

The spatial scale of genetic differentiation in a model organism: the wild yeast *Saccharomyces paradoxus*

Vassiliki Koufopanou*, Joseph Hughes†, Graham Bell‡ and Austin Burt

*Division of Biology and NERC Centre for Population Biology, Imperial College London,
Silwood Park, Ascot, Berks SL5 7PY, UK*

Little information is presently available on the factors promoting genetic divergence in eukaryotic microbes. We studied the spatial distribution of genetic variation in *Saccharomyces paradoxus*, the wild relative of *Saccharomyces cerevisiae*, from the scale of a few centimetres on individual oak trees to thousands of kilometres across different continents. Genealogical analysis of six loci shows that isolates from Europe form a single recombining population, and within this population genetic differentiation increases with physical distance. Between different continents, strains are more divergent and genealogically independent, indicating well-differentiated lineages that may be in the process of speciation. Such replicated populations will be useful for studies in population genomics.

Keywords: speciation; reproductive isolation; gene genealogies; isolation by distance; population genomics

1. INTRODUCTION

Speciation in plants and animals can occur by a number of different pathways, with divergent selection on geographically isolated populations probably being the most common (Coyne & Orr 2004). Much less information is available on speciation in eukaryotic microbes. It is possible that many of the same processes apply as in larger eukaryotes, or alternatively, it has been proposed that small size allows essentially global dispersal, which would prevent divergence between geographically separate populations (Finlay 2002; Fenchel & Finlay 2004). To distinguish between these alternatives, data are required on the genetic divergence of microbial populations, ideally over a broad range of geographical distances. In this paper, we analyse such data for *Saccharomyces paradoxus*, the closest known relative of the model eukaryote *Saccharomyces cerevisiae* (bakers' yeast). Unlike its more famous relative, *S. paradoxus* has never, to our knowledge, been domesticated, and so its history and population genetics should be free from the idiosyncracies of close association with humans.

S. paradoxus and *S. cerevisiae* can be isolated using the same ethanol-containing enrichment media (Sniegowski *et al.* 2002), and there are now strains available for both species from several locations. *S. paradoxus* has mainly

been isolated from the bark of oak trees and surrounding soil, while *S. cerevisiae* is commonly obtained from vineyards, but the two species are also known to co-occur (Naumov *et al.* 1998; Redzepovic *et al.* 2002). Multilocus sequencing of *S. cerevisiae* strains from different locations indicates strong and complex effects of domestication (Fay & Benavides 2005; Aa *et al.* 2006). For *S. paradoxus*, previous work has shown that strains from North America and Far East Asia differ in allozyme frequencies from European isolates, and that hybrids between different continents are partially sterile (Naumov *et al.* 1997; Sniegowski *et al.* 2002). Within a single population of *S. paradoxus*, analyses of DNA sequence variation indicate that the three modes of reproduction observed in the lab—clonality, inbreeding and outcrossing—also occur in nature (Johnson *et al.* 2004). Here, we extend this study to examine the spatial distribution of DNA sequence variation, from the scale of a few centimetres on an oak tree to thousands of kilometres between different continents. We find increasing genetic differentiation with increased distance, both within a tree and between trees, and also at greater geographical distances, between different countries in Europe. Nevertheless, the entire population of Europe represents a single recombining population. However, mixing of populations breaks down entirely when isolates from different continents are considered, strains being far more divergent and showing complete sorting of alleles at all six loci sampled. This indicates a persistent lack of mixing which together with the observed partial sterility of hybrids suggests the presence of three independent lineages in the process of speciation.

* Author for correspondence (v.koufopanou@imperial.ac.uk).

† Present address: DEEB, IBLS, University of Glasgow, Glasgow G12 8QP, UK.

‡ Present address: Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada.

The electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2006.1922> or via <http://www.journals.royalsoc.ac.uk>.

One contribution of 15 to a Discussion Meeting Issue 'Species and speciation in micro-organisms'.

2. MATERIAL AND METHODS

Yeast strains were isolated in the summer of 2003 from the bark of two oak trees in Silwood Park in southeast England.

