Current Biology 17, 407–411, March 6, 2007 ©2007 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2006.12.047

Report

Allopatric Divergence, Secondary Contact, and Genetic Isolation in Wild Yeast Populations

Heidi A. Kuehne,[1,*] Helen A. Murphy,[2] Chantal A. Francis,[2] and Paul D. Sniegowski[2]

[1] Division of Biological Sciences
University of Montana
Missoula, Montana 59812

[2] Department of Biology
University of Pennsylvania
415 South University Avenue
Philadelphia, Pennsylvania 19104

Summary

In plants and animals, new biological species clearly have arisen as a byproduct of genetic divergence in allopatry. However, our understanding of the processes that generate new microbial species remains limited [1] despite the large contribution of microbes to the world's biodiversity. A recent hypothesis claims that microbes lack biogeographical divergence because their population sizes are large and their migration rates are presumably high [2, 3]. In recapitulating the classic microbial-ecology dictum that "everything is everywhere, and the environment selects" [4, 5], this hypothesis casts doubt on whether geographic divergence promotes speciation in microbes. To date, its predictions have been tested primarily with data from eubacteria and archaebacteria [6–8]. However, this hypothesis's most important implication is in sexual eukaryotic microbes, where migration and genetic admixture are specifically predicted to inhibit allopatric divergence and speciation [9]. Here, we use nuclear-sequence data from globally distributed natural populations of the yeast *Saccharomyces paradoxus* to investigate the role of geography in generating diversity in sexual eukaryotic microbes. We show that these populations have undergone allopatric divergence and then secondary contact without genetic admixture. Our data thus support the occurrence of evolutionary processes necessary for allopatric speciation in sexual microbes.

Results and Discussion

Saccharomyces paradoxus is a close, undomesticated relative of the brewer's yeast *S. cerevisiae* and is known to inhabit temperate woodlands across the northern hemisphere, where it can be found on bark, leaf surfaces, and exudates of deciduous trees and in associated soils [10–12]. As part of a larger study aimed at characterizing the genetic structures of natural populations of *S. paradoxus* and *S. cerevisiae*, we analyzed nucleotide-sequence variation in spliceosomal introns from nine unlinked nuclear loci in 62 independent isolates of *S. paradoxus* sampled from multiple sites in two continental regions: eastern North America and eastern Europe/western Asia (Figure 1). (For convenience, we will refer to these two regions simply as "North America" and "Eurasia".) All of the isolates were confirmed as belonging to the biological species *S. paradoxus* through fertile matings to genetically marked *S. paradoxus* tester strains. Phylogenetic analysis of the sequence data revealed two major and unambiguously distinct genetic groups ("A" and "B") that included 61 of the isolates and were separated by large numbers of fixed nucleotide differences (Table 1); a singleton isolate represented a third genetic group ("C"). Group A (15 isolates) was present in both North America and Eurasia, whereas group B (46 isolates) was present only in North America (Figure 1). The singleton isolate representing group C was found in North America.

Eleven of the 15 isolates in group A were from Eurasia; the remaining four (discussed below) were from North America. The 11 Eurasian group A isolates exhibited 11 unique multilocus haplotypes and were scattered without obvious pattern among ten sites separated by 110–3600 km. A Mantel test for association between genetic and geographic distances in these isolates was not significant ($r = -0.02$, $p = 0.45$). Analysis of the pattern of multilocus variation among these eleven isolates has revealed that they are part of a breeding population structured by free recombination at the scale of thousands of kilometers [13] and thus indicates a long residence time for group A in Eurasia. Similarly, the 46 isolates of group B found in North America exhibited 35 unique multilocus haplotypes distributed without obvious pattern among nine sites separated by 70 to nearly 1000 km. A Mantel test for association between genetic and geographic distance within group B was also not significant ($r = 0.12$, $p = 0.09$). Analyses of multilocus variation indicated that group B is part of a breeding population that is broadly structured by free recombination [13], and this finding is consistent with a long residence time for group B in North America. Because group C was represented by a single isolate, inferences about its genetic population and geographic structure were not possible. It is interesting to note, however, that this single isolate was genetically quite distant from and shared no mutations with groups A and B (Table 1).

