

Identification of arbuscular mycorrhizal fungi in soils of the North Caucasus based on Illumina MiSeq data for ITS1 and ITS2 regions

Andrey P. Yurkov^{1*}, Alexey A. Kryukov¹, Anastasia O. Gorbunova¹, Andrey V. Shcherbakov¹, and Peter M. Zhurbenko²

¹ All-Russia Research Institute for Agricultural Microbiology, Laboratory of Ecology of Symbiotic and Associative Rhizobacteria, 196608 Saint Petersburg, Russia

² Komarov Botanical Institute, Laboratory of Biosystematics and Cytology, 197376 Saint Petersburg, Russia

Abstract. The objective of our research was to analyze the efficiency of identification of arbuscular mycorrhizal fungi (AMF) for 2 regions: ITS1 and ITS2 regions of AMF DNA isolated from the soils of the North Caucasus (Karachay-Cherkessia). For the first time the necessity of different AMF species identification using both ITS regions was revealed, but not one region. The research demonstrated: 1) the set of taxa is different using ITS1- and ITS2-based identification; 2) analysis of the ITS1 region reveals a greater number of operational taxonomic units; 3) ITS2 allows identification of AMF at the species level more often. Sample preparation for Illumina MiSeq analysis was optimized. Obligatory stages in the sample preparation were the purification of DNA in the agarose gel in Silica after isolation, as well as separate amplification of ITS1 and ITS2 followed by combining and joint sequencing for each sample. The results showed the highest AMF biodiversity for the 176Te sample from the ecosystem of the subalpine meadow of the southeastern slope of Malaya Hatipara mountain (43°25'48.0"N 41°42'31.0"E; 2401 m above sea level), in which 8 species of AMF were identified (*Archaeospora spainiae*, *Claroideoglomus claroideum*, *Diversispora versiformis*, *Entrophora infrequens*, *Funnelformis mosseae*, *Glomus indicum*, *Paraglomus laccatum*, *Rhizophagus irregularis*).

1 Introduction

Certainly, AMF are able to enhance plant growth due to the transfer of phosphorus, water, and a number of macro and microelements. In this regard, the urgent problem is to select the most effective strains and new AMF species from region with high biodiversity, both to study the mechanisms of their interaction with plants and to use in agriculture. In this regard, the assessment of AMF diversity in the soils of the North Caucasus, as the single “hot spot” of biodiversity in Russia, is the most appropriate. On the other hand, the identification of AMF (both morphological and molecular-genetic identification) has a number of difficulties.

*Corresponding author: yurkovandrey@yandex.ru

Morphological identification often does not allow a clear separation of closely related species [1]. At the same time, views of various authors on the systematics of Glomeromycotina radically differ [2]. The examples of multiple misidentifications of AMF have been revealed [3]. The reason for difficulties in molecular genetic identification is high genetic polymorphism of AMF, including high intraspecific polymorphism. This is due to the ability of AMF to form anastomoses and exchange genetic material, as well as to form a significant number of nuclei. At present there is no approved region for AMF DNA barcoding, despite the active development of the next generation sequencing (NGS) [4-5]. Prior to the widespread NGS, the SSU-ITS1-5.8SrRNA-ITS2-LSU region (or only part of the SSU) was most often used to identify AMF. A mixture of specific primers was selected for PCR, the use of which did not guarantee detection of total AMF polymorphisms. Sanger sequencing was performed for up to 48 clones per sample [6]. With the widespread introduction of NGS from 2014-2015 with application of universal primers, the efficiency of AMF identification and assessment of their genetic polymorphism increased, especially for soil microbiomes. On the other hand, the disadvantage of NGS, in particular, Illumina MiSeq, is the smaller length of analyzed regions (of which ITS2 is most often used) compared to Sanger sequencing. At the same time, a number of authors note to identify the studied taxa the separate analysis of the efficiency of their identification for ITS1 and ITS2 regions is desirable to have been performed. For example, some works mentioned that different sets of taxa can be detected in the analysis of fungal ITS1 and ITS2 [7-8]. Results of some studies, including those we conducted earlier [1, 5, 9], showed the detected AMF sequences very often have significant differences from sequences presented in the databases. For example, the NCBI database contains <50% of AMF species [4-5]. Thus, urgent issues of AMF systematics are the need of research for the high variability of their DNA and the active replenishment of the GenBank (NCBI) with AMF sequences, which will significantly increase the efficiency of AMF species identification. In this regard, the aim of the present study is to optimize the AMF identification methodology in terms of comparing the identification for ITS1 and ITS2 regions with the assessment of species diversity in soil samples of one of the largest centers of world biodiversity – the North Caucasus.

