INVITED TECHNICAL REVIEW

WILEY | MOLECULAR ECOLOGY RESOURCES

High-throughput identification and diagnostics of pathogens and pests: Overview and practical recommendations

Leho Tedersoo[1] | Rein Drenkhan[2] | Sten Anslan[1] | Carmen Morales-Rodriguez[3] | Michelle Cleary[4]

[1]Natural History Museum and Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia

[2]Institute of Forestry and Rural Engineering, Estonian University of Life Sciences, Tartu, Estonia

[3]DIBAF, University of Tuscia, Viterbo, Italy

[4]Southern Swedish Forest Research Centre, Swedish University of Agricultural Sciences, Alnarp, Sweden

Correspondence
Leho Tedersoo, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia.
Email: leho.tedersoo@ut.ee

Abstract
High-throughput identification technologies provide efficient tools for understanding the ecology and functioning of microorganisms. Yet, these methods have been only rarely used for monitoring and testing ecological hypotheses in plant pathogens and pests in spite of their immense importance in agriculture, forestry and plant community dynamics. The main objectives of this manuscript are the following: (a) to provide a comprehensive overview about the state-of-the-art high-throughput quantification and molecular identification methods used to address population dynamics, community ecology and host associations of microorganisms, with a specific focus on antagonists such as pathogens, viruses and pests; (b) to compile available information and provide recommendations about specific protocols and workable primers for bacteria, fungi, oomycetes and insect pests; and (c) to provide examples of novel methods used in other microbiological disciplines that are of great potential use for testing specific biological hypotheses related to pathology. Finally, we evaluate the overall perspectives of the state-of-the-art and still evolving methods for diagnostics and population- and community-level ecological research of pathogens and pests.

KEYWORDS
community ecology, crop disease, forest pathology, high-throughput sequencing, metabarcoding, metagenomics, molecular diagnostics, plant pathogens, sentinel plantations

1 | INTRODUCTION

Globalization and international trade of plants have greatly accelerated the frequency and magnitude of pest and pathogen invasions to agroforestry systems leading to novel encounters with plant hosts (Lenzen et al., 2012; Liebhold, Brockerhoff, Nuñez, Wardle, & Wingfield, 2017). In some rare cases, these invasive antagonists have caused large-scale transformations of native ecosystems and changed the ecological dynamics through local and regional extinction of native host species (Prospero & Cleary, 2017) and collapses of ancient civilizations (Santini, Liebhold, Migliorini, & Woodward, 2018). In addition, climate change facilitates the probability of establishment of introduced pests and pathogens and promotes range expansion of existing populations (Seidl et al., 2017). Botanical gardens and early warning sentinel systems represent means to identify new and emerging risks to natural plant communities and to improve surveillance globally (Barham, 2016; Vettraino et al., 2015).

Besides economic damage and disease in plants and animals including humans (Seyedmousavi et al., 2018), pathogens and pests play a key role in maintaining diversity and primary productivity in

natural ecosystems (Bagchi et al., 2014; Maron, Marler, Klironomos, & Cleveland, 2011). This Janzen–Connell phenomenon occurs mainly through herbivory or root decay by hexapods or microbial pathogens that are specialized on the dominant plant species and selectively increase their mortality at the seedling stage (Liang et al., 2016).

Traditionally, microbial organisms including pathogens have been identified based on symptoms of disease or culture morphology, whereas detection of pests usually relies on morphological characters of representative individuals. Many obligate intracellular pathogens do not grow in pure culture and never form reproductive structures, which render their detection and identification difficult. Furthermore, both microbial pathogens and animal pests may exhibit high intraspecific variability or comprise cryptic species that may strongly differ in niche and aggressiveness (Ashfaq & Hebert, 2016; Tuda, Kagoshima, Toquenaga, & Arnqvist, 2014). Within biological species, genotypes or races may also differ in their pathogenicity (Barnes et al., 2016; Brasier & Kirk, 2010), sometimes depending on the presence of accessory pathogenicity loci and chromosomes (Möller & Stukenbrock, 2017). These inter- and intraspecific differences emphasize the importance of precise detection of the organisms at the level of species and pathotypes or strains therein.

Rapid and accurate identification of pathogenic microorganisms and pests is essential for detection and employment of appropriate mitigation measures (Comtet, Sandionigi, Viard, & Casiraghi, 2015). Since the early 1990s, molecular methods brought a revolution into our understanding about the identity and autecology of microbial species and shed light into the population structure and community composition of microbiome, including pathogens (Abdelfattah, Malacrinò, Wisniewski, Cacciola, & Schena, 2018; Grünwald, McDonald, & Milgroom, 2016; Hyde et al., 2013; Pace, 1997). Several reviews provide an overview about use of the early molecular techniques for the identification of pests, biocontrol agents and microbial pathogens (Gherbawy & Voigt, 2010; Kashyap, Rai, Kumar, Chakdar, & Srivastava, 2017b; Levesque, 2001; McCartney, Foster, Fraaije, & Ward, 2003; Sankaran, Mishra, Ehsani, & Davis, 2010; Schaad, Jones, & Chun, 2001). Information about more recent high-throughput methods and analysis protocols is highly scattered, with emphasis on the overall concept (Bik et al., 2012), genetic markers (Mendes, Garbeva, & Raaijmakers, 2013), specific methods (Oulas et al., 2015) or particular organism groups such as plants and animals (Deiner et al., 2017; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012), insects (Wachi, Matsubayashi, & Maeto, 2018), fungi (Dickie & St. John, 2016; Nilsson et al., 2018; Tedersoo & Nilsson, 2016), bacteria (Lagos et al., 2015; Pollock, Glendinning, Wisedchanwet, & Watson, 2018) and viruses (Adams & Fox, 2016; Mokili, Rohwer, & Dutilh, 2012). Apart from viruses, only Mendes et al. (2013), Geisen (2016) and Abdelfattah et al. (2018) focused on high-throughput identification of prokaryotes, protists and fungi including examples from pathogens.

This article aims to provide an overview about cutting-edge molecular methods available for identification of plant pathogens and pests, focusing on high-throughput identification at the species level, but also visiting genotype- and population-level methods whenever appropriate for diagnosis. Based on research pitfalls and our experience, we offer practical recommendations for the analysis work flow from sampling design through molecular analysis, bioinformatics analysis, taxonomic and functional assignment and statistics (Figure 1). We illustrate the analysis steps and major achievements with examples from pathogens and pests. Finally, we provide a synthesis about the perspectives of population- and species-level analyses for monitoring and efficient diagnostics of pests and pathogens.

2 | THE EMERGING METHODS

In the last 15 years, researchers have taken advantage of the rapid development of high-throughput molecular identification methods to characterize the enormous diversity of microbial life aboveground and belowground. These methods enable identification of thousands of taxa per sample from hundreds of samples simultaneously and facilitate concurrent focus on any groups of organisms and viruses (Bork et al., 2015; Knief, 2014; Mendes et al., 2013; Uroz, Buee, Deveau, Mieszkin, & Martin, 2016). Based on their technical aspects, high-throughput identification methods can be divided into PCR-based quantification methods, hybridization-based methods (e.g., microarrays), second-generation fingerprinting methods (e.g., RAD-seq) and sequence-based methods, for example, metabarcoding, (meta)genomics and (meta)transcriptomics. The first and most influential examples of these methods and their applications in pathogens and pests are concluded in Table 1.

2.1 | Quantification methods

Taxon-specific primers and quantitative PCR (qPCR) methods have been used for two decades to determine the relative and/or absolute abundance of pathogenic organisms (Sanzani, Nicosia, & M.G., Faedda, R., Cacciola, S.O., & Schena, L., 2014; Schena, Nigro, Ippolito, & Gallitelli, 2004). Recent technical advances enable running thousands of sample and template combinations in parallel. For example, Muurinen et al. (2017) performed simultaneous replicated screening for hundreds of antibiotics resistance genes.

Droplet digital PCR (ddPCR) is based on microfluidics technology, which separates the amplification reaction into >20,000 individual droplets and allows absolute quantification of DNA from up to four target organisms or genes simultaneously, with a detection limit of 10^{-5} relative abundance (Hindson et al., 2011). Dreo et al. (2014) showed much greater precision of ddPCR for quantification of two bacterial plant pathogens with optimal and suboptimal primers compared with ordinary qPCR. Currently, ddPCR can be run in 96-well and 384-well plates, but it is technically possible to increase the throughput of samples. It may be also possible to increase the number of fluorescent dyes to be able to multiplex >4 reactions. Overview of the methodology and use in pathology is reviewed in Gutierrez-Aguirre, Rački, Dreo, and Ravnikar (2015).

Quantification of marker or functional genes is possible by spiking approach combined with HTS identification methods. For spiking, known quantity of control DNA or individuals is added to the sample prior to DNA/RNA extraction and the quantity of target organisms
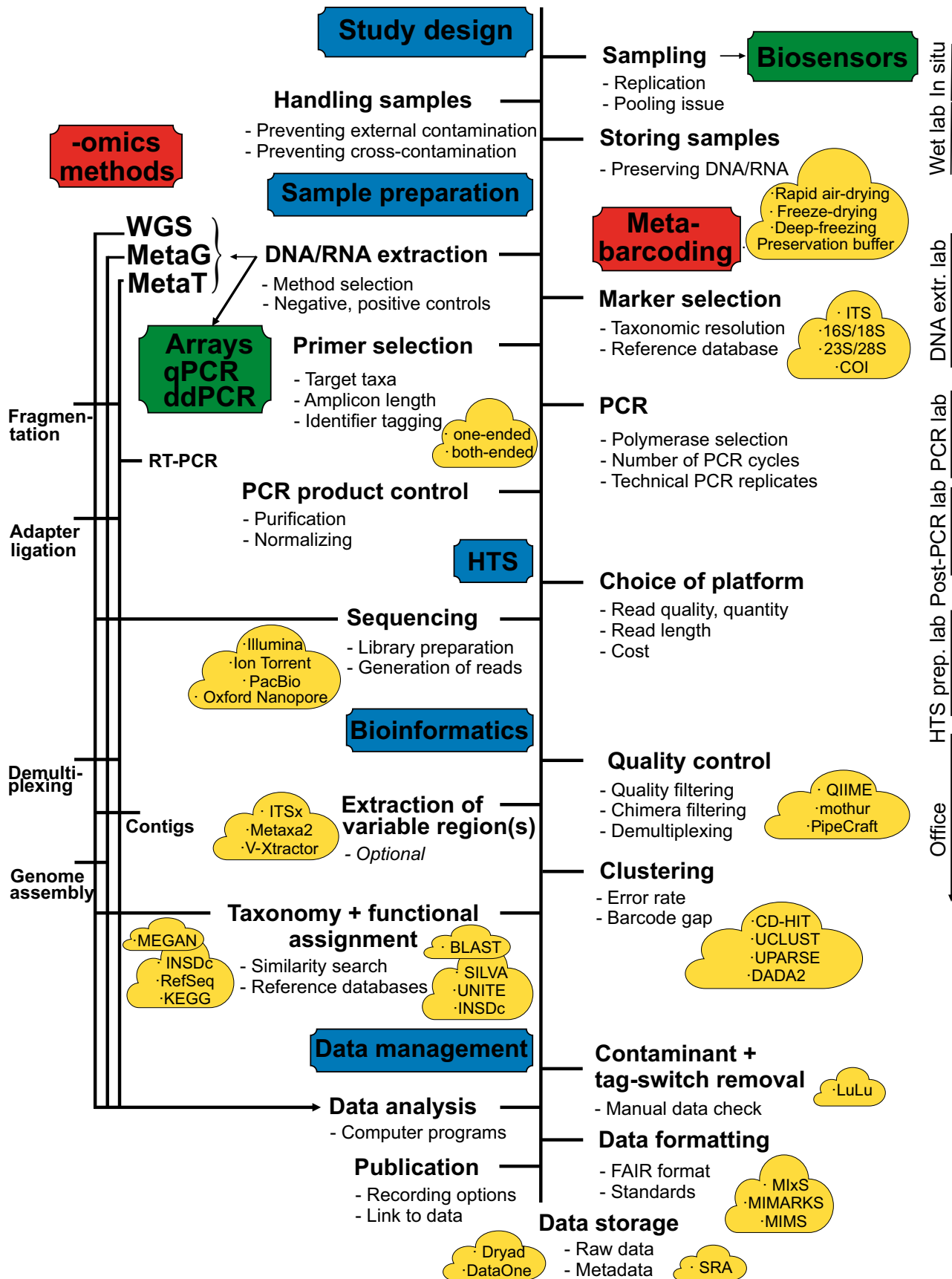
Study design

Handling samples
- Preventing external contamination
- Preventing cross-contamination

-omics methods

Sample preparation

WGS
MetaG
MetaT

DNA/RNA extraction
- Method selection
- Negative, positive controls

Arrays qPCR ddPCR

Primer selection
- Target taxa
- Amplicon length
- Identifier tagging

· one-ended
· both-ended

Fragmen-tation

RT-PCR

PCR product control
- Purification
- Normalizing

Adapter ligation

HTS

Sequencing
- Library preparation
- Generation of reads

· Illumina
· Ion Torrent
· PacBio
· Oxford Nanopore

Bioinformatics

Demulti-plexing

Contigs

· ITSx
· Metaxa2
· V-Xtractor

Extraction of variable region(s)
- *Optional*

Genome assembly

Taxonomy + functional assignment
- Similarity search
- Reference databases

MEGAN
INSDc
·RefSeq
·KEGG

· BLAST
· SILVA
·UNITE
·INSDc

Data management

Data analysis
- Computer programs

Publication
- Recording options
- Link to data

Sampling → **Biosensors**
- Replication
- Pooling issue

Storing samples
- Preserving DNA/RNA

·Rapid air-drying
·Freeze-drying
·Deep-freezing
Preservation buffer

Meta-barcoding

Marker selection
- Taxonomic resolution
- Reference database

· ITS
·16S/18S
·23S/28S
·COI

PCR
- Polymerase selection
- Number of PCR cycles
- Technical PCR replicates

Choice of platform
- Read quality, quantity
- Read length
- Cost

Quality control
- Quality filtering
- Chimera filtering
- Demultiplexing

· QIIME
·mothur
·PipeCraft

Clustering
- Error rate
- Barcode gap

·CD-HIT
·UCLUST
·UPARSE
·DADA2

Contaminant + tag-switch removal
- Manual data check

·LuLu

Data formatting
- FAIR format
- Standards

· MIxS
·MIMARKS
·MIMS

Data storage
- Raw data
- Metadata

· Dryad
·DataOne

· SRA

Wet lab In situ
DNA extr. lab
PCR lab
Post-PCR lab PCR lab
HTS prep. lab
Office

FIGURE 1 Schematic overview of a HTS-based study from sampling through molecular analysis, bioinformatics, publishing and databasing. The most relevant and widely used platforms and methods are indicated [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Examples of high-throughput identification studies that focus on or include pathogens or pests

NGS technology	Sampling area, country	Sampled host species and substrate	Targeted taxa	Target DNA marker	Primers	Main results	Reference
Studies targeting pathogens							
Sanger: MetT	Global	*Homo sapiens*: faeces	RNA viruses	All	None	Plant viruses prevail; humans may act as vectors	Zhang et al. (2005)
454: MetB	Pakistan	Orthoptera, Gastropoda: faeces	Plants	trnL (mini)	g + h	Different plant species prevail in diet of animal pests	Valentini et al. (2009)
454: MetT	Poland + Lithuania	*Lycoperiscon esculentum*: leaves	RNA viruses	All	None	Several previously unrecognized viruses detected	Adams et al. (2009)
Illumina: MetT	Peru	*Manihot esculenta*: leaves	SPFMV, SPCSV viruses	siRNA	None	Sequence of whole genomes using siRNA method	Kreuze et al. (2009)
454: MetT	CA, USA	*Vitis vinifera*: leaves	RNA viruses	All	Various	Sequence of whole genomes from known and unknown viruses	Al Rwahnih, Daubert, Golino, and Rowhani (2009)
454: WGS	Unknown	*Phytophthora* 4 spp.: cultures	*Phytophthora*	All	None	Host jumps are followed by rapid genome evolution in repeat-rich regions	Raffaele, Farrer, Cano, Studholme, and MacLean (2010)
454: MetT	OK, USA + Costa Rica	15 angiosperm families: leaves	dsRNA viruses	All	RT–PCR primer	11 known virus families; thousands of novel host-specific viruses	Roossinck et al. (2010)
Illumina: MetB	WA, Australia	17 wild plant species: leaves	ssRNA Viruses	All	Oligo-d (T) primers for RT–PCR	Multiple novel viruses describe	Wylie, Luo, Li, and Jones (2012)
454: MetB	Italy	*Phytophthora*: mock community	Oomycetes: *Phytophthora* spp.	ITS1	ITS6 + ITS7	HTS method suited for detection of *Phytophthora* spp.	Vettraino, Bonants, Tomassini, Bruni, and Vannini (2012)
Ion Torrent: MetG	South Africa	*Eucalyptus grandis*: leaf, petiole, twig, wood	Fungi	ITS1	ITS1F[a] + ITS2[a]	Community dominated by *Dothidiomycetes* that harbour well-known plant pathogens.	Kemler et al. (2013)
454: MetB	Finland	*Picea abies*: stumps	Basidiomycota	ITS	ITS1F[a] + ITS4B[a]	*Phlebiopsis gigantea* biocontrol does not affect fungal composition, lost 1 year after inoculation; *Heterobasidion* undetected	Terhonen et al. (2013)

(Continues)

TABLE 1 (Continued)

NGS technology	Sampling area, country	Sampled host species and substrate	Targeted taxa	Target DNA marker	Primers	Main results	Reference
Illumina: MetG	China	*Hexapoda*: mixed specimens	Hexapoda	None (mtDNA enriched)	None	High-resolution identification of arthropods	Zhou et al. (2013)
Illumina: MetB	Panama	Woody plants: leaves	Bacteria	16S rRNA	799F + 1115R; PCRII_for + PCRII_rev	Communities dominated by a few core microbiome taxa	Kembel et al. (2014)
454: MetB	Martinique, France	Arthropod predators	Metazoa	COI (mini)	Uni-MinibarF1 + R1	Multiple insect pests make a strong contribution to diet	Mollot et al. (2014)
454: MetB	Spain	Soil, water	*Phytophthora*	ITS1	18Ph2F + 5.8S-1R	Greater species and phylogenetic richness in water samples	Catala, Pérez-Sierra, and Abad-Campos (2015)
Illumina: RAD-seq	Germany	*Fusarium graminearum*: cultures	*Fusarium graminearum*	None	None	High levels of divergence in all populations	Talas and McDonald (2015)
Illumina: MetB	Germany	*Arabidopsis thaliana* tissues	Bacteria, Fungi, Oomycetes	16S rRNA ITS1. ITS2	2 pairs for each group	Pathogens regulate microbiome diversity; fungal and oomycete antagonism to bacteria	Agler et al. (2016)
Illumina: MetB	WA, USA	*Populus trichocarpa*: leaves	Fungi	ITS1	ITS1F[a] + ITS2[a]	Detection of species suppressing and facilitating rust infection	Busby, Peay, and Newcombe (2016)
454: MetB	Primorye, Russia	*Fraxinus mandshurica*: leaf, leaflets, rachises	Fungi (incl. *Hymenoscyphus* spp.)	ITS2	gITS7 + ITS4	Fungal composition similar in infected and uninfected leaves; *H. fraxineus* detected from 33% of samples	Cleary et al. (2016)
PacBio: MetaB	Mexico	*Coffea arabica* +*Hemileia vastatrix*: pustules	Fungi	ITS	ITS1F[a] +ITS4	Communities diverse, differ geographically; potential biocontrol agents detected	James, Marino, Perfecto, and Vandermeer (2016)
Illumina: MetT	North America	Cultures of five fungal pathogens	RNA viruses	All	None	Tens of novel mycoviruses that may have biocontrol properties	Marzano et al. (2016)
Illumina: WGS	Variable	Erysiphales spp.: cultures	Erysiphales (Fungi)	None	None	Interspecific hybridization causes raise of novel pathogens	Menardo et al. (2016)
ON: WGS	Guinea	*Homo sapiens*: tissue samples	Ebola virus	None	None	Virus genome sequenced and detected in <60 min.	Quick et al. (2016)
Illumina: MetB	France	*Raphanus sativus*	Bacteria, Fungi	16S, gyrB, ITS1	515F+806b aF64+aR353 ITS1F[a]+ITS2[a]	*Alternaria* infection alters plant microbiome	Rezki et al. (2016)

