



Article

Comparison of Phylogenetic Tree Topologies for Nitrogen Associated Genes Partially Reconstruct the Evolutionary History of *Saccharomyces cerevisiae*

Manuel Villalobos-Cid ¹, Francisco Salinas ^{2,3,4}, Eduardo I. Kessi-Pérez ^{2,5},
Matteo De Chiara ⁶, Gianni Liti ⁶, Mario Inostroza-Ponta ¹ and Claudio Martínez ^{2,5,*}

¹ Departamento de Ingeniería Informática, Facultad de Ingeniería, Universidad de Santiago de Chile (USACH), Santiago 9170022, Chile

² Centro de Estudios en Ciencia y Tecnología de los Alimentos (CECTA), Universidad de Santiago de Chile (USACH), Santiago 9170022, Chile

³ Millennium Institute for Integrative Biology (iBio), Santiago 7500574, Chile

⁴ Instituto de Bioquímica y Microbiología, Facultad de Ciencias, Universidad Austral de Chile (UACH), Valdivia 5110566, Chile

⁵ Departamento de Ciencia y Tecnología de los Alimentos, Universidad de Santiago de Chile (USACH), Santiago 9170201, Chile

⁶ Université Côte d'Azur, CNRS, INSERM, IRCAN, 06107 Nice, France

* Correspondence: claudio.martinez@usach.cl

Received: 14 November 2019; Accepted: 19 December 2019; Published: 23 December 2019



Abstract: Massive sequencing projects executed in *Saccharomyces cerevisiae* have revealed in detail its population structure. The recent “1002 yeast genomes project” has become the most complete catalogue of yeast genetic diversity and a powerful resource to analyse the evolutionary history of genes affecting specific phenotypes. In this work, we selected 22 nitrogen associated genes and analysed the sequence information from the 1011 strains of the “1002 yeast genomes project”. We constructed a total evidence (TE) phylogenetic tree using concatenated information, which showed a 27% topology similarity with the reference (REF) tree of the “1002 yeast genomes project”. We also generated individual phylogenetic trees for each gene and compared their topologies, identifying genes with similar topologies (suggesting a shared evolutionary history). Furthermore, we pruned the constructed phylogenetic trees to compare the REF tree topology versus the TE tree and the individual genes trees, considering each phylogenetic cluster/subcluster within the population, observing genes with cluster/subcluster topologies of high similarity to the REF tree. Finally, we used the pruned versions of the phylogenetic trees to compare four strains considered as representatives of *S. cerevisiae* clean lineages, observing for 15 genes that its cluster topologies match 100% the REF tree, supporting that these strains represent main lineages of yeast population. Altogether, our results showed the potential of tree topologies comparison for exploring the evolutionary history of a specific group of genes.

Keywords: *Saccharomyces cerevisiae*; nitrogen associated genes; evolutionary history; tree topologies; clean lineages; representative strains

1. Introduction

The yeast *Saccharomyces cerevisiae* (hereinafter, called “*S. cerevisiae*” or simply “yeast”) is a microorganism with wide biotechnological applications, able to conduct the alcoholic fermentation in the production of many food and beverages, particularly wine [1,2]. Moreover, *S. cerevisiae* is a workhorse for molecular biology and genetic studies, being the first eukaryotic genome to be

fully sequenced [3]. Currently, the advances in sequencing technologies have allowed massive sequencing projects, making available a tremendous amount of genomic information, which covered from microorganisms to plants and human genomes [4–6]. In this sense, the sequencing information accumulated so far has unveiled *S. cerevisiae* genome content and population structure [7], and also clarified the relationship between the different species of the *Saccharomyces* genus [8–10].

The first attempts to study the genetic diversity of *S. cerevisiae* were using molecular markers, such as PFGE [11,12], mitochondrial DNA digestion (RFLP-mtDNA) [13] or microsatellites [14,15]. These approaches showed a strong correlation between the molecular patterns and the geographic isolation of the yeast strains, suggesting an ecological relationship between isolates [16,17]. Similarly, the first sequencing approaches were focused on individual genes, where the sequenced information was concatenated to infer phylogeny, revealing the presence of two groups of domesticated *S. cerevisiae* strains (grape wine and sake wine strains) in comparison of isolates from nature [18].

Afterwards, the genome sequencing of 36 yeast isolates from diverse ecological origins unveiled yeast population structure and demonstrated the presence of five clean lineages (defined in terms of presenting unique private SNPs not shared between lineages) within the species: Wine/European (WE), West African (WA), North American (NA), Sake (SA) and Malaysian (MA) [19]. Subsequent massive sequencing projects have confirmed these results, including the 100 yeast genomes project [20] and the 1002 yeast genomes project [21]. This last project represents the most complete catalogue of yeast genetic diversity so far, expanding the number of phylogenetic clusters initially observed in the species to 26 clades [19,21].

After the initial insights into yeast population structure [19], five strains were considered as representatives of the first five clean lineages identified: DBVPG6765 (WE strain) for WE cluster, DBVPG6044 (WA strain) for WA cluster, YPS128 (NA strain) for NA cluster, Y12 (SA strain) for SA cluster and UWOPS03.461.4 (MA strain) for MA cluster. Between them, the MA strain has been seldom utilised in genetic studies, due to its reproductive isolation respect to the other four representative strains [22,23]. Thus, this set of four strains has become a powerful tool for disentangling the genetic basis of quantitative traits in yeasts, allowing to map the causative alleles of phenotypic variation [23,24]. In this context, the WE, WA, NA and SA strains have been utilised as founder (parental) strains of recombinant yeast populations, which have permitted the mapping of QTLs (Quantitative Trait Loci) for multiple phenotypes [25–30].

Given that fermentative phenotypes are in general quantitative traits, different strains have been used to map QTLs related to the fermentation process, such as: fermentation rate [31], aroma compounds production [32] and nitrogen consumption [33–35]. Among these phenotypes, our group has systematically focused its interest on nitrogen associated phenotypes, due to the importance of nitrogen sources for the fermentation process, being its deficiencies the principal cause of stuck and sluggish fermentations (reviewed in [36,37]). In this sense, we have mapped multiple QTLs related to nitrogen consumption using yeast populations derived from the four representative strains (WE, WA, NA and SA), validating the specific causative genes by reciprocal hemizyosity approaches [35,38–40]. However, it is not fully understood which fraction of the genetic diversity observed in the species is represented by the alleles mapped in these QTL experiments, making necessary performing bioinformatic analyses to assess these contributions.