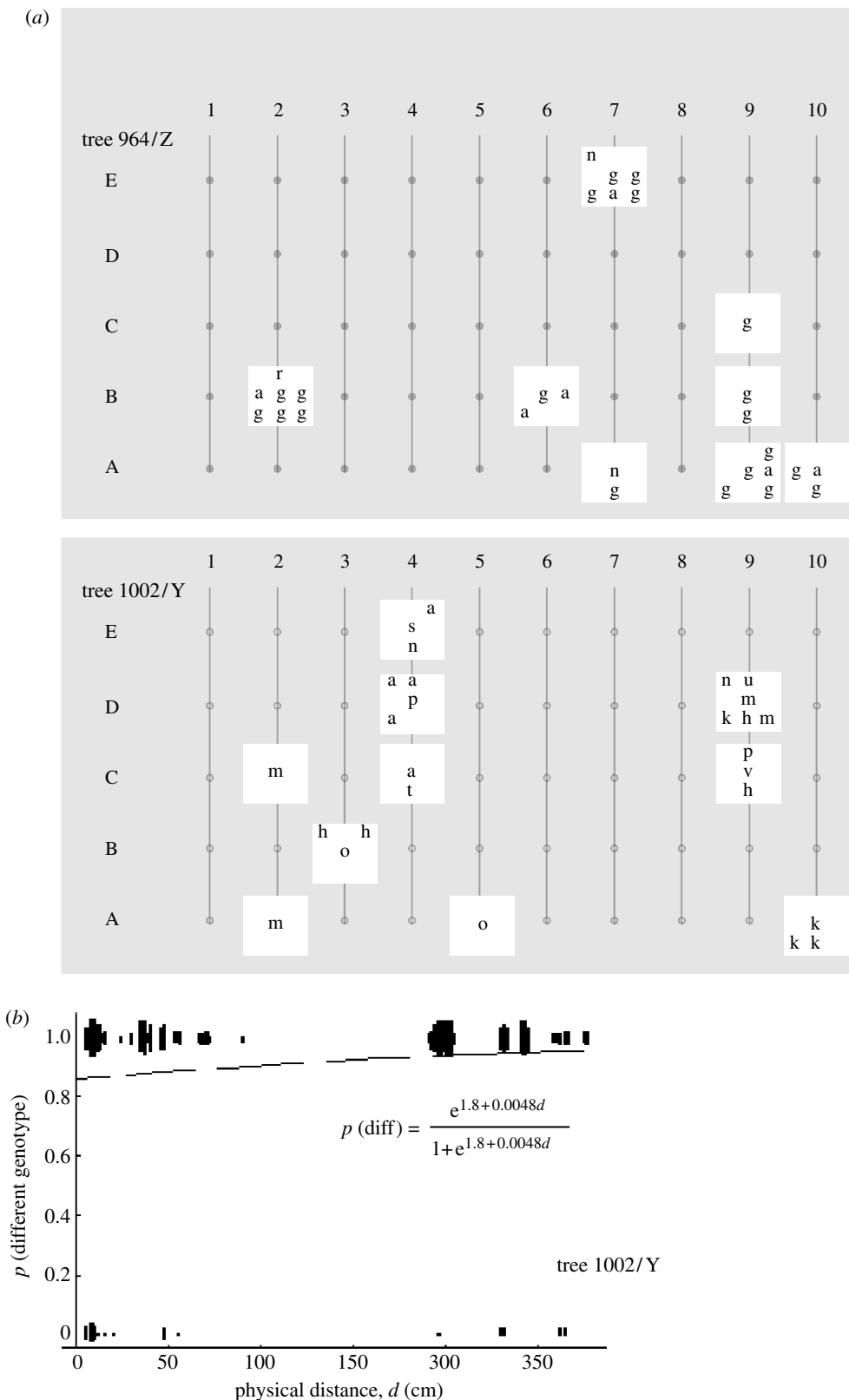


Figure 1. (a) The distribution of genotypes on two trees: 56 strains, 13 different genotypes, different letters indicating a different genotype (see electronic supplementary material). Each line represents a vertical transect, with dots indicating the sites of first sampling. Squares show the positions of the second set of samples collected around the sites found to have *S. paradoxus* in the first sample. (b) The probability of uncovering a different genotype as a function of physical distance on the tree trunk (logistic regression, $p=0.02$). For every pair of strains, whether they have the same or different genotype is plotted as a function of the distance between them. The data for tree 1002/Y only; points have been displaced to make them more obvious.

