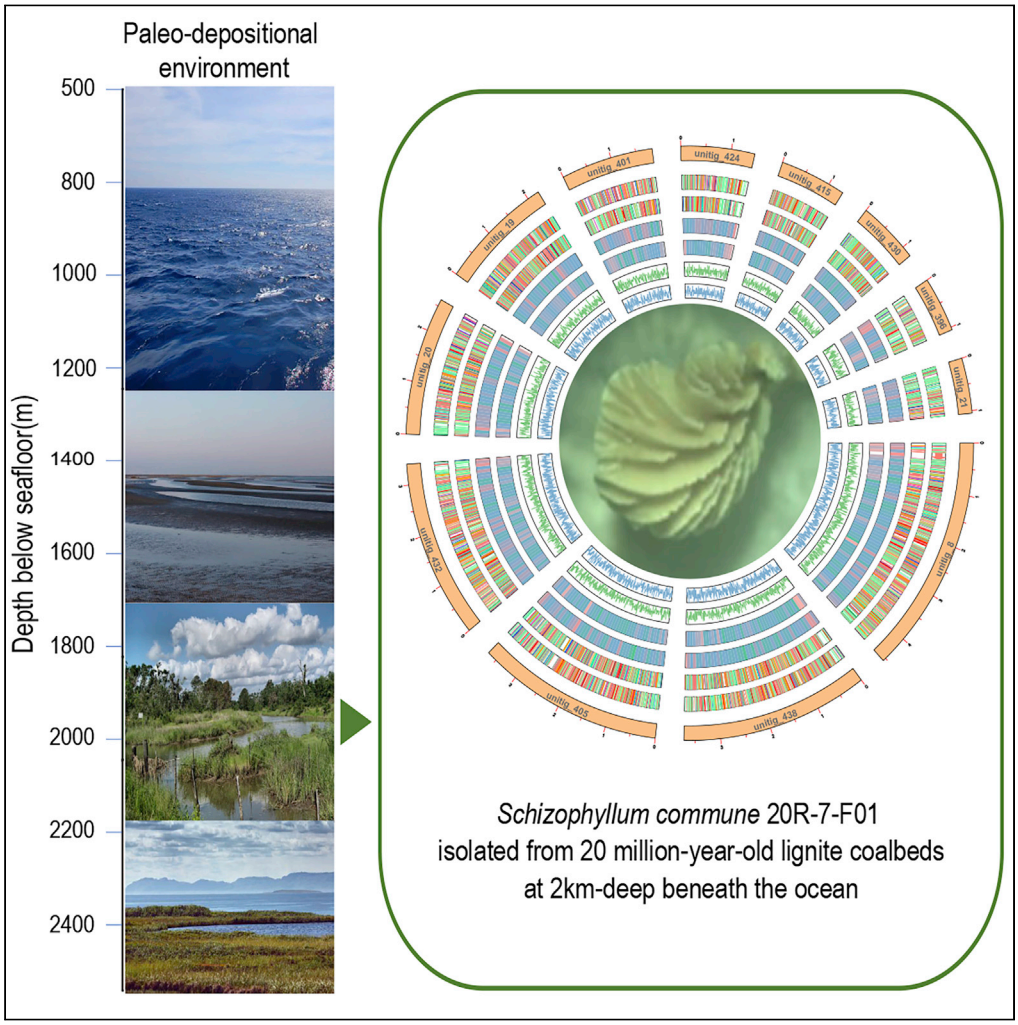
iScience

CellPress
OPEN ACCESS

Article

Genome, genetic evolution, and environmental adaptation mechanisms of *Schizophyllum commune* in deep subseafloor coal-bearing sediments



Xuan Liu, Xin Huang, Chen Chu, ..., Zain Ul Arifeen Muhammad, Fumio Inagaki, Changhong Liu

wanglong@nju.edu.cn (L.W.)
chliu@nju.edu.cn (C.L.)

Highlights

The genome of subseafloor *Schizophyllum commune* is distinct from that of other environmental isolates

The divergence time between strains 20R-7-F01 and H4-8 of *S. commune* is 28–73 Mya

S. commune experiences an evolutionary stasis in the subseafloor environment

DNA repair, transposons, and CAZymes help fungi adapt to the subseafloor environment

Liu et al., iScience *25*, 104417
June 17, 2022
https://doi.org/10.1016/j.isci.2022.104417

iScience

CellPress
OPEN ACCESS

Article

Genome, genetic evolution, and environmental adaptation mechanisms of *Schizophyllum commune* in deep subseafloor coal-bearing sediments

Xuan Liu,[1] Xin Huang,[1] Chen Chu,[1] Hui Xu,[1] Long Wang,[1,*] Yarong Xue,[1] Zain Ul Arifeen Muhammad,[1] Fumio Inagaki,[2,3] and Changhong Liu[1,4,*]

SUMMARY

To understand the genomic evolution and adaptation strategies of fungi to subseafloor sedimentary environments, we *de novo* assembled the genome of *Schizophyllum commune* strain 20R-7-F01 isolated from ~2.0 km-deep, ~20-millionyearsago (Mya) coal-bearing sediments. Phylogenomics study revealed a differentiation time of 28–73 Mya between this strain and the terrestrial type-strain H4-8, in line with sediment age records. Comparative genome analyses showed that FunK1 protein kinase, NmrA family, and transposons in this strain are significantly expanded, possibly linking to the environmental adaptation and persistence in sediment for over millions of years. Re-sequencing study of 14 *S. commune* strains sampled from different habitats revealed that subseafloor strains have much lower nucleotide diversity, substitution rate, and homologous recombination rate than other strains, reflecting that the growth and/or reproduction of subseafloor strains are extremely slow. Our data provide new insights into the adaptation and long-term survival of the fungi in the subseafloor sedimentary biosphere.

INTRODUCTION

Subseafloor sediment is one of the largest and taxonomically diverse microbial habitats and organic carbon pools, with 2.9×10^{29} microbial cells, accounting for 0.18–3.6% of the Earth's total living biomass (Hoshino et al., 2020; Kallmeyer et al., 2012; Parkes et al., 2014; Wormer et al., 2019). Besides the relatively well-studied prokaryotes, a growing number of studies have shown that fungi are also an important component of the subseafloor biosphere, with the deepest distribution down to ~2.5 km (Liu et al., 2017; Pachiadaki et al., 2016). Most of these fungi are affiliated with Ascomycota, Basidiomycota, and Chytridiomycota, whose numbers and distribution are not closely related to depth, but are significantly related to the total organic carbon, electron accepters, temperature, and salinity (Orsi et al., 2013; Rédou et al., 2015). Phylogenetic analysis and the fungal-host co-deposition hypothesis suggest that subseafloor fungi are closely related to terrestrial species, and almost all subseafloor fungi exist in terrestrial or aquatic habitats (Liu et al., 2017; Orsi et al., 2015; Quemener et al., 2020). Previous studies have shown that subseafloor fungi may be alive and metabolically active under the extremely energy-limited conditions of the deep subseafloor (Orsi et al., 2013,2015; Sohlberg et al., 2015). Our previous laboratory investigation also revealed that the subseafloor *S. commune* possess normal growth ability under anaerobic conditions and employ various strategies, such as ethanol fermentation and production of amino acids, increase in mitochondrial number, and autophagosome formation, to adapt to an anaerobic environment (Zain Ul Arifeen et al., 2021a, 2021b). However, the fungus could not produce fruit bodies in the absence of oxygen, suggesting that fungi can only reproduce asexually in deep subseafloor sediments (Zain Ul Arifeen et al., 2021b). All these data indicate that subseafloor fungal species may have unique metabolic activities that helped them sustain life in the deep subseafloor environments for up to ~81 to 48 Mya (Ivarsson, 2012; Ivarsson et al., 2013). Although a large number of studies highlight the diversity and activity of fungi in the deep subseafloor biosphere (Liu et al., 2017; Orsi et al., 2015; Pachiadaki et al., 2016; Quemener et al., 2020; Rédou et al., 2015), how fungi evolved and adapted to the energetically challenging subseafloor environment and have persisted over millions of years remains largely unknown.

[1]State Key Laboratory of Pharmaceutical Biotechnology, Nanjing University, Nanjing 210023, China

[2]Mantle Drilling Promotion Office, Institute for Marine-Earth Exploration and Engineering (MarE3), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama 236-0001, Japan

[3]Department of Earth Sciences, Graduate School of Science, Tohoku University, Sendai 980-8574, Japan

[4]Lead contact

*Correspondence: wanglong@nju.edu.cn (L.W.), chliu@nju.edu.cn (C.L.)

https://doi.org/10.1016/j.isci.2022.104417

Table 1. Genome assembly and annotation summary of *S. commune* strains

Assembly feature	*Strain* 20R-7-F01	*Strain* LoenenD	*Strain* TattoneD	*Strain* H4-8
Genome size (Mbp)	40.79	35.88	36.46	38.48
Coverage (X)	113.1X	112.3X	109.4X	8.29X
Number of Scaffold	162	1,774	1,707	36
N50 (bp) of contigs	1,826,793	54,148	54,683	2,548,518
GC content (%)	57.32	57.50	57.45	56.67
Gene number	10,765	13,827	15,199	13,210
Average gene length (bp)	1,725	1,708	1,692	1,795
Average exon length (bp)	213	247	264	249
Average intron length (bp)	91	76	72	79
Average number of exons per genes	5.9	5.55	5.27	5.7

S. commune is one of the most widely distributed mushrooms in the forests on Earth and an effective wood-degrading basidiomycete, which can produce a large number of hydrolases (xylanase, pectinase, cellulase, and endoglucanase) and oxidoreductase like laccases (Tovar-Herrera et al., 2018). It has also served as a model mushroom to study cell wall biogenesis, hyphal fusion and development, matingtype, heterologous expression of genes, and gene deletions (Ohm et al., 2010). Sequence search of NCBI and JGI databases and previous reports revealed that 58 strains of *S. commune* had been completely or partially sequenced, of which type strain H4-8 had been fully studied (Ohm et al., 2009). Based on culture study of fungal diversity in subseafloor lignite coal-bearing sediment samples collected during the Integrated Ocean Drilling Program (IODP) Expedition 337 (Inagaki et al., 2015), the distribution of *S. commune* has been extended to the sediment down to 2,457 meters below seafloor (mbsf) (Liu et al., 2017). *S. commune*, as one of the predominant subseafloor fungi in those particular geologic horizons, not only has special characteristics of anaerobic growth, but also has a unique anaerobic energy generation pathway, which might endow it to subsist energy-limited subseafloor sedimentary environments (Zain Ul Arifeen et al., 2021a). This long-term physical isolation of *S. commune* in deep subseafloor sediments provides a unique opportunity to investigate how genomic diversity, nucleotide substitution, gene expansion/contraction affect the genetic evolution of *S. commune* over millions of years since they were buried from the ancient forest to lignite-bearing sediment in the subseafloor.