(Continues)

TABLE 1 (Continued)

NGS technology	Sampling area, country	Sampled host species and substrate	Targeted taxa	Target DNA marker	Primers	Main results	Reference
Illumina: MetB	Estonia	Forest nursery soils	Oomycetes	ITS	ITS1Oo + ITS4ngs	HTS conditions suitable for identification of oomycetes	Riit et al. (2016)
454: MetB	Great Britain	*Quercus* spp.: bark, wood	Bacteria	16S rRNA	341f + 805r	Composition differs in healthy and infected tissue by acute oak decline	Sapp et al. (2016)
454: MetB	SA, Canada	4 crops: roots and soil	Fungi	ITS1	ITS1F[a] + ITS2[a]	Pathogens become dominant after crop rotations with legumes	Bainard et al. (2017)
454: MetB	Spain	*Quercus ilex*: roots, soil	Oomycetes, esp. *Phytophthora*	ITS1	ITS6[a] + 5.8S–1R	*Phytophthora* sp. nov dominates	Catala, Berbegal, Pérez-Sierra, and Abad-Camposa (2017)
454: MetB	Norway	*Fraxinus excelsior*: leaves, petioles	Fungi (incl. *Hymenoscyphus* spp.)	ITS1, ITS2	ITS5[a] + ITS2[a]; gITS7 + ITS4	Fungal communities differ by season and infection rate; few reads of *Hymenoscyphus* spp. in ITS1 data	Cross et al. (2017)
454: MetB	Italy	Bark beetles: *Orthotomicus erosus, Xyleborinus saxesenii*	Fungi	ITS2	ITS3[a] + ITS4	Bark beetles carry pathogenic fungi with exported timber	Malacrino et al. (2017)
454: MetB	Finland	*Betula pendula*: leaves	Fungi	ITS2	fITS7[a] + ITS4	No effect of tree diversity and neighbourhood on pathogens	Nguyen et al. (2017)
Ion torrent: ddRAD sequences	Multiple	*Hymenoscyphus fraxineus*: ascocarps, isolates; *H. albidus*: isolates	Fungi: *H. albidus, H. fraxineus*	None	None	High genetic variation in E Asia compared to Europe; E Russia a likely source area for Europe; possibility to assign strain to a population	Sønstebø et al. (2017)
Illumina: MetB	France	*Arabidopsis thaliana*: tissues	Bacteria	gyrB	aF64+aR353	Pathogens co-occur in diseased plants, differ in roots and leaves and seasonally	Bartoli et al. (2018)
PhyloChip	Netherlands	*Arabidopsis thaliana*: roots	Bacteria, Fungi	None	None	Pathogen-induced root microbes induce systemic resistance in offspring	Berendsen et al. (2018)
Illumina: MetB	Slovenia, Germany	*Centaurea* spp.: rhizosphere	Nematodes	18S	3NDf + 1132rmod	MetB outperforms qPCR in nematode identification	Geisen et al. (2018)
ON: MetG	New Zealand	*Rattus norvegicus*: gut	All	None	None	Diet consists of various plants and insects incl. pests	Pearman et al. (2018)

(Continues)

TABLE 1 (Continued)

NGS technology	Sampling area, country	Sampled host species and substrate	Targeted taxa	Target DNA marker	Primers	Main results	Reference
Studies including pathogens							
454: MetT	Europe	Agricultural soils	All (esp. ammonia oxidizers)	None (AmoA)	None (various for AmoA)	Archaea are predominate ammonia oxidizers	Leininger et al. (2006)
454: MetB	Various	Various soils	Bacteria, Archaea	18S	787f + 1492rm	Richness and composition differ in agricultural and forest soils; pathogens undistinguished	Roesch et al. (2007)
454: MetG	China	Vector insect	*Candidatus Liberibacter asiaticus*	None	None	Plant pathogen genome recovered from mixed DNA samples	Duan et al. (2009)
454: MetB	France	6 tree plantations: soil	Fungi	ITS1	ITS1F[a] + ITS2[a]	Putative pathogen *Ceratobasidiaceae* sp. dominant in all sites	Buee et al. (2009)
454: MetB	KS, USA	*Quercus macrocarpa*: leaves in rural versus urban sites	Fungi	ITS1	ITS1F[a] + ITS2[a]	Both pathogens and endophytes common, habitat effect	Jumpponen and Jones (2009)
454: MetB	Americas	Soil	Bacteria	16S	27F + 338R	pH drives soil bacterial composition on a continental scale; pathogens undistinguished	Lauber, Zhou, Gordon, Knight, and Fierer (2010)
454: MetB	FL, USA + unknown	Individuals extracted from soil	Nematoda	18S; 28S	NF1 + 18Sr2b[a]; D3a + D3b	Two markers recover nearly all taxa	Porazinska et al. (2009)
454: MetB	Costa Rica	Forest soil	Eukaryotes	18S; 28S	SSUF04 + SSUR22; NF1 + 18Sr2b[a]	HTS of 18S and 28S rRNA genes can be used for eukaryote diversity studies	Creer et al. (2010)
454: MetB	Global	Indoor dust	Fungi	28S	LROR-F + LR5F	Indoor dust reveals predominance of microfungi incl. plant pathogens	Amend, Seifert, Samson, and Bruns (2010)
454: MetB	Americas	Soil	Eukaryotes (esp. protists)	18S	F515[a] + R1119[a]	Distribution of protist phyla depends on climate; pathogens undistinguished	Bates et al. (2013)
454: MetB	KS, USA	Individuals extracted from soil	Nematoda	18S	NF1 + 18Sr2b[a]	Quantification of taxa semiquantitative at best	Darby, Todd, and Herman (2013)
Illumina: MetB	NY, USA	Green roof and city park soil	Fungi	ITS1	ITS1F[a] + ITS2[a]	Fungal composition differs, pathogens undistinguished	McGuire et al. (2013)
Illumina: MetB	Germany	Grassland soil	Fungi	ITS1	ITSFl2[a] + ITS2[a]	Illumina HTS can be used to recover fungal diversity	Schmidt et al. (2013)

(Continues)

TABLE 1 (Continued)

NGS technology	Sampling area, country	Sampled host species and substrate	Targeted taxa	Target DNA marker	Primers	Main results	Reference
454: MetB	NC, TN, USA	3 tree species: roots	Bacteria, Fungi	16S, 18S, ITS, 28S	various	Soil origin explains endophytes better than host species; pathogens commonly detected	Bonito et al. (2014)
454: MetB	Portugal	Vitis vinifera: leaves	Bacteria, Fungi	16S, ITS2, 28S	V6F, V6R; ITS3[a] + ITS4; D2F[a] + D2R	Fungal richness declines but bacterial richness increases with time	Pinto et al. (2014)
454: MetB	Global	Soil	Fungi	ITS2	ITS3NGSmix + ITS4ngs	Plant pathogen richness peaks in tropics	Tedersoo et al. (2014)
454: MetB	Costa Rica	Reared individuals	Lepidoptera	COI	LepF1 + LepR1	HTS can be used in single-specimen barcoding	Shokralla et al. (2014)
454: MetB	Italy	Olea europaea: fruits, leaves	Fungi	ITS2	ITS3[a] + ITS4	Low diversity and high Colletotrichum abundance in rotten fruits	Abdelfattah, Nicosia, Cacciola, Droby, and Schena (2015)
Illumina: mtMetG	Spain	Coleoptera: bodies	Coleoptera	COI, etc.	None	Larval species and phylogenetic diversity greater in subsoil	Andujar et al. (2015)
Illumina: MetB	AK, USA + Canada	Populus balsamifera: leaves	Fungi	ITS1	ITS1FI2[a] + ITS4	Proportion of pathogens declines northwards	Balint et al. (2015)
454: MetT	Europe	Soil	Eukaryotes	None	None, 18S extracted	High protist diversity incl. many pathogens	Geisen et al. (2015)
Illumina: MetB	USA	Dust	Fungi	ITS1	ITS1F[a] + ITS2[a]	Spore dispersal prediction for pathogens and other fungi	Grantham et al. (2015)
454: MetB	FL, USA	Scolytidae bark beetles: mycangia	Fungi	ITS2	gITS7 + ITS4	Composition differs among species; several yeasts and plant pathogens besides mutualists	Kostovcik et al. (2015)
454: MetB	Italy	Fruit fly: Bactrocera oleae	Fungi	ITS2	ITS3[a] + ITS4	Fruit fly carries spores of both pathogens and biocontrol agents; males and females differ	Malacrino, Schena, Campolo, Laudani, and Palmeri (2015)
Illumina: MetB	Costa Rica	1,010 trapped individuals	Arthropoda	COI	III_LCO1490 + III_C_R; III_B_F + III_HCO2198	Low match between morphology, Sanger sequencing and HTS	Shokralla et al. (2015)

(Continues)

TABLE 1 (Continued)

NGS technology	Sampling area, country	Sampled host species and substrate	Targeted taxa	Target DNA marker	Primers	Main results	Reference
454: MetB	Italy	*Fragaria vesca*: fruits, leaves	Fungi	ITS2	ITS3[a] + ITS4	Fungicides reduce the dominant species *Botrytis cinerea*	Abdelfattah, Wisniewski, Nicosia, and M.G., Cacciola, S.O., & Schena, L. (2016)
454: MetB	Estonia	Forest soil	Eukaryotes	ITS2	ITS3NGSmix + ITS4ngs	Pathogen composition is relatively most affected by plant neighbourhood	Bahram et al. (2016)
Illumina: MetB	Panama	Plantation forest soil	Fungi	28S	NL1 + NL4	Pathogens respond more strongly to vegetation than saprobes and mutualists	Kivlin and Hawkes (2016)
Illumina: MetG	Brazil	Predator arthropods: gut	All	None	None	Insect pests are important food sources	Paula et al. (2016)
454: MetB	Estonia + Finland	Forest soil	Eukaryotes	ITS2	ITS3NGSmix + ITS4ngs	Plant pathogens respond negatively to soil C:N ratio; nematodes have site-specific drivers	Tedersoo, Bahram, et al. (2016)
454: MetB	The Netherlands	Agricultural soils	Fungi	ITS2; ^{13}C enriched	ITS9[a] + ITS4	Pathogen: mycorrhiza ratio of active community declines with succession	Hannula et al. (2017)
Illumina: MetG	Great Britain	*Solanum tuberosum*: rot	All	None	None	Detection of taxonomic and functional profile of potato rot spots	Doonan et al. (2017)
Illumina: MetB	MN, USA	Agricultural soil	Fungi	ITS2	ITS1F[a] + ITS2[a]	Pathogens are relatively more abundant in high-biomass plant communities	Cline et al. (2017)
Ion Torrent: MetB	Greenland	Tundra soil	Fungi	ITS2	fITS7[a] + ITS4	Pathogens were more prevalent in highly stressed habitats	Grau et al. (2017)
Illumina: MetB	North America	*Helianthus annuus*: roots, leaves, seeds	Bacteria, Fungi	16S; ITS1	515F + 806r; ITS1F[a] + ITS2[a]	Proportion of pathogens greater in older than recent varieties	Leff, Lynch, Kane, and Fierer (2017)
Illumina: MetB	NY + MA, USA	Forest soil	Fungi	ITS2	fITS7[a] + ITS4	Pathogen richness is greater in *Alliaria petiolata* infested sites	Anthony, Frey, and Stinson (2017)

(Continues)

TABLE 1 (Continued)

NGS technology	Sampling area, country	Sampled host species and substrate	Targeted taxa	Target DNA marker	Primers	Main results	Reference
454: MetB	Estonia	*Pinus sylvestris* forest soil	Eukaryotes	ITS2	ITS3NGSmix + ITS4ngs	Plant pathogen richness increase at higher soil moisture	Hiiesalu, Bahram, and Tedersoo (2017)
PacBio: MetB	Papua New Guinea + Estonia	Forest and nursery soil	Eukaryotes (esp. fungi, oomycetes)	18S, ITS, 28S	Multiple	Longer amplicons: higher resolution, less artefactual taxa; full ITS-based identification of oomycetes	Tedersoo et al. (2017)
Illumina: MetB	S Europe	Bat *Miniopterus schreibersii*: faeces	Animals	COI, mt16S	ArtF1c + ArtR2c; Coleop16Sc + Coleop16Sd	Pests constitute >50% of prey	Aizpurua et al. (2018)
Illumina: MetB	SA, Australia	Soil in successional habitats	Fungi	ITS (ITS1 analysed)	ITS1F[a] + ITS4	Pathogen abundance declines with ecosystem naturality	Yan et al. (2018)

MetB, metabarcoding; MetG, metagenomics; metT, metatranscriptomics; ON, Oxford Nanopore; WGS, whole-genome sequencing.
[a]Poorly performing primer(s).

or genes is detected based on the relative amount of obtained sequences (Pochon, Bott, Smith, & Wood, 2013; Tkacz, Hortala, & Poole, 2018). In theory, spiking allows absolute quantification of the DNA marker content of any pathogenic organism or gene, but this method has been little tested thus far. Differences in sequence length, G + C/A + T content, DNA secondary structure, etc. (see Technical biases below), may all affect accuracy of the spiking approach.

2.2 | Microarrays

Microarray technology is based on accommodation of multiple target-template hybridization reactions onto small chips using robotics technologies to generate arrays and perform multiple hybridization reactions simultaneously. Microarrays have been widely used for species diagnosis, detection of functional genes and gene expression (Sessitsch et al., 2006). Diagnostic microarrays were the earliest high-throughput identification methods that enabled targeting specific pre-selected taxa of viruses, bacterial and fungal pathogens and pests at species level (Lee et al., 2013; Szemes et al., 2005; Wilson et al., 2002). The first diagnostic (macro)arrays for selected plant pathogenic fungi and oomycetes included just >10 species and enabled quantification of pathogens present at low abundance (<0.1% in mixed DNA samples; Lievens et al., 2005; Szemes et al., 2005). By combining classical antagonism tests and high-density microarrays comprising >106 probes, Mendes et al. (2011) identified rhizosphere microbial taxa and particular genes that are antagonistic to a fungal pathogen *Rhizoctonia solani*. PhyloChip-based analyses revealed that plant infection by pathogenic oomycete *Hyaloperonospora arabidopsidis* enhanced growth of rhizosphere microbes, which triggered systemic resistance and reduced damage in the plants' offspring (Berendsen et al., 2018).

While early microarrays used PCR-amplified templates, fine tuning of sensitivity enabled to detect taxa from genomic DNA (DeAngelis et al., 2011). Microarrays also enable to detect single nucleotide polymorphisms (SNPs), which allow genotyping of plant pathogens and detection of aggressive strains (Lievens, Claes, Vanachter, Cammue, & Thomma, 2006). Although reusable microarrays are cheap to run, provide highly sensitive results rapidly and suffice for monitoring the presence and abundance of specific pathogenic taxa and pathogenicity-related genes from complex samples, their major disadvantage is missing the large proportion of species and functions present in the targeted environment and non-optimal stringency in various probes (Sessitsch et al., 2006). Therefore, microarrays have been replaced by high-throughput sequencing (HTS) methods in the last decade.

2.3 | HTS methods for identification of species

High-throughput sequencing represents several new and emerging technologies that fundamentally differ in their ways of recording nucleotides. Furthermore, these methods exhibit substantial differences in throughput, read length, accuracy and technical biases

TABLE 2 Cons and pros of sequencing platforms

Platform	Read length (average: max; kb)	Error rate (%, per bp): main issues	Throughput (10^6 reads)	Cost (run: library prep.; EUR)	Optimal (suboptimal) use
454: GS-FLX (discontinued)	0.7–1:1	0.1: homopolymer indels, end	1.2 (run)	5,000:200	MetB (MetG, WHG)
Ion Torrent: PGM	0.4:0.45	0.5–1.5: homopolymer indels, start, end	5 (chip 318)	1,000:150	MetB (MetG, WHG)
Ion Torrent: Gene Studio S5[a]	0.5–0.6:0.6	0.5–1.5: homopolymer indels, start, end	12 (chip 530)	n.d.: n.d.	MetB (MetT, MetG, WHG)
Illumina: MiSeq	2 × 0.3 (paired-end: 0.58)	0.01–0.1: substitutions, end	20 (lane)	1,500:50–100	MetB (MetG, MetT, WHG)
Illumina: HiSeq	2 × 0.25 (paired-end: 0.48	0.01–0.1: substitutions, end	300 (lane)	4,500:50–100	MetG, MetT, WHG (MetB)
BGISEQ-5002	2 × 0.15 (paired-end 0.28), single-end: 0.4	0.08–0.5: long switch	600 (lane)	n.d.: n.d.	MetG, MetT, WHG
PacBio: Sequel	30:100	13 (raw); <0.1 (10× consensus): homopolymer indels	0.4 (SMRT cell)	1,500:300	MetB, WHG
Oxford Nanopore: MinION	10–100:800	15 (raw); 3–5 (bidirectional): various	0.1–0.35 (flow cell =run)	500:50	WHG (MetB)
Oxford Nanopore: PromethION[a]	n.d.: n.d.	15 (raw); 3–5 (bidirectional): various	100 (run)	n.d.: n.d.	WHG (MetB, MetG, MetT)

[a]Service not available as of October 2018; hence, the values are approximate [b]Service offered through collaborative contract as of October 2018; prices and terms negotiable

(Knief, 2014; Reuter, Spacek, & Snyder, 2015). During the first five years in the market, HTS platforms usually evolve rapidly in terms of throughput, data quality and reduction in analytical costs, but technical constraints become limiting soon thereafter. Fundamentally new HTS methods are announced almost every year, but a fraction of these gain public attention and approximately half of those appear in the market (Heather & Chain, 2016). Table 2 provides an overview of widely used HTS platforms.

The first commercially available HTS method, 454 pyrosequencing (Roche Diagnostics, Basel, Switzerland), was developed in early 2000s. The 454 technology was >100-fold cheaper (10^{-2} EUR/read) than Sanger sequencing, and the analysis chemistry was rapidly optimized to provide high-quality reads from 50 to 700–1,000 bases at 1.2 million read throughput (Reuter et al., 2015). The 454 technology was rapidly adopted by microbial ecologists who performed ground-breaking discoveries about the ultra-high diversity of prokaryotes (Leininger et al., 2006; Sogin et al., 2006). Anecdotally, much of the diversity turned out to be analytical artefacts, indicating the need for careful quality control and optimization of both sample preparation and analytical steps (Huse, Huber, Morrison, Sogin, & Mark Welch, 2007). Separation of artefacts from rare taxa is still the greatest issue of all HTS technologies. Soon after these pioneering prokaryote studies, 454 pyrosequencing was implemented to identify eukaryotes and to separate potentially pathogenic taxa from other guilds based on taxonomic information from indoor environment, animal samples, soil and foliage (Buee et al., 2009; Jumpponen & Jones, 2009; Luna et al., 2007; McKenna et al., 2008; Wegley, Edwards, Rodriguez-Brito, Liu, & Rohwer, 2007). Several years after implementation, the 454 method was used to identify macroorganisms such as plants from mammal and hexapod diet (Valentini et al.,

2009) and animals including parasitic nematodes and other pests (Creer et al., 2010; Porazinska et al., 2009; Table 1).

