PNAS RESEARCH ARTICLE | EVOLUTION

A global phylogenomic analysis of the shiitake genus *Lentinula*

Sean Sierra-Patev[a,1], Byoungnam Min[b,1], Miguel Naranjo-Ortiz[a,1], Brian Looney[a,1], Zachary Konkel[c], Jason C. Slot[c], Yuichi Sakamoto[d], Jacob L. Steenwyk[e], Antonis Rokas[e], Juan Carro[f], Susana Camarero[f], Patricia Ferreira[g,h], Gonzalo Molpeceres[f], Francisco J. Ruiz-Dueñas[f], Ana Serrano[f], Bernard Henrissat[i,j], Elodie Drula[k,l], Karen W. Hughes[m], Juan L. Mata[n], Noemia Kazue Ishikawa[o], Ruby Vargas-Isla[o], Shuji Ushijima[p], Chris A. Smith[q], John Donoghue[r], Steven Ahrendt[b], William Andreopoulos[b], Guifen He[b], Kurt LaButti[b], Anna Lipzen[b], Vivian Ng[b], Robert Riley[b], Laura Sandor[b], Kerrie Barry[b], Angel T. Martínez[f], Yang Xiao[s], John G. Gibbons[t], Kazuhisa Terashima[p], Igor V. Grigoriev[b,u], and David Hibbett[a,1,2]

Edited by David Hillis, The University of Texas at Austin, Austin, TX; received August 20, 2022; accepted December 22, 2022

Lentinula is a broadly distributed group of fungi that contains the cultivated shiitake mushroom, _L. edodes_. We sequenced 24 genomes representing eight described species and several unnamed lineages of _Lentinula_ from 15 countries on four continents. _Lentinula_ comprises four major clades that arose in the Oligocene, three in the Americas and one in Asia–Australasia. To expand sampling of shiitake mushrooms, we assembled 60 genomes of _L. edodes_ from China that were previously published as raw Illumina reads and added them to our dataset. _Lentinula edodes_ sensu lato (s. lat.) contains three lineages that may warrant recognition as species, one including a single isolate from Nepal that is the sister group to the rest of _L. edodes_ s. lat., a second with 20 cultivars and 12 wild isolates from China, Japan, Korea, and the Russian Far East, and a third with 28 wild isolates from China, Thailand, and Vietnam. Two additional lineages in China have arisen by hybridization among the second and third groups. Genes encoding cysteine sulfoxide lyase (_lecsl_) and γ-glutamyl transpeptidase (_leggt_), which are implicated in biosynthesis of the organosulfur flavor compound lenthionine, have diversified in _Lentinula_. Paralogs of both genes that are unique to _Lentinula_ (_lecsl_ 3 and _leggt_ 5b) are coordinately up-regulated in fruiting bodies of _L. edodes_. The pangenome of _L. edodes_ s. lat. contains 20,308 groups of orthologous genes, but only 6,438 orthogroups (32%) are shared among all strains, whereas 3,444 orthogroups (17%) are found only in wild populations, which should be targeted for conservation.

fungi | domestication | mushrooms | evolution | population genomics

Significance

Lentinula is an economically important group of fungi that includes the cultivated shiitake mushroom. We performed a comparative analysis of shiitake and related species, including 24 new genomes from Asia, Australasia, and the Americas. *Lentinula* is roughly 28 My old and includes four major groups, three in the Americas and one in Asia–Australasia. *Lentinula edodes* (shiitake) comprises three independent lineages that may warrant recognition as species. One lineage of *L. edodes* is represented by a single Nepalese isolate, while the others are broadly distributed across East and Southeast Asia and show evidence of hybridization. Genes that encode enzymes responsible for biosynthesis of sulfur-containing compounds have expanded in *Lentinula* and may contribute to the distinctive flavor of shiitake mushrooms.

Lentinula edodes, also known as shiitake, xiang-gu, nng, or p'yogo (Japanese, Chinese, Vietnamese, and Korean, respectively), is one of the most widely cultivated mushrooms in the world. Shiitake is native to Asia, but *Lentinula* (Agaricales, Omphalotaceae) is distributed on all continents except Europe and Antarctica, as well as New Zealand, Tasmania, Papua New Guinea (PNG), Borneo, and the Caribbean (1–5). *Lentinula* species are white rot wood decayers, with most occurring on oaks and relatives (Fagales). All species of *Lentinula* are edible, but some people have an allergic reaction to the carbohydrate lentinan in uncooked mushrooms, termed "shiitake dermatitis" (6). A distinctive culinary property of the group is a garlicky or radish-like flavor that is due to cyclic organosulfur compounds, such as lenthionine (7, 8). A large body of literature explores the potential health benefits of *L. edodes*, which is a component of Chinese traditional medicine and contemporary "nutraceuticals" (9).

Taxonomy of *Lentinula* is in flux (*SI Appendix*, Table S1). A recent phylogenetic analysis using genes encoding internal transcribed spacers (ITS) of ribosomal RNA (rRNA), large subunit rRNA, and translation elongation factor 1-α (*tef*1-α) resolved 15 terminal groups that correspond to ten described (or provisional) species and five unnamed lineages (10). Trees based on rRNA and *tef*1-α genes were largely congruent, with one major difference: ITS resolved two nonsister lineages within the geographic range of *L. edodes*, which were previously called Group 1 and Group 5 (11), but *tef*1-α resolved only one lineage. Most populations of *Lentinula* in Asia–Australasia are mating compatible, which has led some authors to suggest that they all represent the same species, namely *L. edodes* (12).

Several genomes of *L. edodes* have been published from Chinese, Japanese, and South Korean strains (13–17), as well as one genome of *L. novae-zelandiae* from New Zealand (18). Population genomic analyses of *L. edodes* in China have been performed by Xiao et al. (19) and Zhang et al. (17). Population structure analysis using single-nucleotide polymorphisms (SNPs) in both studies suggested the existence of three populations in China that Zhang et al. (17) termed cultivars, wild 1, and wild 2. Other studies aimed at strain characterization in *L. edodes* have used analyses of mitochondrial genomes (20) and other methods based on DNA polymorphisms (21–25).

Author contributions: S.S.-P., B.M., M.N.-O., B.L., K.B., J.G.G., K.T., I.V.G., A.T.M., and D.H. designed research; S.S.-P., B.M., M.N.-O., B.L., Y.S., and J.D. performed research; K.W.H., J.L.M., N.K.I., R.V.-I., S.U., and C.A.S. contributed new reagents/analytic tools; S.S.-P., B.M., M.N.-O., B.L., Z.K., J.C.S., Y.S., J.L.S., A.R., J.C., S.C., P.F., G.M., F.J.R.-D., A.S., B.H., E.D., S.A., W.A., G.H., K.L., A.L., V.N., R.R., L.S., K.B., A.T.M., Y.X., J.G.G., and I.V.G. analyzed data; K.W.H., J.L.M., N.K.I., R.V.-I., S.U., and K.T. contributed strains; C.A.S. contributed data; and D.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]S.S.-P., B.M., M.N.-O., B.L., and D.H. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: dhibbett@clarku.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2214076120/-/DCSupplemental.

Published February 27, 2023.

Ambiguity about species limits and the lack of a robust phylogenetic framework hinder basic and applied research in *Lentinula*. We conducted a phylogenomic analysis using a geographically diverse set of *Lentinula* isolates, including 24 new genomes. To expand sampling of *L. edodes*, we assembled and annotated 60 genomes from Illumina reads published by Xiao et al. (19). We performed phylogenetic and population structure analyses, assessed evidence of gene flow between lineages, and characterized the pangenomes of *L. edodes*, *L. boryana*, and *L. raphanica*. We also analyzed genes encoding enzymes involved in decay of plant cell walls and biosynthesis of organosulfur compounds. Our results have implications for classification and conservation in *Lentinula* and provide clues to the origin of its unique flavor profile.

Results

Twenty-Four New *Lentinula* Genomes. We sequenced the genomes of 11 *Lentinula* individuals from Asia–Australasia and 13 individuals from the Americas, which correspond to 11 of the lineages that are resolved with rRNA and *tef*1-α genes (*SI Appendix*, Table S1) (10). Seven genomes were assembled using PacBio long reads, while the others were produced using Illumina. Genome completeness as estimated by BUSCO 5.3.2 (Benchmarking Universal Single-Copy Orthologs) (26) is 98.80 to 99.93% with an average of 99.51% (OrthoDB v10, Basidiomycota) (27). All but *L. aciculospora* are haploid genomes. We added four published genomes (three *L. edodes* and one *L. novae-zelandiae*) (13, 16, 18) for a core comparative dataset of 28 *Lentinula* genomes (Dataset S1).

Lentinula genomes average 40.48 Mbp (Figs. 1 and 2). *Lentinula aciculospora* JLM2183 has the largest genome (56.71 Mbp), but almost half the genome (45.97%) is composed of transposable elements. *Lentinula raphanica* has the lowest repeat content (0.58 to 10.86%) but more coding regions with thousands more genes than the other species. The sizes and repeat contents of *Lentinula* genomes are comparable to those of other basidiomycetes (28, 29). To characterize gene presence/absence variation (PAV) across *Lentinula*, we used an orthology-based approach. We considered orthogroups to be present in a strain if at least one protein-coding sequence was assigned to that orthogroup by OrthoFinder (30) without considering gene length or copy number. The core genome of *Lentinula* comprises 5,329 orthogroups that are shared among all included strains, while 9,763 orthogroups show PAV across the dataset without being restricted to any monophyletic group. Orthogroups unique to lineages are rare, except in *L. raphanica*, which has 636 lineage-exclusive orthogroups containing 809 to 868 genes per genome (*SI Appendix*, Figs. S1 and S2).

