



OPEN Misconception of *Schizophyllum commune* strain 20R-7-F01 origin from subseafloor sediments over 20 million years old

Sergiy M. Boiko

Hole C0020A during Ocean Drilling Expedition 337 is the deepest hole in scientific ocean drilling (depth of 2466 m below the seafloor). The presence of microbial and fungal communities serves as firm evidence for life within sediments. The isolated from the core and cultivable in the laboratory strain *S. commune* 20R-7-F01, which has existed in a deep subseafloor environment for over 20 million years, is being considered a high-quality genome model for studying its evolution and environmental adaptation mechanism. We questioned the origin of the strain from the sediments using high mutagenic simple sequence repeats of DNA. The full genome analysis of mono-, di-, tri-, and tetra-motifs of DNA revealed no regularities in the quantitative distribution of motifs in different *S. commune* genomes. At of common trinucleotide motifs loci, strain 20R-7-F01 has the highest percentage of similarity (48.8%) among East Asian strains, which indicates an intensive genetic exchange. According to a multidimensional scaling of 1938 common simple sequence repeats of DNA loci, no signs were found that would indicate the spatial and temporal isolation of the 20R-7-F01 strain. The extreme insufficient of water, and oxygen, high temperature and pressure at the level of 2 km below the ocean floor, and the tetrapolar mating system make it impossible growth of mycelium with different nuclear status and formation of basidiocarps. The general obtained data confirm the terrestrial origin of *S. commune* strain 20R-7-F01 and the territory of the Far East of the Russian Federation (approximately Khabarovsk, Primorsky Krai) is the probable place of origin.

Keywords Samples from sedimentary cores, *Schizophyllum commune* 20R-7-F01, Simple sequence repeats of DNA, Terrestrial origin.

Integrated Ocean Drilling Program Expedition 337 was the first expedition dedicated to subseafloor microbiology that used riser drilling technology. The site examined during the expedition is located off the Shimokita Peninsula, Japan, at a water depth of 1180 m. Seismic profiles suggested the presence of deep, coal-bearing horizons at ~2 km below the seafloor. One of the scientific objectives during Expedition 337 was to explore the distribution of subseafloor life at the greatest depths that have ever been sampled by scientific ocean drilling. Hole C0020A is the deepest hole in the history of scientific ocean drilling, and in September 2012 drilling was terminated at a total drilling depth of 2466 m below seafloor (mbsf)^{1,2}. Microbial communities in deepsea sediment found 1.5–2.5 km below the Pacific Ocean's seafloor at temperatures of 40–60 °C, show that life exists in these extreme conditions³. Isotopic compositions of methane, carbon dioxide, and gas compositions testified to microbial methanogenesis. Peak concentrations of indigenous microbial cells occurred in lignite layers and ranged from <10 to ~10⁴ cells cm⁻³. According to the authors, this indicates that terrigenous sediments retain indigenous community members millions of years after burial. In addition, Liu et al., explored the distribution, diversity, and origins of fungal populations in subseafloor sediments which were collected during Expedition 337⁴. Fungi were isolated from 34 of 47 deep sediment samples collected from 1289 to 2457 mbsf. A total of 69 fungal cultures representing *Ascomycota* (14 genera, 23 species) and *Basidiomycota* (4 genera, 4 species) were obtained. *Penicillium* and *Aspergillus* were dominant genera within the *Ascomycetes*, and genera *Schizophyllum*, *Irpex*, *Bjerkandera*, and *Termitomyces* were identified for *Basidiomycota*. The cultivable fungal populations were not abundant as compared to their terrestrial habitats, however exhibited growth in the laboratory, ~20 million years after their likely burial. The 27 fungal species showed no or very little correlation to the depth and age.

Department of Phytoecology, Institute for Evolutionary Ecology, National Academy of Sciences of Ukraine, 37 Lebedeva Str, Kyiv 03143, Ukraine. email: bsmbio@gmail.com

Authors pointed out that isolated fungi are indigenous, originating from past terrigenous environments, which have persisted, possibly as spores. They proposed uses as a model the high-quality genome of strain *S. commune* 20R-7-F01 subsisting deep subseafloor environment over 20 Mya for the study of its evolution, and environmental adaptation mechanism by comparison with terrestrial strains^{5,6}. In the presented papers, the authors make some controversial statements that prompted us to conduct a deeper analysis of the geographical origin of the unique strain 20R-7-F01 and new strains of *S. commune* for the National Center for Biotechnology Information (NCBI) database.

We used simple sequence repeats of DNA (SSR) or microsatellites for full-genome analysis of strains *S. commune*. SSRs are preferred markers for genetic studies due to their high polymorphism, co-dominant inheritance, high reproducibility, and full genome coverage. The length of the repeated motif in microsatellites is between one and six nucleotides and microsatellite lengths range from a few to thousands of repeats⁷. SSRs have an order of magnitude higher mutation rate, ranging from 10^{-6} to 10^{-2} events per locus per generation, than point mutations^{8,9}. The highly polymorphic SSRs are suitable for detecting the hybridization between closely related species, studying the gene flow, and recent population structure. They are under selection pressure in genomes, present in intergenic and non-coding regions, and a small proportion of SSRs occur within exons. SSRs within the open reading frame (ORF) are less polymorphic, indicating different rates and evolutionary trajectories of their development in different locations in the genome¹⁰. Most repetitive types (mono-, di-, and tetramers) of motifs are usually more common in introns and intergenic regions compared to exons, while trimers and hexamers are more abundant in exons. Microsatellites outside of transcribed regions are used to identify different genotypes, within ORFs or in promoter regions to study differences between populations or closely related species.

The distribution of SSRs in the genomes of the fungus *S. commune* has a certain dependence on the origin of the culture. The results of previous studies show that the GCT motif promotes the divergence of the population from the Russian Federation, while the CTC, GAG and GGA motifs of the population from the United States¹¹. East Asian strains from China and South Korea occupy a separate position, mainly due to the CGA motif. Differences in the quantitative features of the motifs between populations showed the evolutionary transformation of the fungal genomes under the influence of environmental conditions. This fact made it possible to identify specific SSR DNA loci and develop unique primers¹², for them, which allow us to determine the geographical origin of the sample. Using this tactic, three new strains *S. commune*, Loenen D, Tattone D, and 20R-7-F01 in the NCBI database, were used to establish their probable geographic origin. The only information available on the parental strains indicates that *S. commune* strains Tattone and Loenen were collected in Tattone (Corsica, France) and Loenen aan de Vecht (The Netherlands), respectively¹³. The *S. commune* 20R-7-F01 strain is unique because it was isolated from deep-sea sediment cores over 20 million years old and located 2 km below the ocean floor⁴.

Based on the information above, the study aimed to identify the likely geographic origins of new *S. commune* strains, including the unique 20R-7-F01 strain found in coal beds over 20 million years old, and determine its differences from terrestrial strains.

Results

In silico PCR analysis at 34 SSR DNA loci was performed to identify the origins of the *S. commune* strains 20R-7-F01, Loenen D, and Tattone D¹². The analysis of the available PCR products of the experimental strains by the principal component method allowed us to identify the probable populations of their origin (Fig. 1).

It has been clearly established that the strains Loenen_D and Tattone_D belong to the population originating from the territory of the Russian Federation. It is still difficult to establish which part of Europe or Asia they come from, as there is a lack of geographical coordinates for all strains of the Ru population. However, as the data set grows, this issue will be resolved quickly enough.

The most interesting strain is *S. commune* 20R-7-F01, because according to the description, it is more than 20 million years old⁴ and is expected to occupy a separate position. Contrary to expectations, in the global principal component space, it was in the group of strains of the East Asian population (Fig. 1). On the one hand, this is natural, because the strain was isolated from sedimentary soils of the Pacific Ocean floor near Japan, but on the other hand, the depth of the layers ~2000 m below the bottom and another 1180 m of water oceanic column³ should be an insurmountable barrier to the exchange of genetic material. The findings raised a new question: how did a strain isolated for 20 million years come to be close to modern terrestrial specimens of East Asian origin? Perhaps the 34 SSR DNA loci used are not sufficient to detect the difference? To find the answer, a complete SSR analysis of different *S. commune* genomes was applied. In addition to the *S. commune* strain 20R-7-F01, three strains were selected from the Russian Federation population (Ru) - X-9_S78, X-21_S81, X-69_S85; three from the United States population (USA) - H4-8, 14-25_S64, 14-90_S72; and three from the East Asia population (EA) - Japan JCM22674, Chine MG53, South Korea Incheon IUM1114-SS01. Strain USA H4-8 was specifically included as the one used to establish age differences⁵.