Here, we report a high-quality genome sequencing of *S. commune* 20R-7-F01 that has been buried in a 1,966 m-deep lignite coal-bearing sediment to advance our insight of *S. commune* evolution and adaptation to the deep subseafloor extremes. A large-scale single nucleotide polymorphisms (SNP) data from 14 high-coverage genomes of *S. commune* strains derived from different habitats, including nine (6 subseafloor strains; 1 marine strains, 2 terrestrial strains) newly re-sequenced accessions from this study and five terrestrial accessions from a previous study, was used to investigate the genetic diversity and evolutionary history of *S. commune* and the mechanism of its adaptation to the depositional environments in the deep subseafloor biosphere.

RESULTS

Genome assembly and scaffolding and annotation

Whole genome of *S. commune* 20R-7-F01 (CGMCC 11604) was sequenced using a shotgun sequencing strategy and PacBio RSII long-read sequencing platform. A total of 376,337 reads representing a cumulative size of 4.58 Gb were generated with PacBio RSII sequencing (Table S1). The total sequence coverage of the whole genome assembly was ~110×. After filtering out the low-quality reads, we obtained a composite 40.79 Mb genome from the retained 3.11 Gb clean reads. The final assembly comprised 162 scaffolds with N_{50} of 1.83 Mb, with the longest scaffold being 4.5 Mb (Table 1 and Figure 1). The BUSCO analysis suggested that the annotation set was well completed, with 92.8% complete BUSCOs and 3.7% missing BUSCOs. The Quast analysis indicated that the quality of *S. commune* 20R-7-F01 assembly was similar to that of previously sequenced strain H4-8, but superior to that of the strains LoenenD and TattoneD (Tables S2 and S3).

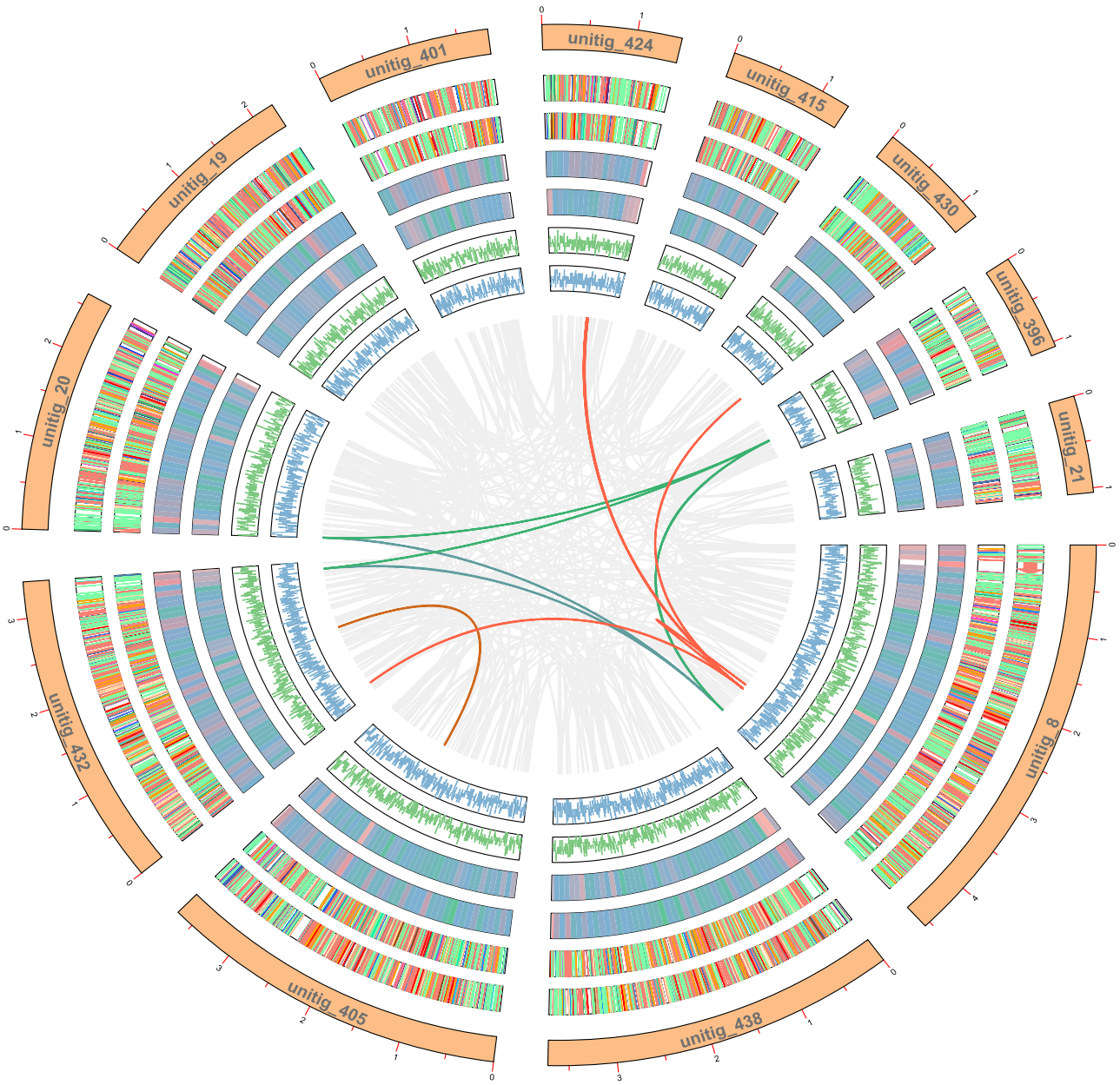


Figure 1. The circos diagram of *S. commune* 20R-7-F01 genome

The outermost layer is the chromosome and its size. The second and third layers are CDS on the positive and negative chains, and the different colors indicate the functional classification of different COGs of the CDS. The fourth and fifth layers are gene density on the positive and negative chains. The sixth and seventh layers are GC content, the green part indicates that the GC content in this area is higher than the whole genome average GC content, and the blue part indicates that the GC content in this area is lower than the whole genome average GC content. Links between all genes represents inparalogs and the bold line indicates the top five genes with the most copies.

Gene annotation on the genome assembly of *S. commune* 20R-7-F01 was performed using genomic and transcriptomic data. Results showed that the genome comprised 10,765 protein-coding genes, 220 tRNA, 78 rRNA, and 69 other ncRNAs fragments. The average length of gene was 1,725.2 bp, and each gene had 5.9 exons on average (Tables S4 and S5). The average length of exon and intron was 213 bp and 91 bp, respectively. The size of the coding region was 18.57 Mb, accounting for 45.53% of the whole genome. 11.06% of the genome sequences contained repeats and the most frequent repetitive elements were LTRs (1.66%) (Table S6). The content of GC in the genome was 57.32%, and GC in the coding region was 59.31%. The genome size is similar to those observed in the existing *S. commune* strains isolated from

the Earth's surface, while the number of protein-encoding genes is lower than other terrestrial strains (Table 1).

Genomic comparison between strains 20R-7-F01 and H4-8

To investigate whether subseafloor strain *S. commune* 20R-7-F01 has unique genetic characteristics, we performed comparative genomic analysis of this strain with the most well-studied terrestrial type strain H4-8. First, a total of 92 gene clusters, including 237 genes, were found only in strain 20R-7-F01, but not in strain H4-8. Except 184 unknown functional genes, other genes are mainly involved in DNA repair, transposon protein, alpha-1,3-glucan synthase, aryl-alcohol dehydrogenase (Table S7). In particular, the number of genes related to DNA replication and repair in strain 20R-7-F01 was 175, 10 more than that in strain H4-8. Second, compared with strain H4-8 (79 bp), strain 20R-7-F01 has larger size of introns (91 bp). Third, a total of 29 gene families, containing 170 genes, mainly encoding FunK1 protein kinase, NADH flavin oxidoreductase, Cytochrome p450 were identified in strain 20R-7-F01, but the copy number of these clusters in strain 20R-7-F01 was at least two more than that of strain H4-8 (Table S8). Fourth, both strains shared 152 conserved syntenic blocks (30.82 Mb), which was accounting for 76 and 80% of the genome size in strain 20R-7-F01 and strain H4-8, respectively. However, a 0.16 Mb inversion occurred on unitig_432 of strain 20R-7-F01, corresponding to the scaffold_7 of strain H4-8 (Figures S1 and S2). In addition, strain H4-8 had 11,343 inserted duplications and 1,021 translocations, while strain 20R-7-F01 had only 9,640 inserted duplications and 958 translocations (Table S9). Fifth, according to Ka/Ks ratio of 16,474 collinear genes, we identified a total of 525 genes that provide evidence of positive selection in strain 20R-7-F01 relative to strain H4-8, which are mainly involved in FunK1 protein kinase, glycosyl hydrolase, as well as regulation of Rho protein signal transduction (Table S10).

Genome evolution and divergence time estimates

To determine the evolutionary relationships of *S. commune* 20R-7-F01 with other closely related fungal isolates, the protein sequences of 11 species/strains were analyzed with a gene family clustering method. A total of 957 single-copy orthologous genes were selected to construct a phylogenomic tree with the maximum likelihood method. Phylogenomic analysis showed that the species in Agaricales were clustered into a branch, of which *S. commune* 20R-7-F01 and *S. commune* H4-8 were a clade. Based on the phylogeny and fossil records, we suggest that *S. commune* 20R-7-F01 diverged from the reference *S. commune* H4-8 genome about 28–73 Mya (Figure 2). A Mitochondrial phylogenetic tree also confirmed that the two strains differentiated about 4.8–22 Mya (Figure S3), which was consistent with the estimated ages of sediment (approx. 20 Mya) from which strain 20R-7-F01 were isolated (Gross et al., 2015). In addition, compared with other fungal species, the subseafloor *S. commune* 20R-7-F01 contained 14 significantly expanded gene families encoding FunK1 kinase, NmrA family, transposon protein, and carboxy-cis, *cis*-muconate cyclase, as well as two contracted gene families encoding pepsin-like aspartate proteases and integrase (Figure 2 and Table S11).

