http://dx.doi.org/10.1016/j.ejsobi.2016.10.005

E. Capra [a, *], R. Giannico [b], M. Montagna [c], F. Turri [a], P. Cremonesi [a], F. Strozzi [b], P. Leone [a], G. Gandini [d], F. Pizzi [a]

a Istituto di Biologia e Biotecnologia Agraria, Consiglio Nazionale delle Ricerche, via Einstein, 26900 Lodi, Italy b Parco Tecnologico Padano, via Einstein, 26900 Lodi, Italy c Universita degli Studi di Milano, Dipartimento di Scienze Agrarie e Ambientali - Produzione, Territorio, Agroenergia, via Celoria 2, 20133 Milano, Italy [d] Universita degli Studi di Milano, Dipartimento di Medicina Veterinaria, via Celoria 10, 20133 Milano, Italy

Keywords:
Metazoa
Soil
18S rRNA
Metagenomics

a b s t r a c t

Reliable evaluations of soil biodiversity represent a key factor in understanding ecosystem services. To date, species-discriminating barcodes efficiently describe bacterial and fungal communities associated with environmental samples, whereas investigations of soil microfauna are often hampered by the lack of a marker region encompassing the taxonomic range of soil organisms. Two new PCR primer sets targeting the V4-V5 and V5-V7 variable regions of the ribosomal 18S RNA (18S rRNA) were designed to be specific for metazoans metabarcoding and capable of detecting the majority of their lineages. In silico and in vivo assays on four soil typologies were carried out to compare the newly developed primer sets with a selection of primers targeting the homologous gene, which were previously used to assess soil metazoan biodiversity. The new primer sets, both on the basis of the in silico and in vivo comparisons, were very selective and consistent when analysing metazoan biodiversity across the tested soil typologies. On the basis of the coverage index and taxonomic resolution, the new primers targeting the ribosomal 18S RNA outperformed the other primers, and they represent a promising tool for assessing soil metazoan biodiversity through metabarcoding approaches.

.

1. Introduction

The soil ecosystem hosts high numbers of organisms [15] and represents a major reservoir of biodiversity. In this complex system, invertebrates are key components both in terms of abundance and in providing ecosystem services, such as actively contributing to soil formation and cycling of elements [18]. Reliable approaches to measure and evaluate biodiversity of soil organisms represent promising tools for a plethora of purposes, encompassing conservation biology investigations, biomonitoring programs, and evaluation of soil alterations. Traditional approaches to investigate biodiversity of soil organisms rely on morphological identification of taxa; the disadvantages of this procedure are the high level of expertise and the time effort required [30]. In recent years, the increasing use of DNA markers in species identification [14,23,25], has changed the morphology-based perspective in biodiversity studies. The metabarcoding approach is the leading candidate [1,3], to investigate organism diversity and to provide ecological information [11,33]. Briefly, the metabarcoding approach consists of detection and identification of taxa from complex environmental matrices (e.g., soil, soil litter, canopy, and water) starting from DNA extracted from bulk samples. Regardless of the high-throughput

sequencing platforms adopted, molecular markers and primers play a crucial role in the recovery of a high proportion of taxa diversity. For some metazoan groups, the most commonly used marker is a portion of the mitochondrial Cytochrome Oxidase subunit I (COI), characterized by a relatively high nucleotide substitution rate [24] which allows for discrimination of organisms at the species level. However, because of these features, it is difficult to identify regions conserved across the main animal lineages to design primers suitable for DNA metabarcoding. Nevertheless, primer sequences amplifying a 50 fragment of COI in metazoans [10] were successfully adapted to high-throughput sequencing technologies to characterize the metazoan biodiversity of bulk samples, including arthropods [33], and coral fish gut content [19].

The use of markers more conserved than the COI gene represents a possible solution to obtain universal PCR primers for metazoans. The small ribosomal subunit 18S RNA (18S rRNA), which possesses a lower nucleotide substitution rate compared to the COI gene, represents an alternative to mitochondrial markers. Nevertheless, a possible flaw of 18S rRNA is low discriminating capability in distinguishing among closely related species.18S rRNA primers were successfully used in biodiversity surveys of marine ecosystems, characterized mainly by polychaetes [8,11], and of soil ecosystems characterized by nematodes and arthropods [26,32].