2 Materials and methods

2.1 Sampling

The soil samples collected at three stationary trial plots in Karachay-Cherkessia to isolate AMF strains:

1) No. 176Te, a subalpine grass-forb meadow of the south-eastern slope of the Malaya Hatipara mountain (43°25'48.0"N 41°42'31.0"E; 2401 m above sea level) with a slope of 25-30°. There are low trees and shrubs (*Pinus sp.*, *Betula sp.*, *Salix sp.*, *Juniperus communis* and *Rosa sp.*). There are cushion mosses and epilithic lichens on the rocky ledges. Dominants: grasses *Festuca sp.* and *Agrostis planifolia*. Subdominants: umbellifers *Heracleum*, *Laserpitium stevenii* and *Bupleurum polyphyllum*.

2) No. 145Te, a subalpine meadow of the Alibek river valley (Alibek mountain; 43°18'05.0"N 41°33'49.0"E; 1957 m above sea level); anthropogenically distressed tall grass-forb meadow with moderate moisture, a wavy microrelief with a surface slope of 10-15°. The meadow is surrounded by low trees *Betula sp.* and *Acer campestre*. Dominants: umbellifers *Eleutherospermum cicutarium*, *Seseli libanotis* and *Anthriscus sylvestris*. Subdominant: *Chamaenerion angustifolium*.

3) No. 189Te, a moist ruderal meadow of the Kuban river valley near the Ordzenikidzevskiy settlement (43°51'38.0"N 41°54'22.0"E; 795 m above sea level), a

floodplain pasture without a slope. There are small shrubs of *Alnus glutinosa*. Dominant: *Ambrosia artemisiifolia*. Subdominant: *Setaria pumila*.

2.2 Illumina MiSeq sequencing and data analysis

Sample preparation for sequencing was performed according to the method described in the work [5] with modifications. DNA extraction from the soil was carried out by the CTAB method after grinding samples with liquid nitrogen. Then, the DNA was additionally purified in Silica agarose gel: this procedure significantly (by 2-3 times) increases the yield of the PCR product. PCR was conducted separately for ITS1 and ITS2 with universal primers for these regions: this technique increases the PCR product yield by 1.5-2 times in comparison with amplification for two regions at the same time. The primers were synthesized in Evrogen (Evrogen, Russia) with adapters for Illumina MiSeq: ITS5 primer (5'-GGAAGTAAAAGTCGTAACAAGG-3') and our modified ITS-6RK reverse primer (5'-CGTTCAAAGATTTCGATGATTCAC-3') for amplification of ITS1 region; ITS3 primer (5'-GCATCGATGAAGAACGCAGC-3') and ITS4 primer (5'-TCCTCCGCTTATTGATATGC-3') for amplification of ITS2. The products were purified in Silica agarose gel after amplification. Amplicon libraries were sequenced in Illumina MiSeq using the MiSeq® Reagent Kit v3 (600-cycle) with two-sided reading (2x300 N) (Illumina, Inc., USA). Sequences were processed using Illumina software (Illumina, Inc.) and USEARCH software [10]. The main steps in USEARCH data processing are described at <https://www.drive5.com/usearch>. Based on the obtained data, a taxonomic analysis of all samples was performed using the MEGA7 software package [11] for ITS1 and ITS2 separately. At the same time, the representation of various taxa, p-distance, and phylogenetic trees constructed by the NJ (Neighbor Joining) and MP (Maximum Parsimony) methods were evaluated. The biodiversity was assessed using the Shannon and Margalef indices.