Genetic relationship and population structure

To reveal the evolutionary process of *S. commune* using a large-scale SNP and insertion/deletion (InDel) dataset generated from 14 high-coverage genomes of *S. commune*, including nine newly re-sequenced accessions and five previously studied terrestrial accessions. Approximately 1.77Gb of high-quality reads were collected using Illumina whole genome resequencing of the nine strains, which was aligned against the reference genome with a mapping rate ranging from 73.39 to 96.55%, resulting in coverage depths ranging from 35 to 71×(Table S12). Combining our data with the five previously sequenced strains (all with >50× coverage depth), we identified a total of 19,755,267 SNPs and 3,753,725 InDels in the 14 accessions. Among them, 5,595,289 SNPs and 1,009,532 InDels were found in subseafloor strains, accounting for 39.53 and 36.79% of terrestrial and marine strains, respectively (Table S13).

Using whole-genome SNP data, we performed a neighbor-joining (NJ) phylogenetic analysis and principal component analysis (PCA) on the 31 samples (15 strains, 16 basidiospores) from different habitats. The NJ tree clearly showed that strains collected from subseafloor sediments notably differ from those obtained from terrestrial habits (Figure 3A). The ITS sequence analysis of 127 strains from NCBI databases further confirmed the differentiation between subseafloor and terrestrial strains of *S. commune* (Figure S4). All these results showed that subseafloor strains of *S. commune* are a distinct subset that do not overlap with other sampled diversity. In addition, the marine strain MCCC 3A00233 was clustered in the terrestrial group, suggesting that this strain may have originated from the land. In addition, the results of PCA analysis

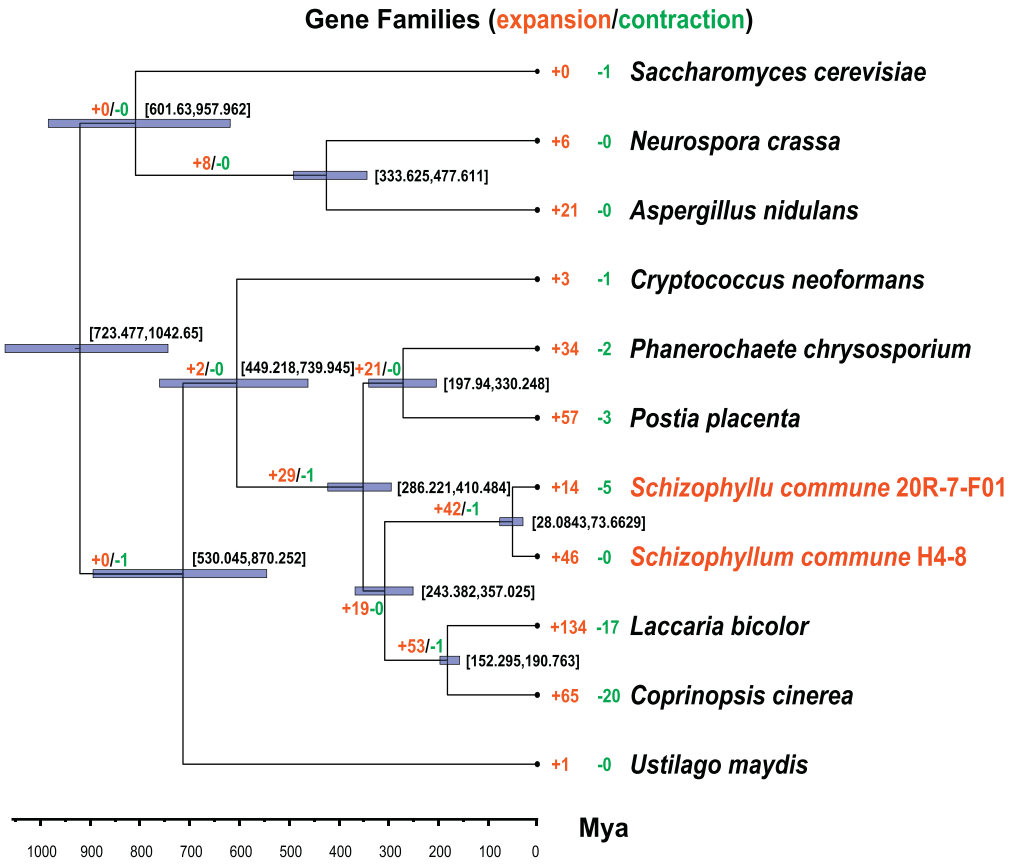


Figure 2. Phylogenetic tree and divergence time of *S. commune* 20R-7-F01

The branch lengths of the phylogenetic tree are scaled to estimated divergence time. The blue bars on the nodes indicate the 95% credibility intervals of the estimated posterior distributions of the divergence times. The overall timeline is shown below the phylogenetic tree.

are consistent to those of NJ tree analysis. Although strains MF and Hom2-8 were relatively far from other terrestrial strains, the first two principal components, PC1 and PC2, could clearly distinguish the subseafloor and terrestrial strains (Figure 3B).

Genetic differentiation and diversity

To understand whether loss-of-function polymorphisms (LoF) such as frameshift and premature stop codon play a selective role in the environmental adaptation of *S. commune*, we calculated these polymorphisms based on SNP datasets. A total of 684 and 2,874 premature stop codons that were caused by SNVs (stop-gain), 18,429 and 66,327 insertion/deletion-induced frameshift variants (frameshift) leading to the disruption of a transcript reading frame, and 101,371 and 392,083 nonsynonymous SNVs were identified in subseafloor and terrestrial strains of *S. commune*, respectively (Table S13). Taking all these stop-gain, frameshift, and nonsynonymous SNVs variants together, we identified 120,484 and 361,284 LoF variants across 9,362 and 10,453 protein-coding genes in subseafloor and terrestrial strains of *S. commune*, respectively. This indicates more unexpected redundancy in terrestrial strains than in subseafloor strains. In addition, 29 LoF genes were only found in the subseafloor strains of *S. commune*, among which two genes were related to transposon protein (Table S14), and the other 1,141 LoF genes were not found in the subseafloor population, involving in structural constituent of ribosome, pyrophosphatase activity, hydrolase activity and ATPase activity (Table S15).

Moreover, genes with LoF mutations showed a correlation with nucleotide diversity (π). The average nucleotide diversity (π) in LoF genes (subseafloor population = 0.0201; terrestrial population = 0.0196) was

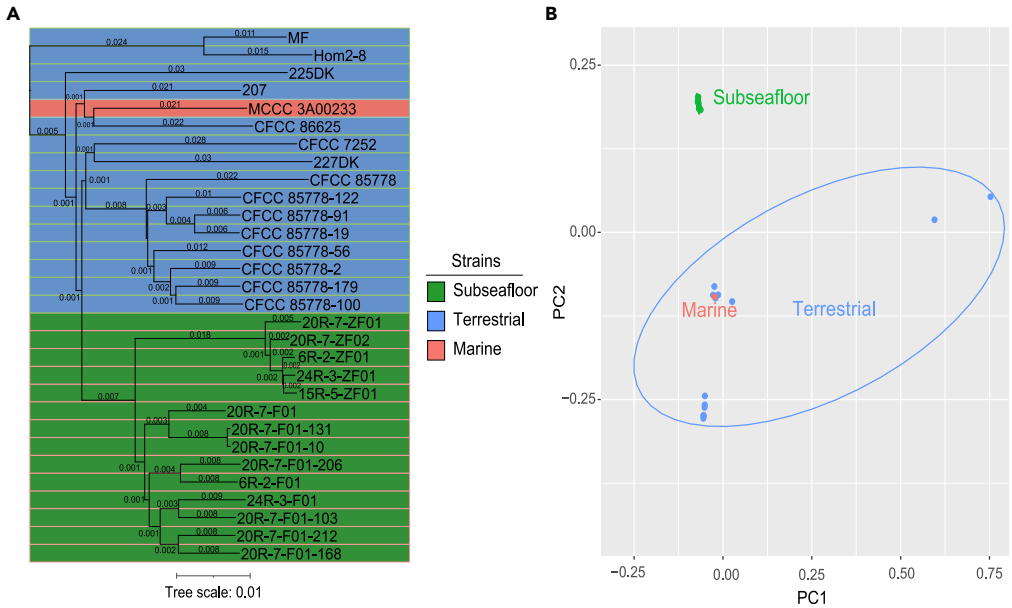


Figure 3. Phylogenetic, population stratification and principal component analyses of *S. commune*

(A) Maximum likelihood phylogenetic tree of subseafloor, marine, and terrestrial *S. commune* population.

(B) PCA analysis for 30 *S. commune* samples.

higher than that in non-LoF genes (subseafloor population = 0.0141; terrestrial population = 0.0112) (Table S16).

Roles of mutation and recombination

To explore the relative effect of nucleotide diversity originating from mutations versus homologous recombination (r/m) on the genetic diversification of populations, we calculate the relative rate of recombination to mutation (R/θ), the mean length of recombined DNA (δ), and the mean divergence of imported DNA (v) for branch tips in *S. commune* genome phylogeny, which allows for a calculation of r/m (r/m = (R/θ) ×δ× v). The result showed that the r/m of subseafloor strains was 0.0526, which was approximately 5,000 times lower than that of terrestrial strains (r/m = 266.66) (Table 2). To avoid the impact of different geographical locations on the analysis results, we choose the small-scale geographical and clonal terrestrial strains (5 strains from Moscow and 4 strains from Florida) (Baranova et al., 2015) to compare with our subseafloor strains by using ClonalFrame software. The result showed that the r/m of Moscow strains was 145.75, and the r/m of Florida strains was 71.10, which were approximately 2,770 and 1,351 times higher than that of subseafloor strains (r/m = 0.0526), respectively (Table S17). To investigate whether the heterozygosity of dikaryotic strains 6R-2-ZF01 (2.96%), 15R-5-ZF01 (3.05%) and 24R-3-ZF01 (2.98%) (Table S13) had an impact on the calculation of r/m value, we compared the r/m value of subseafloor homozygous samples (0.0139) with those of both homozygous and heterozygous samples (0.0526) (Table S18), and found that the r/m value of subseafloor strains was significantly lower than that of terrestrial strains (r/m = 266.66). In addition, we also calculated the index of association (IA) of subseafloor strains and Moscow strains, and found that the frequency of IA in subseafloor strains was concentrated around 0.55, which was higher than that of Moscow strains (0.03) (Figure S5). These consistent results indicated that homologous recombination played a minimal role in diversification of the subseafloor strains.