The type of sample can greatly influence the choice of the region (COI, 18S rRNA) for metazoan metabarcoding. Recently, 18S rRNA metabarcoding was successfully used to characterize the metazoan communities in leaf-litter and soil. Because of the inability of COI to amplify from soil samples, >99% of returned Operational Taxonomic Units (OTUs) were bacterial, and the use of COI was limited in evaluation of Metazoa present in Malaise traps and Canopy fogging[32]. Similarly, the characterization of communities associated with Zostera marina seagrass meadows, using both mitochondrial and nuclear ribosomal genetic markers (COI and 18S rRNA) revealed that the highest proportion of OTUs were nonMetazoa in the case of COI gene, whereas a large number of OTUs were identified as metazoans for 18S rRNA. Furthermore, the18S gene uncovered higher numbers of common species as well as a greater number of overall species when compared to COI [8]. Notwithstanding, the DNA metabarcoding of soil with primers targeting the 18S rRNA resulted in a high proportion of OTUs returned as non-metazoan hits [8,32]. Thus, currently, a set of primers targeting 18S rRNA specific for metazoans is still lacking. Until a few years ago, to study biodiversity associated with environmental samples, most investigators preferred a 454 pyrosequencing platform, as it supplies reads longer than other technologies such as Illumina sequencing. New Illumina platforms and strategies (i.e., Miseq instrument), in combination with 2x300 bp paired-end, allowed sequencing of amplicons with length comparable to that previously obtained by 454 pyrosequencing, with the advantage of higher quality and lower costs per base [2,6,17]. Taking advantage of these developments, Illumina platforms have been recently adopted to characterize soil biodiversity of different taxonomic groups across the whole tree of life, including fungi [29], plants, metazoans, protozoans, and bacteria [12].

The aim of this study was to develop new sets of primers on the 18S rRNA capable of encompassing the wide taxonomic diversity of metazoans inhabiting the soil ecosystem. The new primer sets, suited for Illumina platforms, were compared with previously published primers on four soil typologies.

2. Materials and methods

2.1. Primer design and experimental setup

Two new sets of primers, targeting the 18S rRNA gene, were in silico designed and tested to evaluate their efficiency in recovering soil metazoan diversity. Primers were designed in regions showing variation between metazoan and non-metazoan groups, using Fungi as a reference. The subset included different metazoan taxa: Artropoda (Accession number AY703484, EU432215, AY859604), Annelida (Accession number AF411895, AJ272183), Nematoda (Accession number AY284671, AY284591), Mollusca (Accession number EF489341, FJ917212), Rotifera (Accession number DQ297695, AJ487049), Tardigrada (Accession number FJ435736, DQ839607), Cordata (Accession number AB211066), Platyhelminthes (Accession number AJ270162); and Fungi: Ascomycota (Accession number AY838789, AF258606), Basidiomycota (Accession number AB021676). Small subunit sequences (SSU) from the NCBI database were aligned with BioEdit [13]. The two primer sets, M1041F, M1648R (hereafter Meta1) and M620F, M1260R (hereafter Meta2), were designed using the Primer3 software [31]. These primers were compared with previously published primers targeting the 18S rRNA gene: SSU_FO4, SSU_R22 (hereafter Meta3) [11], and #1, #2_RC (hereafter Meta4) [21] (Table 1).

The performances of in silico primers were evaluated on SSU rRNA SILVA database, by using Test Prime 1.0, allowing for one mismatch (www.arb-silva.de/search/testprime; [16]). Finally, the selected primers (i.e., the newly developed and those previously published) were experimentally compared through metabarcoding of four soil typologies. Three technical replicates for each soil typology and the primer pair were carried out, resulting in a total of 48 samples.

2.2. Sampling and DNA extraction

A total of four soil samples were collected in Northern Italy from cornfield (C), woodland (W), grazeland (G), and heavily grazed pasture (H) after removing vegetation cover. Each sample consisted of soil core of 5 cm diameter and 20 cm of length. After homogenization, the samples were kept at 20 C. DNA was extracted from soil (0.5 g) using Nucleo Spin Soil kit Macherey-Nagel (Düren Germany).

2.3. DNA metabarcoding protocol: PCR amplicons generation, library preparation, Miseq run

DNA was amplified with locus specific primers added to Illumina overhang adapter sequences. Forward overhang: 50 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[locus specific target primer], Reverse overhang: 50 GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG-[locus specific target primer]. The 18S rRNA gene was amplified with the two new primer sets Meta1 and Meta2, and with two previously published primers pairs Meta3 and Meta4. PCRs were performed in 50 ml volumes using Phusion HighFidelity PCR Master Mix (Thermo Fisher Scientific, MA USA) following manufacturer instructions, with 0.2 ml of each primer (100 mM), and 2 ml of genomic DNA (5 ng/ml). For 18S rRNA amplification, the cycling conditions were 30 s at 98 C, followed by 25 cycles of 1 min at 98 C, 30 s at 58 C, 1 min at 72 C, and finally 7 min at 72 C. Finally, amplicons were cleaned-up with Agencourt AMPure XP (Beckman, Coulter Brea, CA) and the sizes were checked with a 2100 Bioanalyser Instrument (Agilent Technologies, Santa Clara, CA). Libraries were prepared with a second stage PCR using Nextera XT Index 1 Primers (N7XX) and Nextera XT Index 2 Primers (S5XX) from the Nextera XT Index kit (FC-131-1001 or FC-1311002), following 16S Metagenomic Sequencing Library Preparation protocol (http://www.illumina.com/content/dam/illuminasupport/documents/documentation/chemistry_documentation/ 16s/16s-metagenomic-library-prep-guide-15044223-b.pdf) with some modifications. The libraries obtained were quantified by Real Time PCR with KAPA Library Quantification Kits (Kapa Biosystems, Inc. MA, United States) pooled in equimolar proportions and sequenced, MiSeq Reagent kit v3 (600 cycles), with paired-end reads of 300 bp.