The genetic distance between the inferred resident populations of *S. paradoxus* in Eurasia (group A) and North America (group B) is substantial. Across the 4926 bp of sequence obtained for the nine loci analyzed, the isolates from these populations exhibited 176 fixed differences and no shared mutations. This striking pattern provides strong evidence that the two populations have evolved independently for tens of millions of generations over tens of thousands of years (see Experimental Procedures). We note that this is not the first time that geographical divergence has been documented in *Saccharomyces*. We showed previously that *S. paradoxus* populations from Europe and Far East Asia are genetically diverged [12] by using allozyme analysis,

*Correspondence: heidi.kuehne@mso.umt.edu

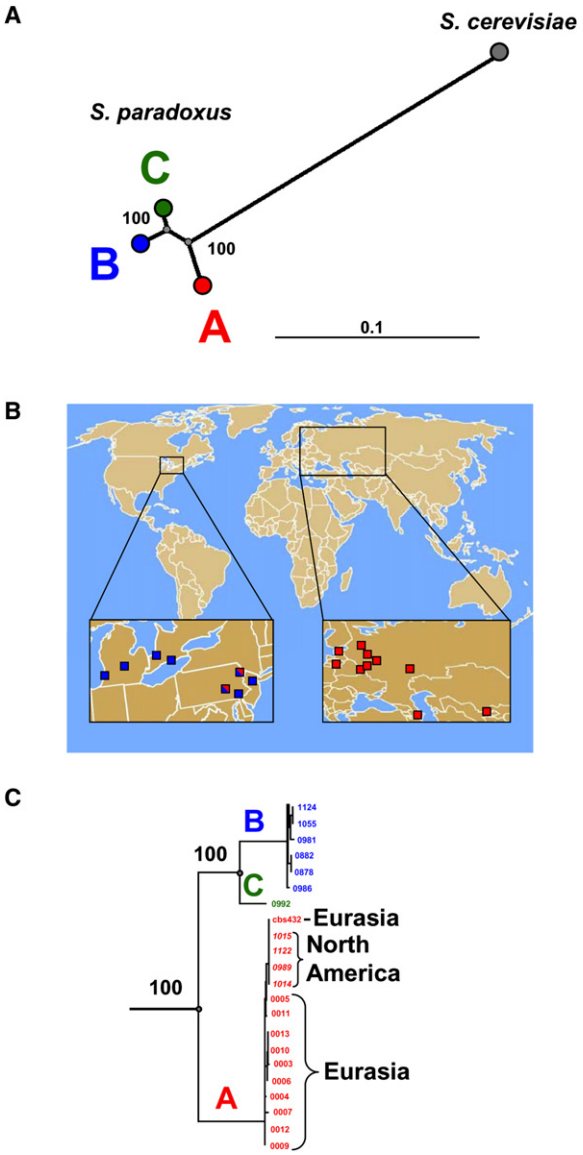

Figure 1. Phylogeny and Collection Sites of *S. paradoxus* Isolates

(A) Maximum-likelihood tree of the three *S. paradoxus* genetic groups rooted on *S. cerevisiae*. The tree was constructed with nine concatenated intron DNA sequences from 57 *S. paradoxus* isolates; orthologous *S. cerevisiae* DNA sequences were obtained from the *Saccharomyces* Genome Database (http://www.yeastgenome.org/). Numbers at internal nodes represent % bootstrap support (1000 replicates); the scale bar represents number of per site substitutions. Complete maximum-likelihood tree is given in the Supplemental Data.

(B) Red symbols represent genetic group A, which was found at all of the collection sites in Eurasia (Estonia, Lithuania, Azerbaijan, Uzbekistan, Russia-Moscow, St. Petersburg, Novgorod, Voronezh, Republic of Tatarstan, and the Tula Region) and at two of the collection sites in North America (Tuscarora Forest, Pennsylvania and Buck Hill Falls, Pennsylvania, both in the U.S. and separated by approximately 217 km). Blue symbols represent genetic group B, which was found at all of the collection sites in North America (U.S.: Media, Buck Hill Falls and Tuscarora Forest in Pennsylvania, New Brunswick in New Jersey, East Lansing and Saugatuck in Michigan; Canada: Grand Bend and Port Rowan in Ontario). The single isolate representing genetic group C (not shown) was collected in Buck Hill Falls, Pennsylvania.

(C) Detail of the maximum-likelihood tree showing genetic group A. The four group A isolates collected in North America and CBS432 were shown as derived relative to other group A isolates in 996 of 1000 replicate analyses, and this finding is consistent with the North American group A lineage being founded by migrants from Eurasia, rather than the other way around.