To estimate genomic diversity of subseafloor and terrestrial strains, we determined the nucleotide diversity (π) and Tajima's D based on SNP variants of 10 re-sequenced strains (Table S19). The results showed that the average nucleotide diversity (π = 0.01758) of the subseafloor strains was approximately ½ that (π = 0.0333) of terrestrial strains, and the nucleotide diversity between the subseafloor strain 20R-7-F01 and other terrestrial strains ranged from 0.068 to 0.104 after Jukes-Cantor corrections (Table S20). Compared with the average π value of homozygous samples (0.01356), the average π value of samples containing both homozygous and heterozygous strains was 0.01758, which was much lower than that (π = 0.0333) of

Table 2. The contribution of recombination and mutation to nucleotide diversity of subseafloor and terrestrial *S. commune* populations

Group	R/θ	δ	v	r/m
Subseafloor	0.00238761	2084.436	0.0105791	0.0527
Terrestrial	0.122791	11831.35	0.183557	266.67

terrestrial strains (Table S21), indicating that the heterozygosity of dikaryotic strains did not affect the results of genomic diversity analysis. Moreover, the subseafloor strains also had a lower ratio of nonsynonymous/synonymous mutation (nonsyn/syn = 0.22) and transitions to transversions (Ts/Tv = 1.93) than that of terrestrial strains (nonsyn/syn = 0.24, Ts/Tv = 1.98), probably due to the extremely slow growth or reproduction of subseafloor strains *in situ* (Table S13).

Selective sweep analysis

Since selective sweep regions usually contain loci related to environment adaptation, we performed a whole-genome screening of the overlapping selective sweep regions by combining Tajima's D with paired F-statistics (*Fst*) analysis. A total of 104 and 244 selective sweep regions with 0.25 Mb (153 gens) and 0.68 Mb (325 genes) genome size were identified in the subseafloor (top/bottom 10% for Tajima's D2.82/0.88) and terrestrial groups (top/bottom 10% for Tajima's D = 1.55/-0.49), *Fst* = 0.4540, respectively. There were 97 genes in the selective sweep regions shared by the subseafloor and terrestrial groups, and 56 and 228 unique genes, respectively. The specific genes in subseafloor group are associated with the activities of selective transposase (4 genes) and hydrolases (2 genes) (Table S22).

DISCUSSION

S. commune is a ubiquitous white rot fungus with a worldwide distribution and has been utilized as a model system for studying mating-type gene function and mushroom development (Ohm et al., 2010). Although the genome of 58 terrestrial strains of *S. commune* has been sequenced to date, only strain H4-8 has been fully annotated (Ohm et al., 2009). Here, we presented for the first time a high-quality genome of *S. commune* 20R-7-F01, which can be used as a model strain of *S. commune* subsisting deep subseafloor environment over 20 Mya for the study of its origination, evolution, and environmental adaptation mechanism by comparison with terrestrial strains. The subseafloor strain contained 10,765 protein coding genes, of which 13.86% had no ortholog in the strain *S. commune* H4-8, indicating that subseafloor *S. commune* are genetically different from their terrestrial counterparts.

Compared with terrestrial strains H4-8, the subseafloor strain 20R-7-F01 has a similar-sized genome, but longer introns and more numbers of genes associated with DNA repair and transposon proteins. The genome size of the subseafloor strain 20R-7-F01 was similar to that of the terrestrial strains, suggesting that hetero recombination or lateral gene transfer played a small role in driving the evolution of the *S. commune* during ~20 million years of depositional history (Nelson-Sathi et al., 2015). It has been reported that the high fidelity of DNA replication and repair mechanisms plays an important role in genome conservation of subsurface microbes (Becraft et al., 2021). As the *in situ* temperature, where strain 20R-7-F01 was isolated, was about 45–50°C, it generally causes damage to DNA, proteins and other biomolecules (Inagaki et al., 2015; Lever et al., 2015; Steen et al., 2013). Therefore, the subseafloor strain possess extra genes related to DNA repair as well as large introns that buffer and regulate fluctuations in protein concentration in fungal cells and protect coding DNA from mutations (Lynch and Conery, 2003; Stival Sena et al., 2014), which might have helped the fungus to survive there and play an essential role in the genome conservation. In addition, five genes encoding transposon proteins in subseafloor *S. commune* 20R-7-F01 were notably expanded compared with other fungi, which indicated that transposon proteins may play an important role in genome plasticity and effective means for organisms to adapt to stressful or extreme environments (Vigil-Stenman et al., 2017; Li et al., 2014).

Meta-omics analyses revealed that subseafloor fungi are not only metabolically active but may play important ecological roles in the subseafloor environments (Quemener et al., 2020; Ivarsson et al., 2018). Little is known about the mechanisms by which fungi adapt to extreme conditions of the subseafloor; however, limited research suggested that fungi sustain life by hydrolyzing their own cell walls (autolysis) under carbon starvation conditions (Emri et al., 2005; Kim et al., 2011). Our previous work showed that the subseafloor

strains were able to utilize lignite/lignin as a sole carbon and energy source (Zain Ul Arifeen et al., 2020), possessed normal growth ability under anaerobic conditions by employing various strategies, including ethanol fermentation and production of amino acids, increase in mitochondrial number, and autophagosome formation (Zain Ul Arifeen et al., 2021a, 2021b). In this study, we also found that subseafloor strain 20R-7-F01 contained more CAZymes coding genes related to cell wall hydrolysis than terrestrial strain H4-8, and these genes may help the fungus obtain energy by degrading the buried organic matter (from terrestrial forest soil to lignite coal) in the sedimentary system (Inagaki et al., 2015). In addition, genes encoding FunK1 kinase, NmrA family, and carboxy-cis, *cis*-muconate cyclase were also found to be expanded in *S. commune* 20R-7-F01 compared to the other fungus, suggesting that NmrA family genes play a vital role in regulating the activity of the GATA transcription factor AreA during nitrogen metabolism in various fungi (Stammers et al., 2001), and indicating that carboxy-cis, *cis*-muconate cyclase is a major enzyme for the dissimilation of aromatic compounds derived from lignin (Thatcher and Cain, 1975). On the contrary, genes encoding retropepsins, pepsin-like aspartate proteases, and integrase were significantly contracted in *S. commune* 20R-7-F01, indicating that *S. commune* 20R-7-F01 could be barely infected by a virus since it isolates from the subseafloor sedimentary environment.

LoF mutations are known to play important roles in adaptation and phenotypic diversification as well as rewire the cell's metabolism (Hottes et al., 2013; Xu et al., 2019), also found that the presence of LoF mutations is correlated with the level of nucleotide diversity, the density of transposable elements (TEs), and gene family size (Xu et al., 2019). However, very little is known about the evolutionary patterns of LoF mutations at the genomic level and the adaptive effects of LoF variants at the population level (Albalat and Cañestro, 2016) in the deep biosphere. Our study first revealed that the frequency and quantity of LoF mutations in subseafloor sedimentary strains of *S. commune* were less than that of terrestrial strains, possibly due to the extremely slow growth and/or proliferation of fungi in the lignite coal-bearing strata, with minimal estimates of microbial turnover time ranging from a few months to over 100 years under the laboratory experimental conditions (Trembath-Reichert et al., 2017). In addition, two LoF genes, related to transposon proteins, were only found in the population of subseafloor strains, which indicated that transposon proteins may play an important role in genome plasticity and effective means for organisms to adapt to stressful or extreme environments (Vigil-Stenman et al., 2017; Li et al., 2014).

The ratio of homologous recombination and nonsynonymous/synonymous mutation can be used to characterize genome evolution (Vos and Didelot, 2009). In this study, we found that the recombination rate and nonsynonymous to synonymous substitution rate of subseafloor strains of *S. commune* were notably reduced compared with that of terrestrial strains. Moreover, similar results were suggested in the subseafloor bacterial population, which had a lower homologous recombination rate than the terrestrial bacterial populations due to the physical isolation of individual cells, reduced cell concentrations, and reduced availability of extracellular DNA for recombination in subseafloor sediments (Lever et al., 2015). Thus, the limited availability of nutrients (Quemener et al., 2020), slow growth/proliferation of hyphae/spores (Pachiadaki et al., 2016; Raghukumar et al., 2004), and few cell numbers (Lever et al., 2015; Quemener et al., 2020) may be the main reasons for the decrease in the genetic diversity of *S. commune* strains in deep subseafloor sediments compared with those on land of this modern world.

Although the strain 20R-7-F01 was isolated from core sediment Unit III and the geological age of this sediment has been determined to be ~20 Mya (Gross et al., 2015), the actual age of the fungus is however difficult to determine. There are great differences in the age of subseafloor fungi reported; for instance, the fossilized fungi isolated from the deep subseafloor basalts (Ivarsson, 2012) and the continental igneous crust (Drake et al., 2019, 2021) had been dated as ~81-48 and ~39 Mya, respectively. However, the age of these fungi might theoretically be much younger than the age of their habitats where active hydrothermal circulation may occur (Ivarsson, 2012; Drake et al., 2019, 2021). This suggests that fungal age based on geological age of the sediments might not represent the accurate age of these anciently deposited fungi to the subseafloor sediment. Interestingly, based on genomic and mitochondrial phylogenetic analysis we found that the divergence time of 20R-7-F01 and H4-8 was 28–73 Mya and 4.8–22 Mya, respectively, which is consistent with the geological age of the sediment (Gross et al., 2015; Inagaki et al., 2012). Although, we could not rule out all possibilities about the origin of fungi in subseafloor sediments, multiple evidences from structural geology, sedimentology, (bio)geochemistry, and

microbiology consistently support that subseafloor fungi are indigenous and have persisted in subseafloor sediments over millions of years (Inagaki et al., 2015; Liu et al., 2017; Zain Ul Arifeen et al., 2021a). However, the arrival of these fungi to subseafloor sediments through underground circulation is not feasible because the occurrence of active fluid circulation is not physically possible under the 2 km-deep subseafloor sedimentary system. In particular, the reverse advection from shallow to deep below the unconformity layer (marine-terrestrial transition) associated with carbonates and conglomerates will not be possible (Inagaki et al., 2012).

