunistic human pathogens, raising the hypothesis that fungi inhabiting certain wild environments are preadapted for opportunistic pathogenicity. More broadly, we advocate for a methodological framework that couples diverse lines of genomic, phenotypic, and ecological data with multiple analytical approaches to investigate the plurality of evolutionary mechanisms underlying ecological adaptation.

## Results

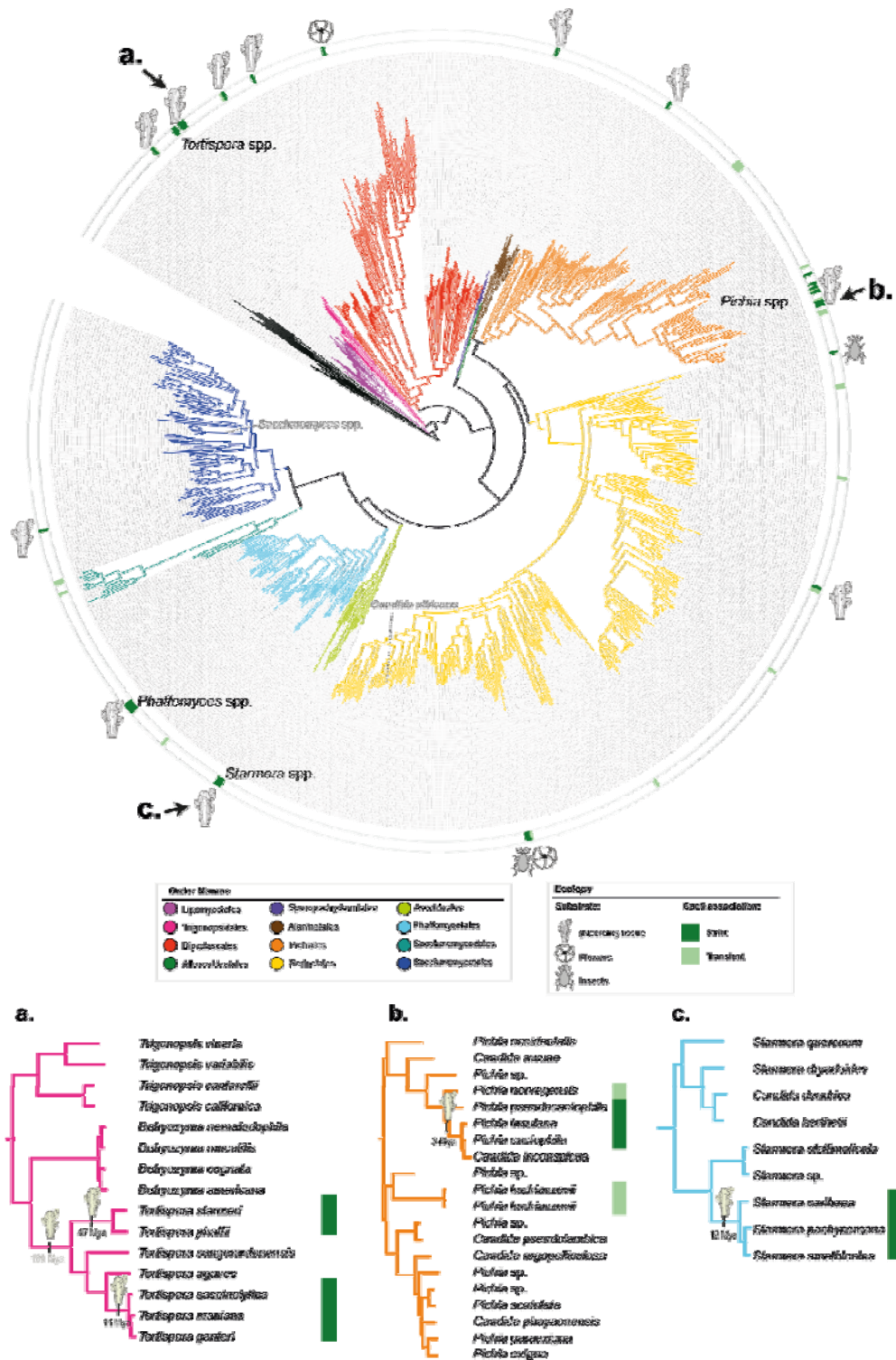
### Yeast cactophily likely evolved independently 17 times

We examined the ecological association of yeasts with the cacti environment across a dataset of 1,154 strains from 1,049 yeast species<sup>51,61</sup>. Yeast-cacti associations vary substantially in their strengths<sup>35</sup>. Some cacti-associated yeast species are more cosmopolitan, being commonly isolated from cacti but also other environments (henceforth referred to as transient), whereas others are strictly cactophilic, defined as those almost exclusively isolated from cacti (Table S1; note that this classification is based on the available ecological information, which may be impacted by sampling bias and other sampling issues – it is possible that strictly cactophilic species could also be found in other, yet unsampled, environments). We observed that strictly cactophilic species are found across almost the entire Saccharomycotina subphylum spanning from the Trigonopsidales (i.e., *Tortispora* spp.)<sup>33</sup> to the Saccharomycetales (i.e., *Kluyveromyces starmeri*)<sup>52</sup> (Fig. 1). Using these species in an ancestral state reconstruction, we inferred a total of 17 origins for the evolution of cactophily (Fig. S1).

Cactophily is found in single species belonging to different orders, but it also involves nearly entire genus (i.e., *Tortispora*). Specifically, seven of the 17 instances of cactophily evolution involve clades containing two or more species while the remaining 10 involve single species, suggesting that different taxa evolved this ecological association at different times (Fig. 1). For instance, cross referencing relaxed molecular clock analyses of the yeast phylogeny<sup>51</sup> with ancestral state reconstructions suggests that cactophily in *Tortispora* likely emerged twice, once in the most recent common ancestor (MRCA) of *T. starmeri*/*T. phaffi* around 47 Mya and in the MRCA of *T. caseinolytica*/*T. mauiana*/*T. ganteri* around 11 Mya (Fig. 1). An alternative hypothesis would place the emergence of cactophily in the MRCA of the genus around 180 Mya, which is inconsistent with the 35 Mya estimated origin of the Cactaceae family<sup>62</sup>. Cactophily in the genus *Starmera* and the *Pichia cactophila* group emerged more recently, most likely around 12 and 3 Mya, respectively (Fig. 1).

Cacti-associated yeasts seemingly exhibit significant niche partitioning (Table S1). For example, *T. ganteri* is typically isolated from columnar cacti, while its close relative *T. caseinolytica* is more commonly found in *Opuntia* spp.<sup>33</sup>. Furthermore, *Pichia cactophila* is considered a generalist cactophilic yeast, being widely distributed across a wide range of cacti species<sup>37</sup>, while closely related *P. heedii* has been predominantly found in association with certain species of columnar cacti<sup>39</sup>. However, many species (e.g., *T. starmeri* or *P. insulana*) alternate between the two types of cacti<sup>33,58</sup>, similar to some *Drosophila* species<sup>41</sup>. Other species, such as *Wickerhamiella cacticola* or *Kodamaea nitidulidarum*, are associated with cacti flowers and/or flower-visiting insects, like beetles, and not with necrotic cacti tissues<sup>59,60,63</sup>.





**Fig. 1. Yeast cactophily originated repeatedly and at different times.** (top) Genome-wide-based phylogeny of the subphylum Saccharomycotina<sup>51</sup> depicting the different types of ecological association of strictly cactophilic yeast species with the cacti environment: (necrotic)

cacti tissues, which include both *Opuntia* spp. and columnar cacti, cacti flowers, and cacti-visiting insects. Cactophilic lineages and well-known yeasts, such as *Saccharomyces* spp. and *Candida albicans*, are noted on the phylogeny. **a.**, **b.**, and **c.**. Subtrees of three cactophilic clades: **a.** *Tortispora*, **b.** *Starmera*, and **c.** *Pichia*. Estimated times of origin, as determined in <sup>51</sup>, for the emergence of cactophily are represented for these three example clades.

### **Detecting signatures of convergent evolution in cactophilic yeasts**

We envision three distinct scenarios that may capture how different yeast lineages convergently evolved cactophily (Fig. 2A):

#### *Scenario I: Convergent phenotypes and genotypes*

Selective pressures associated with the cacti environment (e.g., high temperature, desiccation, or presence of toxic compounds) favor similar phenotypic traits that evolved through the same genomic mechanisms;

#### *Scenario II: Convergent phenotypes through divergent genotypes*

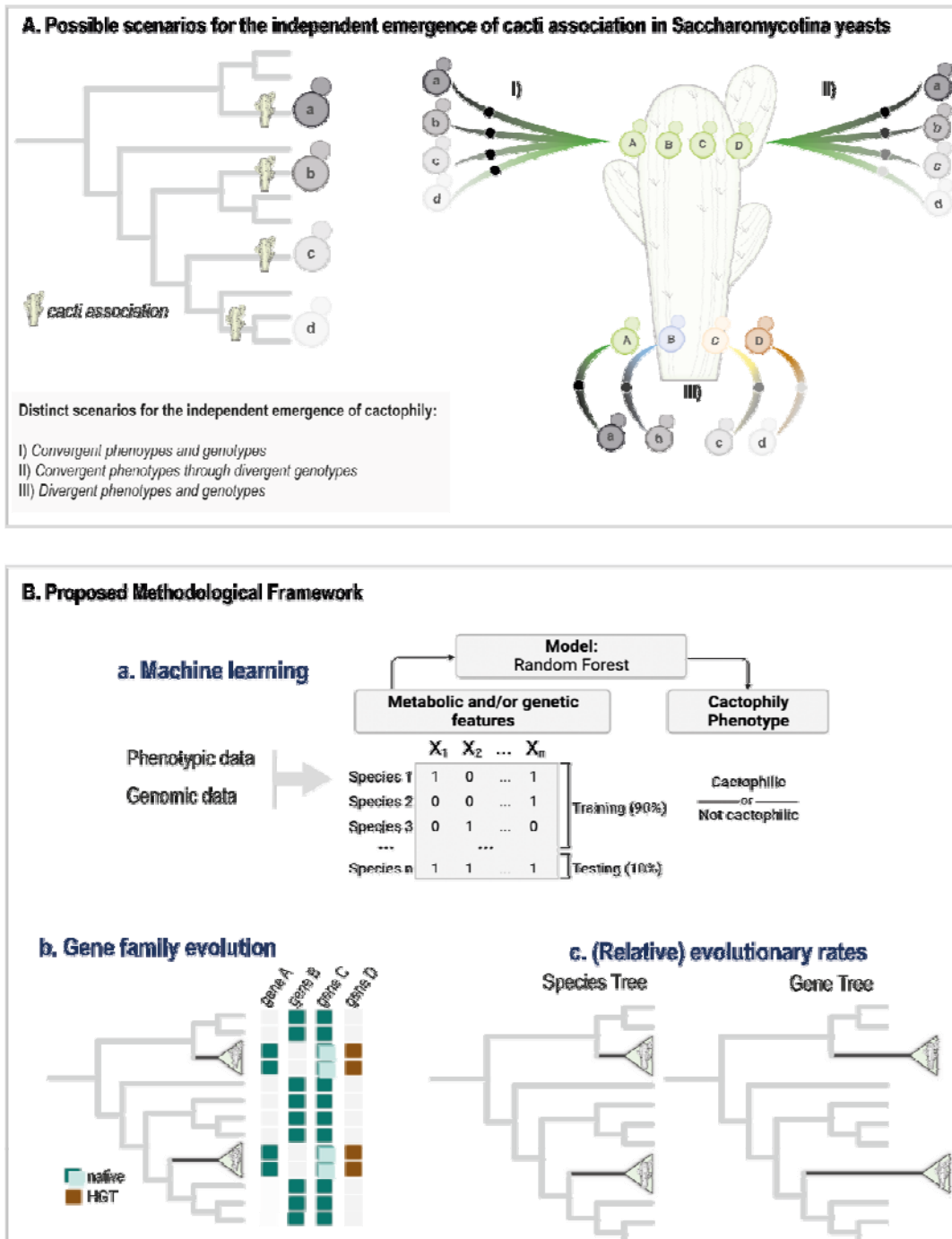
Selective pressures associated with the cacti environment favor similar phenotypic traits across cactophilic species, but different evolutionary mechanisms (e.g., gene duplication, HGT) contribute to phenotypic convergence of different lineages. In this scenario, similar phenotypes emerge through distinct evolutionary trajectories;

#### *Scenario III: Divergent phenotypes and genotypes*

Distinct phenotypic landscapes are explored by distinct clades when thriving in the same environment (niche partitioning); different evolutionary mechanisms contribute to these phenotypes.