Table 1. Fixed Differences between Genetic and Geographic Groups in *S. paradoxus*

	Group A (North America)	Group B (North America)	Group C (North America)
Group A (Eurasia)	3	176	160
Group A (North America)	0	181	165
Group B (North America)	0	0	89

Groups A, B, and C are determined by analysis of the multilocus-sequence data as described in the text. Geographic origins are in parentheses. Entries in the table represent comparisons across all nine intron loci (4926 bp). No two groups shared mutations.

and a recent study of nuclear-sequence data in *S. paradoxus* isolates sampled in Canada and Europe documents strong geographical divergence between these regions [14].

Our current study, however, provides the first evidence of secondary contact with genetic isolation in *Saccharomyces* and indeed in any sexually reproducing, freeliving microbe. As shown in Figure 1, we collected multiple group A isolates of *S. paradoxus* in sympatry with group B isolates at two widely separated sites in North America. Although these four North American group A isolates were identical at the nine intron loci analyzed here, we observed microsatellite divergence among them at a tenth locus (see the Supplemental Data available with this article online). The close genetic similarity and relatively wide geographic dispersion of these group A isolates strongly suggest migration from Eurasia followed by establishment and dispersal within North America. The North American group A isolates were also identical at the nine intron loci to the *S. paradoxus* type strain CBS 432, which was collected in Russia sometime in the early 1900s (G.I. Naumov, personal communication). However, we also observed microsatellite divergence at the tenth locus between CBS 432 and all four of the North American group A isolates. The observed divergences both among the North American group A isolates and between these isolates and CBS 432 rule out the unlikely possibility that they are contaminants from our laboratory's frozen stock of strain CBS 432 (see Supplemental Data). The close similarity of these isolates to each other and to CBS 432, compared with the extensive variation observed among the Eurasian group A isolates, provides compelling evidence that group A's presence in North America is the consequence of a transoceanic migratory event. The North American group A isolates and CBS 432 are diverged from the Eurasian isolates that form the basal part of the group A clade (Figure 1C), and such a finding is also consistent with transoceanic migration from Eurasia to North America. This migratory event had to have occurred a sufficiently long time ago to allow both geographical dispersion and the evolution of microsatellite divergence within the North American subpopulation of group A.

Our field observations indicate that groups A and B have had ample ecological opportunity to interbreed

after secondary contact in North America: We collected members of both groups from the same individual tree and even from within the same restricted (~100 cm^2) exudate location on an individual tree. However, our genetic data provide no evidence of genetic admixture between these sympatric groups: Across all 4926 bp analyzed, these populations were as diverged and as distinct as the resident populations of North America and Eurasia (Table 1). Our results thus indicate that allopatric *S. paradoxus* populations on two different continents, separated by an oceanic barrier, diverged genetically and evolved reproductive barriers that prevented genetic admixture upon migration and secondary contact.

Although we cannot say exactly what has prevented groups A and B from admixing on secondary contact in North America, it is noteworthy that spore viability in crosses between group A and group B in North America is sharply reduced compared to its value in intragroup crosses (Figure 2). (Spore viability is also reduced in crosses between the singleton isolate representing group C and group B.) Previous studies have in fact obtained evidence for reductions in spore viability between geographically distant populations of *S. paradoxus* [11, 15] but not between sympatric populations as observed here. Such partial postzygotic isolation could in principle have arisen in allopatry through genic or karyotypic differences [9], or even as a consequence of simple sequence divergence [15, 16]; its maintenance in sympatry, however, would seem unlikely without the presence of additional isolating factors. Prezygotic barriers to mating have been shown to evolve readily in experimental yeast populations [17], and we have observed that sympatric natural populations of *S. paradoxus* and *S. cerevisiae* exhibit strong differences in mating behavior that may contribute some prezygotic isolation [18]. These observations raise the possibility that some degree of prezygotic isolation could also have evolved between *S. paradoxus* groups A and B, either generally or only where they are found in sympatry, and this is something that we are currently investigating.

Explicit tests of the "no microbial biogeography" hypothesis [2] should be expanded to include plants, especially ferns, mosses, and liverworts [19]. Because ferns, mosses, and liverworts produce copious numbers of spores that disperse readily, predictions made for freeliving microbes under the "no microbial biogeography" hypothesis should also hold for these organisms [19]. Molecular studies of ferns, mosses, and liverworts indicate that strong geographical divergence may be common in these taxa, despite morphological similarity in apparently cryptic species [19–21], although tests of mating isolation have yet to be carried out in these cases. Similarly, the role of microbial dispersal in generating biogeographic diversity might be explored in flowering plants that produce abundant microbial seeds; one such plant is the model organism *Arabidopsis thaliana* [22], which displays strong geographical structure in Europe [23].