Hypervariable species are particularly interested for population genetics (Baranova et al., 2015; Seplyarskiy et al., 2014), and population genetics can examine the distribution of genetic variation and levels of genetic diversity within and between populations (Keats and Sherman, 2014). Genetic diversity of a population is both a factor and an outcome of a wide range of evolutionary processes. In the vast majority of the species, nucleotide diversity (π), the probability that two alleles randomly chosen from the population at a nucleotide site differ from each other, is below 0.03–0.05 (Leffler et al., 2012). However, our results show that *S. commune* exceeds this value, for examples, the nucleotide diversity of the subseafloor strain 20R-7-F01 and other 24 terrestrial haploid strains (Baranova et al., 2015; Seplyarskiy et al., 2014) ranged from 0.07 to 0.11, also the π of the subseafloor strain 20R-7-F01 and the terrestrial strain H4-8 reached about 0.104. These results are consistent with other studies where the diversity at synonymous sites is 0.2 in the American population of *S. commune* and 0.13 in the Russian population, thus confirming that *S. commune* is the most polymorphic among all studied eukaryotic species (Baranova et al., 2015; Seplyarskiy et al., 2014).

Conclusions

In this study, we present the first *de novo* whole genome sequencing and assembly of *S. commune* 20R-7-F01, isolated from a ~20 millionyearsago coal-bearing sediment at ~2.0 km below the ocean floor. Phylogenomic analysis reveals that the divergence between strain 20R-7-F01 and terrestrial type-strain H4-8 is approximately 28–73 Mya. The expansion of the genes encoding FunK1 kinase, NmrA family, transposon protein, and Carboxy-cis, *cis*-muconate cyclase possibly contributed to the extreme environmental adaptation of *S. commune* from the coastal to the deep subseafloor. Nucleotide diversity, substitution rate, and the homologous recombination rate were lower in subseafloor strains than those in terrestrial strains, reflecting that the growth or reproduction of subseafloor strains is extremely slowly. Our data contribute to the study of *S. commune* evolutionary trajectory and phylogenomic diversity among the species and provide novel targets in understanding the molecular adaptation mechanisms of the fungi underlying the geological perturbations associated with the deep subseafloor sedimentary biosphere.

Limitations of the study

Although we report here new insights into the evolutionary history and adaptation mechanism of *S. commune* through comparative genomic analyses, more detailed studies, such as large-scale population genome analysis, will be needed in the future to identify vital gene loci and physiological characteristics associated with environmental adaptation. Also, the adaptation and survival mechanisms of deep subseafloor microbial communities along with increasing depth, pressure, and temperature during depositional regimes (e.g., Heuer et al., 2020; Köster et al., 2021; Beulig et al., 2022) could be addressed by scientific ocean drilling and mimicking cultivation-based laboratory experimentations in the future.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Fungal strains and culture conditions
- METHOD DETAILS
  - De novo genome sequencing and assembly

- Transcriptome sequencing and assembly
- Gene annotation
- Gene family identification and phylogenetic evolution analysis
- Genome resequencing, alignment and variant calling
- Phylogenetic and population structure analyses
- Detection of homologous recombination

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104417.

ACKNOWLEDGMENTS

The authors are grateful to all crews, drilling team members, lab technicians, and shipboard scientists on board the drilling vessel *Chikyu* for supporting core sampling and onboard measurements during IODP Expedition 337. We thank Prof. Kai-Uwe Hinrichs and Prof. Mark A. Lever for their useful discussions. We thank Prof. Sihai Yang of Nanjing University for assistance in genome analyses of strain 20R-7-F01. This work was supported by the National Natural Science Foundation of China (no. 91951121, 41773083, 41973073).

AUTHOR CONTRIBUTIONS

Conceptualization, X.L. and C.H.L.; Resources, X.H., H.X., C.C., Y.R.X., and M.Z.A.; Writing-Original Draft, X.L. and L.W.; Visualization, X.L.; Writing-Review & Editing, X.L., L.W., M.Z.A., F.I. and C.H.L.; Funding Acquisition, C.H.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 5, 2021
Revised: January 30, 2022
Accepted: May 12, 2022
Published: June 17, 2022

REFERENCES

Albalat, R., and Cañestro, C. (2016). Evolution by gene loss. Nat. Rev. Genet. *17*, 379–391. https://doi.org/10.1038/nrg.2016.39.

Baranova, M.A., Logacheva, M.D., Penin, A.A., Seplyarskiy, V.B., Safonova, Y.Y., Naumenko, S.A., Klepikova, A.V., Gerasimov, E.S., Bazykin, G.A., James, T.Y., and Kondrashov, A.S. (2015). Extraordinary genetic diversity in a wood decay mushroom. Mol. Biol. Evol. *32*, 2775–2783. https://doi.org/10.1093/molbev/msv153.

Becraft, E.D., Lau Vetter, M.C.Y., Bezuidt, O.K.I., Brown, J.M., Labonté, J.M., Kauneckaite-Griguole, K., Salkauskaite, R., Alzbutas, G., Sackett, J.D., Kruger, B.R., et al. (2021). Evolutionary stasis of a deep subsurface microbial lineage. ISME J. *15*, 2830–2842. https://doi.org/10.1038/s41396-021-00965-3.

Belloch, C., Querol, A., García, M.D., and Barrio, E. (2000). Phylogeny of the genus Kluyveromyces inferred from the mitochondrial cytochrome-c oxidase II gene. Int. J. Syst. Evol. Microbiol. *50*, 405–416. https://doi.org/10.1099/00207713-50-1-405.

Beulig, F., Schubert, F., Adhikari, R.R., Glombitza, C., Heuer, V.B., Hinrichs, K.U., Homola, K.L., Inagaki, F., Jørgensen, B.B., Kallmeyer, J., et al. (2022). Rapid metabolism fosters microbial survival in the deep, hot subseafloor biosphere. Nat. Commun. *13*, 312. https://doi.org/10.1038/s41467-021-27802-7.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569. https://doi.org/10.1038/nmeth.2474.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271. https://doi.org/10.1093/bioinformatics/btl097.

De Summa, S., Malerba, G., Pinto, R., Mori, A., Mijatovic, V., and Tommasi, S. (2017). GATK hard filtering: tunable parameters to improve variant calling for next generation sequencing targeted gene panel data. BMC Bioinf. *18*, 119. https://doi.org/10.1186/s12859-017-1537-8.

Delcher, A.L., Phillippy, A., Carlton, J., and Salzberg, S.L. (2002). Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. *30*, 2478–2483. https://doi.org/10.1093/nar/30.11.2478.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498. https://doi.org/10.1038/ng.806.

Didelot, X., and Wilson, D.J. (2015). ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput. Biol. *11*, e1004041. https://doi.org/10.1371/journal.pcbi.1004041.

Drake, H., Ivarsson, M., Heim, C., Snoeyenbos-West, O., Bengtson, S., Belivanova, V., and Whitehouse, M. (2021). Fossilized anaerobic and possibly methanogenesis-fueling fungi identified deep within the Siljan impact structure, Sweden. Commun. Earth Environ. *2*, 34. https://doi.org/10.1038/s43247-021-00107-9.

Drake, H., Roberts, N.M.W., Heim, C., Whitehouse, M.J., Siljeström, S., Kooijman, E., Broman, C., Ivarsson, M., and Åström, M.E. (2019). Timing and origin of natural gas accumulation in the Siljan impact structure, Sweden. Nat. Commun. *10*, 4736. https://doi.org/10.1038/s41467-019-12728-y.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157. https://doi.org/10.1186/s13059-015-0721-2.

Emri, T., Molnár, Z., and Pócsi, I. (2005). The appearances of autolytic and apoptotic markers are concomitant but differently regulated in carbon-starving *Aspergillus nidulans* cultures. FEMS Microbiol. Lett. *251*, 297–303. https://doi.org/10.1016/j.femsle.2005.08.015.

Forget, L., Ustinova, J., Wang, Z., Huss, V.A.R., and Franz Lang, B. (2002). Hyaloraphidium curvatum: a linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. Mol. Biol. Evol. *19*, 310–319. https://doi.org/10.1093/oxfordjournals.molbev.a004084.

Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., Wilkinson, A.C., Finn, R.D., Griffiths-Jones, S., Eddy, S.R., and Bateman, A. (2009). Rfam: updates to the RNA families database. Nucleic Acids Res. *37*, D136–D140. https://doi.org/10.1093/nar/gkn766.

Gross, D., Bechtel, A., and Harrington, G.J. (2015). Variability in coal facies as reflected by organic petrological and geochemical data in Cenozoic coal beds offshore Shimokita (Japan) - IODP Exp. 337. Int. J. Coal Geol. *152*, 63–79. https://doi.org/10.1016/j.coal.2015.10.007.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. *9*, R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Heuer, V.B., Inagaki, F., Morono, Y., Kubo, Y., Spivack, A.J., Viehweger, B., Treude, T., Beulig, F., Schubotz, F., Tonai, S., et al. (2020). Temperature limits to deep subseafloor life in the Nankai Trough subduction zone. Science *370*, 1230–1234. https://doi.org/10.1126/science.abd7934.

Hoshino, T., Doi, H., Uramoto, G.I., Wörmer, L., Adhikari, R.R., Xiao, N., Morono, Y., D'Hondt, S., Hinrichs, K.U., and Inagaki, F. (2020). Global diversity of microbial communities in marine sediment. Proc. Natl. Acad. Sci. U. S. A. *117*, 27587–27597. https://doi.org/10.1073/pnas.1919139117.

Hottes, A.K., Freddolino, P.L., Khare, A., Donnell, Z.N., Liu, J.C., and Tavazoie, S. (2013). Bacterial adaptation through loss of function. PLoS Genet. *9*, e1003617. https://doi.org/10.1371/journal.pgen.1003617.

