Fungal multilocus sequence typing — it's not just for bacteria

John W Taylor* and Matthew C Fisher†

Multilocus sequence typing uses nucleotide sequence from several genes to identify individual microbial pathogens. The data obtained for multilocus sequence typing can be used to recognize fungal species and to determine if the fungi are purely clonal, or if they also recombine. Genetic regions with more polymorphisms and microsatellites might be used to recognize populations within species and are well suited to Bayesian methods of assigning unknown individuals to populations of origin. Knowledge of species, populations and reproductive mode can help answer questions common to all emerging diseases: is the disease due to the recent spread of a pathogen, to the emergence of a virulent strain of an existing pathogen, or to a change in the environment that promotes disease?

Addresses
*Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720-3102, USA
e-mail: jtaylor@socrates.berkeley.edu
†Institute of Zoology, Regent's Park, London, NW1 4RY, UK
e-mail: matthew.fisher@ioz.ac.uk

Current Opinion in Microbiology 2003, **6**:351–356

This review comes from a themed issue on
Host–microbe interactions: fungi
Edited by Bruce Klein

1369-5274/$ – see front matter
© 2003 Elsevier Science Ltd. All rights reserved.

DOI 10.1016/S1369-5274(03)00088-2

Abbreviations

MLST	multilocus sequence typing
MLEE	multilocus enzyme electrophoresis
RAPDs	randomly amplified polymorphic DNA
SNP	single nucleotide polymorphisms
MLMT	multilocus microsatellite typing
GCPSR	genealogical concordance phylogenetic species recognition

Introduction: what is multilocus sequence typing?

Multilocus sequence typing (MLST) [1] uses nucleotide sequence from ca. 500 nucleotides of each of ten or so housekeeping genes to characterize genetic diversity in bacterial species that cause human disease. Although originally lacking an acronym, an almost identical technique has been used by mycologists to study basic evolutionary features of human, animal and plant pathogenic fungi [2–5,6••,7,8,9••,10–14]. MLST has become popular because it directly samples the tremendous polymorphism present in nucleotide sequences and because each new study can use and add to all previously obtained data. As a consequence, the dataset for each species continues to improve and can be made available to all interested parties from a source on the internet. MLST has replaced older methods that hide variation, such as MLEE (multilocus enzyme electrophoresis), or that required increasing numbers of comparisons among known genotypes as each new genotype was added, such as DNA–DNA hybridization, or that were sensitive to small changes in the laboratory environment, such as electrophoretic karyotyping or randomly amplified polymorphic DNA (RAPDs) [15]. MLST also has the advantage over single nucleotide polymorphism (SNP) analysis that new polymorphic nucleotide positions in any of the gene fragment sequence can be detected and added to the database. This feature makes it possible to add new individuals from new geographic locations to the study without the danger that variation found to be polymorphic in the initial population will be monomorphic in the newly added ones, as can happen with SNPs [16].