***Lentinula* Has a Conserved White Rot Decay Apparatus.** Substrate use efficiency is a key determinant of productivity in cultivation. Accordingly, we characterized genes encoding plant cell wall degrading enzymes (PCWDEs), including carbohydrate-active enzymes (CAZys) (31) and decay-related oxidoreductases, as well as cytochrome P450s (CYPs) (*SI Appendix*, Table S2, Figs. S3–S12, and Datasets S2 and S3). The numbers of CAZy genes in *L. raphanica* (average 480) and other species of *Lentinula* (average 477) were not significantly different (*P* value of the rank-sum test: 0.52). Overall, the composition of PCWDEs and CYPs is highly uniform across *Lentinula* species, reflecting their conserved ecological role as white rot saprotrophs on hardwood substrates, mainly Fagales.

Enzymes Involved in Biosynthesis of Organosulfur Compounds Have Expanded in *Lentinula*. Cysteine sulfoxide lyase (encoded by *lecsl*) and γ-glutamyl transpeptidase (encoded by *leggt*) are implicated in the biosynthesis of formaldehyde and the cyclic organosulfur compounds that give shiitake its distinctive flavor and aroma (7, 8, 13, 32–34). While a paralog of *lecsl* is reported to facilitate organosulfur production (13), there is little evidence supporting the role of a specific *leggt* orthogroup in the production of organosulfur small molecules. Most *Lentinula* genomes contain three *lecsl* orthogroups (*lecsl* 1 to 3) and five *leggt* orthogroups (*leggt* 1 to 5), totaling upward of six homologues per species (*SI Appendix*, Figs. S13 and S14 and Table S3). *Lentinula* genomes possess 2 to 4 (average 2.2) copies of *lecsl* and 5 to 7 (average 6.7) copies of *leggt*, but the other species of Omphalotaceae in our analysis possess 0 to 1 (average 0.4) *lecsl* gene and 3 to 4 (average 3.6) *leggt* genes. Phylogenetic and syntenic analyses of *lecsl* and *leggt* homologues from all available Agaricomycotina genomes suggest that gene duplications occurred within Omphalotaceae, resulting in paralogs of *lecsl* and *leggt* that are unique to *Lentinula* (*SI Appendix*, Figs S13 and S14). Independent tandem duplications of *leggt* 5 followed the divergence of *L. raphanica* from *L. boryana* and the emergence of *L. edodes*. An additional duplication resulted in a third distal *leggt* 5 paralog in *L. raphanica*. Gene coexpression analysis in a dikaryotic *L. edodes* isolate NBRC111652 indicates that *lecsl* 3 and *leggt* East 5B are the only significantly coordinately expressed *lecsl* and *leggt* copies from this isolate (*SI Appendix*, Fig. S15). These genes are up-regulated in fruiting bodies, suggesting that they may be responsible for biosynthesis of lenthionine and formaldehyde.

***Lentinula* Contains Four Major Groups That Arose in the Paleogene.** The most recent common ancestor (MRCA) of *Lentinula* lived approximately 28.3 (27.5 to 30.4) Mya (Fig. 2

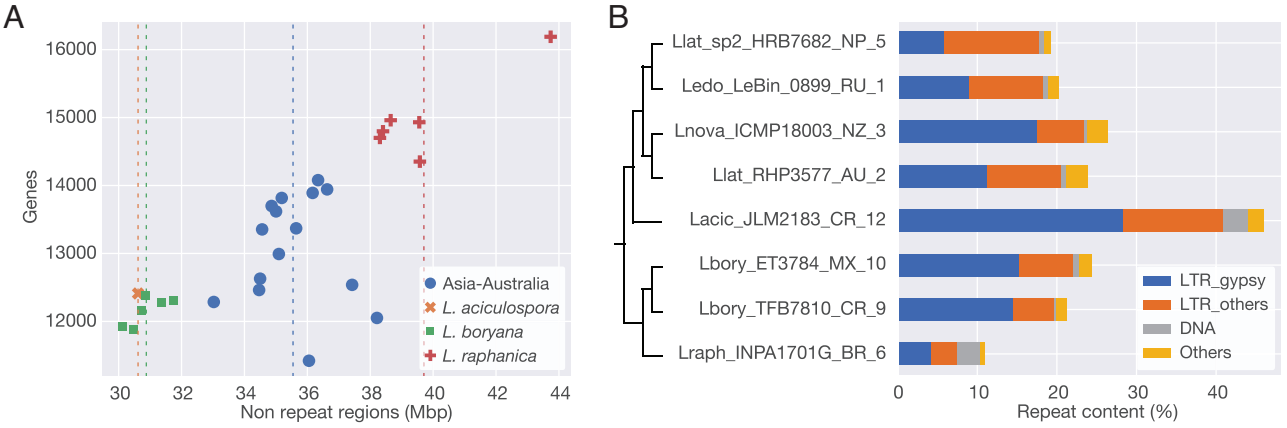

Fig. 1. Genome characteristics of *Lentinula*. (*A*) Nonrepeat regions and number of genes. The vertical lines indicate the average sizes of the groups. (*B*) Repeat contents of the eight PacBio assemblies. Repeat regions were estimated using the JGI annotation pipeline. LTR: long terminal repeats.

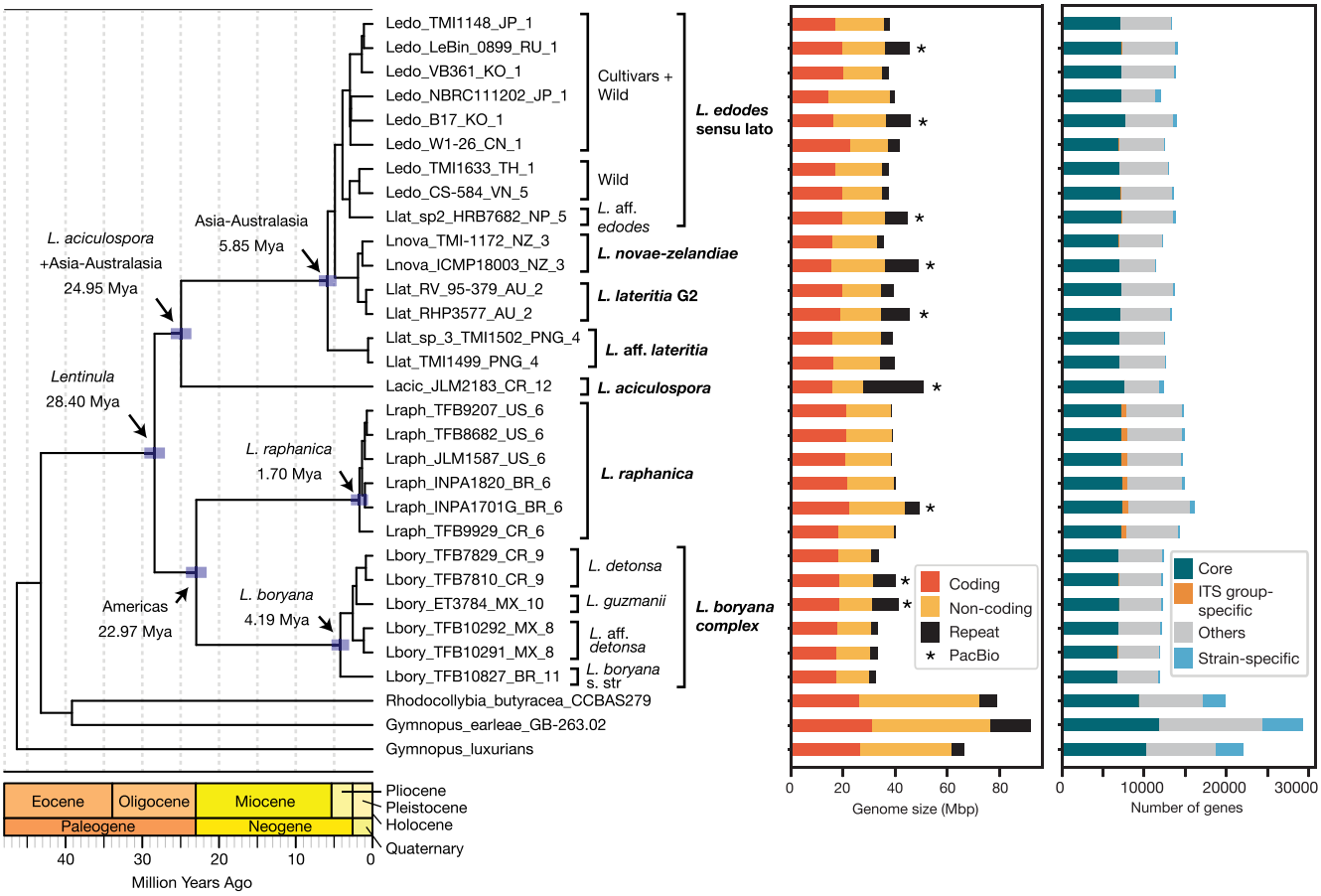

Fig. 2. Phylogenetic relationships, clade ages, genome sizes and contents, and classification of *Lentinula*. Chronogram was produced using BEAST. Node bars indicate 95% h.p.d. of age estimates. All branches receive 100% bootstrap support unless otherwise indicated. KO = Korea, JP = Japan, CN = China, RU = Russia, TH = Thailand, VN = Vietnam, NP = Nepal, AU = Australia, NZ = New Zealand, PNG = Papua New Guinea, CR = Costa Rica, US = United States, BR = Brazil, and MX = Mexico. Numbers following isolate and country codes indicate ITS groups. Core genes are those conserved in all genomes, while ITS-group specific genes include the genes conserved in all members of a group but lacking in others.

and *SI Appendix*, Figs. S16–S18 and Dataset S4). Four major groups of *Lentinula* diverged in the Oligocene. Three occur in tropical or subtropical regions of the Americas and correspond to *L. raphanica*, *L. aciculospora*, and the *L. boryana* species complex (a group of four lineages, including *L. boryana, L. guzmanii, L. detonsa*, and *L.* aff. *detonsa*) (10). The fourth major group occurs in Asia–Australasia. The MRCA of the Asian–Australasian group lived about 6.0 (5.7 to 6.4) Mya, which is slightly older than the MRCA of the *L. boryana* complex (4.2 [3.9 to 4.5] Mya). Four lineages are resolved in the Asian–Australasian clade, which we have labeled *L. edodes* sensu lato (s. lat.; broadly distributed in temperate to tropical Asia), *L. novae-zelandiae* (New Zealand), *L. lateritia* (PNG and Australia), and *L.* aff. *lateritia* (PNG), which is the sister group of the remaining Asian–Australasian lineages. *Lentinula aciculospora*, from Costa Rica, is the sister group of the Asian–Australasian clade.