Huang, X., Duan, N., Xu, H., Xie, T.N., Xue, Y.R., and Liu, C.H. (2018). CTAB-PEG DNA extraction from fungi with high contents of polysaccharides. Mol. Biol. *52*, 621–628. https://doi.org/10.1134/s0026893318040088.

Inagaki, F., Hinrichs, K.U., and Kubo, Y. (2012). Deep coal-bed biosphere off Shimokita: microbial processes and hydrocarbon system associated with deeply buried coal-bed in the ocean. IODP Prel. Rept. *337*. https://doi.org/10.2204/iodp.pr.337.2012.

Inagaki, F., Hinrichs, K.U., Kubo, Y., Bowles, M.W., Heuer, V.B., Hong, W.L., Hoshino, T., Ijiri, A., Imachi, H., Ito, M., et al. (2015). Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor. Science *349*, 420–424. https://doi.org/10.1126/science.aaa6882.

Ivarsson, M. (2012). Subseafloor basalts as fungal habitats. Biogeosciences *9*, 3625–3635. https://doi.org/10.5194/bg-9-3625-2012.

Ivarsson, M., Bengtson, S., Drake, H., and Francis, W. (2018). Fungi in deep subsurface environments. Adv. Appl. Microbiol. *102*, 83–116. https://doi.org/10.1016/bs.aambs.2017.11.001.

Ivarsson, M., Bengtson, S., Skogby, H., Belivanova, V., and Marone, F. (2013). Fungal colonies in open fractures of subseafloor basalt. Geo-Mar Lett. *33*, 233–243. https://doi.org/10.1007/s00367-013-0321-7.

Kallmeyer, J., Pockalny, R., Adhikari, R.R., Smith, D.C., and D'Hondt, S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. Proc. Natl. Acad. Sci. U S A *109*, 16213–16216. https://doi.org/10.1073/pnas.1203849109.

Kamvar, Z.N., Brooks, J.C., and Grünwald, N.J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genet. *6*, 208. https://doi.org/10.3389/fgene.2015.00208.

Kamvar, Z.N., Tabima, J.F., and Grünwald, N.J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ *2*, e281. https://doi.org/10.7717/peerj.281.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

Keats, B.J.B., and Sherman, S.L. (2014). Population genetics. In Reference Module in Biomedical Sciences (Elsevier).

Kim, Y., Islam, N., Moss, B.J., Nandakumar, M.P., and Marten, M.R. (2011). Autophagy induced by rapamycin and carbon-starvation have distinct proteome profiles in *Aspergillus nidulans*. Biotechnol. Bioeng. *108*, 2705–2715. https://doi.org/10.1002/bit.23223.

Köster, M., Kars, M., Schubotz, F., Tsang, M.Y., Maisch, M., Kappler, A., Morono, Y., Inagaki, F., Heuer, V.B., Kasten, S., and Henkel, S. (2021). Evolution of (Bio-) geochemical processes and diagenetic alteration of sediments along the tectonic migration of ocean floor in the Shikoku Basin off Japan. Geochem.Geophy. Geosy. *22*. e2020GC009585. https://doi.org/10.1029/2020gc009585.

Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. *34*, 1812–1819. https://doi.org/10.1093/molbev/msx116.

Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. *35*, 3100–3108. https://doi.org/10.1093/nar/gkm160.

Lee, T.H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics *15*, 162. https://doi.org/10.1186/1471-2164-15-162.

Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Ségurel, L., Venkat, A., Andolfatto, P., and Przeworski, M. (2012). Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. *10*, e1001388. https://doi.org/10.1371/journal.pbio.1001388.

Lever, M.A., Rogers, K.L., Lloyd, K.G., Overmann, J., Schink, B., Thauer, R.K., Hoehler, T.M., and Jørgensen, B.B. (2015). Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. FEMS Microbiol. Rev. *39*, 688–728. https://doi.org/10.1093/femsre/fuv020.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189. https://doi.org/10.1101/gr.1224503.

Li, S.J., Hua, Z.S., Huang, L.N., Li, J., Shi, S.H., Chen, L.X., Kuang, J.L., Liu, J., Hu, M., and Shu, W.S. (2014). Microbial communities evolve faster in extreme environments. Sci. Rep. *4*, 6205. https://doi.org/10.1038/srep06205.

Li, X., Xu, S., Zhang, J., and Li, M. (2021). Assembly and annotation of whole-genome sequence of *Fusarium equiseti*. Genomics *113*, 2870–2876. https://doi.org/10.1016/j.ygeno.2021.06.019.

Liu, C.H., Huang, X., Xie, T.N., Duan, N., Xue, Y.R., Zhao, T.X., Lever, M.A., Hinrichs, K.U., and Inagaki, F. (2017). Exploration of cultivable fungal communities in deep coal-bearing sediments from ~1.3 to 2.5 km below the ocean floor. Environ. Microbiol. *19*, 803–818. https://doi.org/10.1111/1462-2920.13653.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964. https://doi.org/10.1093/nar/25.5.955.

Lynch, M., and Conery, J.S. (2003). The origins of genome complexity. Science *302*, 1401–1404. https://doi.org/10.1126/science.1089370.

Nelson-Sathi, S., Sousa, F.L., Roettger, M., Lozada-Chávez, N., Thiergart, T., Janssen, A., Bryant, D., Landan, G., Schönheit, P., Siebers, B., et al. (2015). Origins of major archaeal clades correspond to gene acquisitions from bacteria. Nature *517*, 77–80. https://doi.org/10.1038/nature13805.

Ohm, R.A., de Jong, J.F., Berends, E., Wang, F., Wösten, H.A.B., and Lugones, L.G. (2010). An efficient gene deletion procedure for the mushroom-forming basidiomycete *Schizophyllum commune*. World J. Microbiol. Biotechnol. *26*, 1919–1923. https://doi.org/10.1007/s11274-010-0356-0.

Ohm, R.A., de Jong, J.F., Lugones, L.G., Aerts, A., Kothe, E., Stajich, J.E., de Vries, R.P., Record, E., Levasseur, A., Baker, S.E., et al. (2009). Genome sequence of the model mushroom *Schizophyllum commune*. Nat. Biotechnol. *28*, 957–963. https://doi.org/10.1038/nbt.1643.

Okane, I., Ono, Y., Ohmachi, K., Aime, M.C., and Yamaoka, Y. (2021). Phylogenetic relationships among fern rust fungi and *Desmella lygodii* comb. nov. Mycoscience. *62*, MYC552. https://doi.org/10.47371/mycosci.2021.06.006.

Orsi, W., Biddle, J.F., and Edgcomb, V. (2013). Deep sequencing of subseafloor eukaryotic rRNA reveals active Fungi across marine subsurface provinces. PLoS One *8*, e56335. https://doi.org/10.1371/journal.pone.0056335.

Orsi, W.D., Richards, T.A., and Santoro, A.E. (2015). Cellular maintenance processes that potentially underpin the survival of subseafloor fungi over geological timescales. Estuar Coast Shelf S *164*, A1–A9. https://doi.org/10.1016/j.ecss.2015.04.009.

Pachiadaki, M.G., Rédou, V., Beaudoin, D.J., Burgaud, G., and Edgcomb, V.P. (2016). Fungal and prokaryotic activities in the marine subsurface biosphere at Peru Margin and Canterbury Basin inferred from RNA-based analyses and microscopy. Front. Microbiol. *7*, 846. https://doi.org/10.3389/fmicb.2016.00846.

Parkes, R.J., Cragg, B., Roussel, E., Webster, G., Weightman, A., and Sass, H. (2014). A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere:geosphere interactions. Mar.Geol. *352*, 409–425. https://doi.org/10.1016/j.margeo.2014.02.009.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.

Quemener, M., Mara, P., Schubotz, F., Beaudoin, D., Li, W., Pachiadaki, M., Sehein, T.R., Sylvan, J.B., Li, J., Barbier, G., et al. (2020). Meta-omics highlights the diversity, activity and adaptations of fungi in deep oceanic crust. Environ. Microbiol. *22*, 3950–3967. https://doi.org/10.1111/1462-2920.15181.

Raghukumar, C., Raghukumar, S., Sheelu, G., Gupta, S.M., Nagender Nath, B., and Rao, B.R. (2004). Buried in time: culturable fungi in a deep-sea sediment core from the Chagos Trench, Indian Ocean. Indian Ocean. Deep-Sea Res. Pt I. *51*, 1759–1768. https://doi.org/10.1016/j.dsr.2004.08.002.

Rédou, V., Navarri, M., Meslet-Cladière, L., Barbier, G., and Burgaud, G. (2015). Species richness and adaptation of marine fungi from deep-subseafloor sediments. Appl. Environ. Microbiol. *81*, 3571–3583. https://doi.org/10.1128/aem.04064-14.

Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics *27*, 2325–2329. https://doi.org/10.1093/bioinformatics/btr355.

Salamov, A.A., and Solovyev, V.V. (2000). Ab initio gene finding in Drosophila genomic DNA. Genome Res. *10*, 516–522. https://doi.org/10.1101/gr.10.4.516.

Seplyarskiy, V.B., Logacheva, M.D., Penin, A.A., Baranova, M.A., Leushkin, E.V., Demidenko, N.V., Klepikova, A.V., Kondrashov, F.A., Kondrashov, A.S., and James, T.Y. (2014). Crossing-over in a hypervariable species preferentially occurs in regions of high local similarity. Mol. Biol. Evol. *31*, 3016–3025. https://doi.org/10.1093/molbev/msu242.

Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31. https://doi.org/10.1186/1471-2105-6-31.

Sohlberg, E., Bomberg, M., Miettinen, H., Nyyssänen, M., Salavirta, H., Vikman, M., and Itävaara, M. (2015). Revealing the unexplored fungal communities in deep groundwater of crystalline bedrock fracture zones in Olkiluoto, Finland. Front. Microbiol. *6*, 573. https://doi.org/10.3389/fmicb.2015.00573.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Stammers, D.K., Ren, J., Leslie, K., Nichols, C.E., Lamb, H.K., Cocklin, S., Dodds, A., and Hawkins, A.R. (2001). The structure of the negative transcriptional regulator NmrA reveals a structural superfamily which includes the short-chain dehydrogenase/reductases. EMBO J. *20*, 6619–6626. https://doi.org/10.1093/emboj/20.23.6619.