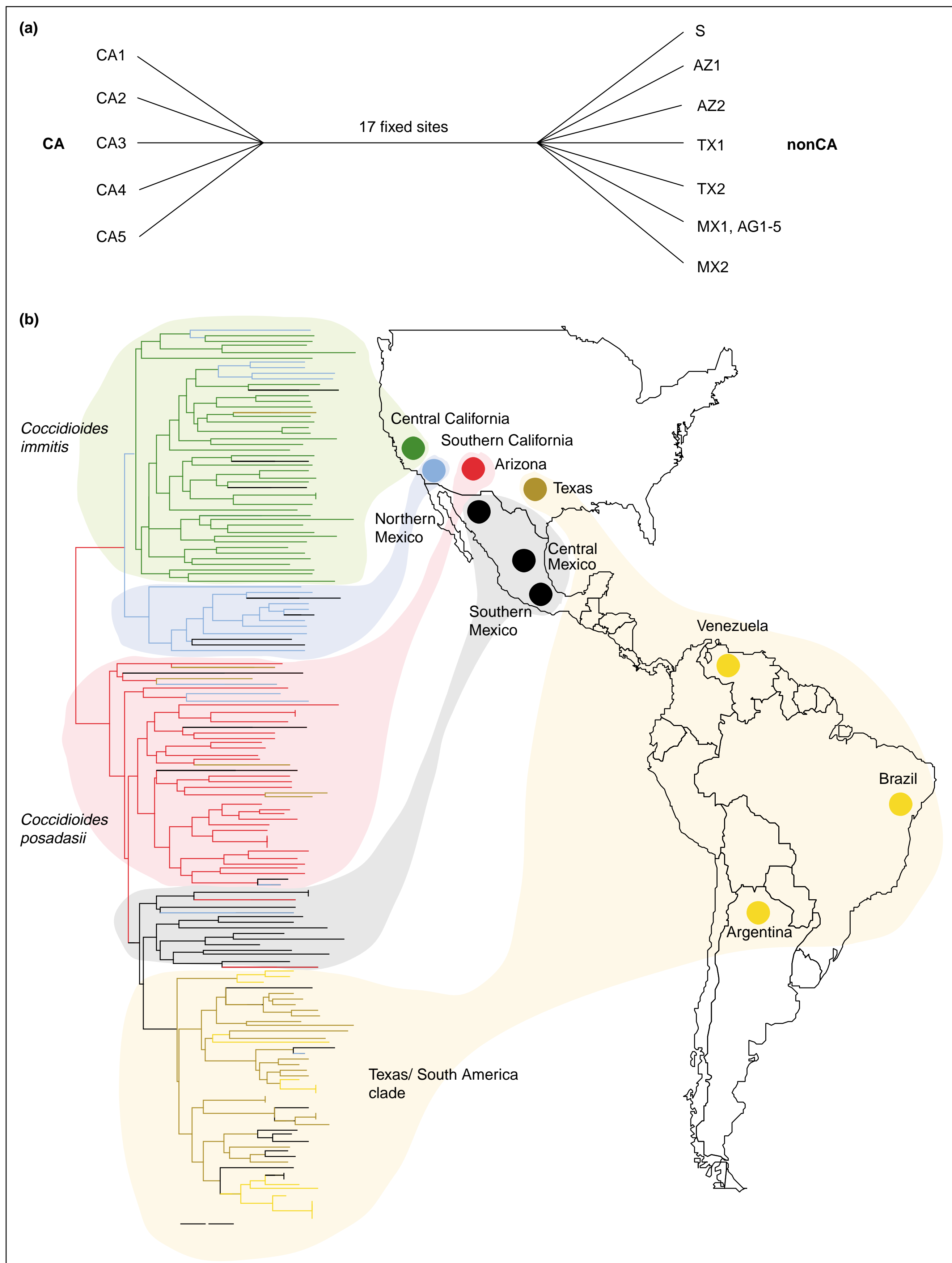
Why is multilocus sequence typing useful?

Species recognition

The first use of MLST in fungi was for species recognition. The traditional methods of fungal species recognition, principally by phenotype or, where possible, by mating tests, are being superceded by phylogenetic methods using nucleotide sequence of multiple gene genealogies. Recognizing species as clades of individuals that are genetically isolated in nature has uncovered new, cryptic species in *Candida albicans* [17], *Coccidioides immitis* [2], *Histoplasma capsulatum* [18] and *Cryptococcus neoformans* [19], as well as many plant pathogenic [11] and toxigenic fungi [20]. Some taxa formerly recognized by phenotype have been shown to be wrong. For example, the plant pathogen *Fusarium oxysporum* f. sp. *cubense* is a mix of four clades, the result of independent jumps to the diagnostic host, banana [21]. Similarly, the horse and donkey pathogen *Histoplasma capsulatum* var. *farciminosum* embraces members of three *Histoplasma* clades, each the result of host jumps to the diagnostic equid hosts (T Kasuga, personal communication). Some of the newly recognized species have important phenotypic differences, such as virulence. One example is *H. capsulatum*, where the North American class 2 clade causes disease in otherwise healthy individuals, but the North American class 1 clade causes disease only in humans that are not immuno-incompetent [18]. Another example is *C. neoformans*, where *C. neoformans* group A is far more common clinically than is *C. neoformans* group D [19].