To explore which scenario(s) best reflect(s) the process of yeast adaptation to the cacti environment, we developed a framework for identifying signatures of adaptation and

convergence from high-throughput genomic and phenotypic data<sup>51</sup> (Fig. 2B). Specifically, we performed: machine learning to identify phenotypic and genetic commonalities that distinguish cactophilic from non-cactophilic yeasts; genome-wide family analyses to identify patterns of gene presence/absence due to gene duplication, gene loss, and HGT; and genome-wide evolutionary rate analyses to detect signatures of convergence in evolutionary rates and of positive selection in individual genes (Fig. 2B), which have been also frequently implicated in adaptive evolution<sup>5,19,21,22,64</sup>. We applied this methodological framework to study convergence in ecological specialization in yeasts but note that it can be applied more generally to study the process of convergent or adaptive evolution.



**Fig. 2. Alternative scenarios for the evolution of yeast cactophily and methodological framework employed in this study.** (A) Methodological framework for investigating signatures of convergent evolution of yeast cactophily (B). A) Association with cacti evolved multiple times in yeasts from distinct genetic backgrounds, which are represented by distinct shades of grey. Convergent adaptation to the cacti environment might have involved convergence at both phenotypic and genomic levels (Scenario I), convergence at the phenotypic level but through

distinct molecular paths (Scenario II) or involved lack of convergence at both phenotypic and genotypic levels (Scenario III). Each of these scenarios is tested using a methodological framework (B) involving (a) machine learning approach wherein a model is trained to distinguish cactophilic from non-cactophilic yeasts from genomic and phenotypic features; (b) gene family evolution to find evidence of gene gains, duplications, losses, and HGT that might have occurred in cactophilic yeasts; (c) evaluation of signatures of positive selection ( $\omega$ ) and of changes in relative evolutionary rates in branches leading to cactophilic clades. This strategy will expose both genomic and phenotypic traits associated with cactophily (a, b) and highlight genes that are evolving slower or faster than the average of the genome or are under positive selection in cactophilic vs. non-cactophilic species (c).

### **Specific metabolic and genomic traits predict cactophily**

To investigate if cactophilic yeasts share similar phenotypic and/or genomic traits, we used a dataset of 1,154 yeast strains<sup>51</sup>, from which 52 are either strictly cactophilic (rarely found in other environments, n=31) or transient (frequently isolated from cacti but cosmopolitan, n=21) (Table S1). Functional genomic annotations (KEGG – 5,043 features) and physiological data (122 features) were retrieved<sup>51</sup> and used as features in a supervised random forest (RF) classifier trained to distinguish cactophilic from non-cactophilic yeasts. By training 20 independent RF runs using randomly selected balanced datasets (52 non-cactophilic species randomized each time and the 52 cactophilic species), we correctly identified an average of 38 cactophilic species (Fig. 3A), yielding an overall accuracy and precision of 72% and 73%, respectively. Species incorrectly assigned in 10 or more independent runs were equally distributed across strictly and transiently cactophilic groups (seven in each) (Table S2). We next repeated the analysis considering only strictly cactophilic species (transient were considered non-cactophilic) and obtained slightly higher accuracy (76%) and precision (76%) (Fig. 3B). Notably, correct classifications were obtained across phylogenetically distantly related genera (e.g., *Tortispora*, *Phaffomyces* or *Pichia*) (Fig. 3A), while incorrect classifications were obtained for species belonging to cactophilic clades in which correct classifications were

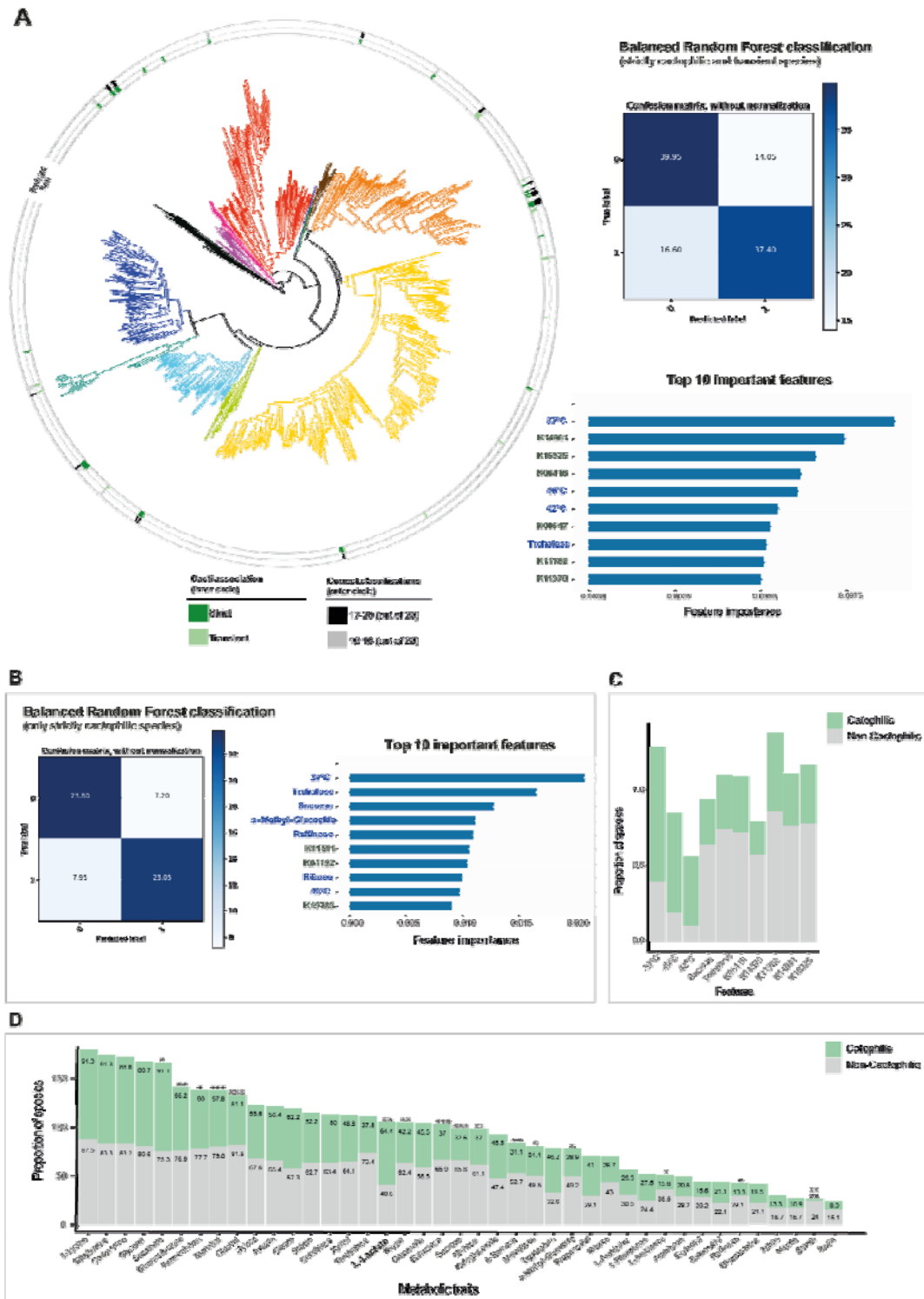
obtained; for instance, despite being closely related, we obtained both correct and incorrect classifications for cactophilic species belonging to the *Phaffomyces* genus (highlighted in light blue, Fig. 3A). These results suggest that the phylogenetic relatedness between some cactophilic species did not interfere with the accuracy of the RF classifier.

We next examined the top features that significantly contributed to the RF classifier. The feature with the highest relative importance was growth at 37°C either when analysing all cactophilic species (Fig. 3A) or just strictly cactophilic species (Fig. 3B). Ability to grow at high temperatures was previously found to be prevalent in cacti-associated yeasts<sup>38,65</sup> and is also an adaptive feature of cactophilic *Drosophila*<sup>48,66</sup>. Growth at 40°C and 42°C were also among the top ten most important features, supporting the hypothesis that thermotolerance is a distinctive feature of cactophilic species. In fact, 90%, 66%, and 46% of cactophilic yeasts can grow at 37°C, 40°C and 42°C, respectively, compared to only 39%, 19% and 10% of non-cactophilic yeasts (Chi-Squared Test,  $p < 0.01$ ) (Fig. 3C).

Analysing the top 100 most important metabolic features, we observed that 74 are less common in cactophilic species compared to non-cactophilic species (Table S2). For instance, trehalose assimilation is more rarely found in cactophilic species (~25%) than non-cactophilic species (~74%) (Chi-Squared Test,  $p$ -value  $< 0.01$ ) (Fig. 3C). Trehalose generally accumulates during numerous stress conditions including heat stress<sup>67,68</sup>. When cells return to a more favourable condition, the accumulated trehalose is hydrolysed into glucose by the trehalase Nth1. Inactivation of *NTH1* by mutations, and therefore subsequent impairment of trehalose hydrolysis, can be one of the outcomes of the heat stress response in experimentally evolved strains of *S. cerevisiae* under high temperature stress<sup>69,70</sup>, suggesting that deficient trehalose hydrolysis can be beneficial under long-term thermal stress conditions. However, *NTH1* is generally present in cactophilic genomes, suggesting that absence of *NTH1* does not explain impairment in trehalose assimilation in these species. Other top important metabolic features, such as sucrose assimilation, are also more rarely found in cactophilic species (Chi-Squared Test,  $p$ -value  $< 0.001$ ) (Table S2, Fig. 3C), echoing the general trend of a narrower spectrum of

carbon sources assimilated by cactophilic species compared to their non-cactophilic counterparts (Fig. 3D). However, some metabolic traits are more frequently found among cacti-associated yeasts, such as assimilation of lactate (Chi-Squared Test,  $p$ -value  $< 0.01$ ) (Fig. 3D), which was previously found to be positively associated with the cacti environment<sup>38,65</sup>.

Among the most important genomic features were presence or absence of genes involved in multiple distinct functions: K15325 (splicing), K06116 (glycerol metabolism), K01192 (N-glycan metabolism), K00547 (amino acid metabolism), K11762 (chromatin remodelling), K11370 (DNA repair), K11511 (DNA repair), or K19783 (post-replication repair). All these features are less common in cactophilic than in non-cactophilic species (Table S2). One interesting exception is K03686, a Hsp40 family protein encoded by 85% of cactophilic and 57% of non-cactophilic species (Table S2).



**Fig. 3. Cactophilic and non-cactophilic yeasts can be predicted from genomic and metabolic data with good accuracy. A) (left) Distribution of correct random forest (RF)**



classifications of cactophilic yeasts across the Saccharomycotina phylogeny. (right) Confusion matrix showing the average number of true positives (37.40), true negatives (39.95), false positives (16.60), and false negatives (14.05) across strictly cactophilic and transient species resulting from 20 independent RF runs. On the bottom, the top ten most important metabolic (presence or absence of growth; in blue) and genomic (presence or absence of KEGG; in grey) features for the RF classifier ranked according to their importance scores are shown. B) Confusion matrix and top ten most important features for the RF classifier using only strictly cactophilic species. C) Distribution of the proportion of the most important features to predict cactophily (strict cactophilic and transient) and non-cactophilic species for the entire dataset of 1,154 yeasts. D) Distribution of the proportion of metabolic traits (representing presence or absence of growth under the conditions shown)<sup>51</sup> across cactophilic (strict and transient) and non-cactophilic species. Only metabolic traits for which no more than 50% of data were missing were considered. Additionally, only metabolic traits exhibiting more than 10% prevalence in one of the groups (cactophilic or non-cactophilic) are shown. Statistically significant differences (Chi-Squared Test; \* p-value < 0.05, \*\* p-value < 0.01, \*\*\* p-value < 0.001) between the proportions of cactophilic and non-cactophilic species able to grow in each carbon/nitrogen source are shown.