Steen, A.D., Jørgensen, B.B., and Lomstein, B.A. (2013). Abiotic racemization kinetics of amino acids in marine sediments. PLoS One *8*, e71648. https://doi.org/10.1371/journal.pone.0071648.

Stival Sena, J., Giguère, I., Boyle, B., Rigault, P., Birol, I., Zuccolo, A., Ritland, K., Ritland, C., Bohlmann, J., Jones, S., et al. (2014). Evolution of gene structure in the conifer *Picea glauca*: a comparative analysis of the impact of intron size. BMC Plant Biol. *14*, 95. https://doi.org/10.1186/1471-2229-14-95.

Thatcher, D.R., and Cain, R.B. (1975). Metabolism of aromatic compounds by fungi: kinetic properties and mechanism of 3-carboxy-cis,cis-muconate cyclase from *Aspergillus niger*. Eur. J. Biochem. *56*, 193–204. https://doi.org/10.1111/j.1432-1033.1975.tb02222.x.

Tovar-Herrera, O.E., Martha-Paz, A.M., Pérez-LLano, Y., Aranda, E., Tacoronte-Morales, J.E., Pedroso-Cabrera, M.T., Arévalo-Niño, K., Folch-Mallol, J.L., and Batista-García, R.A. (2018). *Schizophyllum commune*: an unexploited source for lignocellulose degrading enzymes. Microbiologyopen *7*, e00637. https://doi.org/10.1002/mbo3.637.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111. https://doi.org/10.1093/bioinformatics/btp120.

Trembath-Reichert, E., Morono, Y., Ijiri, A., Hoshino, T., Dawson, K.S., Inagaki, F., and Orphan, V.J. (2017). Methyl-compound use and slow growth characterize microbial life in 2-km-deep subseafloor coal and shale beds. Proc. Natl. Acad. Sci. U S A *114*, E9206–E9215. https://doi.org/10.1073/pnas.1707525114.

Vigil-Stenman, T., Ininbergs, K., Bergman, B., and Ekman, M. (2017). High abundance and expression of transposases in bacteria from the Baltic Sea. ISME J. *11*, 2611–2623. https://doi.org/10.1038/ismej.2017.114.

Vos, M., and Didelot, X. (2009). A comparison of homologous recombination rates in bacteria and archaea. ISME J. *3*, 199–208. https://doi.org/10.1038/ismej.2008.93.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. *40*, e49. https://doi.org/10.1093/nar/gkr1293.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. *35*, 543–548. https://doi.org/10.1093/molbev/msx319.

Wormer, L., Hoshino, T., Bowles, M.W., Viehweger, B., Adhikari, R.R., Xiao, N., Uramoto, G.i., Konneke, M., Lazar, C.S., Morono, Y., et al. (2019). Microbial dormancy in the marine subsurface: global endospore abundance and response to burial. Sci.Adv. *5*, eaav1024. https://doi.org/10.1126/sciadv.aav1024.

Xu, Y.C., Niu, X.M., Li, X.X., He, W., Chen, J.F., Zou, Y.P., Wu, Q., Zhang, Y.E., Busch, W., and Guo, Y.L. (2019). Adaptation and phenotypic diversification in Arabidopsis through loss-of-function mutations in protein-coding genes. Plant Cell *31*, 1012–1025. https://doi.org/10.1105/tpc.18.00791.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide

complex trait analysis. Am. J. Hum. Genet. *88*, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

Yang, Z. (2007). Paml 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.

Zain Ul Arifeen, M., Chu, C., Yang, X., Liu, J., Huang, X., Ma, Y., Liu, X., Xue, Y., and Liu, C. (2021a). The anaerobic survival mechanism of *Schizophyllum commune* 20R-7-F01, isolated from deep sediment 2 km below the seafloor. Environ. Microbiol. *23*, 1174–1185. https://doi.org/10.1111/1462-2920.15332.

Zain Ul Arifeen, M., Ma, Z.J., Wu, S., Liu, J.Z., Xue, Y.R., and Liu, C.H. (2021b). Effect of oxygen concentrations and branched-chain amino acids on the growth and development of sub-seafloor fungus, *Schizophyllum commune* 20R-7-F01. Environ. Microbiol. *23*, 6940–6952. https://doi.org/10.1111/1462-2920.15738.

Zain Ul Arifeen, M., Yang, X.Y., Li, F.F., Xue, Y.R., Gong, P.X., and Liu, C.H. (2020). Growth behaviors of deepsubseafloor *Schizophyllum commune* in response to variousenvironmental conditions. Acta Microbiol. Sin. *60*, 1882–1892.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Total DNA and RNA of *Schiophyllum commune*	This study	NA
Deposited data		
original code	This study	https://github.com/liuxuan-425lab/genome_code
Resequencing data	This study	GeneBank: PRJNA738972
RNA sequence data	This study	GeneBank: PRJNA543698
Genome assembly data	This study	GeneBank: PRJNA544166
Other genome data	Baranova et al. (2015)	GeneBank: PRJNA236351
Other resequencing data	Seplyarskiy et al. (2014)	GeneBank: PRJNA234274
Software and algorithms		
BUSCO	Waterhouse et al. (2018)	https://busco.ezlab.org/
HGAP3	Chin et al. (2013)	http://www.pacb.com/devnet/
Quast	Gurevich et al. (2013)	http://bioinf.spbau.ru/quast
TopHat	Trapnell et al. (2009)	http://ccb.jhu.edu/software/tophat/index.shtml
Cufflinks	Roberts et al. (2011)	http://cole-trapnell-lab.github.io/cufflinks/
RNAmmer	Lagesen et al. (2007)	http://www.cbs.dtu.dk/cgi-bin/nph-sw_request?rnammer
tRNAscan-SE	Lowe and Eddy (1997)	https://github.com/UCSC-LoweLab/tRNAscan-SE
Fgenesh	Salamov and Solovyev (2000)	http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind
EVidenceModeler	Haas et al. (2008)	https://evidencemodeler.github.io/
OrthoFinder	Emms and Kelly (2015)	https://github.com/davidemms/OrthoFinder
Orthomcl	Li et al. (2003)	https://orthomcl.org/orthomcl/app/downloads/software/
MAFFT	Katoh and Standley (2013)	https://github.com/GSLBiotech/mafft
Gblocks	Castresana (2000)	https://github.com/atmaivancevic/Gblocks
RAxML	Stamatakis (2014)	http://evomics.org/learning/phylogenetics/raxml/
PAML	Yang (2007)	http://abacus.gene.ucl.ac.uk/software/paml.html
CAFE	DeBie et al. (2006)	https://github.com/hahnlab/CAFE
MUMmer	Delcher et al. (2002)	http://mummer.sourceforge.net/
BWA-MEM	Li and Durbin (2009)	https://github.com/lh3/bwa
GATK	De Summa et al. (2017)	https://gatk.broadinstitute.org/hc/en-us
Annovar	Wang et al. (2010)	http://www.openbioinformatics.org/annovar/annovar_download_form.php
PLINK	Purcell et al. (2007)	http://zzz.bwh.harvard.edu/plink/
SNPhylo	Lee et al. (2014)	https://github.com/thlee/SNPhylo
GCTA	Yang et al. (2011)	https://yanglab.westlake.edu.cn/software/gcta/
VCFtools	Danecek et al. (2011)	http://vcftools.sourceforge.net/
ClonalFrameML	Didelot and Wilson (2015)	https://github.com/xavierdidelot/ClonalFrameML
MCScanX	Wang et al. (2012)	https://github.com/wyp1125/MCScanX
bcftools	Danecek et al. (2021)	http://github.com/samtools/bcftools
Poppr	Kamvar et al. (2014)	https://github.com/grunwaldlab/poppr

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to the Lead Contact, Changhong Liu (chliu@nju.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The genome assembly data, RNA sequencing data, and resequencing data have been deposited at GenBank and SRA. Accession numbers are listed in the key resources table. The 5 terrestrial samples re-sequencing data were downloaded from the NCBI under the accession number SRP184549, SRP184550, SRP184481, SRP183624, SRP184554.
- All original code has been deposited at Github repository and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Fungal strains and culture conditions

A total of nine *S. commune* strains were used for sequencing in this study. Six strains, including three dikaryotic (6R-2-ZF01, 15R-5-ZF01, 24R-3-ZF01) and three corresponding monokaryotic (6R-2-F01, 20R-7-01, 24R-3-F01) strains, isolated from 1496 ~1993 mbsf sediments during IODP Expedition 337 (Liu et al., 2017). Different subseafloor strains come from different cores, different distances, and different geological ages; for example, subseafloor strain 6R-2-F01 was isolated from sediment core Unit 2 at a depth of 1,496 mbsf with an estimated geological sediment age of Miocene, while strains 15R-5-F01/20R-7-F01/24R-3-F01 were derived from sediment core Unit 3 at depths of 1,924 mbsf, 1966 mbsf and 1,993mbsf; respectively, having geological age of early to late Miocene (Gross et al., 2015). Two terrestrial strains CFCC 7252 and CFCC 86625 were purchased from China Forestry Culture Collection Center, which were isolated from Populus wood in Songshan, Beijing (altitude:700m) and Jurong, Jiangsu (altitude: 300m) of China, respectively, and the other five terrestrial strains (225DK, 227DK, MF, Hom2-8, 207) were obtained from NCBI and JGI database. Moreover, one strain MCCC 3A00233, collected from marine sediment of the Atlantic Ocean, was purchased from Third Institute of Oceanography, State Oceanic Administration, People's Republic of China. Details of habitat of marine and terrestrial strains are shown in Table S12. Details of the habitat and culture conditions of *S. commune* strains have been described previously (Zain Ul Arifeen et al., 2020). For DNA and RNA isolation, mycelium of *S. commune* was grown in 250 mL conical flask containing 150 mL PD (200 g/L potato and 20 g/L glucose) and incubated at 30°C and 200 rpm for 3–5 days.