3 Results and discussion

After sequencing, the average size of the amplicon library was 29917 sequences, and after filtering poor-quality reads, pairing and removing chimeras, it decreased down to 5548. Results showed the following classes of fungi were determined in samples isolated from the soils of various ecosystems of the North Caucasus: Agaricomycetes (Ag), Chytridiomycetes (Ch), Dothideomycetes (Do), Eurotiomycetes (Eu), Glomeromycetes (Gl, AM fungi), Leotiomycetes (Le), Malasseziomycetes (Ma), Mucoromycetes (Mu), Pezizomycetes (Pe), Sordariomycetes (So), Tremellomycetes (Tr), minor taxa (other) and not identified (NI). Significant variations in the frequency of classes in terms of the number of operational taxonomic units (OTU), including copies both for samples and for ITS regions (Figure 1), were revealed. The 176Te sample was characterized by greatest variety of fungus classes. Moreover, the proportion of AMF (class Glomeromycetes) ranged from 0.7-4.9% for ITS1 and 0.6-1.4% for ITS2 of total fungal OTU. This is consistent with the results obtained by other authors, according to the proportion of AMF was about 1.9% of fungal sequences using ITS1-based identification and 0.4% for ITS2 [8]. On the other hand, only 0.17% of AMF was detected by 454 GS-FLX+ sequencing for LSU region in the work of Senes-Guerrero et al. [4].

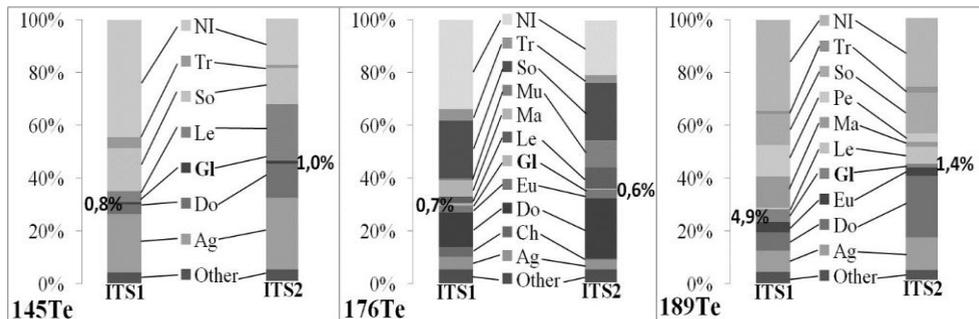


Fig. 1. Identified fungal classes by the OTU proportion for the ITS1 and ITS2 regions

A comparison of identified fungal classes based on ITS1 and ITS2 (Table 1) was shown the number of fungal OTU was significantly higher for the ITS1 region compared to the ITS2 region (by 1.3-3.8 times). Despite this, the number of taxa from class to species was slightly higher for ITS2 region in 189Te sample (1.2-1.7 times), but for other samples (145Te and 176Te) it was slightly higher for ITS1 region (1.1-1.4 times and 1.2-1.5 times, respectively). According to the results obtained by Sommermann et al., the number of fungal OTU was higher for ITS1, while data for ITS2 were more diverse in taxonomic composition [8]. On the contrary, Kichko et al. [7] note the DNA of the genus *Phytophthora* fungus was detected only using universal primers for ITS1 in the absence of ITS2 fragments between analyzed amplicons, despite the fact that the number of OTU was 25% lower for identification by ITS1, than for ITS2 (187 OTU versus 249 OTU, respectively). Thus, identification for only one ITS region will be inaccurate, even at the genus level.