The Illumina (www.illumina.com) and Ion Torrent (www.iontorrent.com) technologies replaced 454 in the early 2010s because of greater throughput at lower costs. Nonetheless, the Ion Torrent is continuously haunted by short read length (up to 450 bp) and fluctuating sequence quality, which has limited its use in analysis of soil and plant samples (see Kemler et al., 2013). Compared with the 454 platform, the Illumina technology provides up to 3,000-fold greater throughput, several times greater accuracy and possibility to sequence reads of up to 550 bp (2 × 300 paired-end option) at relatively low cost, 10^{-5}–10^{-4}-EUR/read. Generation of self-chimeric sequences and long artefactual inserts or deletions represents the main shortfall of Illumina sequencing compared to other platforms (Tedersoo, Anslan, et al., 2015). At present, Illumina sequencing is by far the best option for short DNA/RNA barcodes and metagenomics, considering sequence quality and analytical costs of library preparation and sequencing (Knief, 2014). It will undoubtedly remain the most widely used HTS method by the end of this decade in spite of only negligible technological improvements since 2015. The ultra-high throughput of Illumina sequencing allows analysis of >1,000 samples in a single run at sufficient sequencing depth (Zinger et al., 2017). Illumina sequencing revealed that growing rotations of legume crops greatly increase the pathogen load in soils, with several-year legacy effects (Bainard et al., 2017). Cline et al. (2017) showed that relative abundance of soil pathogens increases with plant biomass in grasslands. In 2015, a paired-end ultra-HTS platform BGISEQ (www.seq500.com/en/), which is similar to the Illumina platform, was released. So far, it has been used for metagenomic detection of human pathogens (Cheng et al., 2018). Given its shorter read

length (2 × 150 paired-end or 400 bases single end), it is currently suboptimal for amplicon-based detection and identification of organisms.

As a major technical advance, much longer DNA fragments spanning tens of kilobases can be sequenced using the Pacific Biosciences (PacBio, www.pacificbiosciences.com) and Oxford Nanopore (www.nanoporetech.com) technologies, which became commercially available in 2011 and 2015, respectively. However, both of these platforms have very high initial error rates (10%–15% per base) that have improved only marginally in the last years. In the PacBio platform, circularized DNA molecules are sequenced multiple times, reducing the error rate to a minimum (0.1%) at 9- to 11-fold consensus (Tedersoo, Tooming-Klunderud, & Anslan, 2017). Given the average raw read length of 30 kb, PacBio allows sequencing of up to 5 kb DNA fragments at satisfactory quality (Heeger et al., 2018). Sequencing of long fragments of a single molecule has become attractive in DNA barcoding; for example, Hebert et al. (2017) reported on sequencing the DNA barcode in around 10,000 arthropod specimens simultaneously. In a pioneer study, PacBio was successfully applied to identify potential mycoparasites of the coffee rust, *Hemileia vastatrix*. In general, long fragments greatly improve identification via greater taxonomic resolution of unconserved regions and phylogenetic analysis based on relatively more conserved regions (Schlaeppi et al., 2016; Tedersoo et al., 2017; Wagner et al., 2016). However, both library preparation and sequencing steps of PacBio are relatively expensive (ca. 300 EUR/library, 10^{-2} EUR/read) compared with Illumina sequencing. Thus, PacBio is the method of choice for metabarcodes >550 bp given sufficient funding.

Application of the nanopore technology in pathology and ecology in general suffers greatly from low sequence quality, although sequencing of both strands ($1D^2$ flow cell) and Intramolecular-ligated Nanopore Consensus Sequencing (INC-Seq) technique (Li et al., 2016) have been developed. These advances reduce the error rate to 2%–5%, which is still insufficient for distinguishing among closely related species. The first applications in ecology also stress the need for longer identifier tags and avoiding clustering methods (Benitez-Paez & Sanz, 2017; Kerkhof, Dillon, Häggblom, & McGuinness, 2017; Krehenwinkel et al., 2018). The Oxford Nanopore MinION platform is therefore mostly used as a cheap option to close gaps, resolve long repeats and merge scaffolds in genome sequencing, or perform whole-genome resequencing. For example, Rhodes et al. (2018) sequenced multiple strains of an opportunistic human pathogen *Candida auris* and suggested (unsubstantiated though) Indian origin for the pathogenic strains. Nanopore consensus sequences have been used to acquire long DNA barcodes from multiple arthropod specimens simultaneously (Krehenwinkel et al., 2018). However, routine metabarcode-based identification of fungi and oomycetes with the MinION device suffers from very low proportion of meaningful sequences, extremely common tag-switching events and highly unequal sequencing depth across samples (K. Loit, K. Adamson, R. Drenkhan, M. Bahram, R. Puusepp and L. Tedersoo, unpublished manuscript). In spite of high error rate, the nanopore technology holds a great promise in disease diagnostics due to the low cost of equipment and analysis time of 1–2 days (Quick et al., 2016). The unique direct RNA sequencing option (without cDNA reverse transcription step) is of particular interest, but it requires testing for environmental samples and analytical biases.

2.4 | Metagenomics and metatranscriptomics

PacBio and Oxford Nanopore are sometimes termed as third-generation sequencing technologies because of long reads and possibility of excluding the PCR Step. Also, the Sanger method (Green Tringe et al., 2005) and multiple HTS methods (Frey et al., 2014) have been used for generating amplification-free shotgun metagenomic and metatranscriptomic sequence data sets from DNA and RNA (through reverse transcription) molecules, respectively, to address taxonomic diversity of animal pathogens. Except for viruses (see below), shotgun metagenomic and metatranscriptomic studies specifically targeting plant pathogens are rare, although several projects unintentionally cover pathogens in addition to free-living microbes and their functions (Fierer et al., 2013; Geisen et al., 2015; Hudson et al., 2015; Tedersoo, Anslan, et al., 2015; Tedersoo, Bahram, et al., 2015). More specifically, Doonan, Denman, McDonald, and Golyshin (2017) sequenced the metagenomes of soft rot spots of *Solanum tuberosum* to identify the relative abundance of multiple microbial organisms and their functional potential. The metatranscriptomic approach revealed that root colonization by the pathogenic *Rhizoctonia solani* altered rhizobacterial communities and induced expression of stress-related genes (Chapelle, Mendes, Bakker, & Raaijmakers, 2016). A nanopore technology-based metagenomic study revealed that the invasive *Rattus norvegicus* consume mostly plants and insects in New Zealand (Pearman et al., 2018).

No universal primers exist for viruses, rendering metagenomics and metatranscriptomics the only suitable methods for detecting previously unrecognized viruses (Mokili et al., 2012; Zhang, Breitbart, Lee, Run, & Wei, 2005). Metagenomic- and metatranscriptomic-based identification of viruses has been recently reviewed in Massart, Olmos, Jijakli, and Candresse (2014) and Roossinck, Martin, and Roumagnac (2015) and Adams and Fox (2016). Roossinck et al. (2015) in particular provide information about alternative analysis work flows for single- and double-stranded DNA and RNA viruses. For dsRNA viruses, metatranscriptomics of the 21–24 base fragments of silencing RNA (siRNA) has become a popular identification tool of various viruses due to ease of analysis and high detection capacity of small RNA analysis (Kreuze et al., 2009; Roossinck et al., 2015).

Because of various biases introduced by primer choice and the PCR amplification process, PCR-free technologies offer great promise to molecular identification of organisms, particularly viruses and bacteria. In spite of generating huge amounts of sequence data, shotgun sequencing of full metagenomes or metatranscriptomes is an inefficient approach to taxonomic identification of eukaryotes (Alberdi, Aizpurua, Gilbert, & Bohmann, 2018; Tedersoo, Anslan, et al., 2015; but see Geisen et al., 2015), because only a tiny fraction of the

sequences is likely to originate from relevant marker genes. Furthermore, metagenome and metatranscriptome analyses suffer from several technical problems. Because organisms differ substantially in their AT:CG ratio, genomic fragments with extreme ratios may be disfavoured in the sequence analyses, depending on analysis platform (Shakya et al., 2013). The metagenomic fragments cover random stretches of the marker genes among other genomic regions, rendering it impossible to address species-level taxonomic richness in natural communities (Tedersoo, Anslan, et al., 2015). The marker-based reference databases such as UNITE (Abarenkov et al., 2010) and SILVA (Quast et al., 2013) contain abundant data for relatively short rRNA gene markers, but much less taxa have full-length reference sequences for comparison. The genomic reference databases are underpopulated and may result in misidentifications up to the level of kingdom (Korsakovsky Pond et al., 2009; Pearman et al., 2018; Tedersoo, Anslan, et al., 2015). Nonetheless, metagenomics and metatranscriptomics offer an option for identifying the taxa and characterizing their functional potential simultaneously (Bork et al., 2015; Fierer et al., 2013).

Mitochondrial metagenomics has become a high-quality alternative in biodiversity studies of soil arthropods due to the high abundance and rapid evolution of mitochondria and the lack of PCR bias. Genomic DNA samples can be effectively enriched for mitochondrial products that are typically fragmented and sequenced on Illumina platform (Zhou et al., 2013). Gomez-Rodríguez, Timmermans, Crampton-Platt, and Vogler (2017) found that mitochondrial metagenomics is more sensitive to taxon recovery compared with COI metabarcoding and it allows addressing intraspecific variation and construction of more robust phylogenies. This method revealed greater species richness and phylogenetic diversity of beetles in subsoil compared with topsoil (Andujar et al., 2015). In principle, mitochondrial metagenomics could be applied to any group of eukaryotic organisms that possess these organelles. The main drawbacks include a large variation in the number of mitochondria across organisms and their different life stages as well as poor representation in reference databases (except animals).

3 | HTS METHODS FOR IDENTIFICATION OF INDIVIDUALS

HTS methods can be used to distinguish between individuals when targeting SNPs in rapidly evolving loci, partial genomes or whole genomes (Fuentes-Pardo & Ruzzante, 2017). These comparative genomics methods shed light into the origin, migration pathways, speciation, host shifts, co-evolution, hybridization and horizontal gene transfer of antagonists and enable to detect virulent genotypes along with their underlying genetic mechanisms (Grünwald et al., 2016; O'Hanlon et al., 2018). Population genomics methods can also be used to detect genomic introgression from other species and evolution of new species by polyploidization and hybridization, which is a common mechanism in the rise of novel pathogenic fungi and oomycetes (Restrepo, Tabima, Mideros, Grünwald, & Matute, 2014). Common metabarcoding techniques may be unable to distinguish recent hybrids from parent taxa, because the hybrids usually carry the nuclear marker gene (haploid

organisms) or mitochondrial marker gene (nearly all eukaryotes) from one of the parents or different nuclear alleles from both parents in case of diploid (incl. dikaryotic) and polyploid organisms.

The early HTS population genomics methods focused on the distribution of SNPs in certain variable regions. For example, Isola et al. (2005) used targeted 454 pyrosequencing to detect mutations underlying resistance to a drug in the human pathogen *Mycobacterium tuberculosis*. Subsequent population-level studies targeted multiple genomic fragments flanking restriction sites across the entire fragmented genome, which is termed as restriction site-associated DNA sequencing (RAD-seq; reviewed in Davey et al., 2011). Several examples of using RAD-seq in plant pathogens are given in Grünwald et al. (2016). This method revealed high genetic variability and recombination in a crop pathogen *Fusarium graminearum*, suggesting that these features facilitate rapid adaptation to resistant cultivars and biocides (Talas & McDonald, 2015). RAD-seq also revealed several coexisting groups of a mutalistic fungus *Rhizophagus irregulare*, most of which were globally distributed (Savary et al., 2018).

With plummeting of HTS costs, partial and whole genomes of pathogenic organisms can be readily determined from pure cultures, host tissues and soil environment. The main advantage of whole-genome sequencing (WGS) is the generation of several orders of magnitude more information about polymorphic sites and a better understanding of their linkage and occurrence in exons and introns (Grünwald et al., 2016). Low-coverage genomes and organisms' marker genes and functional genes can be determined from minute DNA concentrations from old herbarium specimens and roots (Tedersoo, Bahram, et al., 2016; Tedersoo, Liiv, et al., 2016; Yoshida, Burbano, Krause, Thines, & Weigel, 2014). Thus, sequencing of genomes from multiple isolates of the same species has become a common practice in microbiology (Liti et al., 2010) and more recently in plant pathology (Menardo et al., 2016; Table 1). Due to much greater genome size and organization of genetic material into multiple chromosomes, eukaryote genomes are more difficult and costly to sequence and assemble compared with these of prokaryotes. Using a shotgun metagenomic approach, Duan et al. (2009) sequenced the genome of an uncultured plant pathogenic bacterium *Candidatus Liberibacter asiaticus* from its psyllid vector. In a population genomics study, Cooke et al. (2012) detected recent evolution of a highly virulent genotype group within *Phytophthora infestans* and its genetic mechanisms of overcoming hosts' resistance. WGS of the malaria parasite *Plasmodium viviparum* revealed substantial population divergence and endemicity on a global scale (Manske et al., 2012). WGS analyses using the pocket sequencer Oxford Nanopore MinION are becoming a routine for "real-time" identification of strains and cryptic species in pathogenic complexes, which has been demonstrated for *Escherichia coli* and the Ebola virus (Loman, Quick, & Simpson, 2015; Quick et al., 2016).

Optimization of sample preparation and HTS protocols has enabled WGS of single cells (Lasken & McLean, 2014). McLean et al. (2013) sequenced the genome of an opportunistic human pathogen *Porphyromonas gingivalis* from several individual cells in parallel. Due to the small genome size, single-cell sequencing may perform well

for bacteria and archaea, but it is more problematic in eukaryotes. Single-cell WGS revealed that human neurons differ in partial genome copy number and some cells have multiple mutations in specific regions (McConnell et al., 2013). Using microfluidics-based single-cell preparation and sequencing, Gawad, Koh, and Quake (2014) determined SNPs from leukaemia cells and shed light into the mechanisms of cancer development. Nair et al. (2014) demonstrated the potential of single-cell sequencing in protists with small genomes, showing the complexity of infection by *Plasmodium* spp. in human patients and efficiency of this method in distinguishing highly virulent strains. Single-cell WGS is yet to be applied to plant pathogens, but it has greatest perspective in understanding the function of unculturable unicellular pathogens such as members of the early diverging fungal lineages and alveolates (Ahrendt et al., 2018).

4 | TARGETING ACTIVE ORGANISMS

DNA molecules are typically long-lived and remain detectable for several weeks to months *post mortem*, depending on fragment length and habitat properties. Short (<200 bp) DNA fragments may be preserved for millennia in anoxic sediments and permafrost (Allentoft et al., 2012). Conversely, rRNA (including the ITS region) and mRNA of functional genes provide insights into the genes transcribed in a time frame of a few hours to a few days, effectively excluding organisms that are dormant (eggs, spores, sclerotia, cysts, etc.; Rajala, Peltoniemi, Hantula, Mäkipää, & Pennanen, 2011). However, sampling for RNA requires extra care and rapid pre-treatment such as freezing in liquid nitrogen or storing in RNA preservation buffers, which is costly and sometimes inefficient (Rissanen, Kurhela, Aho, Oittinen, & Tiirola, 2010). Targeting RNA is, nonetheless, unavoidable in pathological studies of RNA viruses.

Alternative to RNA, microbial ecologists have used ^{13}C (Radajewski, Ineson, Parekh, & Murrell, 2000) and the nucleotide analogue 3-bromo-deoxyuridine (Hanson, Allison, Bradford, Wallenstein, & Treseder, 2008) incorporation into substrate and assimilation into DNA of organisms that metabolize these enriched substrates. The ^{13}C-based stable isotope probing (SIP) is difficult to perform in natural conditions because of multiple carbon sources diluting the isotopic signal. This is especially relevant for eukaryotic pathogens that grow and accumulate ^{13}C or nucleotide analogues into their DNA slowly and may use much of the labelled carbon for respiration. Nonetheless, ^{13}C incorporated into DNA and fatty acids revealed flow of plant-derived carbon through the soil food web and decline in pathogen-to-mycorrhiza ratio during secondary succession (Hannula et al., 2017). This method could be useful when addressing the pests and pathogens that use recent photosynthesis products amongst other organisms or detecting potential biocontrol agents.

5 | TECHNOLOGICAL BIASES OF HTS METHODS

All molecular identification methods suffer from specific analytical biases. Marker bias may select for organisms that exhibit high copy

numbers. Primer bias discriminates against targets that exhibit primer-template mismatches, particularly in the 3′ end of the primer, which reduces their relative amplification efficiency by 1–2 orders of magnitude (Ihrmark et al., 2012; Tedersoo, Bahram, et al., 2015). Primer bias in the ITS region is important in several animal and plant pathogenic fungal groups (Tedersoo & Lindahl, 2016), nematodes and alveolates. PCR bias represents unequal amplification of target species due to differences in AT:CG ratio, DNA secondary structure and marker length (Ihrmark et al., 2012). Some arthropod and fungal groups exhibit introns in rRNA genes or long ITS1 or ITS2 regions, which may render corresponding taxa entirely unrepresented (Tedersoo, Bahram, et al., 2015). The best example concerns the ash dieback disease agent *Hymenoscyphus fraxineus* that exhibits a long 3′ terminal 18S intron, which renders the species undetectable using the ITS1F/ITS0F forward primers (see Cross et al., 2017). Most Oomycota possess a long ITS2 region, which may be discriminated against in studies targeting all eukaryotes (Riit et al., 2016). Method-specific biases may occur in cloning and molecular identifier tag ligation procedures, where variants with highly skewed AT:CG ratios and specific terminal nucleotides may be favoured or discriminated against (Lindahl et al., 2013). Biases may differ among models and analysis chemistry in the same sequencing platform (Tedersoo et al., 2017).

Tag-switches and chimeric molecules are common technical artefacts in HTS analyses. Chimeric molecules usually develop during PCR, when extension is incomplete (due to low processivity, short extension time or depletion of nucleotides) and these short fragments prime the templates during subsequent cycles. Chimeric molecules develop more commonly between two closely related organisms becoming more abundant with increasing number of PCR cycles and community complexity (Aas, Davey, & Kauserud, 2017; Haas et al., 2011). In essence, tag-switch artefacts are also chimeric molecules that develop between multiplexed samples during the post-PCR library preparation step (Schnell, Bohmann, & Gilbert, 2015).

6 | PRACTICAL RECOMMENDATIONS FOR HTS-BASED RESEARCH

6.1 | Design of HTS studies

Study design depends on the objectives of research. Purely descriptive studies with haphazard sample collection and insufficient replication are difficult to publish and not worth the effort, except sequencing genomes or transcriptomes, or validating novel methods. Testing ecological hypotheses requires a proper well-replicated sampling design. Many researchers seem to forget that technical replicates, multiple spatially autocorrelated subsamples and thousands of recovered OTUs do not serve as independent biological replicates (Prosser, 2010). This is particularly relevant in the geographically structured sampling with inherent hierarchical design and multilevel spatial autocorrelation.

One of the main questions in pathological and microbiological research is whether or not to pool subsamples. Pooling may strongly

reduce analytical costs, but it also reduces small-scale resolution. The answer depends on the research objectives, spatiotemporal scale and nature of the samples of the particular study. If individual samples (e.g., leaf, soil core) are small and expected to represent the community poorly, pooling multiple samples is a viable option. In case of hierarchical design (i.e., structured by block, plot or site), it is useful to pool multiple subsamples when the internal variation is not of interest. However, in most other cases, analysis of multiple independent samples is preferable due to the ability to estimate sampling error and address the importance of spatiotemporal variability. HTS analyses can easily recover slight shifts in taxonomic and gene composition; therefore, multivariate techniques require just 3–4 replicates to detect biologically important shifts in community composition (Balint et al., 2016). An extra replicate should be considered, because it is common to obtain low-quality DNA or a limited number of sequences from some (typically 1%–10%) samples. Analysis of richness and diversity measures and pathogen load requires more samples, because univariate tests have lower statistical power.

6.2 | Sample preparation for HTS analysis

HTS techniques are sensitive to spoiling, external contamination and cross-contamination, hence requiring careful collection, handling and pre-treatment to prevent contamination and overgrowth by fast-growing moulds or DNA/RNA degradation (Lindahl et al., 2013). For pre-treatment and storage, rapid air-drying, freeze-drying, deep-freezing and fixing in preservation buffers work equally fine for DNA. Deep-freezing and storage at −80°C works well for potential analysis of RNA, proteins and fatty acids for further analyses (Rissanen et al., 2010). Dried and frozen samples and frozen DNA solution can be usually preserved for decades with minor losses in quality.

To avoid contamination during various steps of analysis, it is recommended to divide the laboratory space into wet laboratory, DNA extraction and PCR laboratory, post-PCR laboratory and HTS preparation laboratory. The PCR products from previous projects may represent the main source of contamination in HTS studies, because a single floating DNA molecule may be progressively amplified and sequenced. This can be controlled by efficient cleaning of laboratories using UV-light and DNase-containing solutions. Negative controls in all stages of analyses enable to detect and track contamination (Lindahl et al., 2013).

