MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 4395–4407 doi: 10.1111/j.1365-294X.2011.05217.x

Clonality, genetic diversity and support for the diversifying selection hypothesis in natural populations of a flower-living yeast

C. M. HERRERA, M. I. POZO and P. BAZAGA

Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas (CSIC), Avenida Américo Vespucio s/n, Isla de La Cartuja, 41092 Sevilla, Spain

Abstract

Vast amounts of effort have been devoted to investigate patterns of genetic diversity and structuring in plants and animals, but similar information is scarce for organisms of other kingdoms. The study of the genetic structure of natural populations of wild yeasts can provide insights into the ecological and genetic correlates of clonality, and into the generality of recent hypotheses postulating that microbial populations lack the potential for genetic divergence and allopatric speciation. Ninety-one isolates of the flower-living yeast *Metschnikowia gruessii* from southeastern Spain were DNA fingerprinted using amplified fragment length polymorphism (AFLP) markers. Genetic diversity and structuring was investigated with band-based methods and model- and nonmodel-based clustering. Linkage disequilibrium tests were used to assess reproduction mode. Microsite-dependent, diversifying selection was tested by comparing genetic characteristics of isolates from bumble bee vectors and different floral microsites. AFLP polymorphism (91%) and genotypic diversity were very high. Genetic diversity was spatially structured, as shown by AMOVA (Φ_{st} = 0.155) and clustering. The null hypothesis of random mating was rejected, clonality seeming the prevailing reproductive mode in the populations studied. Genetic diversity of isolates declined from bumble bee mouthparts to floral microsites, and frequency of five AFLP markers varied significantly across floral microsites, thus supporting the hypothesis of diversifying selection on clonal lineages. Wild populations of clonal fungal microbes can exhibit levels of genetic diversity and spatial structuring that are not singularly different from those shown by sexually reproducing plants or animals. Microsite-dependent, divergent selection can maintain high local and regional genetic diversity in microbial populations despite extensive clonality.

Keywords: clonality, diversifying selection, floral nectar, fungal microbes, *Metschnikowia gruessii*, population genetics, yeasts

Received 3 March 2011; revision received 8 June 2011; accepted 16 June 2011

Introduction

The magnitude and spatial organization of intraspecific genetic diversity are crucial determinants of micro- and macroevolutionary processes, including the generation of new species. This justifies the vast amounts of effort devoted to investigate genetic diversity and structuring in a countless number of plants and animals (Loveless & Hamrick 1984; Mitton 1997; Nybom & Bartish 2000; Nybom 2004). For organisms of other kingdoms, however, similar information is comparatively scarce. This applies particularly to natural populations of unicellular fungi, for which only a handful of studies have been so far conducted (Johnson *et al.* 2004; Koufopanou *et al.* 2006; Lachance *et al.* 2008; Wardlaw *et al.* 2009). Studies on the population genetics of fungal microbes in natural environments are important for at least the following

Correspondence: Carlos M. Herrera, Fax: +34 954 621125;
E-mail: herrera@ebd.csic.es

two reasons. First, while the majority of plants and animals reproduce primarily by sexual means, natural populations of fungi can fall along a continuum from prevailing sexual reproduction to largely clonal proliferation (Anderson & Kohn 1998; Litvintseva *et al.* 2003, 2006), and in unicellular fungi, sexual reproduction seems particularly complex and variable (Phaff *et al.* 1978; Wardlaw *et al.* 2009; Zeyl 2009; Murphy & Zeyl 2010). Given the expected relationship between genetic diversity and the rate of clonal versus sexual reproduction (Balloux *et al.* 2003), studies on the genetic structure of natural populations of unicellular fungi are apt to contribute fresh insights into the long-standing debate on the evolutionary origin and implications of sexuality, and particularly on the factors that contribute to maintain genetic diversity when recombination is rare or absent (Milgroom 1996; Kassen & Rainey 2004; Hillis 2007; Zeyl 2009). And second, studies on patterns of genetic differentiation of fungal microbes at different spatial scales are essential to evaluate the generality of the recent hypothesis postulating that, because of large population sizes and high dispersal rates, microbial populations will lack significant geographical divergence and thus the potential for allopatric speciation (Finlay 2002; Fenchel & Finlay 2004). Although this modern version of the classical 'everything is everywhere' hypothesis (O'Malley 2007) may hold true for bacteria and protists (but see Whitaker *et al.* 2003), its validity for fungal microbes has been challenged (Lachance 2004; Koufopanou *et al.* 2006; Taylor *et al.* 2006; Kuehne *et al.* 2007; Wardlaw *et al.* 2009). Population genetics studies of unicellular fungi using genetic markers and spatial scales similar to those ordinarily considered in studies of macroorganisms will help to evaluate the universality of the micro–macro dichotomy implied by the 'everything is everywhere' hypothesis. In particular, patterns of spatial genetic structuring similar to those commonly shown by large organisms will militate against that hypothesis (Wardlaw *et al.* 2009).

This paper presents a population genetics study of the specialized flower-living yeast *Metschnikowia gruessii* Giménez-Jurado (Ascomycota, Metschnikowiaceae), based on amplified fragment length polymorphism (AFLP) genotyping of isolates from a relatively small geographical area in southeastern Spain. We address the following specific questions: (1) Do natural populations of *M. gruessii* harbour significant levels of genetic variation, and if they do, is there some spatial component akin to that commonly exhibited by macroorganisms? (2) Does population structure exhibit genetic signatures of regular recombination or, alternatively, are natural populations of this yeast predominantly clonal? (3) Is there evidence of divergent microsite-dependent selection on clonal lineages that could maintain high genetic diversity in the face of clonality, as postulated by the diversifying selection hypothesis (Ellstrand & Roose 1987; Kassen & Rainey 2004)? Question 1 will be addressed by applying ordinary band-based methods (Bonin *et al.* 2007) and model- and nonmodel-based clustering procedures (Falush *et al.* 2007; Burnier *et al.* 2009) to AFLP genotyping data. Tests based on the analysis of multilocus linkage disequilibrium (Tibayrenc *et al.* 1991; Maynard Smith *et al.* 1993; Burt *et al.* 1996) will be used for answering question 2. Question 3 will be examined by comparing the genetic diversity of isolates obtained from the bumble bee yeast vectors with that of isolates from distinct floral microsites (nectar, anthers, corolla), and by looking for individual AFLP markers whose frequency varies significantly among floral microsites. A reduction in genetic diversity of isolates from bumble bee vectors to individual floral microsites, along with the existence of markers whose frequency is microsite-dependent, will provide supporting evidence for the diversifying selection hypothesis.

Materials and methods

Study organism

Metschnikowia gruessii is an ascomycetous yeast known to date only from west-central Europe, the Iberian Peninsula and Balearic Islands, where it has been invariably found associated with floral surfaces, floral nectar and flower-visiting insects like bumble bees (Hautmann 1924; Giménez-Jurado 1992; Herzberg *et al.* 2002; Brysch-Herzberg 2004; Herrera *et al.* 2010; Pozo *et al.* 2011; C. M. Herrera, unpublished data). Colonizing inocula are brought to flowers by foraging bumble bees (Brysch-Herzberg 2004; Herrera *et al.* 2010). Vegetative cells found in flowers are diploid and proliferate profusely by multilateral budding, reaching densities in floral nectar $> 5 \times 10^4$ cells/mm^3 within 2–4 days of colonization. Under stringent culture conditions, they may produce asci and ascospores and reproduce sexually (Giménez-Jurado *et al.* 1995; Miller & Phaff 1998), but sexual reproduction does not seem to occur naturally in the flower-living populations of our southern Spanish study area. During 2008–2010, vegetative cells of *M. gruessii* were recorded in 423 floral nectar samples from 40 plant species from our Sierra de Cazorla study region examined microscopically, but the morphologically distinctive asci were never seen (C. M. Herrera & M. I. Pozo, unpublished data). The tridental form of *M. gruessii* vegetative cell groups, otherwise known as tetrad or 'airplane-like' configuration (Giménez-Jurado *et al.* 1995), distinguishes morphologically this species from other European nectar-living yeasts (Fig. 1). This