Table 1. The number of identified fungal taxa and OTU for the ITS1 and ITS2 regions

Taxon	145Te sample		176Te sample		189Te sample	
	ITS1	ITS2	ITS1	ITS2	ITS1	ITS2
Phylum	4	3	4	5	4	4
Class	11	10	20	17	11	13
Order	21	18	47	32	16	25
Family	37	27	72	62	24	40
Genus	60	46	140	106	34	58
Species	67	54	178	122	37	64
OTU	234	67	621	164	105	84

Taxonomic analysis of AMF was carried out by using phylogenetic trees (Table 2). 145Te sample did not have high biodiversity. Species from the genera *Claroideoglomerus*, *Rhizophagus*, *Acaulospora* and *Entrophospora* were identified. Moreover, only sequences of the genus *Claroideoglomerus* were detected for both ITS regions. At the same time, a significant part of the sequences were not identified at the species level. Among identified taxa, virtual ones can be found, for example, in the genus *Acaulospora*. A few sequences were defined only at the family level.

176Te sample was characterized by the highest biodiversity. The following genera can be revealed among the sequences: *Archaeospora*, *Rhizophagus*, *Diversispora*, *Funneliformis*, *Paraglomerus*, *Septoglomerus*, *Scutellospora*, *Dominikia*, *Claroideoglomerus*, *Glomerus*, *Entrophospora*. It is interesting the identification was successfully carried out to the species only for 8 genera out of 11. Only two genera were identified for both ITS regions: *Scutellospora* and *Paraglomerus*. On other hand, a significant amount of OTU was referred to the genus *Septoglomerus*, despite the fact that these sequences were not identified at the species level.

Sequences related to the following genera were found in 189Te sample: *Claroideoglomus*, *Dominikia*, *Entrophospora*, *Glomus*, *Paraglomus*, *Rhizophagus*. Remarkably, the sets of taxa for the ITS1 and ITS2 regions were completely different for the 189Te. Herewith it can be noted that the proportion of sequences identified at the species level was greater in 189Te sample than in others samples.

Table 2. Identified AMF OTU at the species or genus level

Sample	Identification using ITS1 region		Identification using ITS2 region	
	at the genus level, number of OTU	at the species level, number of OTU, sequence numbers	at the genus level, number of OTU	at the species level, number of OTU, sequence numbers
145Te	<i>Claroideoglomus</i> , 2	<i>C. claroideum</i> , 1, MT447955; <i>C. sp.</i> , MT447954	<i>Claroideoglomus</i> , 2	<i>C. claroideum</i> , 1, MT451854; <i>C. sp.</i> , MT451850
	<i>Rhizophagus</i> , 4	<i>R. irregularis</i> , 3, MT447980-MT447982; <i>R. sp.</i> , MT447992	<i>Entrophospora</i> , 1	<i>E. infrequens</i> , 1, MT451858
	<i>Acaulospora</i> , 1	<i>Ac. sp.</i> , MT447948		
	only at the family level, 5	<i>Glomeraceae sp.</i> , MT447993- MT447997	only at the family level, 0	
Total OTU in 145Te	7	4 (57%)	3	2 (67%)
176Te	<i>Archaeospora</i> , 2	<i>Ar. spainiae</i> , 2, MT447949-MT447950	<i>Claroideoglomus</i> , 3	<i>C. claroideum</i> , 3, MT451851-MT451853
	<i>Rhizophagus</i> , 9	<i>R. irregularis</i> , 3, MT447983- MT447985; <i>R. sp.</i> , MT448000- MT448004	<i>Glomus</i> , 1	<i>Gl. indicum</i> , 1, MT451865
	<i>Diversispora</i> , 3	<i>Di. versiformis</i> , 3, MT447951-MT447953	<i>Entrophospora</i> , 3	<i>E. infrequens</i> , 3, MT451860- MT451862
	<i>Funneliformis</i> , 3	<i>F. mosseae</i> , 3, MT447956- MT447958	<i>Paraglomus</i> , 3	<i>P. laccatum</i> , 3, MT451868- MT451870
	<i>Paraglomus</i> , 3	<i>P. laccatum</i> , 3, MT448008- MT448010	<i>Scutellospora</i> , 1	<i>Sc. sp.</i> , MT451849
	<i>Septoglomus</i> , 28	<i>Se. sp.</i> , MT448012- MT448039		
	<i>Scutellospora</i> , 3	<i>Sc. sp.</i> , MT408934- MT408936		
	<i>Dominikia</i> , 5	<i>Do. sp.</i> , MT447960- MT447964		
only at the family level, 6	<i>Glomeraceae sp.</i> , MT447975- MT447978, MT447998- MT447999	only at the family level, 0		
Total OTU in 176Te	56	14 (25%)	11	10 (91%)