Six alternative phylogenetic topologies with varying placements of *L. aciculospora* and *L.* aff. *lateritia* were rejected using topology constraint tests implemented in IQ-TREE V.1.6.1 (35) (*SI Appendix*, Fig. S19 and Table S4). Moreover, the same tree topology was recovered in analyses using six different subsampled versions of the dataset, which were partitioned according to alignment length, DVMC, median long-branch score, percent parsimony informative sites, saturation, and treeness divided by RCV. Examination of long-branch scores per taxon (36) suggests that long-branch attraction is not a likely source of error (*SI Appendix*, Fig. S20).

***Lentinula* Genomes Represent up to 13 Species.** Bayesian analyses under the multispecies coalescent (MSC) model implemented in STACEY (37) resolved nine to thirteen species clusters in the 28 reference *Lentinula* genomes, with the exact value depending on collapse height (ε) settings (38). At ε = 0.0001, 13 species were resolved corresponding to *L. lateritia*, *L.* aff. *lateritia*, *L. novae-zelandiae*, *L. aciculospora*, four species in the *L. boryana* complex, two species in *L. raphanica*, and three species in *L. edodes* s. lat. (Fig. 3). The three species clusters resolved in *L. edodes* s. lat. include one group in Northeast Asia (six individuals from Japan, China, Korea, and Russia), a second from Southeast Asia (two individuals in Vietnam and Thailand), and a third consisting of a single individual from Nepal (HRB 7682). At ε = 0.0002, *L. raphanica* collapsed to a single species cluster, and *L. lateritia* and *L. novae-zelandiae* were lumped into a single species (yielding eleven species). At ε = 0.0003, *L. edodes* s. lat. also collapsed to a single species cluster (yielding nine species).

Cultivated Strains Are Concentrated in One Lineage of *L. edodes* s. lat. To expand sampling of *L. edodes* s. lat., we assembled and annotated 60 genomes (21 cultivated and 39 wild) from Illumina reads published by Xiao et al. (19) (*SI Appendix*, Fig. S21). Concatenated phylogenetic analyses (39) and species tree analyses (40) yielded the same major topology as that obtained with the core dataset (*SI Appendix*, Figs. S22 and 23). Within *L. edodes* s. lat., three lineages were resolved, including 1) a "cultivars + wild" (CW) group

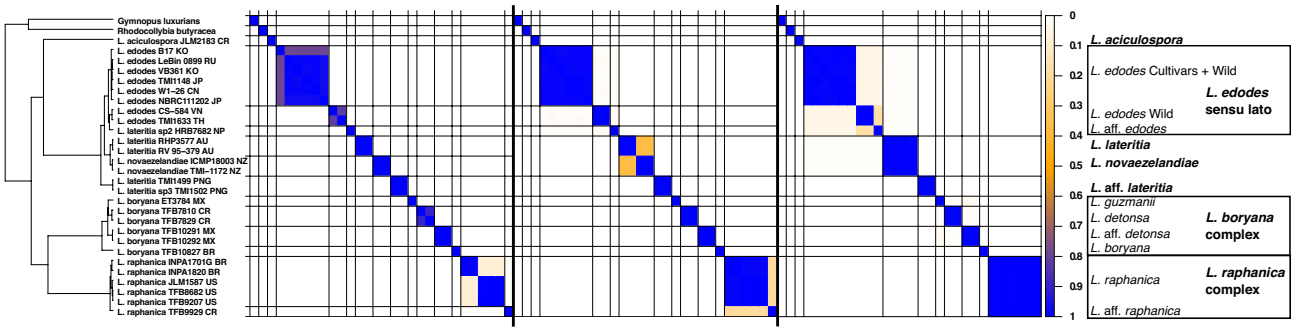

Fig. 3. Species or minimal cluster trees inferred with STACEY with matrices showing posterior probability densities for species clusters obtained at collapse height (ε) settings of 0.0001 (*Left*), 0.0002 (*Middle*), and 0.0003 (*Right*).

containing 20 cultivated isolates and six wild strains from China sampled by Xiao et al. plus the six northeast Asian strains sequenced in this study, 2) a "wild" (W) group containing 26 strains from Xiao et al. (19) plus the two southeast Asian strains reported here, and 3) the Nepalese isolate HRB7682, which was again placed as the sister group of the rest of *L. edodes* s. lat. The CW, W, and Nepalese groups in *L. edodes* s. lat. correspond to the three species resolved with the 28 genomes of the core comparative dataset under the MSC model. In addition to these, a paraphyletic "mixed" (M) assemblage of genomes was resolved, comprising seven wild isolates and one cultivar (ZP85) from the study of Xiao et al. (*SI Appendix*, Figs. S22 and S23). The CW, W, and M groups have broadly overlapping geographic ranges in continental Asia and correspond to the cultivars, wild 1, and wild 2 groups, respectively, delimited by Zhang et al. (17).

***Lentinula edodes* s. lat. Contains At Least Two Admixed Populations.** To assess population structure in *L. edodes* s. lat., we constructed two SNP datasets, one containing 3×10^6 SNPs scored from genome assemblies and the other with 8,777 SNPs scored from Illumina reads (Fig. 4 and *SI Appendix*, Fig. S24). SplitsTree (41) analyses divided the W and CW groups into two widely separated clusters, with M1 and M2 intermediates, which is consistent with gene-based phylogenies (*SI Appendix*, Figs. S22 and S23). In Admixture analyses at $K = 2$ (the optimal value for the number of subpopulations) or $K = 3$, all eight individuals from the M group contain a mixture of SNPs assigned to the CW and W groups, but at $K = 4$, the six isolates of M2 are resolved as a unique population, while the two isolates of M1 (YS84 and YS88) are still resolved as mixtures of the CW and W groups (Fig. 4 and *SI Appendix*, Fig. S24). At $K = 3$ or $K = 4$, the W group is divided into two subpopulations, but the CW group remains homogeneous. The Nepalese isolate HRB7682 is resolved as a member of the W group in Admixture analyses, but a principal component analysis (PCA) of SNP data separates the Nepalese isolate from all other isolates of *L. edodes* s. lat. (Fig. 4).

SNP analyses suggest that M1 and M2 arose via independent hybridization events between the CW and W groups. To reconstruct phylogenetic scenarios that include reticulation, we performed Species Networks applying Quartets (SNaQ) analysis (42) using 196 gene trees from 88 *Lentinula* genomes with two outgroup Omphalotaceae genomes. When M1 and M2 were constrained to be hybrids, the best model supported two introgression events from the CW group into the M group, with a higher degree of introgression into M1 (γ = 48) than M2 (γ = 29) (Fig. 5), which is consistent with the population structure resolved with Admixture. In an unconstrained analysis, where the M groups were not designated as hybrids, the best model suggested introgression between the CW group of *L. edodes* s. lat. and the ancestor, the Asian–Australasian clade, but there is no support for this scenario in other analyses.

ITS Is Polymorphic in *L. edodes* s. lat. Phylogenetic analyses of 51 ITS sequences extracted from genomes or downloaded from GenBank resolved 16 terminal groups in *Lentinula*, including the 15 groups identified by Menolli et al. (10) and the Vietnamese *L. platinedodes* (2) (*SI Appendix*, Fig. S25). Our genome dataset includes exemplars of eleven of the ITS-based lineages (*SI Appendix*, Table S1). Nine of the ITS-based lineages are congruent with genome-based lineages. However, two ITS types that have been called Group 1 and Group 5 (10, 11) are scattered throughout *L. edodes* s. lat. The Group 5 ITS type is present in ten of the 27 genomes in the W group and the Nepalese isolate HRB7682, but all other genomes of *L. edodes* s. lat. have the Group 1 ITS type (Fig. 4 and *SI Appendix*, Fig. S22–S24). Group 1 and Group 5 are nonsister lineages in ITS phylogenies (*SI Appendix*, Fig. S25). Constrained topologies that forced monophyly of Group 1 and Group 5 were rejected using IQ-TREE (*SI Appendix*, Table S5).

The method we used to extract ITS sequences from genomes generates a consensus sequence. However, inspection of Illumina reads mapping to ITS in 60 dikaryotic genomes of *L. edodes* s. lat. published by Xiao et al. (19) revealed cryptic heterogeneity in ITS within individuals. Eight genomes bearing an apparent Group 5 ITS are polymorphic at 8 of the 20 conserved nongapped positions that discriminate the ITS types (*SI Appendix*, Figs. S27 and S28). In contrast, there is no evidence of polymorphism in any of the genomes with an apparent Group 1 ITS and one of the genomes with an apparent Group 5 ITS (YS3353), except at a single position (site 669).