### **HGT and duplication of cell wall degrading enzymes in cactophilic yeasts**

We next looked for genes that might be implicated in cactophily by examining gene family evolution<sup>71</sup> across three groups that contained two or more cactophilic lineages. We constructed three distinct datasets (Table S3) containing species of interest and closest relatives within the Lipomycetales/Dipodascales/Trigonopsidales orders (referred to as LDT group, including *Tortispora* spp., *Dipodascus australiensis*, *Magnusiomyces starmeri*, *Myxozyma mucilagina*, and *Myxozyma neglecta*), Phaffomycetales (including *Starmera* spp. and *Phaffomyces* spp.) and Pichiales (including two distinct *Pichia* spp. cactophilic clades).

We focused on gene duplications and gene gains (e.g., HGT), as gene losses are usually not reliably estimated due to annotation and sampling issues or inaccurate gene family clustering<sup>71</sup>.

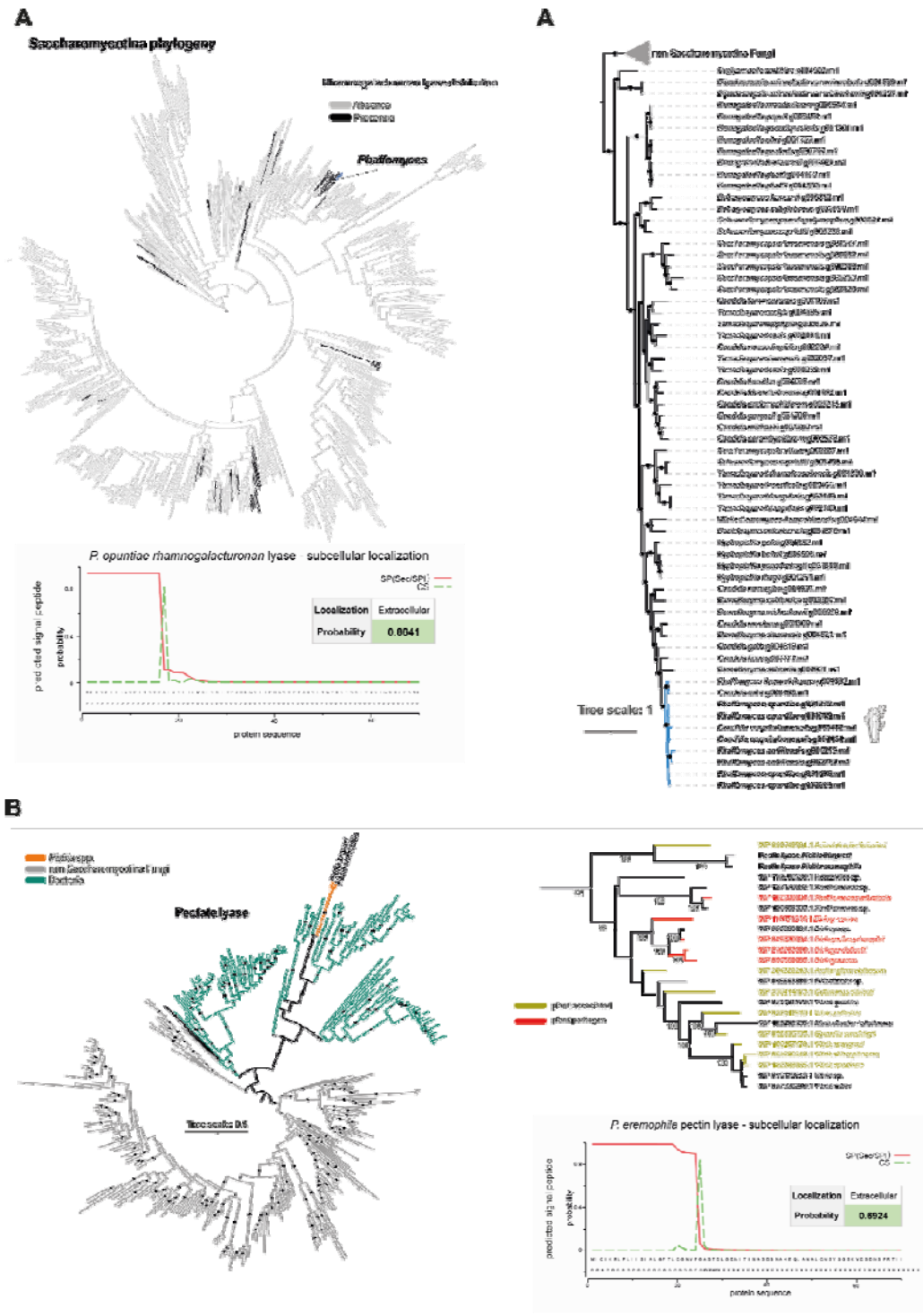
Using gene tree – species tree reconciliation analyses implemented in GeneRax<sup>71</sup> we first examined genes with evidence of duplication in at least one cactophilic species belonging to each group. We cross-referenced the candidate genes with their respective functional annotations (KEGG)<sup>51</sup> and found 11 KEGG families with evidence of duplication in all three groups and 29–37 families with evidence of duplication in two of the three groups (Table S4). Among these gene families, we found duplicated cellulase-encoding genes (K01210) in *Dipodascus australiensis* (LDT group) and *Starmera* species (Phaffomycetales). For the latter species, however, we could not recover complete gene sequences as they were located either at the end of scaffolds or in short scaffolds. Nevertheless, BLASTp searches against the non-redundant (nr) NCBI database indicate that these sequences were endoglucanase-like enzymes similar to those found in filamentous fungi (e.g., KOS23166.1) and are rarely found in yeasts<sup>72</sup>.

Duplication of another gene involved in plant cell wall degradation, encoding a rhamnogalacturonan endolyase (K18195), was detected in the cactophilic *P. antillensis*, *P. opuntiae*, and *Candida coquimbensis* (Phaffomycetales, Table S4). These species contained two copies of this gene compared to their closest relatives, which contained only one (Fig. 4A). These enzymes are responsible for the extracellular cleavage of pectin<sup>73</sup>, which along with cellulose, is one of the major components of plant cell wall. Pectin lyase<sup>74</sup> activities have been rarely reported among yeasts; therefore, we assessed the distribution of the rhamnogalacturonan endolyase across the 1,154 proteomes using a BLASTp search (e-value cutoff  $e^{-3}$ ) and found that it displayed a patchy distribution, being found in fewer than 60 species (Fig. 4A).

Interestingly, by inspecting orthogroups that uniquely contained cactophilic species, we found a gene encoding a pectate/pectin lyase was uniquely found in *P. eremophila* (strictly cactophilic) and *P. kluyveri* (transient), which are commonly isolated from rotting cacti tissues. Sequence similarity searches across the entire dataset of 1,154 yeast genomes confirmed that this gene is absent from all other species. Pectate lyases are also extracellular enzymes involved in pectin hydrolysis and plant cell wall degradation. Consistent with this function, these enzymes are mostly found among plant pathogens and plant-associated fungi and bacteria<sup>75,76</sup>, and their

activity has only been reported in a handful of Saccharomycotina species <sup>74</sup>. Phylogenetic analyses showed that the two yeast sequences are nested deeply within a clade of bacterial pectin lyases (Fig. 4B). The most closely related sequence belongs to *Acinetobacter boissieri* <sup>77</sup>, which has been frequently isolated from plants and flowers, and to *Xanthomonas* and *Dickeya*, two genera of plant pathogenic bacteria <sup>78,79</sup>.

Consistent with their function, HGT-derived pectin lyases as well as cellulases and rhamnogalacturonan endolyases, were predicted to localize to the extracellular space based on primary sequence analyses <sup>72</sup> (Fig. S2, Fig. 4). Pectin lyase enzymatic activity was previously detected in *P. kluyveri* strains associated with coffee fermentation <sup>76,80</sup>, suggesting that the identified HGT-derived pectin lyase is likely responsible for this activity.



**Fig. 4. Duplication and horizontal gene transfer of plant cell wall-degrading enzymes in cactophilic species.** A) Distribution of rhamnogalacturonan endolyases across the 1,154 yeast genomes (presence in black and absence in grey). (on the right) Phylogenetic tree of yeast rhamnogalacturonan endolyase and closest relatives, highlighting the duplication events in

cactophilic *Phaffomyces* species. Phylogeny was constructed in IQ-TREE v2.0.6 (-m TEST, -bb 1,000)<sup>81,82</sup>. Branch support (bootstrap  $\geq 90$ ) is represented as black circles. A BLASTp search against the NCBI nr database (selecting 500 hits) was performed using *P. opuntiae* g001049.m1 protein sequence as query. BLASTp was also performed against the yeast dataset of 1,154 proteomes and all significant hits were retrieved (e-value cutoff  $e^{-3}$ ). In *P. opuntiae*, there are two additional partial sequences (g001496.m1 and g001773.m1, 163 amino acids) that only partially overlap (from 39 to 163 overlapping amino acids) with the remaining nearly complete sequences (g002229.m1: 468 amino acids and g001049.m1: 610 amino acids). Prediction of subcellular localization according to SignalP and Deeploc<sup>72,83</sup> is shown in the panel below. B) Phylogenetic tree of closest related sequences to pectin lyases from *P. eremophila* and *P. kluyveri*. (on the right) Pruned tree highlighting the ecological association of bacteria species harboring the closest related pectin lyase sequences to *P. eremophila/P.kluyveri* proteins. Prediction of subcellular localization<sup>72,83</sup> is shown in the panel below.

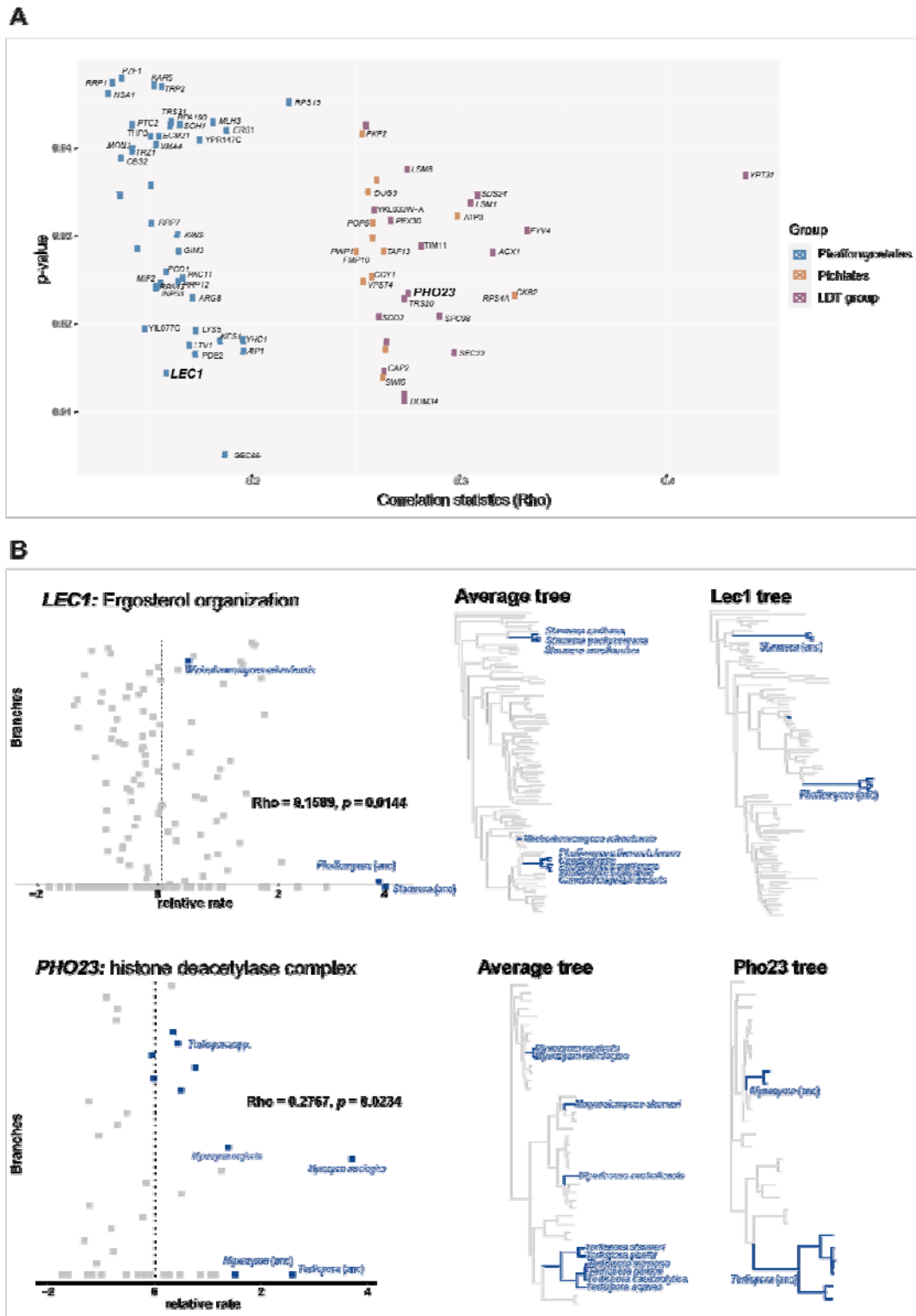
### **Convergent accelerated rates in heat resistance-related genes**

We next specifically looked for evidence of convergent evolutionary rates<sup>84</sup>, another indicator of adaptation<sup>21,22</sup>. For this analysis, we selected all cactophilic species found within the Pichiales, Phaffomycetales and the LDT group.