2.4. Bioinformatics data processing and analyses

Illumina raw reads were trimmed using Trimmomatic v0.32 [4] requiring a minimum base quality of 20 (Phred scale) and a minimum read length of 36 nucleotides. Only trimmed reads were included in downstream analysis. For long amplicons with nonoverlapping paired ends, the first paired read was concatenated to the reverse complement of second paired end read, separating

Primer	Sequence (5o 3o)	Target	From
M1041F	AGAGGTGAAATTCTTGGAYCGY	18S rRNA	This Study
M1648R	ACATCTAAGGGCATCACAGAC	18S rRNA	This Study
M620F	GCAGCCGCGGTAATTCC	18S rRNA	This Study
M1260R	TCRGCTTTGCAACTATACTTCC	18S rRNA	This Study
SSU_FO4	GGCTTGTCTCAAAGATTAAGCC	18S rRNA	Fonseca et al. [11]
SSU_R22	GGCCTGCTGCCTTCCTTGGA	18S rRNA	Fonseca et al. [11]
#1	GCTGGTGCCAGCAGCCGCGGYAA	18S rRNA	Machida and Knowlton [21]
#2_RC	GTCCGTCAATTYCTTTAAGTT	18S rRNA	Machida and Knowlton [21]
Primer combination for 18S rRNA amplification			
M1041F, M1648R (Meta1); M620F, M1260R (Meta2); SSU_FO4, SSU_R22 (Meta3); #1, #2_RC (Meta4)			

Table 1

Sequence of primers and their combination for PCR amplification. Illumina Forward and Reverse overhangs were attached to the primers.

them with a single N base. For partial overlapping reads, the consensus was performed using fastq-join tool (code.google.com/p/ ea-utils/). Joined reads were dereplicated according to USEARCH pipeline [9], with a threshold 2 filtering out reads without replicates. For OTUs selection and identification, reads were clustered at 97% identity using VSEARCH version 1.1.3 (TorbjørnRognesgithub.com/torognes/vsearch), following the USEARCH pipeline (Pipeline 1). Each cluster was identified using BLASTN against the Reference Database described below. A standard overall OTU table was generated and converted into BIOM file format [22] using QIIME1.8.0 utilities [5]. A reference database was created using 16S þ 18S QIIME

formatted SILVA database (Silva111, 99% clustered version) [27]. Organism taxonomy was adapted according to QIIME taxonomy standards with custom Ruby scripts. The coverage of metazoan alpha-diversity associated with each sample was investigated through the rarefaction curves obtained by plotting observed species vs. simulated sequencing, using the QIIME pipeline. To compute the number of OTUs shared across technical replicates, the analysis of commonality was performed and visualized through Venn diagrams, with the Venn Diagrams software (http:// bioinformatics.psb.ugent.be/webtools/Venn/). The effects of primer sets, technical replicates, number of reads, and soil typology on both the total number of OTUs and the number of OTUs shared among samples were analysed by linear model procedure (SAS™ package v. 9.1, SAS Institute, Cary, NC, USA).

USEARCH pipeline required an independent clustering step for each primer couple. This made impossible a direct comparison between OTUs obtained from different primer couples (targeting different region of 18S rRNA). To enable this cross-amplicon comparison, we built a reference-based OTU-picking pipeline using the closed reference OTU picking strategy described in QIIME (Pipeline 2), analysing all couples of primers together. Pipeline 2 did not require any clustering steps because it maps all reads against the reference database by using the Burrows-Wheeler Aligner (BWA) algorithm [20]. Each reference sequence was assumed to be a different OTU because the reference database itself is 97% clustered (Silva111, 97% clustered version). Each OTU abundance was computed with Samtools v1.1 as number of mapped reads over the relative reference sequence. Principal Component analyses (PCA) were carried out to provide spatial illustrations of community structure across soils (beta diversity), by using the QIIME pipeline.