One bioinformatic approach is phylogenetic inference, which attempts to reconstruct a hypothesis that explains the evolutionary relationships between a group of species, strains, genes or proteins. The historical pattern of speciation and divergence allows classifying life according to an evolutionary schema which usually is represented as a phylogenetic tree [41]. For instance, the reconstruction of the tree of life represents the evolutionary relationship between millions of species [42]. Phylogenetic inference has been used in almost every branch of biology, such as botany [43], zoology [44], palaeontology [45] and pharmacology [46], among other fields [47]. Furthermore, phylogeny has been applied to describe the relationships between paralogous gene families, understanding the evolution and epidemiological dynamics of pathogens, studying cell differentiation in cancer and other diseases,

identification of gene function, prediction of the protein tertiary structure, metagenomic sequences classification, and reconstruction of ancestral genomes [48].

Despite the advances in molecular phylogenetic reconstruction methods, the inference process always involves some uncertainty respect to the true historical relationships of the organisms, and phylogeny may include incongruences (i.e., conflicting topologies of the trees for a same taxa) related to analytical or biological factors [41,49]. The analytical factors include biases associated to (i) the selection of the inference criterion, (ii) taxon sampling, and (iii) specific assumptions in the modelling of sequence evolution. Biological factors are related to (i) the biological evidence used to infer phylogeny, (ii) the presence of reticular evolution phenomena (e.g., horizontal gene transfer, incomplete lineage sorting, gene duplication, hybridisation, recombination, among others), (iii) stochastic errors or character sampling biases related to the length of the genes, and (iv) systematic errors due to the presence of noise in the dataset [50]. One way of addressing the biases produced by the mentioned factors is dealing with the problem using a multi objective optimisation approach [51]. However, a complete phylogenetic study requires a comparative analysis to reduce the uncertainty associated to the inference process, understanding the sources of incongruence, and assessing their effects on the resultant tree topologies [50].

In this work, we evaluated the potential of phylogenetic tree topologies comparison for evolutionary history reconstruction of a given set of genes in *S. cerevisiae*. For a proof of concept, we selected 22 genes related to nitrogen consumption phenotypes under fermentation conditions, all of them previously mapped by linkage analyses. For this set of genes, we analysed the sequence information from 1011 yeast strains that were part of the “1002 yeast genomes project”. Utilising the concatenated sequences of these genes, we partially reconstructed the evolutionary history of the species, obtaining a phylogenetic tree with 27% of topology similarity respect to the reference tree (REF) described by [21] (using whole genome sequencing). Additionally, we generated individual phylogenetic trees for each selected gene and compared the tree topologies among them, observing genes with similar tree topologies and suggesting a similar evolutionary history between them. Finally, we pruned the constructed phylogenetic trees to compare topologies between each phylogenetic cluster/subcluster within the population or between a specific subset of strains. Altogether, our results are a proof of concept that shows how tree topologies comparison can be used to explore the global evolutionary history of the species, opening the possibility of a wider evaluation, encompassing the entire *S. cerevisiae* genome.

2. Materials and Methods

2.1. Gene Selection and Genomic Information Obtention

We selected 22 genes involved in nitrogen associated phenotypes during the fermentation process, identified by QTL mapping and validated by reciprocal hemizyosity analyses (i.e., with a demonstrated effect over the phenotype) [26,35,38–40,52]. Importantly, most of the selected genes were mapped using yeast populations where the founder strains are the DBVPG6765 (WE), DBVPG6044 (WA), YPS128 (NA) and Y12 (SA) strains [22,23]. The list of genes selected, including their functional information, QTL detected and bibliographic reference, is shown in Supplementary Table S1.

The genomic information for the 22 selected genes was obtained from the “1002 yeast genome project” [21]. A VCF (Variant Calling Format) file containing the genotype information of the 22 selected genes was transformed into linear sequences utilising a custom script (please see next section). The DBVPG6765 (WE) strain was not included in the “1002 yeast genomes project” and its sequencing information for the 22 selected genes was extracted from SGRP2 Blast server (<http://www.moseslab.csb.utoronto.ca/sgrp/>) [53].

2.2. Bioinformatic Analyses

Computational experiments were performed using R version 3.3.2 and RStudio version 0.99.491 in an Intel (R) Core (TM) i7-3930K CPU 3.20 GHz, 6 cores, 16 GB RAM and 2TB. Computational experiments consider three stages: (i) data processing, (ii) phylogenetic inference, and (iii) comparison analysis.

2.2.1. Data Processing and Phylogenetic Inference

A custom script based on the *vcfR* package v1.8.0 [54] was used to process the 22 gene sequence variations for 1011 strains stored as a VCF file. For each gene, its sequence was compared to the reference genome and produced a multiple sequence alignment using the information of the 1011 strains, which was stored as an individual FASTA file. Then, an integrated phylogenetic tree (Total Evidence tree, TE) was inferred using the direct concatenation of the sequences for these 22 genes stored as different files, according to the total evidence paradigm [55]. This paradigm, also called “combined analysis” or “congruence approach”, combines all the data sets associated to diverse biological evidence before the inference of a phylogenetic tree, for example, by concatenating sequences of multiple genes [56]. This is considered a good alternative when the data sets are statistically congruent (i.e., similar evolutionary rates or non-conflicting phylogenetic trees) [49].

The TE tree was constructed using the Neighbor-joining strategy, following the same procedure described by [21], by applying the *bionj* algorithm available in the *phangorn* R package v2.5.5 [57]. This strategy was used to infer 22 phylogenetic trees which represent the individual evolutionary history of each gene. In addition, the congruence of these individual trees was calculated by using the Congruence Among Distance Matrix (CADM) metric [49]. This metric computes a coefficient of concordance W ($0 \leq W \leq 1$), where 0 represents a complete disagreement between the input trees and 1 means complete agreement.

2.2.2. Comparison of the Phylogenetic Tree Topologies

A quantitative comparison between phylogenetic trees was performed. To do this, a similarity matrix of trees was built by calculating the percentage of clades in common (PC) [58], $PC = 100 \times \left[1 - \frac{RF}{(2m-6)}\right]$, where the *RF* is the topological Robinson-Foulds distance between trees and *m* is the number of strains. The *RF* distance corresponds to the minimum number of editions (merging or splitting nodes) necessary to transform one unrooted tree into a second tree [59]. Then, the Multidimensional Scaling method, available in the *treeSpace* R package v1.1.3.1 [60], was used to graphically compare the tree topologies by exploring the landscape of phylogenetic trees. Two types of comparisons were performed; first, the topologies of all the trees produced were compared: the TE tree, the REF described by [21], the individual trees related to the 22 selected genes, and a randomly created topology (OUT). The OUT tree was inferred to be contrasted with the REF tree, in order to assure that the differences (or similarities) between trees are produced by the feature of the data instead of noise or bias associated to the inference methods.

Then, all trees generated (TE, REF, OUT and individual trees for each gene) were pruned using the *phangorn* R package to perform specific comparisons between (i) the subtrees (cluster/subcluster) associated to the clades defined by [21] and (ii) the representative strains of the clean lineages WE, WA, NA and SA [19]. The *RF* metric compares trees inferred using the same set of elements (strains), thus, the tree topologies coming from the DBVPG6765 (WE) strain were not compared respect to the REF tree, since this strain was not included by [21].