Correlation analyses between relative evolutionary rates (RER) and cactophily across ancestral (anc) and terminal branches revealed changes in evolutionary rate associated with the evolution of cactophily. Specifically, we inferred that 20 / 3,029 gene families in the LDT group are under accelerated evolution (evolving significantly faster than the average gene) and 33 / 3,029 have undergone decelerated evolution (evolving significantly slower than the average gene) (Table S5, Fig 5A). In the Pichiales, 14 / 2,204 showed evidence of acceleration and 25 / 2,204 of deceleration (Fig. 5A). In the Phaffomycetales, we found 32 / 3,550 accelerated genes and 30 / 3,550 decelerated genes (Table S5, Fig. 5A).

The accelerated genes are associated with varied cellular functions. In Pichiales, 7 out of 14 accelerated genes impact heat resistance according to large-scale studies<sup>85</sup>. For instance, *SWI6*, encodes a transcription factor that induces transcription during heat stress, and deletion of this gene causes several impairments in the resistance to multiple stresses, including heat<sup>86</sup> and cold<sup>87</sup>. Among the accelerated genes in Phaffomycetales, we found *LECI*, which was recently associated with ergosterol organization (Fig. 5B)<sup>88</sup>.

Inspecting the literature for phenotypes associated with either null or conditional mutants<sup>85</sup>, we found that, irrespective of their function, 12 / 20 genes that exhibited accelerated rates in the LDT group are involved in either heat and/or desiccation resistance. For instance, loss of function mutations in *CAP2*, which encodes part of a capping complex involved in barbed-end actin filament capping and filamentous growth, are associated with heat sensitivity and abnormal chitin localization leading to aberrant cell morphology<sup>89,90</sup>. Another gene identified as having accelerated evolutionary rates in the LDT group was *PHO23* (Fig. 5B), which was found to be required for the growth of *S. cerevisiae* during heat shock<sup>86</sup>. Genes that underwent decelerated evolution also play multiple roles (Table S5), and some are involved in essential functions such as DNA repair, cell cycle and splicing or encoding ribosomal subunits.



**Fig. 5. Genes involved in maintenance of the cell envelope show accelerated evolution in cactophilic clades.** A) Genes with accelerated evolutionary rates in the three groups of cactophilic species inspected. Correlation statistics (Rho) for converge between accelerated

rates of evolution in specific genes and cactophilic as well as the respective statistical significance of the correlation (p-value) are displayed. B) Relative evolutionary rates (RER) and respective fixed-topology phylogenies for genes related with integrity of the cell membrane (*LEC1*) and heat stress (*PHO23*).

We next assessed the occurrence of positive selection across cactophilic clades using branch-site tests of rates of nonsynonymous (dN) and synonymous substitutions (dS). These tests were performed separately for each of the five cactophilic subclades within the major lineages selected: *Pichia* A and *Pichia* B clades (Pichiales), *Starmera* and *Phaffomyces* clades (Phaffomycetales), and *Tortispora* clade (Trigonopsidales) (Fig. S3). Only genes for which no evidence of positive selection was found in the non-cactophilic sister clades (p-value < 0.05) were considered (Fig. S3). With this conservative approach, we found evidence for positive selection for 292 / 2,384 genes examined in *Tortispora*; 328 / 2,175 in *Pichia* A; 259 / 2,155 in *Pichia* B; 90 / 1,685 in *Phaffomyces*; and 105/1,602 in *Starmera* (Table S6). Importantly, signatures of selection (significantly higher  $\omega$  values than 1 in branches leading to cactophilic clades) can stem from either high dN or low dS values<sup>91</sup>, and we did not specifically distinguish between the two scenarios. Genes under positive selection showed limited overlap; six genes were under positive selection in three clades and 134 genes in two clades, while no genes presented evidence for positive selection in all five clades.

While the distribution of presence / absence of the trehalase gene *NTH1* was not associated with the low prevalence of trehalose assimilation in cactophilic species (Fig. 3C), *NTH1* was among the genes under positive selection in the two *Pichia* clades (Table S6). The 140 genes with evidence of positive selection in two or three clades were enriched in multiple biological functions including chemotropism, translation and splicing, carnitine metabolism, lactate oxidation, and ergosterol biosynthesis (Table S7). More specifically, 6/140 genes were involved in several steps of ergosterol biosynthesis (*ERG1*, *ERG8*, *ERG24*, *ERG26*, *UPC2*, *SIP3*) (Fig. 6). *ERG13* was also under positive selection in the *Starmera* clade. Ergosterol is involved in



stabilising cell membranes during heat stress and therefore has a major role in the tolerance to numerous stresses in fungi<sup>68,92-94</sup>.

Other genes were specifically involved in cell wall biosynthesis and integrity (Table S6). For instance, *CDA2*, which encodes a chitin deacetylase involved in the function of the fungal cell wall<sup>95,96</sup>, was under positive selection in the stem branches of three distinct cactophilic clades (Fig. 6) and was duplicated in the *Tortispora* clade (Fig. 6).

Other cell wall-related functions were found among the group of genes that showed evidence for positive selection in two or more clades, namely chitin synthases *CHS1*, *CHS2*, and *CHS3* and chitin-related genes *CHS5* (involved in Chs3 transport from the Golgi to the plasma membrane) and *CSH7* (involved the export of Chs3 from the endoplasmic reticulum) and *UTR2* (chitin transglycosylase). *DCWI* (encoding a mannosidase required for cell wall formation and resistance to high temperatures)<sup>97</sup>, *STE7* (encoding a signal transducing MAP kinase involved in cell wall integrity and pseudohyphal growth)<sup>98</sup>, and *AYRI* (encoding a bifunctional triacylglycerol lipase involved in cell wall biosynthesis)<sup>99</sup>, were also found among the genes under positive selection in two cactophilic clades. Mutations in these genes have been associated with cell wall defects and increased heat sensitivity in *S. cerevisiae*<sup>100-103</sup> and differential transcriptional responses to heat stress have also been documented for some of these genes and functions<sup>104</sup>.



**Fig. 6. Signatures of convergent molecular evolution across cactophilic clades.** Genes and functions for which detection of distinctive evolutionary alterations (positive selection, duplication, and HGT) in more than one cactophilic clade are highlighted. The distinct evolutionary events are represented by squares filled with different colors, as indicated in the key. Schematic representations of the fungal cell envelope and plant cell wall are shown as many of the genes with signatures of convergence are associated with functions impacting these two structures.

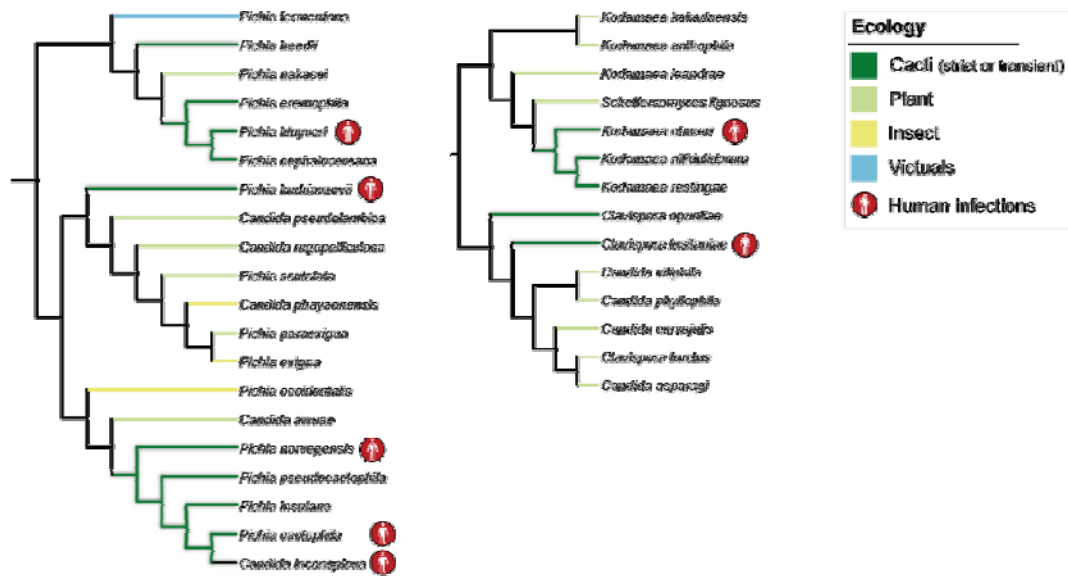
### **Some genes under positive selection in cactophilic clades show evidence of codon optimization**

To infer the transcriptional activity of cactophilic yeasts, we determined gene-wise relative synonymous codon usage (gw-RSCU), a metric which measures biases in codon usage that have been shown to be associated with expression level<sup>105</sup>, and examined the top-ranked genes (95<sup>th</sup> percentile) in cactophilic species (Table S8). Top-ranked genes include many encoding ribosomal subunits and histones, which are known to be highly expressed and codon-optimized in *S. cerevisiae*<sup>106</sup>. We found that the chitin deacetylase gene *CDA2* and genes involved in ergosterol biosynthesis (namely *ERG2*, *ERG5*, *ERG6*, and *ERG11*) were among the genes that fell within the 95<sup>th</sup> percentile rank for gw-RSCU in multiple cactophilic species. To ascertain whether these genes also show signatures of codon optimization in closely related non-cactophilic species, we determined their respective gw-RSCU percentile ranks. While no clear pattern was observed for *ERG* genes (these genes were also highly ranked for gw-RSCU in non-cactophilic species), we observed that *CDA2* is particularly highly ranked in *Phaffomyces*, *Starmera*, and *Pichia* clades compared to their closest relative non-cactophilic species (Fig. S4). *CDA2* showed evidence for positive selection in both *Pichia* clades and *Starmera*, suggesting that distinctive synonymous and/or nonsynonymous might have resulted from translational selection for optimized codons due to higher gene expression.

### **Cactophily as a launching pad for the emergence of opportunistic human pathogens?**

Thermotolerance is a key shared trait by human fungal pathogens<sup>27,107-109</sup>. Interestingly, several cactophilic or closely related species are emerging human opportunistic pathogens (Fig. 7, Fig. S5). Examples include *Candida inconspicua* and *Pichia norvegensis*<sup>110</sup>, which cluster within the *Pichia cactophila* clade (Fig. 7), and *Pichia cactophila*, which was also isolated from human tissue<sup>111</sup>. Cases of fungemia have also been associated with *Pichia kluyveri*<sup>112</sup>, a transient species belonging to a separate clade within the Pichiales. *Kodamaea ohmeri*, which has also been isolated from the cacti environment<sup>52</sup> and is closely related to the cactophilic *Ko. nitidulidarum* and *Ko. restingae*, is also an emerging human pathogen with a significant mortality rate<sup>113-115</sup>.

In addition to thermotolerance, other aspects of the cactophilic lifestyle might be pre-adaptations for human pathogenicity. For instance, the sterol composition of the cell envelope has been implicated in fungal virulence<sup>94,116-118</sup>. Mutations in genes involved in the ergosterol biosynthetic pathway are associated with antifungal resistance<sup>118-121</sup>, making it a major target of antifungal drugs<sup>122</sup>. We showed that the evolutionary rates of several genes involved in the ergosterol biosynthesis across multiple cactophilic clades have changed, suggesting that this pathway might be under a new selection regime in these lineages. However, the impact of these alterations remains to be elucidated.



**Fig. 7. Example cactophilic lineages that contain human opportunistic pathogens.** Ecological associations of cactophilic species and their closest relatives are represented, highlighting examples of species associated with human infections. Ecological information for additional cactophilic species and closest relatives is provided in Fig. S5.