The first set of samples were taken along vertical transects on the lower part of the trunk (running along major troughs in the bark) between the southeast- and southwest-facing sides of the tree. Ten sites were identified along each transect, at 10 cm intervals, and at each site we collected two samples

separated by 1 cm. For this first set of samples, we collected yeasts by pressing 4 mm² of Blu Tack onto the surface of the bark and then peeling it off and transferring it to a sterile collection tube. The location of each sampling site was marked with a pin. For every sample that was positive for

Table 1. Origin and identification of strains.

location	no.	id & origin
UK (Silwood Park, oak tree 1002/Y)	27	Yx.x (this study)
(Silwood Park, oak tree 964/Z)	29	Zx.x (this study) ^a
(Silwood & Great Windsor Park, several oak trees)	9	Qx.x & Tx.x (data from Johnson <i>et al.</i> 2004)
Spain (J. Santa Maria)	3	CECT1076, 10308, 10329
Russia (Leningrad)	1	YPS3 ^b (P. Sniegowski)
(Moscow)	2	YPS4 ^b , 8 ^b (P. Sniegowski)
(Moscow)	3	ATCC96986 ^b , 96983 ^b , 96975
Latvia (Sigulda)	4	SIG1, 3, 6, 8 (this study)
(Riga)	3	RIG3, 10, 11 (this study)
Sweden (Stockholm)	1	STOC3 (this study)
Denmark	1	CBS5829
Unknown	1	CBS432T (<i>S. paradoxus</i> Type)
Far East Asia	9	CBS8436-44
Canada (Mt St Hilaire, Montreal, several oak and other trees & soil)	19	Ax (this study)
Brazil	1	CBS8841T ^b (<i>S. cariocanus</i> Type)

^a Strain T21.4 of Johnson *et al.* (2004) was collected from the same tree.

^b Indicates monosporic, fully homozygous isolates.

S. paradoxus, we then took a second set of eight samples equally spaced around the perimeter of a 10 cm × 10 cm square centred on the positive site. For these second samples, we took 5 mm diameter cores, approximately 1 cm deep into the bark. All the samples were aseptically transferred to 1.5 ml tubes and processed following the protocol of Sniegowski *et al.* (2002). Individual colonies that looked like *S. paradoxus* were then picked and grown on YPD (yeast extract–peptone–dextrose) medium for further polymerase chain reaction (PCR)-based identification (using primers ITS1 and ITS4 to amplify and sequence the ITS1–5.8rRNA–ITS2 region; Johnson *et al.* 2004).

For yeast strains identified as *S. paradoxus*, we extracted DNA and then PCR-amplified and sequenced fragments of six genes involved in the mating reaction. These genes were the same as those previously analysed by Johnson *et al.* (2004), namely *a*-pheromone (*MFA1*), alpha-pheromone (*MFalpha1*), *a*-pheromone receptor (*STE3*), alpha-pheromone receptor (*STE2*), *a*-agglutinin (*AGA2*) and alpha-agglutinin (*SAG1*).

Randomization tests for clonality and estimates of Weir's (1996) Θ statistic for population differentiation were performed using MULTILOCUS (Agapow & Burt 2001).

3. RESULTS AND DISCUSSION

(a) Population structure on individual oak trees

In the first set of isolations, we obtained 100 bark samples from each of the three trees, and the number of *S. paradoxus*-positive samples was about 10% for two of these trees. For the third tree, we did not recover any *S. paradoxus*, even though we had successfully recovered a strain from it previously (T18.2; Johnson *et al.* 2004). We then took a second set of samples around each of the positive sites (figure 1a). Success rate was higher the second time, at 28% (38 positives out of 136 samples for the two trees). The fragments (500–1000 bp) from each of the six genes were amplified and sequenced in all strains. Of approximately 4000 bp sequenced per strain, 16 sites were polymorphic, including a repeat polymorphism in the *MFalpha1* gene (electronic supplementary material).

On one of the trees (1002/Y), 11 out of 27 strains had different genotypes. The probability of uncovering

so few different genotypes in a completely mixed sexual population is $p < 0.001$ (1000 randomizations of the observed data, keeping linkage within the six loci intact; figure 1a). Therefore, we interpret identical genotypes to be members of the same asexually propagating clone. The probability of two isolates having different genotypes increases with physical distance between the samples, indicating localized clonal expansion on the tree or repeated colonizations from multiple different sources (figure 1b). When only different genotypes are considered, there is no correlation between physical and genetic distance. On the second tree (964/Z), there were only four different genotypes among 29 strains, and 20 of these were identical. This precluded any further analysis of differentiation by distance on that tree.