The most difficult part of thorough species recognition is sampling the fungus throughout its range and sampling both clinical and environmental individuals. Clinical

Figure 1



isolates are often readily available, but finding environmental individuals of pathogenic fungi can be very difficult [22]. Sampling can be particularly important when a fungal population has recently expanded, often because of increased host-availability as seen in agriculture or emerging human disease. The source population might be more variable than populations found in crops or immunocompetent humans; if the source is missed, the interpretation may be incorrect. By comparison, characterizing nucleotide variation is straightforward. Where fungi have been thoroughly sampled throughout their range, phylogenetic species recognition has laid the foundation for studies of reproductive mode, pathogen migration, the cause of epidemics and precise epidemiology. A good example of what is possible is provided by studies of the species historically responsible for coccidioidomycosis, *Coccidioides immitis*.

An MLST scheme for *Coccidioides immitis* was first developed by Koufopanou *et al.* [2,3] using 2 384 nucleotide sites from five genes (*CHS1*, *pyrG*, *tcrP*, serine proteinase and *CTS2*) and 17 clinical isolates. Genealogically, each of the five loci was best described by a single most parsimonious tree, and when the five loci were considered together, the isolates were divided into two strongly supported groups. These groups were estimated to have been genetically isolated from one another by 11–12.8 million years (Figure 1a) and have largely non-overlapping geographical distributions, suggesting that their genetic isolation had a biogeographic origin [2,3]. Within each group, the gene genealogies were in conflict, implying that *Coccidioides immitis* comprised two recombining species, but without enough data for a significant result. Using collections of arbitary SNPs and more individuals, Burt *et al.* [8] and Fisher *et al.* [23] confirmed that recombination was occurring within, but not between, each group. Here, the MLST scheme had defined two phylogenetic species by using genealogical concordance phylogenetic species recognition (GCPSR) [24]. A wider sampling of clinical and some environmental isolates from the entire New World distribution of *C. immitis* showed that no further cryptic species existed and a new species was described as a result, *Coccidioides posadasii*, formerly known as non-California *Coccidioides immitis* [25]. Once MLST has identified the cryptic species harbored within morphological species, it then becomes possible to investigate the contribution of genetic variation to phenotype. For instance, although comparisons between *C. immitis* and *C. posadasii* showed no variation in the size of arthroconidia, significant phenotypic variation was seen in averaged growth rates for members of the two species [25]. This finding raises the possibility that further interspecific comparisons might succeed in explaining some of the variation observed in pathogenicity and the symptoms of coccidioidomycosis. Analysis of close relatives of *Coccidioides* species also shows 'cryptic species' recognized by genetic isolation; apparently divergence to form new species is common, but persistence of both newly diverged species is rare [26].

Populations within species

MLST has great potential as a tool to interpret the fine-scale population genetic structure of microbes. In the bacterial gastric pathogen *Helicobacter pylori*, MLST was used to type 370 isolates from 27 geographical and ethnic human groupings [27]. *H. pylori* undergoes frequent genetic recombination, so its genome is a mosaic of small fragments derived from several ancestral populations. However, despite this powerful homogenizing force, significant population genetic structure exists. Assessing the linkage between individual nucleotides and using this to assign isolates to five ancestral populations using the programme STRUCTURE [28], the authors were able to show that patterns of human migration, stemming from an out-of-African origin, explain the current day distribution of *H. pylori*. Such accurate assignment of isolates is, in part, made possible by the high levels of polymorphism found in this bacterium, (i.e. 36% of the sequenced nucleotide sites were polymorphic).

However, the power of MLST runs into problems when the species, or populations, being typed have insufficient genetic variation to differentiate isolates. Such reduced levels of genetic diversity occur as a consequence of evolutionary processes such as recent speciation, population bottlenecks or selective sweeps. Lack of variation has been recognized as a problem in, for instance, *Mycobacterium tuberculosis* where only 1/10 000 nucleotides are polymorphic [29], making it hard to achieve sufficient discriminatory power to distinguish between isolates. In *Coccidioides*, MLST resolves *C. immitis* and *C. posadasii* with ease, but certain populations within the species lack any variation at the MLST loci. For instance, within South America, eleven out of eleven *C. posadasii* isolates were identical at the MLST loci used, an identity that corresponds to an index of diversity of 0% [30]. When