## Discussion

By examining high-throughput genomic, phenotypic, and ecological data for 1,049 yeast species we unveiled multiple (~17) independent occurrences of cacti association. The ability to grow at  $\geq 37^{\circ}\text{C}$  emerged as the strongest predictor of the cactophilic lifestyle. Being a polygenic trait, thermotolerance can arise through multiple distinct evolutionary trajectories<sup>123</sup>. Heat stress generally affects protein folding and cell integrity and involves a complex response from multiple genes impacting the expression of heat shock proteins, the integrity of the cell wall and membranes, production of compatible solutes, repression of protein biosynthesis, and/or temporary interruption of the cell cycle<sup>67</sup>. The expression of genes involved in cell wall biosynthesis and integrity is for instance affected when strains of *S. cerevisiae* are exposed to heat stress<sup>104</sup>. Possibly in association with thermotolerance, we found genes involved in maintaining the cell envelope exhibiting evidence of positive selection, codon optimization, and duplication in multiple cactophilic clades.

We also found genomic fingerprints that indicate phenotypic convergence in the ability to feed on plant material. Acquisition and duplication of plant cell wall-degrading enzymes can be interpreted as adaptive and supports the involvement of cactophilic yeasts in the cacti necrosis process<sup>40</sup>. Interestingly, the contrasting mechanisms employed, HGT of a bacterial pectin lyase and duplication of cellulases and a rhamnogalacturonan lyase, show that phenotypic convergence can arise through disparate molecular mechanisms. While we found possible cases of phenotypic and molecular convergence for thermotolerance (Scenario I, Fig. 2) and cases of possible phenotypic convergence through distinct molecular mechanisms in respect to plant cell wall degrading ability (Scenario II, Fig. 2), we found limited evidence for convergence at both phenotypic and genomic levels (Scenario III, Fig. 2). Based on these data, we conclude that cactophily can originate through multiple phenotypic and genetic changes, some commonly found and some more rare.

It was previously postulated that most cactophilic yeasts evolved from ancestors associated with plants<sup>37</sup> (Fig. S5), indicating that the ability to thrive in plant-related environments was already

present in the ancestors of many of the lineages that evolved cactophilic lifestyle. However, cacti are native to arid and semiarid climates in the Americas<sup>124</sup>, where extremely high (and low) temperatures and low humidity constitute crucial challenges. This reasoning aligns with our finding that thermotolerance is the phenotypic feature with the strongest signature of adaptation to cacti.

Compared to mammals, a lineage that has provided spectacular examples of parallel molecular evolution underpinning the independent emergence of convergent traits<sup>5,8,64,125-127</sup>, yeasts exhibit far higher levels of genetic and physiological diversity<sup>26</sup>. Consequently, the probability of finding overlapping evolutionary paths might be reduced as pleiotropic effects or mutational epistasis might be more prominent across divergent genetic backgrounds<sup>9,10</sup>. It is also possible that collection of additional phenotypic data or evolutionary genomic analyses will revise our understanding of the nature of convergent evolution to cactophily. This caveat notwithstanding, these results present a first snapshot of the study of convergent evolution of an ecological trait in yeasts, employing multiple state-of-the-art methodologies that aim at looking into a wide range of evolutionary mechanisms, phenotypes, and genetic determinants. It also underlines the exceptional value of combining high throughput physiological, genomic, and ecological data to investigate still-pressing questions in evolutionary biology.

## **Materials and Methods**

### **Species selection**

Yeast species associated were selected from Opulente & LaBella et al.<sup>51</sup> by cross referencing ecological information available from the literature<sup>33,35-37,39,52,54,55,58-60,63,128,129</sup>. Two different groups of cactophilic species were determined according to their degree of association with cacti: strictly cactophilic (species that are mainly isolated from cacti and very rarely isolated from other environments) or transiently cactophilic (species frequently found in cacti, but also frequently found in other environments or for which strong association with cacti was not clear from the literature) (Table S1). Species very rarely isolated from this environment were not considered as they could either represent misidentifications or have originated from stochastic events.

### **Inference and dating of cacti association events in the Saccharomycotina**

The number of independent events of cacti association were inferred by performing an ancestral state reconstruction using a continuous-time Markov model for discrete trait evolution implemented in Mr Bayes<sup>130</sup>. A simplified workflow implemented in the R environment was followed<sup>131,132</sup>. Only species classified as strictly cactophilic (Table S1) were considered as exhibiting the trait (cactophily). Transient species were considered as not having the trait. The estimated times for the emergence of cactophily were inferred according to a relaxed molecular clock analyses of the subphylum Saccharomycotina<sup>51</sup>.

### **Machine learning**

To assess whether cactophilic species can be classified based on physiological and/or genomic traits, we used physiological data (for 893/1,154 strains) and genomic data (functional KEGG annotations) for 1,154 yeast strains<sup>51</sup>. Both strictly and transiently cactophilic species were used and classified as “1” (meaning having the trait). All the remaining species in the dataset were classified as “0” (meaning lacking the trait). We trained a machine learning algorithm built by an XGBoost (1.7.3) random forest classifier (XGBRFClassifier()) with the parameters “

max\_depth=12, n\_estimators=100, use\_label\_encoder=False, eval\_metric='mlogloss', n\_jobs = 8" on 90% of the data, and we used the remaining 10% for cross-validation, using RepeatedStratifiedKFold from sklearn.model\_selection (1.2.1)<sup>133,134</sup>. We used RepeatedStratifiedKFold to generate accuracy measures. We used the cross\_val\_predict() function from Sci-Kit Learn to generate the confusion matrixes; these matrices show the numbers of strains correctly predicted to be cactophilic or non cactophilic (True Positives and True Negatives, respectively) and incorrectly predicted (False Positives, which are predicted to be cactophilic but are not; and False Negatives, which are not predicted to be, but are cactophilic). Top features were automatically generated by the XGBRFClassifier using Gini importance, which uses node impurity (the amount of variance in growth on a given carbon source for strains that either have or do not have this trait/feature). This process was repeated for 20 runs with 52 (or 31 for the analysis excluding transiently cactophilic species) randomly selected non-cactophilic species for each run, and then the averages of each result were used in the final confusion matrixes and feature importance graphs.

### **Gene family evolution analyses**

To find genes that are specific to or have expanded in cactophilic clades, orthogroup assignment was performed with Orthofinder v.2.3.8<sup>135</sup> using an inflation parameter of 1.5 and DIAMOND v2.0.13.151<sup>136</sup> as the sequence aligner. Due to the high phylogenetic distance between the major strictly cactophilic clades, in order to optimize the number of orthogroups correctly assigned, this analysis was performed separately for each cactophilic group (LDT group, Pichiales and Phaffomycetales) (Table S3). Closely related non-cactophilic species were included according to the previously reported phylogeny in Opulente & LaBella, *et. al.*, 2023<sup>51</sup>. Species in which more than 30% of the genes were multi-copy were discarded. Gene family evolutionary history was inferred using GeneRax v1.1.0<sup>71</sup>, which incorporates a Maximum-likelihood and species-tree-aware method. For that, orthogroups containing more than ten sequences were aligned with MAFFT v7.402 using an iterative refinement method (L-INS-i)<sup>137</sup>. Pruned species trees for each dataset were obtained from the main Saccharomycotina tree<sup>51</sup>.



using PHYKIT v 1.11.12<sup>138</sup>. The species trees, and alignments were subsequently used as inputs in GeneRax. Briefly, the UndatedDTL probabilistic model was used to compute the reconciliation likelihood that accounts for duplications, transfers, and losses. For simplification, the same model of sequence evolution for all gene families (LG+I+G4) during gene tree inference by GeneRax. Reconciled trees were visualized with Notung v2.9<sup>139</sup>.

Phylogenetic trees were constructed for candidate genes/gene families putatively relevant for niche adaptation.

### **Evolutionary rates**

To determine which genes might exhibit altered evolutionary rates in cactophilic clades/species, we used both branch-site tests of positive selection using codeml implemented in PAML<sup>140</sup> and convergent evolutionary rates analyses implemented in RER converge<sup>84</sup>. For PAML, only strictly cactophilic species were considered and only clades containing three or more cactophilic species were included. Remaining cactophilic (strictly or transient) species were excluded. In this way, five datasets (see Fig. S3) were considered including members of different lineages (two subclades within the Phaffomycetales: *Starmera* and *Phaffomyces*; two within the Pichiales: *Pichia* A and *Pichia* B; and one within Trigonopsidales: *Tortispora*). In *Tortispora* we also included *T. agaves* because, despite not being associated with Cactaceae species, it is associated with plants with similar characteristics (*Agave* spp.)<sup>33</sup>. For all these five datasets, closely related species belonging to each of the families were included based on the species phylogeny based on<sup>51</sup> (Fig. S3). Next, selection of orthogroups using Orthofinder v.2.3.8 was performed for all the species included in the five datasets. Clustering of sequences was based on protein sequence similarity and calculated using DIAMOND v2.0.13.151 using an inflation parameter of 1.5. Single copy orthologs (SCOs) present in all species in each dataset were selected. To avoid masking nonparallel signatures of positive selection, each dataset was separately analysed in PAML<sup>140</sup> using the branch-site model<sup>141</sup> and considering the branch leading to the cactophilic clade as the foreground branch. The likelihood of a gene being under

positive selection was evaluated through a likelihood ratio test [LRT:  $2x(\ln_{HI}-\ln_{HO})$ ] <sup>141</sup> using a p-value threshold of  $<0.001$  (chi-square test with one degree of freedom; critical value = 10.8276). To avoid cases where genes under positive selection in the foreground branch were also under positive selection in closely related species (background branches), branch-site tests were performed in the same way but considering the sister clade as the foreground branch (Fig. S3) Genes for which a significant signal for positive selection was also obtained for the sister clade (p-value threshold of  $<0.05$ ) were excluded. In this way, a conservative approach was employed, only considering genes for which a strong signal for positive selection was found specifically in the cactophilic branches. We further examined which genes were under positive selection in more than one clade. We assessed whether convergent amino acid substitutions occurred in genes that showed evidence of positive selection in three distinct clades (six in total: *CDA2*, *BDF1*, *PET112*, *GRS1*, *CCT3* and *SPN1*) using PCOC <sup>142</sup>.

To further investigate fingerprints of convergence in evolutionary rates, we used RER converge with an extended dataset, in order to include more than one cactophilic clade/species per dataset, so that convergence could be tested (Table S3). In this case, SCOs from the Orthofinder run performed for the gene family evolution analysis were used. To increase the number of orthologs available for analysis, multi-copy orthogroups that were present in at least two species belonging to distinct cactophilic clades within each dataset were selected. Next, SCOs from each multi-copy orthogroup were pruned using OrthoSNAP v0.0.1 <sup>143</sup>. Next, multiple sequence alignments were produced for each multi-copy orthogroup using MAFFT v7.402 (-- localpair) and phylogenies were obtained with FastTree <sup>144</sup>. We next ran OrthoSNAP with default parameters, keeping at least 50% of the species from the original dataset (50% occupancy). For each orthogroup, branch lengths were estimated on a fixed topology obtained from the Saccharomycotina species tree <sup>51</sup> by pruning the species of interest using PHYKIT v 1.11.12 <sup>138</sup>. Each orthogroup was first aligned with MAFFT v7.402 (-- localpair), and the best-fitting model was assessed using IQ-TREE v2.0.6. Branch lengths were determined for each orthogroup, in the fixed tree topology, using RAxML-NG v.0.9.0 <sup>145</sup> under the best protein models inferred by

IQ-TREE. All phylogenies were further analysed with RER converge to find evidence of convergent evolutionary rates in cactophilic species included in each dataset.

Briefly, we tested the hypothesis of convergent evolutionary rates in the ancestral branches leading to the cactophilic clades and/or species. Only phylogenies including a minimum of two foreground species and ten species in total were considered. Genes for which a correlation ratio (Rho- correlation between relative evolutionary rate and trait) higher than 0.25 and a  $p$ -value (association between relative evolutionary rate and trait) lower than 0.05 were obtained were further considered as good candidates for being under convergent accelerated evolution. For those, original trees were manually checked. For Phaffomycetales, we exceptionally considered  $\rho > 0.15$  because we failed to find genes with  $\rho > 0.25$ .

A detailed scheme of the entire workflow can be found in Fig. S6.