189Te	<i>Rhizophagus</i> , 7	<i>R. irregularis</i> , 1, MT447979; <i>R. sp.</i> , MT447986- MT447991	<i>Claroideoglossum</i> , 1	<i>C. claroideum</i> , 1, MT451855
	<i>Paraglossum</i> , 1	<i>P. laccatum</i> , 1, MT448011	<i>Entrophospora</i> , 3	<i>E. infrequens</i> , 3, MT451859, MT451863- MT451864
	<i>Dominikia</i> , 4	<i>Do. sp.</i> , MT447965- MT447968	<i>Glomus</i> , 2	<i>Gl. indicum</i> , 2, MT451866- MT451867
	only at the family level, 6	<i>Glomeraceae sp.</i> , MT447969-MT447974	only at the family level, 0	
Total OTU in 189Te	12	2 (17%)	6	6 (100%)
Average OTU for sample	25.0	6.7 (27%)	6.7	6.0 (90%)

Total analysis of 3 samples did not disclose the presence of OTU of the genus *Rhizophagus* for ITS2, although fungi of this genus are cosmopolitans and therefore are often found in various soils.

In all samples an unusual species *E. infrequens* was found, which was not related to AMF 2 years ago, but is currently characterized as AMF with an uncertain taxonomic position (Diversisporales incertae sedis). The species *E. infrequens* is also interesting because of *E. infrequens* sequence for ITS2 region is much shorter (80-100 bp) than all the other AMF genera, presented in the Genbank (NCBI), that also indicates the unique position of the genus *Entrophospora* in the AMF phylogeny. In the future, a research of representatives of this genus may lead to significant reorganization in the AMF systematics.

According to Table 2, the number of genera and species varied greatly depending on the place of sampling. Evaluation of the diversity indices showed that for 145Te, 176Te and 189Te samples (based on the number of OTU for each AMF species that were detected in the samples in total for both ITS regions) the Shannon index was 1.46, 3.11, and 2.16, and the Margalef index was 1.12, 2.52 and 1.92, respectively. That is, the ecosystem of the subalpine zone of Malaya Hatipara was the most diverse (the Shannon index is 1.4-2.1 times higher, and the Margalef index 1.3-2.3 times higher compared to the other two ecosystems). 14 OTU were identified in 176Te sample at the species level using the ITS1 region, and 10 OTU for ITS2. Moreover, the identified species for two regions were differed. Such differences are characteristic of all samples (Table 2). The largest total number of species was detected in the 176Te sample and was equal to eight: *Ar. spainiae*, *C. claroideum*, *Di. versiformis*, *E. infrequens*, *F. mosseae*, *Gl. indicum*, *P. laccatum*, *R. irregularis*. Literature data indicated a greater possible diversity of AMF. So, studying soil samples taken in the Andes, it was shown that about 60% of analyzed plants formed symbiosis with >10 AMF species each [4]. On the other hand, comparison of identified taxa showed that, on average, OTU were identified at the genus level for ITS1 3.7 (6.7/25) times more than for ITS2 (Table 2). But, meanwhile, the proportion of OTU identified at the species level, on the contrary, was 3.4 times greater (90% / 27%) for ITS2 versus ITS1 (Table 2). These data are consistent with the results of several authors [8].