(b) Distribution of genotypes among different oak trees

The distribution of genotypes on the two trees, which are 500 m apart, also indicates local clonal growth. Out of eight genotypes with more than one representative strain, only two were found on both trees ($p < 0.001$, compared to random throws of 13 genotypes on two trees; figure 1). Clustering of identical genotypes on trees in close proximity was also found by Johnson *et al.* (2004). Furthermore, some clones are identical to those found 6 years ago, demonstrating persistence of particular genotypes over this time-scale. For example, clone Y4 had previously been found on the same tree and also on four other trees within a 500 m radius. Clones Y1 and Z1 had been found on different trees as far as 2 km apart. Thus, clones can extend over a considerable distance.

(c) Population differentiation within Europe

A total of 19 strains from several locations in Continental Europe were analysed, including Russia, Latvia, Sweden, Denmark and Spain (table 1). None of the Silwood/Windsor genotypes was identified in samples from outside the UK. Only three strains had genotypes identical to others (two strains from Spain were identical and two Latvian strains were identical to

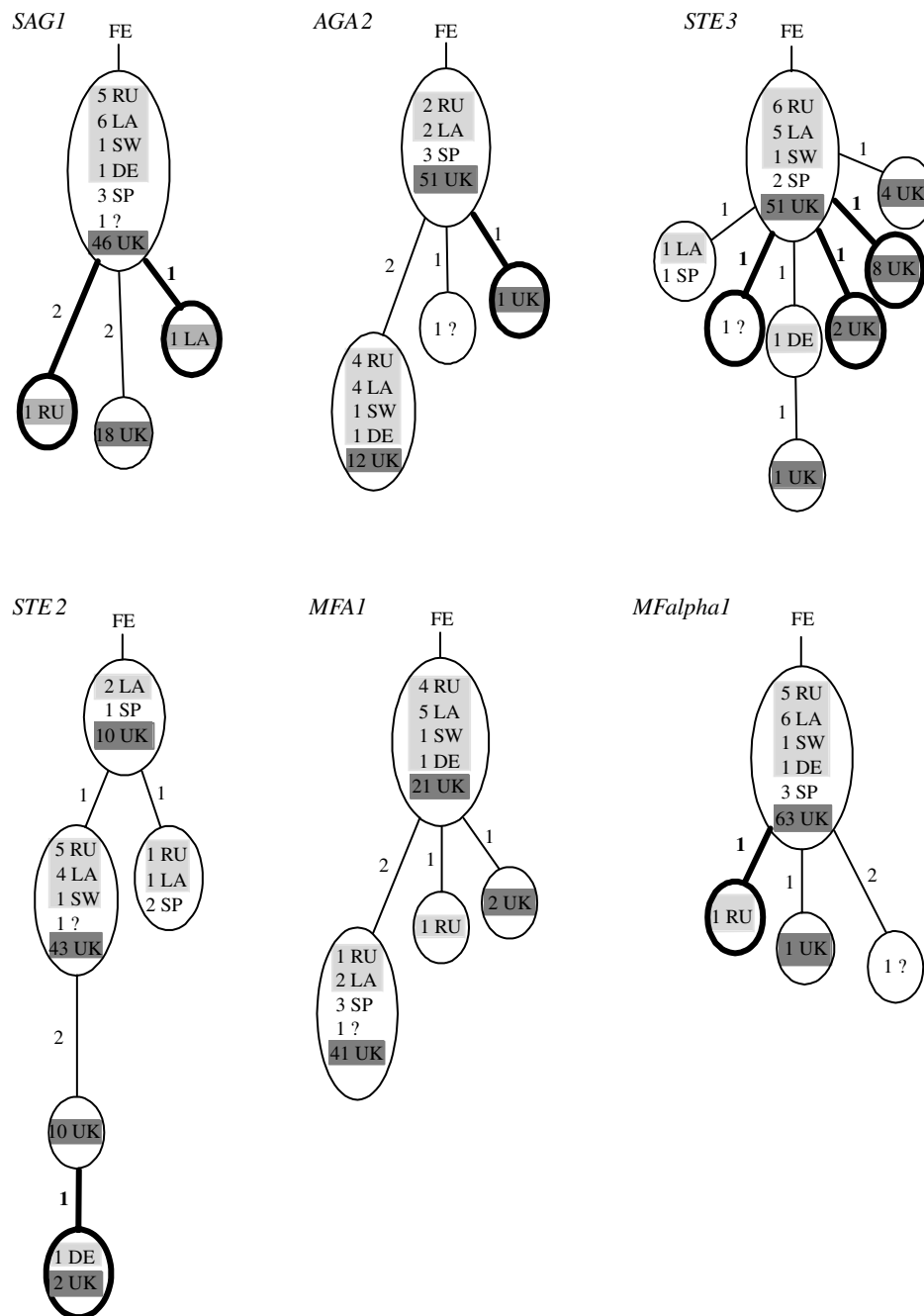


Figure 2. Genealogies for the six gene fragments. A total of 84 strains are shown, 65 from the UK and 19 from other locations in Europe (number of strains followed by location: RU (Russia), LA (Latvia), SW (Sweden), DE (Denmark) and SP (Spain)). Dark-grey shading highlights strains from the UK and light grey from Northeastern Europe. All the trees are perfect, containing no homoplasy, and rooted with strains from Far East Asia; numbers indicate the number of changes on that branch and non-synonymous changes are in bold. Strains from Johnson *et al.* (2004) for which we have only restriction digest data are not included, and for *MFalpha1*, polymorphic sites in the repeat region are also not included.