Wild Populations of *L. edodes* s. lat. Harbor Unique Genetic Diversity. The pangenome for 27 cultivar genomes of *L. edodes* s. lat. contains 16,864 orthogroups, of which 7,248 orthogroups comprise the core genome (i.e., the accessory genome contains 9,616 orthogroups). However, with addition of 39 wild genomes, the pangenome expanded to 20,308 orthogroups, while the core genome declined to 6,483 orthogroups, indicating that wild genomes of *L edodes* s. lat. contain 3,444 orthogroups that are not present in cultivar genomes (Fig. 6). The pangenome rarefaction curve for *L. edodes* s. lat. appears to be approaching saturation. The pan-, core, and accessory genomes of *L. raphanica* contain 12,608, 9,501, and 3,107 orthogroups, while those of the *L. boryana* complex contain 10,658, 8,781, and 1,877 orthogroups, respectively. However, those groups are represented by only six genomes each, and their rarefaction curves do not appear to be saturated (*SI Appendix*, Fig. S29).

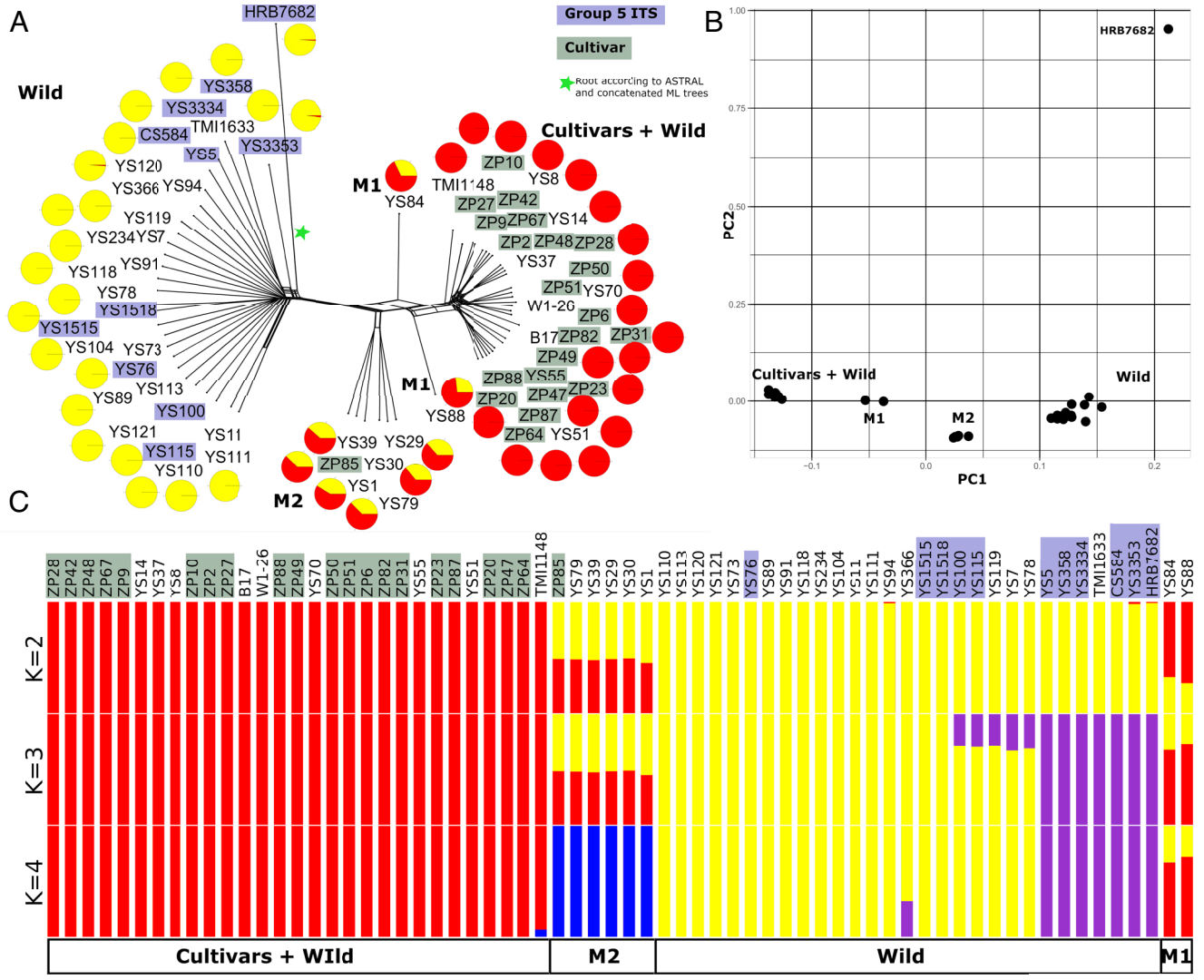

Fig. 4. Population structure in *L. edodes* s. lat. based on 3 × 10^6 SNPs obtained from assemblies. (*A*) SplitsTree graph of SNP data, with Admixture pie graphs at *K* = 2. (*B*) Admixture results at *K* = 2, 3, and 4. (*A* and *B*) Labels for isolates with group 5 ITS type are shaded in purple (others are group 1). Labels for cultivars are shaded in gray (others are wild isolates). (*C*) PCA of SNPs.

Discussion

***Lentinula* Originated in the Tropics and Occurs Mainly on Oaks and Relatives.** Three of the four major lineages of *Lentinula* (*L. aciculospora, L. raphanica,* and the *L. boryana* complex) are found in tropical and subtropical regions in the Americas, while the fourth occurs in Asia and Australasia. Contrary to prior analyses of ITS and *tef*1-α sequences (10), *L. aciculospora* is the sister group of the Asian–Australasian group, which renders the lineages from the American tropics paraphyletic. Genome data are lacking for several tropical species of *Lentinula*, including *L. madagasikarensis* (Madagascar), *L. platinedodes* (Vietnam), *L. guarapiensis* (Paraguay), *L. ixodes* (Brazil) (43), and undescribed specimens from Central Africa (*SI Appendix*, Table S1), which precludes historical biogeographic analyses. Nevertheless, it is most likely that *Lentinula* arose in tropical or subtropical regions, perhaps the Neotropics, as suggested by Menolli et al. (10). *Lentinula edodes* s. lat. is nested within a paraphyletic assemblage of lineages from Australasia, including *L. lateritia*, *L.* aff. *lateritia,* and *L. novae-zelandiae*, suggesting that it represents a unique expansion of *Lentinula* into the north temperate zone.

Most species of *Lentinula* occur on Fagales, such as the shii tree, *Castanopsis cuspidata*, for which shiitake is named. The exceptions include *L. madagasikarensis*, which is associated with Malvales, Malpighiales, and Myrtales (3), and *L. raphanica*, which is reported on eight orders of angiosperms, including Fagales (5, 10). *Lentinula raphanica* is also notable for its broad geographic distribution, from the Gulf Coast of North America to the southern Atlantic coast of Brazil (10). The genome of *L. raphanica* contains 636 orthogroups that are not found in any other *Lentinula* species, including 507 with no known homologues in any organism when aligned with the NCBI nr database (May, 2021; Fig. 2 and *SI Appendix*, Fig. S1). It is not clear whether the expansion of *L. raphanica* onto new substrates in diverse habitats is related to the acquisition of the "orphan" genes in this group. The uniformity of PCWDEs and CYPs across *Lentinula*, including *L. raphanica*, suggests that adaptation to new hosts and environments is not correlated with gains or losses of genes encoding known groups of decay-related enzymes (*SI Appendix*, Figs. S3–S12, Table S2, and Datasets S2 and S3). Fossils and molecular clock analyses suggest that Fagales originated in the Cretaceous period (44), making it plausible that they include the original substrates for *Lentinula*.

Species of *Lentinula* are often described as having flavors and aromas reminiscent of cabbages, radishes, or garlic, which are due to the presence of the organosulfur compound lenthionine.

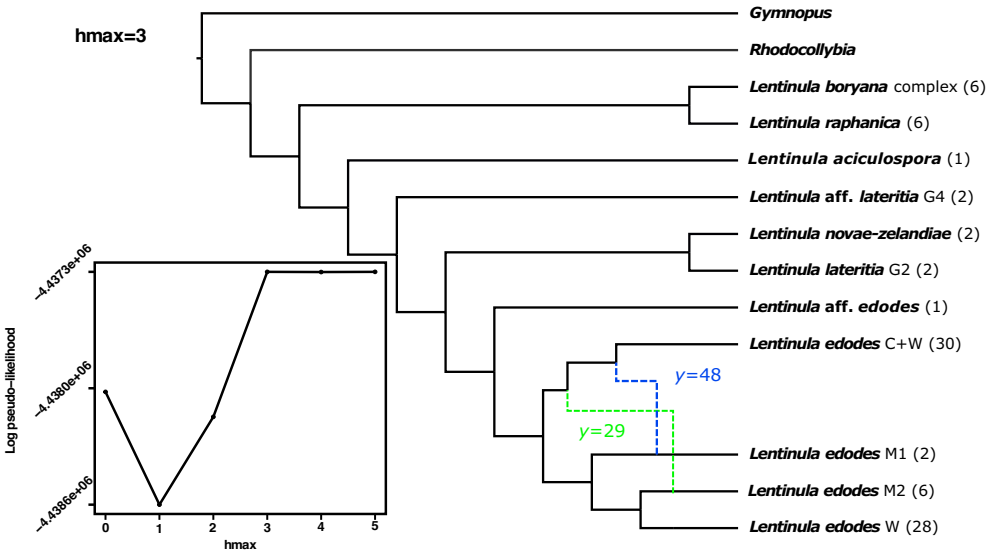

Fig. 5. The best-supported phylogenetic SNaQ network inferred from pseudolikelihood inference with incomplete lineage sorting based on 196 gene trees from an 86-genome dataset and enforcing M1 and M2 clades as hybrids, implemented in PhyloNet. The number of genomes per group is given in parentheses. Both M1 and M2 were resolved in a clade with *L. edodes* W with different amounts of introgression from *L. edodes* C + W. M1 as sister to the M2, and the W clade shows more introgression with the *L. edodes* C + W clade. This pattern may be indicative of cultivated strains being reintroduced to hybridize with natural populations of the W clade. *Inset*: Comparison of model likelihoods for different maximum number of putative hybridizations (hmax) shows the best model at hmax = 3.