Our study of global patterns of genetic variation in yeast provides direct support for the role of biogeography in generating microbial diversity. Our results are consistent with the classical allopatric speciation model that has been validated extensively in plants and animals. Indeed, we cannot rule out the possibility that the resident North American and Eurasian *S. paradoxus* populations are already sufficiently diverged to maintain their identities as separate species where they are currently sympatric in North America. Microbial speciation may not be that different from macrobial speciation after all.

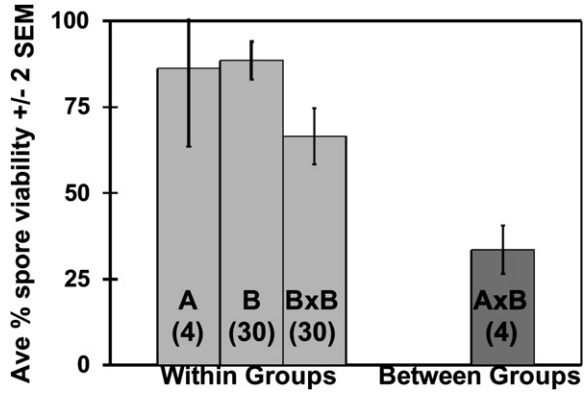

Figure 2. Average Spore Viabilities of Within- and Between-Group Crosses

Error bars show ±2 SEM. Light-gray bars represent within-group spore viabilities measured by homothallic selfing (A and B) and outcrossing within distinct genetic groups (B × B). Dark gray bars represent between-group spore viabilities measured by outcrossing between distinct genetic groups (A × B). No significant difference was observed between homothallic crosses within groups A and group B; significant differences were seen among all three types of genetic crosses (one-way analysis of variance with Tukey-Kramer tests for multiple comparisons, $\alpha = 0.05$). Numbers of replicates for each type of cross (corresponding to different strain combinations) are listed inside each bar. For each replicate, spores were dissected out of at least 16 asci and the number of colonies visible to the unaided eye at 72 hr was counted.

Experimental Procedures

Strains

S. paradoxus isolates were collected from oak trees at nine sites in North America in 1995, 1999, and 2000 and at ten sites in Eurasia in the 1980s. We identified novel isolates by matings between *S. cerevisiae* and *S. paradoxus* tester strains and auxotrophic derivatives of novel isolates. Tester strains were derived from North American isolates described in Sniegowski et al. [11]. Previous *S. paradoxus* isolates were identified with similar tests [11, 24]. Details are given in the Supplemental Data, and complete methods are given in Sniegowski et al. [11].

DNA Sequences, Alignments, and Phylogenetic Analysis

We assembled multilocus haplotypes with sequence variation observed at nine unlinked nuclear loci containing spliceosomal introns. The North American *S. paradoxus* sample included four group A isolates identical at nine intron loci to the *S. paradoxus* type strain CBS 432. We therefore searched further for variation among these isolates and CBS 432 by DNA sequencing four microsatellite loci in each of these five isolates. All loci were amplified with genome preparations made from monosporic cultures derived from the wild isolates and from CBS 432. Further details are given in the Supplemental Data.

We aligned sequences with the ClustalW algorithm. Isolates yielding singleton polymorphisms were resequenced from independent PCR product to verify that the polymorphisms did not result from mistakes incorporated during amplification. Polymorphisms shared among two or more isolates were considered independent evidence of variation and were not verified by additional sequencing.

We used PHYML v2.4.4 [25] to construct a maximum-likelihood tree by using DNA-sequence data concatenated from the nine intron loci. The likelihood tree was rooted on orthologous *S. cerevisiae* sequences. For assessing support for tree topology and branch length estimates, 1000 bootstrap replicates were performed. Further details are given in the Supplemental Experimental Procedures.

Spore Viabilities

Three types of genetic crosses were analyzed for their effect on spore viability: (1) homothallic selfing within North American groups A and B, (2) outcrossing between isolates within North American group B, and (3) outcrossing between isolates between North American groups A and B. All four North American group A isolates and 30 of the 46 North American group B isolates were tested. All outcrosses were made with a highly fertile derivative of a North American group B isolate.