DNA or RNA extraction can be performed by using commercial kits or classical protocols developed since 1990s. Because extraction procedures may affect the recovered community composition (Lindahl et al., 2013), optimal protocols should be selected considering the mass, substrate and target organism (Brooks et al., 2015). Samples can be extracted at low overall mass (0.25 g), which recovers nearly comparable taxonomic richness with high-quantity extraction (10 g) for microorganisms (Song et al., 2015). For small extraction quantities, it is relatively more important to thoroughly homogenize the sample. For ultra-long amplicons and genomic analyses, bead beating should be kept at minimum duration. For metagenomic- and RNA-based analyses, it is of particular importance to minimize the concentration of co-extracted humic acids and saccharides that may interfere with downstream processes. The extracted DNA from soil and other organic-rich substrates may require an extra purification step using filter columns or magnetic beads for optimal performance in (meta)genomics analyses (Bahram et al., 2018).

6.3 | Marker and primer selection

For HTS-based diversity analyses, it is very important to thoroughly consider the DNA/RNA marker based on desired taxonomic resolution. For routine community-level analysis, species-level resolution should always be targeted to avoid bulking together pathogenic organisms with closely related endophytes and saprotrophs (Critescu, 2014; Tedersoo & Nilsson, 2016). Nonetheless, strains of the same antagonist species may differ strongly in pathogenicity, which renders barcoding marker-based detection of pathogens somewhat ambiguous. Bacterial species are typically identified based on the 16S rRNA gene, although this marker effectively enables operation at the subgenus and genus level (Yarza et al., 2014). The ITS region including the tRNA gene between 16S and 23S rRNA genes provides much greater taxonomic resolution, but its highly variable length and AT:GC ratio may generate biased views on composition particularly in HTS studies of prokaryotes (Benitez-Paez & Sanz, 2017; García-Martínez, Acinas, Antón, & Rodríguez-Valera, 1999; Ruegger, Clark, Weger, Braun, & Borneman, 2014). Of eukaryotes, both fungi and oomycetes are routinely studied based on the ITS region, which has been established as a formal barcode for the latter group (Schoch et al., 2012). For protist kingdoms and Metazoa (animals), 18S and 28S rRNA genes, ITS and mitochondrial cytochrome C oxidase I subunit (COI) are all used, with preferences depending on particular groups considering taxonomic resolution and available reference material (Pawlowski, Audic, & Adl, 2012). The official animal barcode COI performs poorly in HTS analyses, because of the lack of conserved regions for inclusive primers and loss of primer-template specificity with multiple degenerations (Geller, Meyer, Parker, & Hawk, 2013). The ITS region and 28S offer comparable species-level resolution and allow use of strong primers in most animal groups such as Nematoda, Acari, Collembola and Hexapoda (Anslan & Tedersoo, 2015; Carneiro, Oliveira Lima, & Correia, 2017; Pacheco da Silva, Bertin, Blin, Germain, & Bernardi, 2014; Yang, Cai, & Cheng, 2011). Particularly for the ITS region, however, databases are sparsely populated with species- and genus-level reference sequences from most arthropod, nematode and protist groups, because these have been excluded as classical barcodes at least partly due to difficulties with Sanger sequencing of heterozygous alleles and poor phylogenetic inference.

Guided by choice of a marker gene, selection of primers is important. Given the high sequencing depth, it is wise to focus on organisms in a broad perspective to secure inclusion of nearly all intended target taxa (Tedersoo & Lindahl, 2016). Because taxon-specific primers are rarely all-inclusive, it is therefore recommended to focus on all eukaryotes when addressing fungi, or Stramenopila when targeting Oomycota. For targeting pathogens in the tissue of

plants or any other specific organisms, researchers may consider primers that exclude the host DNA or including additional blocking primers, especially if targeted pathogens are expected to occur at very low relative abundance. Supplied with 3′ terminal nucleotide modifications and in surplus concentration, blocking primers prevent annealing and elongation of host DNA marker by specifically binding to host DNA downstream of regular primers. This procedure is frequently used for mitochondrial DNA analyses (Vestheim & Jarman, 2008), but it is probably impossible to effectively design for nuclear rRNA genes. On the other hand, cotargeting host DNA enables to determine the relative abundance of pathogen DNA marker relative to host DNA marker that is comparable (i.e., without systematic bias) across samples.

Revisiting the coverage of widely used primers designed in early 1990s (White, Bruns, Lee, & Taylor, 1990) has revealed suboptimal performance and multiple unexpected mismatches to several taxa within the targeted group (Klindworth et al., 2013; Tedersoo & Lindahl, 2016; Tedersoo, Anslan, et al., 2015). While 1–2 non-terminal mismatches are not important in sample preparation for Sanger sequencing, a single mismatch may result in underestimation of the taxon by 1–2 orders of magnitude (Ihrmark et al., 2012; Tedersoo, Bahram, et al., 2015). Accumulated sequence data have enabled molecular ecologists to construct more efficient degenerate primers or primer mixes for prokaryotes, fungi, oomycetes and animals (Apprill, McNally, Parsons, & Weber, 2015; Geller et al., 2013; Nilsson et al., 2018; Parada, Needham, & Fuhrman, 2016; Riit et al., 2016; Tedersoo & Lindahl, 2016; Tedersoo, Bahram, et al., 2016; Tedersoo, Liiv, et al., 2016; Toju, Tanabe, Yamamoto, & Sato, 2012). Table 3 provides a selection of high-affinity primers recommended for use in HTS studies.

In analysis of complex pathological systems, food webs and diet of omnivores, organisms from multiple kingdoms can be targeted simultaneously. Strategies for this include a universal marker such as 18S rRNA or ITS for all eukaryotes, or different markers for each kingdom. Universal primers and primer mixes exist for the rRNA markers (Table 3). Different markers of similar length can be analysed in multiplex or mixed after separate amplification into a common library (de Barba, Boyer, Rioux, Coissac, & Taberlet, 2014). However, markers may yield >2 orders of magnitude difference in average sequencing depth (Tedersoo et al., 2017; Tedersoo, Anslan, et al., 2015), indicating that the relative discrimination factor should be considered beforehand. Metagenomic approach has proven a viable alternative for relative quantification of DNA of target organisms in soil (Bahram et al., 2018) and gut contents (Pearman et al., 2018).

One or both primers used for HTS should be tagged with a molecular identifier to enable multiplexing samples. These tags of typically 6–12 bases should differ from each other by at least 4 bases/indels (e.g., the "error-correcting" Golay identifiers; Lundberg, Yourstone, Mieczkowski, Jones, & Dangl, 2013) to prevent random mutations in tags or impure synthesis to erroneously switch sequences among samples. The tagged primers may also include a platform-specific sequencing primer, but such long oligonucleotides may perform poorly (Lindahl et al., 2013). For Illumina sequencing,

96 combinations of Nextera indexes can be ligated by PCR. Primers tagged with identifiers only are cheaper and can be used for analysis employing any sequencing platform, rendering these usable for many years. To reduce the competition among tagged amplicons in the ligation step, it is advisable to select all identifiers to start with the same nucleotide and use a 2-base linker sequence with no match to any of the templates. Identifier tags that have an AT:CG ratio less or more than 0.25–4 tend to perform poorly (Tedersoo & Nilsson, 2016). It is strongly recommended to add identifier tags to both the reverse and forward primers to minimize the tag-switching (Gohl et al., 2016).

6.4 | PCR

Prior to PCR, it is recommended to quantify DNA and use equal amounts of template for each sample to be able to use the same number of PCR cycles across the study (Gohl et al., 2016). The PCR mix should include a high-affinity and high-processivity polymerase (e.g., Pfu, Phusion, Q5) to minimize incorporation of erroneous nucleotides and generation of partial reads that can be converted to chimeric sequences in subsequent extension cycles. These more expensive polymerases strongly reduce the number of chimeric sequences and artificial taxa comprised of error-infested sequences (D'Amore et al., 2016; Gohl et al., 2016). For HTS analysis, the primer annealing temperature could be reduced by ca 5°C compared to regular PCR to promote amplification of templates with one or two mismatches to primers. The number of PCR cycles should be kept at minimum—so that a relatively weak band or smear of suitable size is seen on an agarose gel. Increasing extension time is also likely to reduce incomplete amplification and hence chimera formation (D'Amore et al., 2016; Lindahl et al., 2013). Low input DNA content results in lower amount of inhibitors and less chimeric sequences (D'Amore et al., 2016; Gohl et al., 2016). Due to stochastic variation, it is recommended to use at least two PCR replicates that can be pooled post-amplification (Alberdi et al., 2018; Lindahl et al., 2013; Tedersoo et al., 2010).

Amplicon purification depends on further analyses and choice of sequencing platform. It is recommended to normalize amplicon concentration across samples to reduce variation in sequencing depth among samples several-fold (Lindahl et al., 2013). The equimolarly mixed amplicons are subjected to platform-specific adapter ligation in the library preparation step. It is recommended to order library preparation from a sequencing service provider to secure their quality standards and leave the risk of failure to the service provider. Researchers should check the quantity and quality requirements from each service provider, because these may differ greatly. The quantity appears to be negotiable, because service providers usually request 5–10 times more material than they use. Due to high demand, it takes typically 1–2 months to receive the sequences. It does not pay off to order bioinformatics service, because companies provide standard quality,.fasta- and.fastq-formatted files. These can be handled using custom options in any bioinformatics platform, whereas the service provider's analysis routine may be suboptimal

TABLE 3 Recommended primers for HTS-based identification of pathogens and pests

Primer[1]	Marker: direction: position	Sequence	Target	Comment	Reference
Prokaryotes					
GM3F[a]	16S: fwd: 8	AGAGTTTGATCMTGGC	Bacteria	excl. minor phyla, Archaea	Muyzer, Teske, Wirsen, and Jannasch (1995)
515fB[b]	16S: fwd: 515	GTGYCAGCMGCCGCGGTAA	Prokaryotes	Microbiome projects	Parada et al. (2016)
806rB[b]	16S: rev: 806	GGACTACNVGGGTWTCTAAT	Prokaryotes	Microbiome projects	Apprill et al. (2015)
926r[b]	16S: rev: 926	CCGYCAATTYMTTTRAGTTT	Prokaryotes	Microbiome projects	Parada et al. (2016)
GM4R[a]	16S: rev: 1492	TACCTTGTTACGACTT	Bacteria	excl. minor phyla, Archaea	Muyzer et al. (1995)
SSU1492Fngs[c]	ITS: fwd: 1492	GTCGTMACAAGGTANCCG	Prokaryotes	excl. minor groups	This paper
189r[c]	ITS: rev: 189	TACTDAGATGTTTCASTTC	Bacteria	excl. minor phyla, Archaea	Hunt et al. (2006)
Eukaryotes (general)					
TAReuk454FWD1[d]	SSU: fwd: 565	CCAGCASCYGCGGTAATTCC	Eukaryote	BioMarks primer	Stoeck et al. (2010)
TAReukREV3[d]	SSU: rev: 970	ACTTTCGTTCTTGATYRA	Eukaryote	BioMarks primer	Stoeck et al. (2010)
1389F[e]	SSU (ITS): fwd: 1630	TTGTACACACCGCCC	Eukaryote, Prokaryote		Amaral-Zettler, McCliment, Ducklow, and Huse (2009)
ITS9MUNngs[e]	SSU (ITS): fwd: 1635	TACACACCGCCCGTCG	Eukaryote	No prokaryotes	Tedersoo and Lindahl (2016)
1510R[de]	SSU: rev: 1780	CCTTCYGCAGGTTCACCTAC	Eukaryote	5′ poor, excl. Sordariomycetes, some nematodes	Amaral-Zettler et al. (2009)
1510Rngs[de]	SSU: rev: 1780	WCBGCDGGTTCACCWAC	Eukaryote	No prokaryotes	This paper
ITS4ngsUni[efghi]	LSU (ITS): rev: 40	CCTSCSCTTANTDATATGC	Eukaryote	No prokaryotes	Tedersoo and Lindahl (2016)
TW13[efghi]	LSU: rev: 645	GGTCCGTGTTTCAAGACG	Eukaryote	Some prokaryotes; excl. dictyostelids	T.J. White, unpublished
TW14ngs[efghi]	LSU: rev: 960	CTATCCTGRGRGAAAYTTC	Eukaryote	excl. Microsporidea	Tedersoo and Lindahl (2016)
Fungi					
ITSOF-T[f]	SSU (ITS): fwd: 90	ACTTGGTCATTTAGAGGAAGT	Fungi	excl. Mucoromycota, Saccharomycetes	Tedersoo et al. (2008)
ITS2ngs[f]	5.8S(ITS1): rev: 35	TTYRCKRCGTTCTTCATCG	Fungi	excl. minor groups	Tedersoo et al. (2017)
gITS7ngs[g]	5.8S (ITS2): fwd: 70	GTGARTCATCRARTYTTTG	Fungi	excl. minor groups	Tedersoo and Lindahl (2016)
LR5-Fung[g]	LSU: rev: 880	CGATCGATTTGCACGTCAGA	Fungi, Metazoa, Stramenopila	Suited for living plant samples	Tedersoo et al. (2008)
Oomycetes					
ITS1Oo[h]	18S (ITS): fwd: 1795	GGAAGGATCATTACCACAC[2]	Oomycetes	incl. some Ochrophyta	Riit et al., (2016)
ITS3Oo[i]	5.8S (ITS2): fwd: 150	AGTATGYYTGTATCAGTGTC[2]	Oomycetes		Riit et al. (2016)
ITS5Oo[h]	18S (ITS): fwd: 1730	CTYRYCRTTTAGAGGAAGGTG	Stramenopila		This paper

(Continues)

TABLE 3 (Continued)

Primer[1]	Marker: direction: position	Sequence	Target	Comment	Reference
ITS2Oo[h]	5.8S (ITS1): rev: 40	GCAGCGKTCTTCATCGRTGT	Oomycetes	incl. Synurophyceae	This paper
Animals					
jgLCO1490[j]	COI: fwd: 1490	TITCIACIAAYCAYAARGAYATTGG	Metazoa	excl. several groups, incl. Amoebozoa	Geller et al. (2013)
jgHCO2198[j]	COI: rev: 2198	TAIACYTCIGGRTGICCRAARAAYCA	Metazoa	excl. several groups, incl. Amoebozoa	Geller et al. (2013)

Notes. The newly reported primers have been designed to cover >99% targeted taxa and tested in silico and complex soil samples. Full set of primers used for bacteria, fungi, oomycetes and eukaryotes in general can be found in Klindworth et al. (2013), Nilsson et al. (2018), Riit et al. (2016) and Adl, Habura, and Eglit (2014), respectively.

[1]Superscript letters indicate matching forward and reverse primer pairs. [2]Correct primer sequences compared to the trimmed ones in the original publication.

(i.e., optimized for bacterial 16S rRNA gene, mouse or human samples) or untransparent.

6.5 | Controls and technical replication

To quantify contamination and technical artefacts such as sequencing errors, chimera formation and tag-switching, it is recommended to run three types of control samples in parallel. Negative control samples should be incorporated during various steps of sample processing (DNA/RNA extraction, PCR; Lindahl et al., 2013). One or more positive controls—preferably organisms not expected in studied communities—might be included to quantify tag-switching (Schnell et al., 2015; Tedersoo et al., 2017). Mock community of known composition may provide information about the rates of chimera formation and efficiency of recovering ingredient taxa (Nguyen, Smith, Peay, & Kennedy, 2015). It is possible to use artificial DNA molecules for positive control and mock community samples, because their length, AT:CG ratio and homopolymer content can be controlled and their concentration can be precisely determined (Palmer, Jusino, Banik, & Lindner, 2017).

Technical replication is unnecessary in most cases, because these observations cannot be used as independent data points in the analysis. However, limited technical replication of a few samples may be feasible to estimate the performance and reproducibility of the method especially for newly developed protocols (Alberdi et al., 2018; Brooks et al., 2015).

6.6 | Quality filtering of HTS data

Analysis and quality filtering of HTS data are by far more challenging than viewing and editing Sanger sequencing reads because of large amounts of data and no clearly readable chromatograms. There is a myriad of available software for bioinformatics data analysis, most of which, such as mothur (www.mothur.org) and QIIME (www.qiime.org), are run 'on command line. Both of these popular bioinformatics platforms are optimized for alignment-based bacterial 16S rRNA gene analysis. QIITA (https://qiita.microbio.me) is a recently developed web-based analysis platform for bioinformatics and analysis workflow optimized for bacteria-targeted and metagenomic research. PipeCraft is a user-friendly software with a graphical interface, multiple options incorporated from other programs, capacity to analyse metabarcoding data from all sequencing platforms and compatibility with Linux, Mac and Windows, which all attract non-bioinformatician users (Anslan, Bahram, Hiiesalu, & Tedersoo, 2017). For analysis of non-alignable markers such as ITS, PipeCraft outperforms other bioinformat'ics pipelines in terms of input data formats, available analysis options and output quality (Anslan et al., 2018). Comprehensive overview about available analysis platforms for amplicon and metabarcoding data is given in Oulas et al. (2015). Bioinformatics analysis of fungal data is reviewed in Nilsson et al. (2018).

Although the output of HTS platforms is converted to the same format, these data differ in the distribution of errors and require different options for analysis (Anslan et al., 2017; Knief, 2014; Laehnemann, Borkhardt, & McHardy, 2015; Reuter et al., 2015). Quality-trimming is the first step of bioinformatics analysis and it is usually based on removing the 3′ end of sequences (or entire sequences) if it falls below specific quality threshold, the optima of which differ by sequencing platform. In a simultaneous sample demultiplexing process, sequences are re-assigned to biological samples based on the molecular identifiers. For demultiplexing sequence data with Golay barcodes, we recommend allowing 1–2 mismatches to tags and 1–2 mismatches to primers to account for random errors and natural primer-template mismatches. Demultiplexing from a single end typically enables to recover 40%–70% of all reads, but approximately a quarter of these are lost when accounting for the other tagged primer as well (Tedersoo et al., 2017). However, dual-tag demultiplexing enables to remove tag-switching artefacts and incomplete sequences (Kozich, Westcott, Baxter, Highlander, & Schloss, 2013).

To reduce computation time, quality-filtered sequences are usually pre-clustered using 100% or 99% identity and subjected to chimera checking. Chimera detection performs best when combining de novo and reference-based methods as implemented in UCHIME (Aas et al., 2017; Edgar, Haas, Clemente, Quince, & Knight, 2011) that can be run in all bioinformatics platforms.

Extraction of variable region(s) may precede or follow chimera check. Extraction of ITS and other variable regions in rRNA genes

enables simultaneous removal of non-target organisms and focus on a shorter but more variable barcode that has improved taxonomic resolution (Bengtsson-Palme et al., 2013; Hartmann, Howes, Abarenkov, Mohn, & Nilsson, 2010).

6.7 | Sequence clustering and Operational Taxonomic Units

Quality-filtered and trimmed sequences are subjected to clustering into OTUs, for which multiple algorithms exist (Kopylova et al., 2016). The best results are obtained when using open-source de novo clustering with single-linkage algorithms (Frøslev et al., 2017; Lindahl et al., 2013). Except for Illumina data, it is recommended to collapse homopolymers to trimers for clustering (Lindahl et al., 2013) or lowering the gap extension penalty, because other platforms are sensitive to indels in homopolymers. Although many protocols recommend removing sequences containing homopolymers of >8 or >10 bases, we do not encourage this practice for the non-coding regions, because many organisms do have naturally long homopolymers in these markers (Potter et al., 2017; Tedersoo et al., 2017).

In spite of different taxonomic resolution, the bacterial 16S and eukaryote 18S, 28S, ITS and COI sequences are typically clustered at 97% sequence identity, which is regarded as a compromise between natural intraspecific and interspecific sequence variation and random sequencing errors. The 97% sequence similarity threshold for all of these marker genes (except COI in some groups) is too conservative for species-level identification of most taxa. For example, some biological species of *Fusarium* display no variation at all in the relatively unconserved ITS region (Park et al., 2011). Therefore, HTS analysis of the ITS + 28S rRNA gene (Walder et al., 2017) and transcription elongation factor 1 subunit α (TEF; Karlsson et al., 2016) have been used to specifically distinguish *Fusarium* spp. With low-resolution markers, analysis of exact sequence variants can be performed using 100% similarity threshold or the DADA2 clustering program (Callahan et al., 2016).