3. Results

3.1. In silico analysis

The two newly developed sets of primers encompass the V5-V7 (Meta1) and V4-V5 (Meta2) hypervariable regions of 18S rRNA, whereas Meta4 amplifies the V4-V5 region and Meta3 the V1-V2 region (Fig.1). In silico estimated amplicon fragment size, excluding the Illumina tag sequence(s), for each primer set using 18S rRNA of Apis mellifera (AY703484) genes, resulted in 595 bps, 609 bps, 435 bps and 638 bps, for Meta1, Meta2, Meta3, and Meta4 primers, respectively. The specificity of the newly developed primer sets for Metazoa was guaranteed by the presence of mismatches in the 18S rRNA gene sequence between metazoan and not-metazoan groups, as in the case of Meta2 that shows a higher primer mismatch compared to Meta4 at the 30 end of V4-V5 region in Fungi (Supplementary material S1). The in silico analysis indicated that only Eukaryota sequences were detected by the four primer sets, whereas different specificity in metazoan selection was observed across the tested primers (Table 2). Meta1 and Meta2 recovered the18S sequence mainly for metazoans, whereas Meta3 and Meta4 also targeted a high proportion of non-metazoan sequences. In silico analysis showed a similar efficiency in recovering OTUs belonging to different Metazoa lineages for Meta1, Meta2 and Meta4, whereas Meta3 poorly recovered Nematoda, Platyhelminthes and Rotifera.

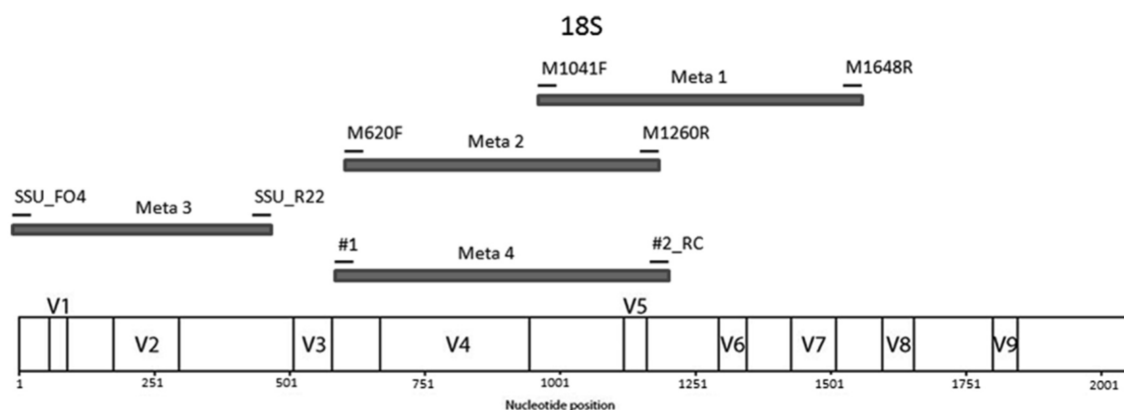

Fig. 1. Relative positions of the newly designed primers and primers described in previous studies depicted on linear maps of 18S rRNA gene. New primers Meta1 (M104F-M164R) and Meta2 (M620F-M1260R), and primers from previous studies Meta3 (SSU_FO4-SSU_R22; [11], Meta4 (#1-#2_RC; [21], were also drawn. Relative positions of variable region (V) based on the secondary structure model of ribosomal RNA genes are also depicted. Figure was adapted from Ref. [21].

Table 2
In silico analysis for the four primer sets: Meta1, Meta2, Meta3, Meta4. For each primer set, the number of hit (matches) to SSU rRNA SILVA database (eligible) and percentage, were determined for Archea, Bacteria and Eukarya (non Metazoa and Metazoa). For Metazoa, the correspondence between (eligible) and (match) for representative taxonomic group was also reported.

	Meta1			Meta2			Meta3			Meta4		
	eligible	match	%	eligible	match	%	eligible	match	%	eligible	match	%
Archaea	69,423	0	0	103,839	0	0	69,833	0	0	103,343	0	0
Bacteria	28,58,901	0	0	28,59,300	0	0	21,27,915	3	0	2859289	3	0
Eukaryota; non Metazoa	332,253	38,748	12	332,504	27,673	8	198,133	109,218	55	390,765	326,887	84
Eukaryota; Metazoa	144,408	115,515	80	146,890	105,630	72	68,063	37,670	55	146,823	126,803	86
Annelida	5593	4795	86	5644	4481	79	3689	3026	82	5644	5364	95
Arthropoda	52,930	46,373	88	53,691	45,206	84	23,731	13,660	58	53,673	46,714	87
Gastrotricha	322	260	81	327	285	87	71	52	73	327	312	95
Mollusca	4567	4185	92	4643	2699	58	3377	2325	69	4643	4429	95
Nematoda	5353	4825	90	5448	3719	68	853	216	25	5438	5049	93
Platyhelminthes	6300	4162	66	6344	3868	61	2668	338	13	6332	5430	86
Rotifera	367	359	98	373	287	77	189	0	0	373	203	54
Tardigrada	363	357	98	363	339	93	114	84	74	363	342	94
Other	68,613	50,199	73	70,057	44,746	64	33,371	17,969	54	70,030	58,960	84

3.2. Primers performance

The MiSeq run generated a total of 19,176,215 paired-end reads. After quality filtration, 17,468,641 reads (91%) were retained with an average of 363,930 reads per sample (range 138,021e1,368,091). Removal of singletons and chimaeras produced on average of 189,535 reads per sample (range 60,534e628,308) that clustered with the reference databases for the 18S rRNA gene. The Meta2 primer set showed the highest metazoan specificity with approximately 100% of reads assigned to metazoan. Meta1 specificity was 62%, whereas this percentage accounted for 25% and 7% of the total reads in the case of Meta3 and Meta4, respectively (Fig. 2, Table 3). The total number of metazoan OTUs differed between primer sets (Supplementary material S2: 3144 OTUs for Meta1; 6876 OTUs for Meta2; 4997 OTUs for Meta3; 549 OTUs for Meta4), and the relative abundance was consistent with soil typology between replicates.