each other and to a Russian strain), i.e. a frequency of 16%. This contrasts with the frequency of 77% in the UK collection and shows limits to the geographical range of clones. Geographical differentiation between genotypes is also indicated by the analysis of allele frequencies among European isolates (population differentiation between strains in Northeastern Europe, UK and Spain, $\theta = 0.25$, $p < 0.001$).

Despite the significant population differentiation, there are no clear and consistent associations of alleles from different regions in Europe when the genealogies of the six loci are compared (figure 2). There is no homoplasy in any of the genealogies, indicating complete

linkage within the 500–1000 bp sequenced for each gene. The strains from different regions are completely intermingled, and there are no common branches among genealogies, showing extensive recombination between loci and mixing of strains throughout Europe.

(d) *Silent versus expressed polymorphism*

A total of 43 sites are polymorphic when Continental and UK strains are combined, 23 of which are in coding regions and 11 of which are non-synonymous (see electronic supplementary material). The frequency of non-synonymous polymorphisms within the European population (48%) is not significantly

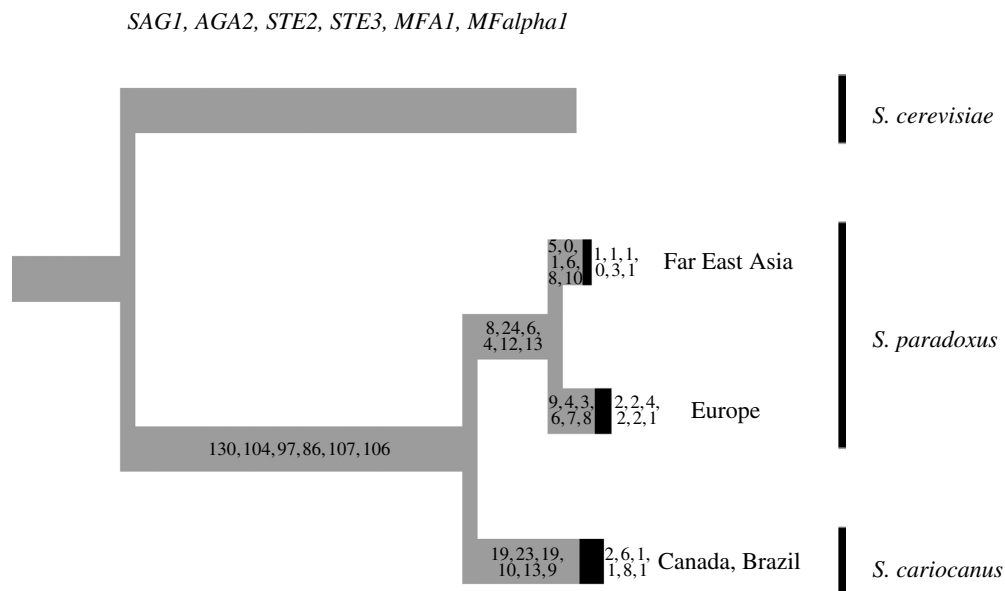


Figure 3. Species tree showing three distinct lineages within *S. paradoxus*; only branches compatible in all six gene genealogies are drawn. Branch lengths are proportional to the average number of changes for the six genes, numbers indicating lengths for *SAG1*, *AGA2*, *STE2*, *STE3*, *MFA1* and *MFalpha1*, respectively. Numbers on the terminal taxa are lengths for the deepest node within the taxon (averages shown in black).