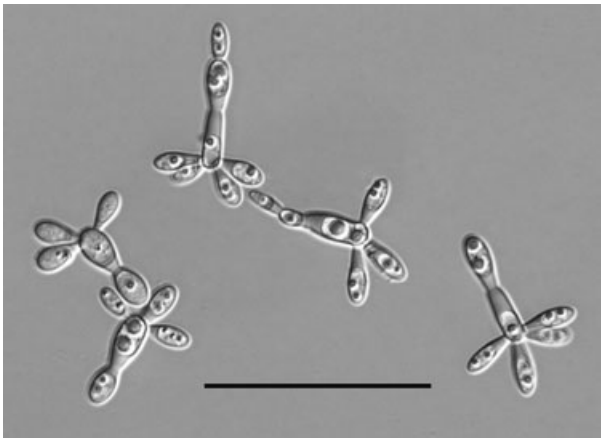

Fig. 1 Vegetative cell groups of *Metschnikowia gruessii* in floral nectar, showing the distinctive tridental or 'airplane-like' shape typical of this species. Nomarski differential-interference microscopy, scale bar 50 μm.

trait has been interpreted as an adaptation for their dispersal by bumble bees (Grüss 1917; Brysch-Herzberg 2004). All isolates considered in this study were identified as *M. gruessii* on the basis of both their morphological characteristics and by two-way sequencing the D1/D2 domain of the 26S ribosomal RNA gene (see Pozo *et al.* 2011), and comparing the results with the corresponding sequence for the type strain (CBS7657) in GenBank (accession number U45737). Fifty-two per cent of our 91 *M. gruessii* isolates had sequences identical to the type strain, 42% had similarities ≥99%, and 6% had similarities between 96% and 98%.

Study area and sampling scheme

This study is based on 91 isolates of *M. gruessii* obtained during spring–summer 2008–2009 from the Sierra de Cazorla region, a well-preserved natural area in Jaén province, southeastern Spain. Owing to resource constraints, a compromise sampling scheme was designed that simultaneously allowed for testing regional genetic differentiation and within-population diversifying selection. For the first goal, a relatively broad area was sparsely sampled by obtaining a small number of isolates from floral nectar per site, which made possible to conduct a conservative test (see Discussion) of genetic differentiation at the geographical scale of the study. For the second goal, bumble bees and different floral microsites were intensively sampled on a narrow temporal window at a single site located in the middle of the sparsely sampled area. Despite our rather limited sampling, however, the number of isolates used in this study is still among the largest considered so far in population genetics studies of wild fungi.

In 2008, 43 isolates were obtained from floral nectar samples of seven plant species collected at nine different localities (Table 1). Collection sites encompassed a broad range of elevations (765–1595 m a.s.l.) and habitat types (mid-elevation mixed pine-oak forests to montane pine woodland). Six sites were located along the upper course of the Guadalquivir River valley and three in the highlands of the adjacent Sierra del Pozo mountain range (Table 1). Distances between nearest and farthest sampling sites were 1.4 and 46.2 km, respectively. Yeast isolates were obtained from single-flower nectar samples using the methods described in detail by Pozo *et al.* (2011), which consisted basically of streaking nectar onto yeast malt plus chloramphenicol agar plates and obtaining isolates from the resulting colonies following morphological criteria. A single isolate was obtained from each individual nectar sample and, with few exceptions, each flower originated from a different plant. In 2009, 30 isolates were obtained during 3–8

Table 1 Collection sites, host plants and sampling microsites for the 91 *Metschnikowia gruessii* isolates considered in this study

					Isolates per sampling microsite		
Collection site	Latitude (ºN)	Longitude (ºW)	Elevation (m a. s. l.)	Host plant	Floral nectar	Other floral microsites*	Bumble bee glossae†
1. Arroyo Frío	37.9302	2.9296	830	*Gladiolus illyricus*	4		
2. Arroyo la Teja	37.9625	2.9099	772	*Prunella grandiflora*	4		
3. Arroyo Saúco	37.9883	2.9061	765	*Teucrium pseudochamaepitys*	4		
4. Fuente Acero	37.8997	2.8535	1485	*Atropa baetica*	6		
5. Hornos	38.2135	2.7169	890	*Phlomis lychnitis*	4		
6. Navahondona	37.8461	2.9623	1595	*Atropa baetica*	4		
7. Parador	37.9142	2.9355	990	*Digitalis obscura*	29	12	18
8. Puente Herrerías	37.8990	2.9397	980	*Antirrhinum australe*	2		
9. Valdecuevas	37.9074	2.8701	1380	*Atropa baetica*	4		

*Corolla plus floral bracts (seven isolates), and anthers (5).
†Isolates originating from the glossae of *Bombus terrestris* (two isolates), *B. pratorum* (11) and *B. pascuorum* (5).

June from floral nectar, anthers, and corolla and floral bract surfaces of 10 neighbouring *Digitalis obscura* (Plantaginaceae) plants at one of the sites sampled in 2008 (Site 7, Table 1). This site was chosen for intensive sampling because of its central location in the sampled area and the abundance of *M. gruessii* in 2008 samples. An additional 18 isolates were obtained on the same dates and plants from the glossae of three bumble bee species caught while foraging at *D. obscura* flowers (*Bombus terrestris*, *B. pratorum* and *B. pascuorum*). Isolates from nectar were obtained as in 2008, and those from corollas, bracts and glossae by gently rubbing these structures onto agar plates. To obtain isolates from pollen, mature anthers were picked with sterile forceps and immersed in sterile water for 48 h. One microlitre of this inoculum was streak-inoculated onto an agar plate. The combined 2008–2009 sample of 91 isolates will be used in the analyses of genetic diversity and tests for clonality. The 2009 subsample, consisting of the 48 isolates originating from bumble bee glossae and floral microsites in Site 7, will be used to test the hypothesis of diversifying selection. In these analyses, isolates from corolla and floral bract surfaces will be combined into a single sample, designated as 'corolla' hereafter.

AFLP genotyping

All *M. gruessii* isolates were DNA fingerprinted using the AFLP technique (see, e.g. Weising *et al.* 2005; Meudt & Clarke 2007; for details on the method and the nature of AFLP markers). Although recent applications of AFLP analyses to fungi have been mostly aimed at discriminating commercial yeast strains (de Barros Lopes *et al.* 1999; Spadaro *et al.* 2008) or assessing broad-scale geographical variation in pathogens (Ivors *et al.* 2004; Kauserud *et al.* 2004; Litvintseva *et al.* 2006), the method has also proven useful to reveal intraspecific differentiation in natural fungal populations (Mueller *et al.* 1996; Arroyo García *et al.* 2002; Laitung *et al.* 2004). In addition, AFLP markers are particularly suitable to genomic scan approaches like the one used here for testing the diversifying selection hypothesis, because they are thoroughly distributed over the whole genome and are often linked to non-neutral, adaptive loci (Bonin *et al.* 2007; Meudt & Clarke 2007), thus being potentially subject to hitchhiking selection (Thomson 1977; Kaplan *et al.* 1989).