Rarefaction curves of only metazoan OTUs showed that the near-plateau phase was reached only for Meta1 and Meta2 primer sets and that species richness varied among soil types. At similar sequencing depth, Meta3 and Meta4 primers did not fully describe the metazoan complexity in all soil types. For each primer, rarefaction curves showed high consistency across technical replicates (Supplementary material S2).

The linear model analysing the effects of primer sets, technical replicates, number of reads, and type of soil explained 94% of the variance in the number of OTUs ($P < 0.0001$). Primer sets, number of reads, and type of soil significantly affected ($P < 0.005$) the number of OTUs. Conversely, no significant differences were observed in number of OTUs across the three technical replicates.

As shown by Venn diagrams (Supplementary material S3), the percentage of OTUs shared by the three technical replicates varied across primer sets, ranging from 76.7± 2.4% in the case of Meta1 (Meta2: 76.6 ± 2.8%; Meta4: 66.7 ± 12.0%) to 52.6 ± 6.3% in the case of Meta3. The linear model analysing the effects of primer sets, technical replicates, number of reads, and type of soil explained 74% of the variance in the percentage of OTUs shared among samples ($P < 0.0005$). No significant differences in the percentage of OTUs shared among samples were observed across technical replicates and soil typologies. Conversely, the effects of primer sets and number of reads were significant ($P < 0.005$).

3.3. Taxonomic compositions

Taxa distributions of metazoan communities were rather consistent across the four primers sets (Fig. 2). The majority of the recovered OTUs were assigned to eight phyla, with Nematoda, Arthropoda and Annelida being the most represented. OTUs classification reached the order level for Arthropoda, Mollusca, and Platyhelminthes, whereas in the case of Annelida, Gastrotricha, Nematoda, Rotifera and Tardigrada it was possible to reach the level of family. PCA analysis, performed on the OTU table, revealed a clear spatial separation between corn and woodland, whereas grazeland and heavily grazed pasture were closely related (Fig. 3). Within each soil typology, the communities described by Meta1, Meta2 and Meta4 clearly clustered, with those identified by Meta3 somehow shifted (Fig. 3).

4. Discussion and conclusions

The primer sets developed in this study, Meta1 and Meta2, were designed to maximize differences between Metazoa and related eukaryotes in the conserved region of 18S rRNA. In detail, the Meta2 reverse primer targets the same 18S rRNA region previously reported to be specific for nematode and other metazoan, by excluding amplification of fungal or plant DNA [28]. The in silico evaluation of the primer sets showed Meta1 and Meta2 had the best balance between specificity and coverage of Metazoa. These results were confirmed by in vivo evaluation on the four soil typologies. The newly designed Meta2 exclusively amplified Metazoa. Although Meta1 also amplified non-metazoan species (38% of the reads), this percentage is far less abundant than that of the Meta3 and Meta4 primers. Meta2 outperformed Meta3 and Meta4 in the number of OTUs across different soils. The identity of the metazoan OTUs was consistent among replicates for all the tested primers. In this study, we were able to compare the community composition across samples for the different primer sets by ad hoc developed bioinformatics pipelines. For each soil typology, the ecological information is partially consistent across primer couples, especially in the case of Meta1, Meta2 and Meta4. Meta3, which targeted a different hypervariable region, clustered separately. Our results are in agreement with those previously published in the case of bacterial 16S rRNA metabarcoding, where microbiota diversity was significantly affected by the targeted variable regions [7].

In conclusion, the results of this study support 18S rRNA metabarcoding as a powerful and repeatable approach for uncovering the composition and structure of soil metazoan communities. The newly designed primers provide higher specificity and equal or metazoan 18S rRNA primers. Some questions such as the low ressuperior reproducibility in comparison with the previous reported olution of 18S metazoan metabarcoding in distinguishing taxa at lower taxonomic levels still remain open. To improve metazoan biodiversity assessment in soil, a combination of new markers and primers should be considered.