All clustering methods generate more OTUs than expected at any barcoding threshold with increasing sequencing depth, indicating accumulation of PCR and sequencing errors into rare "satellite" taxa (Frøslev et al., 2017). This can be ameliorated by performing two or more consecutive clustering steps (Nguyen et al., 2015), post-clustering removal of taxa based on co-occurrence or phylogenetic algorithms (Frøslev et al., 2017; Potter et al., 2017) or focus on longer DNA fragments, where random errors are evened out (Tedersoo et al., 2017). It is further recommended to remove global singletons and perhaps OTUs with <5 or <10 sequences, depending on sequencing depth, as potentially artefactual (Frøslev et al., 2017; Lindahl et al., 2013; Nguyen et al., 2015; Tedersoo et al., 2010).

6.8 | Sequence-based taxonomic identification and taxon communication

Selection of one or more reference databases is essential for sequence-based identification (reviewed in Kashyap, Rai, et al.,

2017b). Since up to 20% of the material in INSDc is of poor quality or misidentified, initiatives such as UNITE (https://unite.ut.ee/), SILVA (www.arb-silva.de) and UniEuk (https://unieuk.org/) have generated databases and reference data sets populated with filtered and third-party annotated sequences. SILVA is focused on nuclear SSU and LSU sequences of prokaryotes and eukaryotes, but both oomycetes and fungi are poorly represented and have problems with taxonomic assignment (Tedersoo Tooming-Klunderud et al., 2017; Tedersoo, Anslan, et al., 2015; Yarza, Yilmaz, Panzer, Glöckner, & Reich, 2017). The UniEuk initiative is focused both on taxonomy and on curation of high-quality 18S rDNA sequences of eukaryotes (Berney et al., 2017). The current version of UNITE includes SSU, ITS and LSU sequence data for all eukaryotes, although only fungal and oomycete ITS sequences have been intensively annotated for taxonomy, sequence quality and ecological metadata. Some pathogenic fungal groups, in particular, have been thoroughly checked, annotated and assigned for type status in UNITE (Nilsson et al., 2014). In the BOLD database (Ratnasingham & Hebert, 2007), curated COI sequence data for animals, Oomycota and other specific groups of protists are maintained. Thus, these databases provide best-suited species-level reference data for general molecular identification of pathogenic organisms. However, researchers focused on more narrow groups such as *Fusarium* or *Phytophthora* could use Fusarium-ID (Park et al., 2011) and the Phytophthora Database (www.phytophthoradb.org) in addition. Animal and human pathogens have annotated sequence data in the ISHAM-ITS database (Irinyi et al., 2015). Metagenomic and metatranscriptomic analyses require inclusion of functional gene and genomics databases for combined taxonomic and functional analysis (Huson, Mitra, Ruscheweyh, Weber, & Schuster, 2011). Detection of viruses amongst genomic, metagenomic and metatranscriptomic reads requires some specific data mining effort. Pipelines such as VirusFinder (Wang, Jia, & Zhao, 2013) and particularly VirFind (Ho & Tzanetakis, 2014) enable to search for virus motifs from custom sequence data sets and EST databases. As a reference for identification, virologists use mostly the RefSeq database of INSDc (O'Leary et al., 2016) and Comprehensive Phytopathogen Genomics Resource (CPGR) database (Hamilton et al., 2011).

For taxonomic assignments, it is most common to use BLAST-based similarity search methods for representative sequences of each OTU (Nilsson et al., 2014; Tedersoo & Nilsson, 2016). The Naive Bayesian Classifier (Porras-Alfaro, Liu, Kuske, & Xie, 2014; Wang, Garrity, Tiedje, & Cole, 2007) is widely used for conservative identification in prokaryotes, but this method has gained little popularity among mycologists due to a low proportion of taxa identified to species or genus level. This has been improved in ProTax-Fungi, which provides statistical assessment of assignment precision to different taxa from species to phylum ranks (Abarenkov et al., 2018). In fungal and oomycete ITS sequences, species, genus, family and order levels can be approximately approximated at >97%–99%, >90%, >85% and >80% ITS sequence similarity, respectively, to the closest identified sequence (Tedersoo et al., 2014; Tedersoo et al., 2017; data in Riit et al., 2016). In bacterial 16S (full-length), these figures are >98%–99%, >94.5%, >86.5% and >82%, respectively (Yarza et al., 2014). Due to different

rates of rRNA gene evolution, there are multiple exceptions, with sordariomycete (Ascomycota) and oomycete species tending to exhibit greater similarity and early diverging fungal groups lower similarity. These differences are even greater among animal and protist groups (Anslan & Tedersoo, 2015; Nassonova, Smirnov, Fahrni, & Pawlowski, 2010; Yang et al., 2011), but remain still poorly understood for most taxa (Pawlowski et al., 2012).

An optional step is to assign functional traits such as pathogenicity information to OTUs, for which database-related tools exist. For Bacteria, an automated pipeline SINAPS (Edgar, 2017) enables to search and predict custom traits using the ProTraits reference database (Brbic et al., 2016). The basic fungal traits can be assigned to taxonomic profiles using a tool in FunGuild database (Nguyen et al., 2016). Its main limitation is genus-level operation, although it alerts that many genera contain both pathogens and saprotrophs or endophytes. As discussed above, the detected "pathogenic" OTUs may be non-pathogenic on non-hosts, rendering the assignments strongly context dependent. Therefore, more accurate metadata with host- and tissue-related traits assigned to species, species hypotheses (see next paragraph) or isolates/sequences are urgently needed.

HTS studies enable to recover tens of thousands of OTUs, most of which cannot be usually assigned to described species, which renders these difficult to communicate across studies. The UNITE and BOLD databases use taxon codes (species hypotheses and BINs, respectively) linked to Digital Object Identifiers (DOIs). These machine-readable DOIs enable communication of both named and unnamed taxa across studies and time (Kõljalg et al., 2013; Kõljalg, Tedersoo, Nilsson, & Abarenkov, 2016; Ratnasingham & Hebert, 2013).

6.9 | Post-bioinformatics data quality control

For soil and raw tissue samples with non-optimal storage conditions, it may be important to estimate sample quality due to potential overgrowth by moulds (Lindahl, Boer, & Finlay, 2010). This can be performed by measuring the average size of extracted DNA/RNA molecules on the gel or calculation of the relative abundance of moulds (fungal orders Hypocreales, Mucorales, Umbelopsidales and Mortierellales). Dominance of a single mould OTU, which is usually associated with reduced taxonomic richness, can be considered indicative of sample spoilage (Tedersoo et al., 2014).

Similarly, it may be feasible to exclude samples with <5- to 10-fold less sequences compared with the median. Such poor recovery may be ascribed to the failure to normalize a sample, poor performance of particular identifier tags and/or dominance of particular organisms in a sample, which are disfavoured in the library preparation, sequencing or quality-filtering steps. In spite of attempts to normalize quantity of amplicons, the number of retrieved sequences typically vary >3-fold. It is common to rarefy all samples to the same minimum sequencing depth, but this loses vast majority of taxonomic information. Therefore, it is recommended to calculate residuals of richness relative to square-root or logarithmic function of sequencing depth (whichever fits better), or use these functions as covariates in uni- and multivariate statistics (Balint et al., 2016).

Due to high sensitivity, HTS data commonly suffer from traces of environmental contamination or tag-switching (see above). Information about the OTUs in control and experimental samples enables evaluation of these technical biases and need for extra quality filtering (Nguyen et al., 2015; Palmer et al., 2017). In case of extensive tag-switching, sequences can be removed according to statistical formulae (Larsson, Stanley, Sinha, Weissman, & Sandberg, 2018). Although the tag-switching artefacts usually account for 0.1%–3% of all sequences (Palmer et al., 2017; Schnell et al., 2015; Tedersoo et al., 2017), these may blur qualitative diversity analyses and particularly network analyses that are sensitive to adding low-abundance OTUs. More importantly, tag-switching may generate false-positive implications of low-level presence of a pathogen or biocontrol organism, especially when these dominate some samples in the library.

6.10 | HTS data analysis

HTS platforms generate enormous OTU-by-sample data matrices, which cannot be sometimes fully loaded into common spreadsheet programs. Therefore, experts use python or perl scripts to navigate and transform the data in text format. These large community matrices also test the limits of statistical software and processors. Many commonly used methods for community phylogenetics, bootstrap resampling and network analysis become computationally prohibitive. Thus, use of computation-efficient algorithms is warranted. To reduce the computation requirements, the data can be compressed by removal of rare species, which typically reduces unexplained variance and promotes statistical power (Põlme et al., 2018), but its effect on potential type I and type II error is not known in multivariate or network analyses.

For multivariate analyses, we recommend downweighing abundant OTUs by Hellinger (square-root) transformation to account for the semiquantitative nature of HTS. Use of qualitative binary data (presence/absence) is not recommended, because of lower fit due to loss of the (semi)quantitative information and artificial equalization of potentially artefactual low-abundance (including tag-switch artefacts) and real high-abundance OTUs (Balint et al., 2016). We recommend use of PERMANOVA for explicit statistical testing of shifts in community composition, because it allows including interactions, random factors and nested designs. ANCOM and Random Forest machine-learning algorithm provide statistical information about the performance of each OTU in the community matrix. General information about multivariate analysis methods suitable for HTS data is given in Buttigieg and Ramette (2014). Notably, the same multivariate techniques are commonly used to analyse standardized microarray and metagenomic and metatranscriptomic data (Thomas, Gilbert, & Meyer, 2012).

In univariate analyses, OTU richness, diversity, colonization, damage and relative abundance of certain taxonomic or functional groups are used as dependent variables. Apart from considering sequencing depth and treatment of rare OTUs, the analyses should follow best statistical practices including appropriate transformations, testing assumptions, etc. Balint et al. (2016) provide an overview about general recommendations to statistical analysis of HTS data, computation-efficient programs and potential pitfalls.

6.11 | HTS data storage and reporting

HTS data sets are stored both as raw data files and elaborated data sets. The raw.fastq files, metadata files and files with identifier tag and primer information are kept in the Short Read Archive (SRA). These files enable users to perform all steps of bioinformatics analyses including generation of OTU table and identification. This is important from several aspects such as confirming earlier findings with updated filtering procedures, addressing additional questions and performing metastudies using standardized filtering procedures. It is, however, discouraged to submit representative sequences of HTS-derived OTUs to public databases because of their short length, potentially artefactual nature and unreliable taxonomic annotation. These environmental sequences would increase the proportion of poorly annotated and erroneous data and complicate identification in subsequent studies.

Curated OTU-by-sample matrices including technical and environmental metadata, representative sequences as well as taxonomic and functional annotations should be deposited in machine-readable FAIR data format in specific data repositories such as Dryad Digital Repository (www.datadryad.org) and DataOne (www.dataone.org). Darwin Core (https://rs.tdwg.org/dwc/) is the main standard for biodiversity data, which is linked to MIxS (https://gensc.org/mixs/) and MIMARKS (https://wiki.gensc.org/index.php?title=MIMARKS) standards by the Genomics Standards Consortium that are relevant to all gene sequence data and marker gene sequence data, respectively. Metagenomic data should follow MIxS and MIMS (https://wiki.gensc.org/index.php?title=MIGS/MIMS) standards (ten Hoopen et al., 2017). The machine-readable FAIR data format allows researchers to understand and rapidly incorporate the data into meta-analyses. Such standardized data sets in digital repositories enable separate DOI-based citations.

In publications, it is important to refer to any additional data in the supplement or data repositories. It is also important to record and describe precisely all analytical steps including specific options in data filtering, because this information provides important details about the data quality and stringency of filtering to the readers. Nilsson et al. (2011) provide thorough recommendations about the required details for molecular and bioinformatics analyses.

7 | PERSPECTIVES

Only a fraction of available high-throughput identification potential has been currently used in plant pathology. This is related to the practical surveillance-oriented work of plant pathologists and entomologists but focus on human and animal subjects by molecular pathologists. Governmental plant health surveillance organizations need to follow certified protocols for diagnosis, which develop slowly due to time-consuming tests. Limited budgets also hinder the possibility of purchasing high-throughput analysis equipment by governmental institutions. Considering analysis costs and time, practicing pathologists would certainly take advantage of qPCR/ddPCR for real-time quantification of specific pathogens and custom microarrays for simultaneous detection and quantification of multiple selected pathogens. In the nearest future, it may be possible to detect multiple organisms including pathogens using high-throughput sequencing on portable pocket-size sequencers as demonstrated for viruses using the Oxford Nanopore MinION platform (Loman et al., 2015). For a simplified procedure, a single working day is essentially required for sample collection, analysis and interpretation of results. Detection of organisms may occur much faster using nanotechnological biosensors that recognize multiple specific volatile molecules in parallel using antibody receptor-based optical or electrochemical detection (Kashyap, Kumar, & Srivastava, 2017a; Khater, Escosura-Muñiz, & Merkoçi, 2017; Sutarlie, Ow, & Su, 2017). Throughput of these biosensors can be greatly increased by using a microchip format for signal detection (Wang, Long, Liu, Wu, & Hu, 2017).

Other high-throughput identification methods are more time-consuming but also more sensitive and thus better suited for research purposes. Metagenomic and metatranscriptomic methods offer great potential when targeting viruses (Zhang et al., 2005) or these together with prokaryote and eukaryote pathogens and pests simultaneously (Chandler, Liu, & Bennett, 2015). Alternatively, nematodes, insect pests, oomycetes and fungi can all be assessed by using a mixture of degenerate primers targeting the same marker or multiplex primers targeting different markers via metabarcoding (de Barba et al., 2014; Tedersoo, Bahram, et al., 2016; Tedersoo, Liiv, et al., 2016). Targeted template capture by use of specific hybridization probes and immunochemical methods allows concentrate marker genes and pathogenesis-related genes of antagonists (Dowle, Pochon, Banks, Shearer, & Wood, 2016) that can be further identified using PCR-free methods.

Because of great intraspecific resolution, high-throughput fingerprinting and population genomics approaches offer enormous potential for diagnosis of aggressive strains or pathotypes and uncover their patterns of dispersal (O'Hanlon et al., 2018) and potential hybridization (Qiu, Cai, Luo, Bhattacharya, & Zhang, 2016). Given appropriate quality filtering, these methods are sensitive enough to distinguish rare alleles and SNPs from noise (Isola et al., 2005) in hundreds of samples in parallel. Whole-genome sequencing and transcriptome analyses complement HTS-based identification methods by shedding light into pathogenesis mechanisms and facilitating generation of vaccines and biocides and selection of biocontrol agents (Grünwald et al., 2016).

For correct identification, community-curated and taxonomically annotated reference databases are urgently needed. Such databases are maintained only for a few most important pathogen groups and cover the main barcoding marker genes (Park et al., 2008, 2011). Sequence databases should share third-party metadata and taxonomic annotations, whenever these are updated in one of these (Nilsson et al., 2014). In spite of a large proportion of erroneous data, INSDc will certainly continue to play a central role in bridging more specific databases encompassing genes from all domains of life. So, let's contribute well-annotated and high-quality sequence data to INSDc to benefit the pathologists research community! This also applies to HTS data sets and data matrices, the great practical and scientific value of

which can be recognized perhaps after several decades. Alongside storing sequence data, it is important to maintain tissue and soil samples that can be resource efficiently kept dried at room temperature. In spite of small size, accumulating DNA samples tend to rapidly fill refrigerators in entire rooms and are vulnerable to technical failure and power cuts. Besides the possibility of morphology-based re-identification and description of new species, both botanical and pathological herbaria provide excellent sources to trace back the evolution and dispersal of pathogens and pests (Drenkhan et al., 2017; Drenkhan, Riit, Adamson, & Hanso, 2016; Yoshida et al., 2014).

Taken together, high-throughput identification techniques offer great promise for detection and rapid identification of new pathogens and diseases in humans as well as tree and crop plantations and early warning systems such as the sentinel nurseries and botanical gardens. HTS has already demonstrated its usefulness in studies of soil- and plant-associated microbial communities for detection of new potential pathogens and potentially invasive species before their introduction to the new environment and contact with new hosts. We predict that rapid monitoring methods such as nanopore sequencing, microarrays and nanotechnological biosensors will become particularly useful for early disease diagnostics and smart application of countermeasures such as biocides and biocontrol agents.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

M.C., L.T. and R.D. developed the concept; S.A. constructed the figure and contributed specific details together with C.M.-R.; and L.T. wrote the article with input from co-authors.

DATA ACCESSIBILITY

This review does not contain original data; all compiled published data are available in the main text body and tables.

ORCID

Leho Tedersoo http://orcid.org/0000-0002-1635-1249

REFERENCES

Aas, A. B., Davey, M. L., & Kauserud, H. (2017). ITS all right mama: Investigating the formation of chimeric sequences in the ITS2 region by DNA metabarcoding analyses of fungal mock communities of different complexities. *Molecular Ecology Resources*, *17*, 730–741. https://doi.org/10.1111/1755-0998.12622

Abarenkov, K., Nilsson, R. H., Larsson, K.-H., Alexander, I. J., Eberhardt, U., Erland, S., … Kõljalg, U. (2010). The UNITE database for molecular identification of fungi – recent updates and future perspectives. *New Phytologist*, *186*, 281–285. https://doi.org/10.1111/j.1469-8137.2009.03160.x

Abarenkov, K., Somervuo, P., Nilsson, R. H., Kirk, P. M., Huotari, T., Abrego, N., & Ovaskainen, O. (2018). Protax-fungi: A web-based tool for probabilistic taxonomic placement of fungal internal transcribed spacer sequences. *New Phytologist*, *220*(2), 517–525.

Abdelfattah, A., Malacrinò, A., Wisniewski, M., Cacciola, S. O., & Schena, L. (2018). Metabarcoding: A powerful tool to investigate microbial communities and shape future plant protection strategies. *Biological Control*, *120*, 1–10. https://doi.org/10.1016/j.biocontrol.2017.07.009

Abdelfattah, A., Nicosia, M. G. L. D., Cacciola, S. O., Droby, S., & Schena, L. (2015). Metabarcoding analysis of fungal diversity in the phyllosphere and carposphere of olive (*Olea europaea*). *PLoS ONE*, *10*, e0131069.

Abdelfattah, A., Wisniewski, M., Nicosia, L. D., & M.G., Cacciola, S.O., & Schena, L., (2016). Metagenomic analysis of fungal diversity on strawberry plants and the effect of management practices on the fungal community structure of aerial organs. *PLoS ONE*, *11*, e0160470.

Adams, I., & Fox, A. (2016). Diagnosis of plant viruses using next-generation sequencing and metagenomic analysis. In A. Wang, & X. Zhou (Eds.), *Current research topics in plant virology* (pp. 323–335). Cham, Switzerland: Springer.

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., … Boonham, N. (2009). Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. *Molecular Plant Pathology*, *10*, 537–545. https://doi.org/10.1111/j.1364-3703.2009.00545.x

Adl, S. M., Habura, A., & Eglit, Y. (2014). Amplification primers of SSU rDNA for soil protists. *Soil Biology and Biochemistry*, *69*, 328–342. https://doi.org/10.1016/j.soilbio.2013.10.024

Agler, M. T., Ruhe, J., Kroll, S., Morhenn, C., & Kim, S.-T., … Weigel, D., (2016). Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biology*, *14*, e1002352.

Ahrendt, S. R., Quandt, C. A., Ciobanu, D., Clum, A., Salamov, A., Andreopoulos, B., … Grigoriev, I. V. (2018). Leveraging single-cell genomics to expand the fungal tree of life. *Nature Microbiology*, https://doi.org/10.1038/s41564-018-0261-0

Aizpurua, O., Budinski, I., Georgiakakis, P., Gopalakrishnan, S., Ibañez, C., Mata, V., … Zrncic, V. (2018). Agriculture shapes the trophic niche of a bat preying on multiple pest arthropods across Europe: Evidence from DNA metabarcoding. *Molecular Ecology*, *27*, 815–825. https://doi.org/10.1111/mec.14474

Al Rwahnih, M., Daubert, S., Golino, D., & Rowhani, A. (2009). Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology*, *387*, 395–401. https://doi.org/10.1016/j.virol.2009.02.028

Alberdi, A., Aizpurua, O., Gilbert, M. T., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*, 134–147. https://doi.org/10.1111/2041-210X.12849

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., … Bunce, M. (2012). The half-life of DNA in bone: Measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B*, *279*, 4724–4733. https://doi.org/10.1098/rspb.2012.1745

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., & Huse, S. M. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE*, *4*, e6372.