The AFLP analysis was performed essentially as originally described by Vos *et al.* (1995), with modifications involving the use of fluorescent dye-labelled selective primers following Applied Biosystems (2005). Restriction–ligation was conducted using *Pst*I/*Mse*I endonuclease mixture and double-stranded adaptors. After an initial screening of primer pair combinations, a total of four *Pst*I + 2/*Mse*I + 2 primer pairs were chosen that provided reliable, consistently scorable results, and each isolate was fingerprinted using these four combinations (Table 2). Fragment separation and detection was made using an ABI PRISM 3100 DNA sequencer. The presence or absence of each marker in each isolate was scored manually by visualizing electrophoregrams with GeneMapper 3.7 software. Genotyping error rate was determined by running intra-plate replicates for a randomly chosen subsample of nine isolates and estimated as the ratio of the total number of loci with discordant scores (all isolates combined) to the product of the number of isolates by the total number of scored loci (Bonin *et al.* 2004). Overall error rate was 4.3%, with slight variations across primer combinations (Table 1).

Data analysis

Analysis of molecular variance (AMOVA; Excoffier *et al.* 1992) was used to partition total genetic diversity into its among- and within-site components. The analysis was based on the pairwise, individual-by-individual genetic distance matrix obtained using a simple matching coefficient, in which any comparison with the same state yields a value of zero, while any comparison of different states yields a value of unity (Huff *et al.* 1993). This is an Euclidean distance (Bonin *et al.* 2007) and is therefore appropriate for AMOVA, which requires an Euclidean metric (Excoffier *et al.* 1992). Pairwise genetic distance among sites (Φ_{st}, an analogue of F_{st}; Excoffier *et al.* 1992) was also obtained from the AMOVA, and their relationship to geographical distance was assessed using Mantel test. Computations were performed with

Table 2 Primer combinations used, number of markers obtained in the size range 150-500 base pairs, observed polymorphism level and estimates of scoring error rates in the amplified fragment length polymorphism analysis of the 91 isolates of *Metschnikowia gruessii* considered in this study

Primer combination	Number of markers	Percentage of polymorphic*	Scoring error rate (%)†
Pst AA–*Mse* CC	40	100	4.3
Pst AT–*Mse* CA	40	85	4.4
Pst AC–*Mse* CA	20	85	5.3
Pst AG–*Mse* CG	18	89	3.6
All combined	118	91	4.3

*A marker was considered polymorphic if at least one isolate showed a variant score.

†Calculated from the nine isolates that were assayed twice as 100 × (number of discordant scores on two analyses)/(number of scored markers × number of individuals).

GENALEX, and statistical significance levels were determined by permutations (Peakall & Smouse 2006).

Spatial genetic structuring was also investigated by examining the assignment of isolates into genetic groups by means of both model- and nonmodel-based clustering methods. For the model-based approach, the program STRUCTURE version 2.3 (Hubisz *et al.* 2009) was used to explore the number of genetic clusters (K) occurring in the sample. A total of 30 replicates were performed of each simulation for K = 1–9, with a burn-in of 50 000, and MCMC of 100 000, assuming admixture and correlated allele frequencies. Because of the extensive multilocus linkage disequilibrium in the sample, binary AFLP scores were treated in these analyses as if they were haploid allelic states. Evanno *et al.*'s (2005) modal ΔK parameter was used as the choice criterion to infer the number of genetic groups. As STRUCTURE may produce anomalous groupings when model assumptions (e.g. Hardy–Weinberg/linkage equilibrium) are violated (Kalinowski 2011), we also performed a nonmodel-based classification of isolates by means of nonhierarchical K-means clustering, largely following the methods outlined in Burnier *et al.* (2009) and Arrigo *et al.* (2010). This technique assigns individuals to a defined number of genetic groups (K) that maximizes a criterion parameter. Two of such parameters were assayed to infer K, namely the intergroup variance ('inertia'; Burnier *et al.* 2009) and the Calinski–Harabasz index (pseudo F-statistic comparing the among-group to the within-group sum of squares of the partition; Borcard *et al.* 2011). We performed 100 000 independent runs starting from random points for each assumed value of K (ranging between 1 and 9) and recording the criterion parameters at each run. Computations for the inertia criterion were performed with the script of Arrigo *et al.* (2010) for the R environment (R Development Core Team 2010), and those for the Calinski–Harabasz index with the cascadeKM function in the vegan package (Oksanen *et al.* 2010).

The number of distinct genotypes occurring in the sample was inferred by examination of the frequency distribution of pairwise genetic distances between isolates ('spectrum of genetic diversity'; Rozenfeld *et al.* 2007), obtained using the GenoType program (Meirmans & Van Tienderen 2004). Even though it was found that all isolates had unique AFLP profiles, isolates belonging to the same clone could appear as different genotypes because of small genotyping differences arising from scoring errors, in which case a bimodal genetic distance distribution with a 'left peak' on or near zero would be expected (Douhovnikoff & Dodd 2003; Meirmans & Van Tienderen 2004; Rozenfeld *et al.* 2007). The height and location of the left peak would depend on the genotype/isolate ratio and the magnitude of scoring error, respectively. The larger the genotype/isolate ratio, the lower the peak's height. At the limit, absence of a left peak would denote a ratio of unity, i.e. each isolate corresponding to a distinct genotype. The smaller the scoring error, the closer to zero will the left peak occur, with errorless scoring leading to a peak on zero. We have used genetic distances between intra-plate replicates of the same isolate to assess the expected location of the left peak in the frequency distributions of genetic distances for our sample. For these analyses, pairwise genetic distances between isolates were computed using both the simple matching coefficient, where double band presence and double band absence are given the same weight, and Jaccard coefficient, which only takes into account the bands present in at least one of the individuals and is therefore unaffected by homoplasic absent bands (Bonin *et al.* 2007). The matrices obtained with the two methods were closely correlated (r = 0.964, P < 0.001, Mantel test with 10 000 repetitions), and only results based on the simple matching coefficient matrix are reported below.

Whether the prevailing reproductive mode of *M. gruessii* in our study region was clonal or sexual was inferred using tests based on the index of association (I_A, a measure of multilocus linkage disequilibrium; Brown *et al.* 1980; Maynard Smith *et al.* 1993) and maximum parsimony tree length (Burt *et al.* 1996), both of which have been often used to test the null hypothesis of extensive recombination in natural fungal populations (Milgroom 1996; Litvintseva *et al.* 2006). The tree length test relies on the notion that, if populations are largely clonal, the phylogenetic tree of isolates will be shorter than if recombination is frequent, because in this latter case, there will be less phylogenetic consistency. Both tests compare parameters for observed data (I_A and tree length) with expectations from freely recombining populations, as estimated by resampling without replacement the observed data set. In the tree length test, loci are treated as phylogenetic characters and alleles as character states, and the length of the most parsimonious tree fitted to the observed data is compared with the length distribution of most parsimonious trees fitted to randomized data sets, with the expectation that extensive clonality would lead to shorter trees. Computations were performed using the programs MultiLocus 1.3 (Agapow & Burt 2001) and Paup* 4.0 (Swofford 1998), and statistical significance assessed with randomization. As linkage disequilibrium can arise when samples from genetically distinct populations are combined, the distribution of isolates among sampling sites was preserved in randomizations by shuffling alleles only among isolates from the same site. In this way, linkage disequilibrium as a result of

population differentiation was maintained in all simulated data sets.

The diversifying selection hypothesis leads to expectations that variable selection among microsites should lead to genetic differences between isolates from distinct floral microsites and that genetic diversity of isolates should decline from bumble bees' mouthparts to individual floral microsites as a consequence of differential, microsite-specific filtering of genotypes. Genetic diversity of isolates from bumble bee glossae and floral microsites was estimated with the Shannon's diversity index (Bonin *et al.* 2007). The genomic scan analysis was based on comparing band frequencies of individual AFLP markers across the three floral microsites. For each marker, a likelihood ratio test was performed on the 2×3 association table between marker score (present–absent) and microsite (nectar, anthers, corolla), and the *P*-value obtained was used to identify significant marker–microsite associations after properly accounting for the multiplicity of significance tests. The latter was accomplished by applying Storey & Tibshirani's (2003) *q*-value method for the estimation of false discovery rates to the set of *P*-values for individual markers obtained from separate tests. Using the QVALUE package (Storey & Tibshirani 2003), we found the largest *q*-value, leading to an expectation of less than one falsely significant marker–microsite association (see, e.g. Herrera & Bazaga 2009 for further details on the application of this method in the context of genomic scans).