(Figure 1 Legend) Recognizing *Coccidioides* species and populations using DNA polymorphisms **(a)** MLST analysis of 17 clinical *Coccidioides* isolates identifies a single branch that subdivides the population into two reproductively isolated groups [2,3]. These groups are estimated to have been genetically isolated from one another by 11–12.8 million years and have largely non-overlapping geographical distributions (Californian and non-Californian), suggesting that their genetic isolation had a biogeographic origin [9••]. **(b)** Applying a nine-locus MLMT scheme confirms the subdivision across the entire New World range of *Coccidioides* and further identifies eight phylogeographic populations (of which five are shown here for clarity). South American isolates group with Texan isolates, suggesting that a relatively recent (ca. 9 000–134 000 years) long-distance dispersal of isolates founded the Latin American populations. Figure 1a modified from [2,3], Figure 1b modified from [9••] (Copyright © 1997, 1998, 2001, National Academy of Sciences U.S.A).

developing an MLST system for fungal species, it is important to use pilot studies to determine whether levels of nucleotide variation are sufficient to answer the questions that are to be asked; if they are not, then more variable loci can be sought.

Microsatellite-based typing schemes

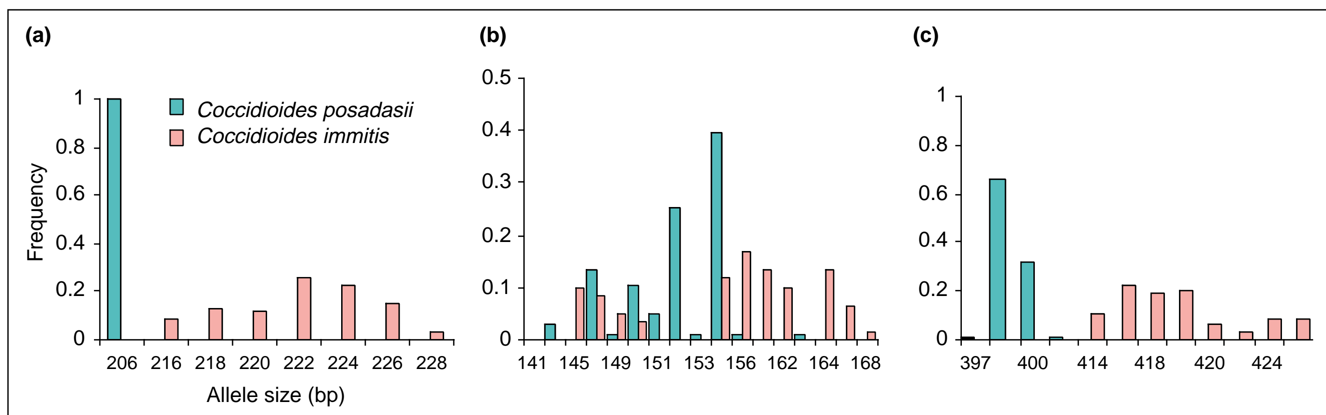
More variable microsatellite loci were sought for *Coccidioides* species, and when the 11 South American *C. posadasii* isolates were re-genotyped using a multilocus microsatellite typing (MLMT) system, eight out of eleven multilocus genotypes were unique, corresponding to an index of diversity of 94.5%. Clearly, the higher mutation rates operating at the microsatellite loci (estimated in yeast to be 10^{-4}–10^{-5} mutations per generation as compared to 10^{-9} for point mutations [31]) allow variability to accumulate at a far higher rate, leading to greater intra-population genetic diversity. In cases where studies need to differentiate between closely related isolates, then microsatellite typing schemes can be highly informative. Typing with microsatellites is similar to MLST in that an entire stretch of sequence is surveyed for genetic variation, albeit for length polymorphisms rather than point mutations, and multilocus genotypes are generated and then added to a database. However, the hypervariability of microsatellite loci and their stepwise mode of mutation, creates alleles that might be identical by size, but not identical by descent, (i.e. loci might be homoplaseous). Also, microsatellites found to be polymorphic in one genetically isolated clade might be monomorphic in others (Figure 2). Studies using microsatellites have shown that both the problems of homoplasy and unexpected monomorphism can be overcome by using many microsatellite loci [32].