3. Results and Discussion

3.1. Reconstructing the Population Structure of *S. cerevisiae* Using a Subset of Nitrogen Associated Genes

Initially, we selected 22 genes associated with the fermentation capacity (fermentation kinetic measured as CO₂ release) and nitrogen consumption (ammonium or amino acids) phenotypes under wine fermentation conditions. These genes were mapped using linkage analysis approaches (QTL mapping) and all of them were validated by reciprocal hemizygoty analyses, confirming their effects over the analysed phenotypes (Supplementary Table S1). Furthermore, the 22 selected genes were identified utilising yeast populations derived from four strains (WE, NA, WA and SA strains) representative of previously described yeast clean lineages [19,23]. Therefore, it is expected that the selected genes represent alleles coming from the main phylogenetic clusters observed in the *S. cerevisiae* population, encompassing a great fraction of the genetic diversity of the species [19].

In order to assess how much of the evolutionary history of the species can be reconstructed with this subset of 22 genes, we used the genomic information from the “1002 yeast genome project” [21]. This allowed us to analyse the genetic information for the 22 selected genes in 1011 different yeast strains isolated from diverse ecological origins. We concatenated the sequence information of these 22 genes and performed a phylogenetic inference to construct a TE tree (Figure 1A). The CADM metric was 0.6, meaning that the evolutionary histories of the genes studied are congruent, supporting the use of the total evidence paradigm in the phylogenetic analysis according to the definition used by [49]. Then, we compared the structure of the TE tree respect to the REF tree described by [21] using whole genome sequencing (Figure 1B), observing in both trees a similar topology with the same set of clades and subclades (Figure 1). We quantified the topology differences observed between the TE and REF trees using the topological Robinson-Foulds distance, showing a 27% of similarity between trees (Figure 2A), an interesting new result with no bibliographic antecedent which marks a comparison value for future studies of the same type. Altogether, these results showed that partial genomic information from nitrogen associated genes partially reconstructed (27%) the evolutionary history of the species. This result clearly contrasts with the obtained for the OUT tree, a randomly created topology that showed no similarity with TE and REF trees (Figure 2A).

3.2. Comparison of Tree Topologies among Genes Revealed Similar Evolutionary Histories

Afterwards, we generated individual trees for each selected gene using the information from the 1011 strains, and then we compared the tree topologies by means of Robinson-Foulds distance among them using as controls the TE and REF trees, in addition to an output (OUT) tree randomly generated (Figure 2A). In general, we observed groups of genes with similar tree topologies, highlighting the tree topologies of *GCN1*, *MDS3* and *RIM15*, which also are the closer ones to the TE and REF trees (Figure 2B). One possible explanation is that these genes have the higher contribution to the topologies of the TE and REF trees, which would imply a stronger contribution in the population structure of the species (Figure 2). However, is important to note that the length and/or conservation degree of the genes under study may introduce some bias. Since we used only the polymorphic sites present in each gene (and not the total length of each one), a potential bias may come only from the conservation degree (understood as the number of polymorphic sites), and this could be evaluated only for the 22 genes used in this proof of concept. A more wide and detailed study, e.g., using all *S. cerevisiae* genes, is necessary to gain a better understanding at this respect.

Other strong association is observed between *EAP1* and *TOR2*, both genes participating of the TORC1 signalling pathway (Figure 2). Interestingly, molecular diversity among the WE, NA, WA and SA strains for TORC1 activation have been recently revealed [61]. However, other TORC1 pathway associated genes (*GTR1*, *NPR1*, *SAP185*, *SCH9* and *SIT4*) showed more different tree topologies, with a disperse localisation in the tree space plot representation in 2D (Figure 2B). Overall, our results confirmed that comparison of tree topologies among different genes is a useful tool to identify genes with similar evolutionary histories.

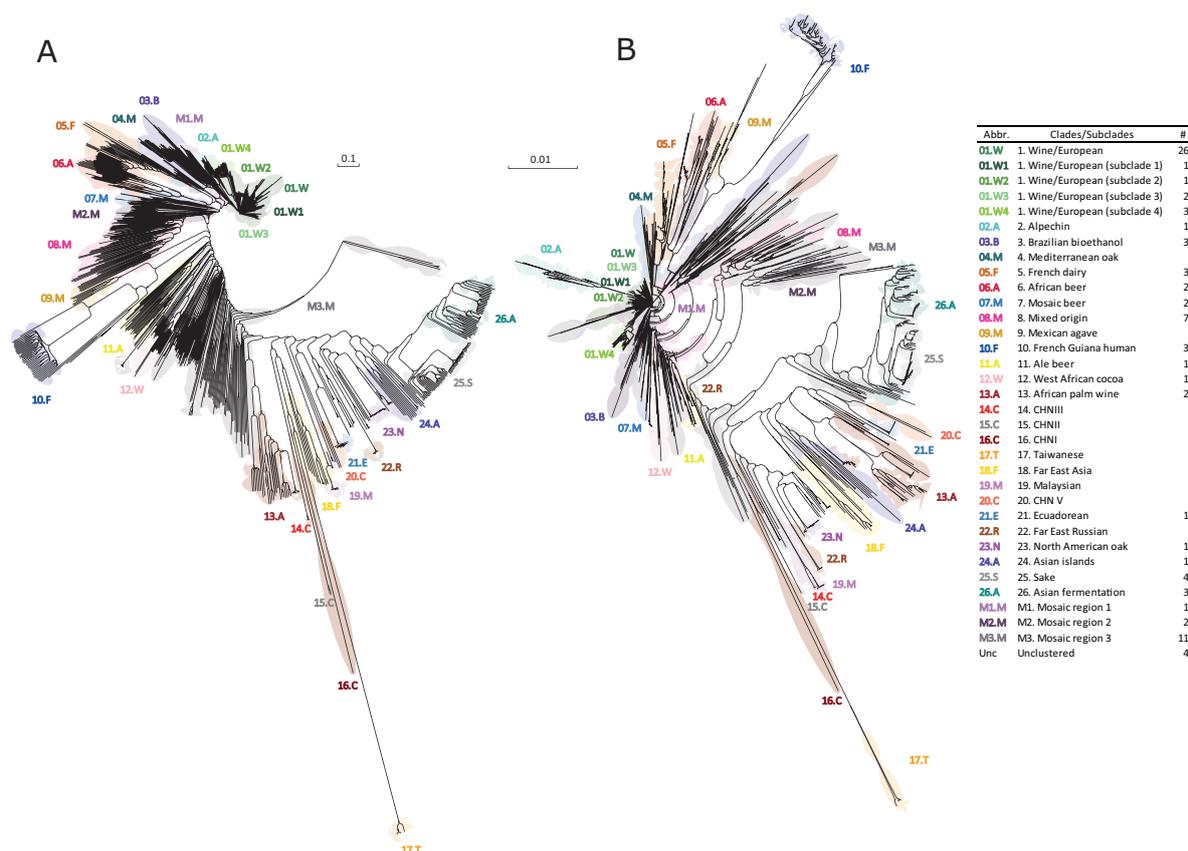


Figure 1. Population structure of *S. cerevisiae* obtained with total and partial genomic information. (A) Neighbor-joining tree described by [21] using the whole genome sequencing information of 1011 yeast strains. This tree was considered as the Reference (REF) tree. (B) Neighbor-joining tree obtained with the concatenated information of the 22 selected genes in 1011 yeast strains. This tree was considered as the Total Evidence (TE) tree. In both trees the phylogenetic clusters and subclusters are represented by the same colour code.