Amend, A. S., Seifert, K. A., Samson, R., & Bruns, T. D. (2010). Indoor fungal composition is geographically patterned and more diverse in

temperate zones than in the tropics. *Proceedings of the National Academy of Sciences USA*, 107, 13748–13753. https://doi.org/10.1073/pnas.1000454107

Andujar, C., Arribas, P., Ruzicka, F., Crampton-Platt, A., Timmermans, A. J. T. N., & Vogler, A. P. (2015). Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology*, 24, 3603–3617. https://doi.org/10.1111/mec.13195

Anslan, S., Bahram, M., Hiiesalu, I., & Tedersoo, L. (2017). PipeCraft: Flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. *Molecular Ecology Resources*, 17, e234–e240. https://doi.org/10.1111/1755-0998.12692

Anslan, S., Nilsson, H., Wurzbacher, C., Baldrian, P., Tedersoo, L., & Bahram, M. (2018). Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. *MycoKeys*, 39, 29–40.

Anslan, S., & Tedersoo, L. (2015). Performance of cytochrome c oxidase subunit I (COI), ribosomal DNA large subunit (LSU) and internal transcribed spacer 2 (ITS2) in DNA barcoding of Collembola. *European Journal of Soil Biology*, 69, 1–7. https://doi.org/10.1016/j.ejsobi.2015.04.001

Anthony, M. A., Frey, S. D., & Stinson, K. A. (2017). Fungal community homogenization, shift in dominant trophic guild, and appearance of novel taxa with biotic invasion. *Ecosphere*, 8, e01951. https://doi.org/10.1002/ecs2.1951

Apprill, A., McNally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, 75, 129–137. https://doi.org/10.3354/ame01753

Ashfaq, M., & Hebert, P. D. (2016). DNA barcodes for bio-surveillance: Regulated and economically important arthropod plant pests. *Genome*, 59, 933–945. https://doi.org/10.1139/gen-2016-0024

Bagchi, R., Gallery, R. E., Gripenberg, S., Gurr, S. J., Narayan, L., Addis, C. E., … Lewis, O. T. (2014). Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature*, 506, 85–88. https://doi.org/10.1038/nature12911

Bahram, M., Kohout, P., Anslan, S., Harend, H., Abarenkov, K., & Tedersoo, L. (2016). Stochastic distribution of small soil eukaryotes resulting from high dispersal and drift in a local environment. *ISME Journal*, 10, 885–896. https://doi.org/10.1038/ismej.2015.164

Bahram, M., Hildebrand, F., Forsland, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., … Bork, P. (2018). Towards the structure and function of the global topsoil microbiome. *Nature*, 560, 233–237.

Bainard, L. D., Navarro-Borrell, A., Hamel, C., Braun, K., Hanson, K., & Gan, Y. (2017). Increasing the frequency of pulses in crop rotations reduces soil fungal diversity and increases the proportion of fungal pathotrophs in a semiarid agroecosystem. *Agriculture, Ecosystems and Environment*, 240, 206–214. https://doi.org/10.1016/j.agee.2017.02.020

Balint, M., Bahram, M., Eren, A. M., Faust, K., Fuhrman, J. A., Lindahl, B., … Tedersoo, L. (2016). Millions of reads, thousands of taxa: Microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews*, 40, 686–700. https://doi.org/10.1093/femsre/fuw017

Balint, M., Bartha, L., O'Hara, R. B., Olson, M. S., Otte, J., Pfenninger, K., & S., Schmitt, I., (2015). Relocation, high-latitude warming and host genetic identity shape the foliar fungal microbiome of poplars. *Molecular Ecology*, 24, 235–248. https://doi.org/10.1111/mec.13018

Barham, E. (2016). The unique role of sentinel trees, botanic gardens and arboreta in safeguarding global plant health. *Plant Biosystems*, 150, 377–380. https://doi.org/10.1080/11263504.2016.1179231

Barnes, I., Van Der Nest, A., Mullett, M. S., Crous, P. W., Drenkhan, R., Musolin, D. L., & Wingfield, M. J. (2016). Neotypification of *Dothistroma septosporum* and epitypification of *D. pini*, causal agents of *Dothistroma* needle blight of pine. *Forest Pathology*, 46, 388–407.

Bartoli, C., Frachon, L., Barret, M., Rigal, M., Huard-Chauveau, C., Mayjonade, B., … Roux, F. (2018). In situ relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*. *ISME Journal*, 12, 2024–2038. https://doi.org/10.1038/s41396-018-0152-7

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., … Nilsson, R. H. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution*, 4, 914–919. https://doi.org/10.1111/2041-210X.12073

Benitez-Paez, A., & Sanz, Y. (2017). Multi-locus and long amplicon sequencing approach to study microbial diversity at species level using the MinION portable nanopore sequencer. *GigaScience*, 6, 1–12. https://doi.org/10.1093/gigascience/gix043

Berendsen, R. L., Vismans, G., Yu, K., Song, Y., Jonge, R., Burgman, W. P., … Pieterse, C. M. (2018). Disease-induced assemblage of a plant-beneficial bacterial consortium. *ISME Journal*, 12, 1496–1505. https://doi.org/10.1038/s41396-018-0093-1

Berney, C., Ciuprina, A., Bender, S., Brodie, J., Edgcomb, V., Kim, E., … Bass D., (2017). UniEuk: Time to speak a common language in Protistology!. *Journal of Eukaryotic Microbiology*, 64, 407–411.

Bik, H. M., Porazinska, D., Creer, S., Caporaso, J. G., Knight, R., & Thomas, W. K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology and Evolution*, 27, 233–243. https://doi.org/10.1016/j.tree.2011.11.010

Bonito, G., Reynolds, H., Robeson, M. S., Nelson, J., Hodkinson, B. P., Tuskan, G., … Vilgalys, R. (2014). Plant host and soil origin influence fungal and bacterial assemblages in the roots of woody plants. *Molecular Ecology*, 23, 3356–3370. https://doi.org/10.1111/mec.12821

Bork, P., Bowler, C., De Vargas, C., Gorsky, G., Karsenti, E., & Wincker, P. (2015). Tara Oceans studies plankton at planetary scale. *Science*, 348, 873. https://doi.org/10.1126/science.aac5605

Brasier, C. M., & Kirk, S. A. (2010). Rapid emergence of hybrids between the two subspecies of *Ophiostoma novo-ulmi* with a high level of pathogenic fitness. *Plant Pathology*, 59, 186–199.

Brbic, M., Piškorec, M., Vidulin, V., Kriško, A., Šmuc, T., & Supek, F. (2016). The landscape of microbial phenotypic traits and associated genes. *Nucleic Acids Research*, 44, 10074–10090. https://doi.org/10.1093/nar/gkw964

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., … Strauss, J. F. (2015). The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology*, 15, 66.

Buee, M., Reich, M., Murat, C., Morin, E., Nilsson, R. H., Uroz, S., & Martin, F. (2009). 454 pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist*, 184, 449–456. https://doi.org/10.1111/j.1469-8137.2009.03003.x

Busby, P. E., Peay, K. G., & Newcombe, G. (2016). Common foliar fungi of *Populus trichocarpa* modify *Melampsora* rust disease severity. *New Phytologist*, 209, 1681–1692.

Buttigieg, P. L., & Ramette, A. (2014). A guide to statistical analysis in microbial ecology: A community-focused living review of multivariate data analyses. *FEMS Microbiology Ecology*, 90, 543–550. https://doi.org/10.1111/1574-6941.12437

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581–584. https://doi.org/10.1038/nmeth.3869

Carneiro, R. M., de Oliveira Lima, F. S., & Correia, V. R. (2017). Methods and tools currently used for the identification of plant parasitic nematodes. In M. M. Shah, & M. Mahamood (Eds.), *Nematology-concepts, diagnosis and control* (pp. 19–35). InTech.

Catala, S., Berbegal, M., Pérez-Sierra, A., & Abad-Camposa, P. (2017). Metabarcoding and development of new real-time specific assays

reveal *Phytophthora* species diversity in holm oak forests in eastern Spain. *Plant Pathology*, 66, 115–123.

Catala, S., Pérez-Sierra, A., & Abad-Campos, P. (2015). The use of genus-specific Amplicon pyrosequencing to assess *Phytophthora* species diversity using eDNA from soil and water in Northern Spain. *PLoS ONE*, 10, e0119311. https://doi.org/10.1371/journal.pone.0119311

Chandler, J. A., Liu, R. M., & Bennett, S. N. (2015). RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. *Frontiers in Microbiology*, 2015, 6.

Chapelle, E. C., Mendes, R., Bakker, P. A., & Raaijmakers, J. M. (2016). Fungal invasion of the rhizosphere microbiome. *ISME Journal*, 10, 265–268. https://doi.org/10.1038/ismej.2015.82

Cheng, J., Hu, H., Kang, Y., Chen, W., Fang, W., Wang, K., … Cao, Q. (2018). Identification of pathogens in culture-negative infective endocarditis with metagenomic analysis. *bioRxiv*, 2018, 336388.

Cleary, M. R., Nguyen, D., Marciulyniene, D., Berlin, A., Vasaitis, R., & Stenlid, J. (2016). Friend or foe? Biological and ecological traits of the European ash dieback pathogen *Hymenoscyphus fraxineus* in its native environment. *Scientific Reports*, 6, 21895.

Cline, L. C., Hobbie, S. E., Madritch, M., Buyarski, C. R., Tilman, D., & Cavender-Bares, J. M. (2017). Resource availability underlies the plant-fungal diversity relationship in a grassland ecosystem. *Ecology*, 99, 204–215. https://doi.org/10.1002/ecy.2075

Comtet, T., Sandionigi, A., Viard, F., & Casiraghi, M. (2015). DNA (meta) barcoding of biological invasions: A powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions*, 17, 905–922. https://doi.org/10.1007/s10530-015-0854-y

Cooke, D. E., Cano, L. M., Raffaele, S., Bain, R. A., Cooke, L. R., Etherington, G. J., … Grünwald, N. J. (2012). Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen. *PLoS Path*, 8, e1002940.

Creer, S., Fonseca, V. G., Porazinska, D. L., Giblin-Davis, R. M., Sung, W., Power, D. M., … Thomas, W. K. (2010). Ultrasequencing of the meiofaunal biosphere: Practice, pitfalls and promises. *Molecular Ecology*, 19, S4–S20. https://doi.org/10.1111/j.1365-294X.2009.04473.x

Critescu, M. E. (2014). From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. *Trends in Ecology and Evolution*, 29, 566–571. https://doi.org/10.1016/j.tree.2014.08.001

Cross, H., Sønstebø, J. H., Nagy, N. E., Timmermann, V., Solheim, H., Børja, I., … Vivian-Smith, A. (2017). Fungal diversity and seasonal succession in ash leaves infected by the invasive ascomycete *Hymenoscyphus fraxineus*. *New Phytologist*, 213, 1405–1417.

D'Amore, R., Ijaz, U. Z., Schirmer, M., Kenny, J. G., Gregory, R., Darby, A. C., … Hall, N. (2016). A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics*, 17, 55.

Darby, B. J., Todd, T. C., & Herman, M. A. (2013). High-throughput amplicon sequencing of rRNA genes requires copy number correction to accurately reflect the effects of management practices on soil nematode community structure. *Molecular Ecology*, 22, 5456–5471.

de Barba, M., Boyer, M. F., Rioux, M. D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources*, 14, 306–323. https://doi.org/10.1111/1755-0998.12188

DeAngelis, K. M., Wu, C. H., Beller, H. R., Brodie, E. L., Chakraborty, R., DeSantis, T. Z., & Andersen, G. L. (2011). PCR amplification-independent methods for detection of microbial communities by the high-density microarray phylochip. *Applied and Environmental Microbiology*, 77, 6313–6322. https://doi.org/10.1128/AEM.05262-11

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., … Pfrender, M. E. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26, 5872–5895. https://doi.org/10.1111/mec.14350

Dickie, I. A., & St. John, M. G. (2016). Second-generation molecular understanding of mycorrhizas in soil ecosystems. In F. Martin (Ed.), *Molecular mycorrhizal symbiosis* (pp. 473–491). Hoboken, NJ: Wiley-Blackwell.

Doonan, J., Denman, S., McDonald, J. E., & Golyshin, P. N. (2017). Shotgun metagenomic sequencing analysis of soft-rot Enterobacteriaceae in polymicrobial communities. In W. Streit, & R. Daniel (Eds.), *Metagenomics: methods and protocols* (p. pp. p. 85–97.). New York, NY: Humana Press.

Dowle, E. J., Pochon, X., Banks, J., Shearer, K., & Wood, S. A. (2016). Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: A case study using freshwater macroinvertebrates. *Molecular Ecology Resources*, 16, 1240–1254. https://doi.org/10.1111/1755-0998.12488

Drenkhan, R., Riit, T., Adamson, K., & Hanso, M. (2016). The earliest samples of *Hymenoscyphus albidus* vs. *H. fraxineus* in Estonian herbaria and the development of new species-specific PCR primers. *Mycological Progress*, 15, 835–844.

Drenkhan, R., Solheim, H., Bogacheva, A., Riit, T., Adamson, K., Drenkhan, T., … Hietala, A. M. (2017). *Hymenoscyphus fraxineus* is a leaf pathogen of local *Fraxinus* species in the Russian Far East. *Plant Pathology*, 66, 490–500.

Dreo, T., Pirc, M., Ramšak, Ž., Pavšič, J., Milavec, M., Žel, J., & Gruden, K. (2014). Optimising droplet digital PCR analysis approaches for detection and quantification of bacteria: A case study of fire blight and potato brown rot. *Analytical and Bioanalytical Chemistry*, 406, 6513–6528. https://doi.org/10.1007/s00216-014-8084-1

Duan, Y., Zhou, L., Hall, D. G., Li, W., Doddapaneni, H., Lin, H., … Dickerman, A. (2009). Complete genome sequence of citrus huanglongbing bacterium, '*Candidatus Liberibacter asiaticus*' obtained through metagenomics. *Molecular Plant-Microbe Interactions*, 22, 1011–1120.

Edgar, R. C. (2017). SINAPS: Prediction of microbial traits from marker gene sequences. *bioRxiv*, 2017, 124156.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27, 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U., Bates, S. T., Lauber, C. L., … Caporaso, J. G. (2013). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences USA*, 109, 21390–21395.

Frey, K. G., Herrera-Galeano, J. E., Redden, C. L., Luu, T. V., Servetas, S. L., Mateczun, A. J., … Bishop-Lilly, K. A. (2014). Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics*, 15, 96.

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, 8, 1188.

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations, and practical recommendations. *Molecular Ecology*, 26, 5369–5406. https://doi.org/10.1111/mec.14264

García-Martínez, J., Acinas, S. G., Antón, A. I., & Rodríguez-Valera, F. (1999). Use of the 16S–23S ribosomal genes spacer region in studies of prokaryotic diversity. *Journal of Microbiological Methods*, 36, 55–64. https://doi.org/10.1016/S0167-7012(99)00011-1

Gawad, C., Koh, W., & Quake, S. R. (2014). Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proceedings of the National Academy of Sciences USA*, 111, 17947–17952. https://doi.org/10.1073/pnas.1420822111

Geisen, S. (2016). Thorough high-throughput sequencing analyses unravels huge diversities of soil parasitic protists. *Environmental Microbiology*, 18, 1669–1672. https://doi.org/10.1111/1462-2920.13309

Geisen, S., Snoek, L. B., ten Hooven, F. C., Duyts, H., Kostenko, O., Bloem, J., … van der Putten, W. H. (2018). Integrating quantitative

morphological and qualitative molecular methods to analyse soil nematode community responses to plant range expansion. *Methods in Ecology and Evolution*, *9*, 1366–1378. https://doi.org/10.1111/2041-210X.12999

Geisen, S., Tveit, A. T., Clark, I. M., Richter, A., Svenning, M. M., Bonkowski, M., & Urich, T. (2015). Metatranscriptomic census of active protists in soils. *ISME Journal*, *9*, 2178–2190. https://doi.org/10.1038/ismej.2015.30

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, *13*, 851–861.

Gherbawy, Y., & Voigt, K. (2010). *Molecular identification of fungi*. Berlin, Germany: Springer.

Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., … Knights, D. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nature Biotechnology*, *34*, 942–949. https://doi.org/10.1038/nbt.3601

Gomez-Rodríguez, C., Timmermans, M. J., Crampton-Platt, A., & Vogler, A. P. (2017). Intraspecific genetic variation in complex assemblages from mitochondrial metagenomics: Comparison with DNA barcodes. *Methods in Ecology and Evolution*, *8*, 248–256. https://doi.org/10.1111/2041-210X.12667

Grantham, N. S., Reich, B. J., Pacifici, K., Laber, E. B., Menninger, H. L., Henley, J. B., … Dunn, R. R. (2015). Fungi identify the geographic origin of dust samples. *PLoS ONE*, *10*, e0122605.

Grau, O., Geml, J., Perez-Haase, A., Ninot, J. M., Semenova-Nelsen, T. A., & Penuelas, J. (2017). Abrupt changes in the composition and function of fungal communities along an environmental gradient in the high Arctic. *Molecular Ecology*, *26*, 4798–4810. https://doi.org/10.1111/mec.14227

Green Tringe, S., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., … Rubin, E. M. (2005). Comparative metagenomics of microbial communities. *Science*, *308*, 554–557. https://doi.org/10.1126/science.1107851

Grünwald, N. J., McDonald, B. A., & Milgroom, M. G. (2016). Population genomics of fungal and oomycete pathogens. *Annual Reviews in Phytopathology*, *54*, 323–346. https://doi.org/10.1146/annurev-phyto-080614-115913

Gutierrez-Aguirre, I., Rački, N., Dreo, T., & Ravnikar, M. (2015). Droplet digital PCR for absolute quantification of pathogens. In C. Lacomme (Ed.). *Plant pathology: Techniques and protocols* (pp. 331–347). New York, NY: Humana Press.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., … Birren, B. W. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, *21*, 494–504. https://doi.org/10.1101/gr.112730.110

Hamilton, J. P., Neeno-Eckwall, E. C., Adhikari, B. N., Perna, N. T., Tisserat, N., Leach, J. E., … Buell, C. R. (2011). The comprehensive phytopathogen genomics resource: a web-based resource for data-mining plant pathogen genomes. *Database*, *2011*, bar053.

Hannula, S. E., Morriën, E., de Hollander, M., van der Putten, W. H., van Veen, J. A., & de Boer, W. (2017). Shifts in rhizosphere fungal community during secondary succession following abandonment from agriculture. *ISME Journal*, *11*, 2294–2304. https://doi.org/10.1038/ismej.2017.90

Hanson, C. A., Allison, S. D., Bradford, M. A., Wallenstein, M. D., & Treseder, K. K. (2008). Fungal taxa target different carbon sources in forest soil. *Ecosystems*, *11*, 1157–1167. https://doi.org/10.1007/s10021-008-9186-4

Hartmann, M., Howes, C. G., Abarenkov, K., Mohn, W. W., & Nilsson, R. H. (2010). V-Xtractor: An open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16S/18S) ribosomal RNA gene sequences. *Journal of Microbiological Methods*, *83*, 250–253. https://doi.org/10.1016/j.mimet.2010.08.008

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*, 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Hebert, P. D., Braukmann, T. W., Prosser, S. W., Ratnasingham, S., Ivanova, N. V., Janzen, D. H., … Zakharov, E. V. (2017). A Sequel to Sanger: amplicon sequencing that scales. *bioRxiv*, *2017*, 191619.