### **Enrichment analyses of genes under positive selection**

Gene ontology (GO) enrichment analyses were performed for the genes under positive selection in each clade ( $N_{Starmera}=105$ ,  $N_{Phaffomyces}=90$ ,  $N_{PichiaA}= 328$ ,  $N_{PichiaB}= 259$ ,  $N_{Tortispora}=293$ ) and for the genes under positive selection in two or more clades ( $N=140$ ).

First, associated GO terms were obtained for all genes using eggNOG-mapper<sup>146</sup>. For in-clade analyses, enrichment analyses were performed using *P. cactophila* (for Pichia A and PichiaB datasets), *Starmera amethionina* (for *Starmera* dataset), *Phaffomyces opuntiae* (for *Phaffomyces* dataset), and *T. caseinolytica* (for *Tortispora* dataset) whole genome annotations as the background. For the analyses involving genes under positive selection in two or more clades, *P. cactophila* genome annotations were used as the background. GO enrichment analyses were performed using the R package topGO 2.28.0<sup>147</sup>. Statistical significance was assessed using Fisher's exact test using the “classic”, “elim”, and default “weight01” methods. Correction for multiple testing was performed the “BH” correction method. The results can be assessed in Table S7.

## Codon Usage Bias

To examine codon optimization in particular genes of cactophilic species, we calculated the gene-wise relative synonymous codon usage (gw-RSCU), implemented in Biokit v0.0.9<sup>105</sup>. This metric was shown to correlate with the tRNA adaptation index (tAI)<sup>105</sup>, which measures the translation efficiency by considering both codon optimization and the intracellular concentration of tRNA molecules<sup>148</sup>. The gw-RSCU was calculated by determining the mean relative synonymous codon usage value for all codons in each gene in the genome based on their genome-wide RSCU values. We ranked the genes with the highest gw-RSCU values (subtracting the standard deviation to the gw-RSCU mean value) and looked at the genes falling into the 95<sup>th</sup> percentile and above (Table S8). Next, gene functions that were relevant for the cactophilic lifestyle were selected, and their gw-RSCU values were inspected in non-cactophilic closest related species. Briefly, a local BLASTp was used to find the putative orthologs by considering a protein identity of > 40%. The top hit was considered to correspond to the orthologous gene; however, whenever multiple hits with similar protein sequence identity were found, the one with the highest rank was considered. The percentile ranking for *CDA2* was determined using the R package *dplyr*<sup>149</sup>.

## References

- 1 Feigin, C. Y. *et al.* Convergent deployment of ancestral functions during the evolution of mammalian flight membranes. *Science Advances* **9**, eade7511, doi:10.1126/sciadv.ade7511.
- 2 Bennici, A. The convergent evolution in plants. *Rivista di biologia* **96**, 485-489 (2003).
- 3 Laehnemann, D. *et al.* Genomics of Rapid Adaptation to Antibiotics: Convergent Evolution and Scalable Sequence Amplification. *Genome Biology and Evolution* **6**, 1287-1301, doi:10.1093/gbe/evu106 (2014).
- 4 Li, G. *et al.* The hearing gene Prestin reunites echolocating bats. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13959-13964, doi:10.1073/pnas.0802097105 (2008).
- 5 Potter, J. H. T. *et al.* Nectar-feeding bats and birds show parallel molecular adaptations in sugar metabolism enzymes. *Current Biology* **31**, 4667-4674.e4666, doi:<https://doi.org/10.1016/j.cub.2021.08.018> (2021).
- 6 Yeaman, S. *et al.* Convergent local adaptation to climate in distantly related conifers. *Science (New York, N.Y.)* **353**, 1431-1433, doi:doi:10.1126/science.aaf7812 (2016).

- 7 Rokas, A. & Carroll, S. B. Frequent and widespread parallel evolution of protein sequences. *Molecular biology and evolution* **25**, 1943-1953, doi:10.1093/molbev/msn143 (2008).
- 8 Liu, Y. *et al.* Convergent sequence evolution between echolocating bats and dolphins. *Current Biology* **20**, R53-R54, doi:<https://doi.org/10.1016/j.cub.2009.11.058> (2010).
- 9 Sackton, T. B. & Clark, N. Convergent evolution in the genomics era: new insights and directions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20190102, doi:10.1098/rstb.2019.0102 (2019).
- 10 Natarajan, C. *et al.* Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science (New York, N.Y.)* **354**, 336-339, doi:10.1126/science.aaf9070 (2016).
- 11 Schrider, D. R., Hahn, M. W. & Begun, D. J. Parallel Evolution of Copy-Number Variation across Continents in *Drosophila melanogaster*. *Molecular biology and evolution* **33**, 1308-1316, doi:10.1093/molbev/msw014 (2016).
- 12 Xu, S., Wang, J., Guo, Z., He, Z. & Shi, S. Genomic Convergence in the Adaptation to Extreme Environments. *Plant communications* **1**, 100117, doi:10.1016/j.xplc.2020.100117 (2020).
- 13 Hughes, A. L. & Friedman, R. Parallel evolution by gene duplication in the genomes of two unicellular fungi. *Genome research* **13**, 794-799, doi:10.1101/gr.714603 (2003).
- 14 Borges, R. *et al.* Gene loss, adaptive evolution and the co-evolution of plumage coloration genes with opsins in birds. *BMC genomics* **16**, 751, doi:10.1186/s12864-015-1924-3 (2015).
- 15 Sharma, V. *et al.* A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nature Communications* **9**, 1215, doi:10.1038/s41467-018-03667-1 (2018).
- 16 Kohler, A. *et al.* Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature genetics* **47**, 410-415, doi:10.1038/ng.3223 (2015).
- 17 Savory, F., Leonard, G. & Richards, T. A. The role of horizontal gene transfer in the evolution of the oomycetes. *PLoS Pathog* **11**, e1004805-e1004805, doi:10.1371/journal.ppat.1004805 (2015).
- 18 Lewis, W. H. *et al.* Convergent Evolution of Hydrogenosomes from Mitochondria by Gene Transfer and Loss. *Molecular biology and evolution* **37**, 524-539, doi:10.1093/molbev/msz239 (2019).
- 19 Partha, R., Kowalczyk, A., Clark, N. L. & Chikina, M. Robust Method for Detecting Convergent Shifts in Evolutionary Rates. *Molecular biology and evolution* **36**, 1817-1830, doi:10.1093/molbev/msz107 (2019).
- 20 Rosenblum, E. B., Parent, C. E. & Brandt, E. E. The Molecular Basis of Phenotypic Convergence. *Annual Review of Ecology, Evolution, and Systematics* **45**, 203-226, doi:10.1146/annurev-ecolsys-120213-091851 (2014).
- 21 Kowalczyk, A., Partha, R., Clark, N. L. & Chikina, M. Pan-mammalian analysis of molecular constraints underlying extended lifespan. *eLife* **9**, e51089, doi:10.7554/eLife.51089 (2020).
- 22 Partha, R. *et al.* Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *eLife* **6**, e25884, doi:10.7554/eLife.25884 (2017).
- 23 Storz, J. F. Causes of molecular convergence and parallelism in protein evolution. *Nature reviews. Genetics* **17**, 239-250, doi:10.1038/nrg.2016.11 (2016).
- 24 Miton, C. M. & Tokuriki, N. How mutational epistasis impairs predictability in protein evolution and design. *Protein science : a publication of the Protein Society* **25**, 1260-1272, doi:10.1002/pro.2876 (2016).

- 25 Li, Y. *et al.* A genome-scale phylogeny of the kingdom Fungi. *Current Biology* **31**, 1653-1665.e1655, doi:<https://doi.org/10.1016/j.cub.2021.01.074> (2021).
- 26 Shen, X. X. *et al.* Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* **175**, 1533-1545.e1520, doi:10.1016/j.cell.2018.10.023 (2018).
- 27 Rokas, A. Evolution of the human pathogenic lifestyle in fungi. *Nature Microbiology* **7**, 607-619, doi:10.1038/s41564-022-01112-0 (2022).
- 28 Gonçalves, C., Coelho, M. A., Salema-Oom, M. & Gonçalves, P. Stepwise Functional Evolution in a Fungal Sugar Transporter Family. *Molecular biology and evolution* **33**, 352-366, doi:10.1093/molbev/msv220 (2016).
- 29 Li, Y. *et al.* HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell* **185**, 2975-2987.e2910, doi:10.1016/j.cell.2022.06.014 (2022).
- 30 Moran, N. A. & Jarvik, T. Lateral Transfer of Genes from Fungi Underlies Carotenoid Production in Aphids. *Science (New York, N.Y.)* **328**, 624-627, doi:doi:10.1126/science.1187113 (2010).
- 31 Starmer, W. T. & Lachance, M.-A. in *The Yeasts (Fifth Edition)* (eds Cletus P. Kurtzman, Jack W. Fell, & Teun Boekhout) 65-83 (Elsevier, 2011).
- 32 Kurtzman, C. P. in *The Yeasts: a taxonomic study* (ed Fell Kurtzman C.P., J.W, Boekhout, T.) Ch. 72, 817-822 (Elsevier, 2011).
- 33 Lachance, M.-A. & Kurtzman, C. P. The yeast genus *Tortispora* gen. nov., description of *Tortispora ganteri* sp. nov., *Tortispora mauiana* f.a., sp. nov., *Tortispora agaves* f.a., sp. nov., *Tortispora sangerardonensis* f.a., sp. nov., *Tortispora cuajiniquilana* f.a., sp. nov., *Tortispora starmeri* f.a., sp. nov. and *Tortispora phaffii* f.a., sp. nov., reassignment of *Candida caseinolytica* to *Tortispora caseinolytica* f.a., comb. nov., emendation of *Botryozyma*, and assignment of *Botryozyma*, *Tortispora* gen. nov. and *Trigonopsis* to the family *Trigonopsidaceae* fam. nov. *International journal of systematic and evolutionary microbiology* **63**, 3104-3114, doi:<https://doi.org/10.1099/ijs.0.052209-0> (2013).
- 34 Barker, J. S. F., Starmer, W. T. & Program, U. S. A. C.-o. S. *Ecological Genetics and Evolution: The Cactus-yeast-Drosophila Model System*. (Academic Press, 1982).
- 35 Ganter, P. F., Morais, P. B. & Rosa, C. A. in *Yeasts in Natural Ecosystems: Diversity* (eds Pietro Buzzini, Marc-André Lachance, & Andrey Yurkov) 225-264 (Springer International Publishing, 2017).
- 36 Starmer, W. T., Fogleman, J. C. & Lachance, M.-A. in *Microbial Ecology of Leaves*. (eds John H. Andrews & Susan S. Hirano) 158-178 (Springer New York).
- 37 Starmer, W. T., Schmedicke, R. A. & Lachance, M.-A. The origin of the cactus-yeast community. *FEMS yeast research* **3**, 441-448, doi:10.1016/S1567-1356(03)00056-4 (2003).
- 38 Lachance, M.-A. & Starmer, W. The community concept and problems of non-trivial characterization of yeast communities. *Coenosis* **1**, 21-28 (1986).
- 39 Starmer, W. T., Aberdeen, V. & Lachance, M.-A. in *Biodiversity and Ecophysiology of Yeasts* (eds Gábor Péter & Carlos Rosa) 485-499 (Springer Berlin Heidelberg, 2006).
- 40 Fogleman, J. C. & Danielson, P. B. Chemical Interactions in the Cactus-Microorganism-Drosophila Model System of the Sonoran Desert1. *American Zoologist* **41**, 877-889, doi:10.1093/icb/41.4.877 (2001).
- 41 Oliveira, D. C. *et al.* Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophila repleta* species group. *Molecular phylogenetics and evolution* **64**, 533-544, doi:10.1016/j.ympev.2012.05.012 (2012).
- 42 Kircher, H. W., Heed, W. B., Russell, J. S. & Grove, J. Senita cactus alkaloids: Their significance to Sonoran Desert *Drosophila* ecology. *Journal of Insect Physiology* **13**, 1869-1874, doi:[https://doi.org/10.1016/0022-1910\(67\)90023-6](https://doi.org/10.1016/0022-1910(67)90023-6) (1967).
- 43 Etges, W. J. Evolutionary genomics of host plant adaptation: insights from *Drosophila*. *Current opinion in insect science* **36**, 96-102, doi:10.1016/j.cois.2019.08.011 (2019).