Results

Genetic diversity

The four AFLP primer combinations produced a total of 118 fragments in the range 150–500 bp that could be unambiguously scored for all the isolates (Table 2). Considerable genetic diversity was represented in the sample, as 91% of markers were polymorphic. Primer combinations varied little in levels of polymorphism, which was invariably very high (range = 85–100%, Table 2). Loci that were monomorphic for the entire data set ($N = 11$) were omitted from all subsequent analyses.

The 91 isolates had unique AFLP profiles, which suggested that most or all isolates corresponded to different genotypes. This interpretation was supported by the frequency distributions of pairwise genetic distances between isolates (Fig. 2). There was not any hint of bimodality or left peaks in these distributions, and the proportion of pairwise distances falling within the observed range for pairwise distances between replicates of the same isolate (range = 1–6, horizontal segments in Fig. 2) was negligible (0.78%, 1.22% and 6.25% of $N = 4095$, 1711 and 32 pairwise distances in Fig. 2a–c, respectively). Using mean pairwise genetic distance between replicates (mean ± SE = 4.1 ± 0.6) as the threshold for isolates to still be considered clonemates (Meirmans & Van Tienderen 2004), the GenoType program led to expectations of 79 and 51 distinct genotypes in the whole sample and in site 7, respectively, or a genotype/isolate ratio of 0.87. Uniqueness of AFLP

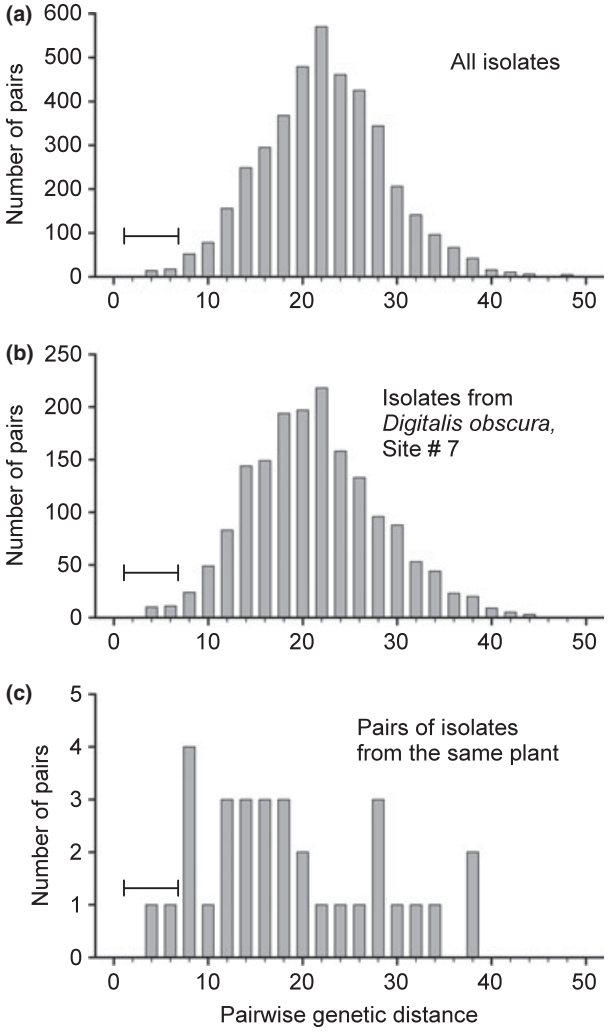

Fig. 2 Frequency distribution of pairwise genetic distances (number of discordant amplified fragment length polymorphism scores) between isolates of *Metschnikowia gruessii* at different spatial scales. (a) Isolates from all sampling sites and microsites combined ($N = 91$ isolates); (b) isolates from *Digitalis obscura* flowers and glossae of visiting bumble bee, Site 7 ($N = 59$ isolates); (c) pairs of isolates originating from different flowers on the same plant, all collection sites and host species combined ($N = 32$ pairs). In each graph, the horizontal segment near the left extreme represents the observed range of genetic distances between replicates of the same isolate in a random subsample of $N = 9$ isolates that were assayed twice in the same plate.

profiles in our sample thus largely reflected actual genotypic differences, with little or no influence of genotyping errors. In addition, genotypic diversity was extensive at all spatial scales considered, because genetic distance distributions for the whole sample (Fig. 2a), isolates from the most thoroughly sampled site 7 (Fig. 2b) and pairs of isolates originating from the same individual host plant (Fig. 2c) were similar in shape and spread.

Spatial structuring

Genetic diversity was spatially structured at the scale of this study, as revealed by the analysis of molecular variance (AMOVA) and model- and nonmodel-based clustering. The AMOVA-based estimate of genetic differentiation between sampling sites, $\Phi_{st} = 0.155$, was significantly greater than zero ($P = 0.01$; permutation test with 1000 repetitions), thus revealing the existence of spatial structuring of genetic diversity in *Metschnikowia gruessii*, with about 16% of total regional diversity being accounted for by variation between sites. Genetic differences between sites, however, did not conform to a simple distance-dependence pattern, as shown by the statistical nonsignificance of the regression between pairwise Φ_{st} values and geographical distance between sites ($r = 0.021$, $P = 0.59$; Mantel test with 1000 permutations).

Results of model- and nonmodel-based clustering were quite similar. STRUCTURE revealed a distinct modal maximum of ΔK at $K = 3$ genetic groups. In *K*-means analyses, the inertia and Calinski-Harabasz criteria yielded similar results. Criteria for $K = 2$ and $K = 3$ were nearly identical, yet group assignments for $K = 3$ matched most closely the distribution of isolates among sampling sites and thus seemed the optimal choice. Membership assignments to the three inferred genetic groups obtained with model- and nonmodel-based clustering revealed similar patterns of geographical structuring (Fig. 3). The highest proportions of membership to group 1 occurred at the six sampling sites located under 1000 m elevation along the Guadalquivir River Valley (sites 1, 2, 3, 5, 7 and 8; elevation 765–990 m), which formed a geographically coherent, fairly homogeneous set with regard to the genetic composition of isolates. Membership to groups 2 and 3 were highest at two (site 6, 1595 m elevation; site 9, 1380 m) and one (site 4, 1485 m) sites each, respectively, all of which were located on the highlands of the Sierra del Pozo mountain range.

Mode of reproduction

The index of association (I_A) and tree length tests yielded similar results, unambiguously rejecting the null hypothesis of free recombination as the prevailing mode of reproduction in the *M. gruessii* populations studied ($P < 0.0001$, randomization tests). Observed I_A was 22.2 standard deviations greater, and the maximum parsimony tree for observed data was 21.8 standard