(Indeed, the specific epithet *raphanica* alludes to radishes.) The unique tandem and distal duplications of genes encoding γ-glutamyl transpeptidase may be responsible for this trait and thus are potential targets for future breeding efforts (*SI Appendix*, Figs. S13–S15).

Mating Compatibility Underestimates Species Diversity in *Lentinula*. The number of species in *Lentinula* has been controversial. To estimate species limits, we performed Bayesian analyses under the MSC model using the 28-genome core comparative dataset. However, such approaches have been criticized on the grounds that they may misinterpret population structure as evidence for species boundaries (45, 46). Moreover, the number of species clusters that we obtained was sensitive to collapse height (ε) settings in STACEY ranging from thirteen at ε = 0.0001 to nine at ε = 0.0003 (Fig. 3). To evaluate alternative scenarios, we referred to prior taxonomy. Six described (or provisional) species were resolved in all STACEY analyses, including *L.* aff. *lateritia*, *L. aciculospora*, and the four species of the *L. bory*ana complex, which Menolli et al. (10) recognized based on anatomical characters and divergence in ITS sequences. The species whose delimitations varied include *L. edodes* s. lat. (resolved as three species at ε = 0.0001 or 0.0002 or one species at ε = 0.0003), *L. raphanica* (two species at ε = 0.0001 and one species at ε = 0.0002 or 0.0003), and *L. lateritia* plus *L. novae-zelandiae* (two species at ε = 0.0001 and one species at ε = 0.0002 or 0.0003).

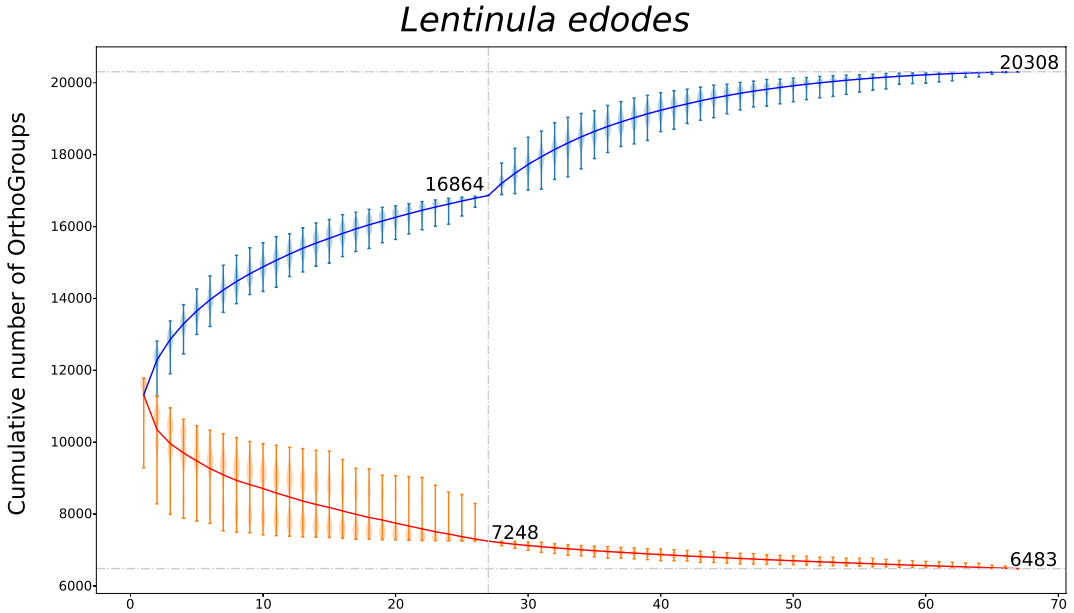

Fig. 6. Rarefaction curve showing pangenome (blue line) and core genome (red line) in *L. edodes* s. lat. X axis indicates number of isolates sampled; Y axis indicates number of orthogroups in each category (an orthogroup may contain one or more gene copies). Core are orthogroups that are present in all isolates up to that point, whereas the pangenome contains orthogroups present in at least one isolate prior to that point. Each defined range of isolates (cultivars: *Left*; wild: *Right*) is produced using 1,000 iterations with random genome addition sequences. Violin plots indicate the distribution of values for the different permutations.

There is no taxonomic precedent for splitting *L. raphanica* or *L. edodes*, but *L. lateritia* and *L. novae-zelandiae* were recognized as discrete species in the monograph of *Lentinula* by Pegler (1). *Lentinula novae-zelandiae* has a stipe covered with fibrillose scales and a wrinkled pileus, whereas *L. lateritia* is smooth, and the two species are geographically isolated. Accordingly, we consider the 13-species scenario obtained at ε = 0.0001 to be the most plausible. Under that scenario, the W and CW groups and isolate HRB7682 from Nepal of *L. edodes* s. lat. would represent three independent species (exemplars of M1 and M2 were not included in the core dataset) (Fig. 3). Geographic ranges of the CW and W groups broadly overlap in central and southwest China (particularly Yunnan, Sichuan, and Shaanxi provinces; *SI Appendix*, Table S1 and Dataset S1) (17), but the potential for movement of cultivars makes it difficult to assess whether strains of the CW group collected in the wild are truly native or are escaped cultivars, as suggested by Zhang et al. (17). SNP and species network analyses suggest that repeated episodes of gene flow from the CW group into the W group have occurred, resulting in the hybrid M1 and M2 groups.

The Nepalese isolate HRB7682 was consistently placed as the sister group of the rest of *L. edodes* s. lat. in phylogenetic analyses, and it was clearly separated from the CW, W, M1, and M2 groups in PCA of SNP data (Figs. 2 and 4 and *SI Appendix*, Fig. S16, S17, S22, and S23). A large geographic distance separates HRB7682 from the other isolates of *L. edodes* s. lat. in our study, but that could be an artifact of limited sampling. Genomic and morphological analyses of additional specimens from across South Asia are needed to assess the ranges and limits of cryptic species within *L. edodes* s. lat. In the absence of such studies, we propose that the Nepalese isolate be classified as *L.* aff. *edodes* and that the CW and W groups be regarded as populations of *L. edodes* sensu stricto (s. str.). Similarly, *L. raphanica* s. lat. should be studied across its range, particularly in Central America, to assess whether it harbors multiple species. At present, lacking morphological evidence, and with limited geographic sampling, it would be premature to describe new species within either *L. edodes* s. lat. or *L. raphanica* s. lat.

Lentinula aff. *lateritia* was consistently resolved as a discrete species under the MSC. *Lentinula* aff. *lateritia* overlaps with *L. lateritia* in PNG (the latter is also distributed in Australia), but they occur in different mountain ranges and at different elevations (12), are nonsister lineages, and are incapable of mating (12). Thus, there is ample evidence to support recognition of *L.* aff. *lateritia* as a unique species, pending morphological description, and designation of a type specimen.

Excluding *L.* aff. *lateritia*, all species of *Lentinula* in Asia–Australasia that have been tested are mating compatible (12) based on the production of clamp connections in test crosses between monokaryons, which have been widely used to assess "biological species" in basidiomycetes (47). Consequently, some authors have suggested that all Asian–Australasian lineages of *Lentinula* represent *L. edodes* (12, 48). Nevertheless, concatenated phylogenies, species tree analyses, and STACEY analyses consistently support the separation of *L. lateritia, L.* aff. *lateritia*, *L. novae-zelandiae*, and *L. edodes* s. lat. as independent lineages (Figs. 2 and 3 and *SI Appendix*, Figs. S22 and S23). Similarly, several species in the *L. boryana* complex are mating compatible based on formation of clamp connections in culture (5), which further demonstrates that the evolution of prezygotic isolation lags behind speciation in *Lentinula*.

ITS Has Low Taxonomic Resolution and Is Polymorphic in *L. edodes* s. lat. ITS is the standard barcode marker for fungi, with a 97 to 99% sequence identity in this region commonly used to delimit species (49, 50). Our sample of *Lentinula* genomes includes 13 groups that may warrant recognition as species. However, using a 97% sequence identity cutoff, we would recognize only four species, corresponding to *L. aciculospora, L. raphanica,* the *L. boryana* species complex, and the entire Asian–Australasian clade (*SI Appendix*, Fig. S26). At a 99% similarity cutoff, 8 species would be resolved, with only two species in Asia–Australasia.

Further complicating the use of ITS as a species marker in *Lentinula* is the existence of two nonsister ITS types in *L. edodes* s. lat., previously called Group 1 and Group 5 (10, 11) (*SI Appendix*, Fig. S25). Phylogenomic analyses suggest that these two ITS types have been retained as a polymorphism within *L. edodes* s. lat. (Fig. 4 and *SI Appendix*, Figs. S22 and S23). Monophyly of the Group 1 plus Group 5 ITS types was rejected, which implies that they did not both arise in the lineage leading to *L. edodes* s. lat. The *L. edodes* s. lat. genomes that bear signatures of both ITS Group 1 and Group 5 are all dikaryons (*SI Appendix*, Fig. S28), suggesting that intragenomic heterogeneity in ITS, where it occurs, may be partitioned among nuclei. Owing to the limitations of ITS, alternative barcode markers should be developed for rapid biodiversity assessment in *Lentinula*, particularly in Asia.