At the first stage, we applied redundant microsatellite selection conditions, which provide an idea of their number in different genomes. Short nucleotide repeats of DNA are the most dynamic in rearrangements and the “youngest” in origin¹⁴. The results show a certain trend for each of the population groups. Thus, representatives of the USA population have the smallest number (350 311–364 262) of microsatellites in their genomes. In representatives of the Ru population, this figure is somewhat higher (366 121–377 654) and it increases significantly in strains of East Asian origin (364 916–482 203) (Fig. 2). Of note is the strain Japan_JCM22674 from the island territory, which is between the USA and Ru populations by this indicator. The strain of interest, 20R-7-F01, has high values of the number of simple DNA repeats (401 055) and is in the general trend of strains of the EA population. This is an evidence of intensive genome rearrangements of *S. commune* strain 20R-7-F01, which are possible under conditions of its physiological activity.

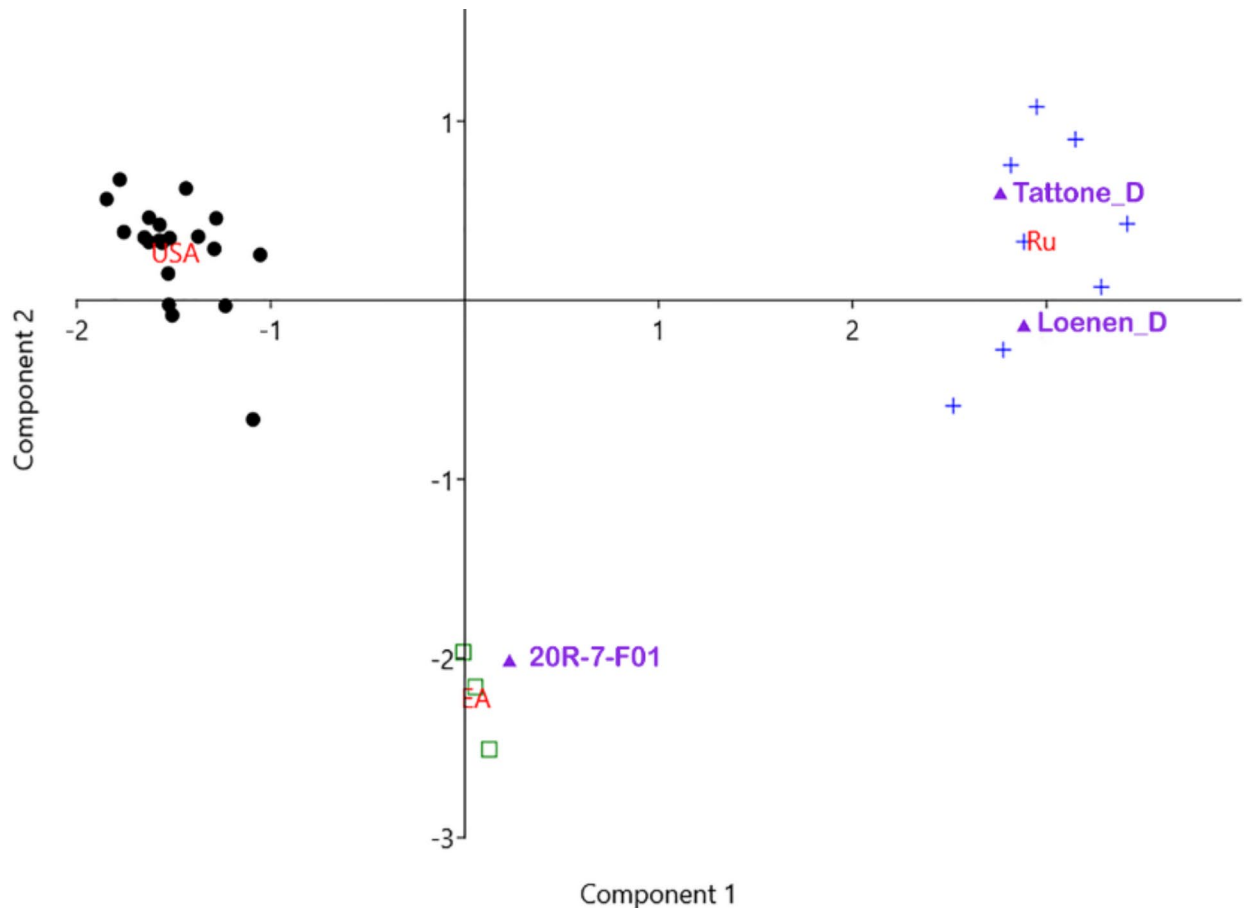


Fig. 1. The global position of new *Schizophyllum commune* strains (Loenen_D, Tattone_D, and 20R-7-F01) in the principal component space by 34 SSR DNA loci.

In the second stage, the number of the most common mono-, di-, tri-, and tetra-motifs in fungal genomes was calculated separately to determine the difference in their number. Hypothetically, a significant age difference and strains belonging to different populations can significantly affect this indicator. The principal component analysis of the data revealed no regularities in the quantitative distribution of motifs in different genomes (Supplementary Fig. 1). The separation of the Chine MG53 strain from the others is associated with a slight quantitative advantage for some types of motifs, as we saw in the previous case. All other strains were randomly distributed in the component space.

The analysis of the loci distribution by different motifs allowed us to establish a certain peculiarity. The more complex the motif is, the smaller its number is observed in the genome. An increase in the number of motif repeats leads to a significant decrease in their concentration in genomes (Supplementary Fig. 2). However, for loci consisting of trinucleotide motifs with a minimum repeat rate of 6–7, the level of decrease becomes slower than for loci consisting of dinucleotide motifs. This trend continues in the future.

Under the third stage, we selected SSR DNA loci in the *S. commune* genomes that were well established both within and outside the population. As expected, simpler loci with single nucleotide motifs dominated in number (Supplementary Fig. 3). The general pattern of their number was similar to the previous results, but this time the USA_H4-8 strain had the lowest number (4059). The highest number of microsatellites was found in the strain Chine_MG53 (6174), and the strain 20R-7-F01 from deep-bottom sediments showed average values (4490). These figures do not indicate the possible isolation of *S. commune* strain 20R-7-F01 for 20 million years.

The general indicators obtained do not allow us to draw comprehensive conclusions, but help to form a vector for further research. We developed primer pairs for SSR DNA loci of all types of motifs for all experimental strains of the EA population. For the USA population, the USA_H4-8 strain was used, and for the Ru population, the Russia X-9 S78 strain. This is due to the expediency of an in-depth analysis of the East Asian group together with strain 20R-7-F01. *In silico* PCR was performed for each *S. commune* genome to determine their relationship (Table 1).

According to the results, the specific SSR loci of strain 20R-7-F01 are mostly common to the East Asian population (Table 2). The maximum number of them forming amplicons (>20%) falls on loci with complex motifs (tri-, tetra-), and the most related are the strains Chine MG53 and South Korea Incheon IUM1114-SS01. The percentage of common loci from mono- and dimotifs is lower (16.1–18.5%), and the Japan JCM22674