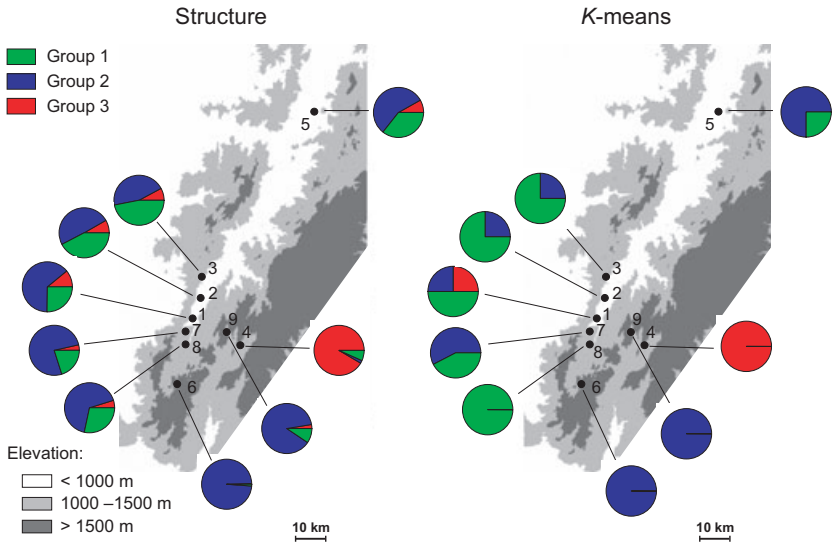

Fig. 3 Population structure of the yeast *Metschnikowia gruessii* in the Sierra de Cazorla study area, southeastern Spain, inferred using the model-based clustering method implemented in the program STRUCTURE (left) and nonmodel-based *K*-means clustering (right). Individual pie charts indicate the estimated (STRUCTURE) or observed (*K*-means) proportion of membership of the isolates at each sampling site for the inferred number of $K = 3$ genetic groups. Numerals refer to sampling site codes in Table 1.

deviations shorter, than the respective expected mean values for simulated freely recombining populations (Fig. 4).

Diversifying selection

In the set of $N = 48$ isolates obtained from site 7 in 2009, the genetic diversity of *M. gruessii* measured with Shannon's information index (I) was significantly heterogeneous among the four sampling sites considered, namely bumble bee glossae and the three floral microsites ($\chi^2 = 11.17$, d.f. = 3, $P = 0.011$; Wilcoxon rank-sum test). Genetic diversity was highest for isolates associated with bumble bee glossae, lowest for those associated with the corolla and intermediate for isolates in nectar and anthers (Fig. 5a). When isolates from all floral microsites were combined into a single sample, the genetic diversity of isolates from bumble bee glossae (mean $I \pm$ SE = 0.338 ± 0.026) did not differ significantly from that of isolates from flowers ($I = 0.312 \pm 0.028$) ($\chi^2 = 0.65$, d.f. = 1, $P = 0.42$; Wilcoxon rank-sum test). These results demonstrate a significant between-floral microsite component to overall genetic diversity, which was corroborated by results of the genomic scan.

Out of a total of 87 polymorphic AFLP markers represented in the sample of isolates from site 7 in 2009, five markers exhibited statistically significant heterogeneity in band frequency among floral microsites, according to our criterion of keeping the expected number of false positives <1 ($\chi^2 \geq 6.79$, d.f. = 2, $P \leq 0.033$, q-value ≤ 0.188; expected number of false positives = $0.188 \times 5 = 0.94$). These markers exhibited broad variation in band frequency among sets of isolates obtained from nectar, anthers and corolla (Fig. 5b).

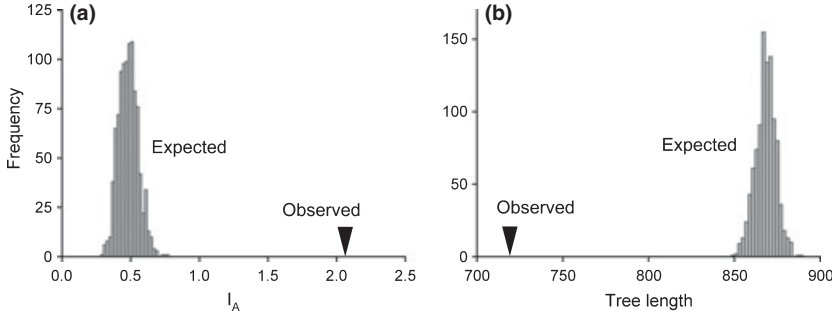

Fig. 4 Summary of the results of two tests for the prevalence of recombination in the sample of *Metschnikowia gruessii* isolates studied. (a) comparison of the observed index of association (I_A) in the data, a measure of multilocus linkage disequilibrium, with a simulated distribution in which an infinite amount of recombination has been imposed on the data by randomly shuffling the alleles amongst individuals, independently for each locus. (b) results of phylogenetic test for linkage disequilibrium, in which the length of the most parsimonious tree fitted to the observed data is compared with the length distribution of the most parsimonious trees for simulated data sets with infinite amount of recombination.

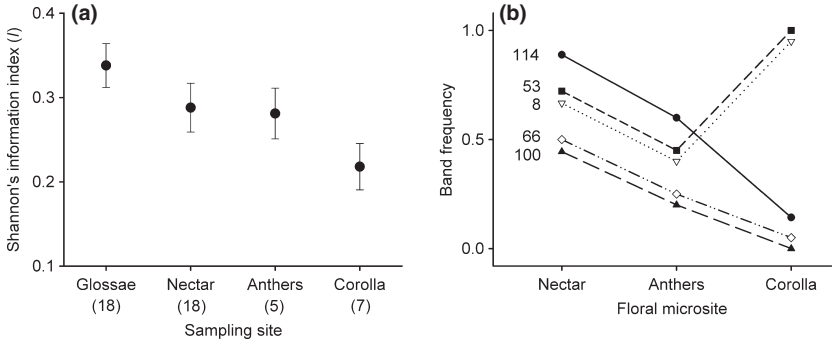

Fig. 5 (a) Variation in genetic diversity, as measured with Shannon's information index, among sets of *Metschnikowia gruessii* isolates obtained from bumble bee glossae and the three floral microsites sampled in 2009 from site 7 (number of isolates per group in parentheses). Vertical segments extend over ± 1 SE around estimates (filled dots). (b) Small-scale variation in band frequency for five amplified fragment length polymorphism markers exhibiting significant variation between groups of isolates originating from the three floral microsites. Each line corresponds to a different marker, identified by primer combination and fragment size (base pairs) as follows: 8, PstAA_MseCC_169; 53, PstAT_MseCA_217; 66, PstAT_MseCA_339; 100, PstAC_MseCA_483; 114, PstAG_MseCG_375.

Discussion

Genetic diversity and structuring

This study has shown that southeastern Spanish populations of the specialized flower-living yeast *Metschnikowia gruessii* harbour considerable genetic diversity over limited spatial scales. Overall polymorphism of AFLP markers (91%) nearly doubled the average polymorphism reported by earlier AFLP-based investigations of genetic variation in fungal taxa, some of which were undertaken over broader geographical areas (mean polymorphism = 50% for 13 species reviewed by Kauserud *et al.* 2004). Although per cent polymorphism is difficult to compare across studies because the investigated loci are usually preselected for high levels of polymorphism (Nybom 2004), the preceding comparison suggests that genetic diversity in *M. gruessii* is comparable or superior to that of most fungi studied so far using similar markers. Considerable genotypic diversity was also represented in our sample, where most or all isolates corresponded to unique multilocus genotypes. It must be noted, however, that our sampling design probably overestimated the genotype/isolate ratio, because only one isolate was recovered from each single-flower nectar sample. Given the fast clonal proliferation of this yeast following the initial colonization of flowers by inocula, our sample would undoubtedly have contained a certain proportion of identical genotypes had a sufficient number of isolates been recovered from each nectar sample.

Irrespective of possible biases in the genotype/isolate ratio, an important aspect revealed by this study is that many different multilocus genotypes of *M. gruessii* coexisted locally on the same flowers and dates. This is illustrated by the results for the thoroughly sampled site 7, where many unique genotypes were recovered from the small number of flowering *Digitalis oscura* plants and foraging bumble bees sampled during a few days in June 2009. This level of local genotypic diversity is far higher than that commonly reported for local populations of fungi, including some yeasts (Anderson & Kohn 1995, 1998; Bonello *et al.* 1998; Koufopanou *et al.* 2006), although the difference might reflect the modest number of isolates examined per locality in most earlier studies rather than a biological reality. In any case, the present results show that when a sufficiently large number of sympatric isolates are genotyped using an ample set of variable markers, local genotypic diversity of fungal populations can turn out to be much higher than traditionally implied (see also Lindqvist-Kreuze *et al.* 2002; Laitung *et al.* 2004).