Using MLMT to type the entire *Coccidioides* collection of 167 isolates supported the original MLST division into two species (Figure 1b). However, the greater levels of genetic variation at these microsatellite-containing loci enabled the two species to be further split into five populations. Within *C. posadasii*, the Latin American population had reduced allelic variation and increased association amongst loci as compared to North American populations. In addition, the significant relationship between geographic and genetic distance seen in North American *C. posadasii* was lost when Latin American genotypes were included in the analysis. Both of these observations support the hypothesis that long distance dispersal of the pathogen occurred into South America between 8 940 and 134 000 years ago, possibly aided by human migration [9••].

Emerging fungal disease and multilocus sequence typing

MLST can be used to help discriminate among the several possible causes of an emerging disease. For example, the fungus may have been in the area all along, and is now causing disease because the host has become susceptible or the pathogen has become more prevalent. In this case, MLST of fungal individuals causing disease should be no different than individuals obtained from the environment, and all fungal individuals in the area of disease outbreak should show population structure similar to that found in other geographic locations. It might also be that the fungus has been in the area all along, but that a new genotype with increased virulence is responsible for an epidemic. In this case, the genotypes of individuals causing disease should show reduced variation compared to environmental isolates and also show association among loci due to the selective sweep. An epidemic of

Figure 2



Microsatellites in *Coccidioides* species: allele distributions at three polymorphic loci. Because of their stepwise mode of mutation, microsatellite loci may suffer from ascertainment bias **(a)** where a locus may be monomorphic within one species. In addition, alleles in separate species might be identical by size, but not by descent (homoplaseous) **(b)**. However, other loci may portray the species designations accurately, as well as being polymorphic within both species **(c)**. Studies using microsatellites have shown that both the problems of homoplasy and unexpected monomorphism can be overcome by using many microsatellite loci. When initially surveying genetic variation for a given species, it is preferable to use both MLST and MLMT because DNA sequence and microsatellites evolve at quite different rates. Figure 2 modified from [25] (Copyright © 2002, Mycologia).

coccidioidomycosis in California in the early 1990s raised suspicions of the emergence of a virulent *C. immitis* genotype. An MLST scheme that specifically targeted 14 SNPs of individual isolates collected during the epidemic showed no significant association of loci and no evidence of clonal spread [23]. Instead of the emergence of a virulent genotype, multivariate statistical analyses suggested that climatic factors could explain this, and past, epidemics of coccidioidomycosis [23]. By contrast, most isolates responsible for the cutaneous histoplasmosis of horses and donkeys caused by *H. capsulatum* var. *farciminosum*, mentioned above, have been shown by MLST to be due to a clonal genotype of *H. capsulatum*. This genotype emerged from a recombining population of South American *H. capsulatum* and has probably spread throughout the Old World [18]. A recent MLST study of the chytrid fungus thought to be responsible for the world-wide decline of amphibians found very low variation in coding regions and suggested that the decline is due to the recent spread of the fungus [5]. Given the low variation, this conclusion seems justified. However, confirmation of low genetic variation with more variable microsatellites (MLMT), or finding of the population from which the fungus spread, would strengthen the case.

Multilocus microsatellite typing and epidemiology

Emergence of coccidioidomycosis in locations outside the endemic areas was observed in the 1990s, likely due to travel to endemic areas that were experiencing increased infection rates. Chaturvedi *et al.* [33] showed that 181 hospitalizations in New York state were due to coccidioidomycosis and multilocus genotyping of five SNP-containing loci demonstrated that 14 of 16 isolates tested were *C. posadasii* and most likely from Arizona. These isolates were re-genotyped using the *Coccidioides* MLMT scheme and the source population for each population was determined by comparing individual MLMT genotypes to the New World database of 167 MLMT genotypes using bayesian assignment tests (BATs) performed by the programme BAYESASS [34]. The increased resolution made possible by using highly polymorphic genetic markers, well-characterized populations and newly developed statistical tools enabled 12 out of 16 isolates to be assigned to source populations with high probability, and all 16 isolates to be assigned to species. With the new method, fewer individuals appear to have had an Arizona origin. Typing fungi considered to be select agents to population of origin may be of more than academic interest, given that forensic epidemiologists are now attempting to type *Bacillus anthracis* to specific laboratories.