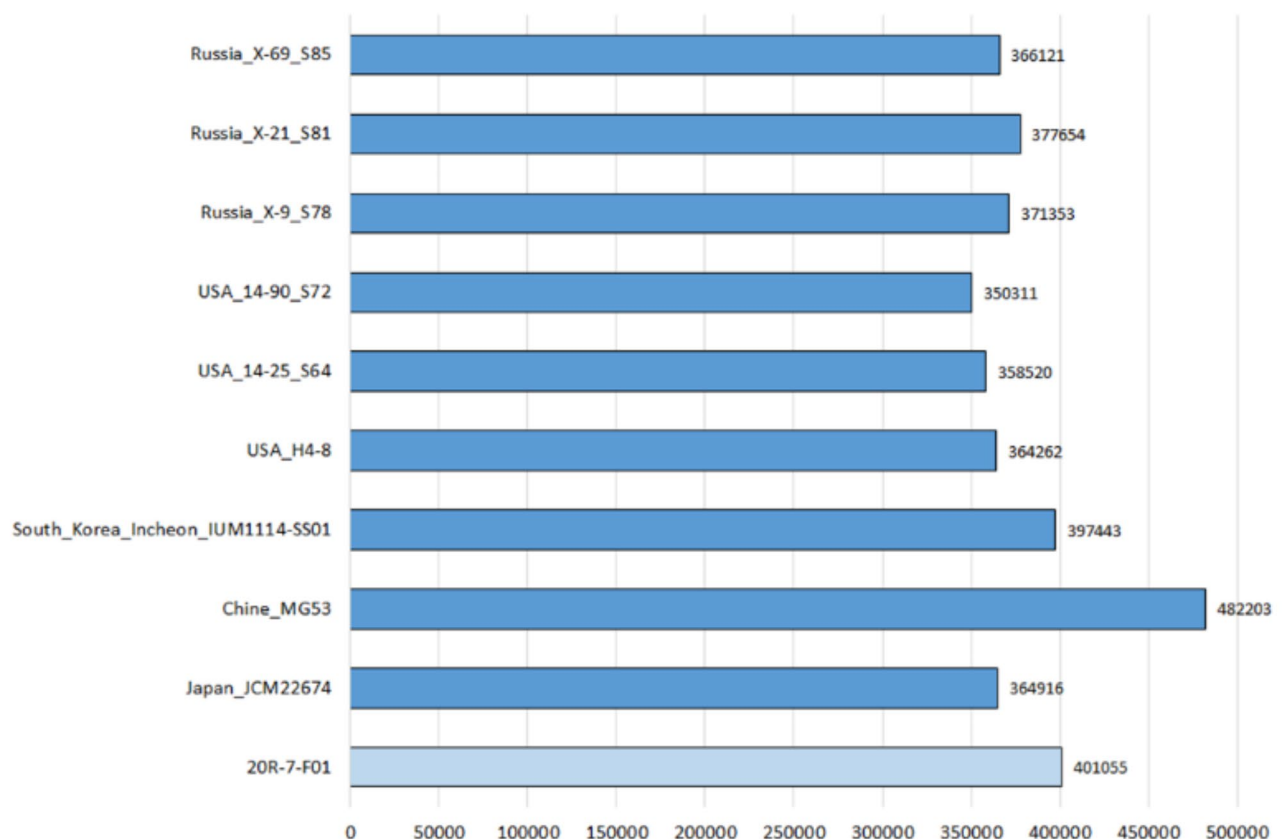


Fig. 2. The number of single nucleotide motifs with fourfold minimum repetition in different genomes of *Schizophyllum commune*.

Strain	Location	GenBank assembly	Genome size
20R-7-F01	?	GCA_024373065.1	40 791 524
Loenen D	?	GCA_023508725.1	35 882 427
Tattone D	?	GCA_023508785.1	36 458 910
JCM 22,674	EA (Japan)	GCA_001599475.1	39 005 850
MG53	EA (China)	GCA_003313685.1	51 157 739
IUM1114-SS01	EA (South Korea)	GCA_014900015.1	40 957 804
H4-8	USA	GCA_000143185.2	38 670 379
14-25_S64	USA	GCA_019143365.1	37 964 127
14-90_S72	USA	GCA_019143045.1	37 225 418
X-9_S78	Russia	GCA_019142955.1	38 585 916
X-21_S81	Russia	GCA_019143075.1	39 038 653
X-69_S85	Russia	GCA_019142975.1	38 217 150

Table 1. List of *S. commune* strains genomes their location and genome sizes in the NCBI database.

strain, which is geographically closest, showed the lowest level of similarity (11–12%). The strains of the Ru population showed a higher percentage of SSR loci similarity than the strains of the USA population. The highest percentage was observed for loci consisting of trimotifs (16–17.4%) and dimotifs (13.9–14.4%). The loci with pentanucleotide motifs, due to their complexity and small number in different genomes, turned out to be of little information, both for strain 20R-7-F01 and for all others. For the strains of the USA population, the largest number of loci shared with strain 20R-7-F01 ranges from 5.3 to 7.8%.

It is expected that the strains of the East Asian population (Japan JCM22674, South Korea Incheon IUM1114-SS01, Chine MG53) have the highest percentage of common SSR DNA loci among them (Tables 3, 4 and 5). It should be noted that the high percentage of common SSR loci of all types of Japan JCM22674 strain with Chine MG53 and South Korea Incheon IUM1114-SS01 strains (Table 3), as well as South Korea Incheon IUM1114-

<i>Schizophyllum commune</i> strains	Motifs				
	mono- (2646*)	di- (195*)	tri- (512*)	tetra- (102*)	penta- (36*)
20R-7-F01	100%	100%	100%	100%	100%
Japan_JCM22674	11,45	11,79	21,29	20,59	5,56
Chine_MG53	16,14	16,92	25,00	22,55	2,78
South_Korea_Incheon_IUM1114-SS01	16,74	18,46	23,24	20,59	2,78
USA_H4-8	1,51	1,54	5,47	6,86	0,00
USA_14-25_S64	1,59	1,54	5,27	7,84	0,00
USA_14-90_S72	1,78	2,05	6,05	5,88	0,00
Russia_X-9_S78	8,35	13,85	16,02	8,82	0,00
Russia_X-21_S81	7,41	13,85	17,38	10,78	0,00
Russia_X-69_S85	7,90	14,36	16,41	9,80	2,78

Table 2. Percentage of amplicons generated by *Schizophyllum commune* samples to specific SSR DNA loci of genome 20R-7-F01. *Number of primer pairs developed for SSR DNA loci.

<i>Schizophyllum commune</i> strains	Motifs				
	mono- (3079*)	di- (171*)	tri- (507*)	tetra- (93*)	penta- (40*)
Japan_JCM22674	100%	100%	100%	100%	100%
20R-7-F01	13,90	9,36	18,15	15,05	5,00
Chine_MG53	19,75	25,73	23,87	21,51	15,00
South_Korea_Incheon_IUM1114-SS01	18,87	23,98	22,49	23,66	10,00
USA_H4-8	2,60	2,34	4,34	4,30	2,50
USA_14-25_S64	2,50	2,92	4,54	5,38	0,00
USA_14-90_S72	2,50	3,51	5,13	4,30	0,00
Russia_X-9_S78	8,54	10,53	12,23	10,75	2,50
Russia_X-21_S81	8,31	12,28	14,00	11,83	2,50
Russia_X-69_S85	8,25	8,77	14,79	11,83	2,50

Table 3. Percentage of amplicons generated by *Schizophyllum commune* samples to specific SSR DNA loci of genome Japan JCM22674. *Number of primer pairs developed for SSR DNA loci.

<i>Schizophyllum commune</i> strains	Motifs				
	mono- (3351*)	di- (191*)	tri- (555*)	tetra- (119*)	penta- (36*)
South_Korea_Incheon_IUM1114-SS01	100%	100%	100%	100%	100%
20R-7-F01	15,61	16,23	20,72	16,81	0,00
Chine_MG53	21,19	22,51	22,34	20,17	22,22
Japan_JCM22674	17,91	24,08	21,08	21,85	8,33
USA_H4-8	2,12	1,05	3,42	5,04	0,00
USA_14-25_S64	2,12	2,62	3,78	5,88	0,00
USA_14-90_S72	2,36	2,09	3,78	4,20	0,00
Russia_X-9_S78	9,55	7,85	12,97	13,45	2,78
Russia_X-21_S81	9,49	10,99	16,58	15,13	2,78
Russia_X-69_S85	9,22	11,52	14,23	13,45	11,11

Table 4. Percentage of amplicons generated by *Schizophyllum commune* samples to specific SSR DNA loci of genome South Korea Incheon IUM1114-SS01. *Number of primer pairs developed for SSR DNA loci.

SS01 strain with Chine MG53 and Japan JCM22674 strains (Table 4). The strain 20R-7-F01 had slightly lower similarity indices, but significantly higher than those of the Ru and USA populations.

The strains of the Ru population showed the highest percentage of common SSR DNA loci among all populations, which indicates the geographical proximity of the research samples (Supplementary Table 2). The strains of the EA population had more common loci than the strains of the USA population. The strains of the USA population showed the lowest percentage of common DNA loci compared to other populations (Supplementary Table 3). Even in the middle of the population, this indicator had moderate values. This, in

Schizophyllum commune strains	Motifs				
	mono- (4377*)	di- (253*)	tri- (658*)	tetra- (137*)	penta- (54*)
Chine_MG53	100%	100%	100%	100%	100%
20R-7-F01	10,67	8,70	21,43	12,41	3,70
South_Korea_Incheon_IUM1114-SS01	15,08	17,00	20,52	16,06	9,26
Japan_JCM22674	13,73	15,81	19,76	12,41	11,11
USA_H4-8	1,74	0,79	3,65	4,38	0,00
USA_14-25_S64	1,62	0,79	3,34	3,65	0,00
USA_14-90_S72	1,53	1,58	3,95	4,38	0,00
Russia_X-9_S78	7,33	6,72	12,46	7,30	0,00
Russia_X-21_S81	7,29	5,93	13,53	10,95	0,00
Russia_X-69_S85	7,84	6,72	14,29	10,22	0,00

Table 5. Percentage of amplicons generated by *Schizophyllum commune* samples to specific SSR DNA loci of genome Chine MG53. *Number of primer pairs developed for SSR DNA loci.

our opinion, is a sign of the considerable remoteness of the sample collection locations from each other. The percentage of common loci with strains from the Ru and EA populations was very low and almost identical. This indicates a significant barrier to gene flow posed by the Pacific Ocean. It should be noted that strain 20R-7-F01 had similar performance to modern terrestrial *S. commune* samples.