METHOD DETAILS

De novo genome sequencing and assembly

Genomic DNA of *S. commune* 20R-7-F01 (CGMCC 11604) was extracted using a CTAB-PEG method (Huang et al., 2018). An Illumina paired-end DNA library of 20 kb insert size was generated using the TruSeq DNA Sample Preparation Kit (Illumina, USA) and the Template Prep Kit (Pacific Biosciences, USA). Genome sequencing was performed using the PacBio RSII long-read sequencing platform at BGI Genomic (Shenzhen, China). A total of 4.58 Gb of raw PacBio data (~113.1 × genome coverage) was obtained, of which 1.47 Gb was removed due to low complexity (length <100 bp), low quality (score <0.8), and adapter and duplication contamination (subreads length <1000 bp; score <0.8). The resultant clean PacBio data were assembled using SMRT Analysis (v2.3.0) HGAP3 workflow to construct scaffold-level contigs (Chin et al., 2013). Integrity of the genome assembly was assessed using BUSCO software (version v3.0.2) (Waterhouse et al., 2018). Quast software was used to evaluate the assembly quality of *S. commune* 20R-7-F01 genome (Gurevich et al., 2013).

Transcriptome sequencing and assembly

Total RNA was extracted from mycelium of *S. commune* 20R-7-F01 using an RNAprep Pure Plant Kit (TIANGEN, Cat. #DP441, China) following the manufacturer's instructions. Magnetic oligo (DT) beads were used to enrich polyA mRNA tails of three independent RNA samples. The cDNA libraries were sequenced using the Illumina HiSeq™ 2000 platform by Personal Biotechnology Company (Shanghai, China). After trimming of low quality reads (Q<20) and adapter contamination, ~72.5 Gb high-quality transcriptome data reads were obtained, which was mapped to genome assembly of strain 20R-7-F01 using TopHat (Trapnell et al., 2009), and performed for gene prediction using Cufflinks (Roberts et al., 2011).

Gene annotation

The repetitive DNA in *S. commune* 20R-7-F01 genome was identified using RepeatModeler and RepeatMasker (http://www.repeatmasker.org/). Mitochondrial sequences were identified by BLASTN search using published mitochondrial genome of *S. commune* (GenBank ID: AF402141.1) (Forget et al., 2002) and were excluded when annotating genes in the nuclear genome. The rRNA, tRNA, and ncRNA were predicted with RNAmmer (Lagesen et al., 2007), tRNAscan-SE (Lowe and Eddy, 1997), and Rfam (Gardner et al., 2009), respectively. Open reading frames were predicted with Fgenesh (Salamov and Solovyev, 2000), Exonerate (Slater and Birney, 2005), Cufflinks, EVidenceModeler (Haas et al., 2008), based on both assembly and RNA-seq data. Gene functions were annotated using NCBI non-redundant protein database (Nr), Eukaryotic Clusters of Orthologous Groups (KOG), SwissProt, Gene Ontology (GO), KEGG, TIGRFAM, and PFAM databases (Li et al., 2021). Mitochondrial genes were predicted according to the annotation of *S. commune* (GenBank ID: AF402141.1) after aligning the local mitochondrial sequences to the published mitochondrial genome using BLASTN.

Gene family identification and phylogenetic evolution analysis

A total of 10 fungal species (i.e., 11 strains) assigned to Basidiomycota including *S. commune* (two strains 20R-7-F01 and H4-8), *Coprinopsis cinerea*, *Laccaria bicolor*, Postia placenta, Phanerochaete chrysosporium, *Cryptococcus neoformans* and *Ustilago maydis*, and Ascomycota such as *Saccharomyces cerevisiae*, *Aspergillus nidulans* and *Neurospora crassa* were used to perform gene family identification. Protein sequences of these species except *S. commune* 20R-7-F01 were downloaded from NCBI and JGI databases, and analyzed by OrthoFinder with default parameters to find orthologous genes (Emms and Kelly, 2015). The orthologous genes between *S. commune* 20R-7-F01 and *S. commune* H4-8 were analyzed using Orthomcl software (Li et al., 2003). The single-copy genes of the 11 analyzed strains were aligned using MAFFT to create a super-gene for each species (Katoh and Standley, 2013). After Gblocks alignment optimization (Castresana, 2000), the conserved blocks of super-genes were used for phylogenetic tree construction using RAxML (Stamatakis, 2014). The divergence time was estimated using the "Independent rates (clock = 2)" model and the "HKY85" model of MCMCTREE program in the PAML package (Yang, 2007), based on the fossil calibration of the most recent common ancestors (MRCA) of *N. crassa* and *A. nidulans* (divergence at 290-470 Mya), and *L. bicolor* and *C. cinerea* (divergence at 120-180 Mya) (Kumar et al., 2017). The MCMC process was run for 1,000,000 iterations, after a burn-in of 2,000 iterations, and the convergence was evaluated by ESS, which was calculated with Tracer software. The mtDNA protein sequences of *A. nidulans*, *C. cinerea*, *C. neoformans*, *L. bicolor*, *N. crassa*, *S. cerevisiae*, *S. commune* and *U. maydis* were downloaded from GenBank (accession numbers JQ435097.1, AACS02000068.1, CP003834.1, MK697670.1, KC683708.1, KP263414.1, AF402141.1, and DQ157700.1, respectively) and also analyzed by OrthoFinder with default parameters to find orthologous genes. Finally, three single copy homologous genes (cytochrome c oxidase subunit 1, cytochrome c oxidase subunit 2, cytochrome c oxidase subunit 3) were identified. The cytochrome c oxidase gene families, as molecular markers, always widely used in molecular evolutionary and phylogenetic relationship studies within and between fungus (Belloch et al., 2000; Okane et al., 2021). Additionally, the same method was used to construct the mitochondrial phylogenetic tree. The expansion and contraction of orthologous gene family were calculated with CAFE and defined as "significantly expanded or contracted" if p-value $\leq$0.01 (De Bie et al., 2006). The positive selected genes (PSGs) between *S. commune* 20R-7-F01 and *S. commune* H4-8 were detected using the MCScanX (i.e., add_ka_and_ks_to_collinearity.pl) (Wang et al., 2012). In addition, synteny analysis of strains 20R-7-F01 and H4-8 was performed using MUMmer sequence alignment package (Delcher et al., 2002). Conserved syntenic blocks between the two strains were identified using the MCScanX package with default parameters based on a minimum requirement of five collinear orthologous genes

Genome resequencing, alignment and variant calling

The genome DNA of *S. commune* strains (Table S12) was extracted and fragmented to generate approximately 300 bp library insert size. The paired-end reads were sequenced on an Illumina HiSeq 2500 platform at BGI Genomic (Shenzhen, China). Raw data were filtered by removing the adapters and low-quality reads (Q<20). All the filtered re-sequencing reads were mapped to the reference genome of *S. commune* 20R-7-F01 using BWA-MEM with default parameters (Li and Durbin, 2009). After removing reads with low mapping quality (MQ < 40), both paired-end and single-end mapped reads were used for SNP detection across entire samples of *S. commune* strains using the GATK toolkit (version 4.0.8.1). SNPs and small InDels (insertion/deletions, 1–50 bp) were called using GATK software (DeSumma et al., 2017), filtered using SelectVariants according to the standard hard filtering criteria: "QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0" (DePristo et al., 2011), and assigned to specific genomic regions and genes using Annovar based on 20R-7-F01 genome annotations (Wang et al., 2010). Additionally, bcftools (Danecek et al., 2021) was used to calculate homozygous and heterozygous sites with "snp.vcf" file as input, and estimated the heterozygosity as "number of heterozygous SNPs"/"genome size". Loss-of-Function (LoF) variants were identified including mutations of stopgain, nonsynonymous SNV, and frameshift (Xu et al., 2019).

Phylogenetic and population structure analyses

A subset of 2,704,270 SNPs obtained from resequencing of 14 *S commune* strains was used for phylogenetic and population structure analysis. Vcf files were converted to hapmap format with custom perl scripts and to PLINK format file by PLINK (Purcell et al., 2007). Under the p-distances model, a neighbor-joining tree of all samples was constructed with SNPhylo software (Lee et al., 2014). GCTA was used to carry out principal component analysis (PCA), and the first 2 eigenvectors were extracted (Yang et al., 2011).

VCFtools (Danecek et al., 2011) was used to estimate the *Fst*, nucleotide diversity and Tajima's D for the whole genome of both subseafloor and terrestrial populations with non-overlapping 10 kb sliding windows. The top 10% of *Fst* with top/bottom 10% of Tajima's D, defined by a sliding window, were selected as candidate selective sweep regions. We further integrated the adjacent selective sweep regions and annotated the corresponding genes across genome of strain 20R-7-F01.

Detection of homologous recombination

The contributions of mutations and recombination to the genomic diversity in the concatenated core genome alignment, the number of recombination events (imports) per genome, and the positions of recombination hot spots, was investigated using ClonalFrameML (Didelot and Wilson, 2015). Nucleotides unaffected by recombination are referred to as unimported and nucleotides subject to recombination are referred to as imported. ClonalFrameML provides the relative rate of recombination to mutation (R/Theta), the mean length of recombined DNA (Delta), and the mean divergence of imported DNA (Nu). These results were used to calculate the relative contribution of recombination versus mutation to the overall genomic diversity (r/m), using the following formula r/m = (R/Theta) ×Delta × Nu. ClonalFrameML was performed in two separate runs containing a core genome alignment that contained (1) only the sediment genomes, and (2) only the terrestrial strains. The resulting r/m values from these two groups (presented in Table 2) were then used to interpret the relative importance of mutations compared to recombination, in the separate groups (e.g., terrestrial strains versus subseafloor strains). Additionally, the same method was used to calculate the homologous recombination of small-scale geographical terrestrial strains (5 strains from Moscow and 4 strains from Florida) (Baranova et al., 2015), which could avoid the impact of different geographical locations on the analysis results. We acknowledge that because the dataset contains genomes covering the diversity of a single fungal genus (*Schizophyllum*), the only detectable recombination most events are from donors from the species under study, so that the main source of recombination is not external. Moreover, we also calculated the index of association for subseafloor population and Russia population by using a function in R package "Poppr" (Kamvar et al., 2014, 2015). The function is named "samp.ia", which can randomly sample sites to calculate the index of association. The specific parameter settings are as follows: the number of snps to be used to calculate

standardized index of association was set 100L; the number of times to perform the calculation was set 100L.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification and statistical analysis used in the genome sequencing and assembly, genome quality assessment, evolutionary analysis and comparative genome analysis can be found in the relevant sections of the method details.