Conservation, Domestication, and the Shiitake Pangenome. An understanding of the genetic diversity of a species and its distribution is important for both conservation and strain development. Across 66 *L. edodes* s. lat. genomes, we found that the core genome includes only 32% of the genetic diversity (orthogroups) of the pangenome (Fig. 6). The accessory genome of *Lentinula* is larger than in most previously described fungi (51, 52).

Within *L. edodes* s. lat., wild genomes contain 3,444 orthogroups that are not present in cultivars (possibly reflecting bottlenecks due to domestication), which could be useful for strain development. Unfortunately, wild populations of *L. edodes* s. lat. are potentially susceptible to loss of diversity due to interbreeding with escaped cultivars (53). The situation is particularly worrisome in South Asia and the adjacent Yunnan and Sichuan provinces of China, which hold a large concentration of the W group of *L. edodes* s. str. and *L.* aff. *edodes*. Wild shiitake is traditionally foraged in Nepal, but current mushroom farms use commercial cultivars (54, 55). Shiitake is also cultivated in New Zealand and Australia, which poses a danger to local populations of *L. novae-zelandiae* and *L. lateritia*, with which it is capable of mating.

There have been at least two domestication events in *L. edodes* s. lat. represented by the cultivated isolate ZP85 in the M2 group and multiple cultivar isolates in the CW group, but it is difficult to estimate the precise number, owing to lack of resolution in the CW group (Fig. 4 and *SI Appendix*, Figs. S22 and S23). Zhang et al. (17) sampled four individuals of *L. edodes* s. lat. from Jiling and Liaoning provinces, adjacent to the Korean Peninsula, which they placed in their "cultivars" group. Our newly sequenced genomes in the CW group include three individuals from Japan, North Korea, and the Russian Far East, which lends support to the view that modern shiitake cultivars originated in Northeast Asia (*SI Appendix*, Table S1). Genome sampling of *L. edodes* s. lat. remains sparse outside of China. To reconstruct the domestication of shiitake, it will be necessary to sequence the genomes of additional wild strains from continental Asia and Japan.

Methods

Sampling. Isolates used in this study were obtained as live or lyophilized cultures from the Tottori Mycological Institute (Tottori, Japan), University of Tennessee fungal culture collection (Knoxville, TN, USA), University of South Alabama (USAM), Duke University fungal culture collection (Durham, NC, USA), National Institute of Amazonian Research

(INPA, Manaus, Brazil), Northwest Mycological Consultants (Corvallis, OR, USA), and Komarov Botanical Institute Basidiomycetes Culture Collection (St. Petersburg, Russia). All field collections were made prior to 1997, except Brazilian isolates of *L. raphanica* and *L. boryana* s. lat., which were obtained under a material transfer agreement between the INPA and Clark University, and *L. aciculospora* JLM 2183, which was collected under contract with CONAGEBIO, MINAE, Costa Rica, permit R-020-2014. Cultures were maintained on malt extract agar and grown on liquid malt extract medium for DNA and RNA isolation at the Clark University.

Genome Sequencing and Annotation. Genome and transcriptome sequencing were performed under the auspices of the Department of Energy Joint Genome Institute (JGI) Community Science Program at the JGI in Walnut Creek and Berkeley, CA. Genomes were sequenced using Illumina and PacBio technologies and assembled with Falcon 0.7.3 and 1.8.8 (PacBio) and SPAdes 3.12.0 and 3.13.0 or Velvet 1.2.07 (Illumina). All genomes were annotated using the JGI Annotation Pipeline and made available via the JGI fungal genome portal MycoCosm (http://mycocosm.jgi.doe.gov/Lentinula) (56). The 60 *L. edodes* s. lat. genomes published as unassembled Illumina reads by Xiao et al. (19) were assembled using SPAdes v.3.13 (57) with reference to the *L. edodes* W1-26 genome (13) and assessed using BUSCO 5.3.2 with the OrthoDB v10 Basidiomycota database (*SI Appendix*).

Analyses of Functional Genes. Genes encoding PCWDEs in a core dataset of 28 *Lentinula* genomes were annotated using the CAZy pipeline (31). Genes encoding decay-related oxidoreductases (class II peroxidases [POD], glucose-choline-methanol oxidoreductases, copper radical oxidases, and multicopper oxidases) were further analyzed by querying *Lentinula* genomes with well-characterized genes from various Agaricomycetes (*SI Appendix*) or by searching according to the EuKaryotic Orthologous Groups (KOG) classification, followed by phylogenetic and sequence-based structural analyses. Phylogenetic analyses of genes encoding class II peroxidases from *Lentinula* and 28 other Agaricales genomes analyzed by Ruiz-Dueñas et al. (58) were performed using five "generic peroxidase" sequences of *Stagonospora nodorum* (Pleosporales) to root the tree. To retrieve and annotate cytochrome P450 monooxygenase genes, 84 hidden Markov models were constructed based on all available P450 amino acid sequences in the NCBI nr and MycoCosm databases (as of August 2018). Sequences were clustered using UCLUST 1.2.22q (59), and model names were applied following Nelson (60).

To reconstruct origins of *lecsl* and *leggt* in *Lentinula*, the 28 *Lentinula* core genomes and all available fungal proteomes in the NCBI and MycoCosm (56) databases (December, 2020) were searched with BLASTp v2.5.0+ (Ye et al. 2006), using known *lecsl* and *leggt* genes as queries (13). *Leggt* Pfam (61) domains (PF01019.24) were retrieved from BLAST homologues with at least 40% Pfam query coverage via hmmsearch v3.3.2. Phylogenetic analyses of *leggt* and *lecsl* genes or Pfam domains were performed using IQ-TREE v2.0.3 (-B 1000) (35) following aligning with MAFFT v7.487 (--auto) (62) and trimming using ClipKIT v1.3.0 (63). Syntenic analysis was conducted by extracting up to 10 Kb up- and downstream of *lecsl* and *leggt* genes and aligning them in clinker v0.0.24 (64).

Transcriptome Analyses. RNA was isolated from a dikaryotic strain of *L. edodes* NBRC111202 using the Qiagen RNeasy Plant Mini Kit. Nine samples were taken, with three replicates each, including mycelium at 2 wk on agar media and 2 mo on sawdust (fruiting) media, whole primordia (≈1 cm), caps and stipes of young (≈3 cm) and mature (>3 cm) fruiting bodies, and whole mushrooms at 0 and 4 d after harvest. RNA-seq reads were analyzed using the CLC genomics workbench. Pearson correlation values were calculated across all replicates for each pairwise comparison of *lecsl* and *leggt* homologues (Bonferroni-corrected $\alpha = 0.05$).

Phylogenomic Analyses. Phylogenetic datasets of single-copy orthologs were constructed using OrthoFinder (30) with the flags -M msa -S blast -T iqtree (unless stated otherwise). Multiple sequence alignments were concatenated using a Biopython-based custom script and trimmed using PhyKIT (65) with default settings. Three datasets were assembled, including 1) a 30-genome/2,258-gene "core" dataset containing 28 *Lentinula* genomes and two Omphalotaceae as outgroups, 2) a 39-genome/294-gene "Agaricales" dataset for molecular clock analyses containing 20 *Lentinula* genomes and 17 other Agaricales, with *Plicaturopsis crispa* (Amylocorticiales) and *Serpula lacrymans* (Boletales) as outgroups, and 3) a 90-genome/196-gene "extended" dataset, with 88 *Lentinula* genomes and two Omphalotaceae as outgroups. Unpartitioned phylogenetic analyses were performed using IQTree (all datasets), and partitioned analyses were performed with RAxML (extended dataset only). Species tree analyses were performed using ASTRAL, with 196 gene trees from the extended dataset as input. Molecular clock analyses were performed using BEAST v2.6.6, with four fossils used for calibration, including *Gondwanagaricites* (Agaricales, 117 Mya), *Archaeomarasmius* (Marasmiaceae, 92 Mya), *Nidula* (Nidulariaceae, 50 Mya), and *Palaeocybe* (Psathyrellaceae, 50 Mya). For further details, see *SI Appendix*.

Topology tests were performed using all three datasets. Topologies focused on *L. aciculospora* placed it as the sister group to 1) all Asian–Australasian lineages (the optimal topology), 2) all other American lineages, 3) *L. boryana*, 4) *L. raphanica*, or 5) all other *Lentinula* lineages. Topologies focused on *L.* aff. *lateritia* placed it as the sister group to (1) all other Asian–Australasian lineages (the optimal topology) or (2) other Australasian lineages (*L.* aff. *lateritia* plus L. *novae-zelandiae*). Topologies were evaluated with the bp-RELL, p-KH, p-SH, and c-ELW tests in IQTree V.1.6.1 (35). As a further sensitivity analysis, the three datasets used in phylogenomic analyses were subsampled according to six metrics: alignment length, degree of violation of the molecular clock (DVMC), median long-branch score, percent parsimony informative sites, saturation, and treeness divided by relative composition variability (RCV). Each statistic is inferred from phylogenetic or alignment properties (e.g., branch lengths and character composition) and is associated with robust phylogenetic signal (63, 66–68). For each metric, partitioned data matrices were made from the best scoring 50% of genes and analyzed using the concatenated and trimmed alignment as input for IQTree v.1.6.1. A file with the topologies to test was generated manually by modifying a previous IQTree run and provided as input with the option -z. IQTree was run with the options -n 0 -nt AUTO -zb 100. To determine if any taxa are susceptible to potential long-branch artifacts, long-branch scores per taxa were also examined (36). Metrics were calculated using PhyKIT, v1.2.1 (65).