For each test strain, we calculated the percentage of common loci observed in at least one other strain of any population (Fig. 3). The strain 20R-7-F01 shows very high rates among East Asian strains. At loci consisting of trinucleotide motifs, it has the highest percentage of similarity (48.8%), and at loci consisting of tetranucleotide motifs, it is ahead of Chine MG53 and South Korea Incheon IUM1114-SS01. This indicates an intensive genetic exchange of *S. commune* strain 20R-7-F01, which contradicts the information about its isolation for 20 million years^{4,5}. The strains of the USA and Ru populations show diametrically opposite indicators, which are fully synchronized with the previous data (Supplementary Tables 2, 3).

Next, we determined how the common loci are distributed by population affiliation of the strains. The sum of the common loci for each strain was determined to be 100%. Loci consisting of di- and trinucleotide motifs were analyzed (Fig. 4). The overall picture of the two types of loci is similar, but the percentage differs significantly. For the East Asian population, the percentage of original loci with dinucleotide motifs decreases and the American loci increase from west to east, except for *S. commune* strain 20R-7-F01 (Fig. 4a). It has the lowest percentage of shared East Asian loci (76.3%) and the highest percentage of Ru population loci (60.5%). The loci shared with the American population amount to 7.9%, which is higher than that of Chine MG53 and lower than that of South Korea Incheon IUM1114-SS01 and Japan JCM22674. The USA H4-8 and Russia X-9_S78 strains are characterized by the dominance of original loci (93.9% and 87.9%, respectively) and a low proportion of loci from other continents.

Loci consisting of trinucleotide motifs show a similar general pattern with a slightly higher percentage of loci from other continents (Fig. 4b). For the strains of the East Asian group, the percentage of original loci approximately remains at the same level as the “dinucleotide” loci. The proportion of East Asian and American loci increases from west to east, except for strain 20R-7-F01. For this strain, there is a higher percentage of Ru population loci, which, by the way, has increased for all strains. In general, we can see that strain *S. commune* 20R-7-F01 falls out of the general picture of the distribution of common loci of both types by the coordinates of the origin.

The analysis of the distribution of common loci of *S. commune* strain 20R-7-F01 by the principal components method among other strains allowed us to identify the two closest cultures. For loci consisting of dinucleotide motifs, the most closely related strain was Incheon IUM1114-SS01 from South Korea (Supplementary Fig. 4a). For more complex loci consisting of trinucleotide motifs, the highest affinity was observed with the Chine MG53 strain (Supplementary Fig. 4b). The final data reflect this result (Supplementary Fig. 4c).

In total, 1938 common SSR DNA loci consisting of di- and trinucleotide motifs were identified for the experimental strains of *S. commune*. Their analysis by AMOVA revealed that genetic variation within populations accounted for 57%, and between populations for 43% (Supplementary Fig. 5). This demonstrates the effectiveness of the use of DNA microsatellites of the fungus *S. commune*, both within and at the inter-population level.

A set of 1938 common SSR DNA loci of *S. commune* strains was used for multidimensional scaling. A system was constructed using the non-metric MDS to show the relatedness of the studied *S. commune* strains (Fig. 5). Stress values and R² correlations confirm the quality of the proposed model. According to the data obtained, all experimental strains clearly fit into the proposed USA, Ru, and EA populations. The strain of *S. commune* 20R-7-F01, which is of the greatest interest, clearly outlined its affiliation with the East Asian population with a vector of approximation to the Ru population. No signs were found that would indicate the spatial and temporal isolation of this strain. The distance between Japan JCM22674 and 20R-7-F01 is almost the same as between Japan JCM22674 and Chine MG53, indicating a terrestrial origin of strain 20R-7-F01.

The construction of minimum spanning tree of non-metric MDS by the Dice and Kulczynski index (Supplementary Fig. 6) of *S. commune* strains allowed us to establish relationships between strains. Stress and correlation indicators allow us to accept the quality of the models. According to the Dice index, the combination

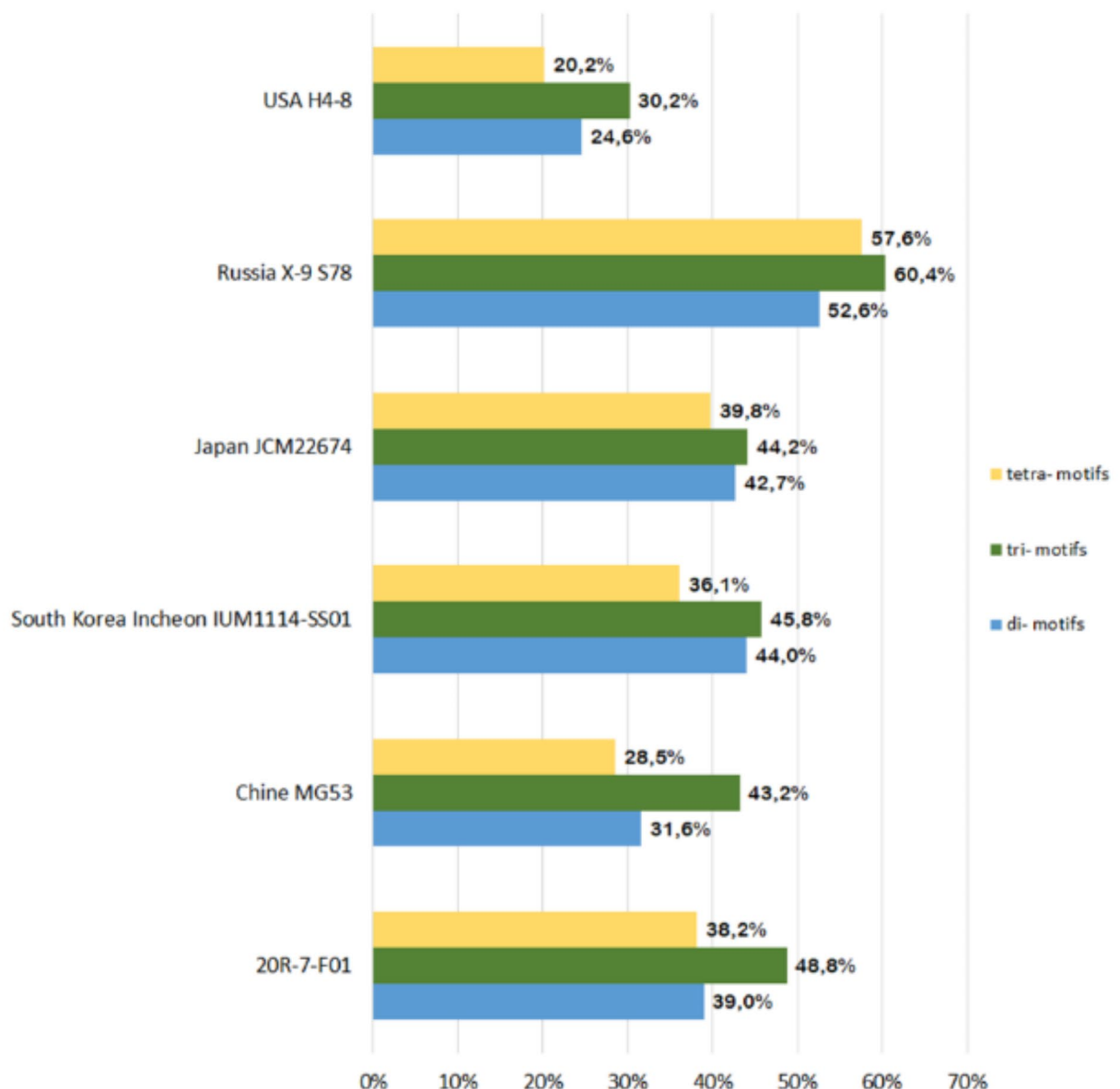


Fig. 3. Percentage of SSR DNA loci of each strain that are shared with other *S. commune* strains.

of USA, Ru, and EA populations is due to the USA H4-8, Russia X-21_S81, and 20R-7-F01 strains (Supplementary Fig. 6a). According to the Kulczynski index, the combination of USA, EA, and Ru populations is due to the USA 14-90 S72 strain, followed by almost all strains of the EA population (except Chine MG53) and the Russia X-21 S81 strain (Supplementary Fig. 6b). The Dice index emphasizes the similarity of the test strains, while the Kulczynski index gives more weight to the difference between the strains and emphasizes it. In both cases, strain 20R-7-F01 is clearly embedded in the linkage system and is a significant link in the system linking EA and Ru populations. This again proves its modern origin and indicates a commensurate rate of evolutionary rearrangements of the genome at the level of other terrestrial *S. commune* samples.

We conducted a phylogenetic analysis on thirty-five experimental strains (Supplementary Table 4) using nucleotide polymorphisms from two DNA amplicons to enhance data certainty. The analysis focused on the polymorphic loci MK1692_USA14-25S64 and MK90_Ru_X21_S81. After aligning amplicons and applying the maximum likelihood method, we constructed the phylogenetic tree (Fig. 6). The structure and branching of the trees confirm our earlier findings, namely the strains grouping occurs according to their origin, and the genome of strain 20R-7-F01 has ongoing, and sync with terrestrial samples changes.