- 44 Fogleman, J. C., Danielson, P. B. & Macintyre, R. J. in *Evolutionary Biology* (eds Max K. Hecht, Ross J. Macintyre, & Michael T. Clegg) 15-77 (Springer US, 1998).
- 45 Guillén, Y. *et al.* Genomics of Ecological Adaptation in Cactophilic *Drosophila*. *Genome Biology and Evolution* **7**, 349-366, doi:10.1093/gbe/evu291 (2015).
- 46 López-Olmos, K., Markow, T. A. & Machado, C. A. Evolution of GSTD1 in Cactophilic *Drosophila*. *Journal of molecular evolution* **84**, 285-294, doi:10.1007/s00239-017-9798-4 (2017).
- 47 Rane, R. V. *et al.* Detoxification Genes Differ Between Cactus-, Fruit-, and Flower-Feeding *Drosophila*. *Journal of Heredity* **110**, 80-91, doi:10.1093/jhered/esy058 (2019).
- 48 Rane, R. V. *et al.* Genomic changes associated with adaptation to arid environments in cactophilic *Drosophila* species. *BMC genomics* **20**, 52, doi:10.1186/s12864-018-5413-3 (2019).
- 49 Matzkin, L. M. The molecular basis of host adaptation in cactophilic *Drosophila*: molecular evolution of a glutathione S-transferase gene (GstD1) in *Drosophila mojavensis*. *Genetics* **178**, 1073-1083, doi:10.1534/genetics.107.083287 (2008).
- 50 Moreyra, N. N. *et al.* Phylogenomics provides insights into the evolution of cactophily and host plant shifts in *Drosophila*. *Molecular phylogenetics and evolution* **178**, 107653, doi:<https://doi.org/10.1016/j.ympev.2022.107653> (2023).
- 51 Opulente, D., A. *et al.* Genomic and ecological factors shaping specialism and generalism across an entire subphylum. *bioRxiv*, 2023.2006.2019.545611, doi:10.1101/2023.06.19.545611 (2023).
- 52 Freitas, L. F. D. *et al.* Yeast communities associated with cacti in Brazil and the description of *Kluyveromyces starmeri* sp. nov. based on phylogenomic analyses. *Yeast (Chichester, England)* **37**, 625-637, doi:<https://doi.org/10.1002/yea.3528> (2020).
- 53 Holzschu, D. L., Phaff, H. J., Tredick, J. & Hedgecock, D. *Pichia pseudocactophila*, a new species of yeast occurring in necrotic tissue of columnar cacti in the North American Sonoran Desert. *Canadian journal of microbiology* **29**, 1314-1322, doi:10.1139/m83-205 (1983).
- 54 MIRANDA, M., HOLZSCHU, D. L., PHAFF, H. J. & STARMER, W. T. *Pichia mexicana*, a New Heterothallic Yeast from Cereoid Cacti in the North American Sonoran Desert. *International journal of systematic and evolutionary microbiology* **32**, 101-107, doi:<https://doi.org/10.1099/00207713-32-1-101> (1982).
- 55 Starmer, W. T., Phaff, H. J., Ganter, P. F. & Lachance, M. A. *Candida orba* sp. nov., a new cactus-specific yeast species from Queensland, Australia. *International journal of systematic and evolutionary microbiology* **51**, 699-705, doi:10.1099/00207713-51-2-699 (2001).
- 56 STARMER, W. T., PHAFF, H. J., MIRANDA, M. & MILLER, M. W. *Pichia cactophila*, a New Species of Yeast Found in Decaying Tissue of Cacti. *International journal of systematic and evolutionary microbiology* **28**, 318-325, doi:<https://doi.org/10.1099/00207713-28-2-318> (1978).
- 57 Barker, J. S. F., East, P. D., Phaff, H. J. & Miranda, M. The ecology of the yeast flora in necrotic *Opuntia* cacti and of associated *Drosophila* in Australia. *Microb Ecol* **10**, 379-399, doi:10.1007/BF02015562 (1984).
- 58 Ganter, P. F., Cardinali, G. & Boundy-Mills, K. *Pichia insulana* sp. nov., a novel cactophilic yeast from the Caribbean. *International journal of systematic and evolutionary microbiology* **60**, 1001-1007, doi:10.1099/ijs.0.014258-0 (2010).
- 59 Lachance, M. A. *et al.* *Wickerhamiella australiensis*, *Wickerhamiella cacticola*, *Wickerhamiella occidentalis*, *Candida drosophilae* and *Candida lipophila*, five new related yeast species from flowers and associated insects. *International journal of systematic bacteriology* **48 Pt 4**, 1431-1443, doi:10.1099/00207713-48-4-1431 (1998).
- 60 Lachance, M.-A. & Starmer, W. T. *Kurtzmaniella* gen. nov. and description of the heterothallic, haplontic yeast species *Kurtzmaniella cleridarum* sp. nov., the

- teleomorph of *Candida cleridarum*. *International journal of systematic and evolutionary microbiology* **58**, 520-524, doi:<https://doi.org/10.1099/ijs.0.65460-0> (2008).
- 61 Groenewald, M. *et al.* A genome-informed higher rank classification of the biotechnologically important fungal subphylum Saccharomycotina. *Studies in Mycology* **105**, 1-22, doi:10.3114/sim.2023.105.01 (2023).
- 62 Arakaki, M. *et al.* Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proceedings of the National Academy of Sciences* **108**, 8379-8384, doi:doi:10.1073/pnas.1100628108 (2011).
- 63 Rosa, C. A. *et al.* *Kodamaea nitidulidarum*, *Candida restingae* and *Kodamaea anthophila*, three new related yeast species from ephemeral flowers. *International journal of systematic and evolutionary microbiology* **49**, 309-318, doi:<https://doi.org/10.1099/00207713-49-1-309> (1999).
- 64 Parker, J. *et al.* Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**, 228-231, doi:10.1038/nature12511 (2013).
- 65 Opulente, D. A. *et al.* Factors driving metabolic diversity in the budding yeast subphylum. *BMC biology* **16**, 26, doi:10.1186/s12915-018-0498-3 (2018).
- 66 Stratman, R. & Markow, T. A. Resistance to thermal stress in desert *Drosophila*. *Functional Ecology* **12**, 965-970, doi:<https://doi.org/10.1046/j.1365-2435.1998.00270.x> (1998).
- 67 Pan, D., Wiedemann, N. & Kammerer, B. Heat Stress-Induced Metabolic Remodeling in *Saccharomyces cerevisiae*. *Metabolites* **9**, doi:10.3390/metabo9110266 (2019).
- 68 Swan, T. M. & Watson, K. Stress tolerance in a yeast sterol auxotroph: role of ergosterol, heat shock proteins and trehalose. *FEMS microbiology letters* **169**, 191-197, doi:10.1111/j.1574-6968.1998.tb13317.x (1998).
- 69 Huang, C.-J., Lu, M.-Y., Chang, Y.-W. & Li, W.-H. Experimental Evolution of Yeast for High-Temperature Tolerance. *Molecular biology and evolution* **35**, 1823-1839, doi:10.1093/molbev/msy077 (2018).
- 70 Chen, B.-C. & Lin, H.-Y. Deletion of NTH1 and HSP12 increases the freeze-thaw resistance of baker's yeast in bread dough. *Microbial Cell Factories* **21**, 149, doi:10.1186/s12934-022-01876-4 (2022).
- 71 Morel, B., Kozlov, A. M., Stamatakis, A. & Szöllősi, G. J. GeneRax: A Tool for Species-Tree-Aware Maximum Likelihood-Based Gene Family Tree Inference under Gene Duplication, Transfer, and Loss. *Molecular biology and evolution* **37**, 2763-2774, doi:10.1093/molbev/msaa141 (2020).
- 72 Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H. & Winther, O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics (Oxford, England)* **33**, 3387-3395, doi:10.1093/bioinformatics/btx431 (2017).
- 73 van den Brink, J. & de Vries, R. P. Fungal enzyme sets for plant polysaccharide degradation. *Applied microbiology and biotechnology* **91**, 1477-1492, doi:10.1007/s00253-011-3473-2 (2011).
- 74 da Silva, E. G., de Fátima Borges, M., Medina, C., Piccoli, R. H. & Schwan, R. F. Pectinolytic enzymes secreted by yeasts from tropical fruits. *FEMS yeast research* **5**, 859-865, doi:10.1016/j.femsyr.2005.02.006 (2005).
- 75 Hugouvieux-Cotte-Pattat, N., Condemine, G. & Shevchik, V. E. Bacterial pectate lyases, structural and functional diversity. *Environmental microbiology reports* **6**, 427-440, doi:10.1111/1758-2229.12166 (2014).
- 76 Atanasova, L. *et al.* Evolution and functional characterization of pectate lyase PEL12, a member of a highly expanded *Clonostachys rosea* polysaccharide lyase 1 family. *BMC Microbiology* **18**, 178, doi:10.1186/s12866-018-1310-9 (2018).



- 77 Álvarez-Pérez, S., Lievens, B., Jacquemyn, H. & Herrera, C. M. *Acinetobacter nectaris* sp. nov. and *Acinetobacter boissieri* sp. nov., isolated from floral nectar of wild Mediterranean insect-pollinated plants. *International journal of systematic and evolutionary microbiology* **63**, 1532-1539, doi:10.1099/ij.s.0.043489-0 (2013).
- 78 Timilsina, S. *et al.* Xanthomonas diversity, virulence and plant-pathogen interactions. *Nature Reviews Microbiology* **18**, 415-427, doi:10.1038/s41579-020-0361-8 (2020).
- 79 Reverchon, S., Muskhelishvili, G. & Nasser, W. Virulence Program of a Bacterial Plant Pathogen: The Dickeya Model. *Progress in molecular biology and translational science* **142**, 51-92 (2016).
- 80 Masoud, W. & Jespersen, L. Pectin degrading enzymes in yeasts involved in fermentation of Coffea arabica in East Africa. *International journal of food microbiology* **110**, 291-296, doi:10.1016/j.ijfoodmicro.2006.04.030 (2006).
- 81 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular biology and evolution* **35**, 518-522, doi:10.1093/molbev/msx281 (2018).
- 82 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. *Molecular biology and evolution* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).
- 83 Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* **37**, 420-423, doi:10.1038/s41587-019-0036-z (2019).
- 84 Kowalczyk, A. *et al.* RERconverge: an R package for associating evolutionary rates with convergent traits. *Bioinformatics (Oxford, England)* **35**, 4815-4817, doi:10.1093/bioinformatics/btz468 (2019).
- 85 Wong, E. D., Karra, K., Hitz, B. C., Hong, E. L. & Cherry, J. M. The YeastGenome app: the Saccharomyces Genome Database at your fingertips. *Database : the journal of biological databases and curation* **2013**, bat004, doi:10.1093/database/bat004 (2013).
- 86 Ruiz-Roig, C., Viéitez, C., Posas, F. & de Nadal, E. The Rpd3L HDAC complex is essential for the heat stress response in yeast. *Molecular microbiology* **76**, 1049-1062, doi:10.1111/j.1365-2958.2010.07167.x (2010).
- 87 Córcoles-Sáez, I., Ballester-Tomas, L., de la Torre-Ruiz, M. A., Prieto, J. A. & Randez-Gil, F. Low temperature highlights the functional role of the cell wall integrity pathway in the regulation of growth in Saccharomyces cerevisiae. *The Biochemical journal* **446**, 477-488, doi:10.1042/bj20120634 (2012).
- 88 Castro, I. G. *et al.* Systematic analysis of membrane contact sites in Saccharomyces cerevisiae uncovers modulators of cellular lipid distribution. *Elife* **11**, doi:10.7554/eLife.74602 (2022).
- 89 Auesukaree, C. *et al.* Genome-wide identification of genes involved in tolerance to various environmental stresses in Saccharomyces cerevisiae. *Journal of applied genetics* **50**, 301-310, doi:10.1007/bf03195688 (2009).
- 90 Amatruda, J. F., Gattermeir, D. J., Karpova, T. S. & Cooper, J. A. Effects of null mutations and overexpression of capping protein on morphogenesis, actin distribution and polarized secretion in yeast. *The Journal of cell biology* **119**, 1151-1162, doi:10.1083/jcb.119.5.1151 (1992).
- 91 Álvarez-Carretero, S., Kapli, P. & Yang, Z. Beginner's Guide on the Use of PAML to Detect Positive Selection. *Molecular biology and evolution* **40**, msad041, doi:10.1093/molbev/msad041 (2023).
- 92 Caspeta, L. *et al.* Altered sterol composition renders yeast thermotolerant. *Science (New York, N.Y.)* **346**, 75-78 (2014).
- 93 Kamthan, A., Kamthan, M. & Datta, A. Expression of C-5 sterol desaturase from an edible mushroom in fisson yeast enhances its ethanol and thermotolerance. *PLOS ONE* **12**, e0173381, doi:10.1371/journal.pone.0173381 (2017).