We further analysed our data set (22 genes in 1011 strains) comparing the tree topologies within each cluster and subcluster described by [21] (REF tree). For this, we pruned all the trees constructed (individual gene, TE and REF trees), allowing us to compare topologies between clades with the same elements (tips or strains). After the pruning, we compared the topology of each cluster and subcluster observed in the REF tree respect to the topology of the cluster and subcluster present in individual gene trees and TE tree (Supplementary Table S2). For some clusters, we observed genes with an apparently strong contribution on its evolutionary history (Supplementary Table S2). This is because, while all genes have their own contribution to the evolutionary history of the species, some of them have a discordant contribution. For instance, for the African beer cluster (that includes 20 strains), the *LYP1* gene reconstructs 59% of the cluster topology (cluster evolutionary history) in the REF tree, meaning that the topology of the African beer cluster for *LYP1* gene is 59% identical to the African beer cluster observed in the REF tree (Supplementary Table S2). Interestingly, for the Wine/European cluster (that includes 268 strains), the TE tree including information for the 22 genes studied reconstructs 25% of the cluster topology in the REF tree (Supplementary Table S2; see subclade 0 and TE column). This result suggests that our topology comparison may be biased by the number of strains present in each cluster/subcluster (Supplementary Table S2), where a low number of strains within a cluster/subcluster increases the probability of inferring the same tree topology. For example, while using four strains it is possible to reconstruct 3 unrooted and 15 rooted trees, increasing the number of strains to six increases the number of possible topologies to 105 unrooted and 945 rooted trees. This bias was reduced by applying the PC metric which includes the normalised

RF scores (see Section 2.2.2). Altogether, our results spotted genes with a possible strong contribution to the cluster/subcluster topology observed in the REF tree. Importantly, the OUT tree (Figure 2 and Supplementary Table S2), included as a control, results different to all the other tree without clades/subclades in common.

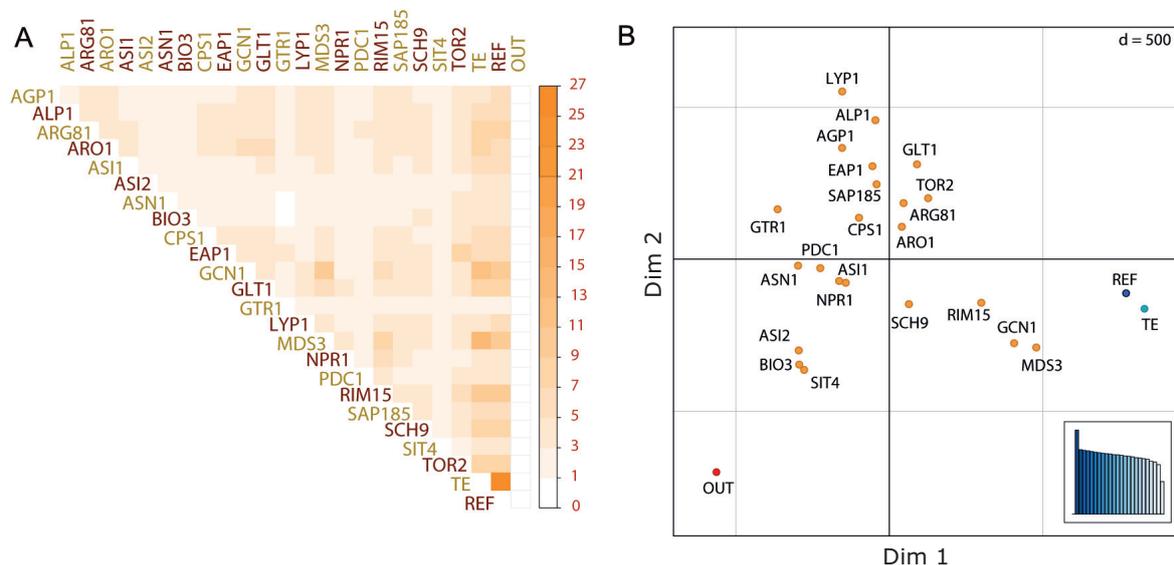


Figure 2. Comparison of tree topologies for nitrogen associated genes. **(A)** Similarity matrix comparing tree topologies between genes. The REF and TE trees were included as controls, in addition to the randomly generated OUT tree. The colour-scale goes between the minimum (0%, white) and maximum (27%, dark orange) similarities observed between individual trees. **(B)** Bi-dimensional representation of the tree spaces obtained from the topology comparison. The multidimensional scaling was performed by using the *Smacof* R package [62].

3.3. Representative Strains from Clean Lineages Reconstruct the Evolutionary History of the Species

Finally, we compared tree topologies in a subset of strains which have been considered as representative of clean lineages within the species (WE, WA, SA and NA) since the first insights into yeast population structure [19]. With now a larger number of strains that have been sequenced and the identification of a total of 26 clades [21], the question of whether these four strains still fully encompass the sequence variation observed in these genes across the entire *S. cerevisiae* population arises. We used again the pruned versions of the constructed trees (individual gene, TE and REF trees) to compare topologies between trees that contain the same elements, but in this case, we compared the topology of the REF tree respect to the topology present in the individual gene trees and TE tree considering only four representative strains of these clean lineages (Supplementary Table S3).