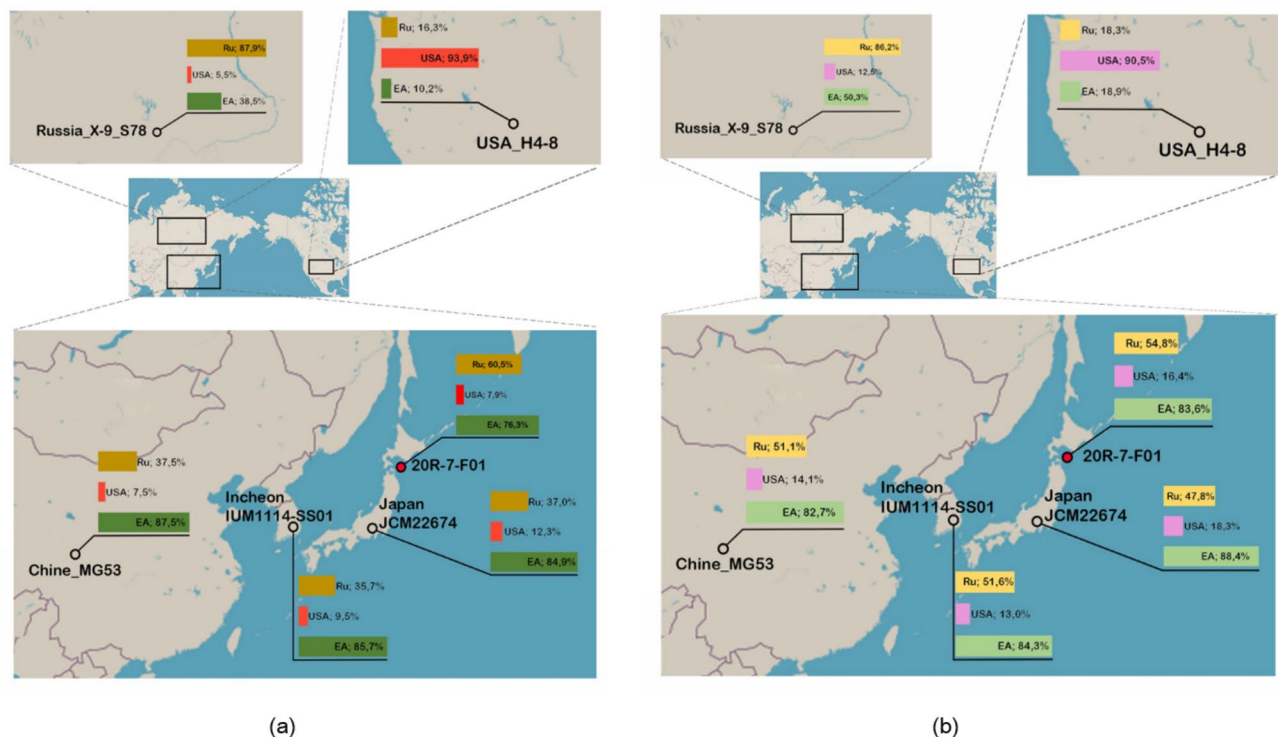


Fig. 4. Percentage of common SSR DNA loci consisting of dinucleotide (a) and trinucleotide (b) motifs in *S. commune* strains of different populations. The figure was prepared using ©mapz.com – Map Data: OpenStreetMap ODbL (<https://www.mapz.com>).

Discussion

Microsatellites, as dynamic elements of the genome, are fully functional markers that allow us to record the distribution of specimens at the local and population levels^{12,15,16}, solve complex issues of the evolution of genetic mechanisms¹⁷, predict the distribution of a species, search for rare genotypes, and develop a strategy for the conservation of endangered species^{11,18,19}. A comprehensive study of microsatellites in the *S. commune* genomes showed that strain 20R-7-F01 is in the general trend of East Asian population, and high quantitative indicators of simple DNA repeats indicate its active genetic rearrangements, which are possible under conditions of physiological activity haplo-dikaryotic stages. The high values of relatedness (>20%) at complex loci (tri- and tetranucleotide motifs) for *S. commune* 20R-7-F01 with strains of the East Asian population, in our opinion, indicate a common distant past. After all, general genetic observations for different organisms show that the rate of microsatellite mutations tends to decrease with increasing motif length. That is, it takes more time for them to appear and become fixed in the genome. Loci consisting of short motifs (mono-, di-) are capable of more dynamic changes and may indicate recent genetic rearrangements. Dinucleotide repeats are believed to be recombination hotspots²⁰, that rapidly adapt to evolutionary demands by restoring genetic variation lost to genetic drift²¹.

The results show that for the strains Japan JCM22674 (Table 3) and South Korea Incheon IUM1114-SS01 (Table 4) a high, approximately equal, percentage of common loci is observed, regardless of the complexity of the motif. For the strains China MG53 (Table 5) and 20R-7-F01 (Table 2) a high percentage of common complex loci (tri- and tetranucleotide motifs) was recorded. The high percentage of SSR loci consisting of trinucleotide motifs can be explained by their wide distribution in the genome. After all, the newly created mutations do not cause a shift in the reading frame during DNA transcription and are less likely to lead to lethal consequences. They accumulate in both coding and non-coding regions of DNA, unlike other types of SSR loci, which are mainly formed in non-coding regions^{10,22}.

In general, if we consider the indicators of relatedness at mono-pentanucleotide loci [Tables 2, 3, 4 and 5] of the strains of the East Asian population, the average percentage is (in descending order): South Korea Incheon IUM1114-SS01–18.1%; Japan JCM22674–17.8%; 20R-7-F01–15.7% and China MG53–13.8%. If we take into account the relatedness among the common loci (Fig. 3), the values are almost equalized (with the exception of China MG53): South Korea Incheon IUM1114-SS01–42%; Japan JCM22674–42.2%; 20R-7-F01–42% and China MG53–34.4%. Thus, according to these data, the *S. commune* strain 20R-7-F01 from sediments does not demonstrate possible spatial and temporal isolation.

Analyzing the distribution of common SSR DNA loci in the *S. commune* genomes with a combination of the probable geographic origin of the samples (Fig. 4) two main conclusions emerge. First, less complex loci consisting of dinucleotide motifs are more dynamic in rearrangements and reflect recent changes in the fungal genomes. Loci consisting of trinucleotide motifs require more time to form a certain pool, are more difficult