Heeger, F., Bourne, E. C., Baschien, C., Yurkov, A., Bunk, B., Spröer, C., … Monaghan, M. T. (2018). Long-read DNA metabarcoding of ribosomal rRNA in the analysis of fungi from aquatic environments. *bioRxiv*, *2018*, 283127.

Hiiesalu, I., Bahram, M., & Tedersoo, L. (2017). Plant species richness and productivity determine the diversity of soil fungal guilds in temperate coniferous forest and bog habitats. *Molecular Ecology*, *26*, 4846–4858. https://doi.org/10.1111/mec.14246

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., … Kitano, T. K. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical Chemistry*, *83*, 8604–8610. https://doi.org/10.1021/ac202028g

Ho, T., & Tzanetakis, I. E. (2014). Development of a virus detection and discovery pipeline using next generation sequencing. *Virology*, *471*, 54–60. https://doi.org/10.1016/j.virol.2014.09.019

Hudson, C. M., Kirton, E., Hutchinson, M. I., Redfern, J. L., Simmons, B., Ackerman, E., … Powell, A. J. (2015). Lignin-modifying processes in the rhizosphere of arid land grasses. *Environmental Microbiology*, *17*, 4965–4978. https://doi.org/10.1111/1462-2920.13020

Hunt, D. E., Klepac-Ceraj, V., Acinas, S. G., Gautier, C., Bertilsson, S., & Polz, M. F. (2006). Evaluation of 23S rRNA PCR primers for use in phylogenetic studies of bacterial diversity. *Applied and Environmental Microbiology*, *72*, 2221–2225. https://doi.org/10.1128/AEM.72.3.2221-2225.2006

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., & Mark Welch, D. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, *8*, R143.

Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N., & Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Research*, *21*, 1552–1560. https://doi.org/10.1101/gr.120618.111

Hyde, K. D., Udayanga, D., Manamgoda, D. S., Tedersoo, L., Larsson, E., Abarenkov, K., … Nilsson, R. H. (2013). Incorporating molecular data in fungal systematics: A guide for aspiring researchers. *Current Research in Environmental and Applied Mycology*, *3*, 1–32. https://doi.org/10.5943/cream/3/1/1

Ihrmark, K., Bödeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenk, J., … Lindahl, B. D. (2012). New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology*, *83*, 666–677. https://doi.org/10.1111/j.1574-6941.2012.01437.x

Irinyi, L., Serena, C., Garcia-Hermoso, D., Arabatzis, M., Desnos-Ollivier, M., & Vu, D. (2015). International society of human and animal mycology (ISHAM)-ITS reference DNA barcoding database—The quality controlled standard tool for routine identification of human and animal pathogenic fungi. *Medical Mycology*, *55*, 313–337.

Isola, D., Pardini, M., Varaine, F., Niemann, S., Rüsch-Gerdes, S., Fattorini, L., … Orrù, G. (2005). A Pyrosequencing assay for rapid recognition of SNPs in *Mycobacterium tuberculosis* embB306 region. *Journal of Microbiological Methods*, *62*, 113–120. https://doi.org/10.1016/j.mimet.2005.02.004

James, T. Y., Marino, J. A., Perfecto, I., & Vandermeer, J. (2016). Identification of putative coffee rust mycoparasites via single-molecule DNA sequencing of infected pustules. *Applied and Environmental Microbiology*, *82*, 631–639.

Jumpponen, A., & Jones, K. L. (2009). Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. *New Phytologist*, *184*, 438–448.

Karlsson, I., Edel Hermann, V., Gautheron, N., Durling, M. B., Kolseth, A. K., Steinberg, C., … Friberg, H. (2016). Genus-specific primers for study of *Fusarium* communities in field samples. *Applied and Environmental Microbiology*, *82*, 491–501. https://doi.org/10.1128/AEM.02748-15

Kashyap, P. L., Kumar, S., & Srivastava, A. K. (2017a). Nanodiagnostics for plant pathogens. *Environmental Chemistry Letters*, *15*, 7–13.

Kashyap, P. L., Rai, P., Kumar, S., Chakdar, H., & Srivastava, A. K. (2017b). DNA barcoding for diagnosis and monitoring of fungal plant pathogens. In B. P. Singh, & V. K. Gupta (Eds.), *Molecular markers in mycology* (pp. 87–122). Cham, Switzerland: Springer.

Kembel, S. W., O'Connor, T. K., Arnold, H. K., Hubbell, S. P., Wright, S. J., & Green, J. l. (2014). Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proceedings of the National Academy of Sciences USA*, *111*, 13715–13720. https://doi.org/10.1073/pnas.1216057111

Kemler, M., Garnas, J., Wingfield, M. J., Gryzenhout, M., Pillay, K. A., & Slippers, B. (2013). Ion Torrent PGM as tool for fungal community analysis: A case study of endophytes in *Eucalyptus grandis* reveals high taxonomic diversity. *PLoS ONE*, *8*, e81718.

Kerkhof, L. J., Dillon, K. P., Häggblom, M. M., & McGuinness, L. R. (2017). Profiling bacterial communities by MinION sequencing of ribosomal operons. *Microbiome*, *5*, 116.

Khater, M., de la Escosura-Muñiz, A., & Merkoçi, A. (2017). Biosensors for plant pathogen detection. *Biosensors and Bioelectricity*, *93*, 72–86. https://doi.org/10.1016/j.bios.2016.09.091

Kivlin, S. N., & Hawkes, C. V. (2016). Tree species, spatial heterogeneity, and seasonality drive soil fungal abundance, richness, and composition in Neotropical rainforests. *Environmental Microbiology*, *18*, 4662–4673. https://doi.org/10.1111/1462-2920.13342

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

Knief, C. (2014). Analysis of plant microbe interactions in the era of next generation sequencing technologies. *Frontiers in Plant Science*, *2014*, 5.

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., … Larsson, K.-H. (2013). Towards a unified paradigm for sequence-based identification of Fungi. *Molecular Ecology*, *22*, 5271–5277. https://doi.org/10.1111/mec.12481

Kõljalg, U., Tedersoo, L., Nilsson, R. H., & Abarenkov, K. (2016). Digital identifiers for fungal species. *Science*, *352*, 1182–1183. https://doi.org/10.1126/science.aaf7115

Kopylova, E., Navas-Molina, J. A., Mercier, C., Xu, Z. Z., Mahé, F., He, Y., … Knight, R. (2016). Open-source Sequence Clustering Methods Improve the State of the Art. *mSystems*, *1*, e00003–e15.

Korsakovsky Pond, S., Wadhawan, S., Chairomonte, F., Ananda, G., Chung, W.-Y., Taylor, J., … Team, G. (2009). Windshield splatter analysis with the Galaxy metagenomic pipeline. *Genome Research*, *19*, 2144–2153. https://doi.org/10.1101/gr.094508.109

Kostovcik, M., Bateman, C. C., Kolarik, M., Stelinski, L. L., Jordal, B. H., & Hulcr, J. (2015). The ambrosia symbiosis is specific in some species and promiscuous in others: Evidence from community pyrosequencing. *ISME Journal*, *9*, 126–135. https://doi.org/10.1038/ismej.2014.115

Kozich, J. J., Westcott, S. L., Baxter, K., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, *79*, 5112–5120. https://doi.org/10.1128/AEM.01043-13

Krehenwinkel, H., Pomerantz, A., Henderson, J. B., Kennedy, S. R., Lim, J. Y., Swamy, V., … Prost, S. (2018). Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *bioRxiv*, *2018*, 358572.

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., & Simon, R. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology*, *388*, 1–7. https://doi.org/10.1016/j.virol.2009.03.024

Laehnemann, D., Borkhardt, A., & McHardy, A. C. (2015). Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. *Briefings in Bioinformatics*, *2015*, 1–26.

Lagos, L., Maruyama, F., Nannipieri, P., Mora, M. L., Ogram, A., & Jorquera, M. A. (2015). Current overview on the study of bacteria in the rhizosphere by modern molecular techniques: A mini-review. *Journal of Soil Science and Plant Nutrition*, *15*, 504–523.

Larsson, A. J., Stanley, G., Sinha, R., Weissman, I. L., & Sandberg, R. (2018). Computational correction of index switching in multiplexed sequencing libraries. *Nature Methods*, *15*, 305–307. https://doi.org/10.1038/nmeth.4666

Lasken, R. S., & McLean, J. S. (2014). Recent advances in genomic DNA sequencing of microbial species from single cells. *Nature Reviews in Genetics*, *15*, 577–584. https://doi.org/10.1038/nrg3785

Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R., & Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiology Letters*, *307*, 80–86. https://doi.org/10.1111/j.1574-6968.2010.01965.x

Lee, W. S., Choi, H., Kang, J., Kim, J. H., Lee, S. H., Lee, S., & Hwang, S. Y. (2013). Development of a DNA microarray for species identification of quarantine aphids. *Pest Management Science*, *69*, 1399–1406. https://doi.org/10.1002/ps.3520

Leff, J. W., Lynch, R. C., Kane, N. C., & Fierer, N. (2017). Plant domestication and the assembly of bacterial and fungal communities associated with strains of the common sunflower, *Helianthus annuus*. *New Phytologist*, *214*, 412–423.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., … Schleper, C. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, *442*, 806–809. https://doi.org/10.1038/nature04983

Lenzen, M., Moran, D., Kanemoto, K., Foran, B., Lobefaro, L., & Geschke, A. (2012). International trade drives biodiversity threats in developing nations. *Nature*, *486*, 109–112. https://doi.org/10.1038/nature11145

Levesque, C. A. (2001). Molecular methods for detection of plant pathogens—What is the future? *Canadian Journal of Plant Pathology*, *23*, 333–336.

Li, C., Chng, K. R., Boey, E. J., Ng, A. H., Wilm, A., & Nagarajan, N. (2016). INC-Seq: Accurate single molecule reads using nanopore sequencing. *GigaScience*, *5*, 34. https://doi.org/10.1186/s13742-016-0140-7

Liang, M., Liu, X., Gilbert, G. S., Zheng, Y., Luo, S., Huang, F., & Yu, S. (2016). Adult trees cause density-dependent mortality in conspecific seedlings by regulating the frequency of pathogenic soil fungi. *Ecology Letters*, *19*, 1448–1456. https://doi.org/10.1111/ele.12694

Liebhold, A. M., Brockerhoff, E. G., Nuñez, M. A., Wardle, D., & Wingfield, M. J. (2017). *Biological Invasions*, *19*, 3073–3077.

Lievens, B., Brouwer, M., Vanachter, A. C., Lévesque, C. A., Cammue, B., & Thomma, B. P. (2005). Quantitative assessment of phytopathogenic fungi in various substrates using a DNA macroarray. *Environmental Microbiology*, *7*, 1698–1710. https://doi.org/10.1111/j.1462-2920.2005.00816.x

Lievens, B., Claes, L., Vanachter, A. C., Cammue, B. P., & Thomma, B. P. (2006). Detecting single nucleotide polymorphisms using DNA arrays for plant pathogen diagnosis. *FEMS Microbiology Letters*, *255*, 129–139. https://doi.org/10.1111/j.1574-6968.2005.00074.x

Lindahl, B. D., de Boer, W., & Finlay, R. D. (2010). Disruption of root carbon transport into forest humus stimulates fungal opportunists at the expense of mycorrhizal fungi. *ISME Journal*, *4*, 872–881. https://doi.org/10.1038/ismej.2010.19

Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., … Kauserud, H. (2013). Fungal community analysis by

high-throughput sequencing of amplified markers – a user's guide. *New Phytologist*, *199*, 288–299. https://doi.org/10.1111/nph.12243

Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., & James, S. A. (2010). Population genomics of domestic and wild yeasts. *Nature*, *458*, 337–341. https://doi.org/10.1038/nature07743

Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, *12*, 733–735. https://doi.org/10.1038/nmeth.3444

Luna, R. A., Fasciano, L. R., Jones, S. C., Boyanton, B. L., Ton, T. T., & Versalovic, J. (2007). DNA pyrosequencing-based bacterial pathogen identification in a pediatric hospital setting. *Journal of Clinical Microbiology*, *45*, 2985–2992. https://doi.org/10.1128/JCM.00630-07

Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, *10*, 999–1002. https://doi.org/10.1038/nmeth.2634

Malacrino, A., Rassati, D., Schena, L., Mehzabin, R., Battisti, A., & Palmeri, V. (2017). Fungal communities associated with bark and ambrosia beetles trapped at international harbours. *Fungal Ecology*, *28*, 44–52. https://doi.org/10.1016/j.funeco.2017.04.007

Malacrino, A., Schena, L., Campolo, O., Laudani, F., & Palmeri, V. (2015). Molecular analysis of the fungal microbiome associated with the olive fruit fly *Bactrocera oleae*. *Fungal Ecology*, *18*, 67–74. https://doi.org/10.1016/j.funeco.2015.08.006

Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., … Ouedraogo, J. B. (2012). Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*, *487*, 375–379. https://doi.org/10.1038/nature11174

Maron, J. L., Marler, M., Klironomos, J. N., & Cleveland, C. C. (2011). Soil fungal pathogens and the relationship between plant diversity and productivity. *Ecology Letters*, *14*, 36–41. https://doi.org/10.1111/j.1461-0248.2010.01547.x

Marzano, S.-Y.-L., Nelson, B. D., Ajayi-Oyetunde, O., Bradley, C. A., Hughes, T. J., Hartman, G. L., … Domier, L. L. (2016). Identification of diverse mycoviruses through metatranscriptomics characterization of the viromes of five major fungal plant pathogens. *Journal of Virology*, *90*, 6846–6863. https://doi.org/10.1128/JVI.00357-16

Massart, S., Olmos, A., Jijakli, H., & Candresse, T. (2014). Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Research*, *188*, 90–96. https://doi.org/10.1016/j.virusres.2014.03.029

McCartney, H. A., Foster, S. J., Fraaije, B. A., & Ward, E. (2003). Molecular diagnostics for fungal plant pathogens. *Pest Management Science*, *59*, 129–142.

McConnell, M. J., Lindberg, M. R., Brennand, K. J., Piper, J. C., Voet, T., Cowing-Zitron, C., … Gage, F. H. (2013). Mosaic copy number variation in human neurons. *Science*, *342*, 632–637. https://doi.org/10.1126/science.1243472

McGuire, K. L., Payne, S. G., Palmer, M. I., Gillikin, C. M., Keefe, D., Kim, S. J., … Fierer, N. (2013). Digging the New York skyline: Soil fungal communities in green roofs and city parks. *PLoS ONE*, *8*, e58020.

McKenna, P., Hoffmann, C., Minkah, N., Aye, P. P., Lackner, A., Liu, Z., … Bushman, F. D. (2008). The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS Path*, *4*, e20. https://doi.org/10.1371/journal.ppat.0040020

McLean, J. S., Lombardo, M. J., Ziegler, M. G., Novotny, M., Yee-Greenbaum, J., Badger, J. H., … Hall, A. P. (2013). Genome of the pathogen *Porphyromonas gingivalis* recovered from a biofilm in a hospital sink using a high-throughput single-cell genomics platform. *Genome Research*, *23*, 867–877. https://doi.org/10.1101/gr.150433.112

Menardo, F., Praz, C. R., Wyder, S., Ben-David, R., Bourras, S., & Matsumae, H. (2016). Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. *Nature Genetics*, *48*, 201–205. https://doi.org/10.1038/ng.3485

Mendes, R., Garbeva, P., & Raaijmakers, J. M. (2013). The rhizosphere microbiome: Significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiology Reviews*, *37*, 634–663. https://doi.org/10.1111/1574-6976.12028

Mendes, R., Kruijt, M., de Bruijn, I., Dekkers, E., van der Voort, M., Schneider, J. H., … Raaijmakers, J. M. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science*, *332*, 1097–1100. https://doi.org/10.1126/science.1203980

Mokili, J. L., Rohwer, F., & Dutilh, B. E. (2012). Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology*, *2*, 63–77. https://doi.org/10.1016/j.coviro.2011.12.004

Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews in Microbiology*, *15*, 756.

Mollot, G., Duyck, P. F., Lefeuvre, P., Lescourret, F., Martin, J. F., & Piry, S. (2014). Cover cropping alters the diet of arthropods in a banana plantation: A metabarcoding approach. *PLoS ONE*, *9*, e93740. https://doi.org/10.1371/journal.pone.0093740

Muurinen, J., Stedtfeld, R., Karkman, A., Pärnänen, K., Tiedje, J., & Virta, M. (2017). Influence of manure application on the environmental resistome under Finnish agricultural practice with restricted antibiotic use. *Environmental Science and Technology*, *51*, 5989–5999. https://doi.org/10.1021/acs.est.7b00551

Muyzer, G., Teske, A., Wirsen, C. O., & Jannasch, H. W. (1995). Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Archives of Microbiology*, *164*, 165–172. https://doi.org/10.1007/BF02529967

Nair, S., Nkhoma, S. C., Serre, D., Zimmerman, P. A., Gorena, K., Daniel, B. J., … Cheeseman, I. H. (2014). Single-cell genomics for dissection of complex malaria infections. *Genome Research*, *24*, 1028–1038. https://doi.org/10.1101/gr.168286.113

Nassonova, E., Smirnov, A., Fahrni, J., & Pawlowski, J. (2010). Barcoding Amoebae: Comparison of SSU, ITS and COI genes as tools for molecular identification of naked lobose amoebae. *Protist*, *161*, 102–115. https://doi.org/10.1016/j.protis.2009.07.003

Nguyen, N. H., Smith, D., Peay, K., & Kennedy, P. (2015). Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytologist*, *205*, 1389–1393. https://doi.org/10.1111/nph.12923

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., … Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community data sets by ecological guild. *Fungal Ecology*, *20*, 241–248. https://doi.org/10.1016/j.funeco.2015.06.006

Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., & Tedersoo, L. (2018). Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nature Reviews in Microbiology*. https://doi.org/10.1038/s41579-018-0116-y

Nilsson, R. H., Hyde, K. D., Pawłowska, J., Ryberg, M., Tedersoo, L., Aas, A. B., … Abarenkov, K. (2014). Improving ITS sequence data for identification of plant pathogenic fungi. *Fungal Diversity*, *67*, 11–19. https://doi.org/10.1007/s13225-014-0291-8

Nilsson, R. H., Tedersoo, L., Lindahl, B., Kjoller, R., Carlsen, T., Quince, C., … Kauserud, H. (2011). Towards standardization of the description and publication of next-generation sequencing datasets of fungal communities. *New Phytologist*, *191*, 314–318. https://doi.org/10.1111/j.1469-8137.2011.03755.x

O'Hanlon, S. J., Rieux, A., Farrer, R. A., Rosa, G. M., Waldman, B., Bataille, A., … Martin, M. D. (2018). Recent Asian origin of chytrid fungi causing global amphibian declines. *Science*, *360*, 621–627. https://doi.org/10.1126/science.aar1965

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., … McVeigh, R. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*, D733–D745. https://doi.org/10.1093/nar/gkv1189

Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G. A., Papaniko-laou, N., Kotoulas, G., … Iliopoulos, I. (2015). Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinformatics and Biology Insights*, 9, 75.

Pace, N. R. (1997). A molecular view of microbial diversity and the bio-sphere. *Science*, 276, 734–740. https://doi.org/10.1126/science.276.5313.734

Pacheco da Silva, V. C., Bertin, A., Blin, A., Germain, J.-F., & Bernardi, D. (2014). Molecular and morphological identification of mealybug spe-cies (Hemiptera: Pseudococcidae) in Brazilian vineyards. *PLoS ONE*, 9, e103267.

Palmer, J. M., Jusino, M. A., Banik, M. T., & Lindner, D. L. (2017). Non-bi-ological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data. *PeerJ*, 6, e4925.

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base mat-ters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ-mental Microbiology*, 18, 1403–1414. https://doi.org/10.1111/1462-2920.13023

Park, B., Park, J., Cheong, K.-C., Choi, J., Jung, K., Kim, D., … Kang, S. (2011). Cyber infrastructure for *Fusarium*: Three integrated platforms supporting strain identification, phylogenetics, comparative genomics and knowledge sharing. *Nucleic Acids Research*, 39, D640–D646. https://doi.org/10.1093/nar/gkq1166

Park, J., Park, B., Veeraraghavan, N., Jung, K., Lee, Y. H., Blair, J. E., … Park, S. Y. (2008). Phytophthora database: A forensic database sup-porting the identification and monitoring of *Phytophthora*. *Plant Dis-ease*, 92, 966–972.