Genetic diversity of *M. gruessii* was spatially structured at the scale of this study, as shown by the statistical significance of Φ_{st}, an AMOVA-based estimator of F_{st} that provides reasonably accurate estimates of genetic differentiation when used with dominant AFLP markers (Bonin *et al.* 2007). The observed value ($\Phi_{st} = 0.155$) reflects a substantial level of genetic differentiation, implying that about 16% of regional genetic variability was attributed to differences among sampling sites, a figure comparable to those reported for other fungal (Arroyo García *et al.* 2002; Laitung *et al.* 2004) and plant populations (Nybom 2004) sampled over equivalent spatial scales. It must be stressed that the AMOVA analysis was able to reveal significant genetic differentiation among sites despite the low statistical power (i.e. high type II error) expected a priori from the small number of isolates obtained per sampling locations (e.g. Ryman *et al.* 2006). The sparse sampling of the study area thus reinforce, rather than weaken, our conclusion that genetic variation in *M. gruessii* was spatially structured at the scale of this study.

The absence of a significant relationship between pairwise genetic and geographical distance between sampling sites (see also Lindqvist-Kreuze *et al.* 2002; Laitung *et al.* 2004; Wardlaw *et al.* 2009; for similar results), along with the broad scatter of Φ_{st} values for any given geographical distance (data not shown), would suggest a lack of regional equilibrium between gene flow and drift, with drift being more influential than gene flow (Hutchison & Templeton 1999). This parsimonious interpretation, however, must be taken with caution in view of the results of clustering analyses. Irrespective of the analytical technique used, clustering analyses revealed that isolates fell into three distinct genetic groups whose distribution roughly matched gross landscape features (mid-elevation Guadalquivir valley vs. Sierra del Pozo highlands). In addition to the interplay between drift and gene flow, the observed pattern of genetic structuring could reflect historical or ecological contingencies, and constraints on gene flow imposed by environmental features (e.g. sharp habitat boundaries or topographical barriers). The relative importance of these factors as determinants of genetic structuring in *M. gruessii* would only be elucidated by future studies combining a denser sampling of the study region with the application of genetic markers suitable for phylogeographical inference (e.g. Koufopanou *et al.* 2006; Wardlaw *et al.* 2009). In the meantime, our results join a growing list of studies showing that, when adequately sampled, wild populations of free-living fungal microbes can exhibit patterns of genetic diversity, structuring and differentiation that are not singularly different from those ordinarily documented for plants and animals (Johnson *et al.* 2004; Koufopanou *et al.* 2006; Lachance *et al.* 2008; Wardlaw *et al.* 2009). These findings undermine the purported universality of premises underlying recent flavours of the 'everything

is everywhere' hypothesis, particularly the assumption that far-reaching dispersal and unrestricted gene flow will constrain geographical divergence and the potential for allopatric speciation in microbes (Finlay 2002; Fenchel & Finlay 2004).

Clonality and genetic diversity

A common approach to study the mode of reproduction of natural fungal populations has been to examine whether observed patterns of genetic variability are consistent with clonality or sexual reproduction (Tibayrenc *et al.* 1991; Maynard Smith *et al.* 1993; Burt *et al.* 1996; Milgroom 1996; Litvintseva *et al.* 2006), and this was the approach adopted here. Results of the index of association (I_A) and tree length tests were unambiguous in rejecting the null hypothesis of random mating in *M. gruessii*, despite the fact that our sampling design (a single isolate recovered per nectar sample) most likely led to substantial underestimation of the level of linkage disequilibrium (Prugnolle & De Meeûs 2010). This strongly suggests that clonality is the predominant or sole reproductive mode in the populations of *M. gruessii* studied and that sexual reproduction must be rare or absent, because even infrequent recombination events can have significant effects on population structure and reduce linkage disequilibrium considerably (Burt *et al.* 1996; Milgroom 1996). This conclusion is consistent with our microscopical observations of many thousands of free-living cells of this yeast in floral nectar, revealing profuse vegetative proliferation of cells but never sexual forms (C. M. Herrera & M. I. Pozo, unpublished data), a negative evidence that provides compelling support for prevailing asexuality (Birky 2010). The possibility remains, however, that sexual reproduction can take place occasionally away from flowers (e.g. in bumble bee nests). Our results for *M. gruessii* contribute to further illustrate the variability in mode of reproduction exhibited by natural populations of yeasts of the *Metschikowia* clade, which in the few species studied so far ranges from predominantly sexual to strictly clonal (Lachance *et al.* 2008; Wardlaw *et al.* 2009). This variability of reproductive modes, combined with a broad range of life histories, ecological features, and biogeographical affinities (Giménez-Jurado *et al.* 1995; Miller & Phaff 1998; Lachance *et al.* 2001, 2006), makes this species-rich group an excellent study system for investigating the ecological correlates, and genetic and evolutionary implications, of clonality vs. sexuality within a relatively homogeneous phylogenetic background.

The most intriguing result of this study was that prevailing clonality in *M. gruessii* did not entail any dramatic genetic impoverishment of its populations in terms of multilocus genotypes, nucleotide polymorphism or population differentiation, as predicted from theoretical models (Balloux *et al.* 2003). High local genotypic diversity has been rarely reported from clonal fungal populations (Anderson & Kohn 1995, 1998), although this could be a consequence of limited sampling as noted above. In thoroughly sampled populations of clonal plants, high local genotypic diversity is fairly common (Ellstrand & Roose 1987; Rozenfeld *et al.* 2007). High local genotypic diversity in clonal organisms has been traditionally interpreted in terms of the diversifying selection hypothesis, postulating that if the fitness of different clones differs across microsites and/or environmental conditions (i.e. there is a significant genotype × environment interaction), environmental heterogeneity will favour the long-term persistence of different clonal lineages at different sites, thus eventually maintaining overall genotypic diversity (Burdon 1980; Ellstrand & Roose 1987; Kassen & Rainey 2004). Properly testing this hypothesis in natural environments confronts a number of practical and conceptual difficulties, as stressed by Ellstrand & Roose (1987) for clonal plants. In the case of microorganisms, experimental studies conducted in artificial laboratory microcosms have often shown that complex, heterogeneous environments are more likely to maintain genotypic diversity than simpler environments (Kassen & Rainey 2004; Replansky & Bell 2009; Slater *et al.* 2010), but we are not aware of any test for microbes in natural scenarios. Results for *M. gruessii*, demonstrating a significant between-floral microsite component to local genetic diversity and the existence of microsite-dependent divergent selection, provide compelling evidence for the diversifying selection hypothesis. Our results also illustrate the potential of genomic scans to unveil fine-scale genotype x environment interactions in natural scenarios (see also Herrera & Bazaga 2009).

The interpretation of genetic heterogeneity of isolates across floral microsites as supporting the diversifying selection hypothesis rests on the assumption that all genotypes had similar probabilities of colonizing all microsites, so that eventual differences between microsites in band frequency of some AFLP markers reflected a differential ability of genotypes to thrive in contrasting microenvironments. Circumstantial evidence supports this assumption. Bumble bees were the near-exclusive pollinators of *D. obscura* at site 7 (C. M. Herrera & M. I. Pozo, unpublished data) and the vectors of *M. gruessii* to flowers, which they 'inoculate' with yeasts while probing for nectar and pollen (Brysch-Herzberg 2004; Canto *et al.* 2008; Herrera *et al.* 2010). Given their close proximity, it is most likely that the initial distribution of yeast inocula among floral microsites occurred at random with regard to genotype. The similarity in genetic

diversity between isolates from bumble bee glossae and those from all floral microsites combined is consistent with this hypothesis. Following their initial arrival to flowers, yeast inocula will face different microenvironmental conditions depending on particularities of the floral microsite (e.g. water deficit, nutrient availability, exposure and plant secondary metabolites). Subsequent filtering of genotypes (i.e. selection) according to their differential capacities to grow in such contrasting microenvironments will lead to the observed pattern of reduced genetic diversity within microsites and genetic divergence between microsites in non-neutral or linked to non-neutral loci.