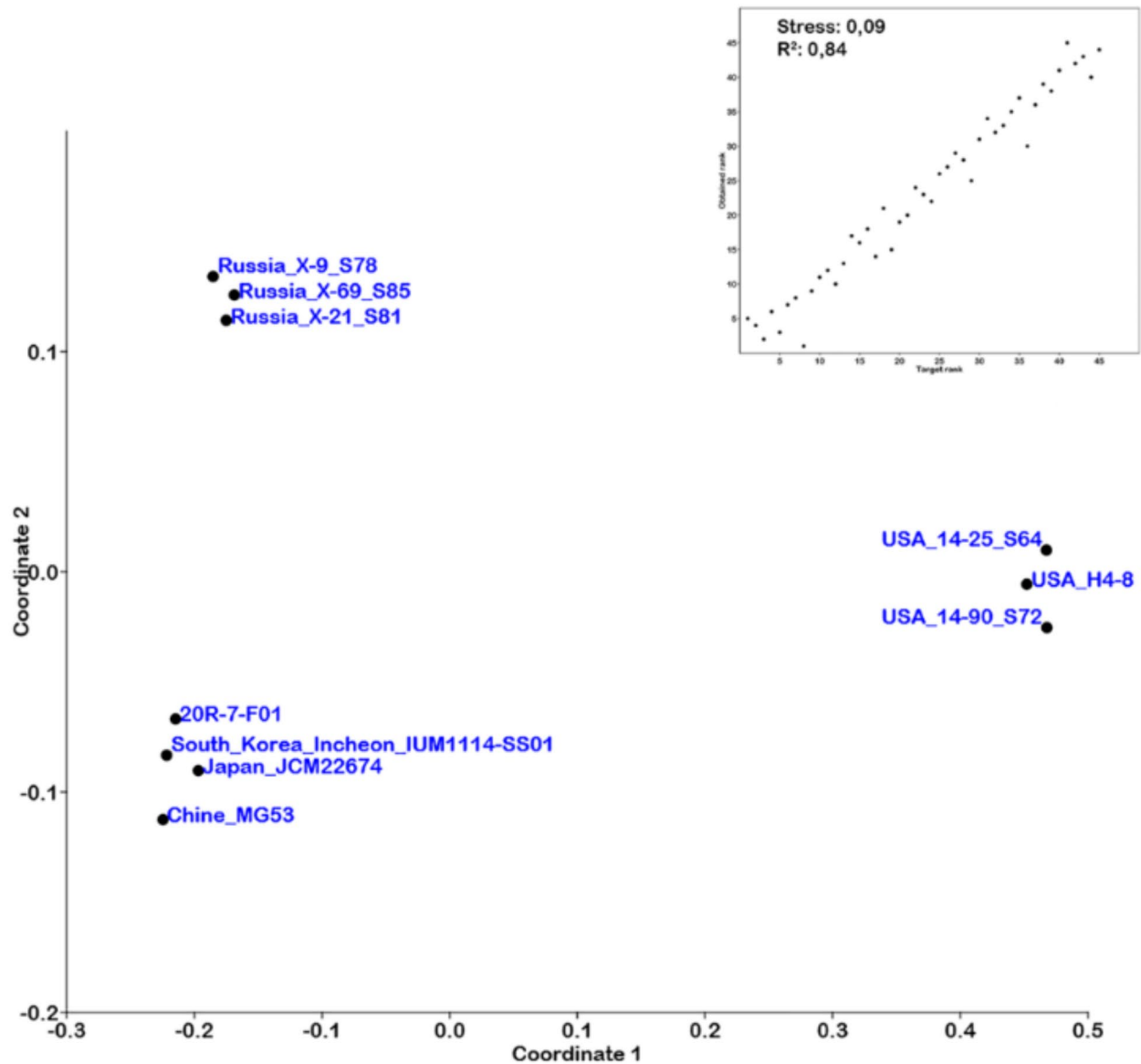


Fig. 5. Non-metric multidimensional scaling (Dice similarity index) of *S. commune* strains based on *in silico* PCR of 1938 SSR DNA loci.

to change, and indicate a change in the genetic profile over a longer period of time. This is confirmed by the example of strains of the East Asian population (Fig. 4a, b). According to the indicators of simple common loci, we can see the location China, which acts “today” as the center of distribution of the original loci (Fig. 4a). If we take into account more complex loci, we observe a high percentage of original loci in geographic locations that may have been previous centers (Japan), especially when there is a certain geographic isolation (Fig. 4b). The percentage of original loci will decrease with intense gene flow.

Second, given the distribution of the percentage of common loci of different origins, we can substantiate the genesis of *S. commune* strain 20R-7-F01. The spread of loci of American origin is primarily ensured by the waters of the Pacific Ocean, so it is quite logical that the largest percentage of them in the East Asian population will be observed in geographically close locations, such as Japan, then the coastline of the Asian continent (Fig. 4a). Going deeper into the mainland will lead to a decrease in the proportion of common loci of American origin. As for the *S. commune* strain 20R-7-F01, it occupies an intermediate position (7.9%) between the South Korean strain Incheon IUM1114-SS01 (9.5%) and China MG53 (7.5%). The percentage of shared loci with samples of the Ru population is the highest (60.5%) among East Asian strains, and at the same time, we observe the lowest percentage of original loci (76.3%). These values are in favor of the terrestrial origin of strain 20R-7-F01. Taking

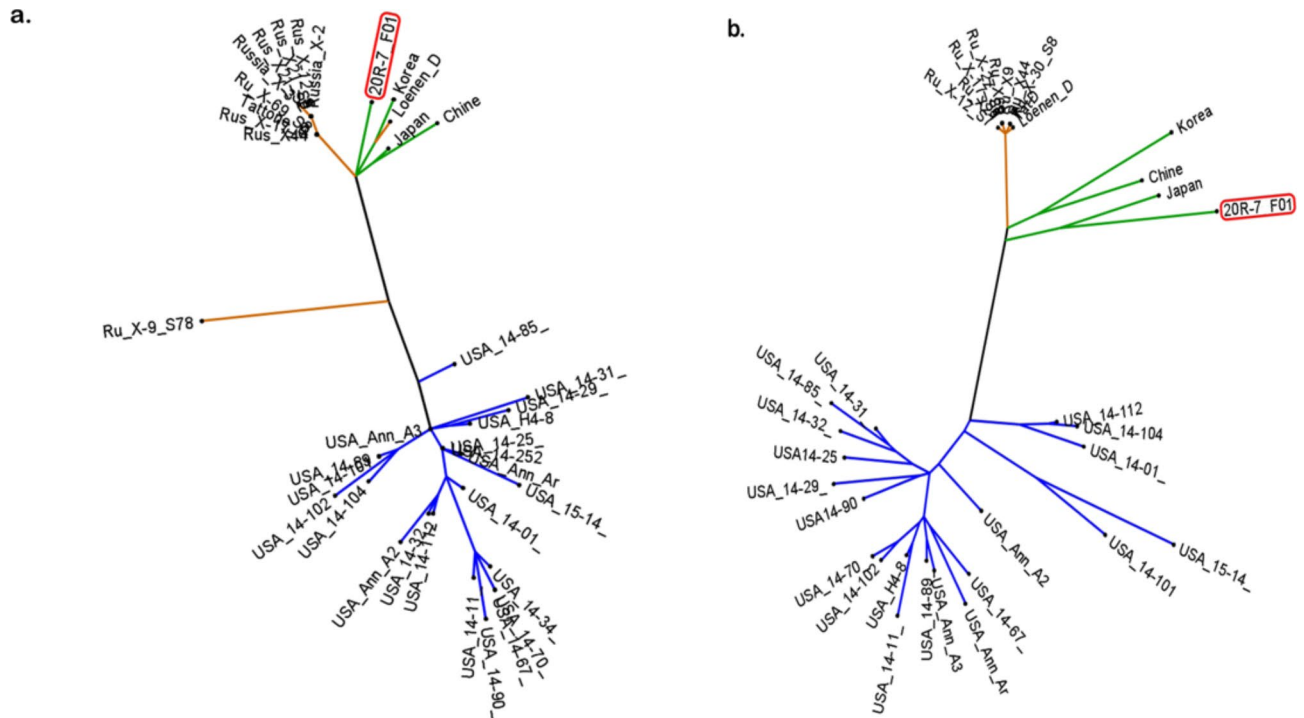


Fig. 6. Maximum likelihood phylogenetic tree based on DNA amplicon sequences by loci MK1692_USA14-25S64 (a) and MK90_Ru_X21_S81 (b).

into account the data presented in this work, we can state that *S. commune* strain 20R-7-F01 shows high genetic integration with strains of the East Asian population, is modern and, given the proportion of common DNA loci, has a probable geographical origin from the Far East of the Russian Federation (approximately Khabarovsk, Primorsky Krai).

During the drilling of well C0020A, four blocks were identified based on the depth and composition of sedimentary rocks. According to Expedition 337, Block III, from which *S. commune* strain 20R-7-F01 was isolated, consists of sandstone, silt, shale, coal, siltstone, and mud, with a porosity value of 55%^{1,23}. Interstitial water was detected from 15 cores of Block II (1256.5–1826.5 mbsf), 7 cores of Block III (1826.5–2046.5 mbsf), and 2 cores of Block IV (2046.5–2466 mbsf). The maximum interstitial water content observed at these depths was 33 ml/dm³ and only for sandstone. All other rocks did not contain interstitial water. The availability of water is one of the main factors for the normal course of physiological processes in fungi. The water content in the mycelium and/or fruiting body reaches 90% and higher. For successful cultivation of *S. commune* on solid media, their relative humidity should be 50–85%^{24,25}. In the case of the sedimentary soil sample (sandstone), the interstitial water content is approximately 3%. This is extremely insufficient for the physiological needs of the fungus, and it should be borne in mind that the high osmotic potential of this solution, due to dissolved compounds, makes this water almost inaccessible.

Temperature is another important factor for the fungus's vital activity. The optimum temperature for the growth of the mesophilic fungus *S. commune* is 20–32°C²⁶, but the fungus is able to tolerate higher temperatures. They have a critical negative effect on the structure of proteins, the functionality of membranes and the genetic apparatus of the cell, so the thermostability of the fungus is limited in time. Numerous studies have shown that at a temperature of 45°C the growth processes of the fungus *S. commune* are almost completely inhibited, and the temperature of 50°C is lethal^{27,28}. It should be noted that such results are observed at short cultivation periods and optimal values of humidity, pH, nutrient and oxygen concentrations. When several factors are critical, the lethal temperature will decrease. If we consider the *in situ* temperature conditions of the *S. commune* strain 20R-7-F01, then with a temperature gradient of 24°C/km, the temperature at the level of ~2 km below the ocean floor will be at the average 48°C¹. Such a temperature at least leads to inhibition of physiological processes of the fungus *S. commune*, and under conditions of constant acute water and oxygen deficiency, it is fatal. According to the authors⁴ representatives of *Ascomycota* and *Basidiomycota* could be preserved in a state of quiescent or as spores. But then the question arises: how, according to our data, did the exchange of genetic material occur between the *S. commune* 20R-7-F01 strain and other representatives of the East Asian population? How did high temperature and dehydration not lead to the destruction of DNA, biological membranes, and deactivation of enzymes over 20 million years? In our opinion, the discrepancies presented and the data we have presented are in favor of the terrestrial origin of strain 20R-7-F01.