Paula, D. P., Linard, B., Crampton-Platt, A., Srivathsan, A., Timmermans, M. J. T. N., Sujii, E. R., & Vogler, A. P. (2016). Uncovering trophic interactions in arthropod predators through DNA shotgun-sequencing of gut contents. *PLoS ONE*, 11, e0161841.

Pawlowski, J., Audic, S., & Adl, S. (2012). CBOL Protist Working Group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biology*, 10, e1001419.

Pearman, W., Smith, A. N., Breckell, G., Dale, J., Freed, N. E., & Silander, O. K. (2018). New tools for diet analyses: nanopore sequencing of metagenomic DNA from stomach contents to quantify diet in an invasive population of rats. *bioRxiv*, 2018, 363622.

Pinto, C., Pinho, D., Sousa, S., Pinheiro, M., Egas, C., & Gomes, A. C. (2014). Unravelling the diversity of grapevine microbiome. *PLoS ONE*, 9, e85622. https://doi.org/10.1371/journal.pone.0085622

Pochon, X., Bott, N. J., Smith, K. F., & Wood, S. A. (2013). Evaluating detection limits of Next-Generation Sequencing for the surveillance and monitoring of international marine pests. *PLoS ONE*, 8, e73935.

Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The madness of microbiome: Attempting to find consensus "best practice" for 16S microbiome studies. *Applied and Environmental Microbiology*, 84, e02627–e2717. https://doi.org/10.1128/AEM.02627-17

Põlme, S., Bahram, M., Jacquemyn, H., Kennedy, P., Kohout, P., Moora, M., … Tedersoo, L. (2018). Host preference and network properties in biotrophic plant–fungal associations. *New Phytologist*, 217(3), 1230–1239. https://doi.org/10.1111/nph.14895

Porazinska, D. L., Giblin-Davis, R. M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., … Thomas, K. (2009). Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources*, 9, 1439–1450. https://doi.org/10.1111/j.1755-0998.2009.02611.x

Porras-Alfaro, A., Liu, K. L., Kuske, C. R., & Xie, G. (2014). From genus to phylum: Large-subunit and internal transcribed spacer rRNA operon regions show similar classification accuracies influenced by database composition. *Applied and Environmental Microbiology*, 80, 829–840. https://doi.org/10.1128/AEM.02894-13

Potter, C., Tang, C. Q., Fonseca, V., Lallias, D., Gaspar, J. M., Thomas, K., & Creer, S. (2017). De novo species delimitation in metabarcoding datasets using ecology and phylogeny. *PeerJ Preprints*, 5, e3121v1. https://doi.org/10.7287/peerj.preprints.3121v1

Prospero, S., & Cleary, M. (2017). Effects of host variability in the spread of invasive forest diseases. *Forests*, 8, 80.

Prosser, J. I. (2010). Replicate or lie. *Environmental Microbiology*, 12, 1806–1810.

Qiu, H., Cai, G., Luo, J., Bhattacharya, D., & Zhang, N. (2016). Extensive horizontal gene transfers between plant pathogenic fungi. *BMC Biol-ogy*, 14, 41.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., … Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database pro-ject: Improved data processing and web-based tools. *Nucleic Acids Research*, 41, D590–D596.

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., … Ouédraogo, N. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530, 228–232. https://doi.org/10.1038/nature16996

Radajewski, S., Ineson, P., Parekh, N. R., & Murrell, J. C. (2000). Stable-isotope probing as a tool in microbial ecology. *Nature*, 403, 646–649. https://doi.org/10.1038/35001054

Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., & MacLean, D. (2010). Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science*, 330, 1540–1543. https://doi.org/10.1126/science.1193070

Rajala, T., Peltoniemi, M., Hantula, J., Mäkipää, R., & Pennanen, T. (2011). RNA reveals a succession of active fungi during the decay of Norway spruce logs. *Fungal Ecology*, 4, 437–448. https://doi.org/10.1016/j.funeco.2011.05.005

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (https://www.barcodinglife.org). *Molecular Ecology Resources*, 7, 355–364.

Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-based registry for all animal species: The Barcode Index Number (BIN) system. *PLoS ONE*, 8, e66213.

Restrepo, S., Tabima, J. F., Mideros, M. F., Grünwald, N. J., & Matute, D. R. (2014). Speciation in fungal and oomycete plant pathogens. *Annual Reviews in Phytopathology*, 52, 289–316. https://doi.org/10.1146/annurev-phyto-102313-050056

Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-throughput sequencing technologies. *Molecular Cell*, 58, 586–597. https://doi.org/10.1016/j.molcel.2015.05.004

Rezki, S., Campion, C., Iacomi-Vasilescu, B., Preveaux, A., Toualbia, Y., & Bonneau, S. (2016). Differences in stability of seed-associated micro-bial assemblages in response to invasion by phytopathogenic microorganisms. *Peer Journal*, 4, e1923.

Rhodes, J., Abdolrasouli, A., Farrer, R. A., Cuomo, C. A., Aanensen, D. M., Armstrong-James, D., … Schelenz, S. (2018). Genomic epidemiology of the UK outbreak of the emerging human fungal pathogen *Candida auris*. *Emerging Microbial Infections*, 7, 43.

Riit, T., Tedersoo, L., Drenkhan, R., Runno-Paurson, E., Kokko, H., & Anslan, S. (2016). Oomycete-specific ITS primers for identification and metabarcoding. *MycoKeys*, 14, 17–30. https://doi.org/10.3897/mycokeys.14.9244

Rissanen, A. J., Kurhela, E., Aho, T., Oittinen, T., & Tiirola, M. (2010). Storage of environmental samples for guaranteeing nucleic acid yields for molecular microbiological studies. *Applied Microbiology and Biotechnology*, 88, 977–984. https://doi.org/10.1007/s00253-010-2838-2

Roesch, L., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K. M., Kent, A. D., … Triplett, E. W. (2007). Pyrosequencing enumerates and con-trasts soil microbial diversity. *ISME Journal*, 1, 283–290. https://doi.org/10.1038/ismej.2007.53

Roossinck, M. J., Martin, D. P., & Roumagnac, P. (2015). Plant virus metagenomics: Advances in virus discovery. *Phytopathology*, 105, 716–727. https://doi.org/10.1094/PHYTO-12-14-0356-RVW

Roossinck, M. J., Saha, P., Wiley, G. B., Quan, J., White, J. D., Lai, H., … Roe, B. A. (2010). Ecogenomics: Using massively parallel pyrosequencing to understand virus ecology. *Molecular Ecology*, 19, 81–88. https://doi.org/10.1111/j.1365-294X.2009.04470.x

Ruegger, P. M., Clark, R. T., Weger, J. R., Braun, J., & Borneman, J. (2014). Improved resolution of bacteria by high throughput sequence analysis of the rRNA internal transcribed spacer. *Journal of Microbiological Methods*, 105, 82–87. https://doi.org/10.1016/j.mimet.2014.07.001

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, 72, 1–13. https://doi.org/10.1016/j.compag.2010.02.007

Santini, A., Liebhold, A., Migliorini, D., & Woodward, S. (2018). Tracing the role of human civilization in the globalization of plant pathogens. *ISME Journal*, 12, 647–652. https://doi.org/10.1038/s41396-017-0013-9

Sanzani, S. M., Nicosia, L. D., & M.G., Faedda, R., Cacciola, S.O., & Schena, L., (2014). Use of quantitative PCR detection methods to study biocontrol agents and phytopathogenic fungi and oomycetes in environmental samples. *Journal of Phytopathology*, 162, 1–13. https://doi.org/10.1111/jph.12147

Sapp, M., Lewis, E., Moss, S., Barrett, B., Kirk, S., Elphinstone, J. G., & Denman, S. (2016). Metabarcoding of bacteria associated with the acute oak decline syndrome in England. *Forests*, 7, 95.

Savary, R., Masclaux, F. G., Wyss, T., Droh, G., Corella, J. C., Machado, A. P., … Sanders, I. R. (2018). A population genomics approach shows widespread geographical distribution of cryptic genomic forms of the symbiotic fungus *Rhizophagus irregularis*. *ISME Journal*, 12, 17–30. https://doi.org/10.1038/ismej.2017.153

Schaad, N. W., Jones, J. B., & Chun, W. (2001). *Laboratory guide for the identification of plant pathogenic bacteria*. St. Paul, MI: APS Press.

Schena, L., Nigro, F., Ippolito, A., & Gallitelli, D. (2004). Real-time quantitative PCR: A new technology to detect and study phytopathogenic and antagonistic fungi. *European Journal of Plant Pathology*, 110, 893–908. https://doi.org/10.1007/s10658-004-4842-9

Schlaeppi, K., Bender, S. F., Mascher, F., Russo, G., Patrignani, A., Camenzind, T., … van der Heijden, M. G. A. (2016). High-resolution community profiling of arbuscular mycorrhizal fungi. *New Phytologist*, 212, 780–791. https://doi.org/10.1111/nph.14070

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, 15, 1289–1303. https://doi.org/10.1111/1755-0998.12402

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., … Fungal Barcoding Consortium. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences USA*, 109, 6241–6246. https://doi.org/10.1073/pnas.1117018109

Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., … Reyer, C. P. O. (2017). Forest disturbances under climate change. *Nature Climate Change*, 7, 395–402. https://doi.org/10.1038/nclimate3303

Sessitsch, A., Hackl, E., Wenzl, P., Kilian, A., Kostic, T., Stralis-Pavese, N., … Bodrossy, L. (2006). Diagnostic microbial microarrays in soil ecology. *New Phytologist*, 171, 719–736. https://doi.org/10.1111/j.1469-8137.2006.01824.x

Seyedmousavi, S., Bosco, S. D., De Hoog, S., Ebel, F., Elad, D., Gomes, R. R., … Guillot, J. (2018). Fungal infections in animals: A patchwork of different situations. *Medical Mycology*, 56, S165–S187.

Shakya, M., Quince, C., Campbell, J. H., Yang, Z. K., Schadt, C. W., & Podar, M. (2013). Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities. *Environmental Microbiology*, 15, 1882–1899. https://doi.org/10.1111/1462-2920.12086

Shokralla, S., Gibson, J. F., Nikbakht, H., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2014). Next-generation DNA barcoding to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources*, 14, 892–901.

Shokralla, S., Porter, T. M., Gibson, J. F., Dobosz, R., Janzen, D. H., Hallwachs, W., … Hajibabaei, M. (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports*, 5, 9687. https://doi.org/10.1038/srep09687

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., … Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences USA*, 103, 12115–12120. https://doi.org/10.1073/pnas.0605127103

Song, Z., Schlatter, D., Kennedy, P., Kinkel, L. L., Kistler, H. C., Nguyen, N., & Bates, S. T. (2015). Effort versus reward: Preparing samples for fungal community characterization in high-throughput sequencing surveys of soils. *PLoS ONE*, 10, e0127234. https://doi.org/10.1371/journal.pone.0127234

Sønstebø, J. H., Vivian-Smith, A., Adamson, K., Drenkhan, R., Solheim, H., & Hietala, A. (2017). Genome-wide population diversity in Hymenoscyphus fraxineus points to an eastern Russian origin of European Ash dieback. *bioRxiv*, 2017, 154492.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H.-W., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, 19, S21–S31. https://doi.org/10.1111/j.1365-294X.2009.04480.x

Sutarlie, L., Ow, S. Y., & Su, X. (2017). Nanomaterials-based biosensors for detection of microorganisms and microbial toxins. *Biotechnology Journal*, 12, 4.

Szemes, M., Bonants, P., de Weerdt, M., Baner, J., Landegren, U., & Schoen, C. D. (2005). Diagnostic application of padlock probes—multiplex detection of plant pathogens using universal microarrays. *Nucleic Acids Research*, 33, e70.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21, 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Talas, F., & McDonald, B. A. (2015). Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. *BMC Genomics*, 16, 996. https://doi.org/10.1186/s12864-015-2166-0

Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., … Abarenkov, K. (2015). Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys*, 10, 1–43.

Tedersoo, L., Bahram, M., Cajthaml, T., Põlme, S., Hiiesalu, I., Anslan, S., … Abarenkov, K. (2016). Tree diversity and species identity effects on soil fungi, protists and animals are context-dependent. *ISME Journal*, 10, 346–362.

Tedersoo, L., Bahram, M., Põlme, S., Anslan, S., Riit, T., Kõljalg, U., … Abarenkov, K. (2015). Response to comment on "Global diversity and geography of soil fungi": Analytical biases in microbial diversity studies. *Science*, 359, 936.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., … Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science*, 346, 1078. https://doi.org/10.1126/science.1256688

Tedersoo, L., Jairus, T., Horton, B. M., Abarenkov, K., Suvi, T., Saar, I., & Kõljalg, U. (2008). Strong host preference of ectomycorrhizal fungi in a Tasmanian wet sclerophyll forest as revealed by DNA barcoding and taxon-specific primers. *New Phytologist*, 180, 479–490. https://doi.org/10.1111/j.1469-8137.2008.02561.x

Tedersoo, L., Liiv, I., Kivistik, P. A., Anslan, S., Kõljalg, U., & Bahram, M. (2016). Genomics and metagenomics technologies to recover

ribosomal DNA and single-copy genes from old fruitbody and ecto-mycorrhiza specimens. *MycoKeys*, 13, 1–20.

Tedersoo, L., & Lindahl, B. (2016). Fungal identification biases in microbiome projects. *Environmental Microbiology Reports*, 8, 774–779. https://doi.org/10.1111/1758-2229.12438

Tedersoo, L., & Nilsson, R. H. (2016). In In: Martin, F. (Ed.). Molecular Mycorrhizal Symbiosis. (Ed.), (pp. 301–322).

Tedersoo, L., Nilsson, R. H., Abarenkov, K., Jairus, T., Sadam, A., Saar, I., … Kõljalg, U. (2010). 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist*, 188, 291–301. https://doi.org/10.1111/j.1469-8137.2010.03373.x

Tedersoo, L., Tooming-Klunderud, A., & Anslan, S. (2017). PacBio metabarcoding of fungi and other eukaryotes: Biases and perspectives. *New Phytologist*, 217, 1370–1385.

ten Hoopen, P. T., Finn, R. D., Bongo, L. A., Corre, E., Fosso, B., Meyer, F., … Willassen, N. P. (2017). The metagenomic data life-cycle: Standards and best practices. *GigaScience*, 6, 1–11. https://doi.org/10.1093/gigascience/gix047

Terhonen, E., Sun, H., Buee, M., Kasanen, R., Paulin, L., & Asiegbu, F. O. (2013). Effects of the use of biocontrol agent (*Phlebiopsis gigantea*) on fungal communities on the surface of *Picea abies* stumps. *Forest Ecology and Management*, 310, 428–433. https://doi.org/10.1016/j.foreco.2013.08.044

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics-a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2, 3.

Tkacz, A., Hortala, M., & Poole, S. (2018). Absolute quantitation of microbiota abundance in environmental samples. *Microbiome*, 6, 110.

Toju, H., Tanabe, A. S., Yamamoto, S., & Sato, H. (2012). High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS ONE*, 7, e40863.

Tuda, M., Kagoshima, K., Toquenaga, Y., & Arnqvist, G. (2014). Global genetic differentiation in a cosmopolitan pest of stored beans: Effects of geography, host-plant usage and anthropogenic factors. *PLoS ONE*, 9, e106268.

Uroz, S., Buee, M., Deveau, A., Mieszkin, S., & Martin, F. (2016). Ecology of the forest microbiome: Highlights of temperate and boreal ecosystems. *Soil Biology and Biochemistry*, 103, 471–488.

Valentini, A., Miquel, C., Nawaz, M. A., Bellemain, E. V., Coissac, E., Pompanon, F., … Swenson, J. E. (2009). New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: The trnL approach. *Molecular Ecology Resources*, 9, 51–60.

Vestheim, H., & Jarman, S. N. (2008). Blocking primers to enhance PCR amplification of rare sequences in mixed samples - a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, 5, 12.

Vettraino, A. M., Bonants, P., Tomassini, A., Bruni, N., & Vannini, A. (2012). Pyrosequencing as a tool for the detection of *Phytophthora* species: Error rate and risk of false Molecular Operational Taxonomic Units. *Letters in Applied Microbiology*, 55, 390–396.

Vettraino, A., Roques, A., Yart, A., Fan, J. T., Sun, J. H., & Vannini, A. (2015). Sentinel trees as a tool to forecast invasions of alien plant pathogens. *PLoS ONE*, 10, e0120571.

Wachi, N., Matsubayashi, K. W., & Maeto, K. (2018). Application of next-generation sequencing to the study of non-model insects. *Entomological Science*, 21, 3–11. https://doi.org/10.1111/ens.12281

Wagner, J., Coupland, P., Browne, H. P., Lawley, T. D., Francis, S. C., & Parkhill, J. (2016). Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification. *BMC Microbiology*, 16, 274.

Walder, F., Schlaeppi, K., Wittwer, R., Held, A. Y., Vogelgsang, S., & van der Heijden, M. G. (2017). Community profiling of *Fusarium* in combination with other plant-associated fungi in different crop species using SMRT sequencing. *Frontiers in Plant Science*, 2017, 8.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wang, Q., Jia, P., & Zhao, Z. (2013). VirusFinder: Software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. *PLoS ONE*, 8, e64465. https://doi.org/10.1371/journal.pone.0064465

Wang, J., Long, J., Liu, Z., Wu, W., & Hu, C. (2017). Label-free and high-throughput biosensing of multiple tumor markers on a single light-addressable photoelectrochemical sensor. *Biosensors and Bioelectronics*, 91, 53–59. https://doi.org/10.1016/j.bios.2016.12.029

Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., & Rohwer, F. (2007). Metagenomic analysis of the microbial community associated with the coral Porites astreoides. *Environmental Microbiology*, 9, 2707–2719. https://doi.org/10.1111/j.1462-2920.2007.01383.x

White, T. J., Bruns, T. D., Lee, S. B., & Taylor, J. W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T. J. White (Eds.), *PCR protocols: A guide to methods and applications* (pp. 315–322). San Diego, CA: Academic Press.

Wilson, K. H., Wilson, W. J., Radosevich, J. L., DeSantis, T. Z., Viswanathan, V. S., Kuczmarski, T. A., & Andersen, G. L. (2002). High-density microarray of small-subunit ribosomal DNA probes. *Applied and Environmental Microbiology*, 68, 2535–2541. https://doi.org/10.1128/AEM.68.5.2535-2541.2002

Wylie, S. J., Luo, H., Li, H., & Jones, M. G. (2012). Multiple polyadenylated RNA viruses detected in pooled cultivated and wild plant samples. *Archives of Virology*, 157, 271–284. https://doi.org/10.1007/s00705-011-1166-x

Yan, D., Mills, J. G., Gellie, N. J., Bissett, A., Lowe, A. J., & Breed, M. F. (2018). High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. *Biological Conservation*, 217, 113–120. https://doi.org/10.1016/j.biocon.2017.10.035

Yang, B., Cai, J., & Cheng, X. (2011). Identification of astigmatid mites using ITS2 and COI regions. *Parasitology Research*, 108, 497–503. https://doi.org/10.1007/s00436-010-2153-y

Yarza, P., Yilmaz, P., Panzer, K., Glöckner, O., & Reich, M. (2017). A phylogenetic framework for the kingdom Fungi based on 18S rRNA gene sequences. *Marine Genomics*, 36, 33–39. https://doi.org/10.1016/j.margen.2017.05.009

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., … Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews in Microbiology*, 12, 635–645. https://doi.org/10.1038/nrmicro3330

Yoshida, K., Burbano, H. A., Krause, J., Thines, M., & Weigel, D. (2014). Mining herbaria for plant pathogen genomes: Back to the future. *PLoS Path*, 10, e1004028.

Zhang, T., Breitbart, M., Lee, W. H., Run, J., & Wei, C. L. (2005). RNA viral community in human feces: Prevalence of plant pathogenic viruses. *PLoS Biology*, 4, e3.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., … Huang, Q. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, 2, 4.

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., DeBarba, M., … Chave, J. (2017). Soil community assembly varies across body sizes in a tropical forest. *bioRxiv*, 2017, 154278.