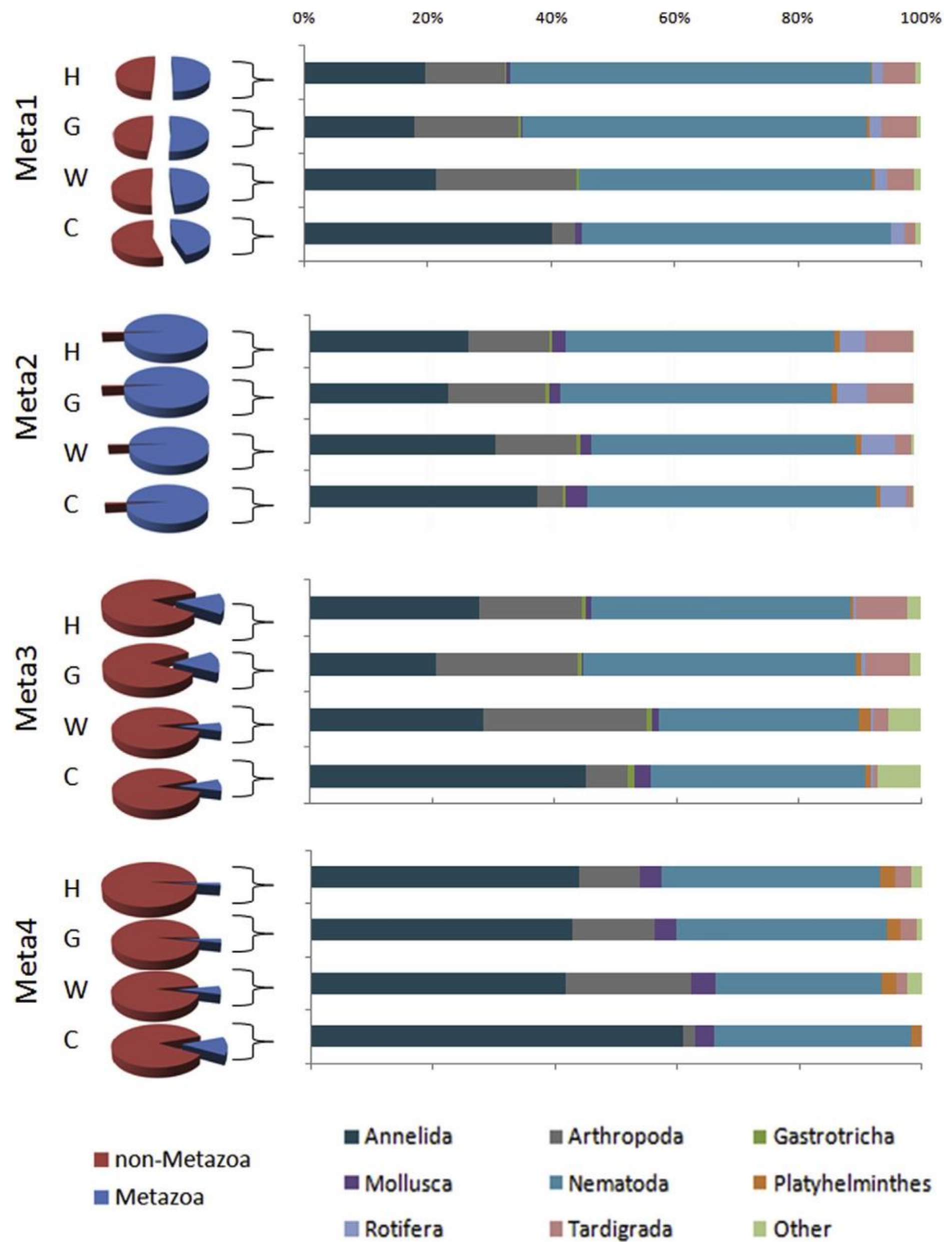

Fig. 2. Percentage of Operational Taxonomic Units (OTUs) recovered from the samples (C: Cornfield, W: Woodland, G: Grazeland, H: Heavy grazed pasture) using different 18S rRNA primer sets (Meta1, Meta2, Meta 3, Meta4). The pie chart indicates the percentage of OTUs assigned to metzoan and non metazoan, whereas the bar chart reports the percentage of OTUs assigned to different taxa for each sample.

Table 3
OTUs assigned to Metazoa (mean between replicates) and percentage of reads assigned to Metazoa (% of replicate mean ± SD) through 18S rRNA metabarcoding obtained with the four primer sets (Meta1, Meta2, Meta3 and Meta4) for different soil samples (C: Cornfield, W: Woodland, G: Grazeland, H: Heavy grazed pasture).