Sexual reproduction in fungi provides recombination of genetic material from two sexual partners. Basidiomycetes are characterized by a haploid-dikaryotic life cycle. Basidioma can form exclusively on dikaryotic mycelium. As a result of meiosis, four haploid basidiospores are exogenously formed at the ends

of the basidia, which are an effective form of fungal spread in nature. To form a fertile dikaryotic mycelium, two compatible haploid mycelia must fuse. The fungus *S. commune* has a tetrapolar mating system, i.e., the formation of a dikaryon requires the fusion of haploid specimens that differ in two sexual compatibility factors with multiple alleles (AxBx, AyBy)^{29–31}. The tetrapolar mating system reduces the level of inbreeding to 25% among the offspring. The alternation of nuclear status forms a complex life cycle of *S. commune* and requires free spread of basidiospores. The conditions at a depth of ~2 km below the ocean floor (high pressure, density of sedimentary soil, temperature + 48°C, lack of water, oxygen, high concentrations of mineral compounds) make it impossible for intensive spread of basidiospores, growth of mycelium with different nuclear status and formation of basidiocarps. And according to our data, these processes should have occurred at the same rate as in terrestrial strains, which is impossible.

Liu et al. conducted a phylogenomic analysis of strain 20R-7-F01 in comparison with the most studied terrestrial strain of *S. commune*, H4-8⁵. Based on genomic and mitochondrial phylogenetic analyses, the authors found that the divergence time of 20R-7-F01 and H4-8 was 28–73 million years ago and 4.8–22 million years ago, respectively, which is consistent with the geological age of the sediment. Although the authors indicate that they cannot determine the origin of the fungi, they argue that the underwater fungi are indigenous and have been preserved in the submarine sediments for millions of years based on data from structural geology, sedimentology, biogeochemistry, and microbiology^{3,4}. The H4-8 strain of *S. commune*, like many others, has no geographic origin data in the NCBI database. Our previous studies have shown that it belongs to the US population¹². According to the presented results strains of the US population have the lowest affinity to strain 20R-7-F01, indicating the longest time of their divergence among all studied strains. In addition, Liu et al. point out that the entry of the fungal pool into submarine sediments through subsurface circulation is impossible, since active fluid circulation is physically impossible in the submarine sedimentary system at a depth of 2 km. We fully agree with this and have our own assumption about the ingress of fungi into the core samples. Block III (1826.5–2046.5 mbsf) is represented by several coal horizons consisting of different types of detrital coal with xylitic coal beds. Carbonaceous detritus is characterized by the dominance of finely ground material with a grain size of less than 0.1 cm, which originates from the wood of Angiosperms, as well as mosses and grasses. Charcoal xylitol is the residue of coniferous wood with fragments longer than 1 cm³². Lignite has a complex and branched pore structure³³, which is why it has good adsorption properties³⁴. These features are used in the removal of heavy metals^{35,36} and organic pollutants and dyes from wastewater^{37,38}. During drilling, a drilling fluid is used, which is fed into the drill hole under pressure, the value of which correlates with the depth. Given the significant pore surface of the lignite, it is expected that it will absorb part of the “external” biological material in the form of spores and fragments of fungal hyphae. This is confirmed by the Expedition 337 report, which states that the concentration of microbial cells in the drilling mud was consistently > 10⁸ cells/cm³. This is an order of magnitude higher than in pelagic seawater used during other expeditions³⁹, and two to five orders of magnitude higher than expected local cell concentrations in sediments more than 1000 m below the seafloor⁴⁰. In addition, during core extraction, pressure changes occur, which provoke phase changes in dissolved gases and lead to core expansion, especially in the plane of fractures, which also leads to contamination with drilling fluid. In our opinion, these factors are the reason for the contamination of cores with samples of fungi and microorganisms.

Conclusion

These data fully confirm the terrestrial origin of *S. commune* strain 20R-7-F01 and refute its genetic isolation by its inclusion in underwater sediments that are more than 20 million years old. The distribution of common SSR DNA loci of different origins in the genome of *S. commune* 20R-7-F01 strain indicates its genesis. According to the results of the analysis, the territory of the Far East of the Russian Federation (approximately Khabarovsk, Primorsky Krai) is the probable place of origin of strain 20R-7-F01. A large group of fungi⁴, which was simultaneously isolated from coal layers (1289–2457 mbsf), raises concerns about their aboriginality. The widespread mesophilic representatives of the genera *Irpex*, *Bjerkandera*, *Termitomyces* (Basidiomycota) are similar to *S. commune* in their lifestyle and, in our opinion, have similar ways of entering the core structure. The results of our study raise the urgent question of the correspondence between the age of origin of mycological and microbiological findings and the age of sedimentary cores on a global scale. Taking into account numerous scientific studies and findings using the drilling method, data containing incorrect information about the origin of samples can be accumulated. This, in turn, can lead to an erroneous judgment of the direction of evolutionary changes in a species, a mismatch in the dynamics of genetic rearrangements, a disruption in the synchronization of genome evolution with environmental factors, and so on. For this reason, it is necessary to be very careful about the age correspondence of samples and sediments, especially for objects in *status vitae*.

Methods

Genome assembly of fungus *S. commune*

Assembly of the full genome of the *S. commune* basidiomycete fungus was downloaded from the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/assembly/?term=Schizophyllum+commune>). The samples used in the study are listed in Table 1. Three conditional populations were defined according to the origin of culture: USA, Ru, and EA.

In silico SSRs identification, primers design, and PCR

The available genomes of *S. commune* were used, for search SSRs, markers design, and *in silico* PCR using the software package Genome-wide Microsatellite Analyzing Tool Package (GMATA v2.3)⁴¹. The identified motifs were one to five nucleotides in size, and the minimum repeat unit depended on the stages.

- *The first stage.* The mode search was set at a minimum number of repeats of four mononucleotides. This allowed us to cover as large a significant part of the nucleotide repeats of each genome as possible.
- *The second stage.* The mode search was set at the following parameters: mono-, di-, tri-, and tetra- motifs from four to ten repeats. The number of each motif type was counted for each strain.
- *The third stage.* The mode search was set at a minimum number of repeats of seven mononucleotides, six dinucleotides, five trinucleotides, four tetranucleotides, and four pentanucleotides.

At the stage of assessing the belonging of *S. commune* strains to a specific population, the set of genetic markers was 34 DNA loci (Supplementary Table 1), and the total number of strains with a complete DNA assembly was 35 (Supplementary Table 4)^{12,42}.

Statistical analysis

Genetic data of *S. commune* strains from three populations were analyzed by using GenAlEx v6.5, PAST v4.13 software^{43,44}. AMOVA, was performed to determine the partitioning of genetic variance among and within the populations. The principal components analysis (PCA) to find hypothetical variables accounting for as much as possible of the variance in the multivariate data and visualize the genetic diversity generated by the SSR markers was used⁴⁵. For help to discover patterns, similarities, and differences among the data points, and interpret them in terms of meaningful dimensions we used multidimensional scaling (MDS). Non-metric MDS uses an iterative procedure to find the best configuration of points that maximizes the correlation between the original and the reduced distances. Nonmetric MDS can handle ordinal data, and it can produce a stress value and a coefficient of determination that measures the quality of the fit⁴⁶.

The analysis of the *in silico* sequencing data and nucleotide alignment of amplicons was performed using BioEdit v. 7.7.1⁴⁷. Phylogenetic analyses by amplicons were performed using DNA Maximum Likelihood method⁴⁸.

Data availability

All data supporting the findings of this study are available within the paper and its Supplementary Information.