For this, we initially selected four strains generally used as representatives of these *S. cerevisiae* clean lineages: DBVPG6765 (WE), DBVPG6044 (WA), YPS128 (NA) and Y12 (SA) [19,23]. However, since topology comparison among trees requires the same elements, the DBVPG6765 strain was replaced in the TE tree by the DBVPG1106 strain, a genetically similar strain also belonging to the Wine/European cluster. This was due to the absence of the DBVP6765 strain in the “1002 yeast genome project” (Supplementary Figure S1). Then, we proceed with the topology comparison among trees only considering the information from these strains (Table 1 and Supplementary Table S3). We observed for the TE tree and 15 individual gene trees that their topologies were 100% identical to the REF tree (Table 1 and Supplementary Table S3), suggesting that information for these four representative strains reflects the global evolutionary history of the species. We performed as a control the same analysis using the information for the 22 genes of six strains isolated from Chile by our lab group that belong to Wine/European cluster (subclade 0; Supplementary Table S2), and that were included in the “1002 yeasts genome project”. The results showed for all the genes evaluated a 0% of identity

respect the REF tree (Table 1 and Supplementary Table S3), suggesting that these strains are not able to reconstruct the evolutionary history observed in the REF tree, which is an expected result since these strains belong to the same clade and are probably very similar genetically. Altogether, our results confirmed the ability of these four representative strains (DBVPG1106, DBVPG6044, K12 and YPS128) to reconstruct the global evolutionary history of the species, an astounding idea considering that the number of clusters have expanded up to 26 [21]. Nevertheless, other complementary approaches (e.g., analysis of non-synonymous SNPs across the “1002 yeast genomes project” strains) could help to better support this idea.

Table 1. Tree topologies comparison considering representative strains (DBVPG1106, DBVPG6044, K12 and YPS128) and Chilean strains. N: Number of strains. TE: Total Evidence phylogenetic tree. REF: Reference phylogenetic tree. OUT: Randomly generated tree.

Strain/Origin	N	TE	REF	OUT
Others	1001	27	100	0
Chile	6	0	100	0
Representatives of clean lineages	4	100	100	0

Since we replaced in tree topologies comparison the DBVPG6765 strain by the DBVPG1106 strain as representative of the Wine/European cluster, we included the DBVPG6765 strain in the TE tree using the information from the SGRP database (Figure 3A). The DBVPG6765 strain is grouped with the Wine/European clade but seems to have a greater phylogenetic distance: this is probably due to the inclusion of the *RIM15* gene within the analysis, since the DBVPG6765 strain carries a unique allele with polymorphisms not present in other strains of its clade [26,53]. Afterwards, we localised the representative strains in the 2D tree space plot, observing that DBVPG6765, DBVPG6044, K12 and YPS128 strains were grouped together, being separated from the DBVPG1106 strain and distant to other strains (Figure 3B and Supplementary Figure S1). This suggests that DBVPG6765, DBVPG6044, K12 and YPS128 strains have a similar phylogenetic distance between them (Figure 3B), supporting its use as representative strains of clean lineages in *S. cerevisiae*, which in turn validate all previous efforts using this set of four strains to perform QTL mapping studies. Interestingly, this conclusion echoes some previous results obtained using a similar approach but at a smaller scale [63].

In conclusion, the comparison of tree topologies for nitrogen associated genes allowed us to partially reconstruct the evolutionary history of the species and identify genes with similar evolutionary trajectories. We also compared the topologies for each phylogenetic cluster or subcluster present in the REF tree, observing genes with cluster/subcluster topologies of high similarity to the REF tree. Finally, we compared tree topologies for a subset of representative strains from clean lineages present in the *S. cerevisiae* population structure, showing that topologies for the TE tree and 15 individual gene trees match in 100% the REF tree topology, supporting the idea that these representative strains reflect the global population structure of the species. Overall, we have shown the potential to assess the evolutionary history of the species by combining tree topologies comparison and the “1002 yeast genomes project” information, which in turn may lead to a more thorough exploration of *S. cerevisiae* evolution at a genomic level. Analysis of genes from other functional groups would help deepen in the usefulness of tree topology comparison to infer similar evolutionary histories for genes associated with a particular cellular function and, in this context, this proof of concept opens the possibility to perform similar analysis using the whole *S. cerevisiae* genome, and then examine the evolutionary patterns of multiple genes with different cellular functions, confirming or rejecting that genes with similar functions have similar evolutionary histories.