		C	W	G	H
Meta1	Metazoa OTUs	595	830	837	882
	% (Metazoa OTUs/total OTUs)	70.4 ± 2.9	61.0 ± 3.3	48.7 ± 0.9	66.0 ± 1.2
Meta2	Metazoa OTUs	1040	1686	2055	2095
	% (Metazoa OTUs/total OTUs)	99.9 ± 0.0	99.8 ± 0.0	99.5 ± 0.0	99.9 ± 2.1
Meta3	Metazoa OTUs	636	1016	1902	1443
	% (Metazoa OTUs/total OTUs)	40.5 ± 1.2	19.9 ± 0.4	18.8 ± 0.2	19.5 ± 1.1
Meta4	Metazoa OTUs	200	121	126	112
	% (Metazoa OTUs/total OTUs)	14.7 ± 1.4	6.7 ± 0.0	3.1 ± 0.4	2.3 ± 0.7

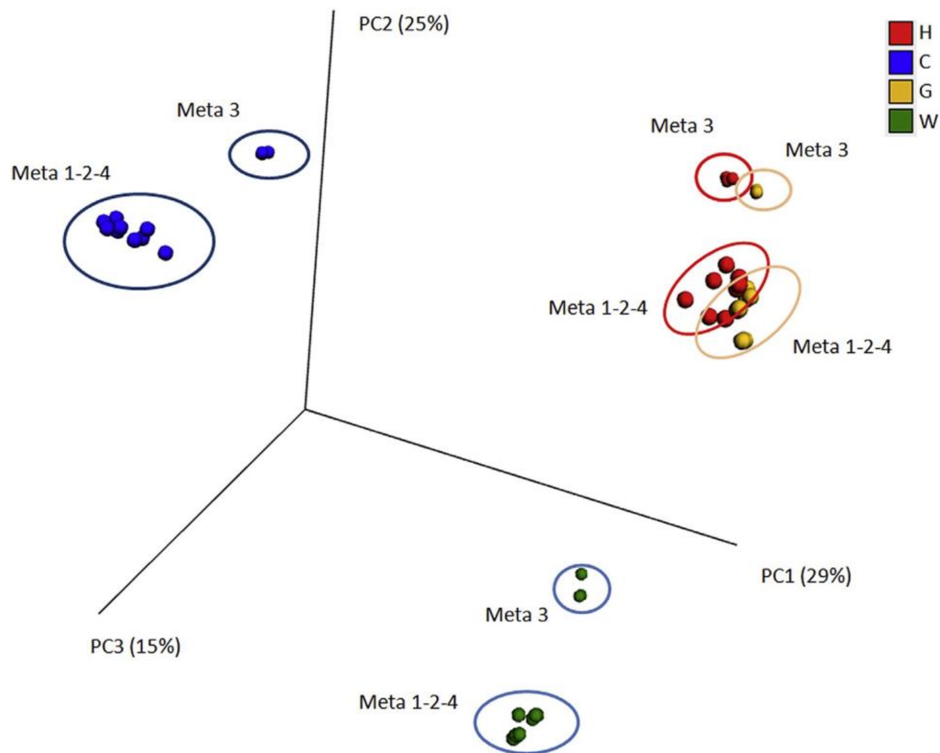

Fig. 3. Principal component analysis (PCA) of community structure (beta diversity) across soils: C: Cornfield (blue), W: Woodland (green), G: Grazeland (yellow), H: Heavy grazed pasture (red). The OTU incidences were obtained for each 18S rRNA primer sets (Meta1, Meta2, Meta 3, Meta4) and technical replicate using the pipeline 2. Ovals are used to help in visualizing compositional differences among sample types.

Acknowledgments : The research was supported by the Agreement Regione Lombardia/CNR, Project FilAgro "Strategie Innovative e sostenibili per la filiera agroalimentare", and by MIUR GenHome project "Technological Resort for the advancement of animal genomic research".

References

[1] D.J. Baird, M. Hajibabaei, Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing, Mol. Ecol. 21 (2012) 2039e2044.

[2] A.K. Bartram, M.D.J. Lynch, J.C. Stearns, G. Moreno-Hagelsieb, J.D. Neufeld, Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads, Appl. Environ. Microb. 77 (2011) 3846e3852.

[3] H.M. Bik, D.L. Porazinska, S. Creer, J.G. Caporaso, R. Knight, W.K. Thomas, Sequencing our way towards understanding global eukaryotic biodiversity, Trends Ecol. Evol. 27 (2012) 233e243.

[4] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for illumina sequence data, Bioinformatics 30 (2014) 2114e2120.

[5] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Pen~a, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D. Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. Mc-Donald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Turnbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of highethroughput community sequencing data, Nat. Methods 7 (2010) 335e336.

[6] J.G. Caporaso, C.L. Lauber, W.A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S.M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J.A. Gilbert, G. Smith, R. Knight, Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms, ISME J. 6 (2012) 1621e1624.

[7] M.J. Claesson, Q. Wang, O. O'Sullivan, R. Greene-Diniz, J.R. Cole, R.P. Ross, P.W. O'Toole, Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions, Nucleic Acids Res. 38 (2010) e200.

[8] D.A. Cowart, M. Pinheiro, O. Mouchel, M. Maguer, J.J. Grall, Mine, S. ArnaudHaond, Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities, PLoS One 10 (2015) e0117562.

[9] R.C. Edgar, B.J. Haas, J.C. Clemente, C. Quince, R. Knight, UCHIME improves sensitivity and speed of chimera detection, Bioinformatics 27 (2011) 2194e2200.