Conclusions and future studies

Fungal MLST and MLMT are proving valuable to studies of human and plant pathogenic fungi and the diseases that they cause. These studies have shown that fungal species and populations are more numerous than estimated from phenotypic data. An important inference is that although fungi can disperse over great distances, in many cases they are constrained to far smaller geographic areas, presumably by local adaptation. Most existing work has been in the developed world. Studies are needed of fungi causing disease in developing areas, and of fungi that have not caused, but could cause, disease in developed areas. To complement the studies of fungi causing disease, which are likely to have been affected by human activity, similar studies are needed of nonpathogenic fungi in natural settings.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou JJ, Zurth K, Caugant DA *et al.*: **Multilocus sequence typing — a portable approach to the identification of clones within populations of pathogenic microorganisms**. *Proc Natl Acad Sci USA* 1998, **95**:3140-3145.

2. Koufopanou V, Burt A, Taylor JW: **Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis***. *Proc Natl Acad Sci USA* 1997, **94**:5478-5482.

3. Koufopanou V, Burt A, Taylor JW: **Correction: Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis***. *Proc Natl Acad Sci USA* 1998, **95**:8414.

4. Carbone I, Kohn LM: **A microbial population–species interface: nested cladistic and coalescent inference with multilocus data**. *Mol Ecol* 2001, **10**:947-964.

5. Morehouse EA, James TY, Ganley ARD, Vilgalys R, Berger L, Murphy PJ, Longcore JE: **Multilocus sequence typing suggests the chytrid pathogen of amphibians is a recently emerged clone**. *Mol Ecol* 2003, **12**:395-403.

6. •• Anderson JB, Wickens C, Khan M, Cowen LE, Federspiel N, Jones T, Kohn LM: **Infrequent genetic exchange and recombination in the mitochondrial genome of *Candida albicans***. *J Bacteriol* 2001, **183**:865-872.

The authors combined analysis of nuclear and mitochondrial DNA variation and proposed that evidence for recombination in natural individuals of this morphologically asexual fungus reflects ancient matings and that recent reproduction is exclusively clonal.

7. Bougnoux ME, Morand S, d'Enfert C: **Usefulness of multilocus sequence typing for characterization of clinical isolates of *Candida albicans***. *J Clin Microbiol* 2002, **40**:1290-1297.

8. Burt A, Carter DA, Koenig GL, White TJ, Taylor JW: **Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis***. *Proc Natl Acad Sci USA* 1996, **93**:770-773.

9. •• Fisher MC, Koenig GL, White TJ, San-Blas G, Negroni R, Alvarez IG, Wanke B, Taylor JW: **Biogeographic range expansion into South America by *Coccidioides immitis* mirrors New World patterns of human migration**. *Proc Natl Acad Sci USA* 2001, **98**:4558-4562.

The authors used hypervariable markers, microsatellites and a collection of 167 individuals from throughout the range of the fungus. The microsatellites recognize populations that correlate with geographic distance, showing that fungal populations may occupy ranges smaller than might be guessed from their potential for dispersal. The data also point to a recent migration of the fungus into Latin America.

10. Carbone I, Anderson JB, Kohn LM: **Patterns of descent in clonal lineages and their multilocus fingerprints are resolved with combined gene genealogies**. *Evolution* 1999, **53**:11-21.

11. O'Donnell K: **Molecular phylogeny of the *Nectria haematococca-Fusarium solani* species complex**. *Mycologia* 2000, **92**:919-938.