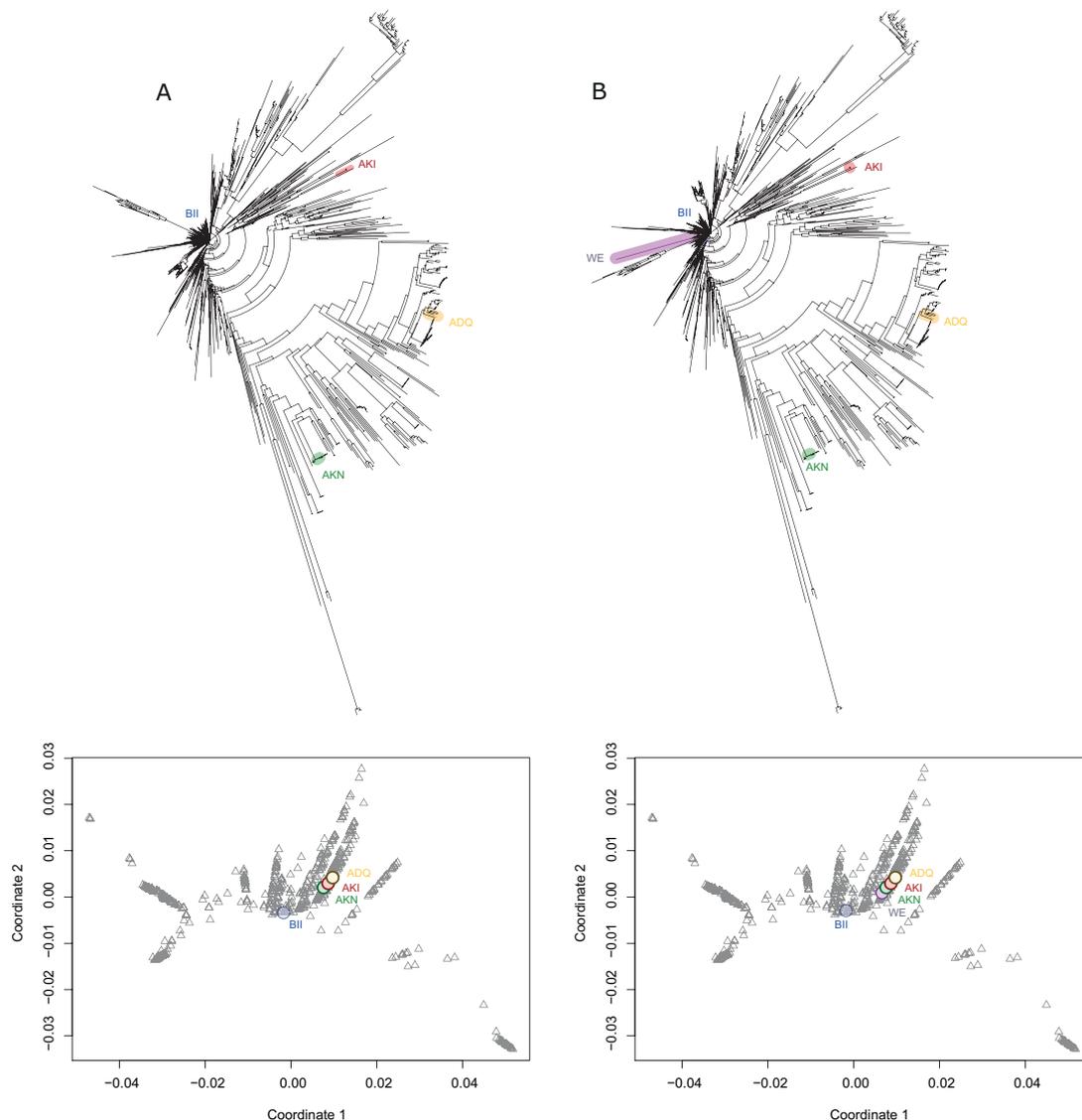


Figure 3. Representative yeast strains in the context of the population structure of the species. (A) TE tree highlighting the position of four strains (BII: DBVPG1106; AKI: DBVPG6044; ADQ: K12 and AKN: YPS128), which are representative of four yeast clean lineages. (B) TE tree considering also the DBVPG6765 (WE) strain. Below each TE tree bi-dimensional tree space representation is showed, obtained from the phylogenetic distance matrix among strains. The positions of the representative strains are highlighted.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2076-2607/8/1/32/s1>.

Author Contributions: Conceptualization, M.V.-C. and F.S.; methodology, M.V.-C. and F.S.; software, M.V.-C. and M.D.C.; validation, M.V.-C.; formal analysis, M.V.-C.; investigation, M.V.-C., F.S. and E.I.K.-P.; resources, G.L., M.I.-P. and C.M.; data curation, M.V.-C. and M.D.C.; writing—original draft preparation, M.V.-C., F.S. and E.I.K.-P.; writing—review and editing, M.V.-C., F.S. and E.I.K.-P.; visualization, M.V.-C., F.S. and E.I.K.-P.; supervision, F.S. and C.M.; project administration, F.S. and C.M.; funding acquisition, G.L., M.I.-P. and C.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Universidad de Santiago de Chile [grant USA1799-VRIDEI 081998SS-PAP] to E.I.K.-P.; CONICYT/PCI [grant REDI170239], CONICYT/FONDECYT [grant 11170158] and Instituto Milenio iBio - Iniciativa Científica Milenio MINECON to F.S.; CONICYT/FONDECYT [grant 1150522] to C.M.; and DICYT-VRIDEI [grant USACH 061819IP] to M.V.-C. and M.I.-P.

Acknowledgments: We thank Angela Contreras for her initial work on VCF files.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Querol, A.; Fernández-Espinar, M.; del Olmo, M.; Barrio, E. Adaptive evolution of wine yeast. *Int. J. Food Microbiol.* **2003**, *86*, 3–10. [[CrossRef](#)]
2. Bisson, L. The Biotechnology of Wine Yeast. *Food Biotechnol.* **2004**, *18*, 63–96. [[CrossRef](#)]
3. Goffeau, A.; Barrell, B.; Bussey, H.; Davis, R.; Dujon, B.; Feldmann, H.; Galibert, F.; Hoheisel, J.; Jacq, C.; Johnston, M.; et al. Life with 6000 Genes. *Science* **1996**, *274*, 546, 563–567. [[CrossRef](#)] [[PubMed](#)]
4. Altshuler, D.; Durbin, R.; Abecasis, G.; Bentley, D.; Chakravarti, A.; Clark, A.; Collins, F.; De La Vega, F.; Donnelly, P.; Egholm, M.; et al. A map of human genome variation from population-scale sequencing. *Nature* **2010**, *467*, 1061–1073.
5. Alonso-Blanco, C.; Andrade, J.; Becker, C.; Bemm, F.; Bergelson, J.; Borgwardt, K.M.; Cao, J.; Chae, E.; Dezaan, T.M.; Ding, W.; et al. 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. *Cell* **2016**, *166*, 481–491. [[CrossRef](#)] [[PubMed](#)]
6. Mukherjee, S.; Seshadri, R.; Varghese, N.; Eloe-Fadrosh, E.; Meier-Kolthoff, J.; Göker, M.; Coates, R.C.; Hadjithomas, M.; Pavlopoulos, G.; Paez Espino, D.; et al. 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat. Biotechnol.* **2017**, *35*, 676–683. [[CrossRef](#)] [[PubMed](#)]
7. Liti, G. The fascinating and secret wild life of the budding yeast *S. cerevisiae*. *eLife* **2015**, *4*, e05835. [[CrossRef](#)]
8. Borneman, A.; Pretorius, I. Genomic Insights into the *Saccharomyces sensu stricto* Complex. *Genetics* **2015**, *199*, 281–291. [[CrossRef](#)]
9. Libkind, D.; Hittinger, C.T.; Valerio, E.; Goncalves, C.; Dover, J.; Johnston, M.; Goncalves, P.; Sampaio, J.P. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14539–14544. [[CrossRef](#)]
10. Scannell, D.R.; Zill, O.A.; Rokas, A.; Payen, C.; Dunham, M.J.; Eisen, M.B.; Rine, J.; Johnston, M.; Hittinger, C.T. The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the *Saccharomyces sensu stricto* Genus. *G3(Bethesda)* **2011**, *1*, 11–25. [[CrossRef](#)]
11. Schwartz, D.C.; Cantor, C.R. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **1984**, *37*, 67–75. [[CrossRef](#)]
12. Vezinhet, F.; Blondin, B.; Hallet, J.N. Chromosomal DNA patterns and mitochondrial DNA polymorphism as tools for identification of enological strains of *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **1990**, *32*, 568–571. [[CrossRef](#)]
13. Querol, A.; Barrio, E.; Huerta, T.; Ramon, D. Molecular monitoring of wine fermentations conducted by active dry yeast strains. *Appl. Environ. Microbiol.* **1992**, *58*, 2948–2953. [[PubMed](#)]
14. Baleiras-Couto, M.M.; Eijmsa, B.; Hofstra, H.; Huis in't Veld, J.H.; van der Vossen, J. Evaluation of Molecular Typing Techniques To Assign Genetic Diversity among *Saccharomyces cerevisiae* Strains. *Appl. Environ. Microbiol.* **1996**, *62*, 41–46. [[PubMed](#)]
15. Perez, M.A.; Gallego, F.J.; Hidalgo, P. Evaluation of molecular techniques for the genetic characterization of *Saccharomyces cerevisiae* strains. *FEMS Microbiol. Lett.* **2001**, *205*, 375–378. [[CrossRef](#)]
16. Legras, J.L.; Merdinoglu, D.; Cornuet, J.M.; Karst, F. Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Mol. Ecol.* **2007**, *16*, 2091–2102. [[CrossRef](#)] [[PubMed](#)]
17. Martinez, C.; Gac, S.; Lavin, A.; Ganga, M. Genomic characterization of *Saccharomyces cerevisiae* strains isolated from wine-producing areas in South America. *J. Appl. Microbiol.* **2004**, *96*, 1161–1168. [[CrossRef](#)]
18. Fay, J.; Benavides, J. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* **2005**, *1*, 66–71. [[CrossRef](#)]
19. Liti, G.; Carter, D.M.; Moses, A.M.; Warringer, J.; Parts, L.; James, S.A.; Davey, R.P.; Roberts, I.N.; Burt, A.; Koufopanou, V.; et al. Population genomics of domestic and wild yeasts. *Nature* **2009**, *458*, 337–341. [[CrossRef](#)]
20. Strobe, P.K.; Skelly, D.A.; Kozmin, S.G.; Mahadevan, G.; Stone, E.A.; Magwene, P.M.; Dietrich, F.S.; McCusker, J.H. The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* **2015**, *25*, 762–774. [[CrossRef](#)]
21. Peter, J.; De Chiara, M.; Friedrich, A.; Yue, J.X.; Pflieger, D.; Bergstrom, A.; Sigwalt, A.; Barre, B.; Freil, K.; Llored, A.; et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **2018**, *556*, 339–344. [[CrossRef](#)] [[PubMed](#)]