different from the frequency of non-synonymous fixed differences between European and either Far Eastern strains (38%, 31 sites in total) or *S. cerevisiae* (43%, 381 sites). Most of the non-synonymous changes (9 out of 11) are singletons (i.e. found only in a single genotype; electronic supplementary material), in contrast to synonymous changes (3 out of 12; $G=7.9$, $p<0.01$). This difference is consistent with purifying selection acting more strongly on the non-synonymous changes to keep them at low frequency.

(e) Lack of heterozygosity in *S. paradoxus*

Though nearest neighbours are more likely to be clonemates than those further away, there is still a high probability of finding distinct, unrelated genotypes as close as 5 cm apart, thus potentially allowing high rates of outcrossing. However, and despite the high genotypic diversity, we did not find a single heterozygote in our sample, implying that matings rarely occur between different clones. Matings may occur predominantly within the ascus, or, more extremely, between mother and a mitotically produced daughter cell, through autodiploidization, with only very rare incidents of outcrossing (Johnson *et al.* 2004). Very low rates of outcrossing are still sufficient to break linkage disequilibrium in populations, thus creating patterns similar to what we observed here. Lack of heterozygosity was also found in North American *S. paradoxus* (Sniegowski *et al.* 2002; contrasting with much higher heterozygosities for vineyard-associated strains of *S. cerevisiae*; Mortimer 2000) and in the highly selfing *Arabidopsis thaliana* (Nordborg *et al.* 2005).

(f) The global structure of *S. paradoxus*: three distinct lineages

To examine whether mixing of *S. paradoxus* also occurs at a global scale, we analysed sequences from Far East Asia and Canada (table 1). Only one out of nine Asian strains sequenced was identical to another strain from

Asia (strains from several locations), but 16 out of 19 Canadian strains were identical to one of the two genotypes that were in turn fairly divergent from each other (strains from one location; note only three genes sequenced for some strains). The Canadian strains are very different from all other European strains (5% divergence), but nearly identical to the *S. cariocanus* type strain from Brazil (only 1 bp difference in *STE2* and 14 instead of 10 (TA) repeats in the *MFA1* gene fragment). The Asian strains are less divergent from the European *S. paradoxus*, at 1.5% (figure 3). Neither group shares any polymorphisms with strains from Europe, indicating complete sorting of alleles and no recombination between strains at these very distant locations (with the exception of the STOC3 strain sharing alleles at two sites in the *MFalpha1* repeat, probably owing to homoplasy in the repeat; electronic supplementary material). The absence of shared polymorphisms also indicates that there has not been any long-term balancing selection acting on these mating genes. The high divergence and lack of shared polymorphism indicate that strains from the New World and the Far East represent populations that are reproductively isolated from European *S. paradoxus*. Differentiation between strains from different continents was also found in studies of allozymes and transposable element insertions (Naumov *et al.* 1997; Liti *et al.* 2005). These findings are consistent with the reduced viability observed in crosses between North American and European strains of *S. paradoxus* (Sniegowski *et al.* 2002; Greig *et al.* 2003). Curiously, no reduced viability was observed when *S. cerevisiae* strains from the two continents were crossed, suggesting lower spatial differentiation in this domesticated species. Similar results, and a correlation between sequence divergence and reproductive isolation, have been obtained in an analysis of sequence polymorphism in six other genes (G. Liti, E. Louis & D. Barton 2006, personal communication)

4. CONCLUSIONS

We have demonstrated increasing genetic divergence with geographical distance at the following four different spatial scales: on an individual oak tree; between oak trees separated by 500 m; among countries in Europe; and among continents. Such results are expected when dispersal tends to be local and argues against a global dispersal pattern. These results are not surprising, particularly as yeast cells and spores are not typically air borne. In our experience, a Petri dish left uncovered in the lab is unlikely to be colonized by a yeast cell, even if a great deal of yeast is being cultured in the same room. Other fungi also typically show greater genetic distance between isolates collected further apart, including those with air-borne spores (e.g. Burt *et al.* 1997; Koufopanou *et al.* 1997; Taylor *et al.* 2006).