We assessed homothallic fertility by sporulating the original, prototrophic cultures in accordance with standard methods [26], dissecting 64 spores out of 16 asci partially digested by zymolyase, and arraying the spores onto YPD agar. Spore viability was recorded as the proportion of colonies visible to the unaided eye after 72 hr at 30°C. We assessed outcrossed fertility by using the same procedures but on matings of auxotrophic strains derived from the original, prototrophic cultures. Spore-viability data were analyzed with a one-way analysis of variance (ANOVA) with the Tukey-Kramer honestly-significant-differences procedure for testing the significance of multiple pairwise comparisons between group A and group B homothallic crosses and among the three types of genetic crosses.

Because the auxotrophic strains were derived by UV mutagenesis, we explored the possibility that the UV treatment could have caused differences in spore viability between the homothallic and outcrossed treatments. We conducted an independent subset of crosses both within group B and between groups A and B by using a genetic method not involving UV mutagenesis (details given in Supplemental Experimental Procedures). Spore-viability data from these crosses were analyzed with a mixed-model, nested ANOVA with planned comparisons. The overall effect of crossing between groups was qualitatively the same as observed in the UV-mutagenized crosses, with crosses between groups A and B yielding significantly lower spore viabilities (Figure S1).

Geographic- and Genetic-Distance Analysis

We used a Mantel procedure [27] to test for spatial autocorrelation within the recombining populations represented by group A in Eurasia and group B in North America. A matrix of pairwise genetic distances was generated in PAUP*4.0 [28] by the application of the Tamura-Nei model of molecular evolution to seven concatenated intron sequences for all isolates within each recombining population. A corresponding matrix of pairwise geographic distances was constructed with distances between collection sites. Mantel tests were performed with the software program zt [29]. The statistical significance of each test was evaluated by 10000 random iterations.

Genetic Divergence Estimate

We estimated the divergence time between the geographically isolated A and B populations by using the equation $k = 2\mu t$, where k is the average number of substitutions per site between groups, t is time in generations, and μ is the per-site-per-generation mutation rate. k was estimated as 0.00132 for these two populations, with pooled DNA sequence data from the nine intron loci [16]. μ was assumed to be 1.84×10^{-10}, according to Fay and Benavides [30]. With these figures, the estimated number of generations since divergence is calculated as $t = 3.6 \times 10^{6}$. The number of generations per year in wild *Saccharomyces* populations is unknown. Fay and Benevides [30] assumed that there were 2920 generations per year in *S. cerevisiae* populations inhabiting temperate and tropical regions; applying this number to the above calculation yields an absolute divergence time of approximately 38000 years for the geographically isolated A and B populations. However, 38000 years is likely to be an underestimate because the true number of generations per year in wild yeast populations, especially in north temperate regions, may well be considerably lower than assumed in Fay and Benevides [30]. Furthermore, the entire north temperate region was affected by a major glaciation event during this time (peaking approximately 20000 years bp), and this is likely to have further reduced the number of generations per year.

Supplemental Data

Supplemental Data include additional Experimental Procedures, two figures, and four tables and are available with this article online at http://www.current-biology.com/cgi/content/full/17/5/407/DC1/.

Acknowledgments

We thank G.I. Naumov for *S. paradoxus* strains, P. Petraitis for statistical advice, and R.F. Rosenzweig for comments on the manuscript. This work was supported by grants from the National Science Foundation and the University of Pennsylvania Research Foundation to P.D.S. and from the Global Climate and Energy Project to R. F. Rosenzweig.

Received: September 21, 2006
Revised: December 22, 2006
Accepted: December 27, 2006
Published online: February 15, 2007