22. Cubillos, F.; Louis, E.; Liti, G. Generation of a large set of genetically tractable haploid and diploid *Saccharomyces* strains. *FEMS Yeast Res.* **2009**, *9*, 1217–1225. [[CrossRef](#)] [[PubMed](#)]
23. Cubillos, F.; Billi, E.; Zörgö, E.; Parts, L.; Fargier, P.; Omholt, S.; Blomberg, A.; Warringer, J.; Louis, E.; Liti, G. Assessing the complex architecture of polygenic traits in diverged yeast populations. *Mol. Ecol.* **2011**, *20*, 1401–1413. [[CrossRef](#)]
24. Liti, G.; Louis, E.J. Advances in quantitative trait analysis in yeast. *PLoS Genet.* **2012**, *8*, e1002912. [[CrossRef](#)] [[PubMed](#)]
25. Cubillos, F.; Parts, L.; Salinas, F.; Bergström, A.; Scovacricchi, E.; Zia, A.; Illingworth, C.; Mustonen, V.; Ibstedt, S.; Warringer, J.; et al. High-Resolution Mapping of Complex Traits with a Four-Parent Advanced Intercross Yeast Population. *Genetics* **2013**, *195*, 1141–1155. [[CrossRef](#)] [[PubMed](#)]
26. Kessi-Pérez, E.I.; Araos, S.; García, V.; Salinas, F.; Abarca, V.; Larrondo, L.F.; Martínez, C.; Cubillos, F.A. RIM15 antagonistic pleiotropy is responsible for differences in fermentation and stress response kinetics in budding yeast. *FEMS Yeast Res.* **2016**, *16*, fow021. [[CrossRef](#)]
27. Marullo, P.; Aigle, M.; Bely, M.; Masneuf-Pomarede, I.; Durrrens, P.; Dubourdieu, D.; Yvert, G. Single QTL mapping and nucleotide-level resolution of a physiologic trait in wine *Saccharomyces cerevisiae* strains. *FEMS Yeast Res.* **2007**, *7*, 941–952. [[CrossRef](#)]
28. Parts, L.; Cubillos, F.A.; Warringer, J.; Jain, K.; Salinas, F.; Bumpstead, S.J.; Molin, M.; Zia, A.; Simpson, J.T.; Quail, M.A.; et al. Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* **2011**, *21*, 1131–1138. [[CrossRef](#)]
29. Salinas, F.; Cubillos, F.A.; Soto, D.; Garcia, V.; Bergstrom, A.; Warringer, J.; Ganga, M.A.; Louis, E.J.; Liti, G.; Martínez, C. The genetic basis of natural variation in oenological traits in *Saccharomyces cerevisiae*. *PLoS ONE* **2012**, *7*, e49640. [[CrossRef](#)]
30. Kessi-Pérez, E.I.; Salinas, F.; González, A.; Su, Y.; Guillamón, J.M.; Hall, M.N.; Larrondo, L.F.; Martínez, C. KAE1 Allelic Variants Affect TORC1 Activation and Fermentation Kinetics in *Saccharomyces cerevisiae*. *Front. Microbiol.* **2019**, *10*, 1686. [[CrossRef](#)]
31. Ambroset, C.; Petit, M.; Christian, B.; Isabelle, S.; Delobel, P.; Guérin, C.; Chiapello, H.; Nicolas, P.; Bigey, F.; Dequin, S.; et al. Deciphering the Molecular Basis of Wine Yeast Fermentation Traits Using a Combined Genetic and Genomic Approach. *G3 (Bethesda Md.)* **2011**, *1*, 263–281. [[CrossRef](#)] [[PubMed](#)]
32. Steyer, D.; Ambroset, C.; Brion, C.; Claudel, P.; Delobel, P.; Sanchez, I.; Erny, C.; Blondin, B.; Karst, F.; Legras, J.L. QTL mapping of the production of wine aroma compounds by yeast. *BMC Genom.* **2012**, *13*, 573. [[CrossRef](#)] [[PubMed](#)]
33. Brice, C.; Sanchez, I.; Bigey, F.; Legras, J.L.; Blondin, B. A genetic approach of wine yeast fermentation capacity in nitrogen-starvation reveals the key role of nitrogen signaling. *BMC Genom.* **2014**, *15*, 495. [[CrossRef](#)] [[PubMed](#)]
34. Ibstedt, S.; Stenberg, S.; Bagés, S.; Gjuvslund, A.; Salinas, F.; Kourtchenko, O.; Karloss, J.; Blomberg, A.; Omholt, S.; Liti, G.; et al. Concerted Evolution of Life Stage Performances Signals Recent Selection on Yeast Nitrogen Use. *Mol. Biol. Evol.* **2014**, *32*, 153–161. [[CrossRef](#)] [[PubMed](#)]
35. Jara, M.; Cubillos, F.A.; García, V.; Salinas, F.; Aguilera, O.; Liti, G.; Martínez, C. Mapping Genetic Variants Underlying Differences in the Central Nitrogen Metabolism in Fermenter Yeasts. *PLoS ONE* **2014**, *9*, e86533. [[CrossRef](#)] [[PubMed](#)]
36. Alexandre, H.; Charpentier, C. Biochemical aspects of stuck and sluggish fermentation in grape must. *J. Ind. Microbiol. Biotechnol.* **1998**, *20*, 20–27. [[CrossRef](#)]
37. Pretorius, I.S. Tailoring wine yeast for the new millennium: Novel approaches to the ancient art of winemaking. *Yeast* **2000**, *16*, 675–729. [[CrossRef](#)]
38. Brice, C.; Cubillos, F.A.; Dequin, S.; Camarasa, C.; Martínez, C. Adaptability of the *Saccharomyces cerevisiae* yeasts to wine fermentation conditions relies on their strong ability to consume nitrogen. *PLoS ONE* **2018**, *13*, e0192383. [[CrossRef](#)]
39. Cubillos, F.A.; Brice, C.; Molinet, J.; Tisné, S.; Abarca, V.; Tapia, S.M.; Oporto, C.; García, V.; Liti, G.; Martínez, C. Identification of Nitrogen Consumption Genetic Variants in Yeast Through QTL Mapping and Bulk Segregant RNA-Seq Analyses. *G3 Genes Genomes Genet.* **2017**, *7*, 1693–1705. [[CrossRef](#)]
40. Salinas, F.; de Boer, C.G.; Abarca, V.; Garcia, V.; Cuevas, M.; Araos, S.; Larrondo, L.F.; Martínez, C.; Cubillos, F.A. Natural variation in non-coding regions underlying phenotypic diversity in budding yeast. *Sci. Rep.* **2016**, *6*, 21849. [[CrossRef](#)]