Note that many AMF species were not identified for the ITS1 region (Table 2). This fact can be explained by insufficient number of sequences in the GenBank for accurate identification, as well as the high polymorphism of ITS sequences of AMF. Thus, the most accurate identification of AMF in soil samples will be due to combining the results of sequencing in both regions.

4 Conclusion

Obtained results for the soil samples, that were collected at several points of single center of world biodiversity in Russia – in the North Caucasus, showed the effectiveness of AMF identification for two ITS regions. Noted, only one region is used for the genetic identification of soil fungi in most studies, and usually it is the ITS2. The present results showed the need for separate amplification for the ITS1 and ITS2 regions with subsequent joint sequencing. An important step in the preparation of AMF soil samples for sequencing is additional purification of DNA in the agarose gel. This approach revealed the high biodiversity of AMF in 176Te sample both in terms of species and generic composition and in terms of diversity indices in comparison with the 145Te and 189Te samples. A comparative analysis of identification results for ITS1 and ITS2 showed the set of detected taxa for one of the regions is incomplete. It is interesting, the number of AMF OTU detected for ITS1 was higher (on average by 2-3 times) than for ITS2. At the same time, the identification at the species level for the ITS1 region was more efficient. Analysis of amplicon libraries showed a significant number of OTU identified at the genus or family level only. Probably, significant number of AMF species has not been studied, and the methods for their identification require improvement.

This work was supported by the Russian Foundation for Basic Research grant No. 19-29-05275 MK and partially grant No. 20-016-00245 A. The research was carried out using the equipment of the Core Center “Genomic Technologies, Proteomics and Cell Biology” in All-Russia Research Institute for Agricultural Microbiology.

References

1. A. A. Kryukov, A. P. Yurkov, *Mycology and Phytopathology* **52**, 38 (2018)
2. J. W. Spatafora, Y. Chang, G. L. Benny, K. Lazarus, M. E. Smith, M. L. Berbee, G. Bonito, N. Corradi, I. Grigoriev, A. Gryganskyi, T. Y. James, K. O'Donnell, R. W. Roberson, T. N. Taylor, J. Uehling, R. Vilgalys, M. M. White, J. E. Stajich, *Mycologia* **108**, 1028 (2016)
3. A. Schußler, M. Kruger, C. Walker, *PloS One* **6** (2011). DOI: 10.1371/journal.pone.0023333
4. C. Senes-Guerrero, A. Schußler, *Fungal Divers* **77**, 317 (2016). DOI: 10.1007/s13225-015-0328-7
5. A. A. Kryukov, A. O. Gorbunova, E. M. Machs, Y. V. Mikhaylova, A. V. Rodionov, P. M. Zhurbenko, A. P. Yurkov, *Vavilov journal of genetics and breeding* **24**, 158 (2020). DOI: 10.18699/VJ19.38-o
6. M. Kruger, H. Stockinger, C. Kruger, A. Schubler, *New Phytol.* **183**, 212 (2009). DOI: 10.1111/j.1469-8137.2009.02835.x
7. A. A. Kichko, T. S. Aksenova, V. M. Shapkin, A. O. Zverev, A. V. Khutti, E. E. Andronov, *Agricultural biology* **54**, 990 (2019). DOI: 10.15389/agrobiology.2019.5.990rus
8. L. Sommermann, J. Geistlinger, D. Wibberg, A. Deubel, J. Zwanzig, D. Babin, A. Schlüter, I. Schellenberg, *PLoS One* **13**, (2018). DOI: 10.1371/journal.pone.0195345
9. A. P. Yurkov, A. A. Kryukov, A. O. Gorbunova, A. P. Kozhemyakov, G. V. Stepanova, E. M. Machs, A. V. Rodionov, M. F. Shishova, *Ecological Genetics* **16**, 11 (2018). DOI: 10.17816/ecogen16211-23
10. R. C. Edgar, *Bioinformatics* (Oxford, England, 2010)
11. S. Kumar, G. Stecher, K. Tamura, *Mol. Biol. Evol.* **33**, 1870 (2016). DOI: 10.1093/molbev/msw054