- 94 Oliveira, F. F. M. *et al.* Erg6 affects membrane composition and virulence of the human fungal pathogen *Cryptococcus neoformans*. *Fungal genetics and biology : FG & B* **140**, 103368, doi:10.1016/j.fgb.2020.103368 (2020).
- 95 Christodoulidou, A., Bouriotis, V. & Thireos, G. Two Sporulation-specific Chitin Deacetylase-encoding Genes Are Required for the Ascospore Wall Rigidity of *Saccharomyces cerevisiae*\*. *Journal of Biological Chemistry* **271**, 31420-31425, doi:<https://doi.org/10.1074/jbc.271.49.31420> (1996).
- 96 Mouyna, I. *et al.* What Are the Functions of Chitin Deacetylases in *Aspergillus fumigatus*? *Frontiers in Cellular and Infection Microbiology* **10**, doi:10.3389/fcimb.2020.00028 (2020).
- 97 Kitagaki, H., Ito, K. & Shimoi, H. A temperature-sensitive *dcw1* mutant of *Saccharomyces cerevisiae* is cell cycle arrested with small buds which have aberrant cell walls. *Eukaryotic cell* **3**, 1297-1306, doi:10.1128/ec.3.5.1297-1306.2004 (2004).
- 98 Lesage, G. & Bussey, H. Cell wall assembly in *Saccharomyces cerevisiae*. *Microbiology and molecular biology reviews : MMBR* **70**, 317-343, doi:10.1128/mubr.00038-05 (2006).
- 99 Ahn, K. W. *et al.* Deletion of *GBG1*/*AYR1* Alters Cell Wall Biogenesis in *Saccharomyces cerevisiae*. *Mycobiology* **38**, 102-107, doi:10.4489/myco.2010.38.2.0102 (2010).
- 100 Trautwein, M. *et al.* *Arf1p*, *Chs5p* and the ChAPs are required for export of specialized cargo from the Golgi. *The EMBO journal* **25**, 943-954, doi:10.1038/sj.emboj.7601007 (2006).
- 101 Sinha, H. *et al.* Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. *Genetics* **180**, 1661-1670, doi:10.1534/genetics.108.092932 (2008).
- 102 Lam, K. K. *et al.* Palmitoylation by the DHH protein Pfa4 regulates the ER exit of Chs3. *The Journal of cell biology* **174**, 19-25, doi:10.1083/jcb.200602049 (2006).
- 103 Lesage, G. *et al.* An interactional network of genes involved in chitin synthesis in *Saccharomyces cerevisiae*. *BMC genetics* **6**, 8, doi:10.1186/1471-2156-6-8 (2005).
- 104 Caspeta, L. & Nielsen, J. Thermotolerant Yeast Strains Adapted by Laboratory Evolution Show Trade-Off at Ancestral Temperatures and Preadaptation to Other Stresses. *mBio* **6**, e00431-00415, doi:doi:10.1128/mBio.00431-15 (2015).
- 105 Steenwyk, J. L. *et al.* BioKIT: a versatile toolkit for processing and analyzing diverse types of sequence data. *Genetics* **221**, iyac079, doi:10.1093/genetics/iyac079 (2022).
- 106 Hershberg, R. & Petrov, D. A. General Rules for Optimal Codon Choice. *PLoS genetics* **5**, e1000556, doi:10.1371/journal.pgen.1000556 (2009).
- 107 Bhabhra, R. & Askew, D. S. Thermotolerance and virulence of *Aspergillus fumigatus*: role of the fungal nucleolus. *Medical Mycology* **43**, S87-S93, doi:10.1080/13693780400029486 (2005).
- 108 Bloom, A. L. M. *et al.* Thermotolerance in the pathogen *Cryptococcus neoformans* is linked to antigen masking via mRNA decay-dependent reprogramming. *Nature Communications* **10**, 4950, doi:10.1038/s41467-019-12907-x (2019).
- 109 Robert, V., Cardinali, G. & Casadevall, A. Distribution and impact of yeast thermal tolerance permissive for mammalian infection. *BMC biology* **13**, 18, doi:10.1186/s12915-015-0127-3 (2015).
- 110 Guitard, J. *et al.* Invasive infections due to *Candida norvegensis* and *Candida inconspicua*: report of 12 cases and review of the literature. *Medical Mycology* **51**, 795-799, doi:10.3109/13693786.2013.807444 (2013).
- 111 Czaika, V. *et al.* Epidemiology and changes in patient-related factors from 1997 to 2009 in clinical yeast isolates related to dermatology, gynaecology, and paediatrics. *International journal of microbiology* **2013**, 703905, doi:10.1155/2013/703905 (2013).

- 112 De Greef, F., Hayette, M. P. & Rossi, C. Fungemia caused by *Pichia kluyveri*: An uncommon yeast infection. *New microbes and new infections* **49-50**, 101018, doi:10.1016/j.nmni.2022.101018 (2022).
- 113 Zhou, M., Li, Y., Kudinha, T., Xu, Y. & Liu, Z. *Kodamaea ohmeri* as an Emerging Human Pathogen: A Review and Update. *Frontiers in microbiology* **12**, 736582, doi:10.3389/fmicb.2021.736582 (2021).
- 114 Li, Z.-M. *et al.* Gut-derived fungemia due to *Kodamaea ohmeri* combined with invasive pulmonary aspergillosis: a case report. *BMC Infectious Diseases* **22**, 903, doi:10.1186/s12879-022-07866-6 (2022).
- 115 Zhou, M. *et al.* Identification and antifungal susceptibility profiles of *Kodamaea ohmeri* based on a seven-year multicenter surveillance study. *Infection and drug resistance* **12**, 1657-1664, doi:10.2147/idr.s211033 (2019).
- 116 Alcazar-Fuoli, L. & Mellado, E. Ergosterol biosynthesis in *Aspergillus fumigatus*: its relevance as an antifungal target and role in antifungal drug resistance. *Frontiers in microbiology* **3**, doi:10.3389/fmicb.2012.00439 (2013).
- 117 Choy, H. L., Gaylord, E. A. & Doering, T. L. Ergosterol distribution controls surface structure formation and fungal pathogenicity. *bioRxiv*, doi:10.1101/2023.02.17.528979 (2023).
- 118 Sanglard, D., Ischer, F., Parkinson, T., Falconer, D. & Bille, J. *Candida albicans* mutations in the ergosterol biosynthetic pathway and resistance to several antifungal agents. *Antimicrob Agents Chemother* **47**, 2404-2412, doi:10.1128/aac.47.8.2404-2412.2003 (2003).
- 119 Branco, J. *et al.* Impact of ERG3 mutations and expression of ergosterol genes controlled by UPC2 and NDT80 in *Candida parapsilosis* azole resistance. *Clinical Microbiology and Infection* **23**, 575.e571-575.e578, doi:<https://doi.org/10.1016/j.cmi.2017.02.002> (2017).
- 120 Fisher, M. C. *et al.* Tackling the emerging threat of antifungal resistance to human health. *Nature Reviews Microbiology* **20**, 557-571, doi:10.1038/s41579-022-00720-1 (2022).
- 121 Rhodes, J. *et al.* Population genomics confirms acquisition of drug-resistant *Aspergillus fumigatus* infection by humans from the environment. *Nature Microbiology* **7**, 663-674, doi:10.1038/s41564-022-01091-2 (2022).
- 122 Rodrigues, M. L. The Multifunctional Fungal Ergosterol. *mBio* **9**, doi:10.1128/mBio.01755-18 (2018).
- 123 Lehnen, M., Ebert, B. E. & Blank, L. M. Elevated temperatures do not trigger a conserved metabolic network response among thermotolerant yeasts. *BMC Microbiology* **19**, 100, doi:10.1186/s12866-019-1453-3 (2019).
- 124 Guerrero, P. C., Majure, L. C., Cornejo-Romero, A. & Hernández-Hernández, T. Phylogenetic Relationships and Evolutionary Trends in the Cactus Family. *Journal of Heredity* **110**, 4-21, doi:10.1093/jhered/esy064 (2019).
- 125 Yeager, M. & Hughes, A. L. Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunological reviews* **167**, 45-58, doi:10.1111/j.1600-065x.1999.tb01381.x (1999).
- 126 Bazykin, G. A. *et al.* Extensive parallelism in protein evolution. *Biol Direct* **2**, 20, doi:10.1186/1745-6150-2-20 (2007).
- 127 Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science (New York, N. Y.)* **307**, 1928-1933, doi:10.1126/science.1107239 (2005).
- 128 Heed, W. B., Starmer, W. T., Miranda, M., Miller, M. W. & Phaff, H. J. An Analysis of the Yeast Flora Associated with Cactiphilic *Drosophila* and their Host Plants in the Sonoran Desert and Its Relation to Temperate and Tropical Associations. *Ecology* **57**, 151-160, doi:<https://doi.org/10.2307/1936406> (1976).

- 129 Kurtzman, C. P., Fell, J. W. & Boekhout, T. *The Yeasts: a taxonomic study 5th edition*. (Elsevier Science, 2011).
- 130 Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic biology* **61**, 539-542, doi:10.1093/sysbio/sys029 (2012).
- 131 Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217-223, doi:<https://doi.org/10.1111/j.2041-210X.2011.00169.x> (2012).
- 132 Steven, H. MBASR: Workflow-simplified ancestral state reconstruction of discrete traits with MrBayes in the R environment. *bioRxiv*, 2021.2001.2010.426107, doi:10.1101/2021.01.10.426107 (2021).
- 133 Chen, T. & Guestrin, C. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, San Francisco, California, USA, 2016).
- 134 Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- 135 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238, doi:10.1186/s13059-019-1832-y (2019).
- 136 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12**, 59-60, doi:10.1038/nmeth.3176 (2015).
- 137 Katoh, K. & Standley, D. M. MAFFT: iterative refinement and additional methods. *Methods in molecular biology (Clifton, N.J.)* **1079**, 131-146, doi:10.1007/978-1-62703-646-7\_8 (2014).
- 138 Steenwyk, J. L. *et al.* PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics (Oxford, England)* **37**, 2325-2331, doi:10.1093/bioinformatics/btab096 (2021).
- 139 Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol* **7**, 429-447, doi:10.1089/106652700750050871 (2000).
- 140 Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular biology and evolution* **24**, 1586-1591, doi:10.1093/molbev/msm088 (2007).
- 141 Jeffares, D. C., Tomiczek, B., Sojo, V. & dos Reis, M. A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. *Methods in molecular biology (Clifton, N.J.)* **1201**, 65-90, doi:10.1007/978-1-4939-1438-8\_4 (2015).
- 142 Rey, C., Guéguen, L., Sémon, M. & Boussau, B. Accurate Detection of Convergent Amino-Acid Evolution with PCOC. *Molecular biology and evolution* **35**, 2296-2306, doi:10.1093/molbev/msy114 (2018).
- 143 Steenwyk, J. L. *et al.* OrthoSNAP: A tree splitting and pruning algorithm for retrieving single-copy orthologs from gene family trees. *PLoS biology* **20**, e3001827, doi:10.1371/journal.pbio.3001827 (2022).
- 144 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **5**, e9490, doi:10.1371/journal.pone.0009490 (2010).
- 145 Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics (Oxford, England)* **35**, 4453-4455, doi:10.1093/bioinformatics/btz305 (2019).
- 146 Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**, D309-d314, doi:10.1093/nar/gky1085 (2019).

- 147 Alexa, A. & Rahnenführer, J. Gene set enrichment analysis with topGO. *Bioconductor Improv* **27**, 1-26 (2009).
- 148 Sabi, R., Volvovitch Daniel, R. & Tuller, T. stAlcalc: tRNA adaptation index calculator based on species-specific weights. *Bioinformatics (Oxford, England)* **33**, 589-591, doi:10.1093/bioinformatics/btw647 (2017).
- 149 Wickham, H. & François, R. *dplyr: A Grammar of Data Manipulation*. (2014).