41. Som, A. Causes, consequences and solutions of phylogenetic incongruence. *Brief. Bioinform.* **2015**, *16*, 536–548. [[CrossRef](#)] [[PubMed](#)]
42. Hinchliff, C.; Smith, S.; Allman, J.; Burleigh, J.; Chaudhary, R.; Coghill, L.; Crandall, K.; Deng, J.B.; Drew, B.; Gazis, R.; et al. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12764–12769. [[CrossRef](#)] [[PubMed](#)]
43. Smith, S.A.; Brown, J.W. Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **2018**, *105*, 302–314. [[CrossRef](#)] [[PubMed](#)]
44. Fedosov, A.; Puillandre, N.; Herrmann, M.; Dgebuadze, P.; Bouchet, P. Phylogeny, systematics, and evolution of the family Costellariidae (Gastropoda: Neogastropoda). *Zool. J. Linn. Soc.* **2017**, *179*, 541–526.
45. Moon, B.C. A new phylogeny of ichthyosaurs (Reptilia: Diapsida). *J. Syst. Palaeontol.* **2017**, *17*, 129–155. [[CrossRef](#)]
46. Guzman, E.; Molina, J. The predictive utility of the plant phylogeny in identifying sources of cardiovascular drugs. *Pharm. Biol.* **2018**, *56*, 154–164. [[CrossRef](#)]
47. Yang, Z.; Rannala, B. Molecular phylogenetics: Principles and practice. *Nat. Rev. Genet.* **2012**, *13*, 303–314. [[CrossRef](#)]
48. Stamatakis, A. Phylogenetics: Applications, Software and Challenges. *Cancer Genom. Proteom.* **2005**, *2*, 301–305.
49. Campbell, V.; Legendre, P.; Lapointe, F.J. The performance of the Congruence Among Distance Matrices (CADM) test in phylogenetic analysis. *BMC Evol. Biol.* **2011**, *11*, 64. [[CrossRef](#)]
50. Garamszegi, L. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*; Springer: Berlin/Heidelberg, Germany, 2014. [[CrossRef](#)]
51. Villalobos-Cid, M.; Dorn, M.; Ligabue-Braun, R.; Inostroza-Ponta, M. A Memetic Algorithm Based on an NSGA-II Scheme for Phylogenetic Tree Inference. *IEEE Trans. Evol. Comput.* **2019**, *23*, 776–787. [[CrossRef](#)]
52. Molinet, J.; Cubillos, F.A.; Salinas, F.; Liti, G.; Martinez, C. Genetic variants of TORC1 signaling pathway affect nitrogen consumption in *Saccharomyces cerevisiae* during alcoholic fermentation. *PLoS ONE* **2019**, *14*, e0220515. [[CrossRef](#)] [[PubMed](#)]
53. Bergström, A.; Simpson, J.; Salinas, F.; Barré, B.; Parts, L.; Zia, A.; Ba, A.; Moses, A.; Louis, E.; Mustonen, V.; et al. A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes. *Mol. Biol. Evol.* **2014**, *31*, 872–888. [[CrossRef](#)] [[PubMed](#)]
54. Knaus, B.J.; Grünwald, N.J. vcfr: A package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **2017**, *17*, 44–53. [[CrossRef](#)] [[PubMed](#)]
55. Levasseur, C.; Lapointe, F.J. Total evidence, average consensus and matrix representation with parsimony: What a difference distances make. *Evol. Bioinform. Online* **2007**, *2*, 1–5. [[CrossRef](#)]
56. McVay, J.D.; Carstens, B.C. Phylogenetic Model Choice: Justifying a Species Tree or Concatenation Analysis. *J. Phylogenet. Evol. Biol.* **2013**, *1*, 1–8.
57. Schliep, K.P. Phangorn: Phylogenetic analysis in R. *Bioinformatics* **2010**, *27*, 592–593. [[CrossRef](#)]
58. Gadagkar, S.; Rosenberg, M. Inferring Species Phylogenies From Multiple Genes: Concatenated Sequence Tree Versus Consensus Gene Tree. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2005**, *304*, 64–74. [[CrossRef](#)]
59. Robinson, D.; Foulds, L. Comparison of phylogenetic trees. *Math. Biosci.* **1981**, *53*, 131–147. [[CrossRef](#)]
60. Jombart, T.; Kendall, M.; Almagro-García, J.; Colijn, C. treespace : Statistical exploration of landscapes of phylogenetic trees. *Mol. Ecol. Resour.* **2017**, *17*, 1385–1392. [[CrossRef](#)]
61. Kessi-Pérez, E.I.; Salinas, F.; Molinet, J.; González, A.; Muñoz, S.; Guillamón, J.M.; Hall, M.N.; Larrondo, L.F.; Martínez, C. Indirect monitoring of TORC1 signalling pathway reveals molecular diversity among different yeast strains. *Yeast* **2019**, *36*, 65–74. [[CrossRef](#)]
62. De Leeuw, J.; Mair, P. Multidimensional Scaling using majorization: SMACOF in R. *J. Stat. Softw.* **2009**, *31*, 1. [[CrossRef](#)]
63. Ramazzotti, M.; Berná, L.; Stefanini, I.; Cavalieri, D. A computational pipeline to discover highly phylogenetically informative genes in sequenced genomes: Application to *Saccharomyces cerevisiae* natural strains. *Nucleic Acids Res.* **2012**, *40*, 3834–3848. [[CrossRef](#)] [[PubMed](#)]