We hypothesize that the simultaneous operation at various ecological scales of site-dependent, diversifying selection similar or more extreme than that documented here at the within-flower level must be a central factor in the long-term maintenance of regional genetic diversity in clonal *M. gruessii*. In our study area, this yeast occurs during April–October in the flowers of at least 40 different species of insect-pollinated plants belonging to many plant families (Pozo *et al.* 2011; C. M. Herrera, unpublished data). Many biotic and abiotic floral features potentially influencing yeast growth, including nectar characteristics, are expected to vary widely across the yeast's disparate group of host plants, many of which coexist locally, flower synchronously and share pollinators. Differences between yeast genotypes in their ability to grow in the nectar from different plant species, or even different individuals of the same species, will ultimately contribute to maintain high regional genotypic diversity in the face of prevailing clonality. Our hypothesis leads to predict that (i) yeast isolates obtained from locally coexisting plant species or individuals that flower simultaneously and share pollinators should exhibit measurable genetic differences in the form of variation in putatively non-neutral loci; (ii) these genetic differences should be related to phenotypic variation among genotypes in their physiological ability to cope with variable microenvironments; and (iii) regional genetic diversity of floral yeast populations should be directly related to the regional diversity of host plants. We will test these expectations in future studies.

Acknowledgements

We thank Laura Cabral and Esmeralda López for laboratory assistance; Pedro A. Tíscar and the Centro de Capacitación y Experimentación Forestal de Vadillo-Castril for laboratory space and facilities; Clara de Vega and Mónica Medrano for discussion; Nils Arrigo for the *R* script for *K*-means computations; and three reviewers for useful suggestions. Computer-intensive analyses were carried out at the SVG cluster of the Centro de Supercomputación de Galicia (CESGA). Permission to work in the Sierra de Cazorla was granted by the Consejería de Medio Ambiente, Junta de Andalucía. Support for this work provided by grants CGL2006-01355, CGL2010-15964 and EXPLORA CGL2007-28866-E/BOS (Ministerio de Educación y Ciencia, Gobierno de España). M.I.P. was supported by a predoctoral fellowship from the Spanish Ministerio de Educación y Ciencia.

References

Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, **1**, 101–102.

Anderson JB, Kohn LM (1995) Clonality in soilborne, plant-pathogenic fungi. *Annual Review of Phytopathology*, **33**, 369–391.

Anderson JB, Kohn LM (1998) Genotyping gene genealogies and genomics bring fungal population genetics above ground. *Trends in Ecology and Evolution*, **13**, 444–449.

Applied Biosystems (2005) *AFLP Plant Mapping Protocol*. Applied Biosystems, Foster City, California.

Arrigo N, Felber F, Parisod C *et al.* (2010) Origin and expansion of the allotetraploid *Aegilops geniculata*, a wild relative of wheat. *New Phytologist*, **187**, 1170–1180.

Arroyo García R, Martínez Zapater JM, García Criado B, Zabalgogeazcoa I (2002) Genetic structure of natural populations of the grass endophyte *Epichloë festucae* in semiarid grasslands. *Molecular Ecology*, **11**, 355–364.

Balloux F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics*, **164**, 1635–1644.

de Barros Lopes M, Rainieri S, Henschke PA, Langridge P (1999) AFLP fingerprinting for analysis of yeast genetic variation. *International Journal of Systematic Bacteriology*, **49**, 915–924.

Birky CW (2010) Positively negative evidence for asexuality. *Journal of Heredity*, **101**, S42–S45.

Bonello P, Bruns TD, Gardes M (1998) Genetic structure of a natural population of the ectomycorrhizal fungus *Suillus pungens*. *New Phytologist*, **138**, 533–542.

Bonin A, Bellemain E, Eidesen PB *et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.

Borcard D, Gillet F, Legendre P (2011) *Numerical Ecology with R*. Springer, New York, NY.

Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics*, **96**, 523–536.

Brysch-Herzberg M (2004) Ecology of yeasts in plant-bumblebee mutualism in Central Europe. *FEMS Microbiology Ecology*, **50**, 87–100.

Burdon JJ (1980) Intra-specific diversity in a natural population of *Trifolium repens*. *Journal of Ecology*, **68**, 717–735.

Burnier J, Buerki S, Arrigo N *et al.* (2009) Genetic structure and evolution of Alpine polyploid complexes: *Ranunculus kuepferi* (Ranunculaceae) as a case study. *Molecular Ecology*, **18**, 3730–3744.

Burt A, Carter DA, Koenig GL, White TJ, Taylor JW (1996) Molecular markers reveal cryptic sex in the human pathogen

Coccidioides immitis. *Proceedings of the National Academy of Sciences, USA*, **93**, 770–773.

Canto A, Herrera CM, Medrano M, Pérez R, García IM (2008) Pollinator foraging modifies nectar sugar composition in *Helleborus foetidus* L. (Ranunculaceae): an experimental test. *American Journal of Botany*, **95**, 315–320.

Douhovnikoff V, Dodd RS (2003) Intra-clonal variation and a similarity threshold for identification of clones: application to *Salix exigua* using AFLP molecular markers. *Theoretical and Applied Genetics*, **106**, 1307–1315.

Ellstrand NC, Roose ML (1987) Patterns of genotypic diversity in clonal plant species. *American Journal of Botany*, **74**, 123–131.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Fenchel T, Finlay BJ (2004) The ubiquity of small species: patterns of local and global diversity. *BioScience*, **54**, 777–784.

Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. *Science*, **296**, 1061–1063.

Giménez-Jurado G (1992) *Metschnikowia gruessii* sp. nov., the teleomorph of *Nectaromyces reukaufii* but not of *Candida reukaufii*. *Systematic and Applied Microbiology*, **15**, 432–438.

Giménez-Jurado G, Valderrama MJ, Sá-Nogueira I, Spencer-Martins I (1995) Assessment of phenotypic and genetic diversity in the yeast genus *Metschnikowia*. *Antonie van Leeuwenhoek*, **68**, 101–110.

Grüss J (1917) Die Anpassung eines Pilzes (*Anthomyces reukaufii*) an den Blütenbau und den Bienenrüssel. *Berichte der Deutschen Botanischen Gesellschaft*, **35**, 746–761.

Hautmann F (1924) Über die Nektarhefe *Anthomyces reukaufii*. *Archiv für Protistenkunde*, **48**, 213–244.

Herrera CM, Bazaga P (2009) Quantifying the genetic component of phenotypic variation in unpedigreed wild plants: tailoring genomic scan for within-population use. *Molecular Ecology*, **18**, 2602–2614.

Herrera CM, Canto A, Pozo MI, Bazaga P (2010) Inhospitable sweetness: nectar filtering of pollinator-borne inocula leads to impoverished, phylogenetically clustered yeast communities. *Proceedings of the Royal Society B*, **277**, 747–754.

Herzberg M, Fischer R, Titze A (2002) Conflicting results obtained by RAPD-PCR and large-subunit rDNA sequences in determining and comparing yeast strains isolated from flowers: a comparison of two methods. *International Journal of Systematic and Evolutionary Microbiology*, **52**, 1423–1433.

Hillis DM (2007) Asexual evolution: can species exist without sex? *Current Biology*, **17**, R543–R544.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Huff DR, Peakall R, Smouse PE (1993) RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloë dactyloides* (Nutt.) Engelm]. *Theoretical and Applied Genetics*, **86**, 927–934.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.