[10] O. Folmer, M. Black, W. Hoeh, R. Lutz, R. Vrijenhoek, DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates, Mol. Mar. Biol. Biotechnol. 3 (1994) 294e299.

[11] V.G. Fonseca, G.R. Carvalho, W. Sung, H.F. Johnson, D.M. Power, S.P. Neill, M. Packer, M.L. Blaxter, P.J. Lambshead, W.K. Thomas, S. Creer, Second-generation environmental sequencing unmasks marine metazoan biodiversity, Nat. Commun. 1 (2010) 98.

[12] S. Giampaoli, A. Berti, R.M. Di Maggio, E. Pilli, A. Valentini, F. Valeriani, G. Gianfranceschi, F. Barni, L. Ripani, V. Romano Spica, The environmental biological signature: NGS profiling for forensic comparison of soils, Forensic Sci. Int. 240 (2014) 41e47.

[13] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucl. Acid. Symp. 41 (1999) 95e98.

[14] P.D.N. Hebert, E.H. Penton, J. Burns, D.H. Janzen, W. Hallwachs, Ten species in one: DNA barcoding reveals cryptic species in neotropical skipper butterfly Astraptesfulgerator, Proc. Natl. Acad. Sci. 101 (2004) 14812e14817.

[15] V.H. Heywood, Global Biodiversity Assessment, Cambridge University Press, Cambridge, 1995.

[16] A. Klindworth, E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, F.O. Glo€ckner, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies, Nucleic Acids Res. 41 (2013) e1.

[17] J.J. Kozich, S.L. Westcott, N.T. Baxter, S.K. Highlander, P.D. Schloss, Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq illumina sequencing platform, Appl. Environ. Microb. 79 (2013) 5112e5120.

[18] P. Lavelle, T. Decaens, M. Aubert, S. Barot, M. Blouin, F. Bureau, P. Margerieb, P. Moraa, J.P. Rossic, Soil invertebrates and ecosystem services, Eur. J. Soil Biol. 42 (2006) S3eS15.

[19] M. Leray, Y.J. Yang, P.C. Meyer, C.S. Mills, N. Agudelo, V. Ranwez, J.T. Boehm, R.J. Machida, A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents, Front. Zool. 10 (2013) 34.

[20] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler Transform, Bioinformatics 26 (2010) 589e595.

[21] R.J. Machida, N. Knowlton, PCR primers for metazoan nuclear 18S and 28S ribosomal DNA sequences, PLoS One 7 (2012) e46180.

[22] D. McDonald, J.C. Clemente, J. Kuczynski, J.R. Rideout, J. Stombaugh, D. Wendel, A. Wilke, S. Huse, J. Hufnagle, F. Meyer, R. Knight, J.G. Caporaso, The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome, Gigascience 1 (2012) 7.

[23] M. Montagna, V. Mereghetti, V. Lencioni, B. Rossaro, Integrated taxonomy and DNA barcoding of Alpine midges (Diptera: Chironomidae), PLoS One 11 (2016) e0149673.

[24] A. Papadopoulou, I. Anastasiou, A.P. Vogler, Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration, Mol. Biol. Evol. 27 (2010) 1659e1672.

[25] M. Pentinsaari, P.D.N. Hebert, M. Mutanen, Barcoding beetles: a regional survey of 1872 species reveals high identification success and unusually deep interspecific divergences, PLoS One 9 (2014) e108651.
[26] D.L. Porazinska, W. Sung, R.M. Giblin-Davis, W.K. Thomas, Reproducibility of read numbers in high-throughput sequencing analysis of nematode community composition and structure, Mol. Ecol. Resour. 10 (2010) 666e676.
[27] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F.O. Glo€ckner, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, Nucleic Acids Res. 41 (2013) D590eD596.
[28] R. Sapkota, M. Nicolaisen, High-throughput sequencing of nematode communities from total soil DNA extractions, BMC Ecol. 15 (2015) 3.
[29] P.A. Schmidt, M. Balint, B. Greshake, C. Bandowa, J. Ro€mbke, I. Schmitt, Illumina metabarcoding of a soil fungal community, Soil Biol. Biochem. 65 (2013) 128e132.
[30] D. Tautz, P. Arctander, A. Minelli, R.H. Thomas, A.F. Vogler, DNA points the way ahead in taxonomy, Nature 418 (2002) 479.
[31] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Fairclothm, M. Remm, S.G. Rozen, Primer3enew capabilities and interfaces, Nucleic Acids Res. 40 (2012) e115.
[32] C. Yang, X. Wang, J.A. Miller, M. de Blecourt, Y. Ji, C. Yang, R.D. Harrison, D.W. Yu, Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator, Ecol. Indic. 46 (2014) 379e389.
[33] D.W. Yu, Y.Q. Ji, B.C. Emerson, X.Y. Wang, C.X. Ye, C.Y. Yang, Z. Ding, Biodiversity Soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring, Methods Ecol. Evol. 3 (2012) 613e623.