Species Delimitation. STACEY v.1.2.5, implemented in BEAST v.2.6.3, was used to assess species limits under the MSC model (37, 38, 69). SortaDate (70) was used to select the best 200 genes in the core dataset using the ML tree inferred from the core dataset as a topological guide tree. An exponential relaxed clock model was used, and all priors and operator values were kept at default values. The analysis was run independently three times sampling trees every 5,000 generations for 100 million generations to assess convergence and ensure that ESS values are above 200, which were summarized in Tracer v1.7.2 (71). Trees from the three runs were combined in LogCombiner with a 10% burn-in. The proportion of species assignments among the posterior trees was summarized using SpeciesDelimitationAnalyzer with a burn-in of 0, a sim cutoff of 1.0, and various collapse heights. The resulting posterior distribution of trees was visualized using the plot.simmatrix() function in R (72).

ITS Analyses. ITS sequences were extracted from genome sequences or were taken from the studies of Menolli et al. (10) and earlier studies (11), except the sequence of *L. platinedodes*, which was downloaded directly from the GenBank. Maximum likelihood trees were constructed using RAxML (73) hosted at vital-it.ch with default settings and 1,000 bootstrapping replicates; bayesian trees were computed using Mr.Bayes version 3.2.7 (74), and pairwise distances among sequences were calculated. Topology tests were conducted in IQ-TREE using an unedited ITS alignment (1,182 nt) or an alignment that was trimmed using ClipKIT (945 nt) (63).

Population Structure Analyses. PacBio sequenced *L. edodes* isolate HRB7683 was utilized as the reference genome for variant calling. The repeat masked assembly (Lenedo1_AssemblyScaffolds_Repeatmasked.fasta.gz from JGI) was indexed with bwa index (BWA), and paired-end Illumina reads (from QC) were mapped with bwa mem (BWA) using parameters "-M -t16" and samtools view using "-buS" (SAMtools). BAM files were sorted with SAMtools sort and read groups added with bamaddrg (2020 Erik Garrison). Sorted BAM files were again indexed with samtools index, and the reference genome indexed with samtools faidx. FreeBayes-parallel was run on 10,000-bp subsets of the indexed reference genome using parameter "-C 10" to set a minimum of 10 observations per variant. Linkage disequilibrium analysis was conducted using LDkit v1.0 (75), and SNPs were filtered using VCFtools including "--thin 5,000 --remove-filtered-all --max-missing-count 6 --maf 0.03 --recode --recode-INFO-all" to eliminate low-occurrence SNPs and subsample at a spacing of 5,000 bp to avoid linkage. SNP arrays were converted to multifasta alignments with VCF kit and analyzed with SplitsTree using the NeighborNet algorithm.

For population structure analyses, the subsampled and filtered vcf file from vcftools was converted to .bed and .bim formats with associated index and list files using plink --vcf with the "allow extra chromosomes" (--aec) option. Admixture v1.3.2 was then run for K (population) values between 2 and 15 using the --cv option for five cross-validation runs per K value. The K value resulting in the lowest cross-validation error was then selected for visualization and further validation. Figures were created using R's barplot and pie functions from Admixture ancestry percentage output. Unrooted phylogenetic networks based on SNPs were constructed using SplitsTree (41). Variant tables in VCF format were processed with PLINK (76) using option --aec, and PCA results were plotted in R using ggplot2 (77).

Species Network Analyses. Phylogenetic networks for the extended dataset were inferred in SNaQ using the maximum pseudolikelihood approach (42). A constrained analysis using the -h option to designate groups M1 and M2 in the *L. edodes* complex as hybrids. Networks were visualized in IcyTree (78).

Pangenome Analyses. Analyses on the distribution of orthogroups were computed using the OrthoFinder output and a series of custom scripts that we have deposited at GitHub.

Data, Materials, and Software Availability. Newly sequenced genomes are available through the JGI MycoCosm portal (https://mycocosm.jgi.doe.gov/Lentinula/Lentinula.info.html) (79) and have been deposited in the DDBJ/ENA/GenBank databases. Genome statistics, including NCBI accession numbers, and data compiled for the analyses can be found at Open Science Framework (https://osf.io/594jv/; see *SI Appendix* for details) (80), and the code created for this project and descriptions of workflows have been deposited at GitHub (https://github.com/MANaranjo/Sierra-Patev-et-al-2022) (81) and Bitbucket (https://bitbucket.org/berkeleylab/jgi-lentinula/src/master/) (82).

ACKNOWLEDGMENTS. This research was supported by the NSF awards DEB-1456588 to D.H. and DEB-1442113 to A.R., the Department of Energy award 0000238431 to D.H., the GENOBIOREF (BIO2017-86559-R) project of the Spanish Ministry of Science and Innovation (cofinanced by FEDER funds) to F.J.R.-D., S.C., and A.T.M., grant PID2019-103901GB-I00 of the Spanish Ministry of Science and Innovation to P.F., and the NIH award R56AI146096 to A.R. J.L.S. and A.R. were supported by the HHMI and the Burroughs Wellcome Fund. The work (proposal: 10.46936/10.25585/60001048) conducted by the US Department of Energy JGI (https://ror.org/04xm1d337), a DOE Office of Science User Facility, was supported by the Office of Science of the US Department of Energy operated under contract no. DE-AC02-05CH11231.

Author affiliations: [a]Biology Department, Clark University, Worcester, MA 01610; [b]U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [c]Department of Plant Pathology, Ohio State University, Columbus, OH 43210; [d]Iwate Biotechnology Research Center, Kitakami, Iwate 024-0003, Japan; [e]Department of Biological Sciences and Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN 37235; [f]Centro de Investigaciones Biológicas "Margarita Salas," Consejo Superior de Investigaciones Científicas, Madrid E-28040, Spain; [g]Department of Biochemistry and Molecular and Cellular Biology, University of Zaragoza, 50009 Zaragoza, Spain; [h]Institute of Biocomputation and Physics of Complex Systems, University of Zaragoza, 50018 Zaragoza, Spain; [i]DTU Bioengineering, Technical University of Denmark 2800, Kgs. Lyngby, Denmark; [j]Department of Biological Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia; [k]Architecture et Fonction des Macromolécules Biologiques, CNRS, Université 13288, Marseille, France; [l]INRAE, UMR 1163, Biodiversité et Biotechnologie Fongiques 13009, Marseille, France; [m]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996; [n]Department of Biology, University of South Alabama, Mobile, AL 36688; [o]Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia, Petrópolis, Manaus AM 69067-375, Brazil; [p]The Tottori Mycological Institute, Japan Kinoko Research Center Foundation, Tottori 689-1125, Japan; [q]Manaaki Whenua - Landcare Research, Auckland 1072, New Zealand; [r]Northwest Mycological Consultants, Corvallis, OR 97330; [s]Institute of Applied Mycology, Huazhong Agricultural University, Wuhan, Hubei 430070, China; [t]Department of Food Science, University of Massachusetts, Amherst, MA 01003; and [u]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA 94720