The separation of isolates from different continents into genealogically independent populations suggests that in these yeasts, as in most plants and animals, speciation can occur allopatrically. It will be interesting to see whether isolates collected from locations between Europe and Far Eastern Asia mostly fall into one of the two groups, with a relatively distinct hybrid zone, or whether there is a continuum of divergence.

Whichever pattern is found, the existence of these genealogically distinct populations further adds to the appeal of *S. paradoxus* as a model system for population genetic studies. The existence of these independent replicate populations will be ideal for studies in population genomics to distinguish the random from the repeatable.

Richy Hetherington collected the isolates from Latvia and Sweden. The research was funded by the BBSRC.

REFERENCES

- Aa, E., Townsend, J. P., Adams, R. I., Nielsen, K. M. & Taylor, J. W. 2006 Population structure and gene evolution in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **6**, 702–715.
- Agapow, P. & Burt, A. 2001 Indices of multilocus linkage disequilibrium. *Mol. Ecol. Notes* **1**, 101–102.
- Burt, A., Dechairo, B. M., Koenig, G. L., Carter, D. A., White, T. J. & Taylor, J. W. 1997 Molecular markers reveal differentiation among isolates of *Coccidioides immitis* from California, Arizona and Texas. *Mol. Ecol.* **6**, 781–786. (doi:10.1046/j.1365-294X.1997.00245.x)
- Coyne, J. A. & Orr, H. A. 2004 *Speciation*. Sunderland, MA: Sinauer.
- Fay, J. C. & Benavides, J. A. 2005 Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* **1**, 66–71. (doi:10.1371/journal.pgen.0010005)
- Fenchel, T. & Finlay, B. J. 2004 The ubiquity of small species: patterns of local and global diversity. *BioScience* **54**, 777–784. (doi:10.1641/0006-3568(2004)054[0777:TUOSSP]2.0.CO;2)
- Finlay, B. J. 2002 Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–1063. (doi:10.1126/science.1070710)
- Greig, D., Travisano, M., Louis, E. J. & Borts, R. H. 2003 A role for the mismatch repair system during incipient speciation in *Saccharomyces*. *J. Evol. Biol.* **16**, 429–437. (doi:10.1046/j.1420-9101.2003.00546.x)
- Johnson, L. J., Koufopanou, V., Goddard, M. R., Hetherington, R., Schafer, S. M. & Burt, A. 2004 Population genetics of the wild yeast *Saccharomyces paradoxus*. *Genetics* **166**, 43–52. (doi:10.1534/genetics.166.1.43)
- Koufopanou, V., Burt, A. & Taylor, J. W. 1997 Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis*. *Proc. Natl Acad. Sci. USA* **94**, 5478–5482. (doi:10.1073/pnas.94.10.5478)
- Liti, G., Peruffo, A., James, S. A., Roberts, I. N. & Louis, E. J. 2005 Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex. *Yeast* **22**, 177–192. (doi:10.1002/yea.1200)
- Mortimer, R. K. 2000 Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Res.* **10**, 403–409. (doi:10.1101/gr.10.4.403)
- Naumov, G. I., Naumova, E. S. & 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.
- Naumov, G. I., Naumova, E. S. & Sniegowski, P. D. 1998 *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* are associated with exudates of North American oaks. *Can. J. Microbiol.* **44**, 1045–1050. (doi:10.1139/cjm-44-11-1045)
- Nordborg, M. *et al.* 2005 The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, 1289–1299. (doi:10.1371/journal.pbio.0030196)
- Redzepovic, S., Orlic, S., Sikora, S., Majdak, A. & Pretorius, I. S. 2002 Identification and characterisation of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* strains isolated from Croatian vineyards. *Lett. Appl. Microbiol.* **35**, 305–310. (doi:10.1046/j.1472-765X.2002.01181.x)
- Sniegowski, P. D., Dombrowski, P. G. & Fingerhahn, 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.
- Taylor, J. W., Turner, E., Townsend, J. P., Dettman, J. R. & Jacobson, D. 2006 Eukaryotic microbes, species recognition and the geographical limits of species: examples from the kingdom Fungi. *Phil. Trans. R. Soc. B* **361**, 1947–1963. (doi:10.1098/rstb.2006.1923)
- Weir, B. S. 1996 *Genetic data analysis II*. Sunderland, MA: Sinauer.