12. Steenkamp ET, Wingfield BD, Desjardins AE, Marasas WFO, Wingfield MJ: **Cryptic speciation in *Fusarium subglutinans***. *Mycologia* 2002, **94**:1032-1043.
13. Xu JP, Vilgalys R, Mitchell TG: **Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans***. *Mol Ecol* 2000, **9**:1471-1481.
14. Geiser DM, Pitt JI, Taylor JW: **Cryptic speciation and recombination in the aflatoxin producing fungus *Aspergillus flavus***. *Proc Natl Acad Sci USA* 1998, **95**:388-393.
15. Taylor JW, Geiser DM, Burt A, Koufopanou V: **The evolutionary biology and population genetics underlying fungal strain typing**. *Clin Microbiol Rev* 1999, **12**:126-146.
16. Burt A, Dechairo BM, Koenig GL, Carter DA, White TJ, Taylor JW: **Molecular markers reveal differentiation among isolates of *Coccidioides immitis* from California, Arizona and Texas**. *Mol Ecol* 1997, **6**:781-786.
17. Sullivan DJ, Westerneng TJ, Haynes KA, Bennett DE, Coleman DC: ***Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals**. *Microbiology* 1995, **141**:1507-1521.
18. Kasuga T, Taylor JW, White TJ: **Phylogenetic relationships of varieties and geographical groups of the human pathogenic fungus *Histoplasma capsulatum* Darling**. *J Clin Microbiol* 1999, **37**:653-663.
19. Franzot SP, Salkin IF, Casadevall A: ***Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates**. *J Clin Microbiol* 1999, **37**:838-840.
20. Cruse M, Telerant R, Gallagher T, Lee T, Taylor JW: **Cryptic species in *Stachybotrys chartarum***. *Mycologia* 2002, **94**:814-822.
21. O'Donnell K, Kistler HC, Cigelink E, Ploetz RC: **Multiple evolutionary origins of the fungus causing Panama disease of banana: Concordant evidence from nuclear and mitochondrial gene genealogies**. *Proc Natl Acad Sci USA* 1998, **95**:2044-2049.
22. Greene DR, Koenig G, Fisher MC, Taylor JW: **Soil isolation and molecular identification of *Coccidioides immmitis***. *Mycologia* 2000, **92**:406-410.
23. Fisher MC, Koenig GL, White TJ, Taylor JW: **Pathogenic clones versus environmentally driven population increase: Analysis of an epidemic of the human fungal pathogen *Coccidioides immitis***. *J Clin Microbiol* 2000, **38**:807-813.
24. Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC: **Phylogenetic species recognition and species concepts in fungi**. *Fungal Genet Biol* 2000, **31**:21-32.
25. Fisher MC, Koenig GL, White TJ, Taylor JT: **Molecular and phenotypic description of *Coccidioides posadasii* sp nov., previously recognized as the non-California population of *Coccidioides immitis***. *Mycologia* 2002, **94**:73-84.
26. Koufopanou V, Burt A, Szaro T, Taylor John W: **Gene genealogies, cryptic species, and molecular evolution in the human pathogen *Coccidioides immitis* and relatives (Ascomycota, Onygenales)**. *Mol Biol Evol* 2001, **18**:1246-1258.
27. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI *et al.*: **Traces of human migrations in *Helicobacter pylori* populations**. *Science* 2003, **299**:1582-1585.
28. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data**. *Genetics* 2000, **155**:945-959.
29. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM: **Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination**. *Proc Natl Acad Sci USA* 1997, **94**:9869-9874.
30. Hunter PR, Gaston MA: **Numeric index of the discriminatory ability of typing systems: an application of Simpson's index of diversity**. *J Clin Microbiol* 1988, **26**:2465-2466.
31. Henderson ST, Petes TD: **Instability of simple sequence DNA in *Saccharomyces cerevisiae***. *Mol Cell Biol* 1992, **12**:2749-2757.
32. Fisher MC, Koenig G, White TJ, Taylor JW: **A test for concordance between the multilocus genealogies of genes and microsatellites in the pathogenic fungus *Coccidioides immitis***. *Mol Biol Evol* 2000, **17**:1164-1174.
33. Chaturvedi V, Ramani R, Gromadzki S, Rodeghier B, Chang HG, Morse DL: **Coccidioidomycosis in New York State**. *Emerg Infect Dis* 2000, **6**:25-29.
34. Fisher MC, Rannala B, Chaturvedi V, Taylor JW: **Disease surveillance in recombining pathogens: Multilocus genotypes identify sources of human *Coccidioides* infections**. *Proc Natl Acad Sci USA* 2002, **99**:9067-9071.