1. D. N. Pegler, The genus *Lentinula* (Tricholomataceae tribe Collybieae). *Sydowia* **36**, 227–239 (1983).
2. L. X. Tham, N. N. Chuong, P. N. Duong, B. H. Thiem, The biogeographical speciations of shiitake *Lentinula edodes* and a new species *Lentinula platinedodes* sp. nov. found in Cat Tien, South Vietnam. *Acad. J. Biol.* **33**, 29–39 (2012).
3. B. P. Looney, B. Buyck, N. Menolli Jr., E. Randrianjohany, D. Hibbett, *Lentinula madagasikarensis* sp. nov., a relative of shiitake mushrooms from Madagascar. *Fungal Syst. Evol.* **8**, 1–8 (2021).
4. J. L. Mata, R. H. Petersen, A new species of *Lentinula* (Agaricales) from Central America. *Mycoscience* **41**, 351–355 (2000).
5. J. L. Mata, R. H. Petersen, K. W. Hughes, The genus *Lentinula* in the Americas. *Mycologia* **93**, 1102–1112 (2001).
6. A. H. Nguyen *et al.*, Clinical features of shiitake dermatitis: A systematic review. *Int. J. Dermatol.* **56**, 610–616 (2017).
7. K. Yasumoto, K. Iwami, H. Mitsuda, Enzyme-catalized evolution of lenthionine from lentinic acid. *Agric. Biol. Chem.* **35**, 2070–2080 (1971).
8. Y. Liu *et al.*, A novel cysteine desulfurase influencing organosulfur compounds in *Lentinula edodes*. *Sci. Rep.* **5**, 10047 (2015).
9. R. Salwan, S. Katoch, V. Sharma, "Recent developments in shiitake mushrooms and their nutraceutical importance" in *Fungi in Sustainable Food Production* (2021), Chapter 10, pp. 165–180, 10.1007/978-3-030-64406-2_10.
10. N. Menolli Jr., *et al.*, Global phylogeny of the Shiitake mushroom and related *Lentinula* species uncovers novel diversity and suggests an origin in the Neotropics. *Mol. Phylogenet. Evol.* **173**, 107494 (2022).
11. D. S. Hibbett, Y. Fukumasa-Nakai, A. Tsuneda, M. J. Donoghue, Phylogenetic diversity in shiitake inferred from nuclear ribosomal DNA sequences. *Mycologia* **87**, 618–638 (1995).
12. N. Shimomura, K. Terashima, K. Hasebe, Intersterility between populations of *Lentinula edodes* from Papua New Guinea. *Mycoscience* **50**, 240–243 (2009).
13. L. Chen *et al.*, Genome sequence of the edible cultivated mushroom *Lentinula edodes* (shiitake) reveals insights into lignocellulose degradation. *PLoS One* **11**, e0160336 (2016).
14. S. G. Park *et al.*, Long-read transcriptome data for improved gene prediction in *Lentinula edodes*. *Data Brief* **15**, 454–458 (2017).
15. Y. Sakamoto *et al.*, *Lentinula edodes* genome survey and postharvest transcriptome analysis. *Appl. Environ. Microbiol.* **83**, e02990–16 (2017).
16. D. Shim *et al.*, Whole genome de novo sequencing and genome annotation of the world popular cultivated edible mushroom, Lentinula edodes. *J. Biotechnol.* **223**, 24–25 (2016).
17. J. Zhang *et al.*, Population genomics provides insights into the genetic basis of adaptive evolution in the mushroom-forming fungus *Lentinula edodes*. *J. Adv. Res.* **38**, 91–106 (2022).
18. C. A. Smith, Macrosynteny analysis between *Lentinula edodes* and *Lentinula novae-zelandiae* reveals signals of domestication in *Lentinula edodes*. *Sci. Rep.* **11**, 1–11 (2021).
19. Y. Xiao *et al.*, Population genomic analysis uncovers environmental stress-driven selection and adaptation of *Lentinula edodes* population in China. *Sci. Rep.* **6**, 36789 (2016).
20. X. Song *et al.*, Mitogenome types of two *Lentinula edodes* sensu lato populations in China. *Sci. Rep.* **9**, 9421 (2019).
21. L.-Z. Fu *et al.*, Evaluation of genetic diversity in *Lentinula edodes* strains using RAPD, ISSR and SRAP markers. *World J. Microbiol. Biotechnol.* **26**, 709–716 (2009).
22. K. Terashima, T. Matsumoto, K. Hasebe, Y. Fukumasa-Nakai, Genetic diversity and strain-typing in cultivated strains of *Lentinula edodes* (the shii-take mushroom) in Japan by AFLP analysis. *Mycol. Res.* **106**, 34–39 (2002).
23. X. Xiang *et al.*, Genetic diversity and population structure of Chinese *Lentinula edodes* revealed by InDel and SSR markers. *Mycol. Progress* **15**, 1–13 (2016).
24. Y. Xiao, W. Liu, Y.-Y. Lu, W.-B. Gong, Y.-B. Bian, Applying target region amplification polymorphism markers for analyzing genetic diversity of *Lentinula edodes* in China. *J. Basic Microbiol.* **50**, 475–483 (2010).
25. R. Zhang *et al.*, Strain-typing of *Lentinula edodes* in China with inter simple sequence repeat markers. *Appl. Microbiol. Biotechnol.* **74**, 140–145 (2007).
26. F. A. Simao, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
27. E. V. Kriventseva *et al.*, OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807–D811 (2019).
28. M. A. Naranjo-Ortiz, T. Gabaldón, Fungal evolution: Cellular, genomic and metabolic complexity. *Biol. Rev.* **95**, 1198–1232 (2020).
29. R. Castanera, A. Borgognone, A. G. Pisabarro, L. Ramírez, Biology, dynamics, and applications of transposable elements in basidiomycete fungi. *Appl. Microbiol. Biotechnol.* **101**, 1337–1350 (2017).
30. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
31. V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
32. Y. Liu *et al.*, Purification and characterisation of two enzymes related to endogenous formaldehyde in *Lentinula edodes*. *Food Chem.* **138**, 2174–2179 (2013).
33. X. Lei *et al.*, Effects of GGT and C-S lyase on the generation of endogenous formaldehyde in *Lentinula edodes* at different growth stages. *Molecules* **24**, 4203 (2019).
34. W. Li *et al.*, Analysis of volatile compounds of *Lentinula edodes* grown in different culture substrate formulations. *Food Res. Int.* **125**, 108517 (2019).
35. B. Q. Minh *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
36. T. H. Struck, TreSpEx-detection of misleading signal in phylogenetic reconstructions based on tree information. *Evol. Bioinform. Online* **10**, 51–67 (2014).
37. G. Jones, Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *J. Math Biol.* **74**, 447–467 (2017).
38. G. Jones, Z. Aydin, B. Oxelman, DISSECT: An assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics* **31**, 991–998 (2015).
39. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
40. S. Mirarab *et al.*, ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–548 (2014).
41. D. H. Huson, Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2005).

42. C. Solis-Lemus, C. Ane, Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet* **12**, e1005896 (2016).
43. J. J. Oliveira *et al.*, *Lentinula ixodes* comb. nov. (Omphalotaceae, Agaricales) including new records in Brazil. *Mycoscience* **63**, 254–266 (2022).
44. Y. Xing, R. E. Onstein, R. J. Carter, T. Stadler, H. Peter Linder, Fossils and a large molecular phylogeny show that the evolution of species richness, generic diversity, and turnover rates are disconnected. *Evolution* **68**, 2821–2832 (2014).
45. J. Sukumaran, M. T. Holder, L. L. Knowles, Incorporating the speciation process into species delimitation. *PLoS Comput. Biol.* **17**, e1008924 (2021).
46. J. Sukumaran, L. L. Knowles, Multispecies coalescent delimits structure, not species. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1607–1612 (2017).
47. R. H. Petersen, There's more to a mushroom than meets the eye: Mating studies in the Agaricales. *Mycologia* **87**, 1–17 (1995).
48. N. Shimomura, K. Hasebe, N. Fukumasa, M. Komatsu, Intercompatibility between geographically distant strains of Shiitake. *Rept. Tottori Mycol. Inst.* **30**, 26–29 (1992).
49. C. L. Schoch *et al.*, Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6241–6246 (2012).
50. D. Vu *et al.*, Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal species and higher taxon delimitation. *Stud. Mycol.* **92**, 135–154 (2019).
51. C. G. P. McCarthy, D. A. Fitzpatrick, Pan-genome analyses of model fungal species. *Microb. Genom.* **5**, e000243 (2019).
52. C. Plissonneau, F. E. Hartmann, D. Croll, Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* **16**, 5 (2018).
53. D. Hibbett, M. J. Donoghue, Implications of phylogenetic studies for conservation of genetic diversity in shiitake mushrooms. *Conserv. Biol.* **10**, 1321–1327 (1996).
54. M. Christensen, S. Bhattarai, S. Devkota, H. O. Larsen, Collection and use of wild edible fungi in Nepal. *Econ. Bot.* **62**, 12–23 (2008).
55. M. Adhikari, Addition and correction to the knowledge on edibility of wild mushrooms in Nepal: A discussion. *Bull. Department Plant Res.* **36**, 1–15 (2014).
56. I. V. Grigoriev *et al.*, MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**, D699–D704 (2014).
57. A. Bankevich *et al.*, SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
58. F. J. Ruiz-Duenas *et al.*, Genomic analysis enlightens agaricales lifestyle evolution and increasing peroxidase diversity. *Mol. Biol. Evol.* **38**, 1428–1446 (2021).
59. R. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* (2010), in press.
60. D. R. Nelson, The cytochrome P450 homepage. *Hum. Genomics* **4**, 59 (2009).
61. S. El-Gebali *et al.*, The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
62. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **15**, 3059–3066 (2002).
63. J. L. Steenwyk, T. J. Buida 3rd, Y. Li, X. X. Shen, A. Rokas, ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.* **18**, e3001007 (2020).
64. C. L. M. Gilchrist, Y. H. Chooi, Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* **18**, btab007 (2021), 10.1093/bioinformatics/btab007.
65. J. L. Steenwyk *et al.*, PhyKIT: A broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics* **37**, 2325–2331 (2021).
66. M. J. Phillips, D. Penny, The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylogenet. Evol.* **28**, 171–185 (2003).
67. X. X. Shen *et al.*, Reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data. *G3* **6**, 3927–3939 (2016).
68. H. Philippe *et al.*, Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol.* **9**, e1000602 (2011).
69. R. Bouckaert *et al.*, BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
70. S. A. Smith, J. W. Brown, J. F. Walker, So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PLoS One* **13**, e0197433 (2018).
71. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
72. P. Alstrom *et al.*, Multiple species delimitation approaches applied to the avian lark genus *Alaudala*. *Mol. Phylogenet. Evol.* **154**, 106994 (2021).
73. A. Stamatakis, P. Hoover, J. Rougemont, A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* **57**, 758–771 (2008).
74. F. Ronquist, J. P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
75. Y. Tang *et al.*, LDkit: A parallel computing toolkit for linkage disequilibrium analysis. *BMC Bioinformatics* **21**, 1–8 (2020).
76. S. Purcell *et al.*, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
77. R. A. M. Villanueva, Z. J. Chen, ggplot2: Elegant graphics for data analysis (2nd ed.). *Measurement* **17**, 160–167 (2019).
78. T. G. Vaughan, IcyTree: Rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics* **33**, 2392–2394 (2017).
79. Anonymous, Lentinula genomes. JGI MycoCosm. https://mycocosm.jgi.doe.gov/Lentinula/Lentinula.info.html. Deposited 20 August 2022.
80. D. Hibbett, B. Min. Supplementary data for Patev *et al.* Lentinula phylogenomics. Open Science Framework. https://osf.io/594jv/. Deposited 3 December 2022.
81. M. Naranjo-Ortiz, S. Sierra-Patev, Sierra-Patev-et-al-2022. GitHub. https://github.com/MANaranjo/Sierra-Patev-et-al-2022. Deposited 11 October 2022.
82. Berkeley Lab. Untitled. Bitbucket. https://bitbucket.org/berkeleylab/jgi-lentinula/src/master/. Deposited 18 October 2022.