Received: 26 June 2024; Accepted: 23 December 2024

Published online: 24 February 2025

References

1. Inagaki, F. et al. Deep coalbed biosphere off Shimokita: microbial processes and hydrocarbon system associated with deeply buried coalbed in the ocean. *IODP Prel. Rept.* **337**, 1–62 (2012).
2. Inagaki, F. et al. *Core Summary for IODP Hole 337-C0020A. SIO7 Data Center, J-CORES Database; Ocean Drilling Program Center for Deep Earth Exploration (JAMSTEC)*. (2013).
3. Inagaki, F. et al. DEEP BIOSPHERE. Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor. *Science* **349**, 420–424 (2015).
4. Liu, C. H. et al. 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 (2017).
5. Liu, X. et al. Genome, genetic evolution, and environmental adaptation mechanisms of *Schizophyllum commune* in deep subseafloor coal-bearing sediments. *iScience* **25**, 104417 (2022).
6. Arifeen, Z. U. M. et al. The anaerobic survival mechanism of *Schizophyllum commune* 20R-7-F01, isolated from deep sediment 2 km below the seafloor. *Environ. Microbiol.* **23**, 1174–1185 (2021).
7. Srivastava, S., Avvaru, A. K., Sowpati, D. T. & Mishra, R. K. patterns of microsatellite distribution across eukaryotic genomes. *BMC Genom.* **20**, 153 (2019).
8. Gemayel, R., Cho, J., Boeynaems, S. & Verstrepen, K. J. Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes* **3**, 461–480 (2012).
9. Zavodna, M., Bagshaw, A., Brauning, R. & Gemmell, N. J. The effects of transcription and recombination on mutational dynamics of short tandem repeats. *Nucleic Acids Res.* **46**, 1321–1330 (2018).
10. Gonthier, P. et al. Selection processes in simple sequence repeats suggest a correlation with their genomic location: insights from a fungal model system. *BMC Genom.* **16**, 1107 (2015).
11. Boiko, S. M. The trends in the spread of simple sequence repeats in the genomes of *Schizophyllum commune*. *Mycologia* **115**, 288–298 (2023).
12. Boiko, S. M. Identification of novel SSR markers for predicting the geographic origin of fungus *Schizophyllum commune* Fr. *Fungal Biol.* **126**, 764–774 (2022).
13. Marian, I. M. et al. The transcription factor Roc1 is a key regulator of cellulose degradation in the wood-decaying mushroom *Schizophyllum commune*. *mBio* **13**, e0062822 (2022).
14. Chakraborty, R., Kimmel, M., Stivers, D. N., Davison, L. J. & Deka, R. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1041–1046 (1997).
15. Heck, D. W., Pethybridge, F. & Hay, S. J. Enabling Population Biology Studies of *Stemphylium vesicarium* from Onion with Microsatellites. *Plant Dis.* <https://doi.org/10.1094/PDIS-04-23-0706-RE> (2023).
16. Terefe, T., Boshoff, W. H. P., Park, R. F., Pretorius, Z. A. & Visser, B. Wheat stem rust surveillance reveals two new races of *Puccinia graminis* f. sp. *tritici* in South Africa during 2016 to 2020. *Plant Dis.*, 37580885 (2023).
17. Fu, C., Liu, M. & PMCID: PMC10475575. Genome-wide identification and molecular evolution of NAC gene family in *Dendrobium nobile*. *Front. Plant. Sci.*, 37670854 (2023).
18. Tseng, M. C., Lee, Y. H., Yen, T. B. & Li, S. M. Genome-wide characterization of microsatellites in cobia *Rachycentron canadum* (Linnaeus, 1766): survey and analysis of their abundance and diversity. *J. Fish. Biol.* **PMID**, 37658731 (2023).
19. Li, S. Q., Zhang, C. & Gao, X. F. Geographic isolation and climatic heterogeneity drive population differentiation of *Rosa chinensis* var. *Spontanea* complex. *Plant Biol.* **25**, 620–630 (2023).
20. Bailey, A. D., Pavelitz, T. & Weiner, A. M. The microsatellite sequence (CT)_n x (GA)_n promotes stable chromosomal integration of large tandem arrays of functional human U2 small nuclear RNA genes. *Mol. Cell. Biol.* **18**, 2262–2271 (1998).
21. Foster, P. L. & Trimarchi, J. M. Adaptive reversion of a frameshift mutation in *Escherichia coli* by simple base deletions in homopolymeric runs. *Science* **265**, 407–409 (1994).

22. Huang, X. et al. Genome-wide survey and analysis of microsatellites in waterlily, and potential for polymorphic marker development. *Genes* **13**, 1782 (2022).
23. Inagaki, F. et al. IODP Expedition 337: deep Coalbed Biosphere off Shimokita – Microbial processes and hydrocarbon system associated with deeply buried coalbed in the ocean. *Sci. Drill.* **21**, 17–28 (2016).
24. Arboleda Valencia, J. W. et al. Holocellulase activity from *Schizophyllum commune* grown on bamboo: a comparison with different substrates. *Curr. Microbiol.* **63**, 581–587 (2011).
25. Salmon, D. N. et al. A bioprocess for the production of phytase from *Schizophyllum commune*: studies of its optimization, profile of fermentation parameters, characterization and stability. *Bioprocess Biosyst. Eng.* **35**, 1067–1079 (2012).
26. Singh, S. et al. Characterization and development of cultivation technology of wild split gill *Schizophyllum commune* mushroom in India. *Sci. Hort.* **289**, 110399 (2021).
27. Castillo, G., Nihoul, A. & Demoulin, V. Correlation between the *in vitro* growth response to temperature and the habitat of some lignicolous fungi from Papua New Guinea coastal forests. *Cryptog. Mycol.* **25**, 57–81 (2004).
28. Hanafusa, Y. et al. First isolation of *Schizophyllum commune* in a harbor seal (*Phoca vitulina*). *Med. Mycol.* **54**, 492–499 (2016).
29. Kothe, E. Mating types and pheromone recognition in the Homobasidiomycete *Schizophyllum commune*. *Fungal Genet. Biol.* **27**, 146–152 (1999).
30. Kothe, E., Gola, S. & Wendland, J. Evolution of multispecific mating-type alleles for pheromone perception in the homobasidiomycete fungi. *Curr. Genet.* **42**, 268–275 (2003).
31. Specht, C. A. Isolation of the B alpha and B beta mating-type loci of *Schizophyllum commune*. *Curr. Genet.* **28**, 374–379 (1995).
32. Zakrzewski, M., Schertel, A., Brus, G., Wagner, M. & Sciazko, A. A three-dimensional reconstruction of coal microstructure using the Cryo-FIB-SEM technique. *Fuel* **275**, 117919 (2020).
33. Jiang, P. et al. Pore characteristics and its heterogeneity of lignite reservoir in the Erlian Basin of Inner Mongolia, China. *Energy Explor. Exploit.* **40**, 1555–1572 (2022).
34. Qi, Y., Hoadley, A. E., Chaffee, A. L. & Garnier, G. Characterisation of Lignite as an industrial adsorbent. *Fuel* **90**, 1567–1574 (2011).
35. Havelcová, M., Mizera, J., Sýkorová, I. & Pekar M. Sorption of metal ions on lignite and the derived humic substances. *J. Hazard. Mater.* **161**, 559–564 (2009).
36. Pentari, D., Perdikatsis, V., Katsimicha, D. & Kanaki, A. Sorption properties of low calorific value Greek lignites: removal of lead, cadmium, zinc and copper ions from aqueous solutions. *J. Hazard. Mater.* **168**, 1017–1021 (2009).
37. Allen, S. J., Whitten, L. J., Murray, M. & Duggan, O. The adsorption of pollutants by peat, lignite and activated chars. *J. Chem. Technol. Biotechnol.* **68**, 442–452 (1997).
38. Venkata Mohan, S., Chandrasekhar Rao, N. & Karthikeyan, J. Adsorptive removal of direct azo dye from aqueous phase onto coal based sorbents: a kinetic and mechanistic study. *J. Hazard. Mater.* **90**, 189–204 (2002).
39. Lever, M. A. et al. Trends in basalt and sediment core contamination during IODP Expedition 301. *Geomicrobiol. J.* **23**, 517–530 (2006).
40. Parkes, R. J., Cragg, B. A. & Wellsbury, P. Recent studies on bacterial populations and processes in subseafloor sediments: a review. *Hydrogeol. J.* **8**, 11–28 (2000).
41. Wang, X. & Wang, L. G. M. A. T. A. An integrated software package for genome-scale SSR mining, marker development and viewing. *Front. Plant. Sci.* **7**, 1350 (2016).
42. Boiko, S. M. Design of microsatellite markers for *Schizophyllum commune* (Agaricales, Basidiomycota) based on analysis of its genome. *Ukr. Bot. J.* **79**, 27–34 (2022).
43. Peakall, R. & Smouse, P. E. GenALEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539 (2012).
44. Hammer, Ø., Harper, D. A. T. & Ryan, P. D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electron.* **4**, 9 (2001).
45. Jackson, D. A. Stopping rules in principal components analysis: a comparison of heuristic and statistical approaches. *Ecology* **74**, 2204–2214 (1993).
46. Taguchi, Y. H. & Oono, Y. Relational patterns of gene expression via non-metric multidimensional scaling analysis. *Bioinformatics* **21**, 730–740 (2005).
47. Hall, T., Biosciences, I. & Carlsbad, C. BioEdit: an important Software for Molecular Biology. *GERF Bull. Biosci.* **2**, 60–61 (2011).
48. Xia, X. Maximum likelihood in Molecular Phylogenetics. In: *Bioinformatics and the Cell*. Springer, Cham. (2018).

Author contributions

S.M.B. designed research, performed research, analyzed data, and wrote the paper.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-84457-2>.

Correspondence and requests for materials should be addressed to S.M.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025