References

1. Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., et al. (2006). Microbial biogeography: Putting microorganisms on the map. Nat. Rev. Microbiol. *4*, 102–112.
2. Finlay, B.J., and Clarke, K.J. (1999). Ubiquitous dispersal of microbial species. Nature *400*, 828.
3. Finlay, B.J. (2002). Global dispersal of free-living microbial eukaryote species. Science *296*, 1061–1063.
4. Baas-Becking, L.G.M. (1934). Geobiologie of Inleiding tot de Milieukunde (The Hague: Van Stockum & Zon).
5. Beijerinck, M.W. (1913). Jaarboek van de Koninklijke Akademie v. Wetenschoppen (Amsterdam: Muller).
6. Foti, M., Ma, S., Sorokin, D.Y., Rademaker, J.L.W., Kuenen, J.G., and Muyzer, G. (2006). Genetic diversity and biogeography of haloalkaliphilic sulphur-oxidizing bacteria belonging to the genus Thioalkalivibrio. FEMS Microbiol. Ecol. *56*, 95–101.
7. Papke, R.T., Ramsing, N.B., Bateson, M.M., and Ward, D.M. (2003). Geographical isolation in hot spring cyanobacteria. Environ. Microbiol. *5*, 650–659.
8. Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science *301*, 976–978.
9. Coyne, J.A., and Orr, H.A. (2004). Speciation (Sunderland, MA: Sinauer Associates).
10. Johnson, L.J., Koufopanou, V., Goddard, M.R., Hetherington, R., Schafer, S.M., and Burt, A. (2004). Population genetics of the wild yeast *Saccharomyces paradoxus*. Genetics *166*, 43–52.
11. Sniegowski, P.D., Dombrowski, P.G., and Fingerman, E. (2002). *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. FEMS Yeast Res *1*, 299–306.
12. Naumov, G.I., Naumova, E.S., and Sniegowski, P.D. (1997). Differentiation of European and Far East Asian populations of *Saccharomyces paradoxus* by allozyme analysis. Int. J. Syst. Bacteriol. *47*, 341–344.
13. Kuehne, H.A. (2005). The Genetic Structure and Biogeography of Natural Saccharomyces Populations (Philadelphia: University of Pennsylvania).
14. Koufopanou, V., Hughes, J., Bell, G., and Burt, A. (2006). The spatial scale of genetic differentiation in a model organism: The wild yeast *Saccharomyces paradoxus*. Philos. Trans. R. Soc. Lond. B Biol. Sci. *361*, 1941–1946.
15. Greig, D., Travisano, M., Louis, E.J., and Borts, R.H. (2003). A role for the mismatch repair system during incipient speciation in Saccharomyces. J. Evol. Biol. *16*, 429–437.
16. Hunter, N., Chambers, S.R., Louis, E.J., and Borts, R.H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J. *15*, 1726–1733.

17. Leu, J.Y., and Murray, A.W. (2006). Experimental evolution of mating discrimination in budding yeast. Curr. Biol. *16*, 280–286.
18. Murphy, H.A., Kuehne, H.A., Francis, C.A., and Sniegowski, P.D. (2006). Mate choice assays and a potential mechanism for pre-zygotic isolation in natural yeast populations. Biology Letters *2*, 553–556.
19. Foissner, W. (2006). Biogeography and dispersal of micro-organisms: A review emphasizing protists. Acta Protozool. *45*, 111–136.
20. Shaw, A.J. (2001). Biogeographic patterns and cryptic speciation in bryophytes. J. Biogeogr. *28*, 253–261.
21. Wall, D.P. (2005). Origin and rapid diversification of a tropical moss. Evolution Int. J. Org. Evolution *59*, 1413–1424.
22. Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C.J., and Koornneef, M. (1999). Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA *96*, 4710–4717.
23. Ostrowski, M.F., David, J., Santoni, S., Mckhann, H., Reboud, X., Le Corre, V., Camilleri, C., Brunel, D., Bouchez, D., Faure, B., et al. (2006). Evidence for a large-scale population structure among accessions of *Arabidopsis thaliana*: Possible causes and consequences for the distribution of linkage disequilibrium. Mol. Ecol. *15*, 1507–1517.
24. Naumov, G.I., Naumova, E.S., and Sniegowski, P.D. (1998). *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* are associated with exudates of North American oaks. Can. J. Microbiol. *44*, 1045–1050.
25. Guindon, S., and Gaccuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. *52*, 696–704.
26. Rose, M.D., Winston, F., and Hieter, P. (1990). Methods in Yeast Genetics: A Laboratory Course Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
27. Sokal, R.R., and Rohlf, F.J. (1995). Biometry: The Principles and Practice of Statistics in Biological Research, Third Edition (San Francisco: W.H. Freeman).
28. Swofford, D.L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Fourth Edition (Sunderland, MA: Sinauer Associates).
29. Bonnet, E., and Van de Peer, Y. (2002). zt: A software tool for simple and partial Mantel tests. Journal of Statistical Software *7*, 1–12.
30. Fay, J.C., and Benavides, J.A. (2005). Evidence for domesticated and wild populations of *Saccharomyces cerevisiae.* PLoS Genet *1*, 66–71.

Accession Numbers

All intron-containing gene sequences discussed in this paper have been deposited into GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) as accession numbers EF196095-EF196647.