Ivors KL, Hayden KJ, Bonants PJM, Rizzo DM, Garbelotto M (2004) AFLP and phylogenetic analyses of North American and European populations of *Phytophthora ramorum*. *Mycological Research*, **108**, 378–392.

Johnson LJ, Koufopanou V, Goddard MR *et al.* (2004) Population genetics of the wild yeast *Saccharomyces paradoxus*. *Genetics*, **166**, 43–52.

Kalinowski ST (2011) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity*, **106**, 625–632.

Kaplan NL, Hudson RR, Langley CH (1989) The ''hitchhiking effect'' revisited. *Genetics*, **123**, 887–899.

Kassen R, Rainey PB (2004) The ecology and genetics of microbial diversity. *Annual Review of Microbiology*, **58**, 207–231.

Kauserud H, Schmidt O, Elfstrand M, Hogberg N (2004) Extremely low AFLP variation in the European dry rot fungus (*Serpula lacrymans*): implications for self/nonself-recognition. *Mycological Research*, **108**, 1264–1270.

Koufopanou V, Hughes J, Bell G, Burt A (2006) The spatial scale of genetic differentiation in a model organism: the wild yeast *Saccharomyces paradoxus*. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **361**, 1941–1946.

Kuehne HA, Murphy HA, Francis CA, Sniegowski PD (2007) Allopatric divergence, secondary contact and genetic isolation in wild yeast populations. *Current Biology*, **17**, 407–411.

Lachance MA (2004) Here and there or everywhere? *BioScience*, **54**, 884.

Lachance MA, Starmer WT, Rosa CA *et al.* (2001) Biogeography of the yeasts of ephemeral flowers and their insects. *FEMS Yeast Research*, **1**, 1–8.

Lachance MA, Anderson TM, Starmer WT (2006) A new subclade of haplontic *Metschnikowia* species associated with insects of morning glory flowers in Africa and description of *Metschnikowia aberdeeniae* sp nov. *International Journal of Systematic and Evolutionary Microbiology*, **56**, 1141–1145.

Lachance MA, Lawrie D, Dobson J, Piggott J (2008) Biogeography and population structure of the Neotropical endemic yeast species *Metschnikowia lochheadii*. *Antonie van Leeuwenhoek*, **94**, 403–414.

Laitung B, Chauvet E, Feau N *et al.* (2004) Genetic diversity in *Tetrachaetum elegans*, a mitosporic aquatic fungus. *Molecular Ecology*, **13**, 1679–1692.

Lindqvist-Kreuze H, Koponen H, Valkonen JPT (2002) Variability of *Peronospora sparsa* (syn. *P. rubi*) in Finland as measured by amplified fragment length polymorphism. *European Journal of Plant Pathology*, **108**, 327–335.

Litvintseva AP, Marra RE, Nielsen K *et al.* (2003) Evidence of sexual recombination among *Cryptococcus neoformans* serotype A isolates in sub-Saharan Africa. *Eukaryotic Cell*, **2**, 1162–1168.

Litvintseva AP, Thakur R, Vilgalys R, Mitchell TG (2006) Multilocus sequence typing reveals three genetic

subpopulations of *Cryptococcus neoformans* var. *grubii* (Serotype A), including a unique population in Botswana. *Genetics*, **172**, 2223–2238.

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, **15**, 65–95.

Maynard Smith J, Smith NH, Orourke M, Spratt BG (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences, USA*, **90**, 4384–4388.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances *Trends in Plant Science*, **12**, 106–117.

Milgroom MG (1996) Recombination and the multilocus structure of fungal populations. *Annual Review of Phytopathology*, **34**, 457–477.

Miller MW, Phaff HJ (1998) *Metschnikowia* Kamienski. In: *The Yeasts, a Taxonomic Study* (eds Kurtzman CP, Fell JW), 4th edn, Elsevier, Amsterdam, The Netherlands. 256–267

Mitton JB (1997) *Selection in Natural Populations*. Oxford University Press, Oxford, UK.

Mueller UG, Lipari SE, Milgroom MG (1996) Amplified fragment length polymorphism (AFLP) fingerprinting of symbiotic fungi cultured by the fungus-growing ant *Cyphomyrmex minutus*. *Molecular Ecology*, **5**, 119–122.

Murphy HA, Zeyl CW (2010) Yeast sex: surprisingly high rates of outcrossing between asci. *PLoS ONE*, **5**, e10461.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, **13**, 1143–1155.

Nybom H, Bartish IV (2000) Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. *Perspectives in Plant Ecology, Evolution and Systematics*, **3**, 93–114.

Oksanen J, Blanchet FG, Kindt R *et al.* (2010). vegan: Community Ecology Package. R package version 1.17-5. Available over the Internet at http://cran.r-project.org/web/packags/vegan/

O'Malley MA (2007) The nineteenth century roots of 'everything is everywhere'. *Nature Reviews Microbiology*, **5**, 647–651.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Phaff HJ, Miller MW, Mrak EM (1978) *The Life of Yeasts*, 2nd edn. Harvard University Press, Cambridge, Massachusetts.

Pozo MI, Herrera CM, Bazaga P (2011) Species richness of yeast communities in floral nectar of southern Spanish plants. *Microbial Ecology*, **61**, 82–91.

Prugnolle F, De Meeûs T (2010) Apparent high recombination rates in clonal parasitic organisms due to inappropriate sampling design. *Heredity*, **104**, 135–140.

R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Replansky T, Bell G (2009) The relationship between environmental complexity, species diversity and productivity in a natural reconstructed yeast community. *Oikos*, **118**, 233–239.

Rozenfeld AF, Arnaud-Haond S, Hernández-García E *et al.* (2007) Spectrum of genetic diversity and networks of clonal organisms. *Journal of the Royal Society Interface*, **4**, 1093–1102.

Ryman N, Palm S, André C *et al.* (2006) Power for detecting genetic divergence: differences between statistical methods and marker loci. *Molecular Ecology*, **15**, 2031–2045.

Slater FR, Bruce KD, Ellis RJ, Lilley AK, Turner SL (2010) Determining the effects of a spatially heterogeneous selection pressure on bacterial population structure at the sub-millimetre scale. *Microbial Ecology*, **60**, 873–884.

Spadaro D, Sabetta W, Acquadro A *et al.* (2008) Use of AFLP for differentiation of *Metschnikowia pulcherrima* strains for postharvest disease biological control. *Microbiological Research*, **163**, 523–530.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences, USA*, **100**, 9440–9445.

Swofford DL (1998) *PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). Version 4*. Sinauer Associates, Sunderland, Massachusetts.

Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D (2006) Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **361**, 1947–1963.

Thomson G (1977) Effect of a selected locus on linked neutral loci. *Genetics*, **85**, 753–788.

Tibayrenc M, Kjellberg F, Arnaud J *et al.* (1991) Are eukaryotic microorganisms clonal or sexual? A population genetics vantage *Proceedings of the National Academy of Sciences, USA*, **88**, 5129–5133.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Wardlaw AM, Berkers TE, Man KC, Lachance MA (2009) Population structure of two beetle-associated yeasts: comparison of a New World asexual and an endemic Nearctic sexual species in the *Metschnikowia* clade. *Antonie van Leeuwenhoek*, **96**, 1–15.

Weising K, Nybom H, Wolff K, Kahl G (2005) *DNA Fingerprinting in Plants. Principles, Methods, and Applications*, 2nd edn. CRC Press, Boca Raton, Florida, USA.

Whitaker RJ, Grogan DW, Taylor JW (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science*, **301**, 976–978.

Zeyl C (2009) The role of sex in fungal evolution. *Current Opinion in Microbiology*, **12**, 592–598.

Data accessibility

AFLP fingerprints deposited at DRYAD entry doi:10.5061/dryad.92h23.